Immersive audio – psycho-acoustics, principles, recording methods

Günther Theile

theile@tonmeister.de

Immersive audio – psycho-acoustics, principles, recording methods

- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- Head tracking
- Binaural Room Synthesis (BRS)
- Virtual headphone

Immersive audio – psycho-acoustics, principles, recording methods

- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- Head tracking
- Binaural Room Synthesis (BRS)
- Virtual headphone

Immersive experience - desirable effects :

- Representation of the auditory scene environment
- Support of spatial perception incl. height, distance, perspective
- Involvement of the listener in the action

Particularly important in today's high-quality
live, event, sport, game, movie, documentary productions

Atmo

Typs:

- Long reverberation (perceived as enveloping portion)
- Non-reflected spatially diffuse sound (e. g. rustling leaves in the forest)
- Ambient noise outdoors (e. g. trafic, roughly localizable)
- Scenery of sources (e. g. applause, crowd in the stadium)
- Spatially distributed off-sources (irrelvant directions, precisely localizable)

Diffusivity

Typs:

- Long lasting reverberation (perceived as enveloping portion)
- Non-reflected spatially diffuse sound (e. g. rustling leaves in the forest)
- Ambient noise outdoors (e. g. trafic, roughly localizable)
- Scenery of sources (e. g. applause, crowd in the stadium)
- Spatially distributed off-sources (irrelvant directions, precisely localizable)

Space

Spatial transmission formats – basic characteristics

Comparison of attribute profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 22.2	WFS* (HOA*)	BINAURAL TECHNIQUES
Front direction, robustness	●	●●	●●	●●	●
Surround direction		●	●●	●●	●●
Elevation			●***		●●
Height			●●		●●
Distance / depth	(●)**	●	●	●●	●●
Proximity to the head				(●)	●●
Intra-active perspective				●●	
Spatial impression	(●)**	●	●●	●	●●
Envelopment		●	●●	●	●●
Timbre	●●	●●	●●	●	●●

*horizontal arrays **simulated depth/spatial impression ***unstable; at the sweetspot only

Spatial transmission formats – basic characteristics

Comparison of attribute profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 22.2	WFS* (HOA*)	BINAURAL TECHNIQUES
Front direction, robustness	●	●●	●●	●●	●
Surround direction		●	●●	●●	●●
Elevation			●***		●●
Height			●●		●●
Distance / depth	(●)**	●	●	●●	●●
Proximity to the head				(●)	●●
Intra-active perspective				●●	
Spatial impression	(●)**	●	●●	●	●●
Envelopment		●	●●	●	●●
Timbre	●●	●●	●●	●	●●

*horizontal arrays **simulated depth/spatial impression ***unstable; at the sweetspot only

Note:

There are no auditory events between the loudspeakers and the listener.
A head close auditory environment can not be imaged.
Sources close to the microphone are perceived close to the loudspeakers, not to the listener.

Example „applause“:
The atmo-microphone too close to the audience enhances the perception of the loudspeaker distance as the boundary of the imaging scene (effect of "acoustic empty bubble").

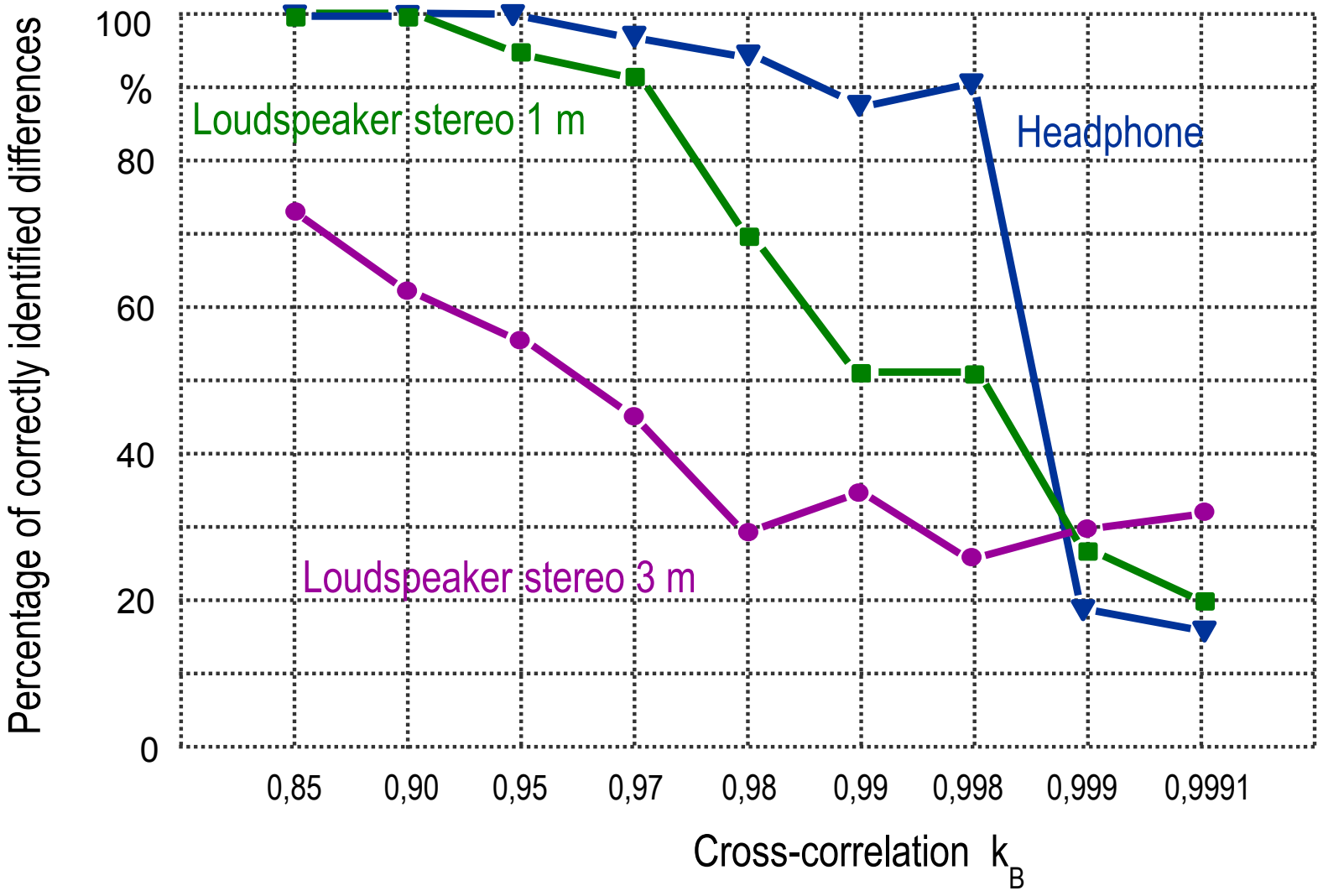
- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- Head tracking
- Binaural Room Synthesis (BRS)
- Virtual headphone

Headphone reproduction

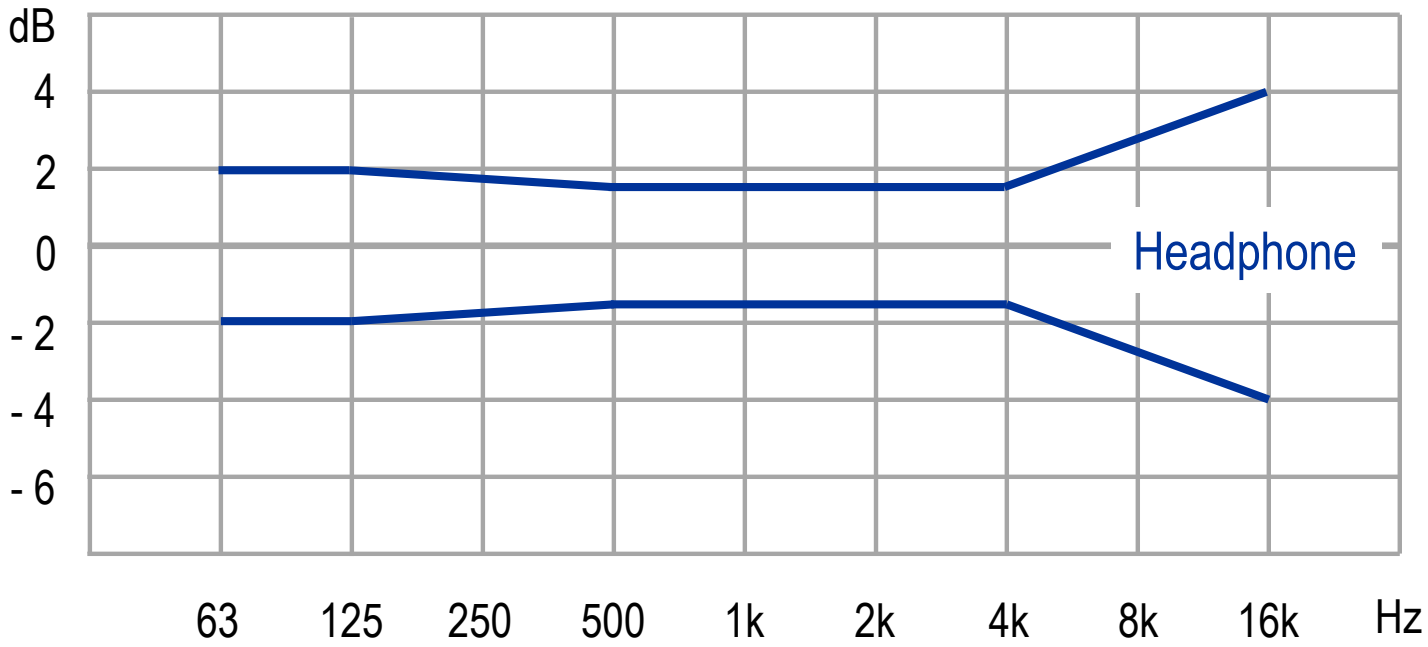
Headphones are the preferred choice for many professional and consumer applications:

- Advantages over loudspeaker reproduction, both in practice and as a matter of principle.
- Provision of controlled and highly consistent listening conditions independent of location.
- The direct coupling of sound transducer and ear entrance provides ideal preconditions for high-fidelity, defined playback, particularly in terms of **imaging precision** and **tone colour**.

Imaging precision of headphone and loudspeaker

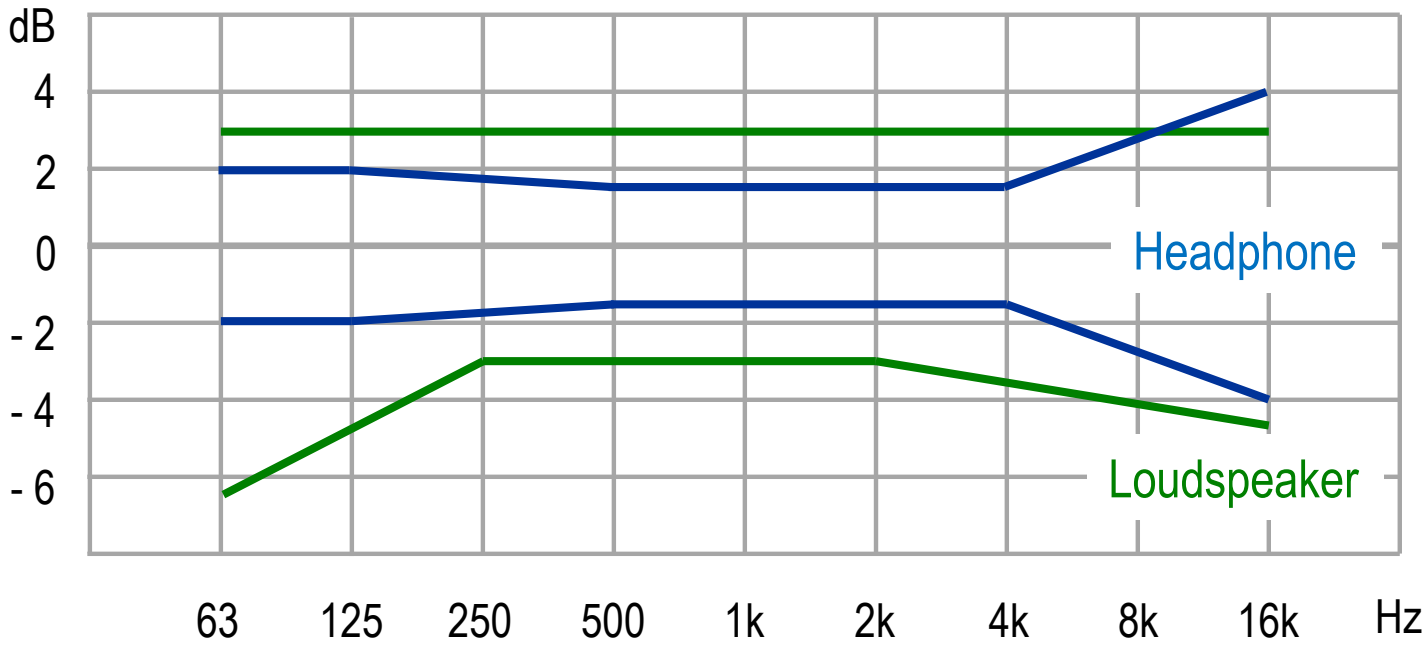


Tolerance corridor of frequency headphone response



Diffuse field transfer function of headphones according to ITU-Rec. BS 708

Tolerance corridors of frequency responses compared

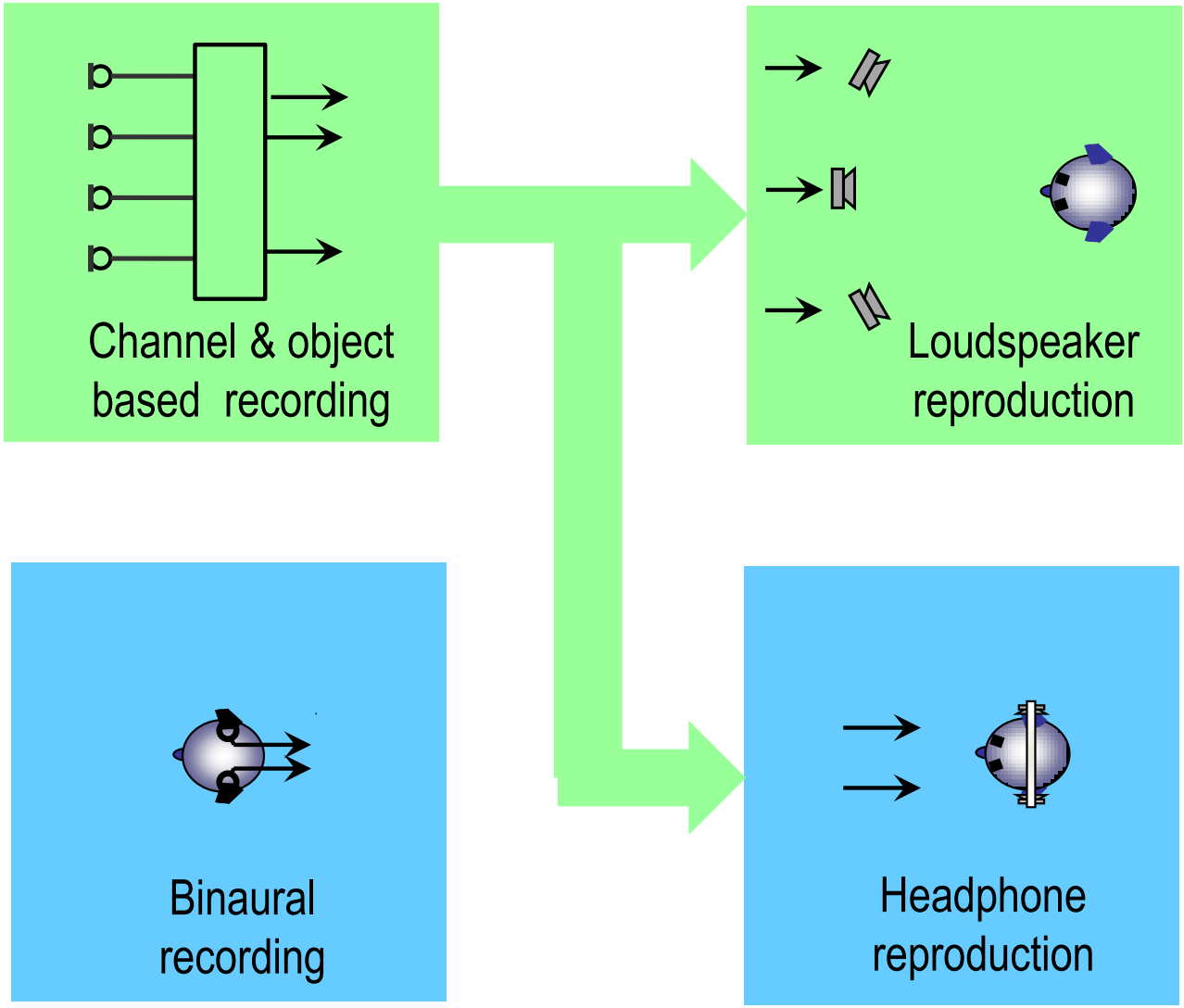


Diffuse field transfer function of headphones according to ITU-Rec. BS 708

Sound pressure level at listening position of listening room according to ITU-Rec. BS 1116

Headphone reproduction

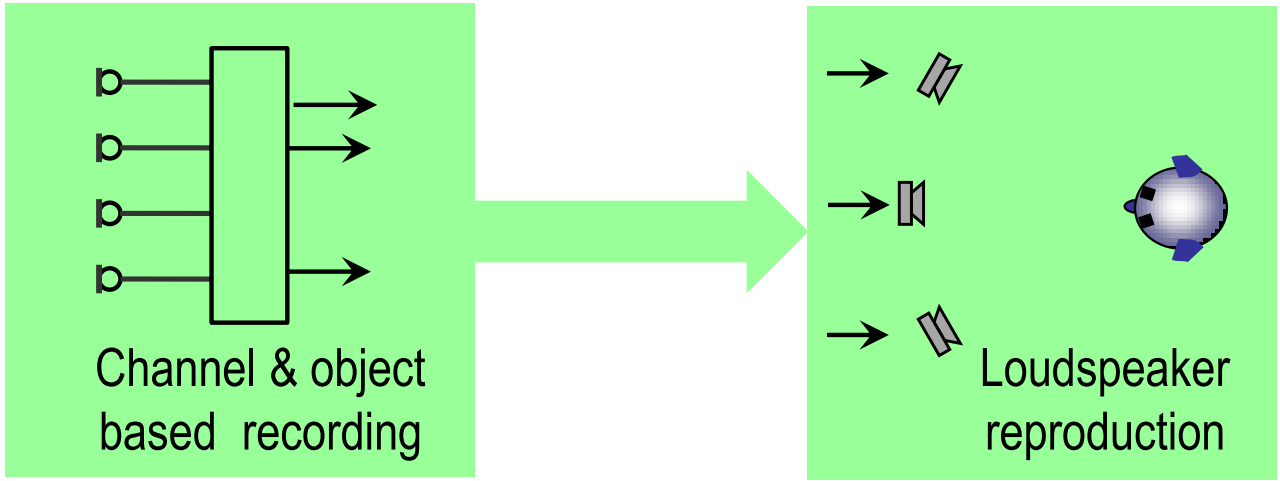
Compatibilty requires neutral headphone equalization standard !



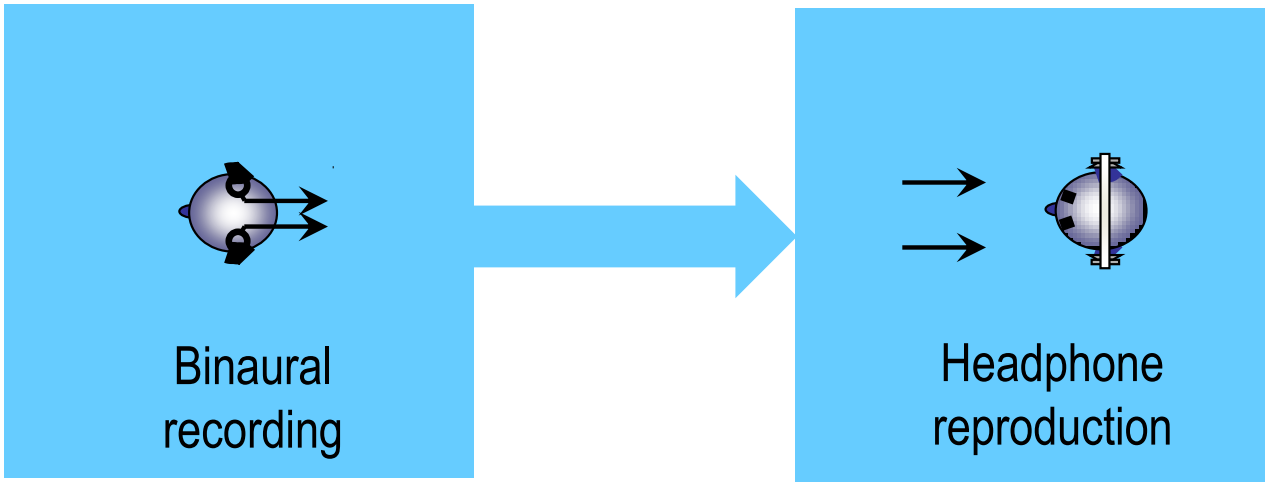
Calibration:
Linear sound pressure level at listening position
(ITU-Rec. BS 1116)

Calibration:
Linear headphone diffuse field transfer function
(ITU-Rec. BS 708)

Headphone reproduction



Calibration:
Linear sound pressure level at listening position (ITU-Rec. BS 1116)



Calibration:
Linear headphone diffuse field transfer function (ITU-Rec. BS 708)

ANNEX 2

Diffuse-field frequency response of studio monitor headphones

Measurement specification

1. Diffuse-field frequency response

Why?

ANNEX 2

Diffuse-field frequency response of studio monitor headphones

Measurement specification

1. Diffuse-field frequency response

Directionally neutral coupling transducer – ear canal

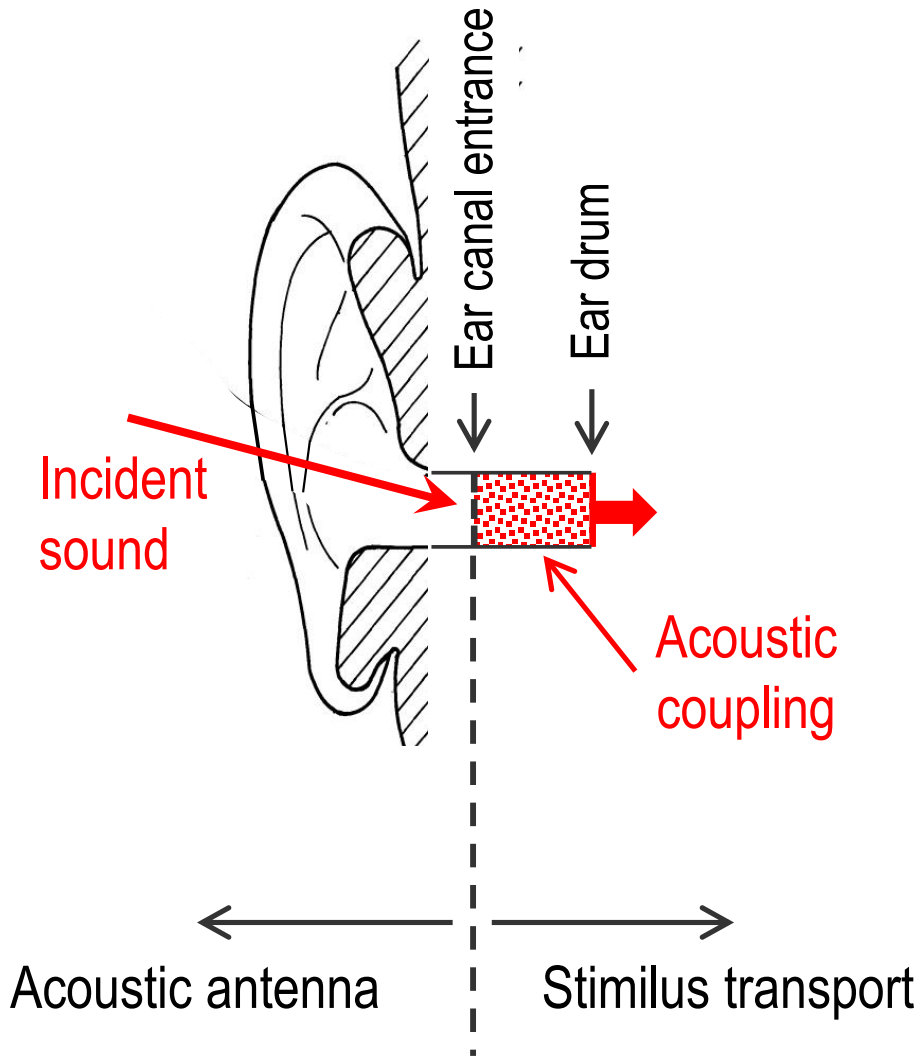
→ Headphone transfer function = 1

→ Integral over all free-field transfer functions (HRTFs)

Ear canal entrance input - natural hearing

The acoustic antenna is effective:

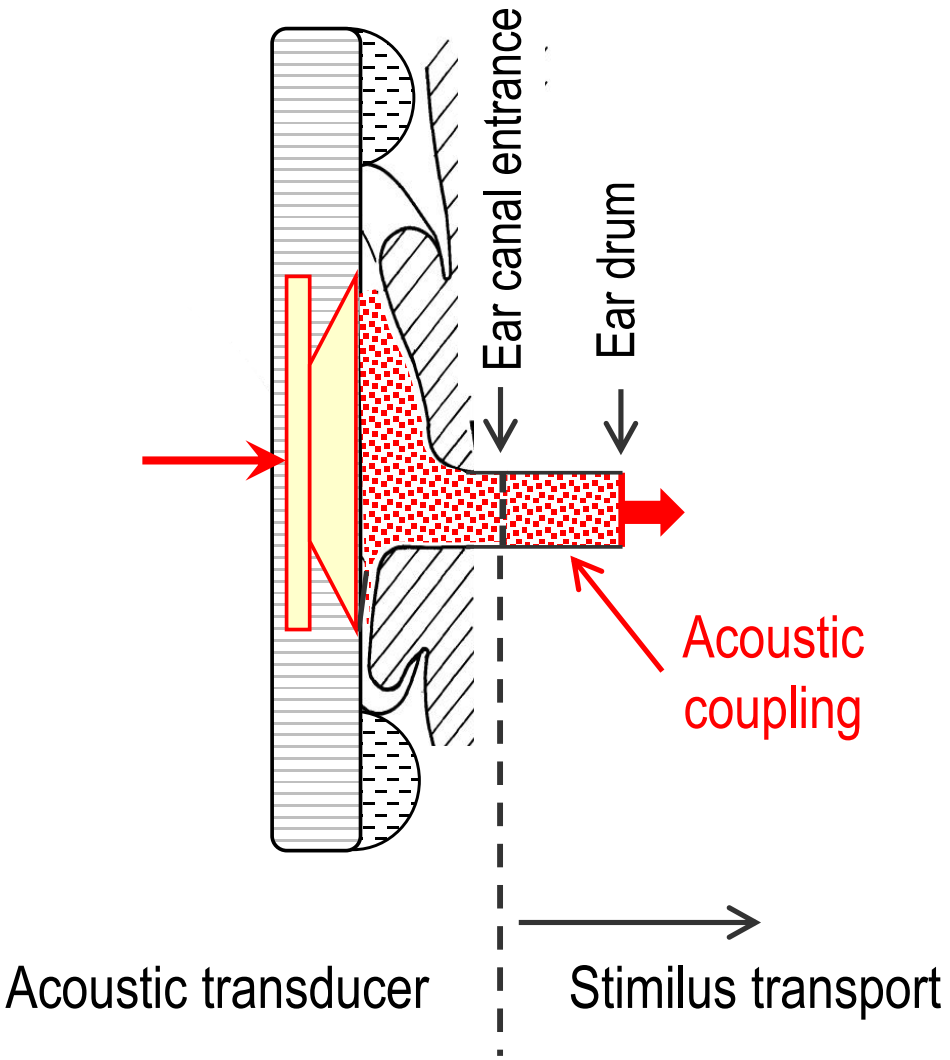
→ HRTF



Ear canal entrance input - headphone

The acoustic antenna is ineffective:

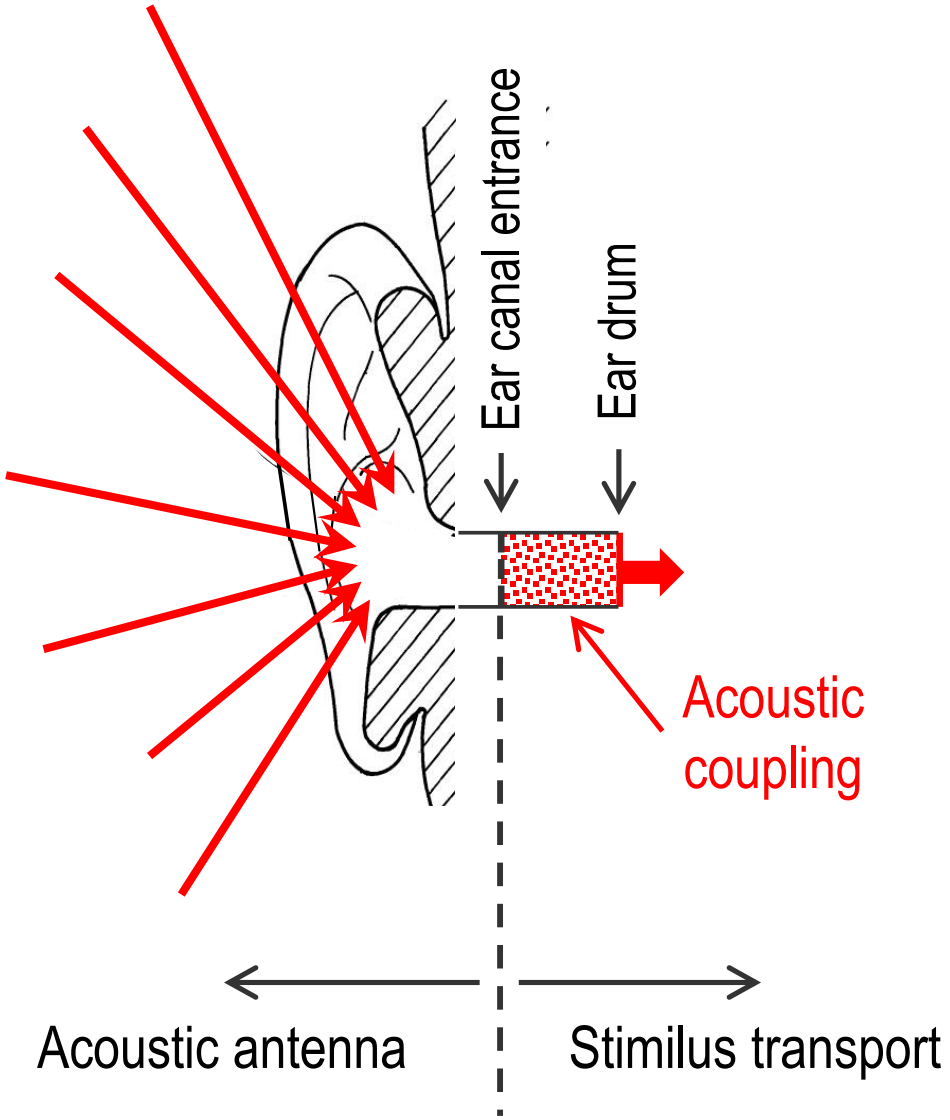
→ No distant source, no HRTF



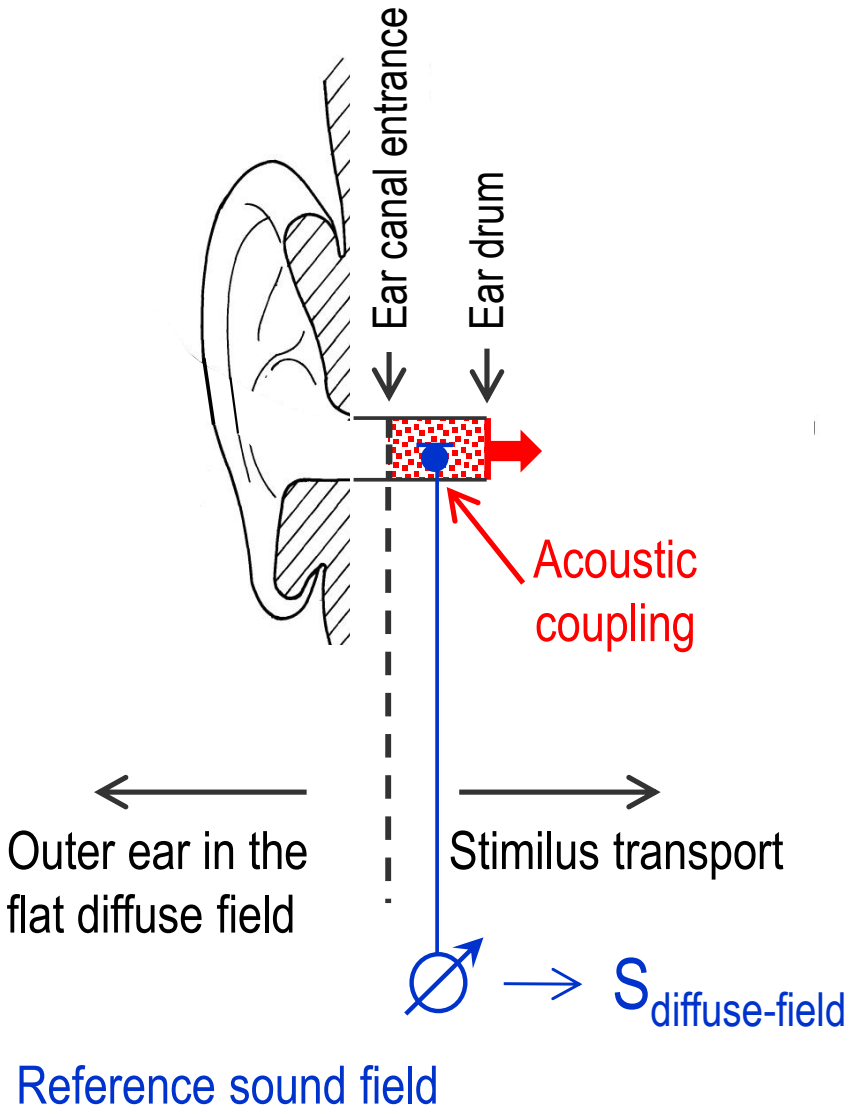
Ear canal entrance input - diffuse sound field

→ Required reference sound field:

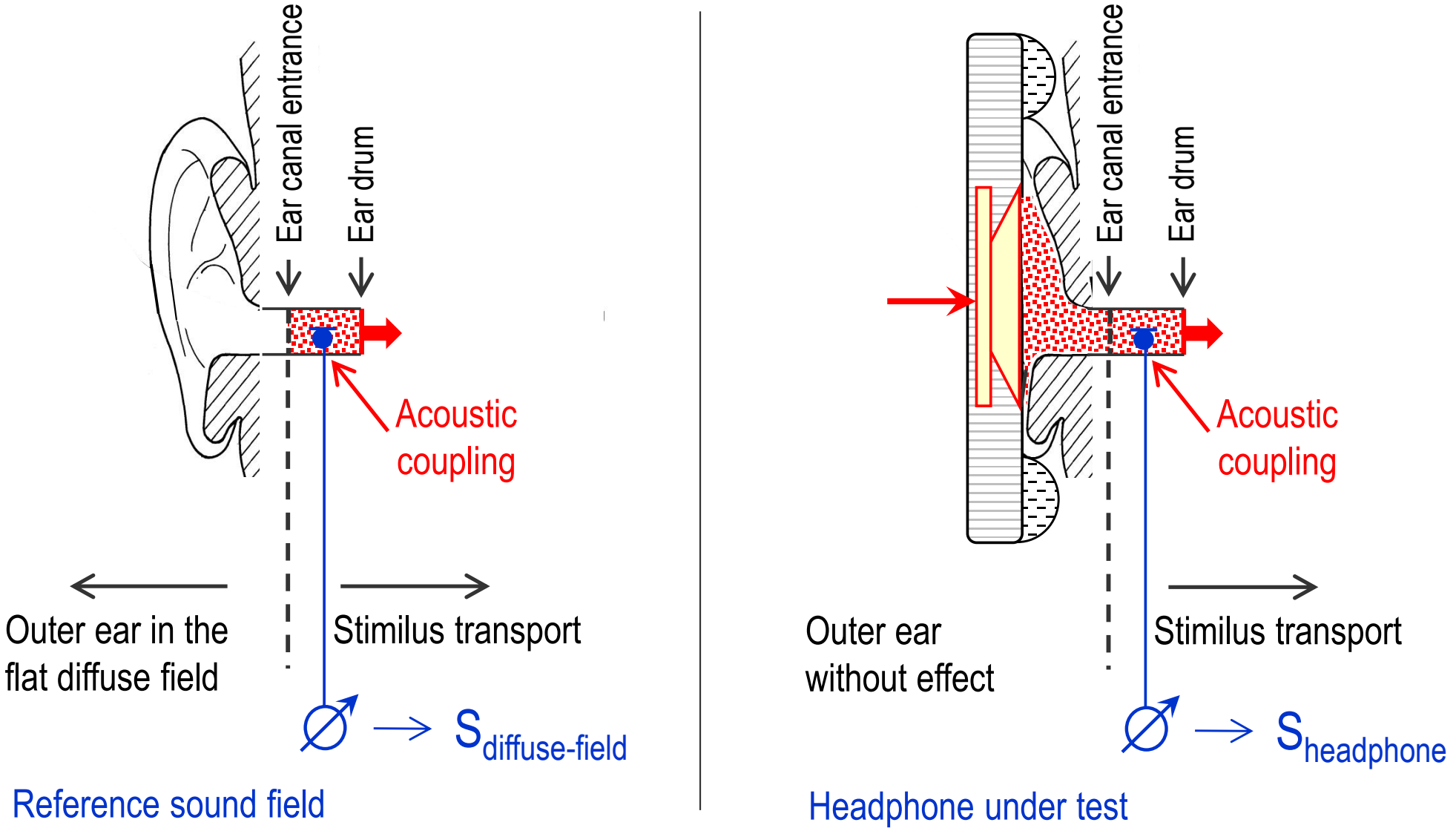
- sufficient spatial diffusivity
- flat frequency response !



ITU-R BS.708 - method of measurement

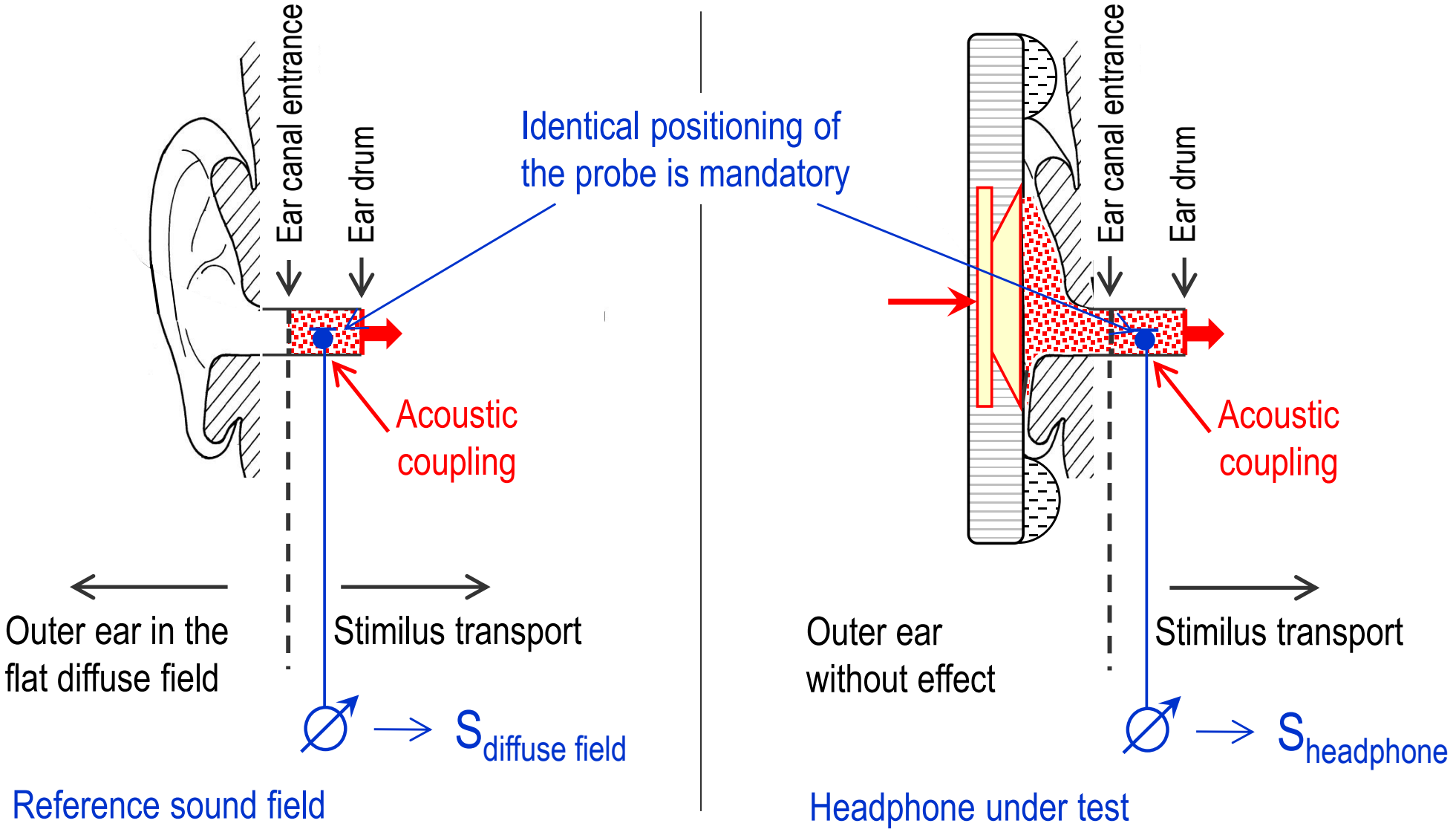


ITU-R BS.708 - method of measurement



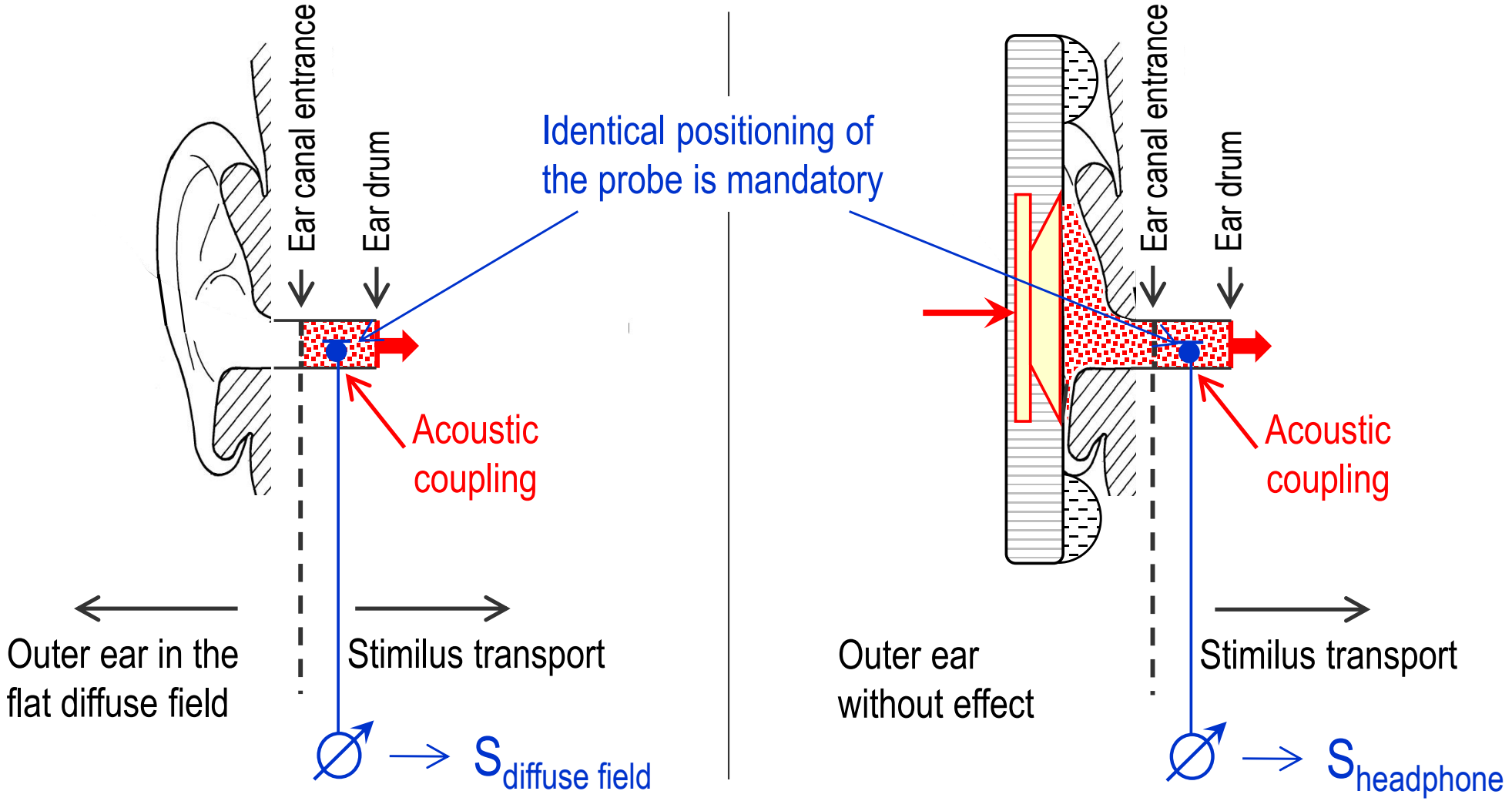
Comparison of output levels of the probe, third-octave measurement

ITU-R BS.708 - method of measurement



Comparison of output levels of the probe, third-octave measurement

ITU-R BS.708 - method of measurement



$$S_{EQ} = S_{diffuse\ field} - S_{headphone}$$

→ Averaging the results from 16 persons in each frequency band

ANNEX 2

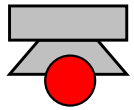
Diffuse-field frequency response of studio monitor headphones

Measurement specification

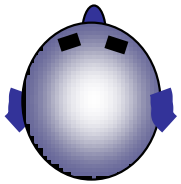
2. No free field loudness comparison measurement

Why?

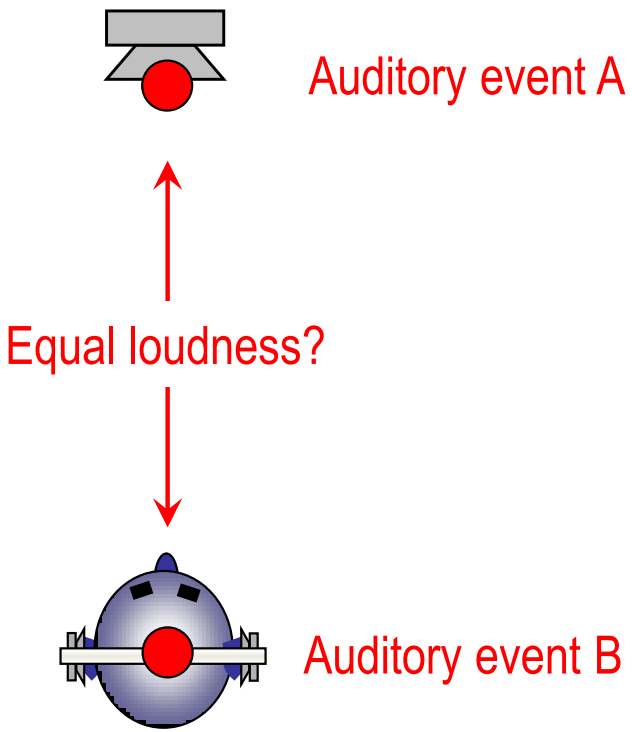
Loudness comparison (free field equalization)



Auditory event A



Loudness comparison (free field equalization)



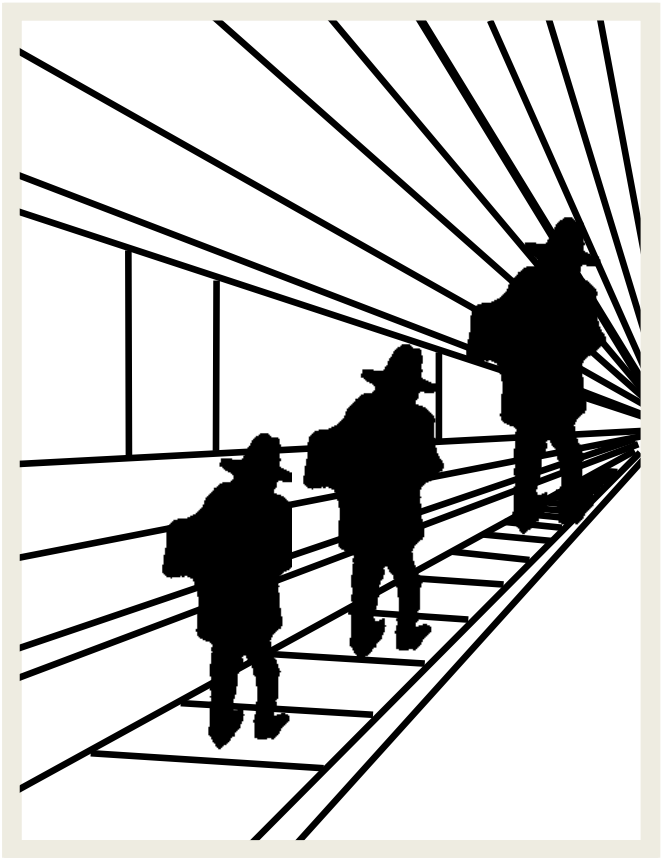
Figures of equal size

Figures of equal size

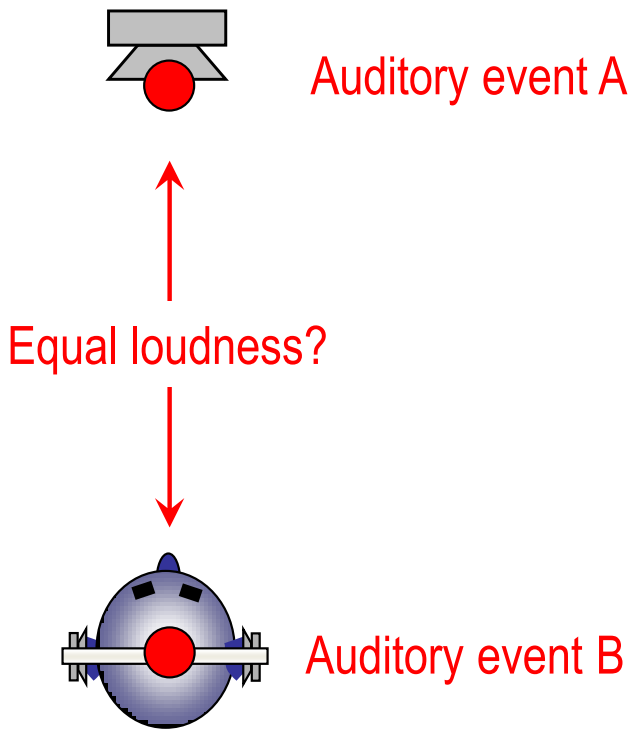
Figures of equal size

Headphone diffuse-field equalization

Figures of equal size ?



Loudness comparison (free field equalization)



Loudness comparison vs. probe measurement

Sound level in the ear canal for equal loud sounds (free-field 0° front)
[Fastl, Schmid, Theile, Zwicker, DAGA 1985]

Higher sound level in case of headphone reproduction

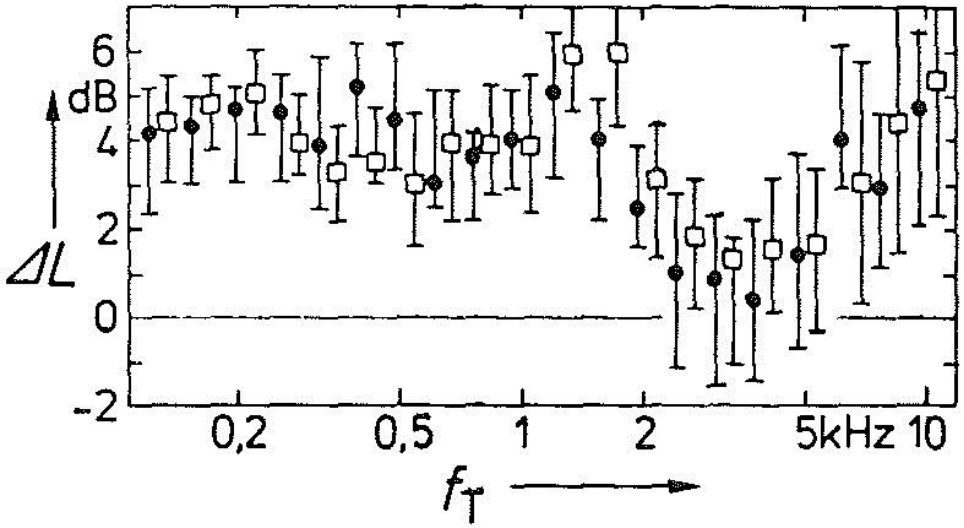
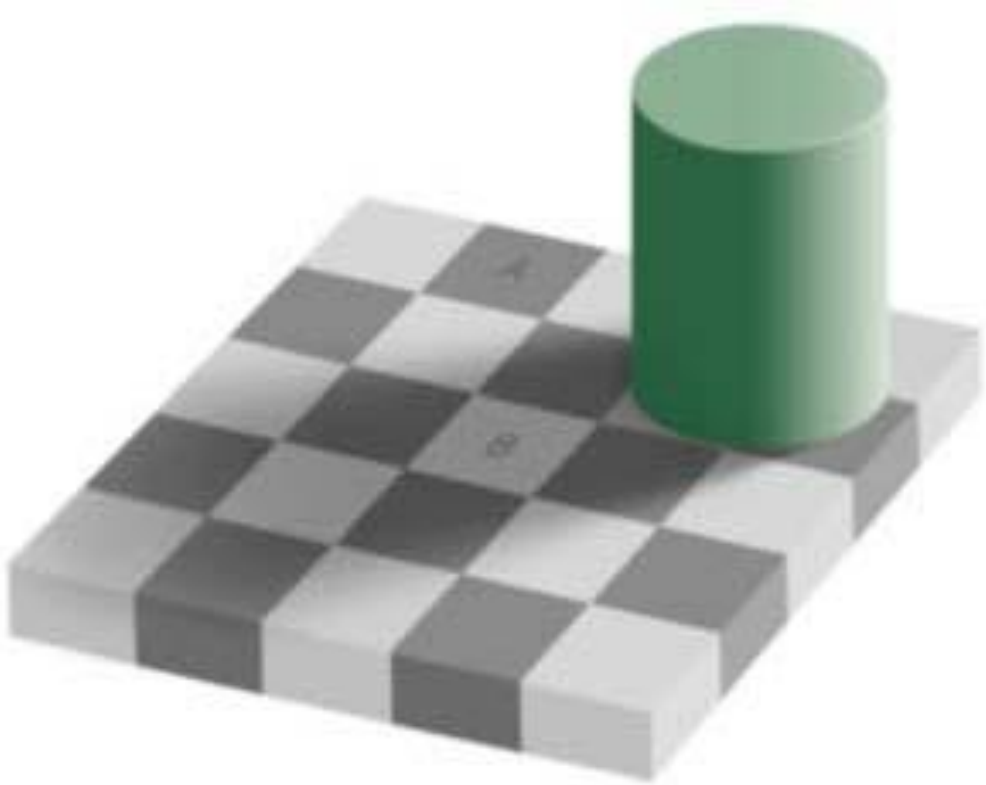


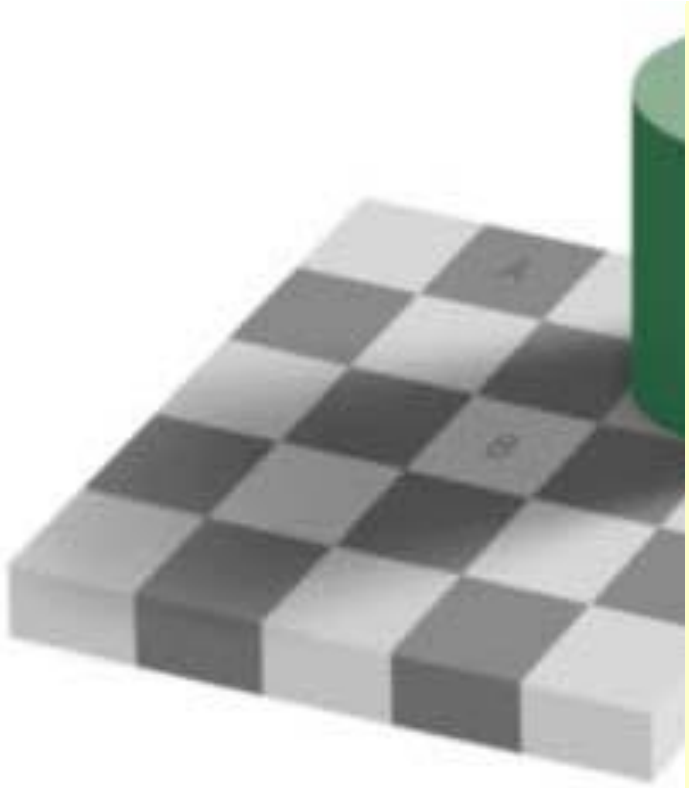
Fig.1: Lautheitsvergleich von Sinustönen, die im Reflexionsarmen Raum aus Kopfhörern bzw. Lautsprechern dargeboten werden. Pegeldifferenz $\Delta L = L_{GKH} - L_{GLS}$ der bei gleicher Lautheit im Gehörgang gemessenen Schallpegel für Kopfhörerbeschallung (L_{GKH}) bzw. Lautsprecherbeschallung (L_{GLS}) als Funktion der Testtonfrequenz f_T. Zentralwerte mit wahrscheinlichen Schwankungen für acht Versuchspersonen. Ausgefüllte Kreise: Geschlossener, elektrodynamischer Hörer, Quadrate: Offener, elektrostatischer Hörer.

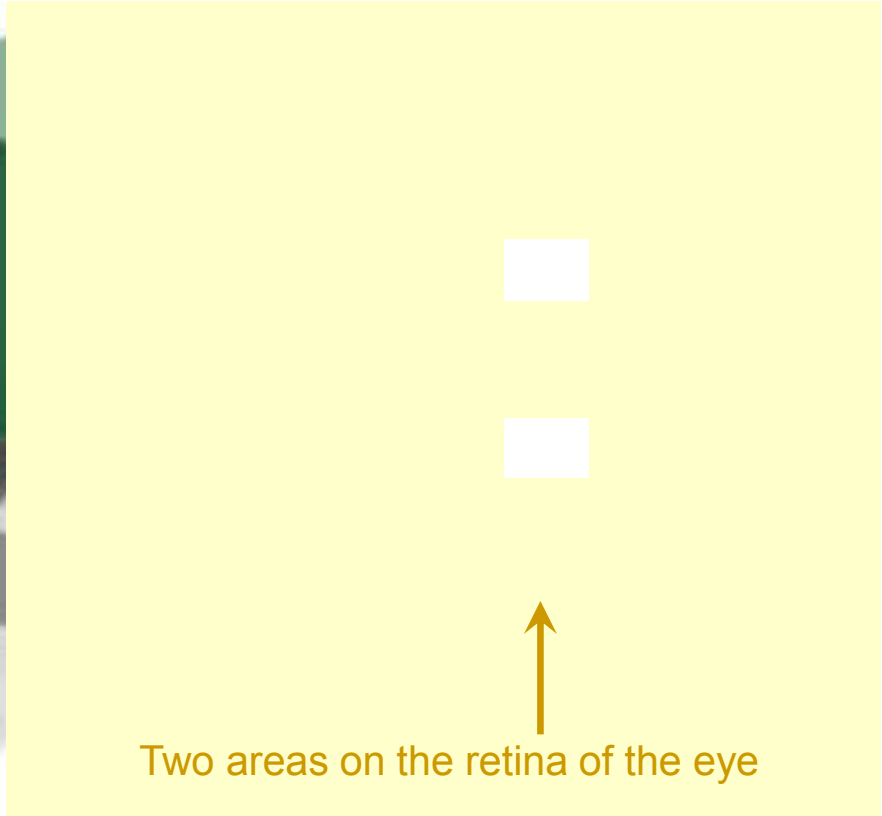
Associative pattern recognition in visual perception



Same lightness of A and B – retina area stimuli without spatial information

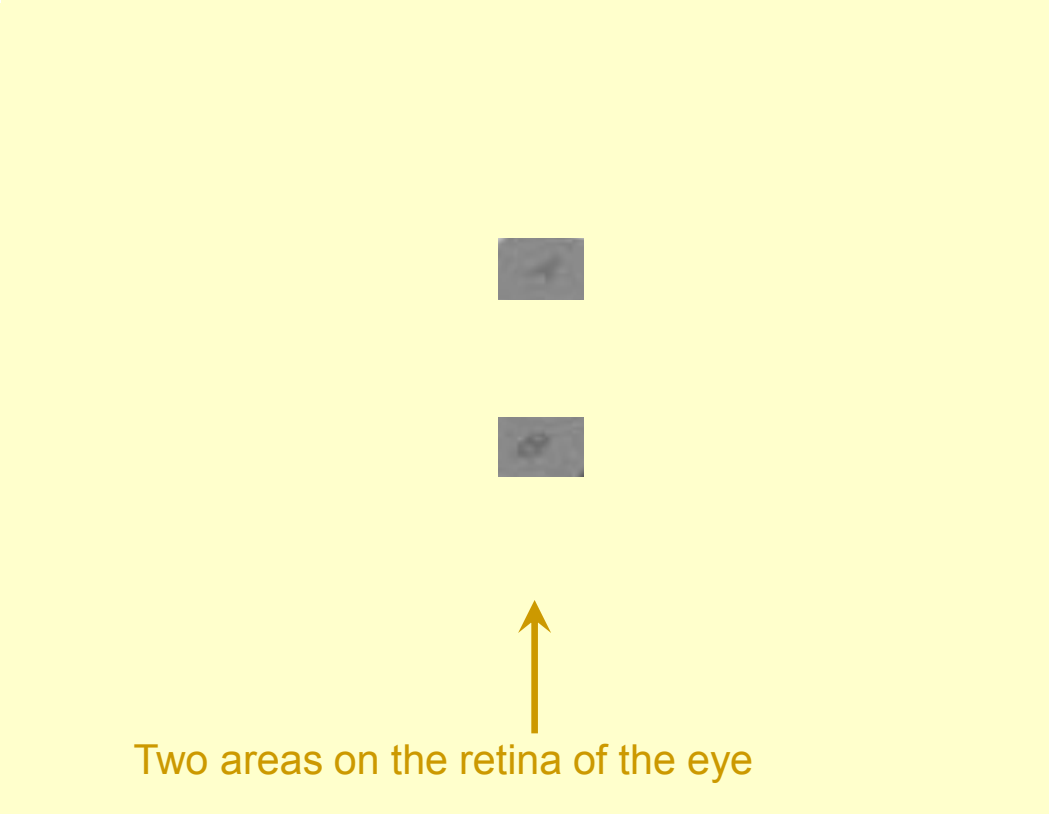
Associative pattern recognition in visual perception





Same lightness of A and B – retina area stimuli without spatial information

Associative pattern recognition in visual perception



Same lightness of A and B – retina area stimuli without spatial information

- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- Head tracking
- Binaural Room Synthesis (BRS)
- Virtual headphone

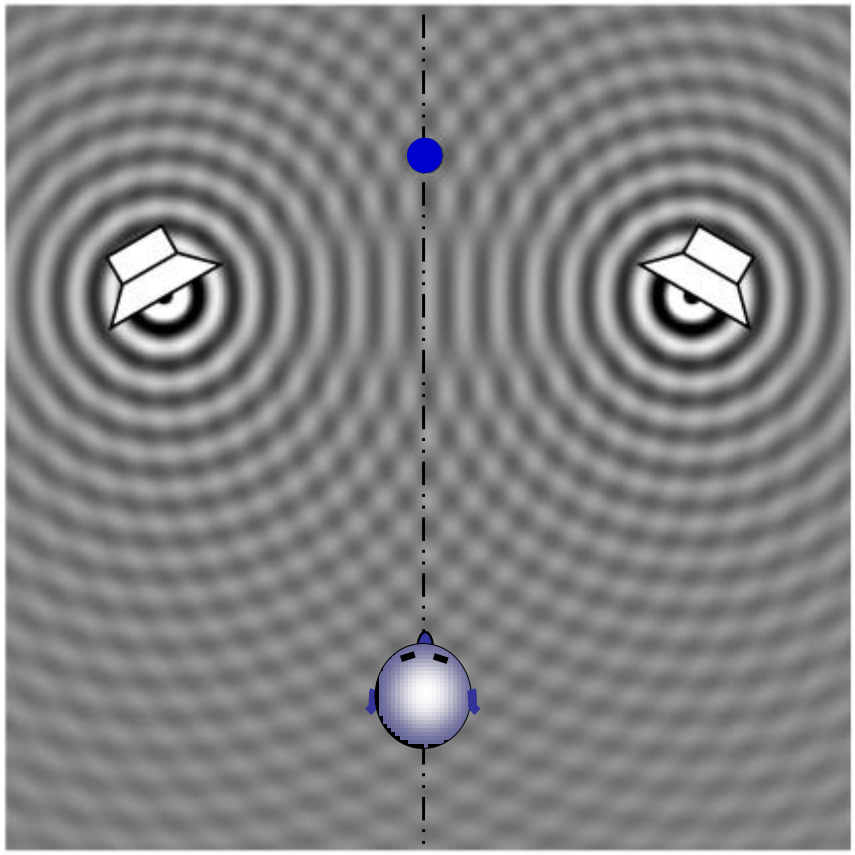
Spatial transmission methods

- (Multi-channel) loudspeaker stereophonic representation
- Multi-channel loudspeaker representation of single sources
- Sound field reconstruction of acoustic scenes
- Binaural reconstruction of the ear input signals

Hybrid systems can combine advantages of both methods

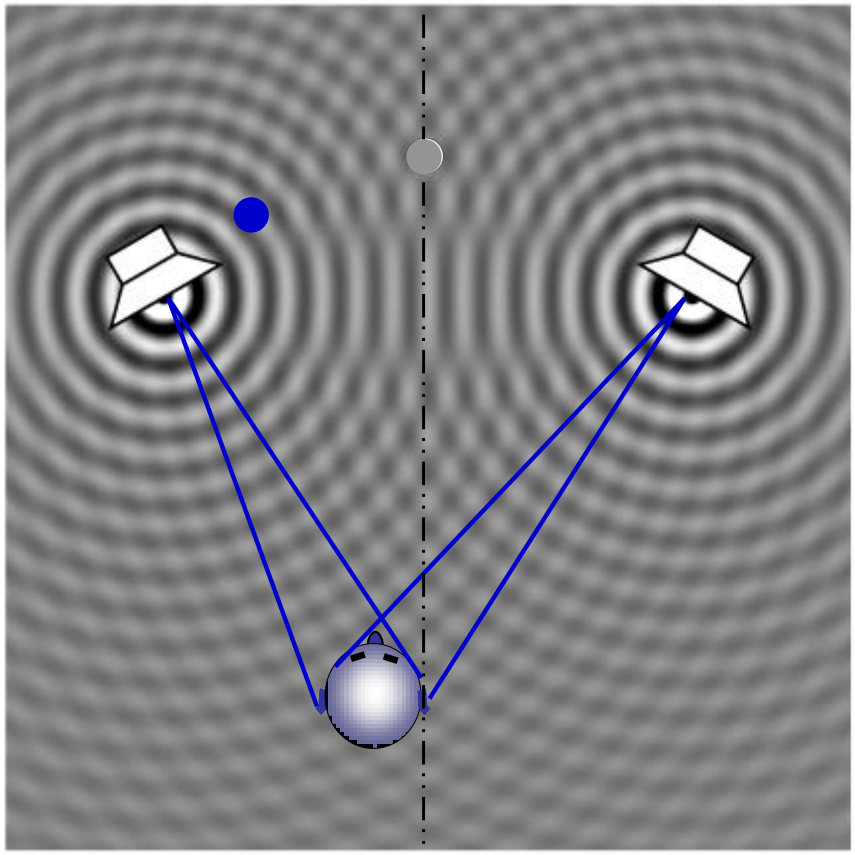
Two-channel stereo, interchannel time difference Δ t = 0 ms

Phantom sound source



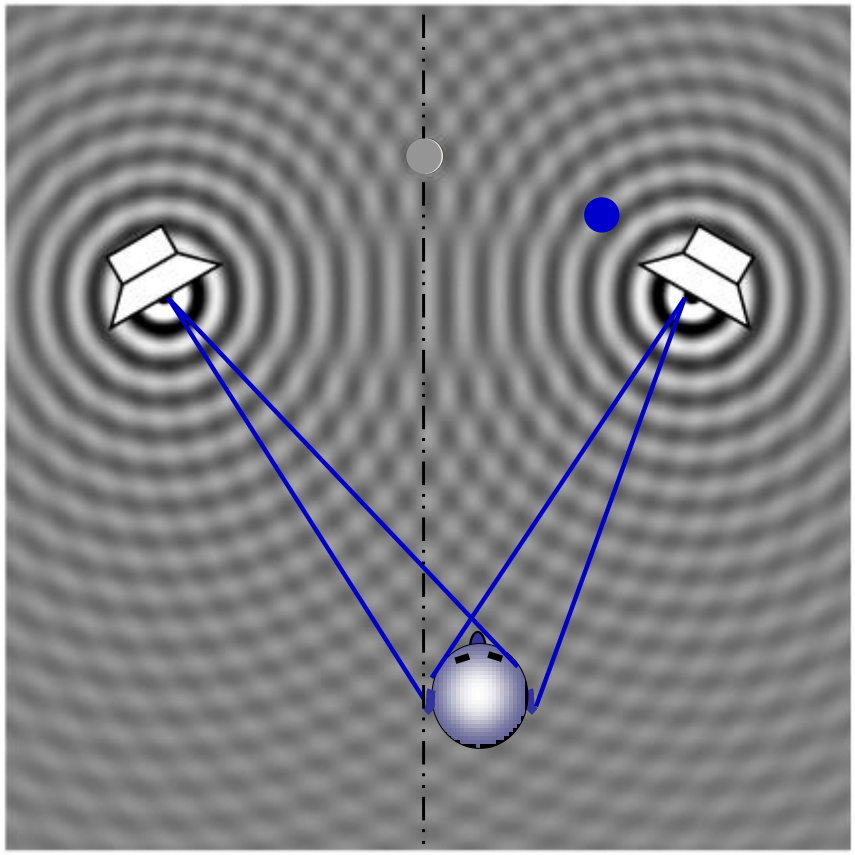
Two-channel stereo, interchannel time difference $\Delta t = -0{,}5$ ms

Phantom sound source, – 25°



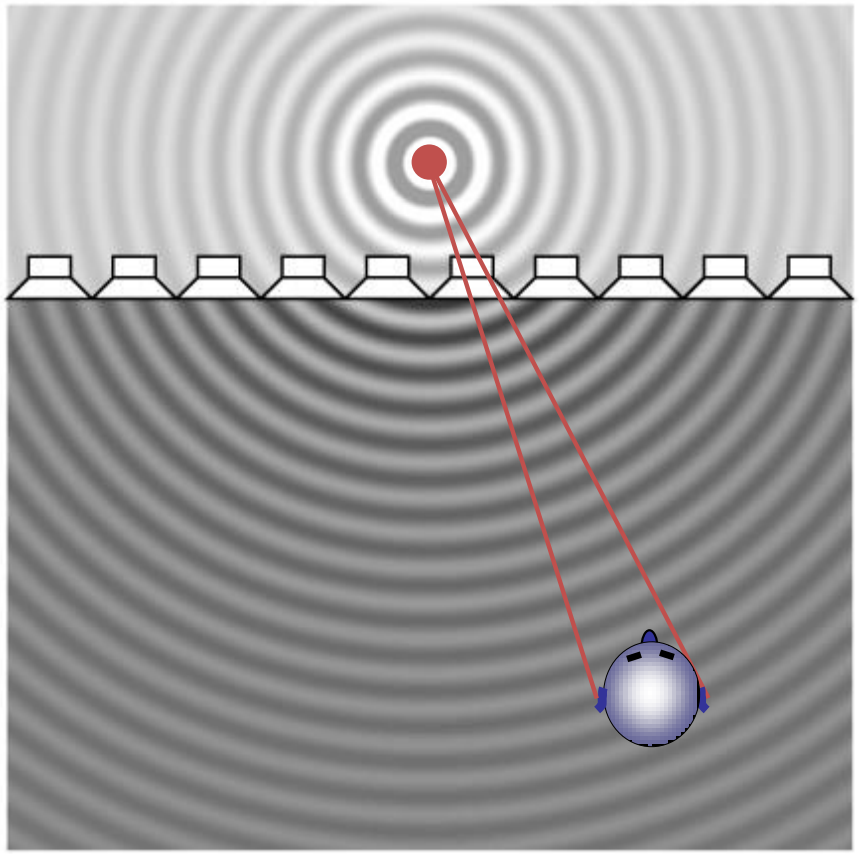
Two-channel stereo, interchannel time difference Δ t = + 0,5 ms

Phantom sound source, + 25°



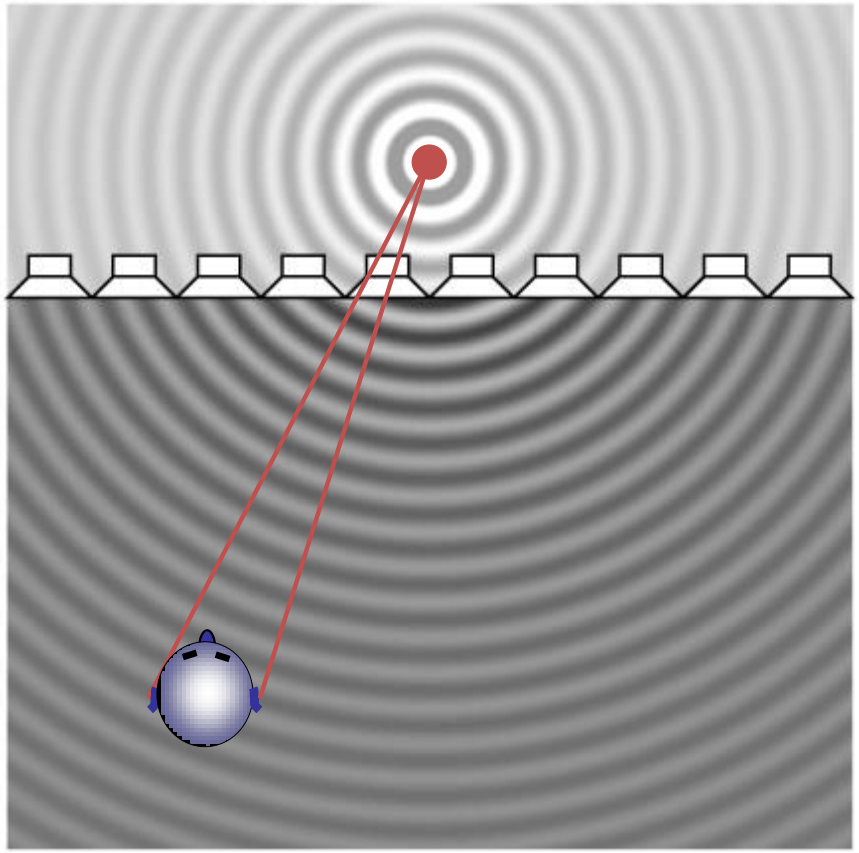
Virtual sound source, basic principal

Virtual sound source, – 25°



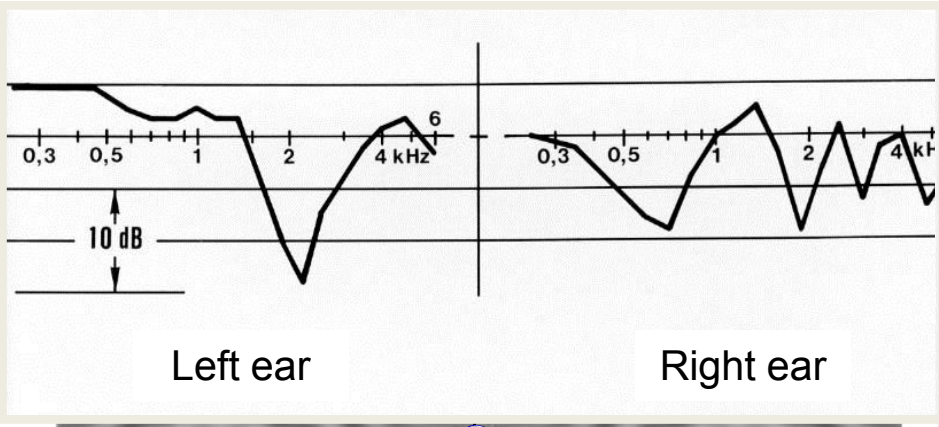
Virtual sound source, basic principal

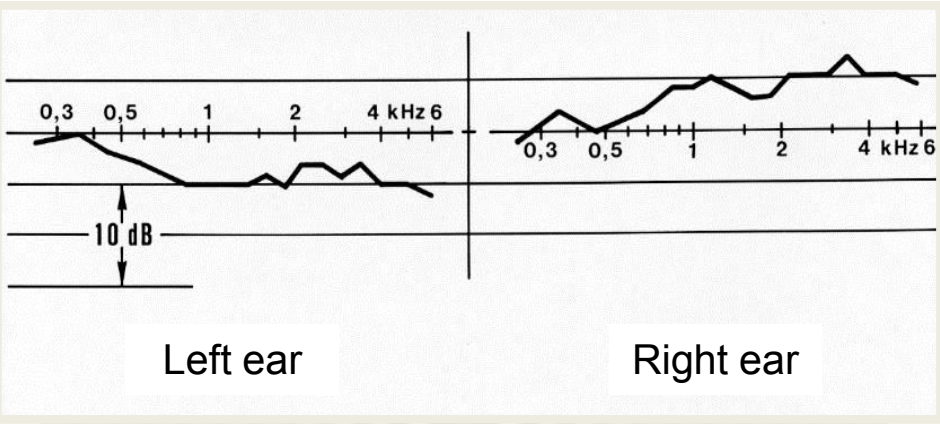
Virtual sound source, + 25°

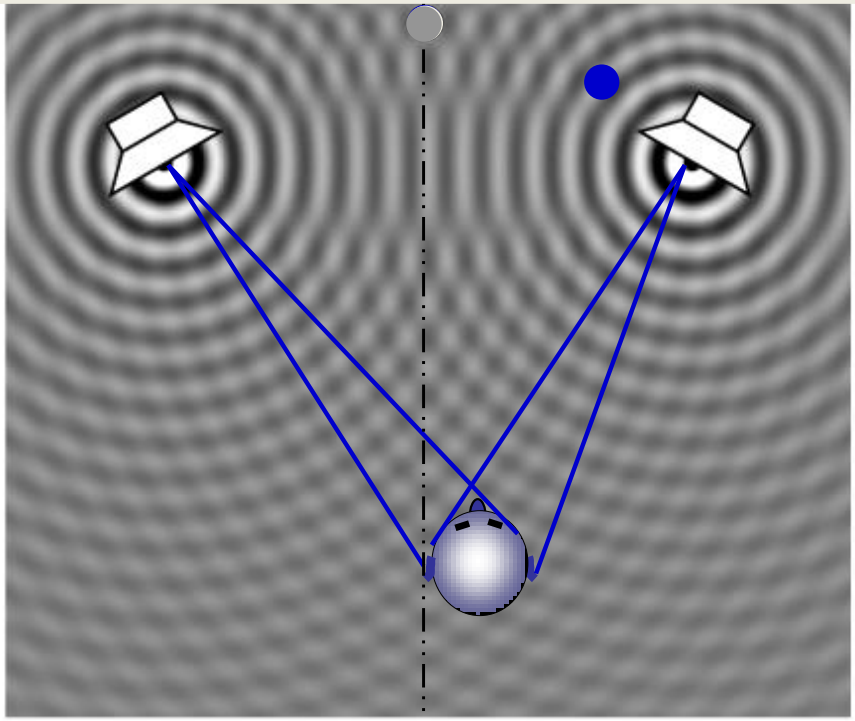


Stereophony vs. sound field synthesis

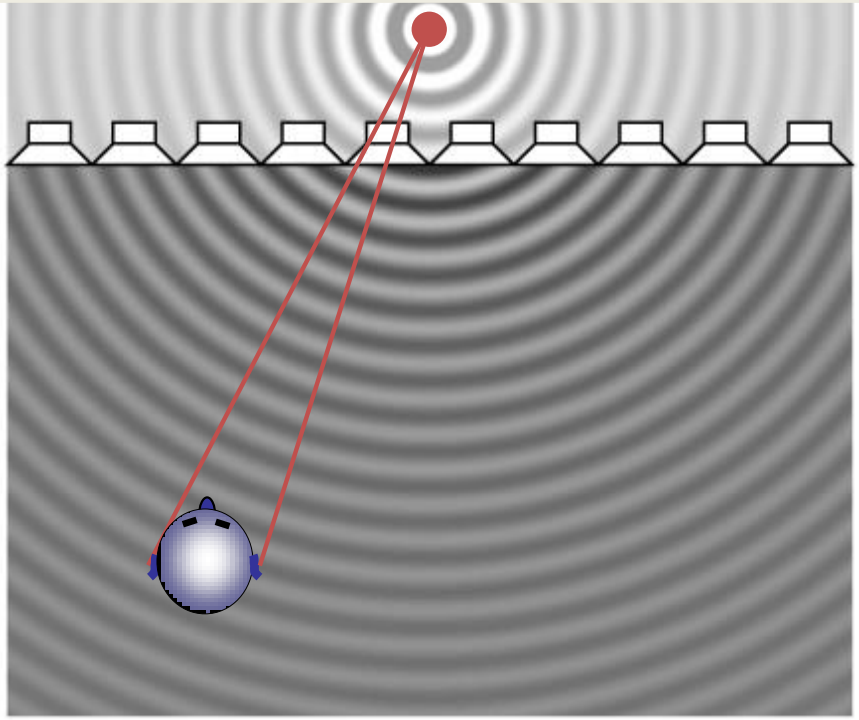
Perceived direction + 25°







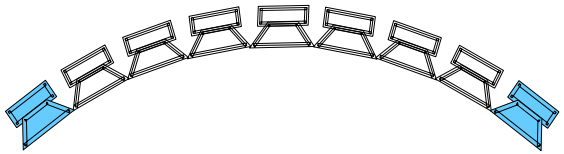
Phantom sound source



Virtual sound source

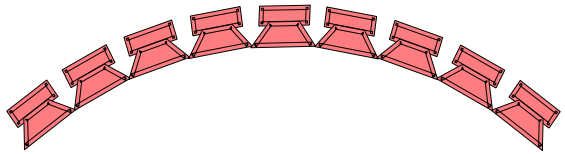
Source rendering

Phantom sound source



Two loudspeakers
produce a stereophonic effect

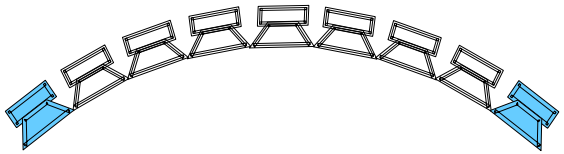
Virtual sound source



A loudspeaker array
produces a natural sound field

Source rendering

Phantom sound source

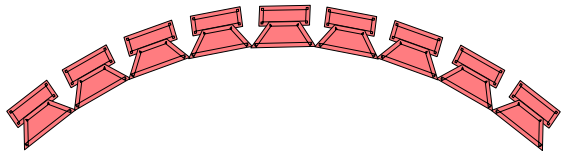


Phsychoacoustic approach

Guideline:

Utilization of stereophonic phenomena
as perfect as possible

Virtual sound source



Physical approach

Guideline:

Reproduction of binaural ear signals
as perfect as possible

Wave Field Synthesis (WFS) ⟷ Ambisonics (HOA)

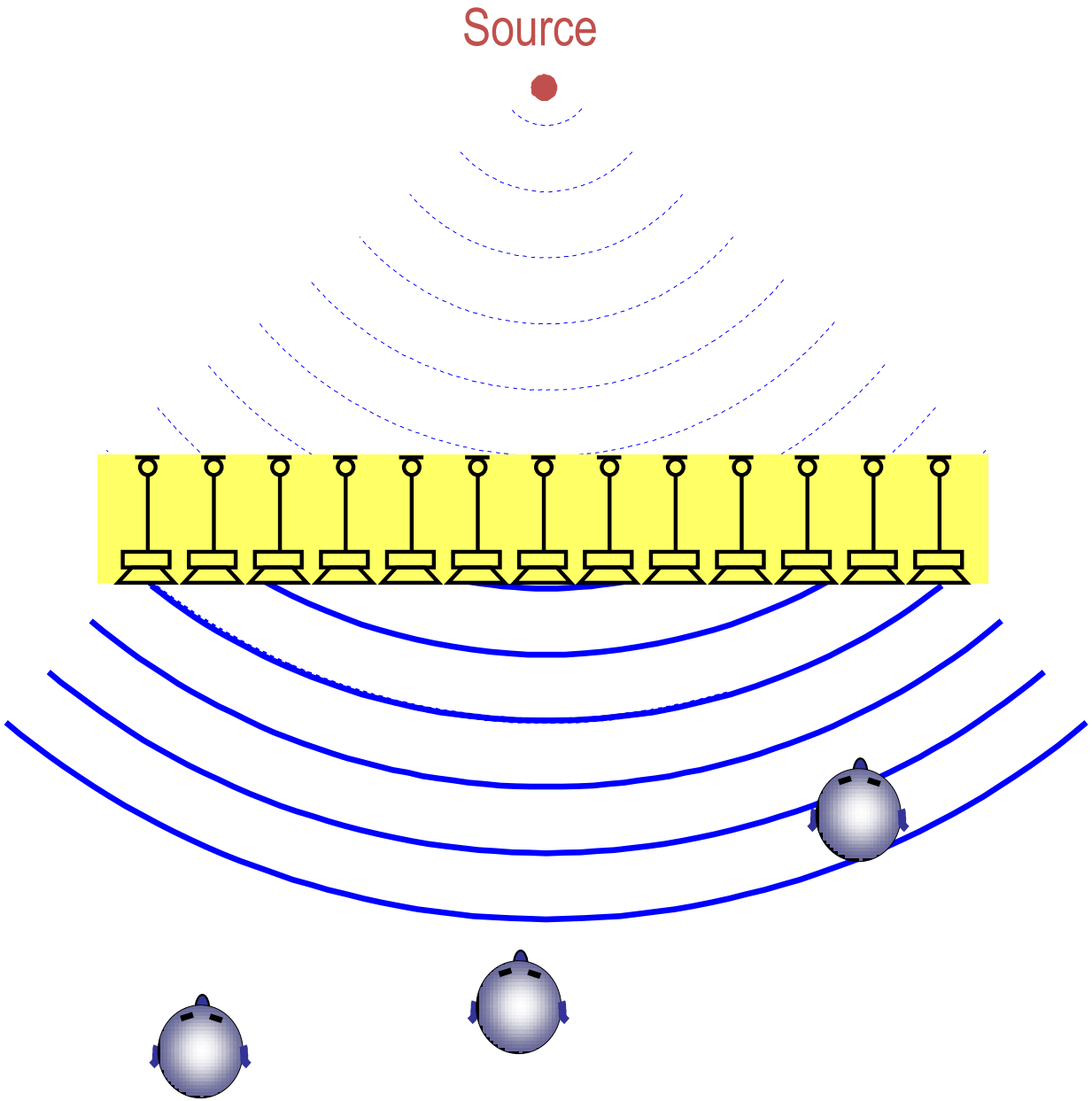
Common mathematical description of the desired sound field

Different technical compromises due to practical constraints:

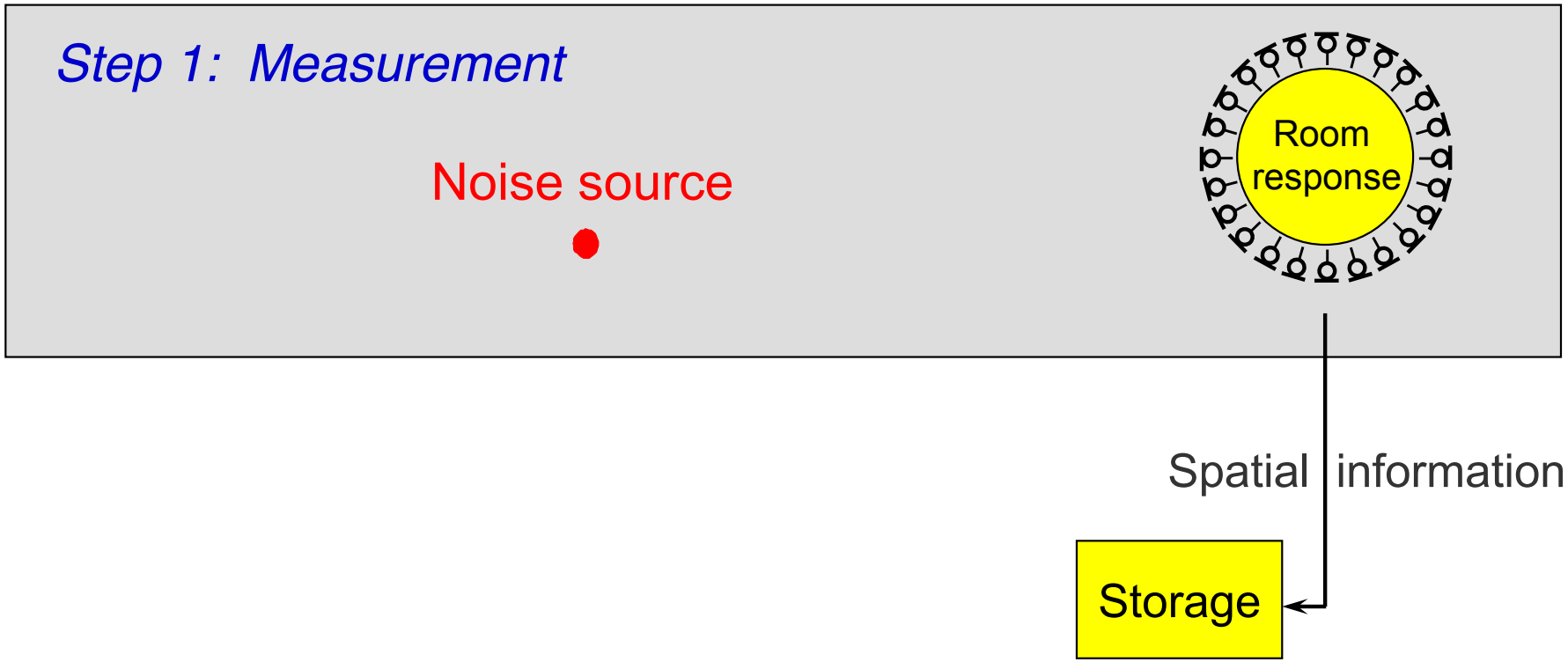
WFS	HOA
Listening area depends on array size	Listening area depends on number of array loudspeakers
Alias frequency depends on distances of array loudspeakers	No alias frequency

Virtual sound source, basic principal

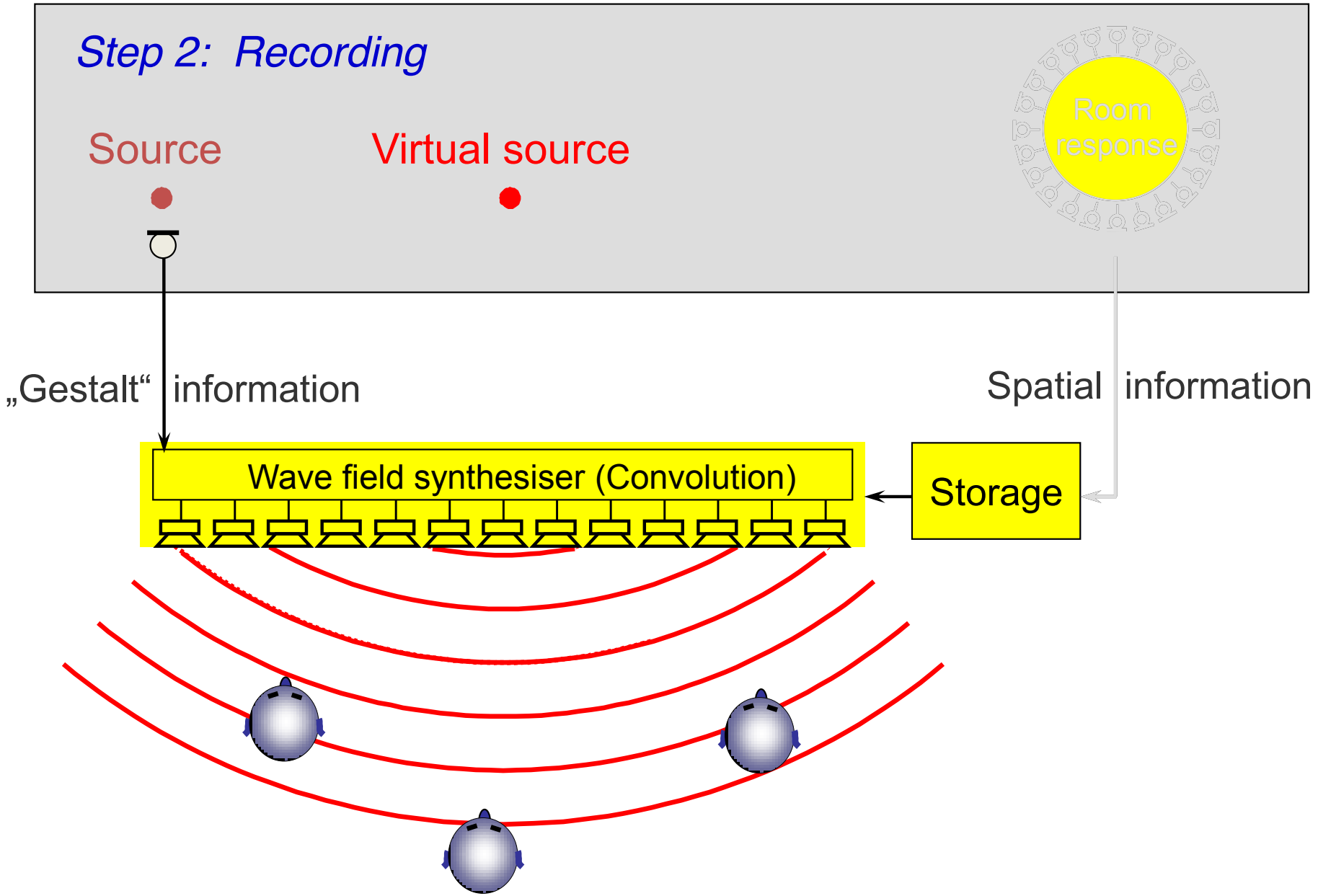
„Acoustic curtain“



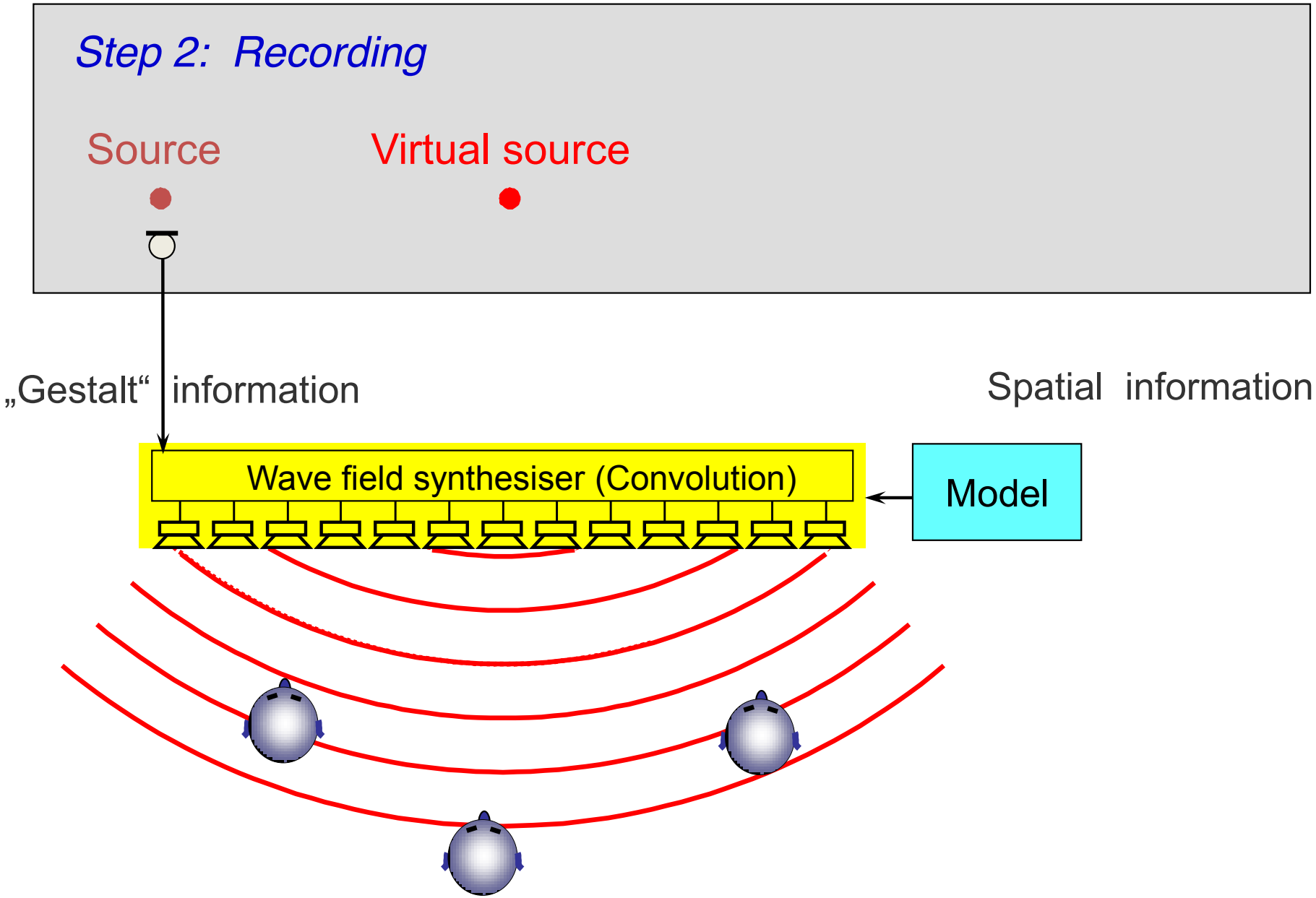
Object-based capturing / rendering (example WFS)



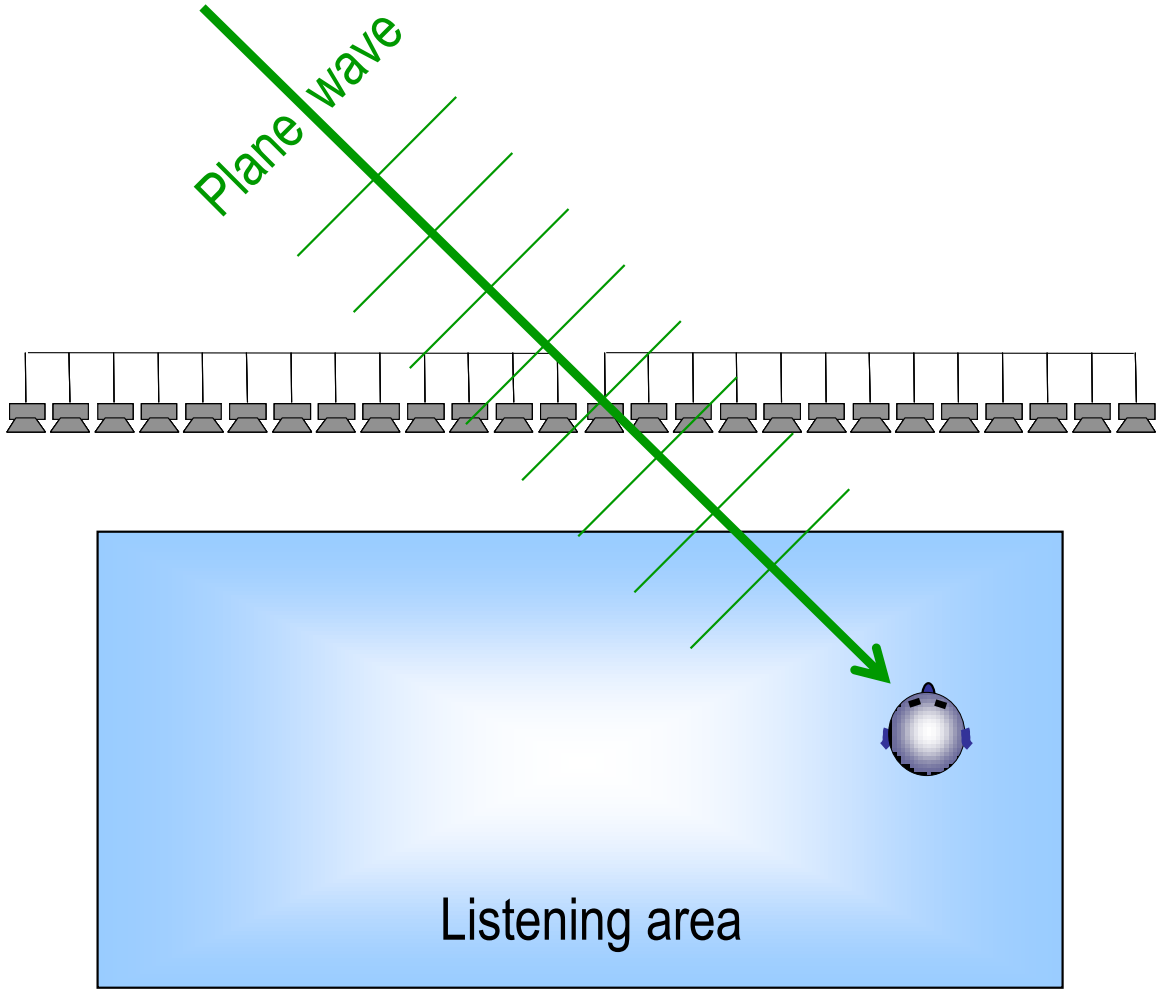
Object-based capturing / rendering (example WFS)



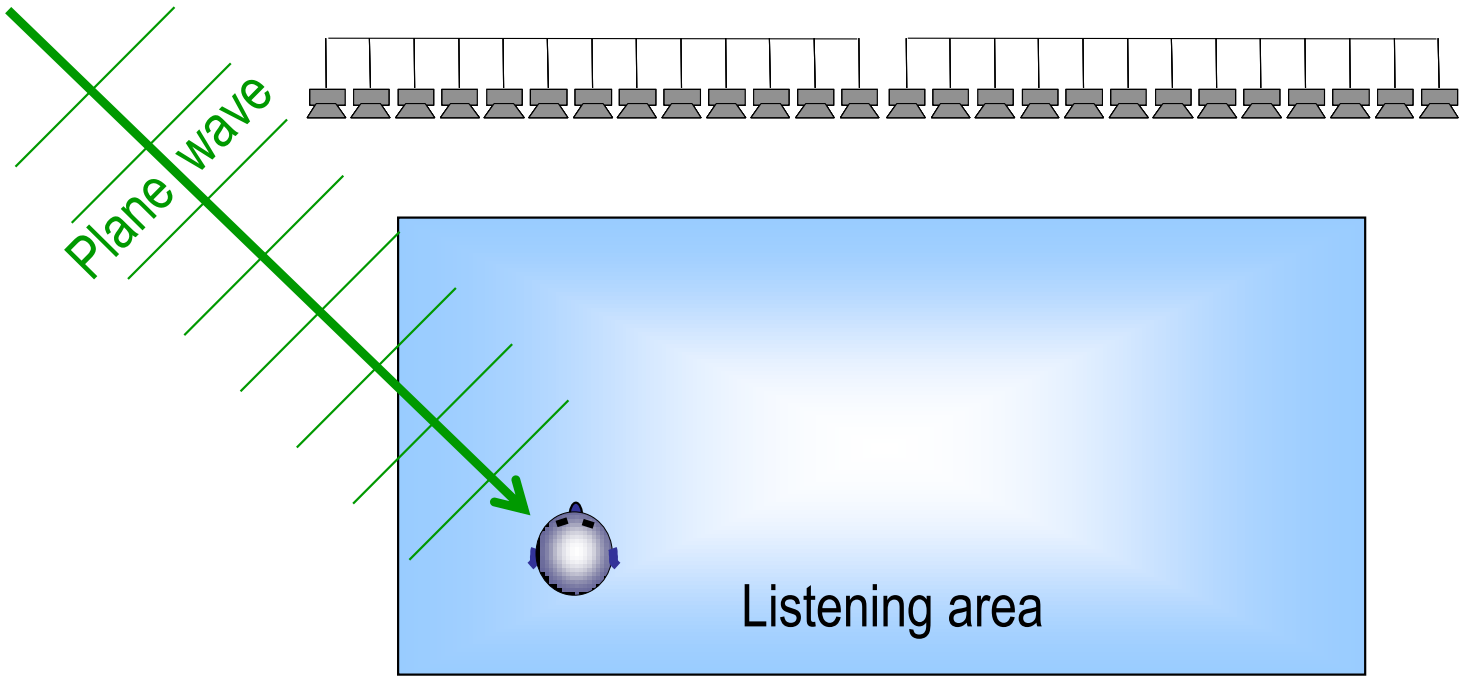
Object-based capturing / rendering (example WFS)



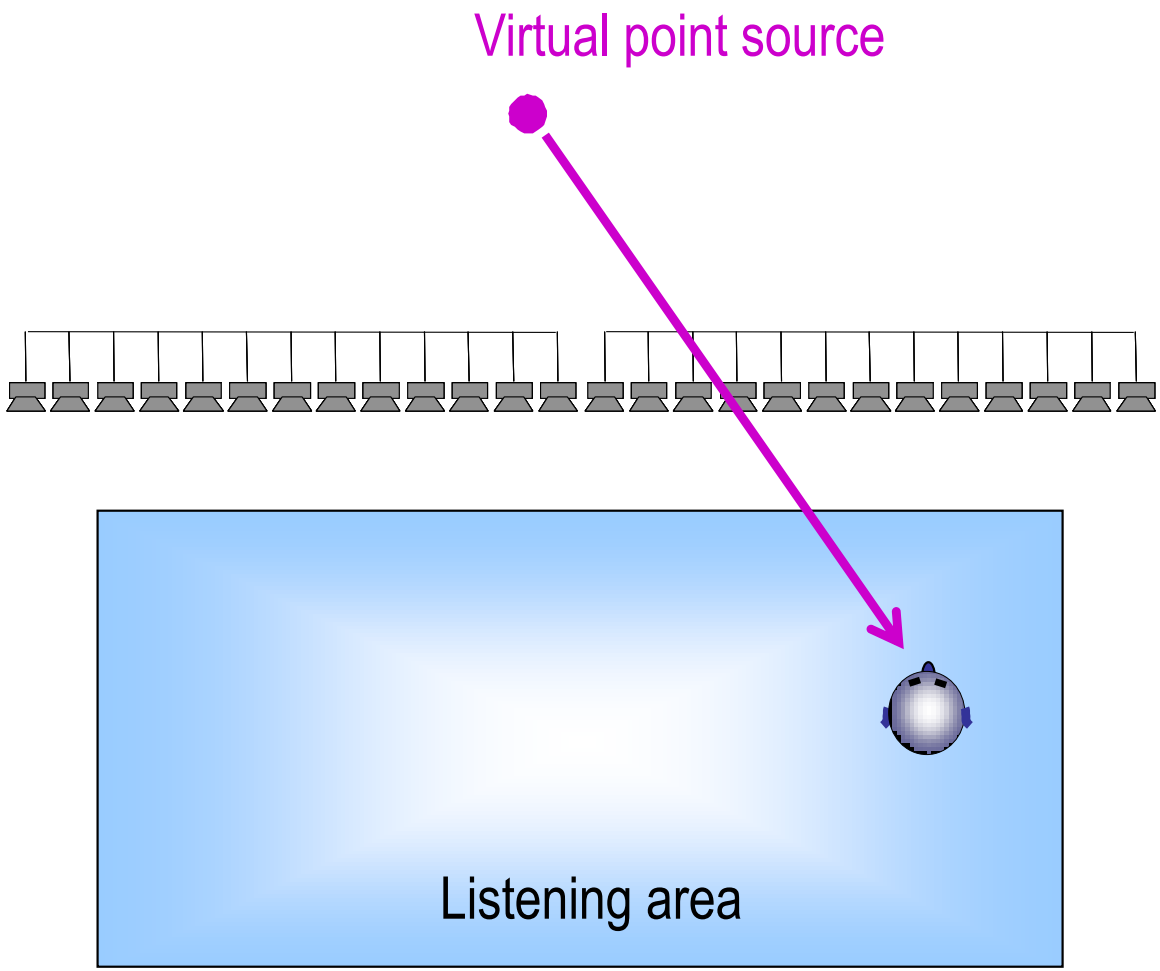
Rendering a perspective acoustic scene



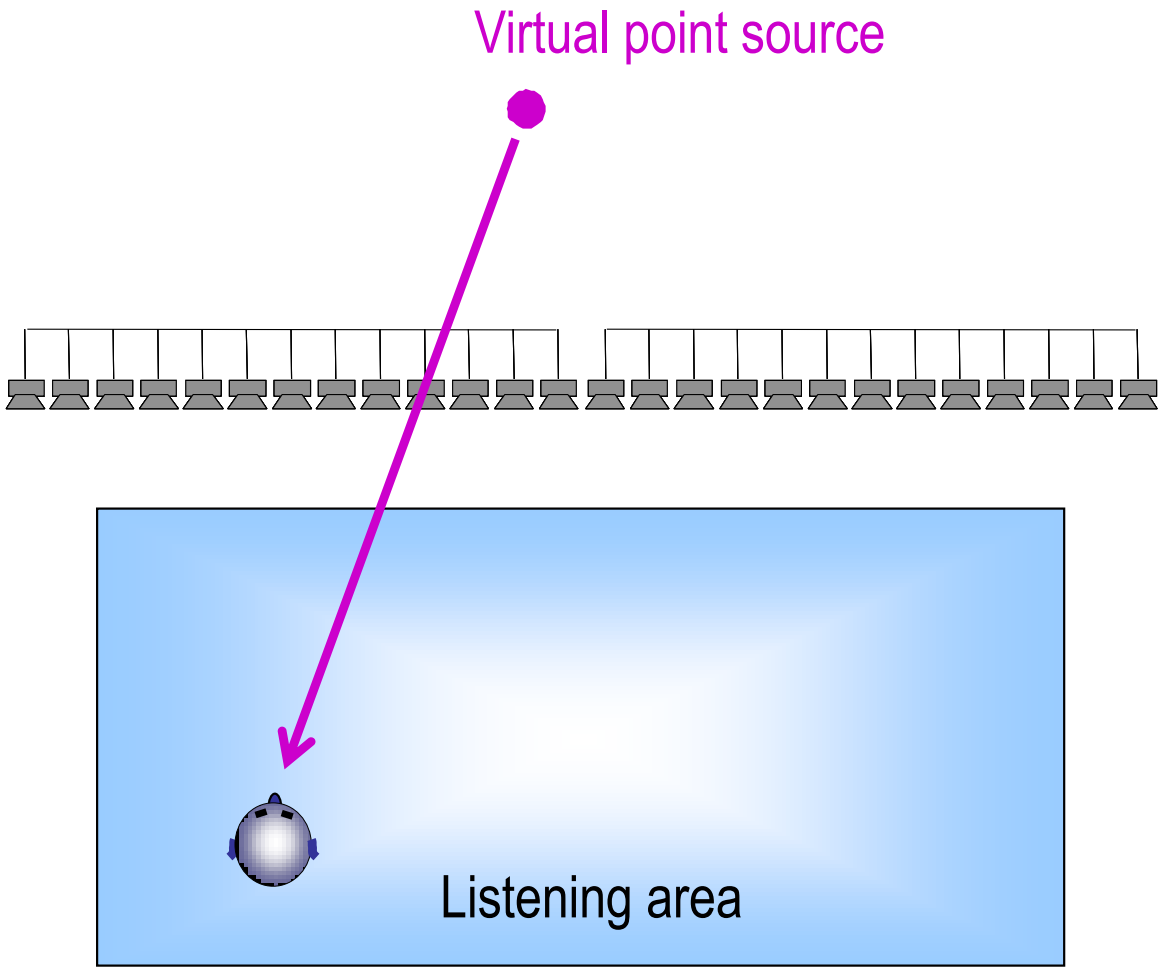
Rendering a perspective acoustic scene



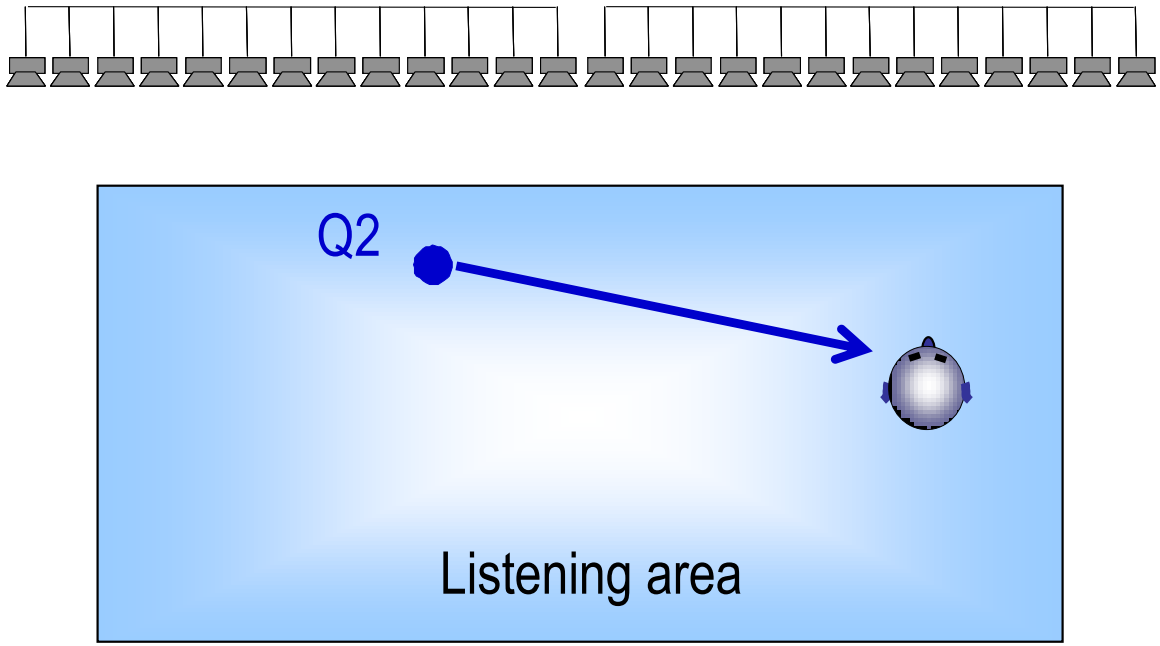
Rendering a perspective acoustic scene



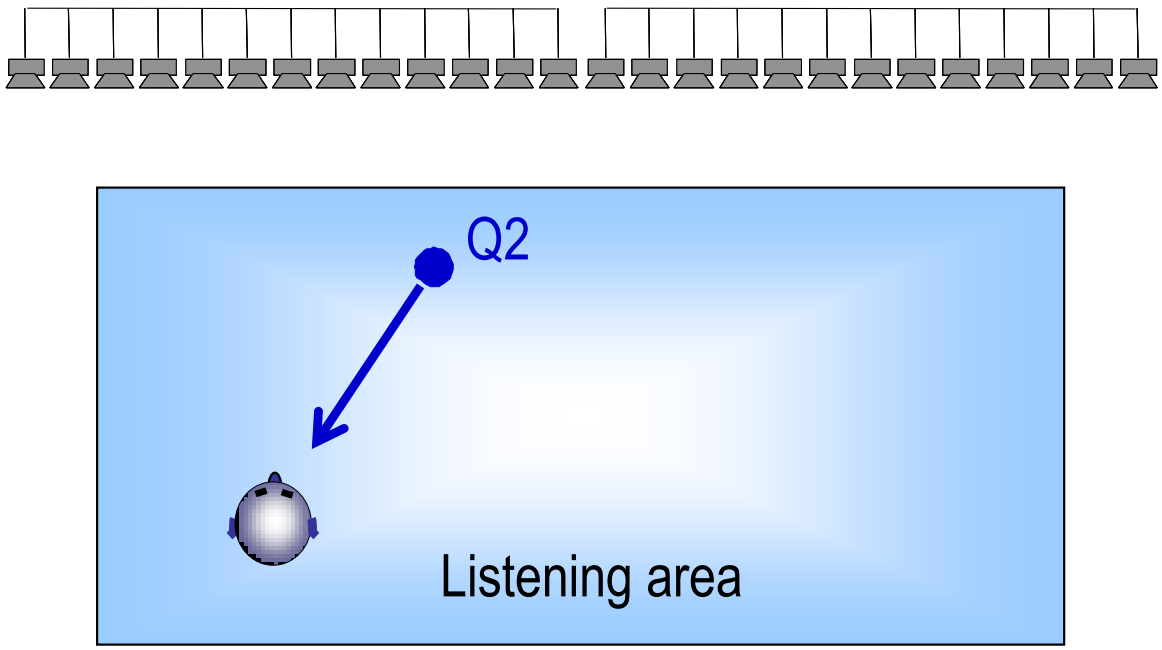
Rendering a perspective acoustic scene



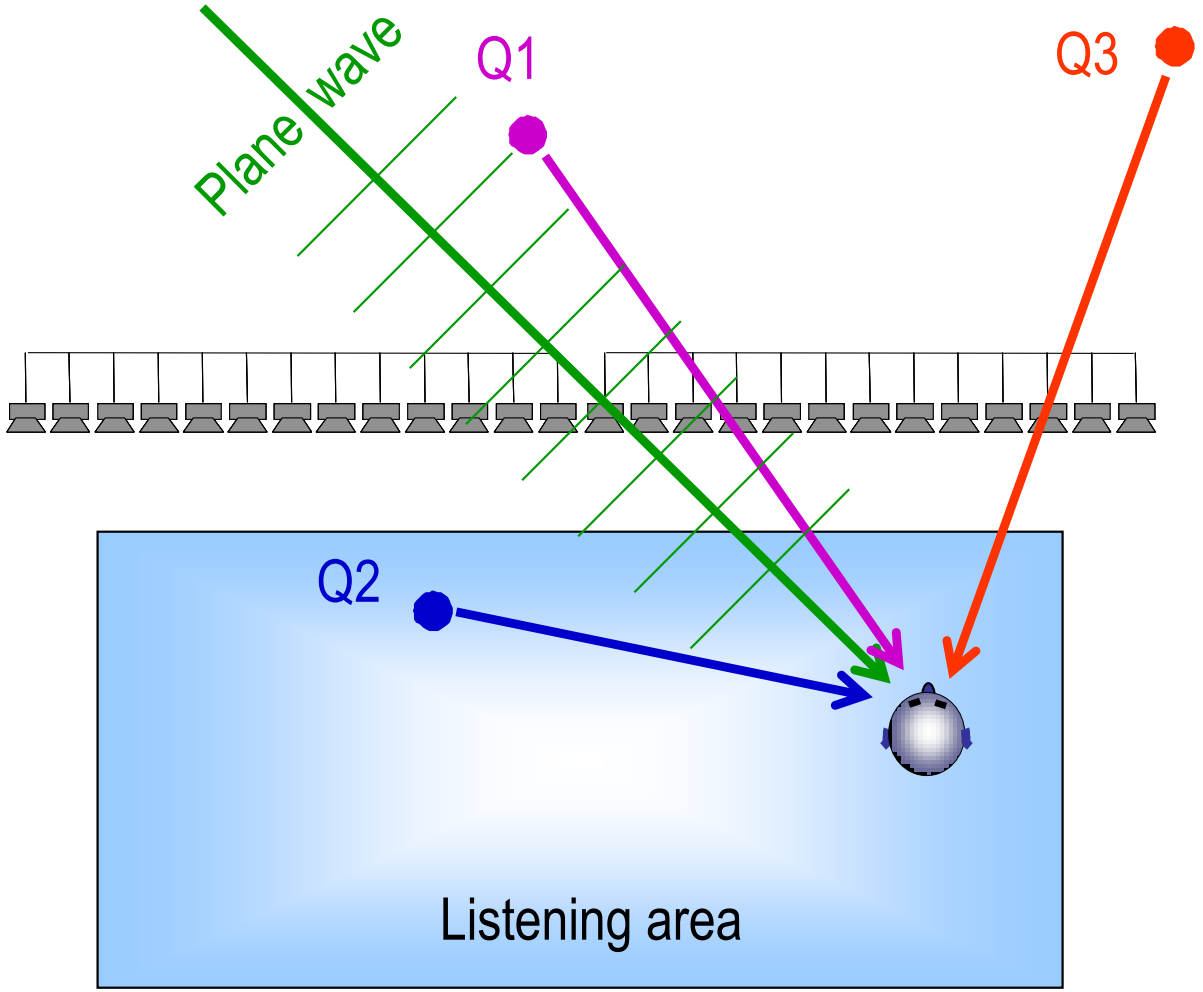
Rendering a perspective acoustic scene



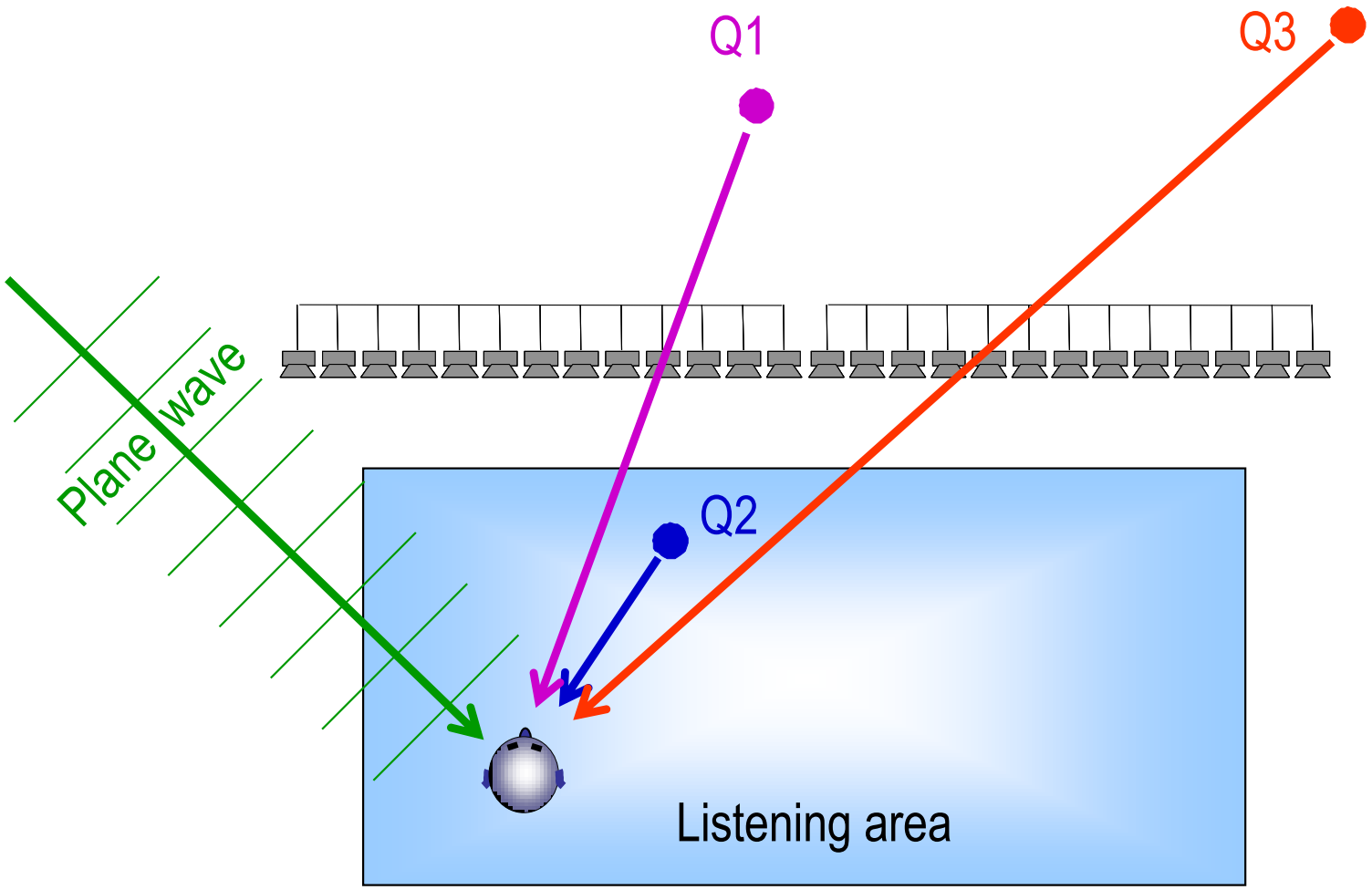
Rendering a perspective acoustic scene



Rendering an intra-active perspective acoustic scene



Rendering an intra-active perspective acoustic scene



Verband Deutscher Tonmeister e.V.

Comparison of attribute profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 22.2	WFS* (HOA*)	BINAURAL TECHNIQUES
Front direction, robustness	●	●●	●●	●●	●
Surround direction		●	●●	●●	●●
Elevation			●***		●●
Height			●●		●●
Distance / depth	(●)**	●	●	●●	●●
Proximity to the head				(●)	●●
Intra-active perspective				●●	
Spatial impression	(●)**	●	●●	●	●●
Envelopment		●	●●	●	●●
Timbre	●●	●●	●●	●	●●

*horizontal arrays **simulated depth/spatial impression ***unstable; at the sweetspot only

Practical constraints:

- Spatial aliasing, due to the array discretisation
- Spatial interference, due to limitation of array dimensions
- Sensitivity to room acoustics
- Restriction to the horizontal plane
- Source depending transmission capacity, due to object-based recording
- Extended or moving sources, ambience, due to object-based recording
- High technical effort

Practical constraints:

- Spatial aliasing, due to the array discretisation
- Spatial interference, due to limitation of array dimensions
- Sensitivity to room acoustics
- Restriction to the horizontal plane
- Source depending transmission capacity, due to object-based recording
- **Extended or moving sources, ambience, due to object-based recording**
- High technical effort

Constraints due to object-based recording

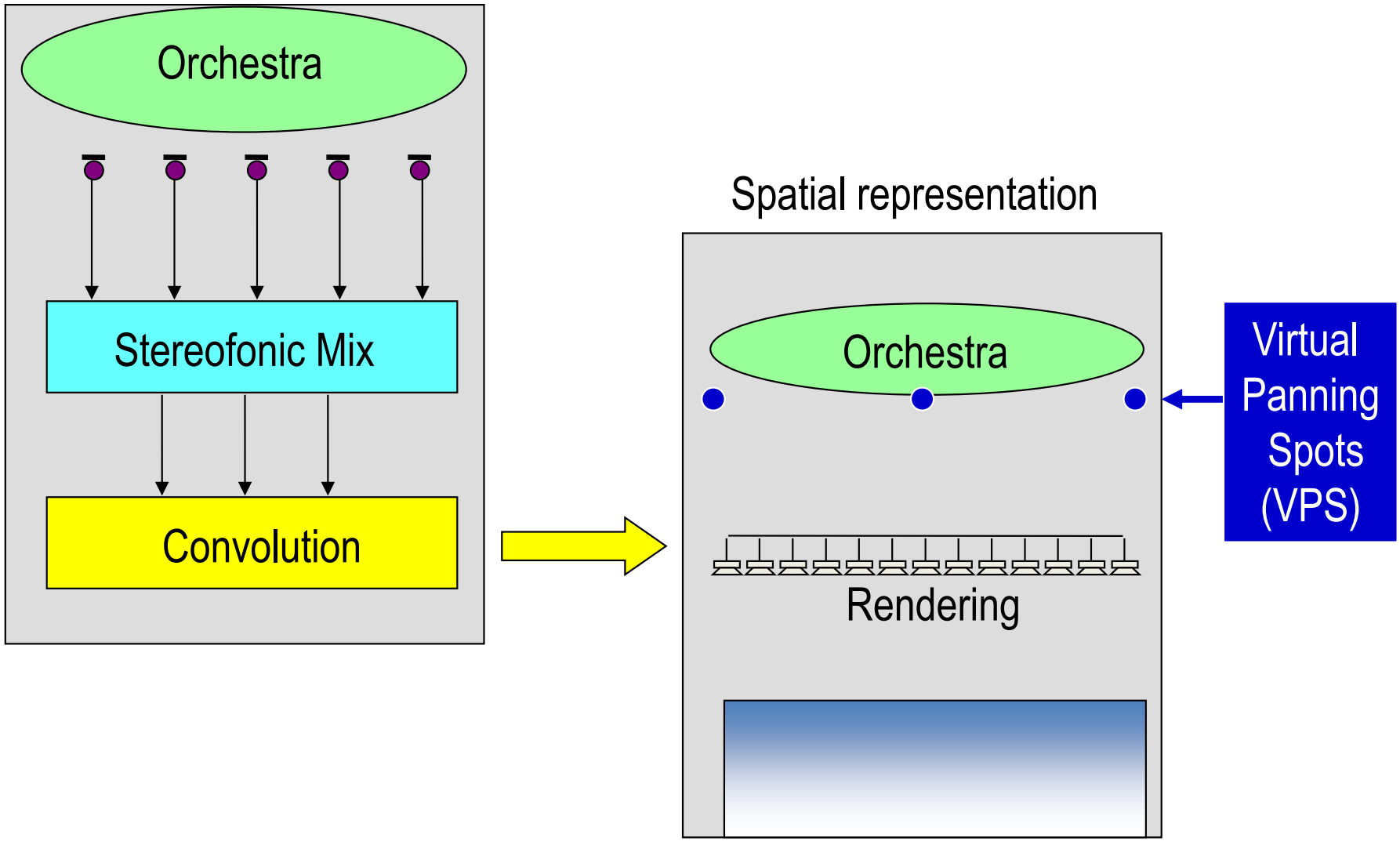
- Imaging of spacious (broad) sound sources
- Imaging of acoustic environment
- Imaging of moving sources
- Source depending transmission capacity

Spatial transmission methods

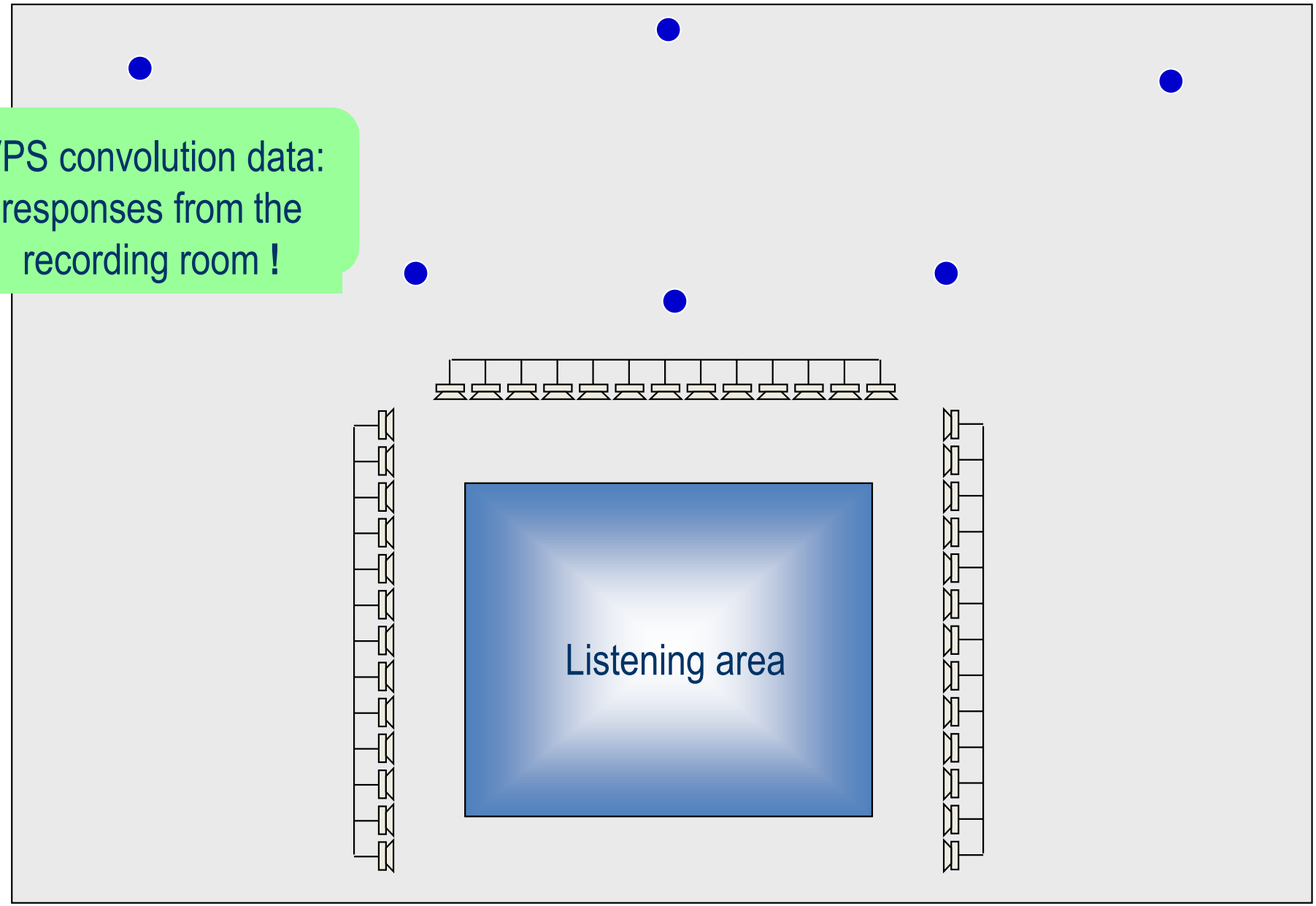
- (Multi-channel) loudspeaker stereophonic representation
- Multi-channel loudspeaker representation of single sources
- Sound field reconstruction of acoustic scenes
- Binaural reconstruction of the ear input signals

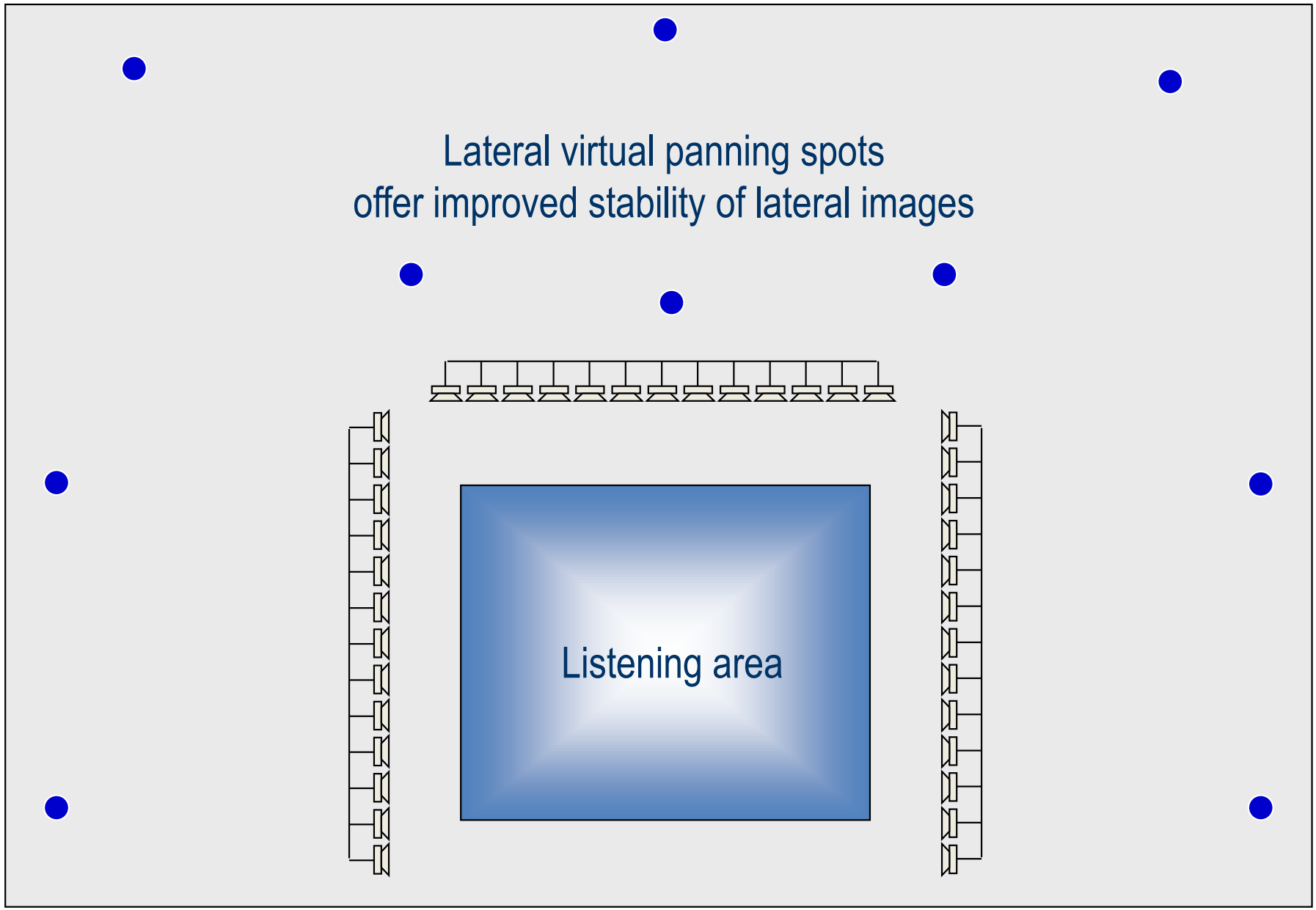
Hybrid systems can combine advantages of both methods

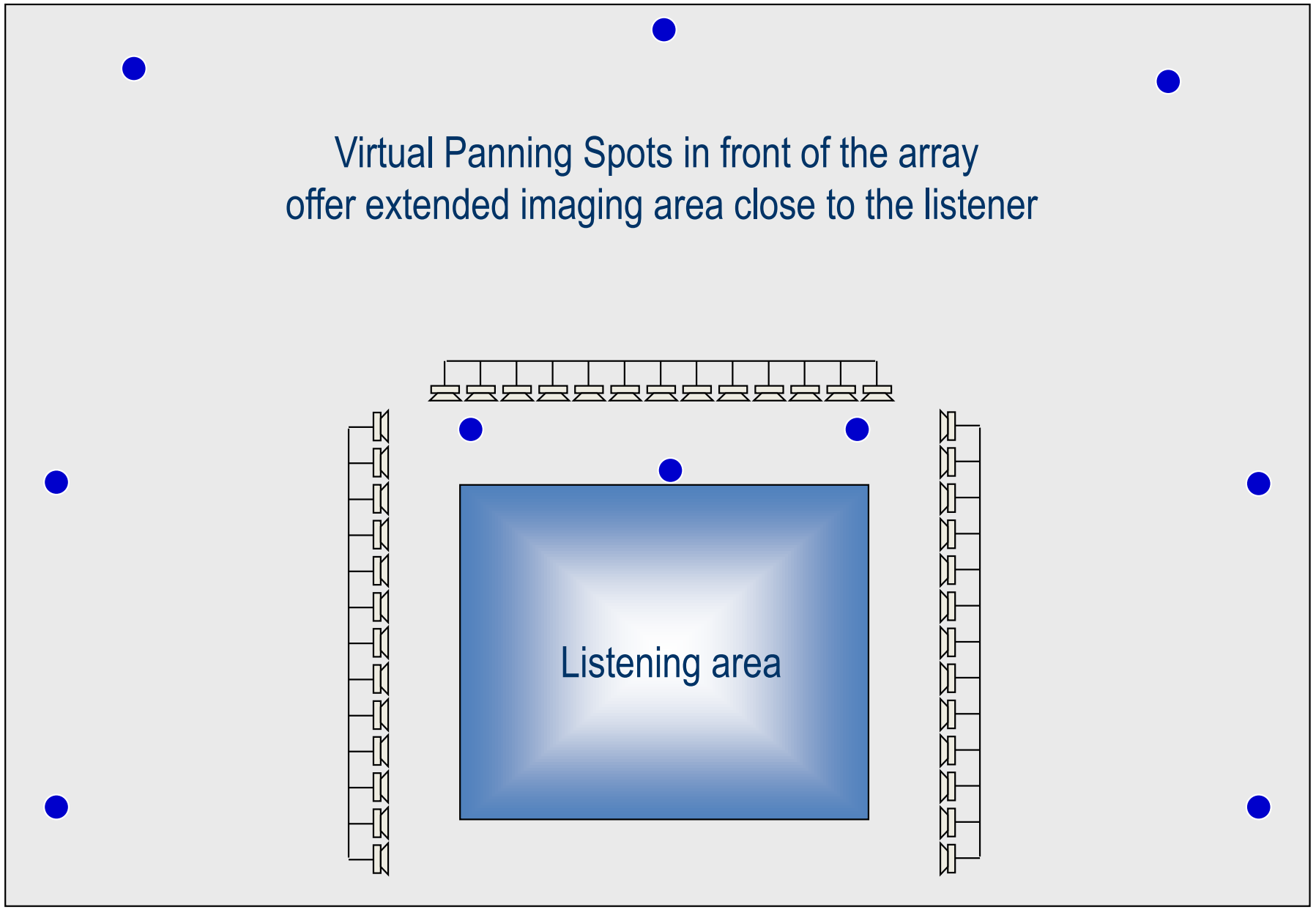
Solution: Virtual Panning Spots



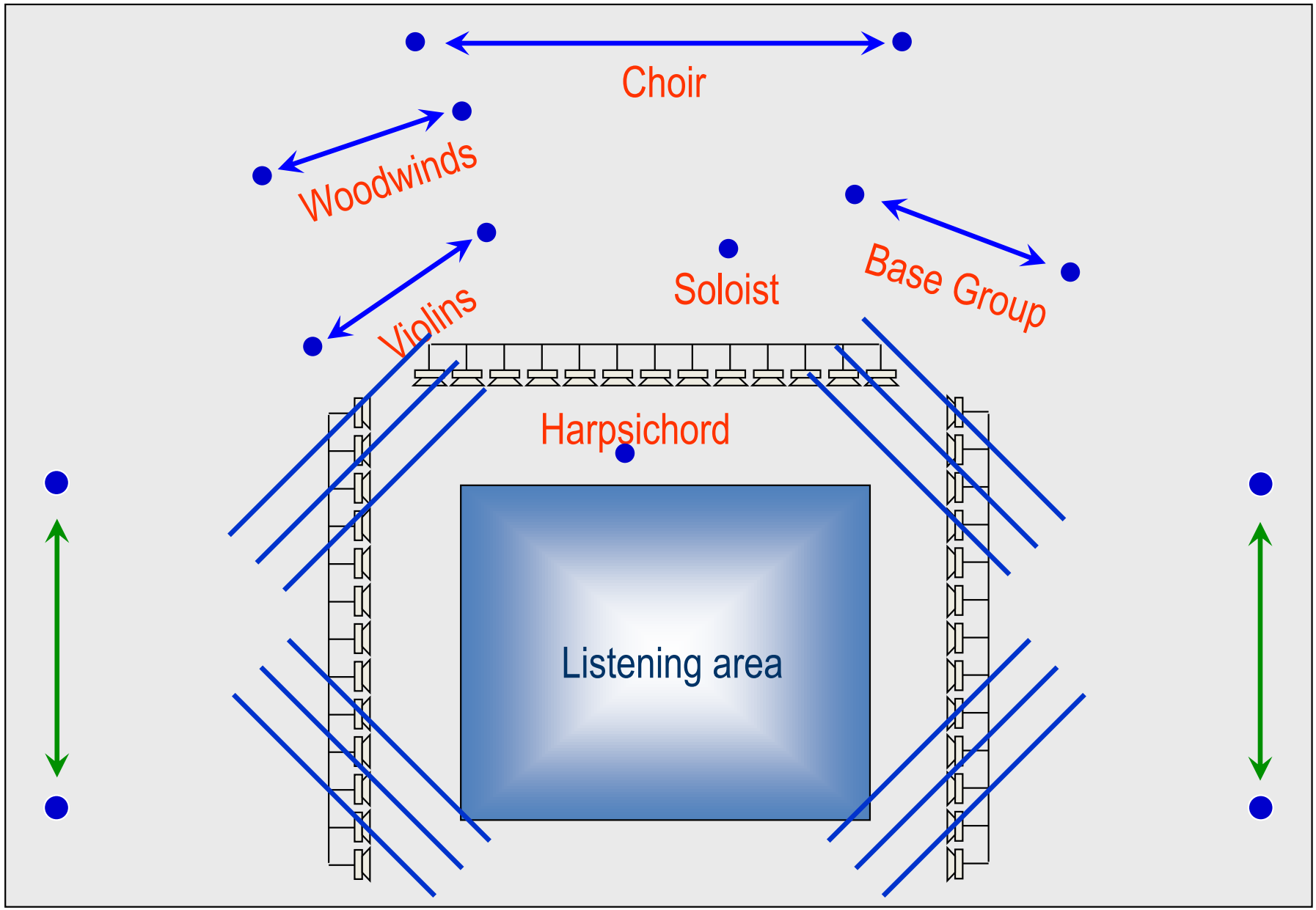
VPS convolution data: responses from the recording room !




Lateral virtual panning spots
offer improved stability of lateral images
Listening area


Virtual Panning Spots in front of the array
offer extended imaging area close to the listener
Listening area

Wave Field Synthesis

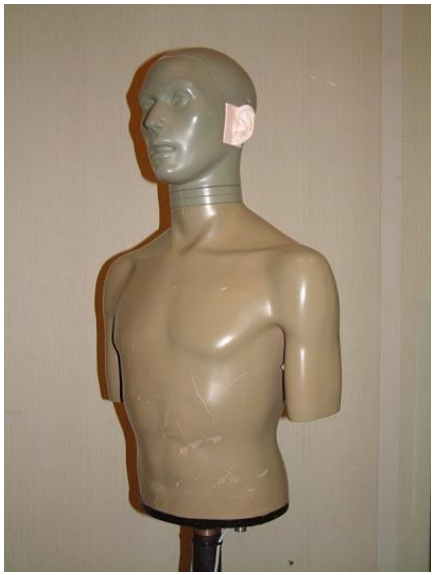

Choir
Woodwinds
Violins
Soloist
Base Group
Harpsichord
Listening area

Spatial transmission methods

- (Multi-channel) loudspeaker stereophonic representation
- Multi-channel loudspeaker representation of single sources
- Sound field reconstruction of acoustic scenes
- Binaural reconstruction of the ear input signals

Dummy heads, actual models



HMS III - HEAD acoustics

HUGO - RWTH Aachen

MANIKIN MK1 - Neutrik-Cortex

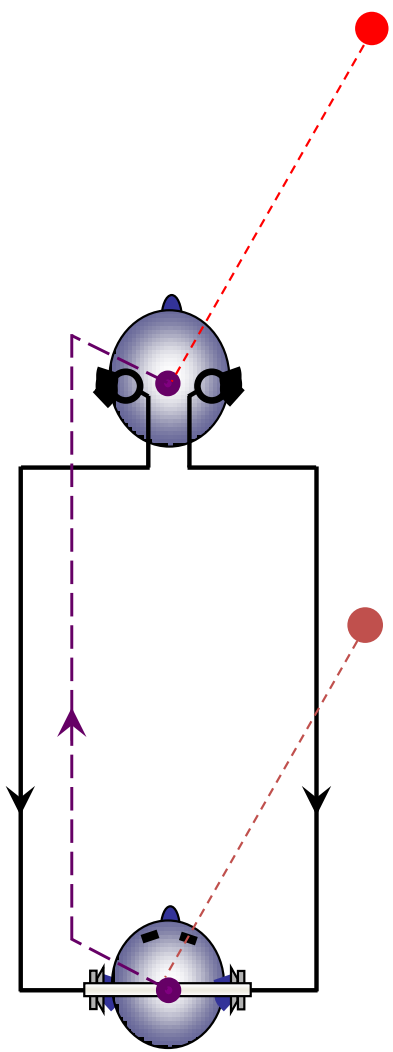
KEMAR - Knowles

KU 81 - Neumann

KU 100 - Neumann

- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- **Head tracking**
- Binaural Room Synthesis (BRS)
- Virtual headphone

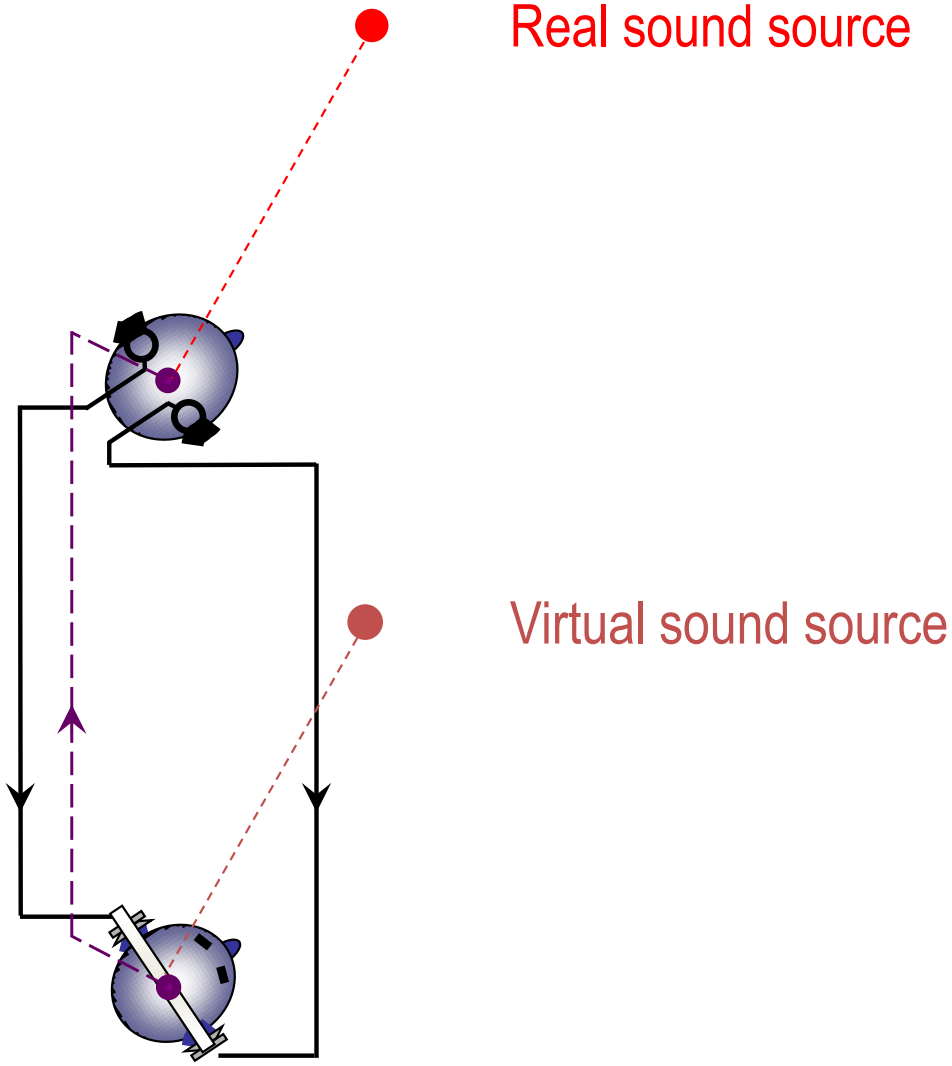
Dummy head (head tracking required!)



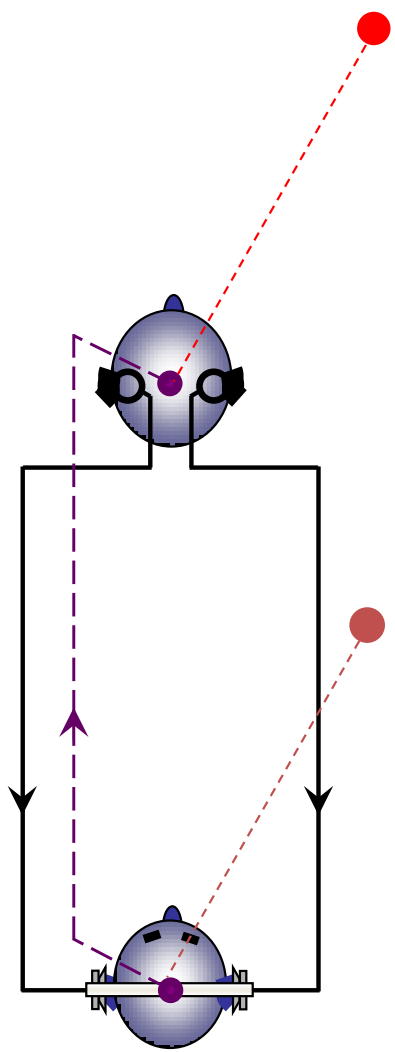
Real sound source

Virtual sound source

Dummy head (head tracking required!)



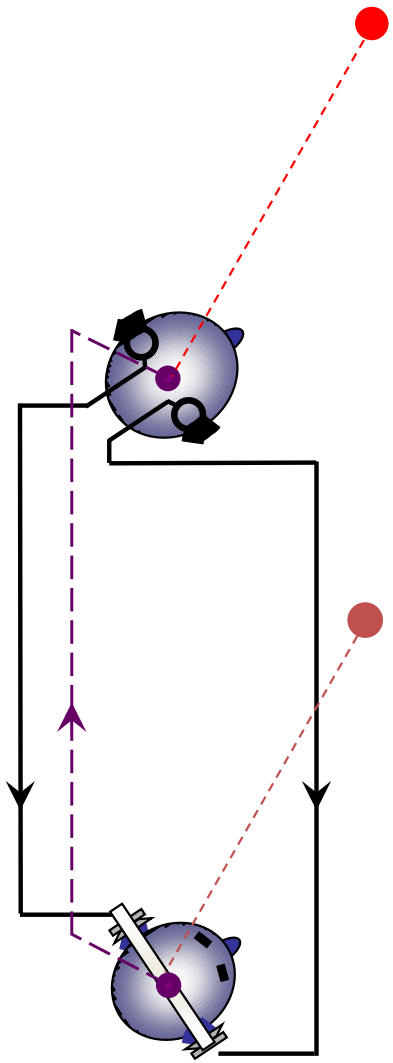
Dummy head (head tracking required!)



Real sound source

Virtual sound source

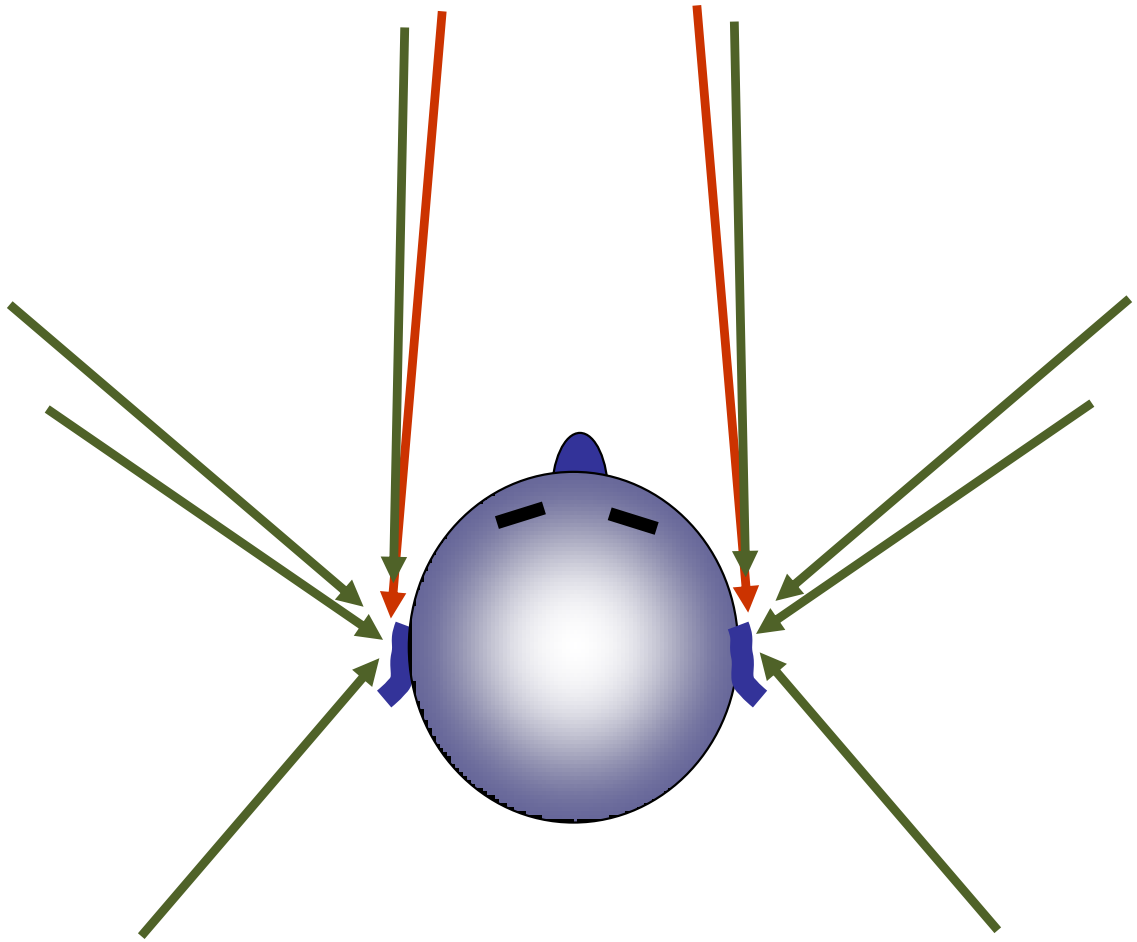
Test: Importance of dynamic cues



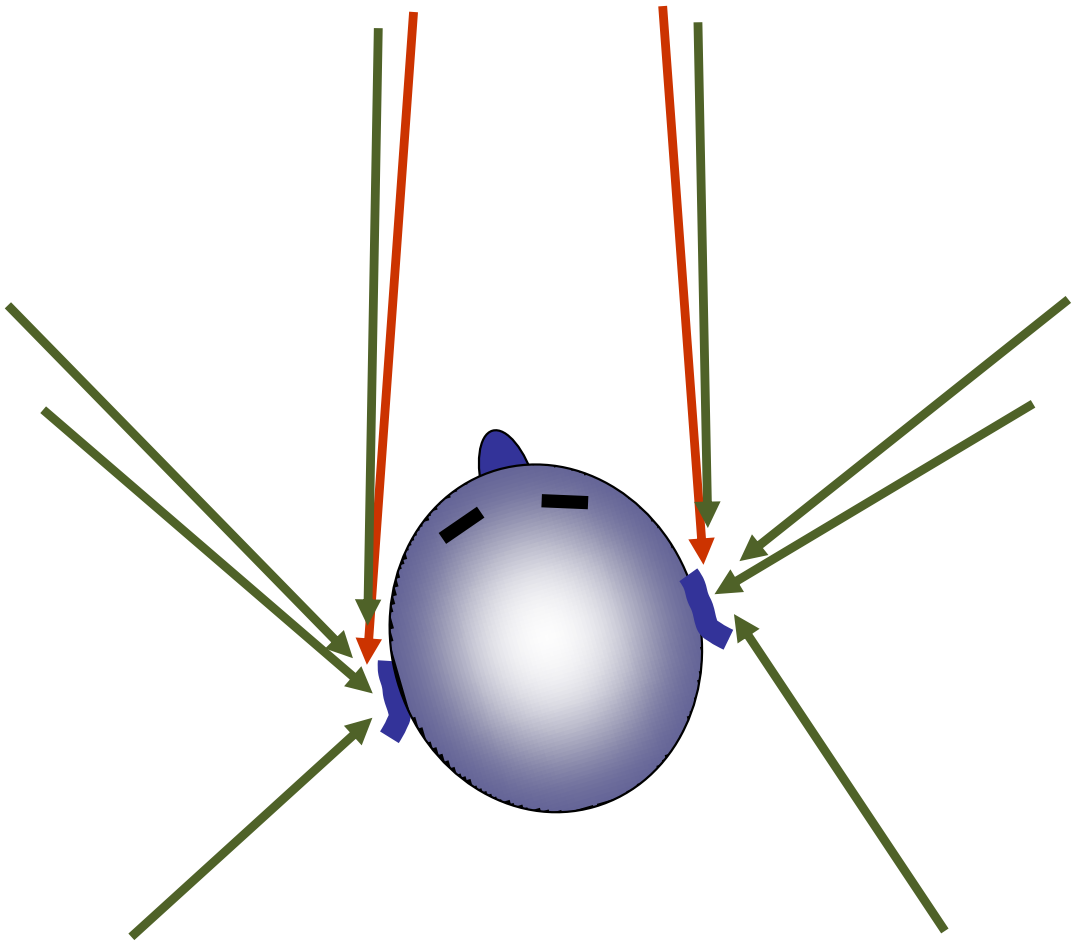
Real sound source

Virtual sound source

Dummy head (head tracking required!)

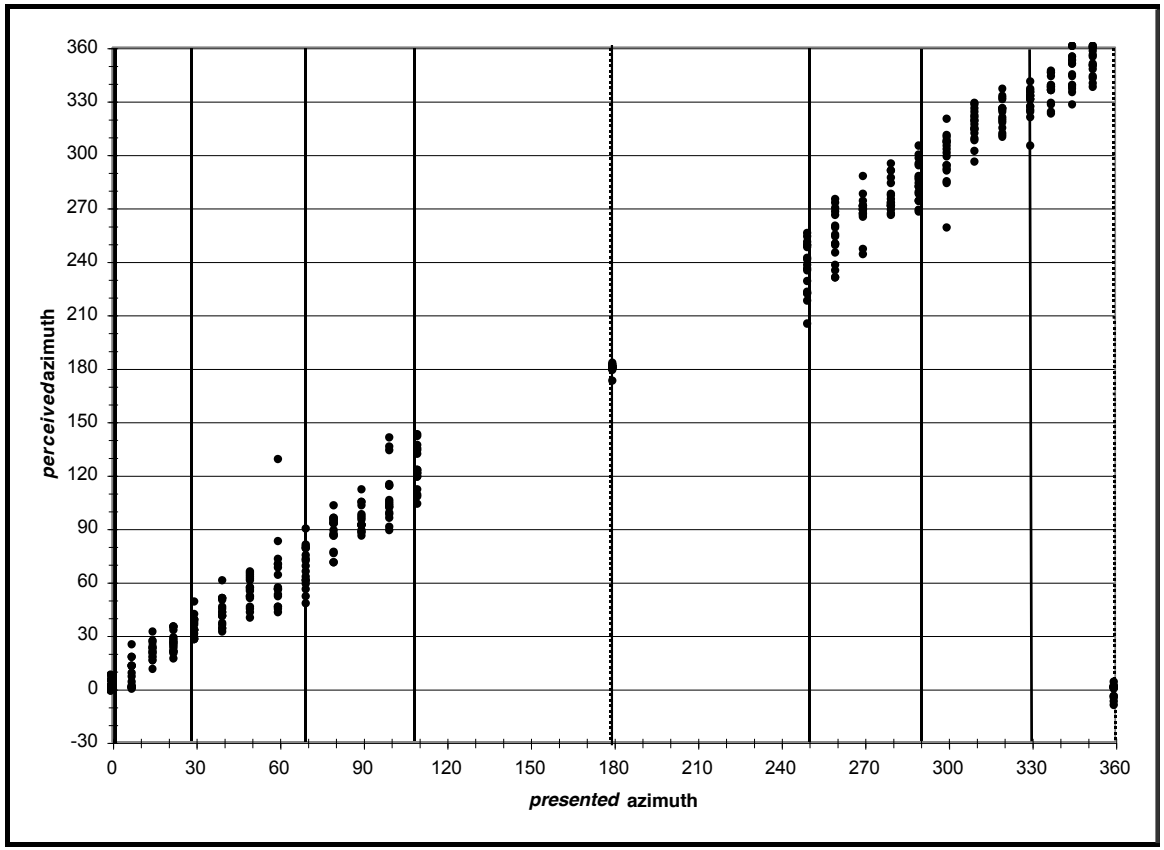


Dummy head (head tracking required!)



Importance of dynamic cues

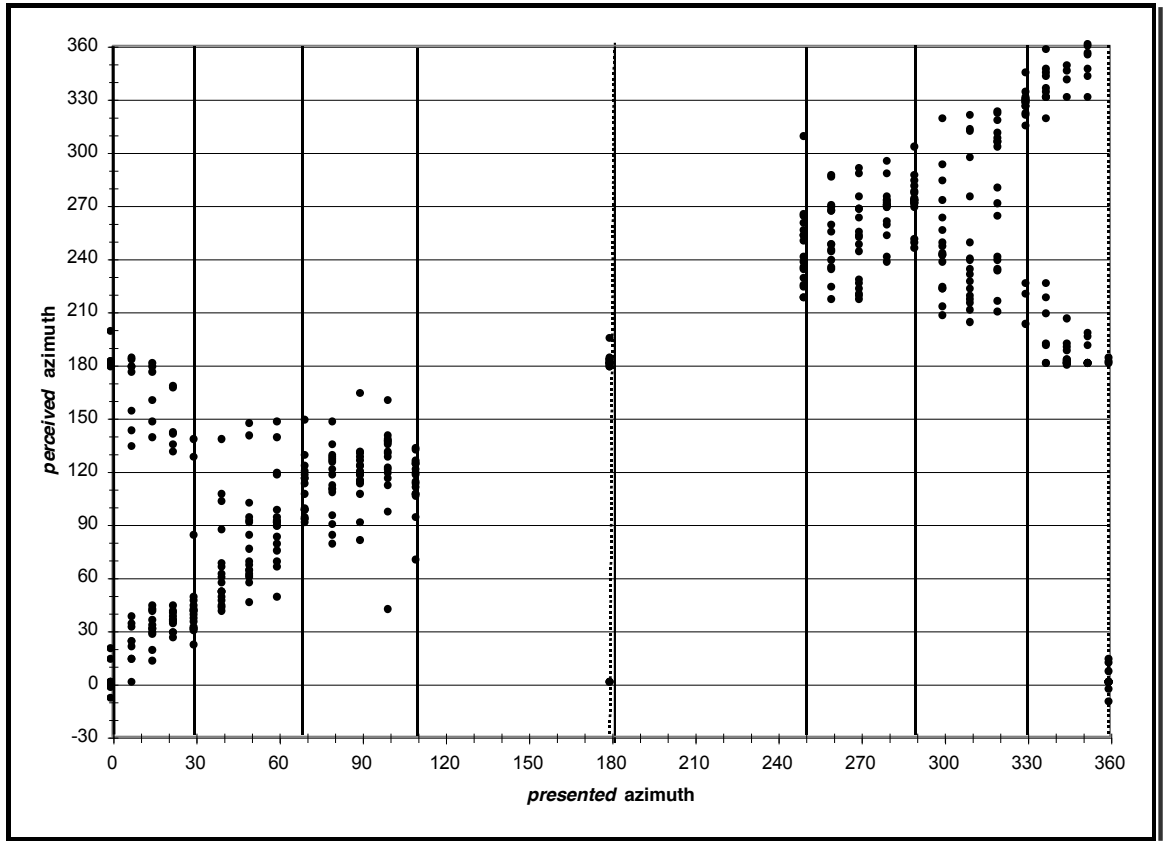
Listening tests horizontal head movements:
Mackensen, Reichenauer, Theile (TMT 1998)



Importance of dynamic cues

Listening tests horizontal head movements:
Mackensen, Reichenauer, Theile (TMT 1998)

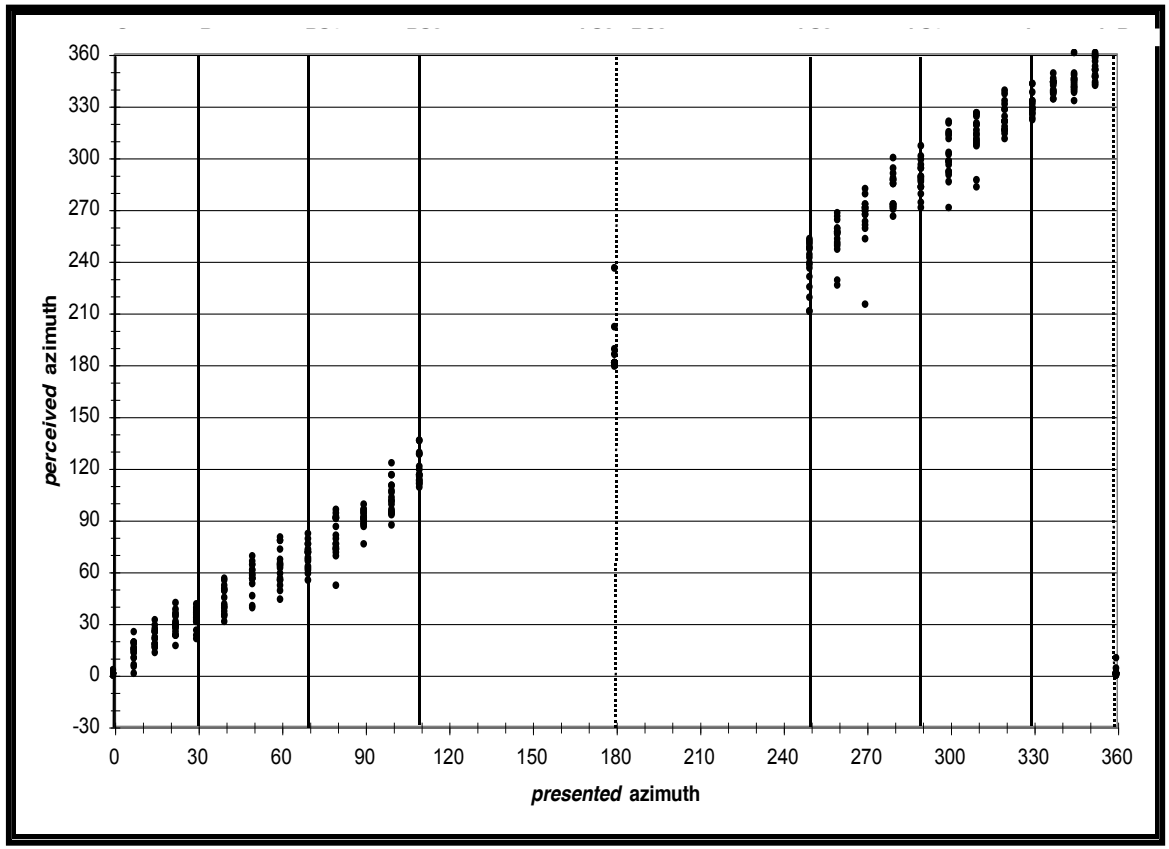
Dummy-head
without
head-tracking



Importance of dynamic cues

Natural listening (studio)

Listening tests horizontal head movements:
Mackensen, Reichenauer, Theile (TMT 1998)



Practical constraints in case of spatial transmission applications:

- Personal reproduction (headphone or „crosstalk cancelation“)
- Head tracking is required
- Exact equalization is required (individual is optimal)
- Original loudness and dynamic is required
- Limited spatial design
- Unsatisfying aesthetic loudspeaker compatibility
- Expensive productions, no archives

Practical constraints in case of spatial transmission applications

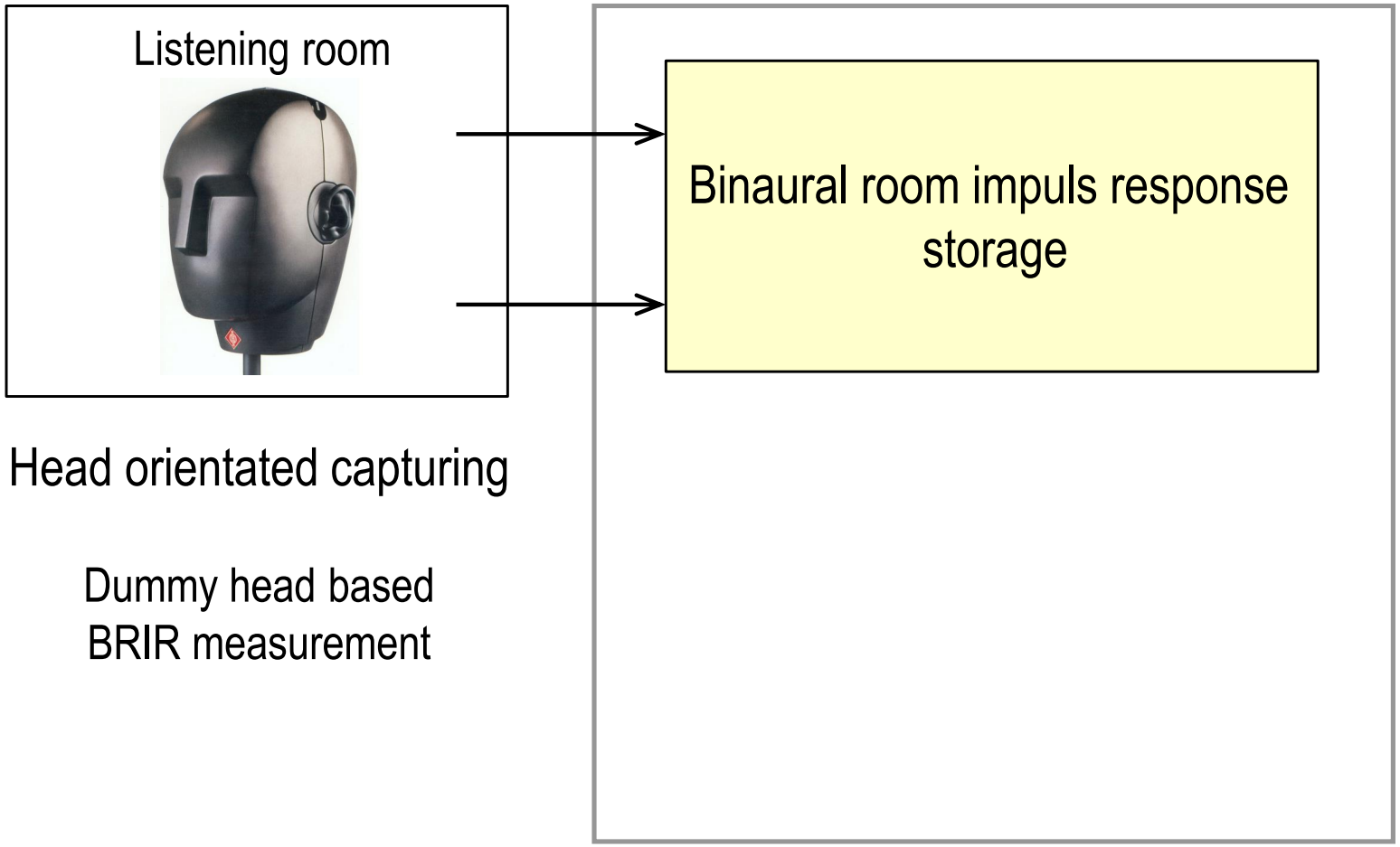
However, in case of pure spatial reproduction applications:

High quality binaural representations of loudspeaker setups are established technologies and used extensively in the professional world for monitoring surround sound.

Developments für consumer applications (mobile, home, games).

- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- Head tracking
- **Binaural Room Synthesis (BRS)**
- Virtual headphone

Binaural Room Synthesis (BRS) - the virtual listening room

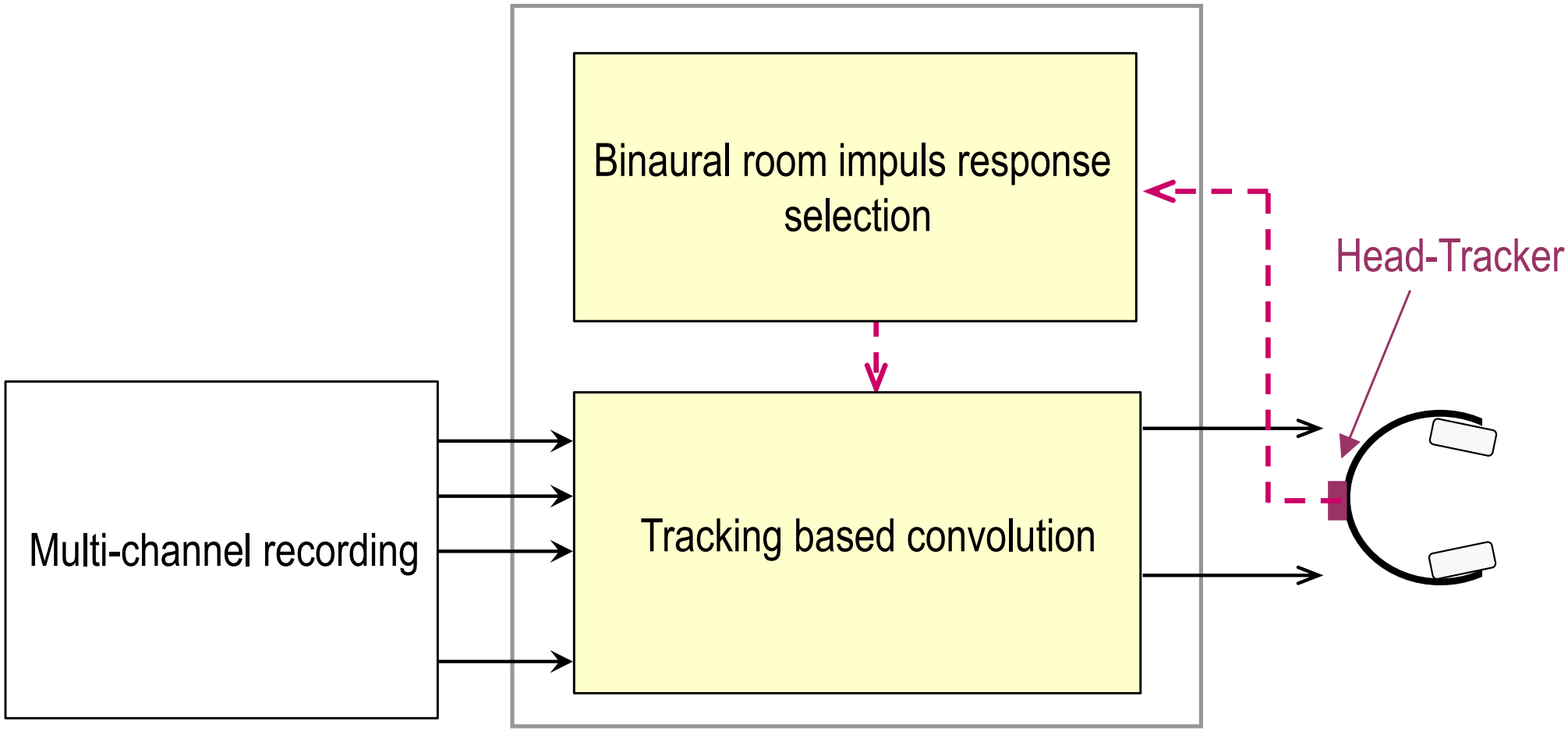


Head orientated capturing

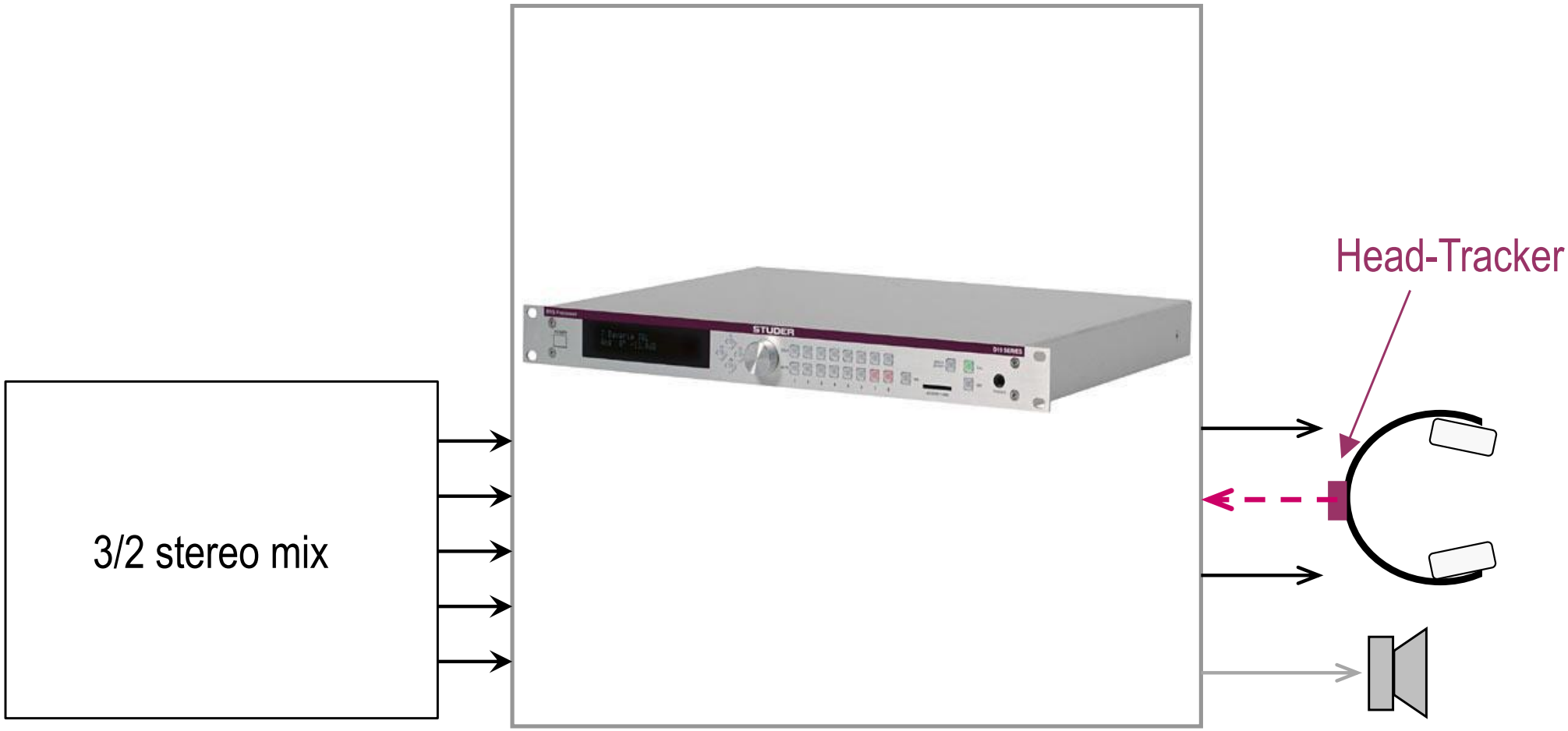
Dummy head based
BRIR measurement

BRS Processor (*Studer / IRT 1999*)

Binaural Room Synthesis (BRS) - the virtual listening room

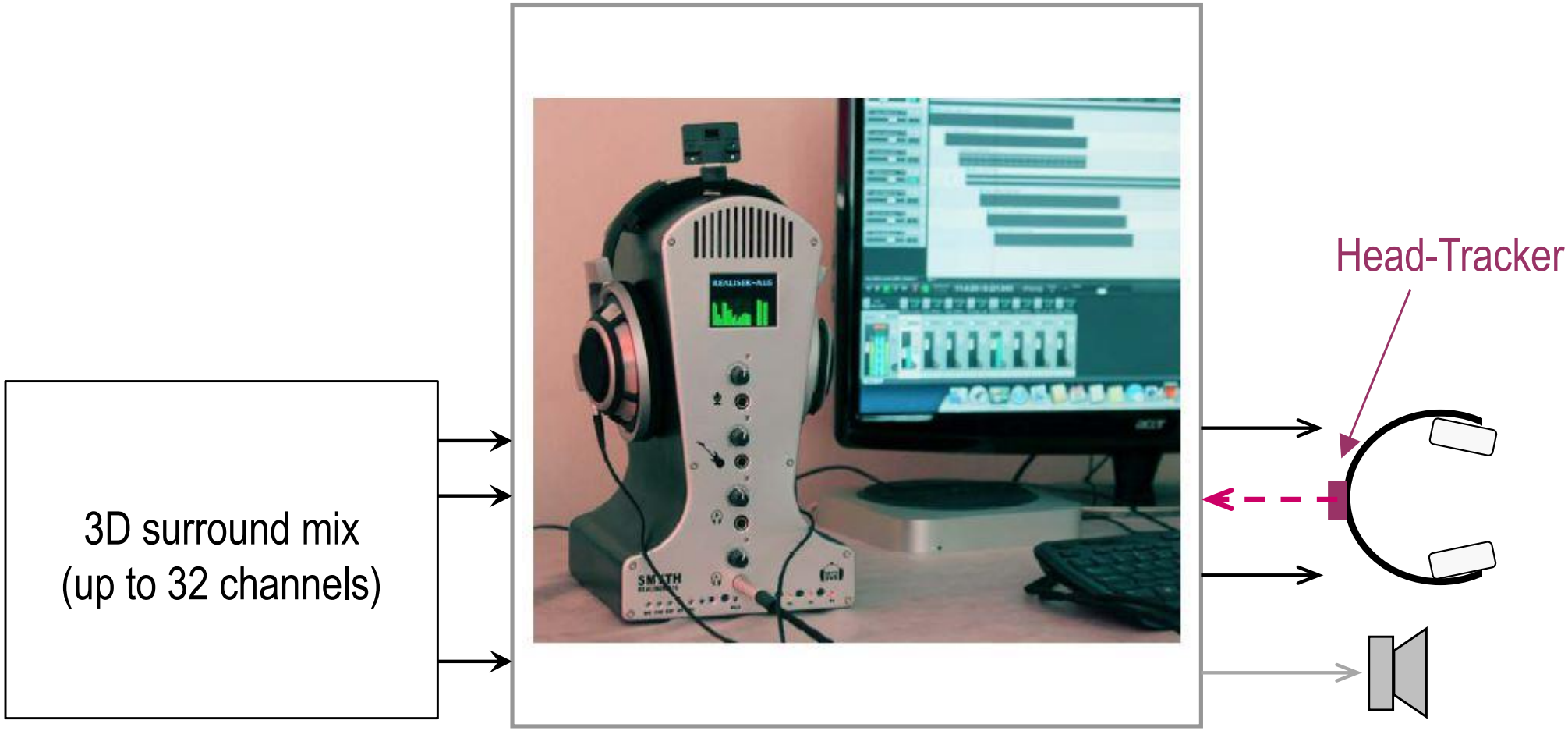


Binaural Room Synthesis (BRS) - the virtual listening room



First BRS Processor (*Studer / IRT 1999*)

Binaural Room Synthesis (BRS) - the virtual listening room

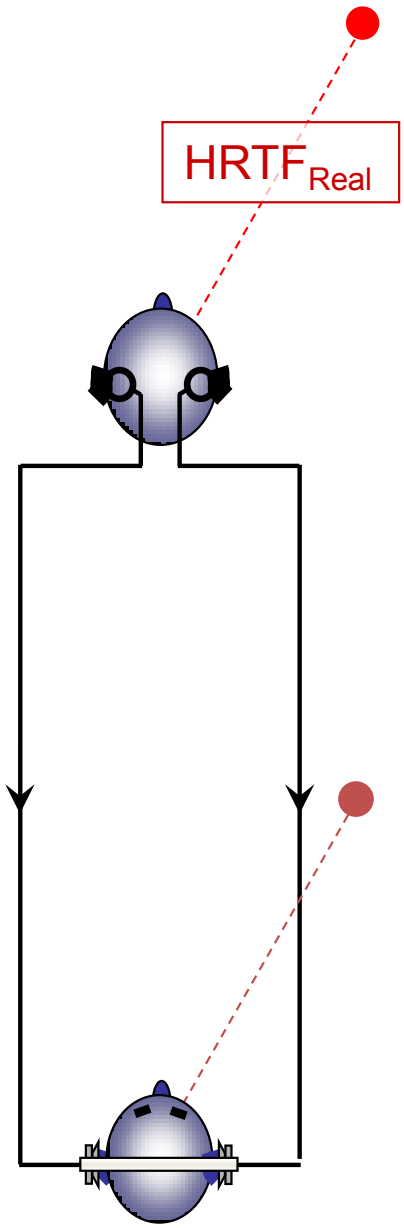


BRS Smyth Realiser A16 32-ch

- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- Head tracking
- Binaural Room Synthesis (BRS)
- Virtual headphone

Dummy head

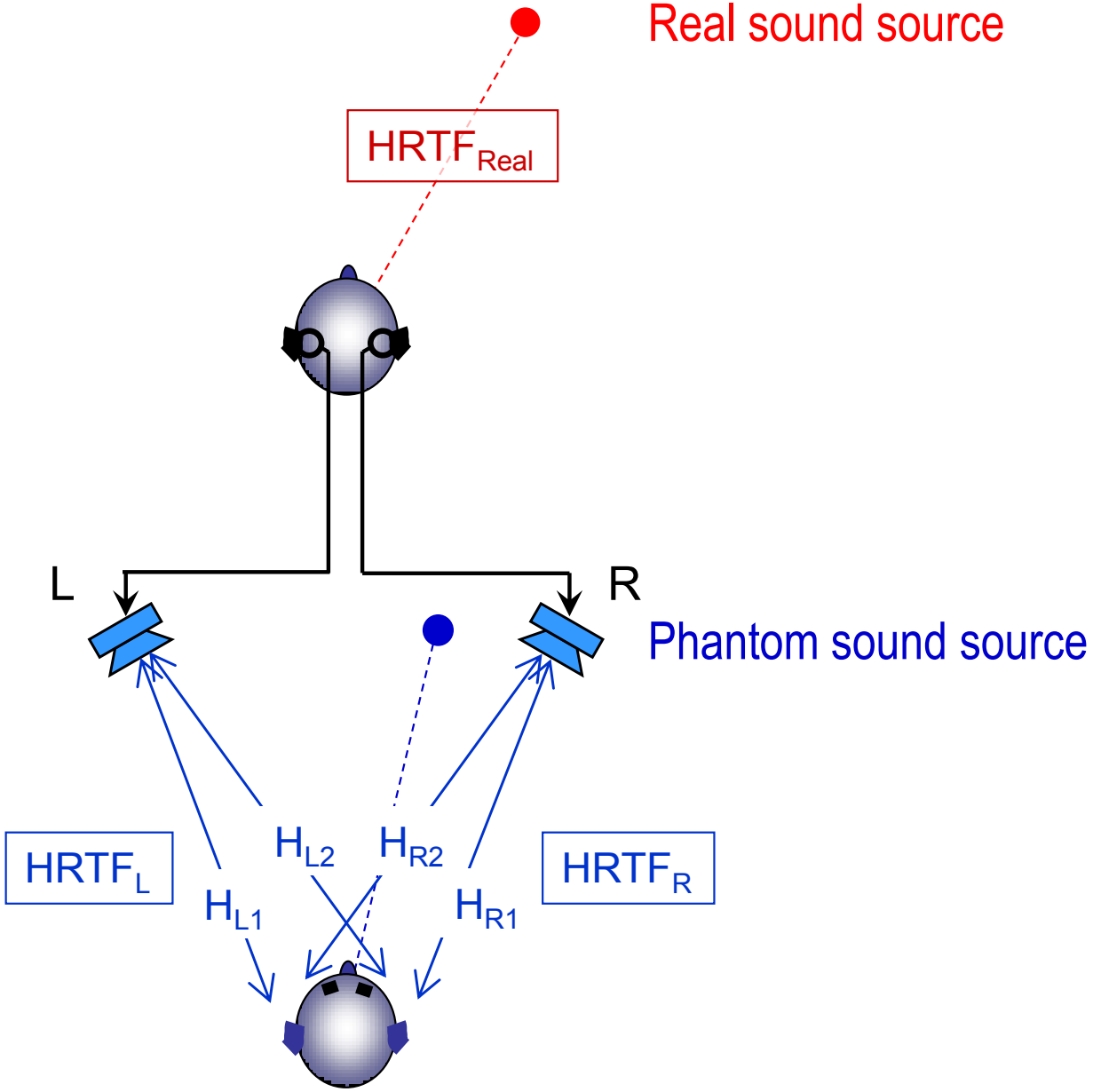
Real sound source



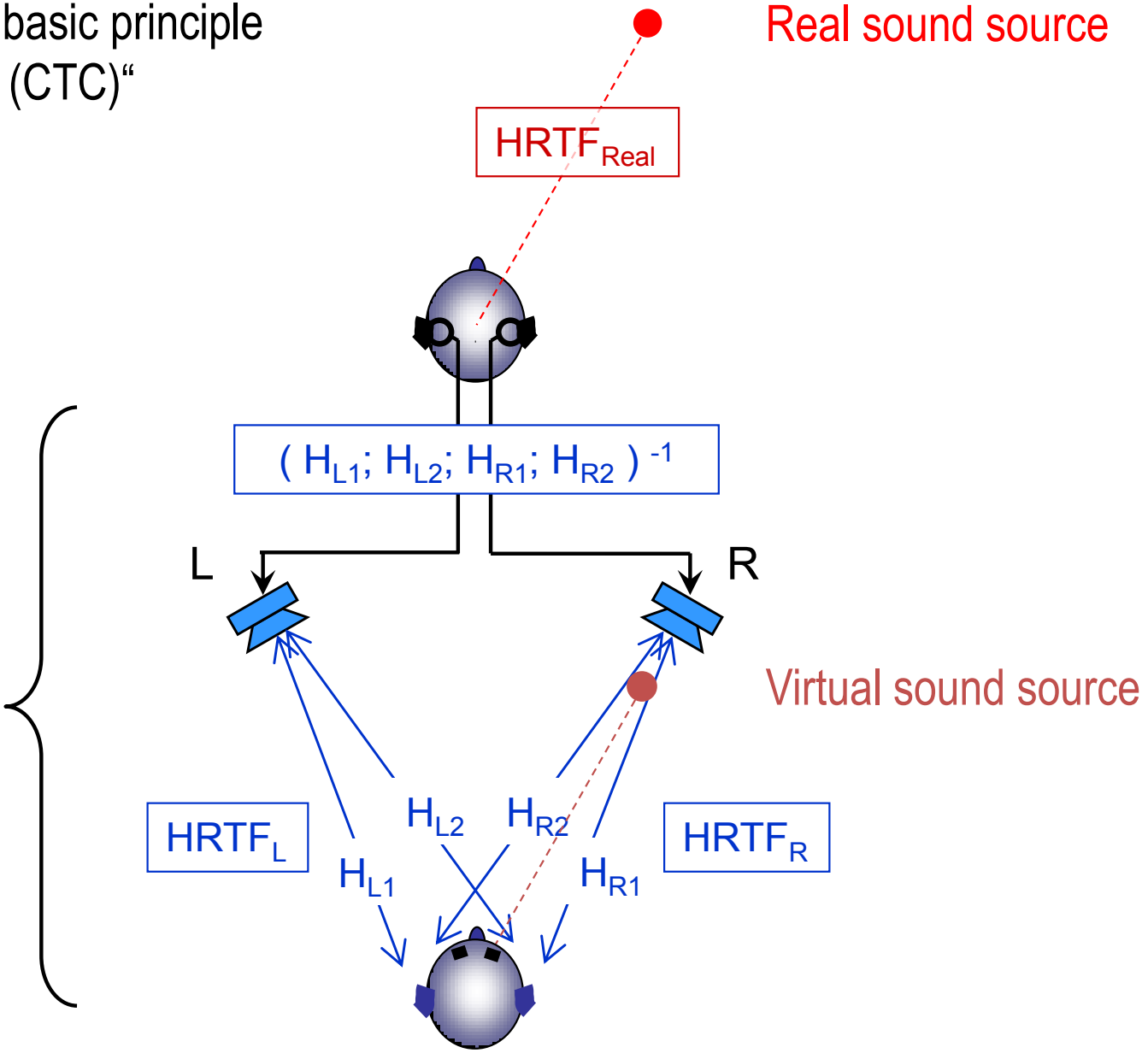
Transfer function = 1

Virtual sound source

Stereophonic imaging

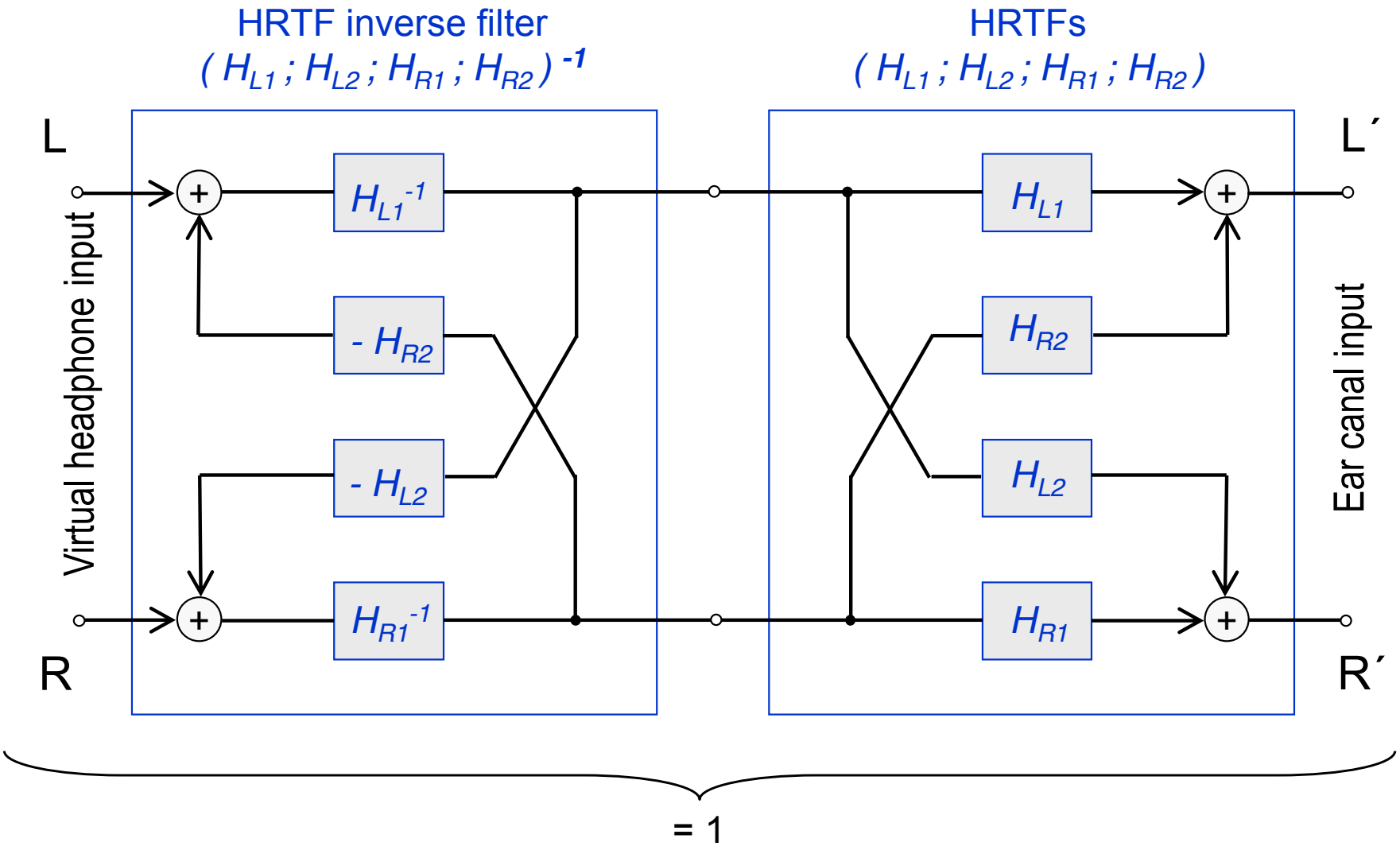


HRTF inverse filtering, basic principle
„Crosstalk Cancelation (CTC)“



Transfer function = 1

HRTF inverse filtering

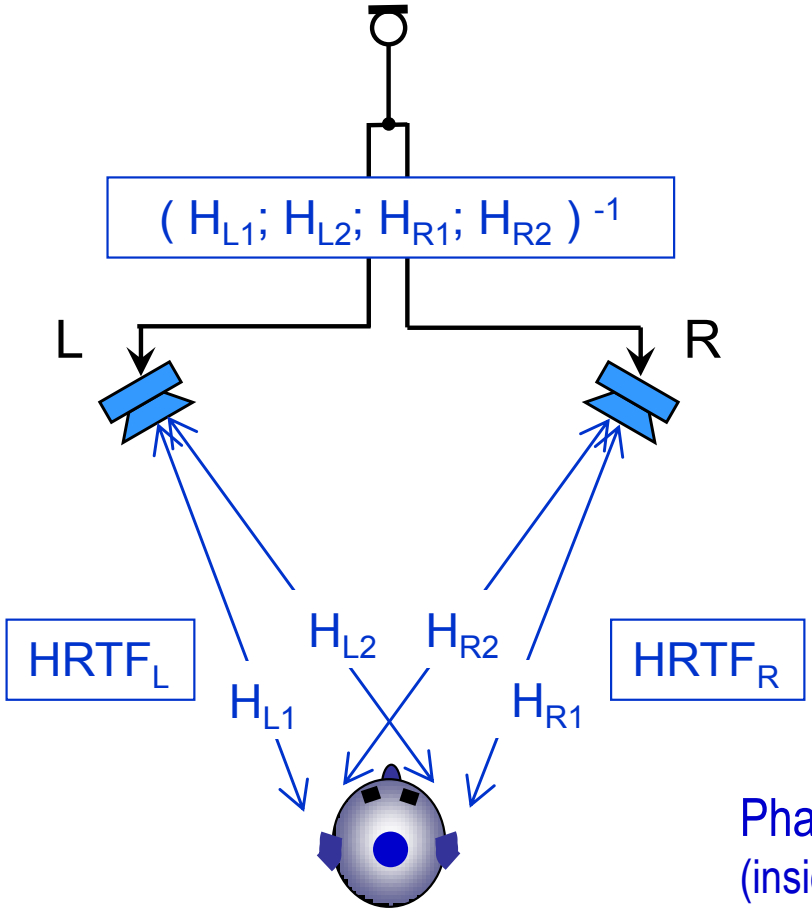


Virtual headphone input: Monophonic signals L & R

Checking the
virtual headphone:

Monophonic signals L & R
should result in
in-head-localization
due to transfer function = 1

Virtual headphone
transfer function = 1



Phantom sound source
(inside localisation)

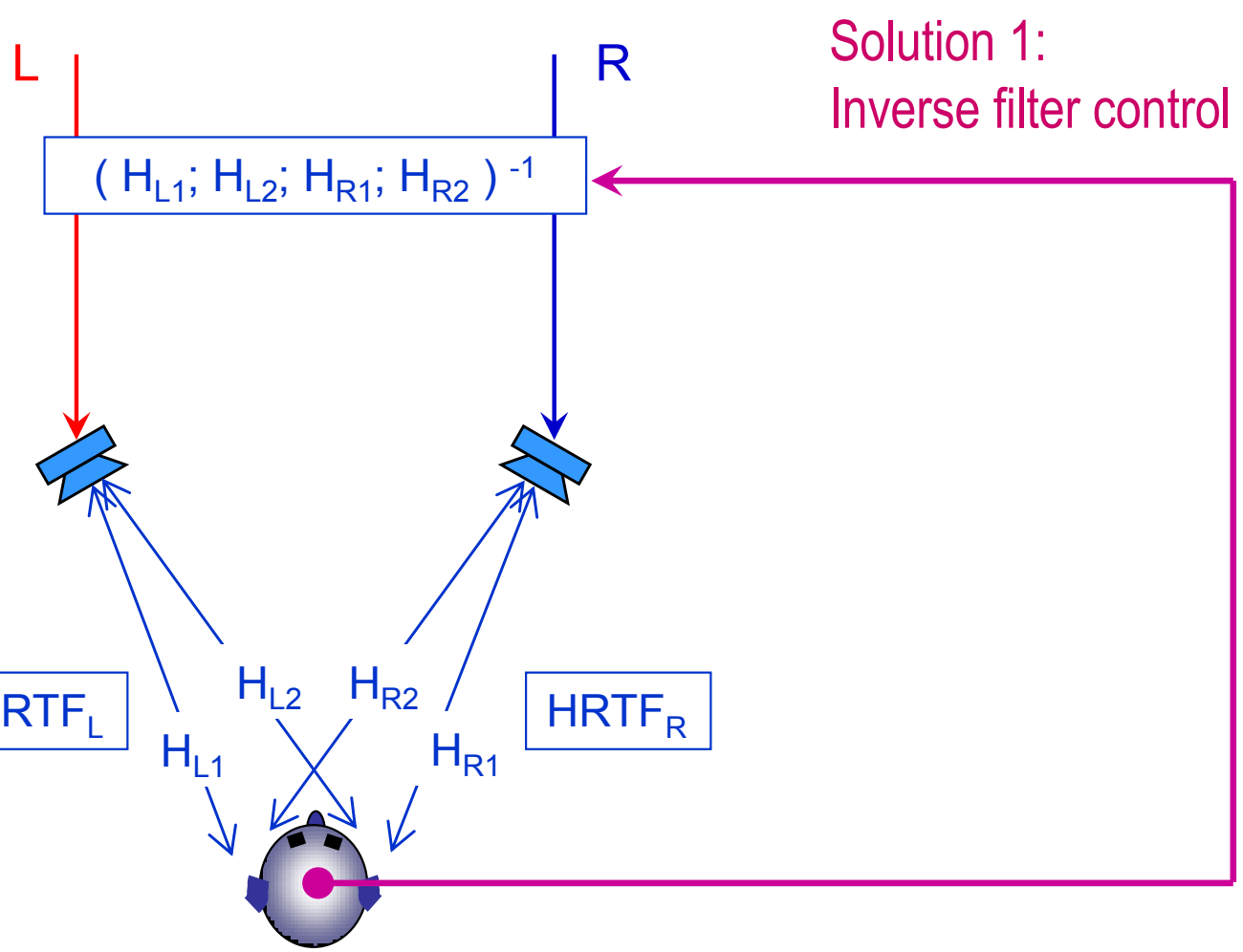
Spatial transmission methods

- (Multi-channel) loudspeaker stereophonic representation
- Multi-channel loudspeaker representation of single sources
- → Sound field reconstruction of acoustic scenes
- → Binaural reconstruction of the ear input signals

Hybrid systems can combine advantages of both methods

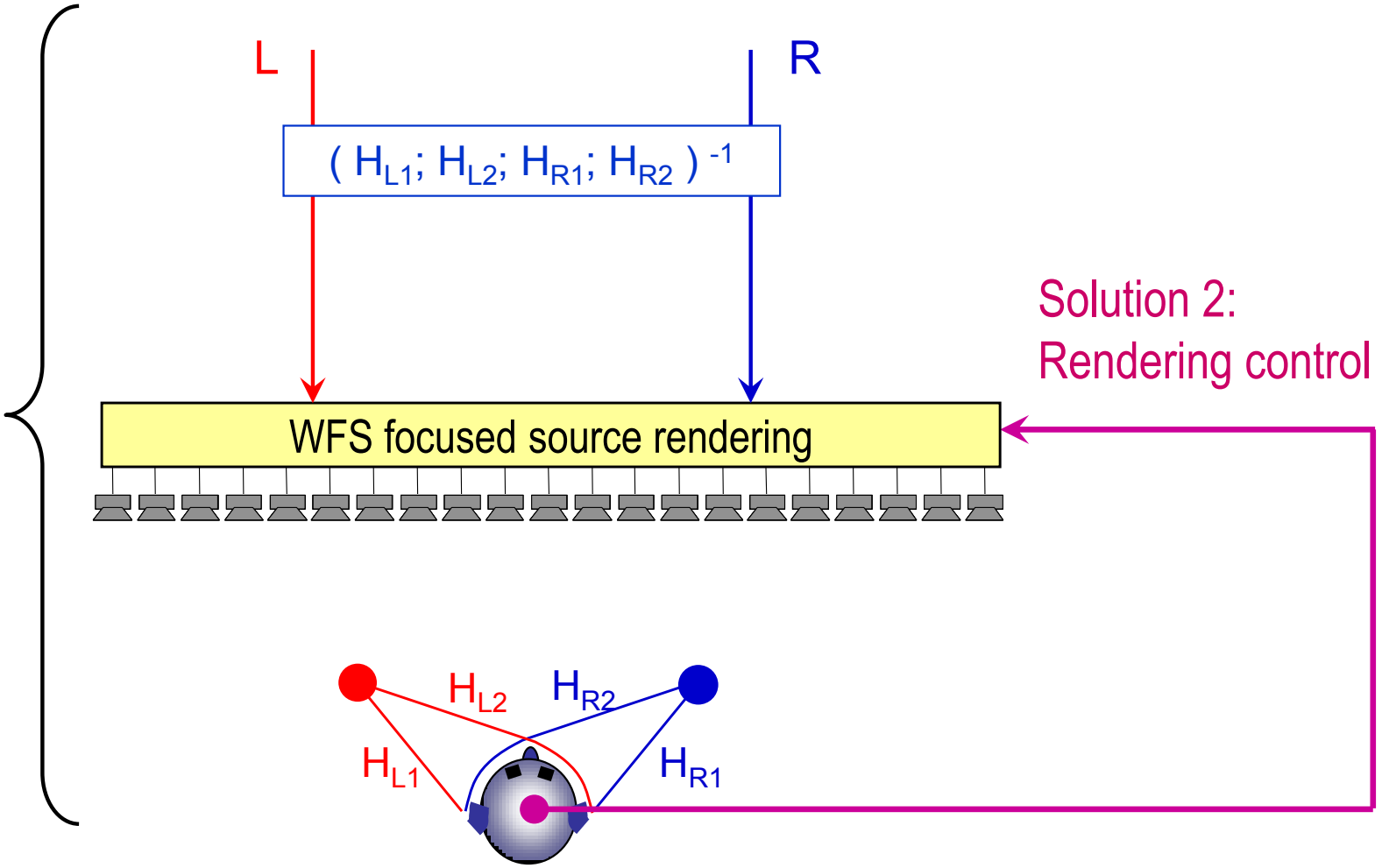
Precise inverse filtering is required → Head tracking !

Virtual headphone
transfer function = 1



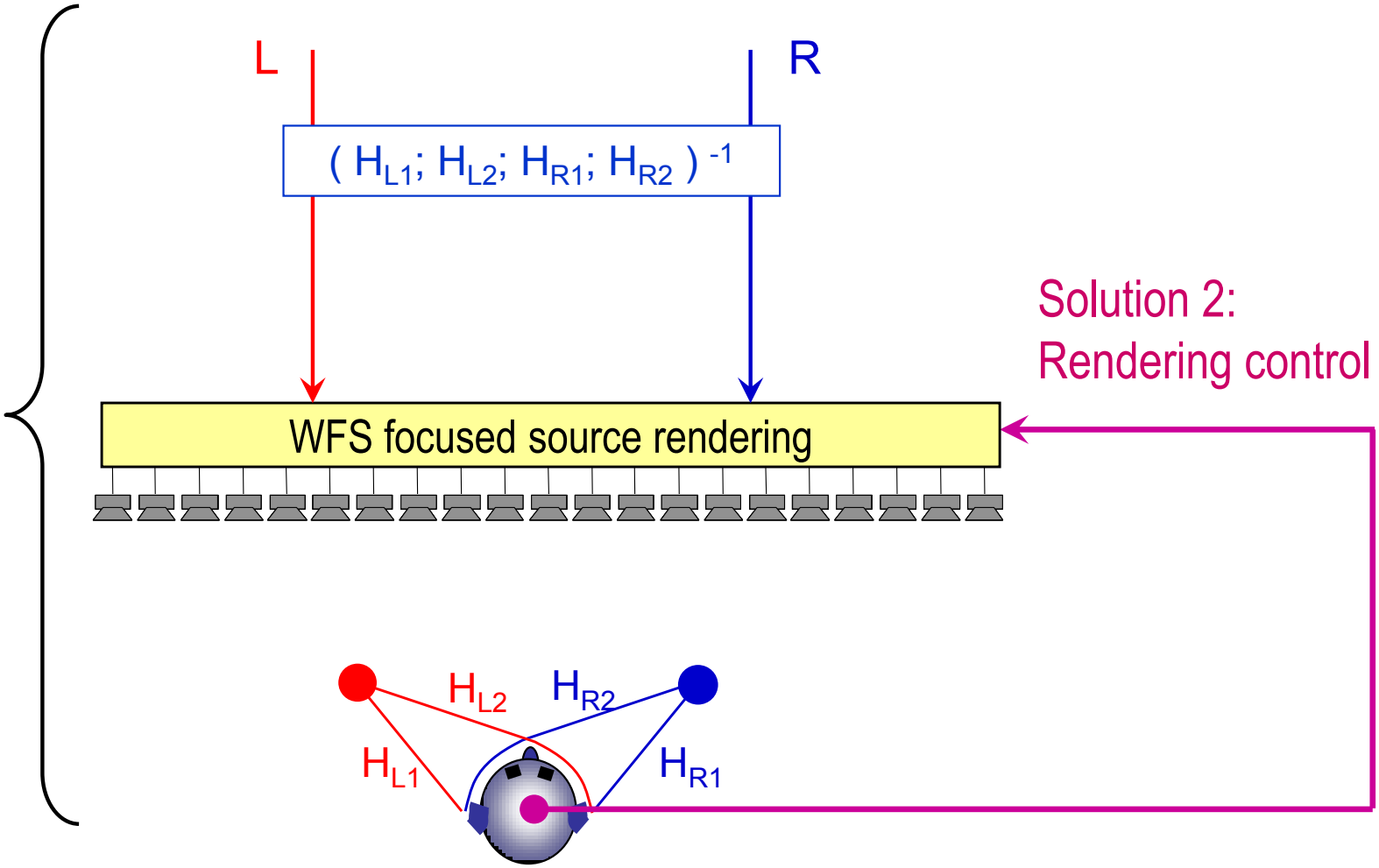
Precise inverse filtering is required → Head tracking !

Virtual headphone
transfer function = 1



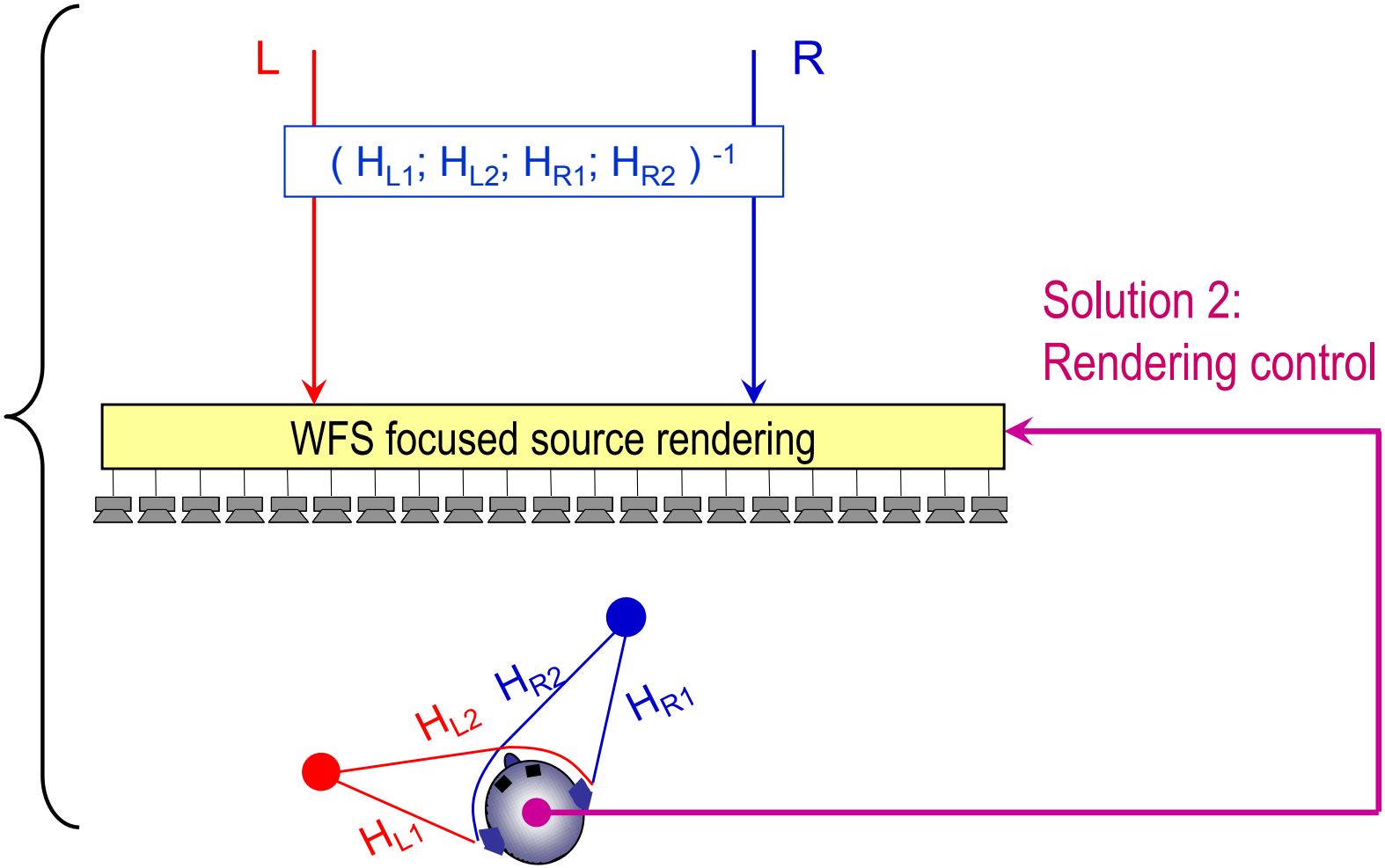
HRTF inverse filter based on WFS (*Wittek 2005*)

Virtual headphone transfer function = 1



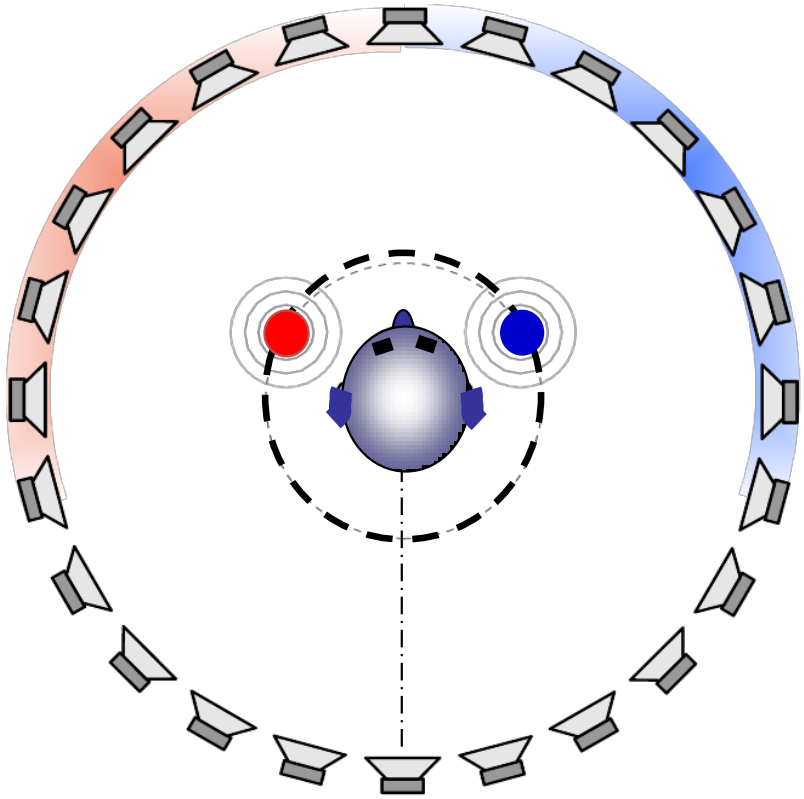
HRTF inverse filter based on WFS (*Wittek 2005*)

Virtual headphone transfer function = 1



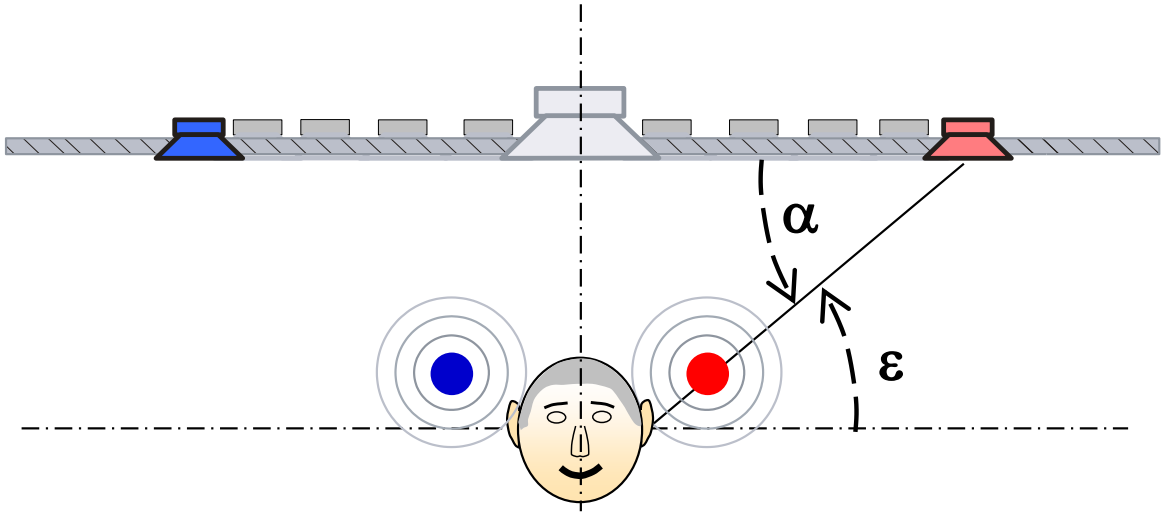
HRTF inverse filter based on WFS (*Wittek 2005*)

Focused sources rendering – circle array



HRTF inverse filter based on WFS (*Wittek 2005*)

Focused sources rendering – circle array



HRTF inverse filter based on WFS (*Wittek 2005*)

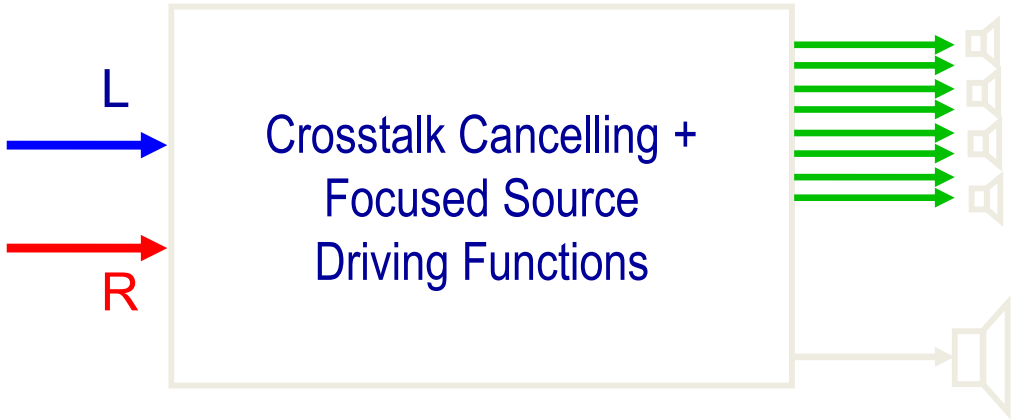
- WFS based focused source rendering
- Head orientation tracking + head position tracking
- Static 0°-HRTF filtering
- Diffusfield equalization

→ Headphone listening without headphone

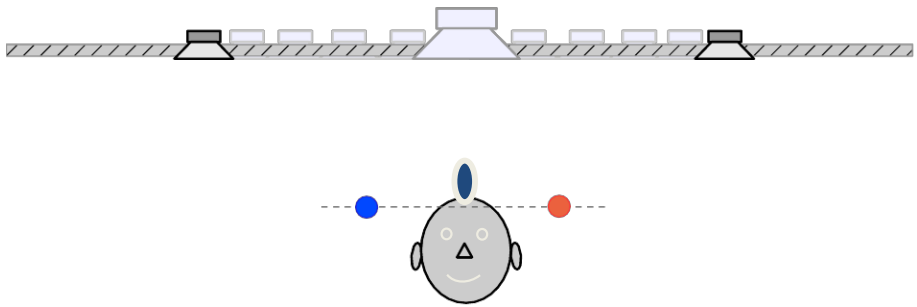
Headphone listening without headphone

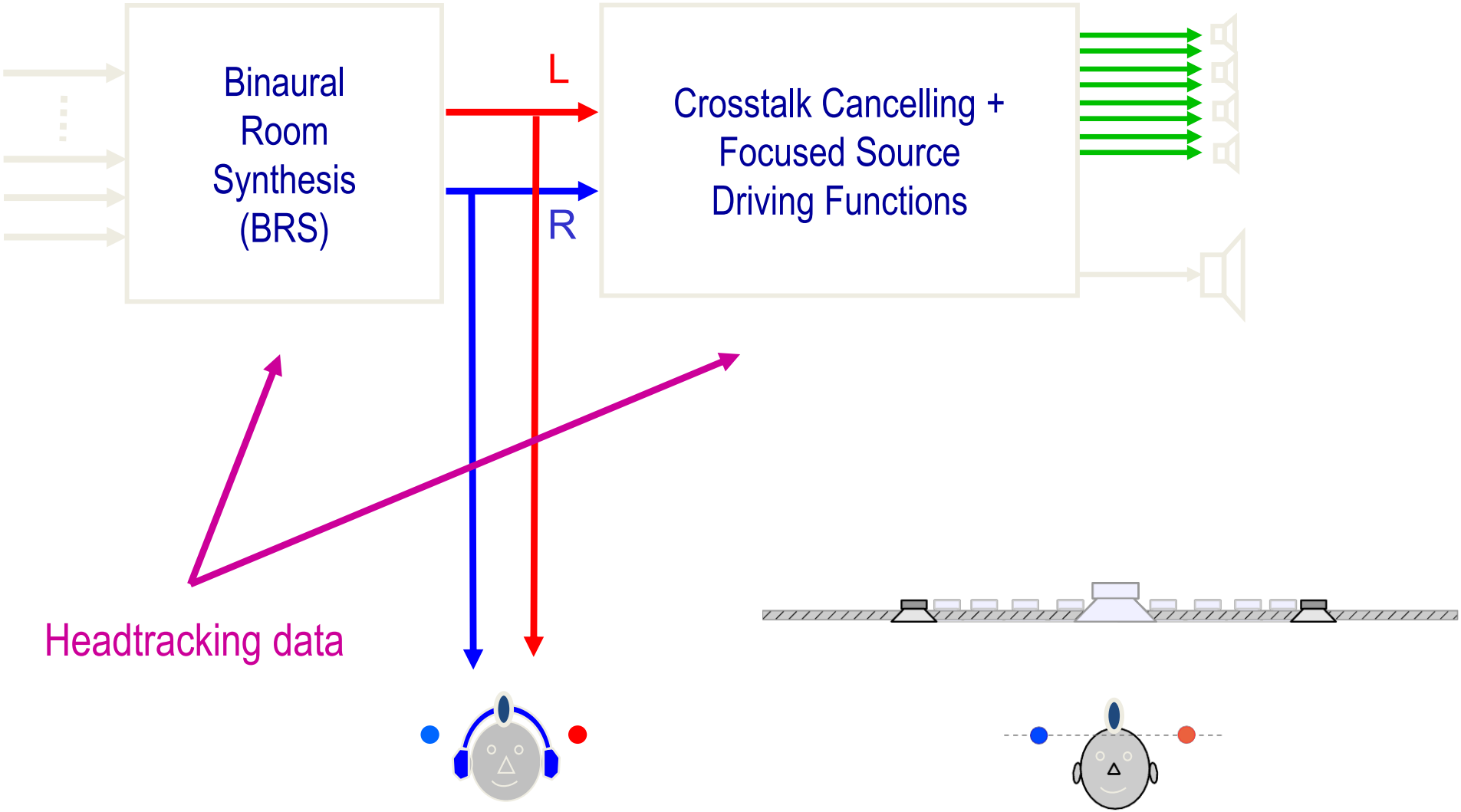
Circle array above listener - experimental setup

Study by
Menzel / Wittek 2005

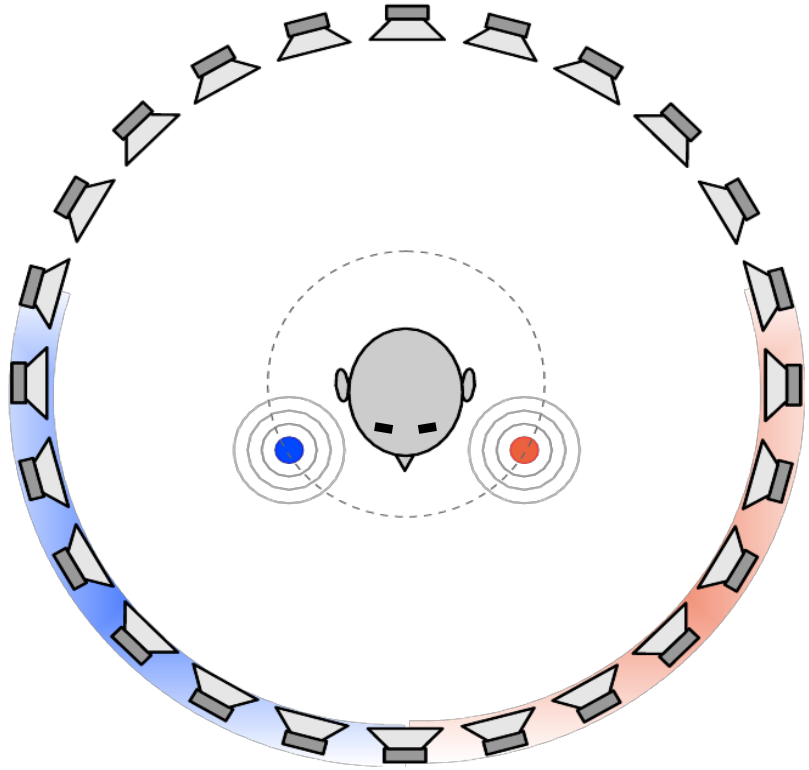


Headtracking data

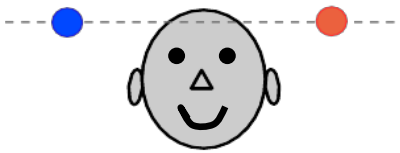


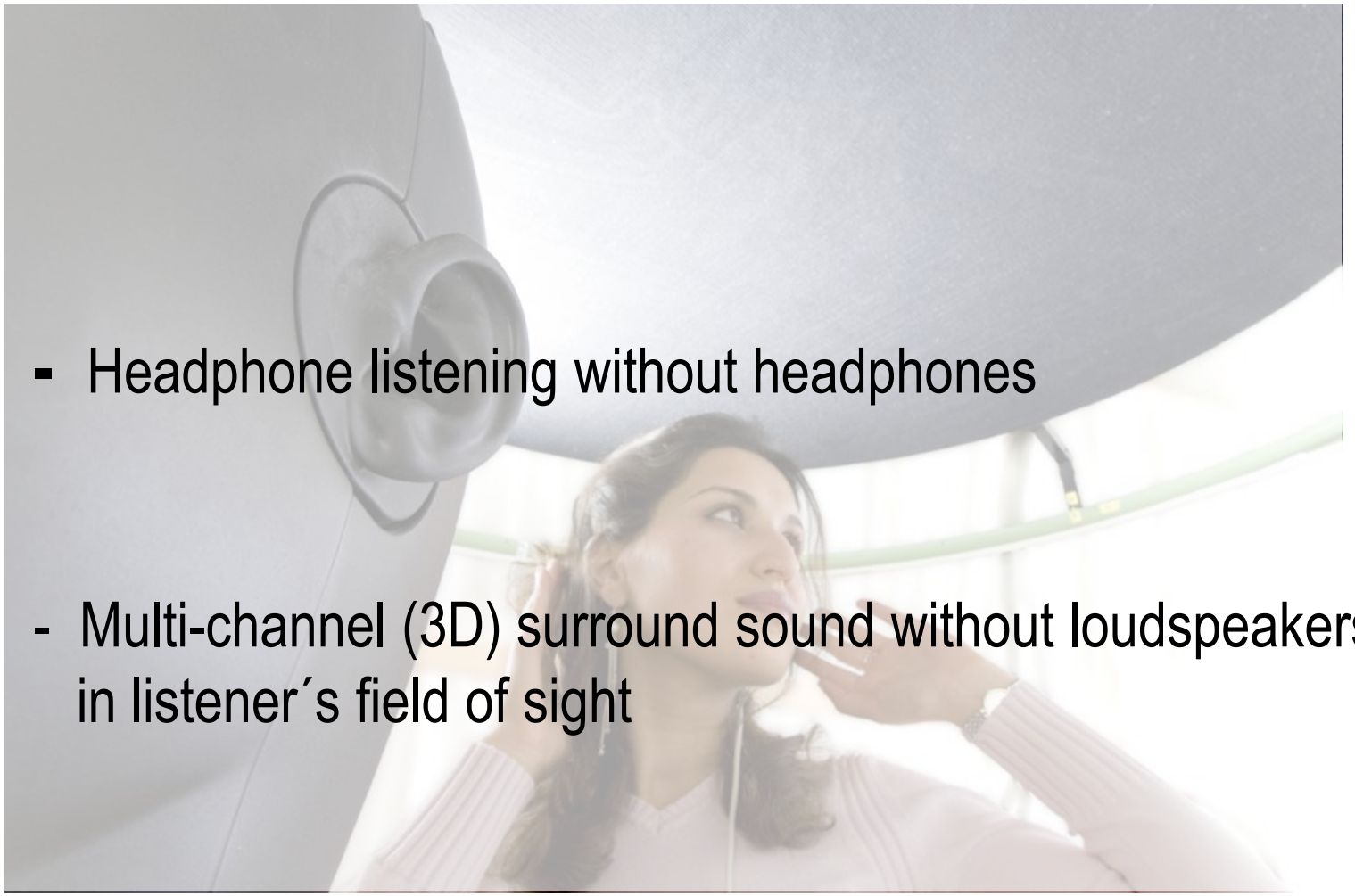

Binaural
Room
Synthesis
(BRS)
L
R
Crosstalk Cancelling +
Focused Source
Driving Functions
Headtracking data

Virtual source



Virtual source




- Headphone listening without headphones
- Multi-channel (3D) surround sound without loudspeakers in listener´s field of sight


Applications:
Automotive
Living room
Virtual Reality
Augmented reality

Audio for Virtual Reality – Binaural Audio Principles

- Immersive sound
- Headphone reproduction
- Spatial transmission methods
  - Stereophony
  - Sound field synthesis (WFS / HOA)
  - Binaural techniques
- Head tracking
- Binaural Room Synthesis (BRS)
- Virtual headphone

29. Tonmeistertagung

VDT International Convention

tmt29

expertise in audio media

Welcome to the 29th Tonmeistertagung!

17th to 20th of November 2016 | Cologne

Congress Centrum Nord, Cologne, Germany
www.tonmeister.de

www.tonmeister.de/tmt

Lectures | Workshops | Round Tables
Exhibition | Products | Demos
Contacts | Experiences | Exchange

29. Tonmeistertagung

VDT International Convention

tmt29

expertise in audio media

Welcome to the 29th Tonmeistertagung!

17th to 20th of November 2016 | Cologne

Congress Centrum Nord, Cologne, Germany
www.tonmeister.de

Virtual Reality Topics:

40 Workshops and Lectures

- 3D Audio
- Binaural Audio
- Immersive Sound
- Video Meets Audio

www.tonmeister.de/tmt

Vielen Dank!

theile@tonmeister.de

theile@hauptmikrofon.de