ETH zürich

Empa
Materials Science and Technology

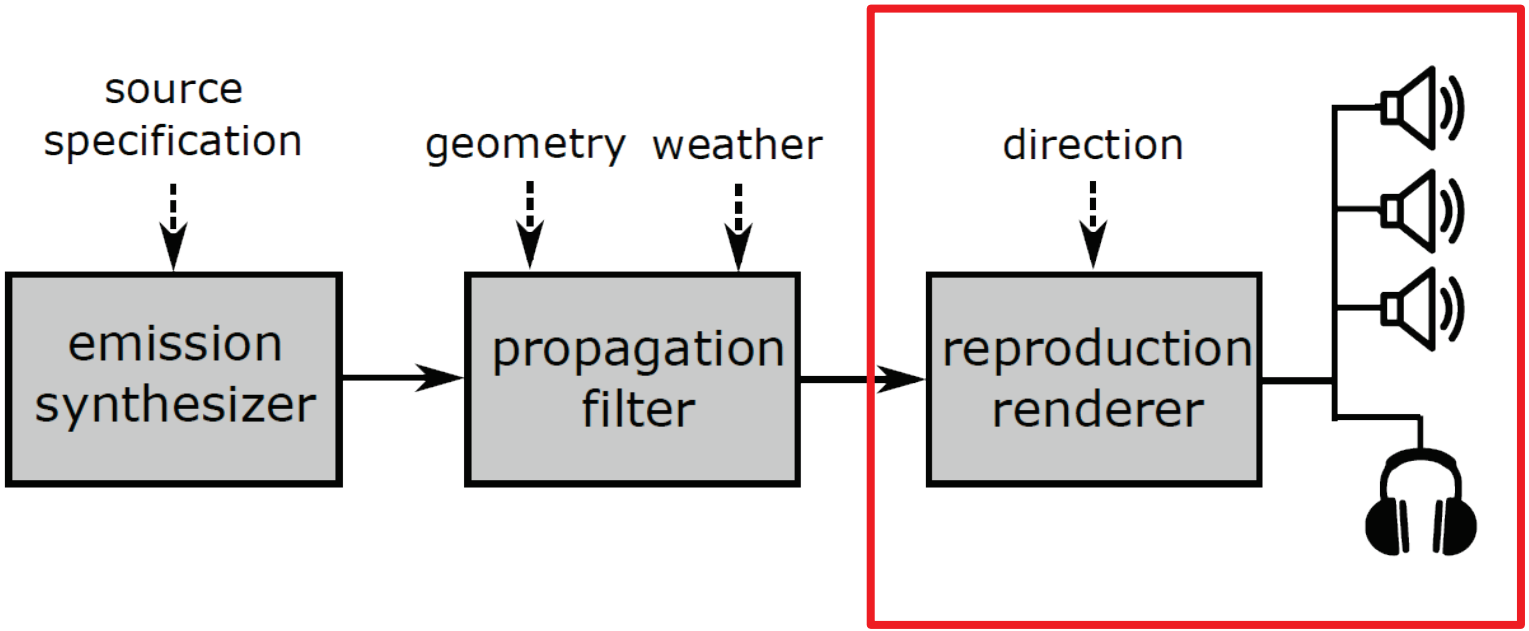
Acoustics II

Auralization: Sound reproduction

Reto Pieren

2024

Flexible auralization approach



Sound reproduction techniques

Sound reproduction: Overview

- Transducer type: Headphones vs. Loudspeakers
- Speaker feed creation
  a) Direct recording with suitable microphone setup, e.g. ORTF for stereo
  b) Reproduction rendering (calculation)

  ↓ increasing flexibility
- Audio formats
  a) Channel-based (stereo, 5.1, binaural)
  b) Scene-based (Ambisonics, HOA)
  c) Object-based (3D audio e.g. Dolby Atmos)

  ↓ increasing flexibility

[ITU 2019. Recommendation ITU-R BS.2076-2 – Audio definition model, International Telecommunication Union (ITU), Geneva, Switzerland.]

ETH zürich
Empa
Materials Science and Technology

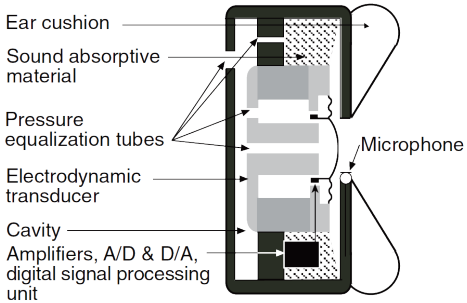
Headphone reproduction

Headphone reproduction

- Transducers directly in front of the outer ear or in ear canal
- No left/right cross-talk
- Electrodynamic and electrostatic transducers
- Fit: Circumaural, supra-aural or ear-fitting

Headphone reproduction

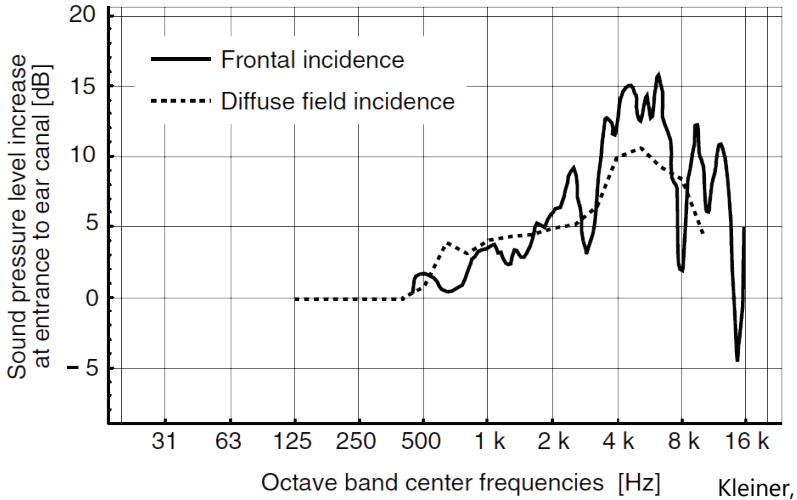
- Acoustical design → sound insulation
  - Closed-back or open-back (perforated outer shell)
  - Passive or noise cancelling
    → active noise control with microphone



[Kleiner, M. 2012. Acoustics and Audio Technology, 3rd edition, J.Ross Publishing]

Headphone reproduction

- Head related transfer function (HRTF) missing
  → equalization necessary (non-flat frequency response)
- HRTF depends on angle of incidence. What reference for frequency response?



Kleiner, 2012

- 2 types: Free-field and diffuse-field equalized

Binaural reproduction

1. Recording with head-and-torso simulator (HATS)

2. Binaural technology: HRTF filtering (convolution with head-related impulse responses)

ETH zürich
Empa
Materials Science and Technology

Binaural rendering

- Challenges:
  - Personalization of HRTFs, e.g. based on anthropometric data (head size, ear shape) → estimation from photos?
  - Interpolation of HRTFs between measured angles
  - Head tracking to dynamically adjust angles of incidence → time-variant HRTF filtering

Binaural rendering: Listening examples with different HRTFs

- Exercise:
  - Which trajectory is the source following?
  - Which measured HRTF best fits my own?

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

ETH zürich
Empa
Materials Science and Technology

Sound reproduction: Loudspeakers vs. headphones

- Pros and cons:

		
Portability	–	+
Room influence	–	+
Disturbing noise	–	+
Intrusion	+	–
Calibration	+	–
Subject's influence	+	–

R. Pieren

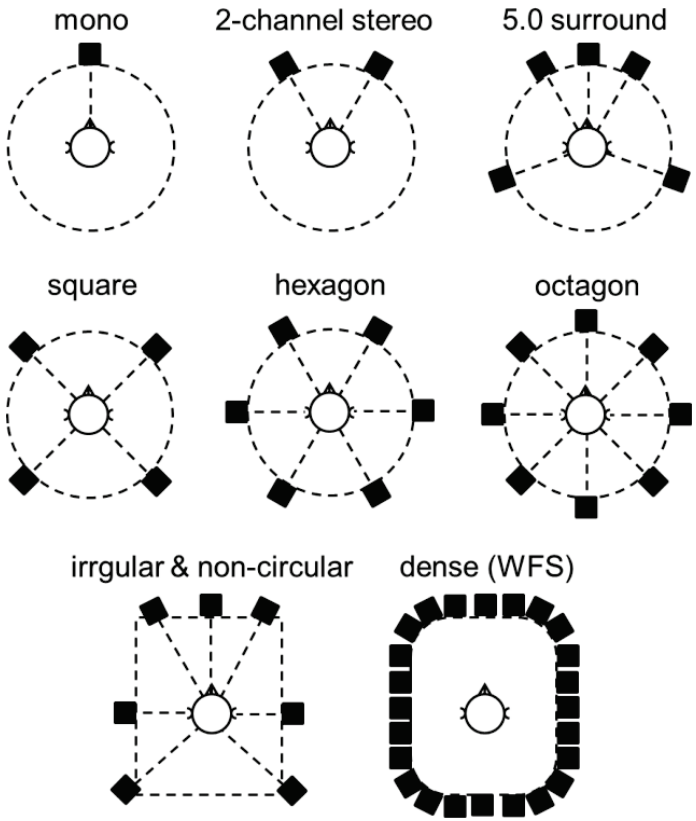
ETH zürich
Empa
Materials Science and Technology

Multichannel loudspeaker reproduction

Multichannel loudspeaker reproduction

Various:

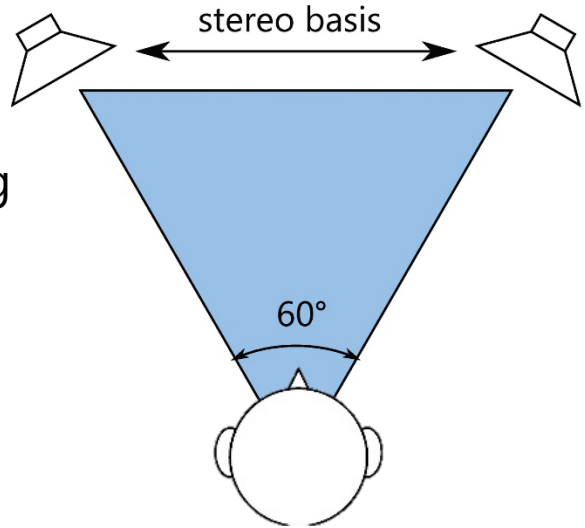
- Speaker arrangements
- Audio formats
- Rendering strategies



2-channel stereophony

2-channel stereophonic reproduction

- Stereo loudspeaker arrangement:
  - equilateral triangle: loudspeaker pair and listening position
  - Stereo basis (base width) = listening distance, preferred 2-3 m



[ITU 2014. Recommendation ITU-R BS.1116-2 - Methods for the subjective assessment of small impairments in audio systems, International Telecommunication Union (ITU), Geneva, Switzerland.]

ETH zürich
Empa
Materials Science and Technology

2-channel stereophonic reproduction

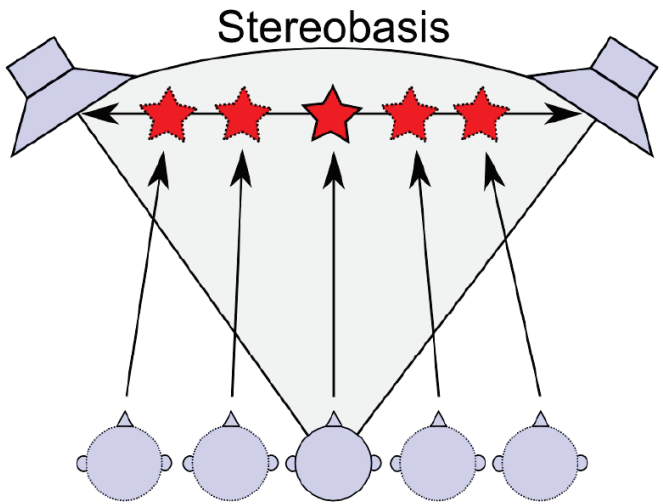
- Cross-talk right speaker → left ear and vice versa
- Duplex theory by Rayleigh: Localization by level differences and time of arrival differences
- Creation of **phantom sources**: Illusion of sources on the stereo basis (same signal to both speakers → localization to the front)

2-channel stereophonic reproduction

- Reason for geometry: tradeoff between front localization and cross-talk
- Listening room requirements: no disturbing reflections, no disturbing room modes

2-channel stereophony: Localization error

- Listening experiment regarding localization error due to asymmetrical listening position



clicks at center

2-channel stereophony: Localization error

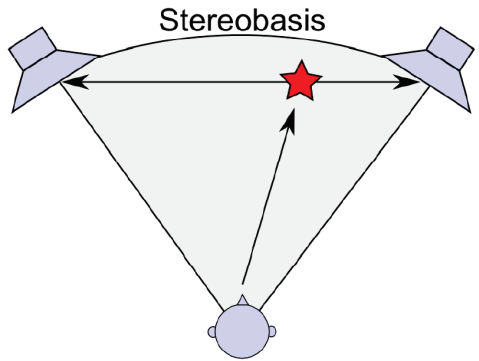
- Localization error due to asymmetrical listening position
  - High sensitivity
  - Example:
    - stereo basis = listening distance = 3 m
    - lateral offset from line of symmetry: +/- 10 cm

    → localization error: 50 cm

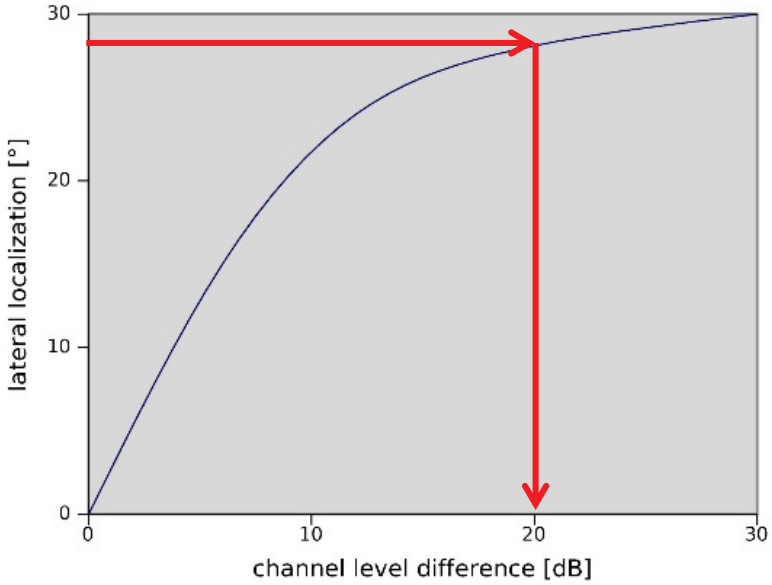
2-channel stereophony: Localization by level information

- Listening experiment: Source localization for different level differences left/right

0 dB	6 dB	12 dB	18 dB	24 dB	30 dB

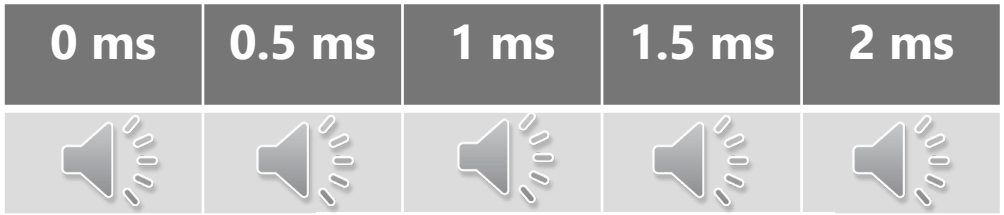


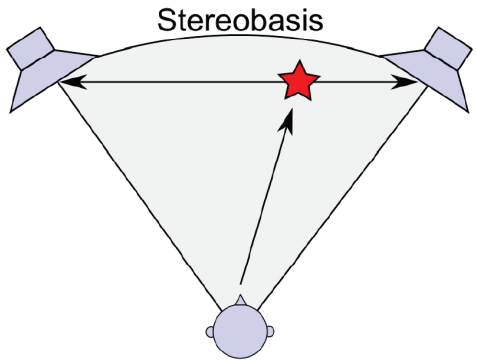
2-channel stereophony: Localization by level information



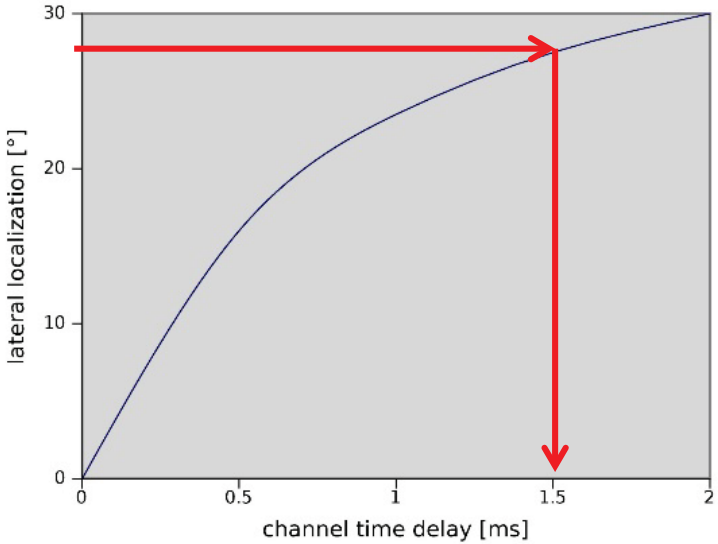
2-channel stereophony: Localization by time information

- Listening experiment: Source localization due to time of arrival differences



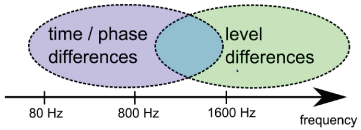


2-channel stereophony: Localization by time information



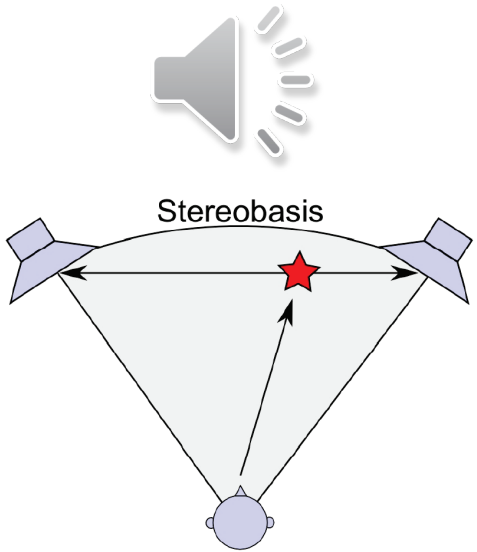
- $c \cdot 1.5$ ms = 50 cm path length difference ≫ ear distance

  → Cross-talk and localization by level difference

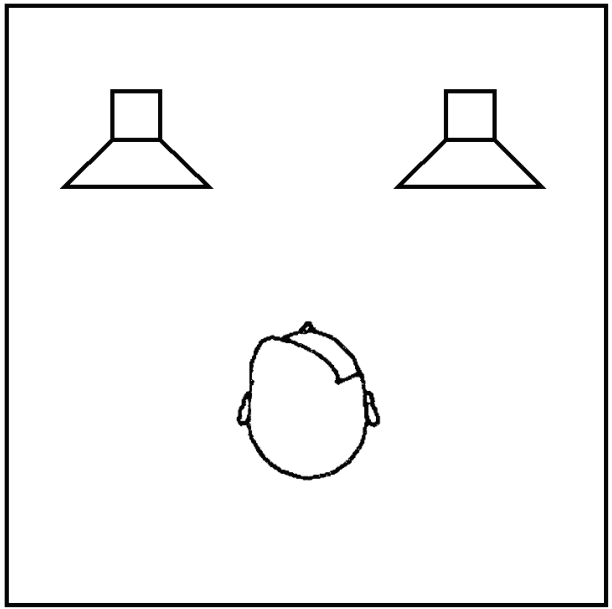


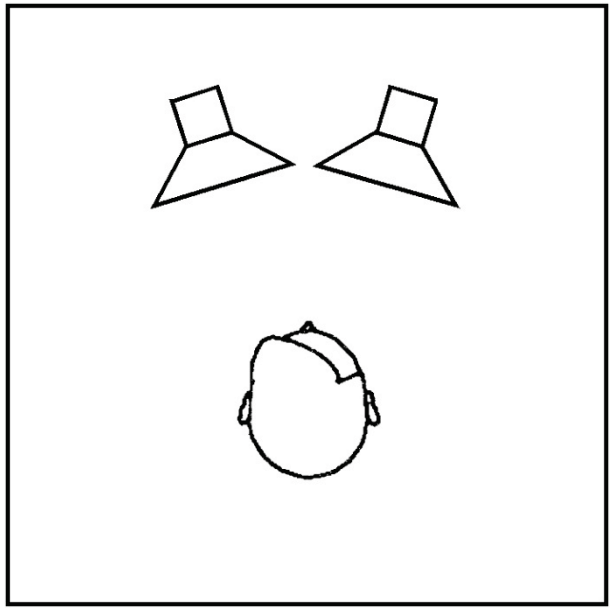
2-channel stereophony: Diverging level and time information

- Listening experiment: Source localization with opposing level and time of arrival information
  - 6 dB and 0.5 ms

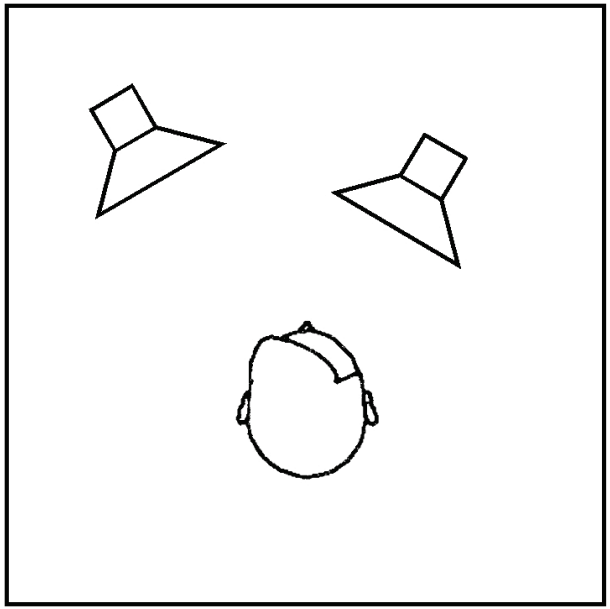


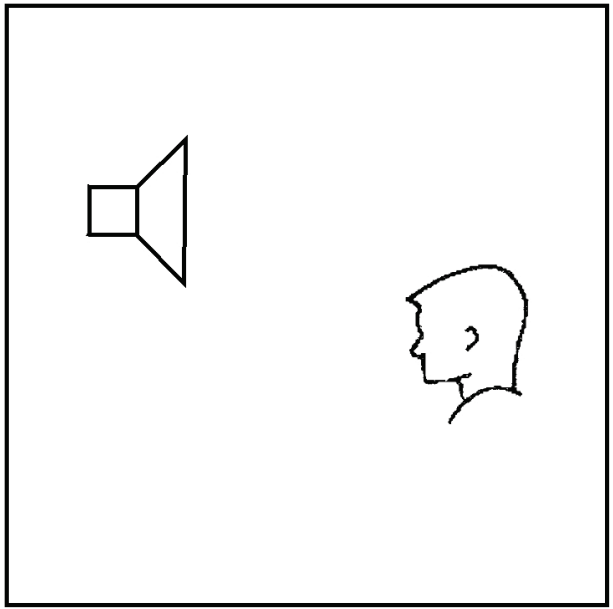
2-channel stereophony: Mounting errors



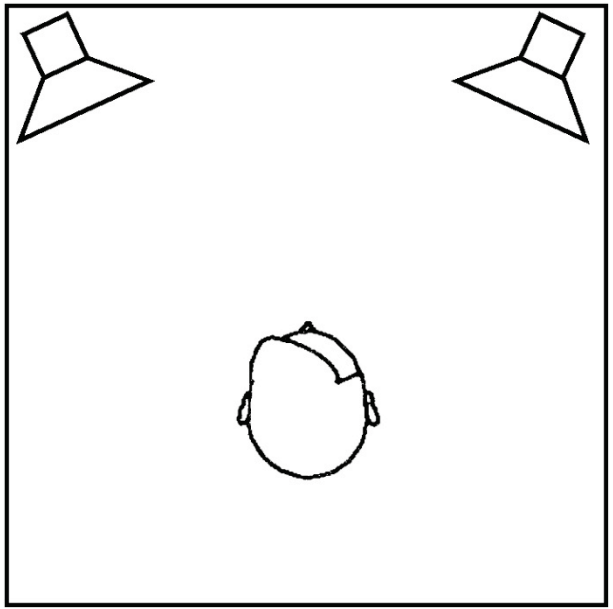


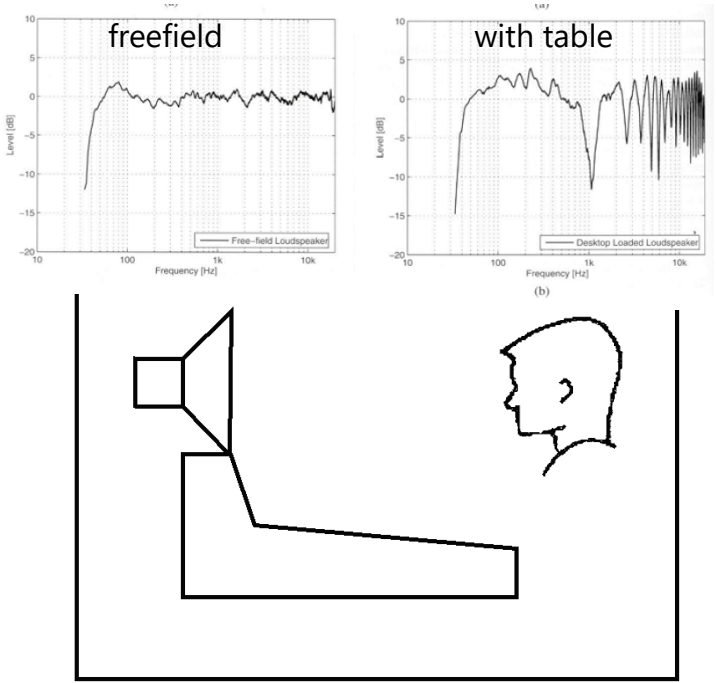
2-channel stereophony: Mounting errors





2-channel stereophony: Mounting errors





Stereo reproduction rendering

Stereo reproduction rendering

- Creation of left and right speaker feed from an object-based audio format:

$$\begin{pmatrix} s_0(t), & \varphi_0\,(t) \\ \vdots & \vdots \\ s_N(t), & \varphi_N\,(t) \end{pmatrix}$$

  - audio signal $s(t)$
  - corresponding azimuth angle of incidence $\varphi(t)$ [=0° frontal]
- Methods
  a) Pair-wise amplitude panning
  b) Virtual microphones
  c) Transaural stereo = Cross-Talk Cancellation (CTC)

ETH zürich
Empa
Materials Science and Technology

Pair-wise amplitude panning

- Panning
  - «creating a **pan**orama»
  - distribution of signal to multiple channels. Analog processing using a panning potentiometer (pan pot)
- Duplicating and amplitude weighting of signals depending on angle:

$$y_{\mathrm{L}}(t) = g_{\mathrm{L}}\big(\varphi(t)\big) \cdot s(t)$$
$$y_{\mathrm{R}}(t) = g_{\mathrm{R}}\big(\varphi(t)\big) \cdot s(t)$$

→ Localization by level differences

ETH zürich
Empa
Materials Science and Technology

Pair-wise amplitude panning

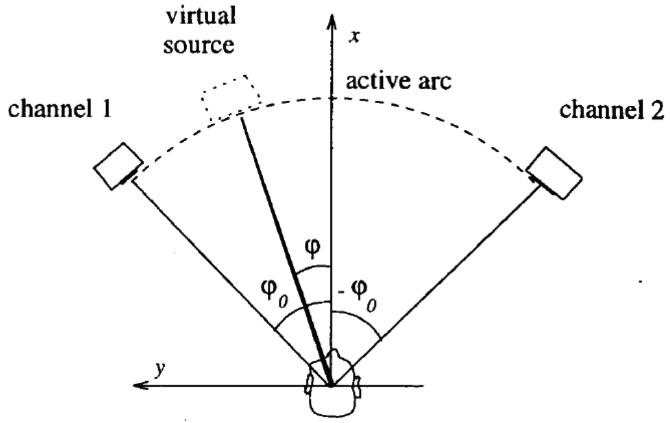
- Usually frequency-independent
- Different panning laws in use to calculate the gains g_{L} and g_{R} depending on
  - Directional perception model
  - Normalization

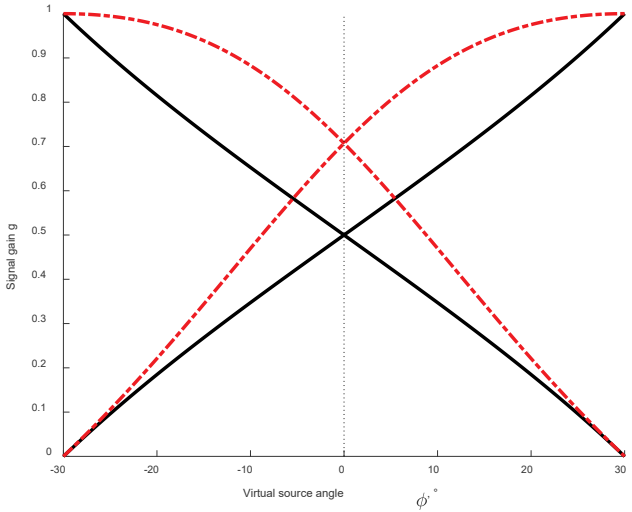
Pair-wise amplitude panning

- Tangent law:

$$\frac{\tan\varphi}{\tan\varphi_0} = \frac{g_\mathrm{L} - g_\mathrm{R}}{g_\mathrm{L} + g_\mathrm{R}}$$

- Normalization: $g_\mathrm{L}^P + g_\mathrm{R}^P = 1$
  - P=1: coherent summation of speaker
  - P=2: incoherent summation of speaker
  - Which one is right?





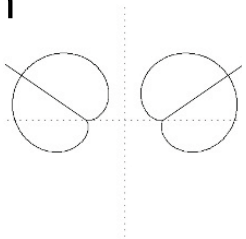
Virtual microphones

- Mimik microphone setup as used in sound recording
  (→ Acoustics II: Recording techniques)
- Model mic directivities, orientation and location
- e.g. ORTF:

$$y_{\mathrm{L}}(t) = \frac{1 + \cos(\varphi(t) - 55°)}{2} s\big(t + \Delta t(t)\big)$$

$$y_{\mathrm{R}}(t) = \frac{1 + \cos(\varphi(t) + 55°)}{2} s(t)$$

$$\Delta t(t) = \frac{0.17 \sin \varphi(t)}{c_0}$$



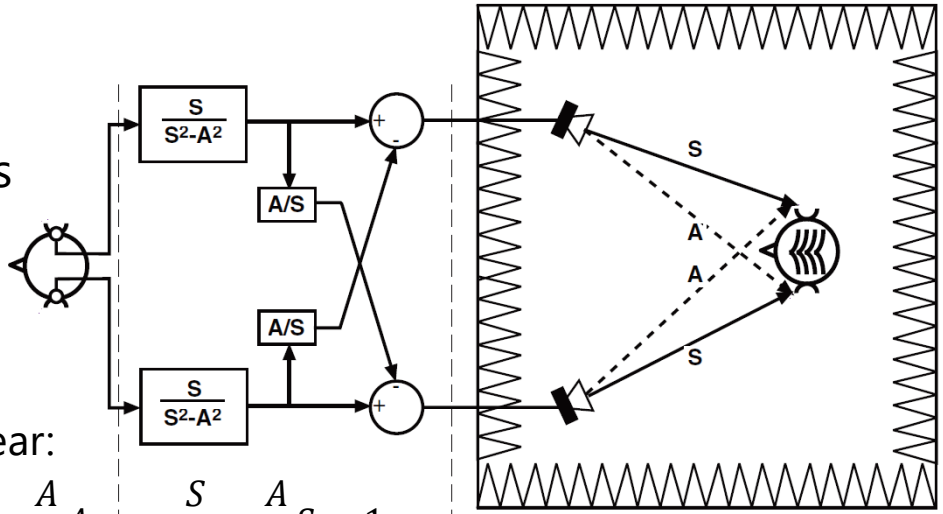
Cardioid pattern
Relative delay

→ XY?
Δt=0 and 55°→65°.

Transaural stereo = Cross-talk cancellation (binaural via loudspeakers)

- Binaural signals → loudspeakers → surround impression
- Inverse filters to pre-compensate cross-talks
- Difficulties:
  - Estimate paths
  - Head movements

  → tracking

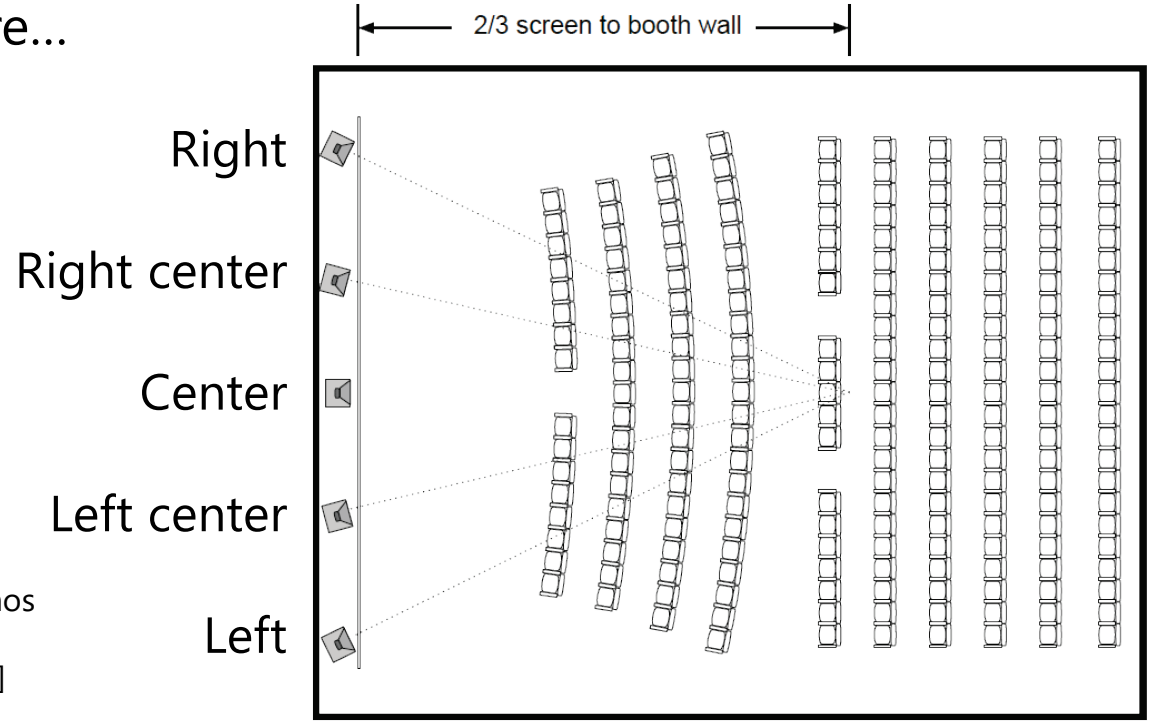


Sum of 4 signal paths per ear:

$$\frac{S}{S^2-A^2}S+\frac{S}{S^2-A^2}A-\frac{S}{S^2-A^2}\frac{A}{S}A-\frac{S}{S^2-A^2}\frac{A}{S}S=1$$

Stabilization of phantom sources

- Adding center speaker → 3-channel stereo
- And more…

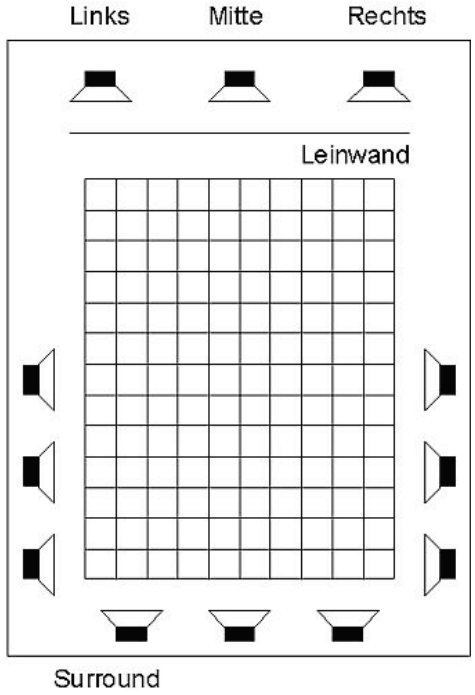


[Dolby 2015. Dolby Atmos Specifications – Issue 3, Dolby Laboratories, Inc.]

Surround

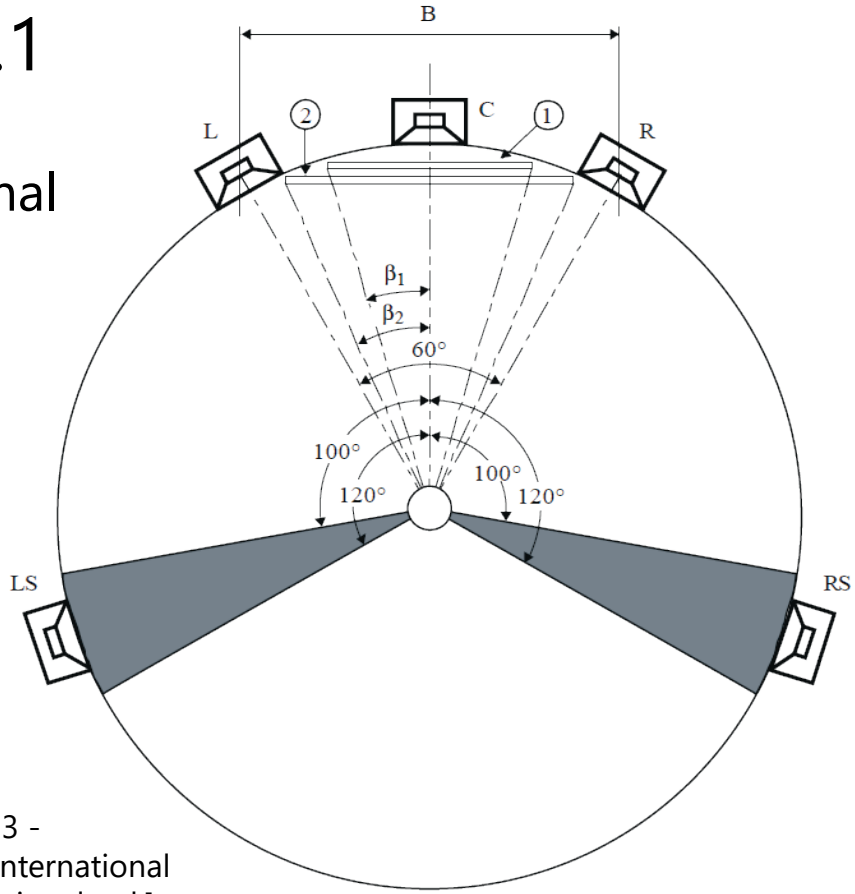
Surround sound 5.1

- 1992: launch of digital 5.1 Surround sound for movie theaters
- 1997: Dolby Digital 5.1 on DVD for consumer applications
- 6 discrete channels
  - Left, Center, Right
  - Surround Left, Surround Right
  - LFE = Low Frequency Effects (up to 120 Hz)



Surround sound 5.1

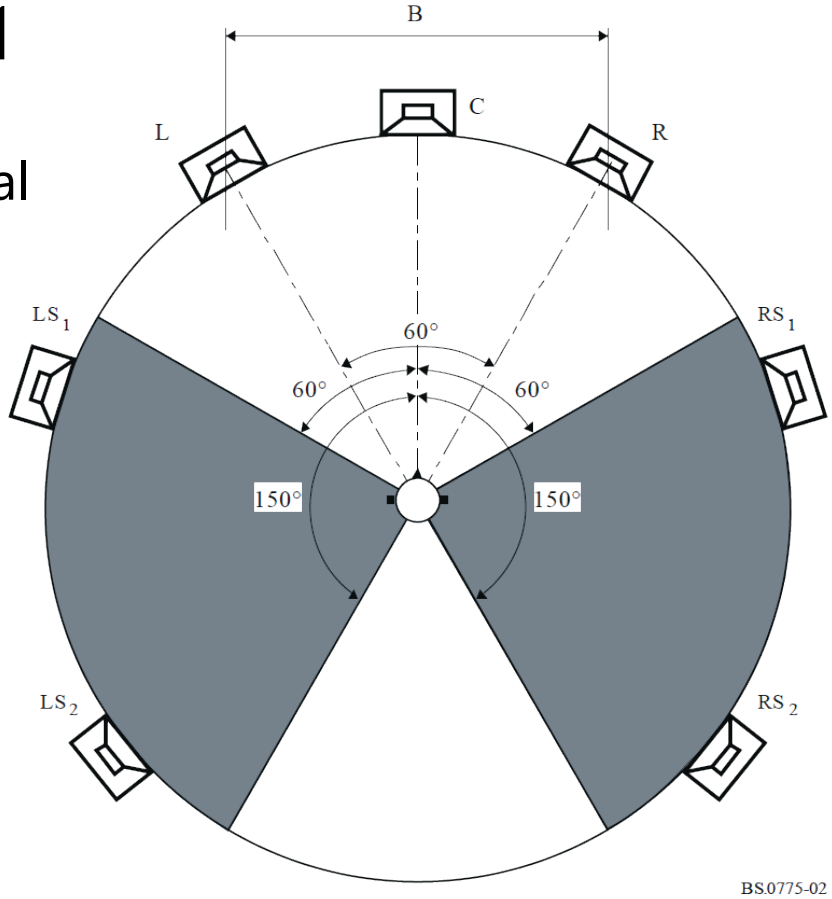
- Arrangement for optimal listening (production, broadcasting)
- + Subwoofer for LFE channel



[ITU 2012. Recommendation ITU-R BS.775-3 - Multichannel stereophonic sound system, International Telecommunication Union (ITU), Geneva, Switzerland.]

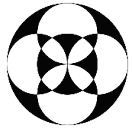
Surround sound 7.1

- Arrangement for optimal listening
- + Subwoofer for LFE channel



Ambisonics

Ambisonics



- Multi-channel recording and reproduction technique
- Generalization of Blumlein and MS stereophony: Extension to Surround and 3D
- Invented by Michael Gerzon and others in the 1970s

  [Gerzon, M.A., 1973. Periphony: With-Height Sound Reproduction. Journal of the Audio Engineering Society 21(1):2–10.]
- Lately increasing interest and applications in VR

  [Zotter, F. & Frank, M. 2019. Ambisonics - A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality, Springer Open.]

ETH zürich
Empa
Materials Science and Technology

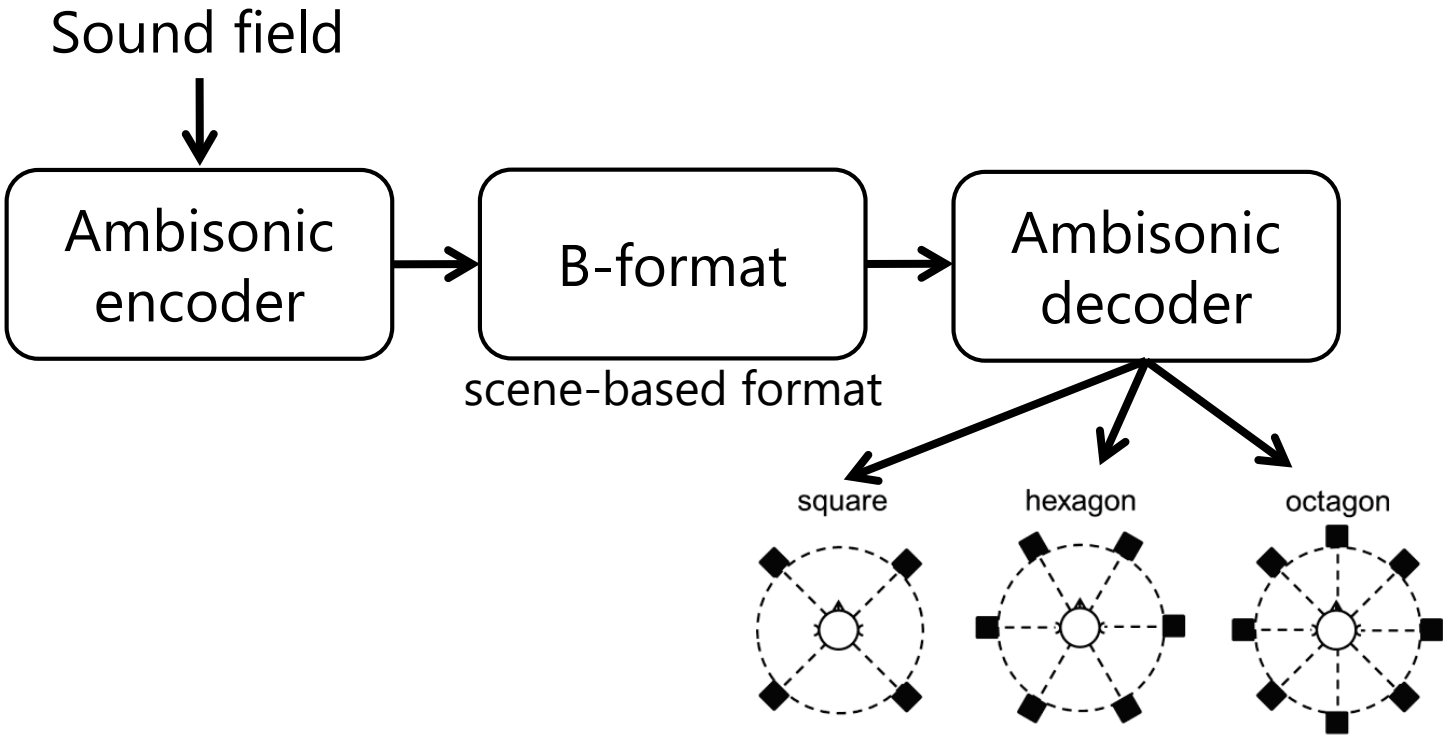
Ambisonics: Concept

- Decomposition of incident sound field at one position into sum of spherical harmonics
  - 1 monopole part C_0^0
  - 3 dipole parts C_1^{-1}, C_1^0, C_1^1
  - Higher order parts $C_n^{-j}, \ldots, C_n^j$

} First-order Ambisonics

- Description independent of loudspeaker arrangement
  → scene-based format

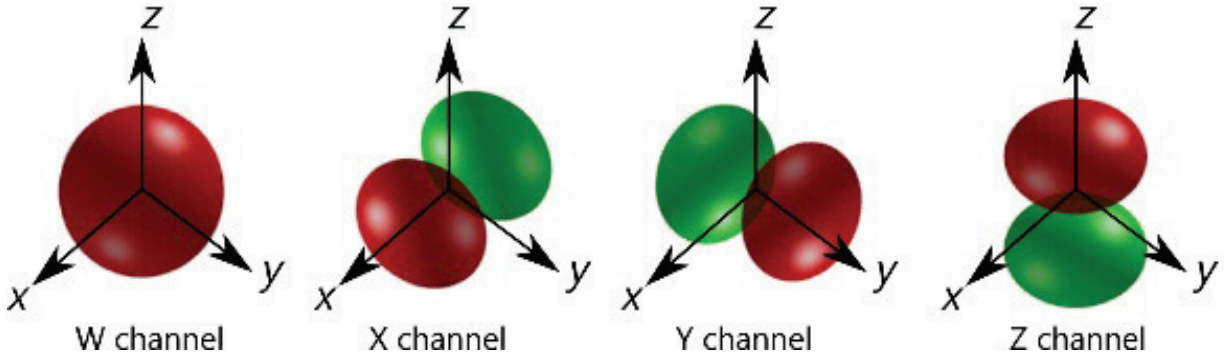
Ambisonics: Processing



→ regular loudspeaker arrangements highly preferred

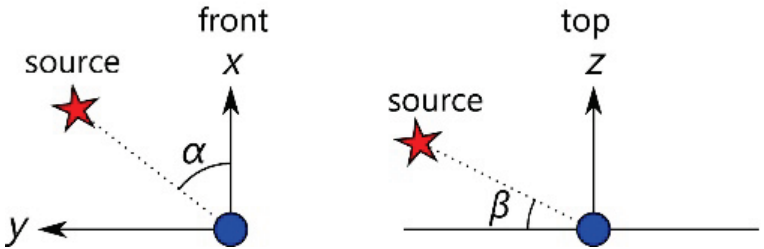
First-Order Ambisonics (FOA)

- B-format with 4 channels
  - W channel: monopole part, omni (scaled sound pressure)
  - X channel: figure-of-eight directivity in x-direction (pressure gradient)
  - Y channel: figure-of-eight directivity in y-direction
  - Z channel: figure-of-eight directivity in z-direction



ETH zürich
Empa
Materials Science and Technology

First-Order Ambisonics (FOA): Encoding

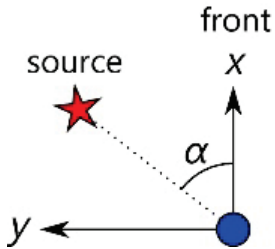


- Encoding of signal $p(t)$ from source under azimuth angle $\alpha(t)$and elevation angle $\beta(t)$(= an object-based format) into B-format

$$
\begin{aligned}
W &= \frac{1}{\sqrt{2}} \cdot p(t) \\
X &= \cos\alpha \cos\beta \cdot p(t) \\
Y &= \sin\alpha \cos\beta \cdot p(t) \\
Z &= \sin\beta \cdot p(t)
\end{aligned}
$$

ETH zürich
Empa
Materials Science and Technology

2D First-Order Ambisonics (FOA): Encoding



- Encoding of signal $p(t)$ from source under azimuth angle α into B-format

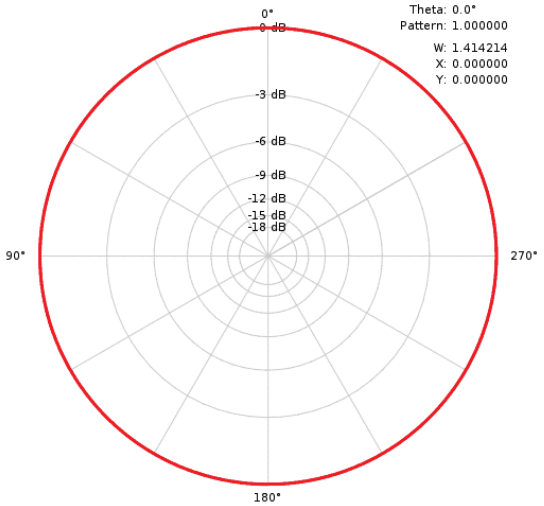
$$
\begin{aligned}
W &= \frac{1}{\sqrt{2}} \cdot p(t) \\
X &= \cos\alpha \cdot p(t) \\
Y &= \sin\alpha \cdot p(t) \\
Z &= 0
\end{aligned}
$$

ETH zürich
Empa
Materials Science and Technology

First-Order Ambisonics (FOA): Virtual microphones

- From W,X,Y,Z B-format channels, the output of any first order virtual microphone with arbitrary orientation can be generated
  → generalization of XY, MS, Blumlein stereo

- An acoustical scenery encoded in B-format can be manipulated in its direction at will (rotated around x, y, z-axis → matrix mixing)

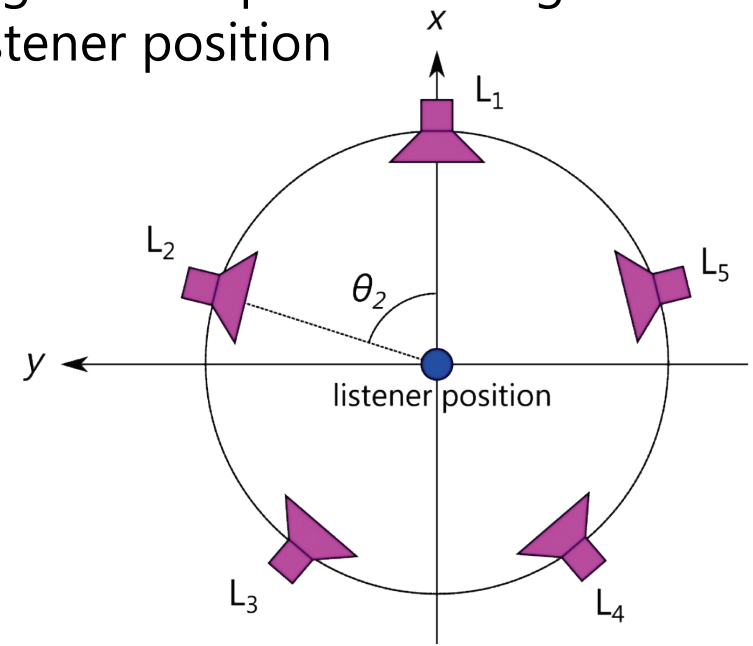


First-Order Ambisonics (FOA): Decoding to loudspeaker array

- Reproduction of Ambisonic audio by multiple loudspeakers
- Decoding = Calculation of individual speaker feed
- Different decoder types

2D First-Order Ambisonics (FOA): Decoding to regular loudspeaker array

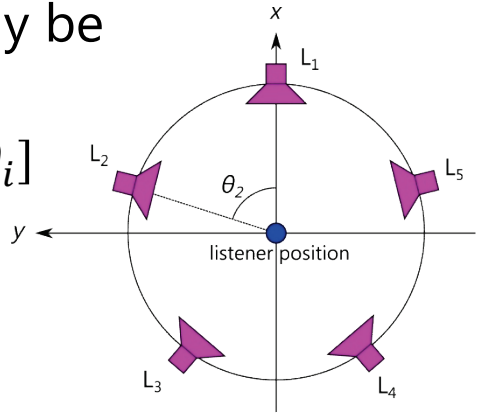
- Assumption: Regular 2D speaker arrangement on circle around listener position



2D First-Order Ambisonics (FOA): Decoding to regular loudspeaker array

- Signal F_i fed to loudspeaker i can generally be expressed as

$$F_i = K_1 \cdot W + K_2[X \cos\theta_i + Y \sin\theta_i]$$



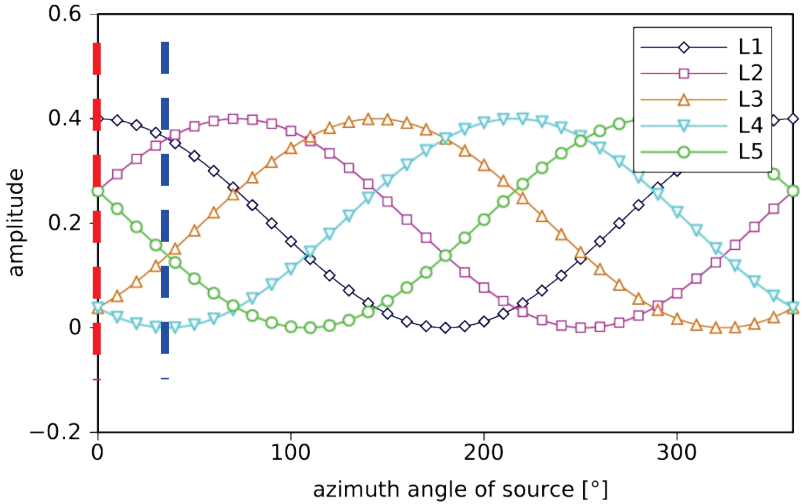
  - θ_i: angle of speaker i with respect to x axis
  - K_1: steers monopole component
  - K_2: steers directional component

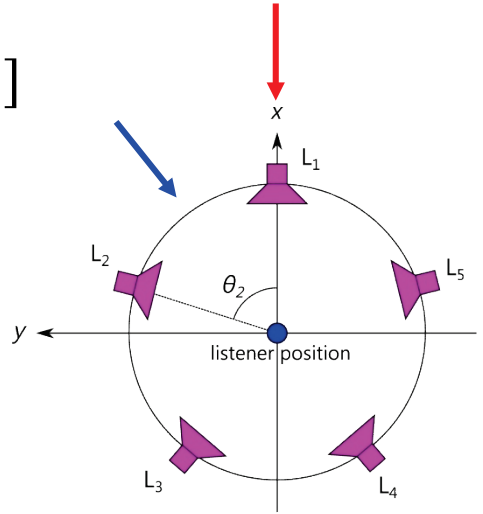
$\frac{K_1}{K_2}$ determines the decoder type

2D First-Order Ambisonics (FOA): Cardioid decoder (in-phase decoder)

$$F_i = K_1 \cdot W + K_2[X \cos\theta_i + Y \sin\theta_i]$$

$$\frac{K_1}{K_2} = \sqrt{2}$$





2D First-Order Ambisonics (FOA): Cardioid decoder (in-phase decoder)

- All speaker signals are in phase
  → no signal canceling by out-of-phase summation

- Localization is relatively robust with respect to changes in listener position

→ A larger listening area with proper localization

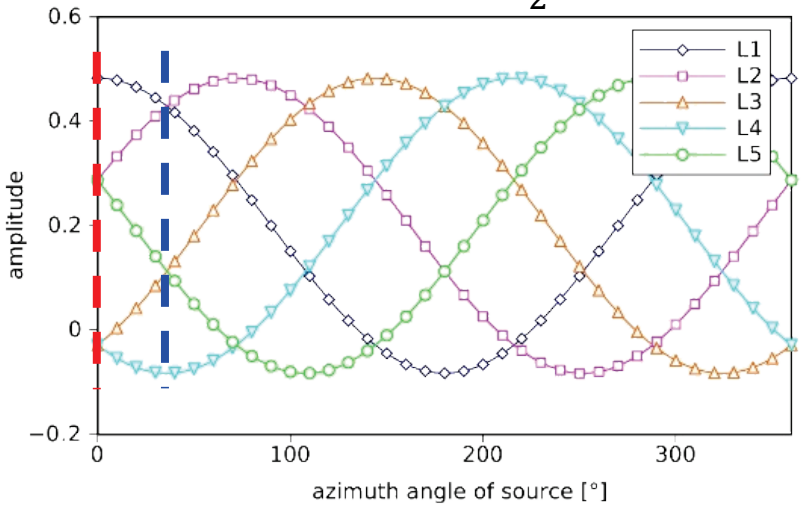
- Coloration due to many simultaneously active speakers (moderate directivity of cardioid)

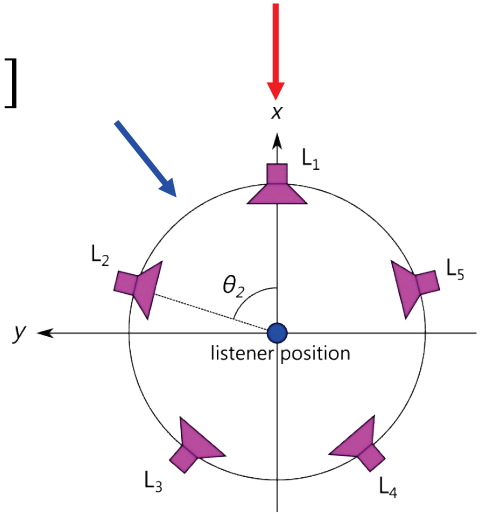
ETH zürich
Empa
Materials Science and Technology

2D First-Order Ambisonics (FOA): Energy localization vector decoder

$$F_i = K_1 \cdot W + K_2[X \cos\theta_i + Y \sin\theta_i]$$

$$\frac{K_1}{K_2} = 1$$





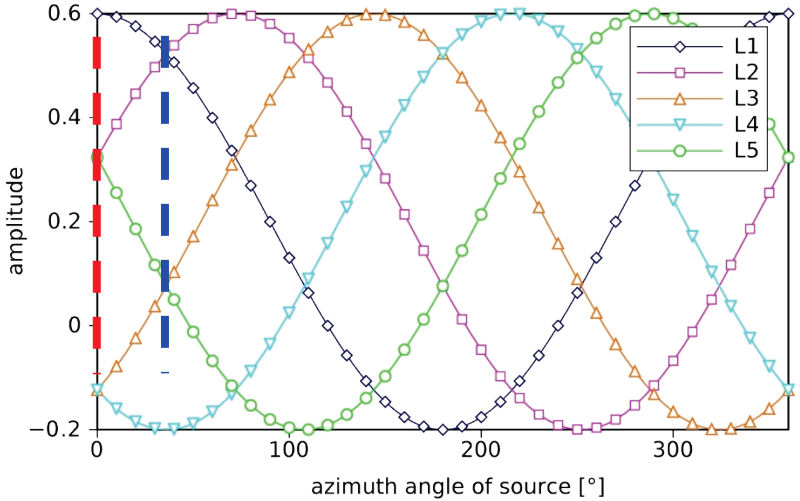
2D First-Order Ambisonics (FOA): Energy localization vector decoder

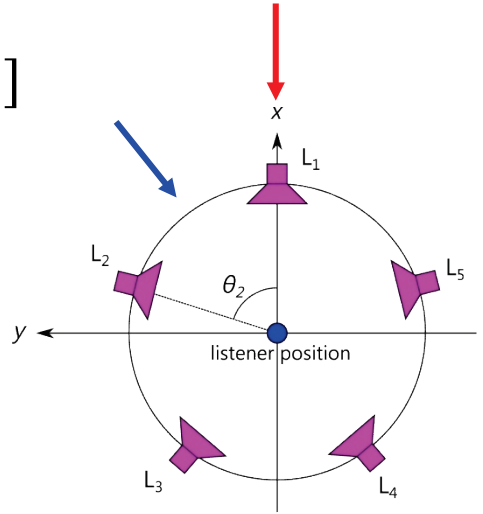
- Some signal canceling by out-of-phase summation
- Optimal localization in the frequency range from 700 to 4000 Hz where inter-aural level differences are of importance

2D First-Order Ambisonics (FOA): Velocity localization vector decoder

$$F_i = K_1 \cdot W + K_2[X \cos\theta_i + Y \sin\theta_i]$$

$$\frac{K_1}{K_2} = \frac{1}{\sqrt{2}}$$





2D First-Order Ambisonics (FOA): Velocity localization vector decoder

- Strong signal canceling by out-of-phase summation
- Optimal localization in the low frequency range where inter-aural time differences are of importance

First-Order Ambisonics (FOA): Decoding in general

- Compromise between localization (main lobe width), source stability and coloration
- Optimal decoder setting is dependent on frequency
  → K_1; K_s frequency dependent → phase-matched shelf-filters

- Difficulty with phase sensitive summation at high frequencies
  - Spatial separation of the two ears
  - Head acts as sound field distortion

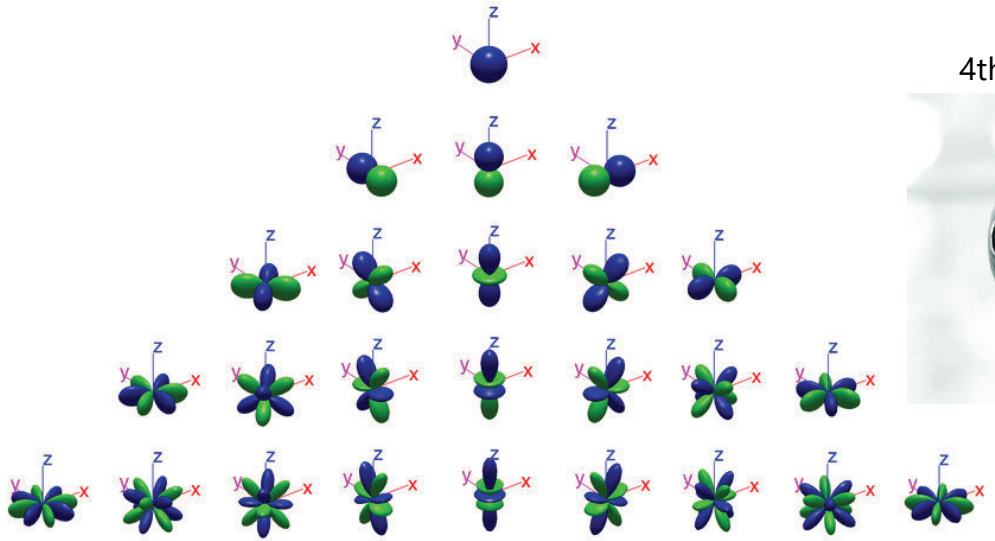
  → high-frequency compensation or HOA

ETH zürich
Empa
Materials Science and Technology

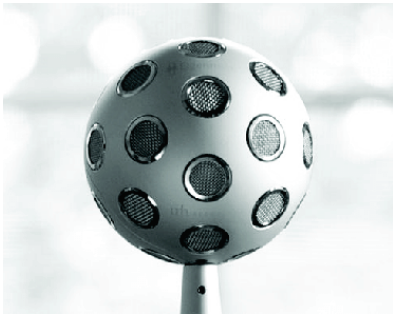
Higher Order Ambisonics (HOA)

- Higher directional resolution by higher number of channels
  → higher complexity in encoding and decoding

Order	Channels
0	1
1	4
2	9
3	16
4	25



4th order HOA mic

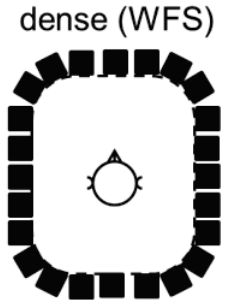


ETH zürich
Empa
Materials Science and Technology

Wave Field Synthesis (WFS)

Wave field synthesis (WFS)

- Stereo and Surround yield optimal results in sweet spot only
- Wave field synthesis:
  - Sound field reconstruction in an extended volume (or area)
  - Principle: Creation of artificial wavefronts synthesized by large number of individually driven loudspeakers (Huygens principle)
  - Pressure reconstruction in bounded region = Control pressure and normal particle velocity on boundary



[Spors, S., et al. 2008. The Theory of Wave Field Synthesis Revisited, Proceedings of the 124th AES Convention, Amsterdam.]

ETH zürich
Empa
Materials Science and Technology

Wave field synthesis (WFS)

- Mathematical basis: Kirchhoff-Helmholtz integral (→ Acoustics I)

$$\check{p}(x, y, z, \omega) = \frac{1}{4\pi} \int_S \left(\underbrace{\mathrm{j}\omega\rho_0 \check{v}_S(\omega)}_{\text{monopole and dipole secondary sources}} \frac{\mathrm{e}^{-\mathrm{j}kr}}{r} + \underbrace{\check{p}_S(\omega) \frac{1 + \mathrm{j}kr}{r^2} \cos\phi \mathrm{e}^{-\mathrm{j}kr}}_{\text{monopole and dipole secondary sources}} \right) \mathrm{d}S$$

- Different methods how to reduce and realize the sources, and derive the source steering functions for a given pressure

Wave field synthesis: Limitations

- Spatial discretization
  - Realization of continuous secondary source distributions on bounding surface with discrete elements (=real loudspeakers)
  - Spatial aliasing for small wavelenghts

Wave field synthesis: Limitations

- Room reflections
  - Sound field influenced by reflections at room boundaries and objects
  - Remedy:
    - Absorption
    - Near field installation (within critical distance)
    - Compensation in synthesis partially possible

ETH zürich
Empa
Materials Science and Technology

Wave field synthesis: Limitations

- Requirements regarding number of sources
  - Typical: restriction to horizontal plane (listening area instead of volume), 2D region → sources on surrounding path

Wave field synthesis: state-of-the-art

- ~100 installations woldwide
- Still experimental character
- Ongoing research

Fraunhofer HHI, Berlin

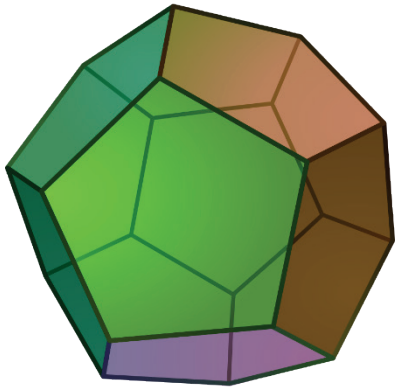
TU Berlin

3D audio


AuraLab

3D immersive sound («3D audio»)

- 3D loudspeaker arrangements → source above head (e.g. aircraft, reflection from ceiling)
- Many different arrangements
  - Spherical or upper hemispherical
  - Stacks of horizontal layers
  - Irregular or regular, e.g. Platonic solids
- Reproduction rendering
  a) Higher Order Ambisonics (HOA) decoding
  b) Vector Base Amplitude Panning (VBAP) from object-based audio format

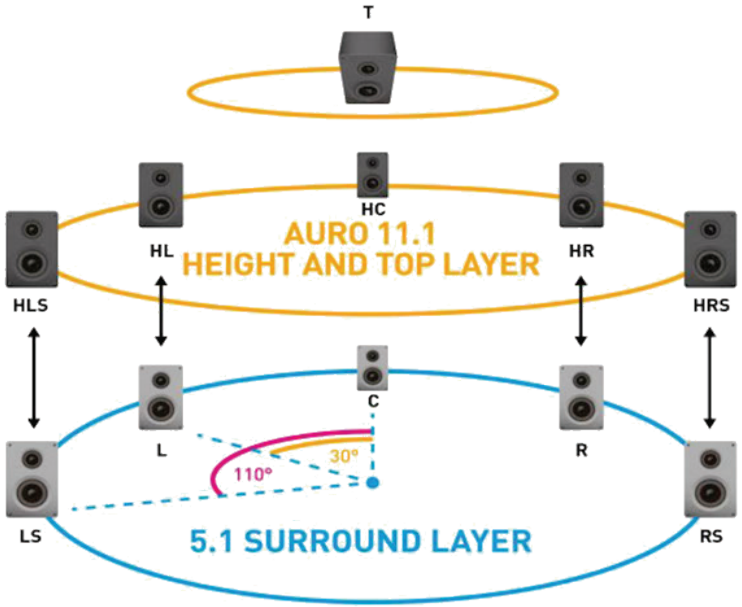


Dodecahedron

ETH zürich
Empa
Materials Science and Technology

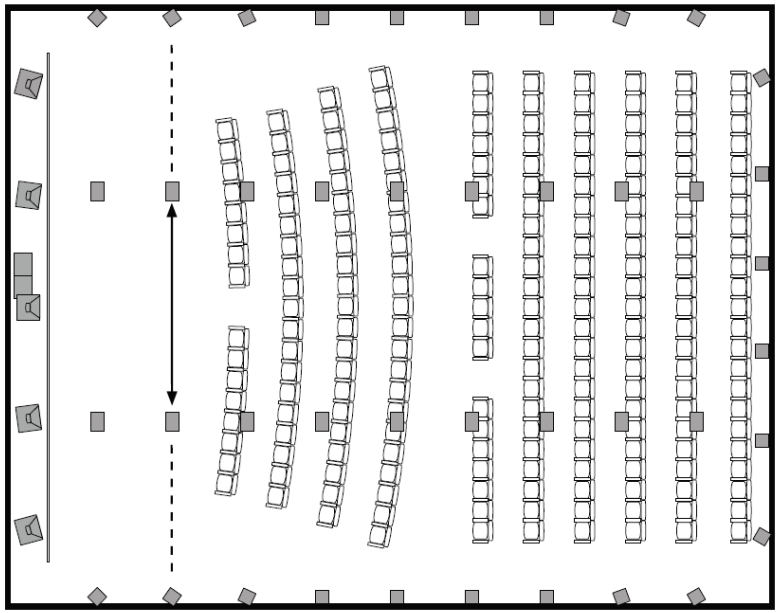
3D audio in cinemas and home theaters

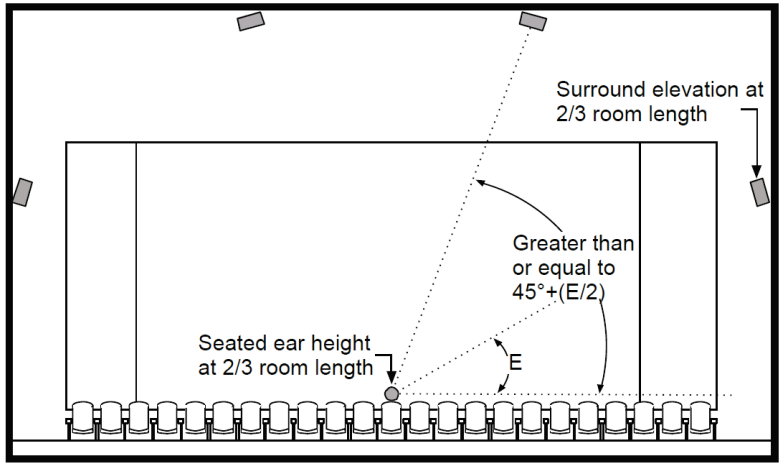
- Typical: irregular, upper hemisphere layouts
- Commercial formats
  - Auro-3D (channel-based)
    - 3 height layers, 11.1 channels
  - DTS:X (fully object-based)
  - Dolby Atmos (channel- plus object-based)



3D audio in cinemas and home theaters

- Dolby Atmos: 2 height layers





[Dolby 2015. Dolby Atmos Specifications – Issue 3, Dolby Laboratories, Inc.]

Vector Base Amplitude Panning (VBAP)

- Original algorithm by Ville Pulkki in 1997

  [Pulkki, Ville 1997. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. Journal of the Audio Engineering Society 45(6).]

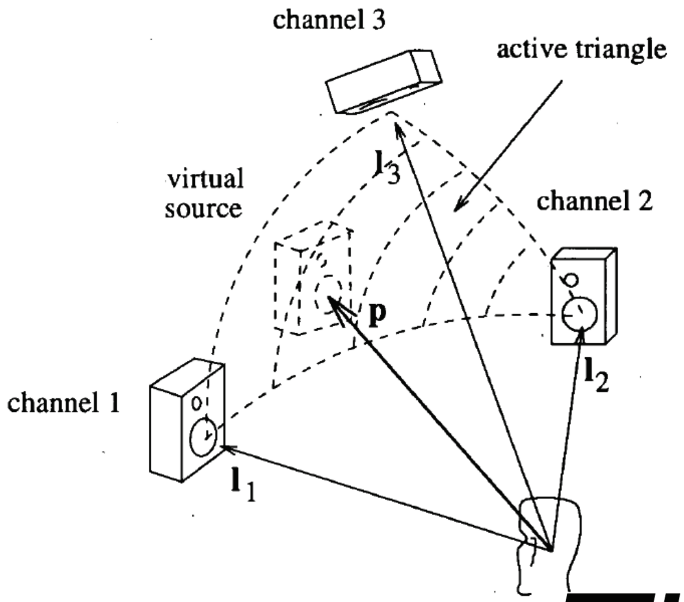
- Widely used in various (improved) forms
- Panning for arbitrary 3D loudspeaker arrays (with constant listening distance)

Vector Base Amplitude Panning (VBAP)

- Generalization of pair-wise panning with tangent law to 3D
  - Pair-wise → triplet-wise
- Procedure
  1. Determining «active triangle»
  2. 3 gain factors via matrix operation

  projection of vector $\boldsymbol{p}$ onto vector base defined by $\boldsymbol{l_1, l_2, l_3}$
  3. Scaling to satisfy gain normalization

$$g_1^2 + g_2^2 + g_3^2 = C$$



Combinations: Virtual loudspeaker playback

Virtual loudspeaker playback

- Mimiking loudspeaker playback over headphones
- Interpret each speaker as a source → apply HRTFs
- Needs head tracking (inertial, optical sensors) to dynamically adjust HRTFs

ETH zürich
Empa
Materials Science and Technology

Virtual loudspeaker playback

- Applications:
  - Listening/Mixing 5.1 Surround sound over headphones
    (e.g. Windows Sonic (Win10), Dolby Atmos for Headphone, Waves Nx)
  - Efficient head-rotation in VR (e.g. YouTube spatial audio)

1. B-format → real-time scene rotation according to head rotation by dynamic matrix mixing
2. Ambisonic decoding to virtual regular loudspeaker array rotating with listener
3. static HRTF filtering

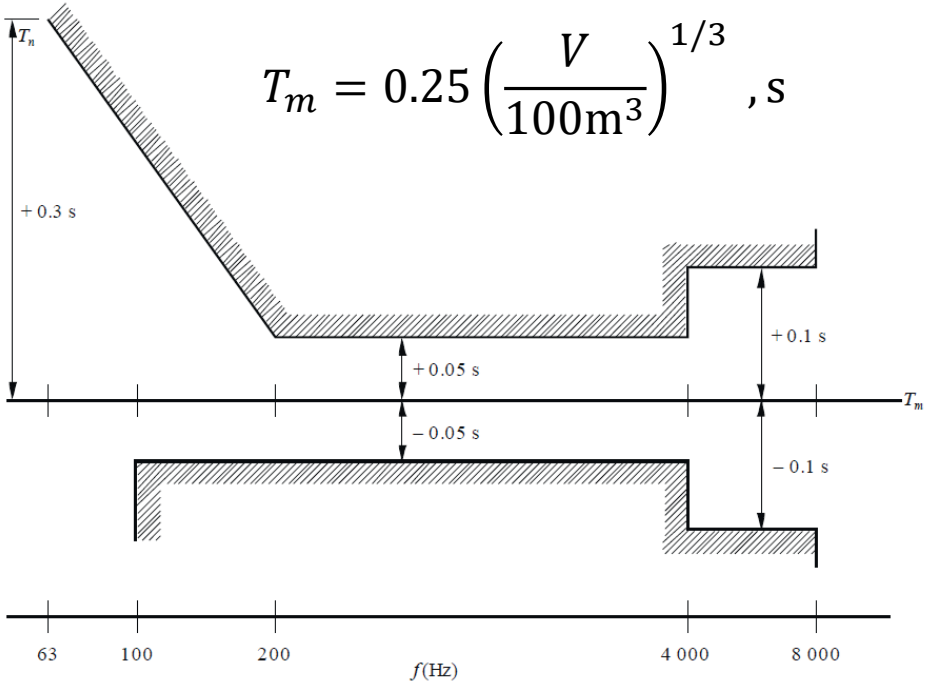
} static binaural decoding

ETH zürich
Empa
Materials Science and Technology

Room influence on reproduction

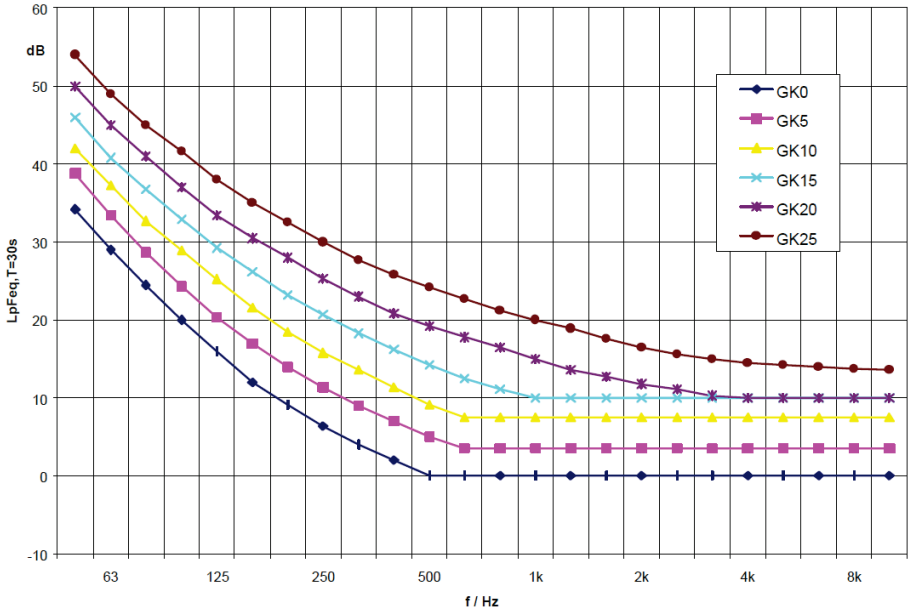
- Example: Reverberation time tolerance limits for assessment of audio systems (ITU-R BS.1116-2)

$$T_m = 0.25\left(\frac{V}{100\mathrm{m}^3}\right)^{1/3}, \mathrm{s}$$



Requirements on background noise

- Example requirements: Noise rating curves (GK) for radio and TV studio rooms

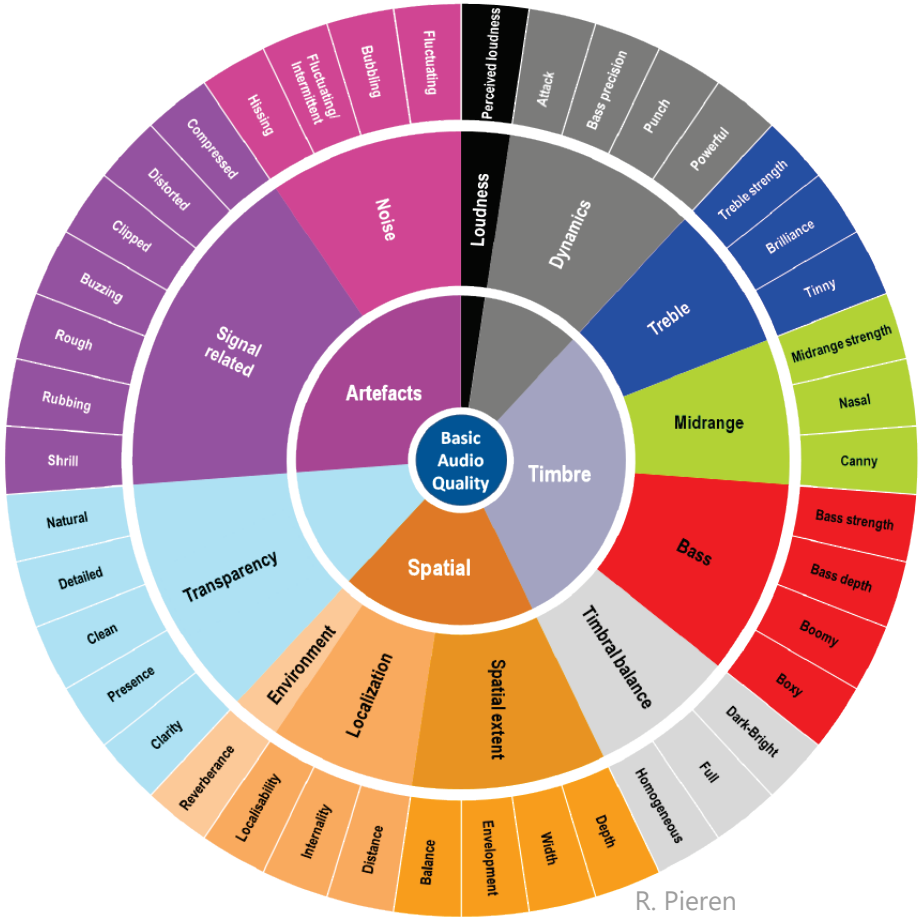


Listening room: GK5-GK15

[IRT 1995. Akustische Information 1.11-1 – Höchstzulässige Schalldruckpegel von Dauergeräuschen in Studios und Bearbeitungsräumen bei Hörfunk und Fernsehen, Institut für Rundfunktechnik, München.]

ETH zürich
Empa
Materials Science and Technology

Audio quality assessment: «Audio Wheel» by ITU



- Many defined attributes to be assessed

[ITU 2017. Recommendation ITU-R BS.2399-0 – Methods for selecting and describing attributes and terms, in the preparation of subjective tests, International Telecommunication Union (ITU), Geneva, Switzerland.]

ETH zürich
Empa
Materials Science and Technology

Take home messages

- Large variety of sound reproduction techniques
- No reproduction system for all purposes
- Still active field of research and engineering (HOA, WFS, 3D audio, HRTF, head tracking)

ETH zürich
Empa
Materials Science and Technology