Audio developments based on psychoacoustics

– Tutorial –

(version 6.1)

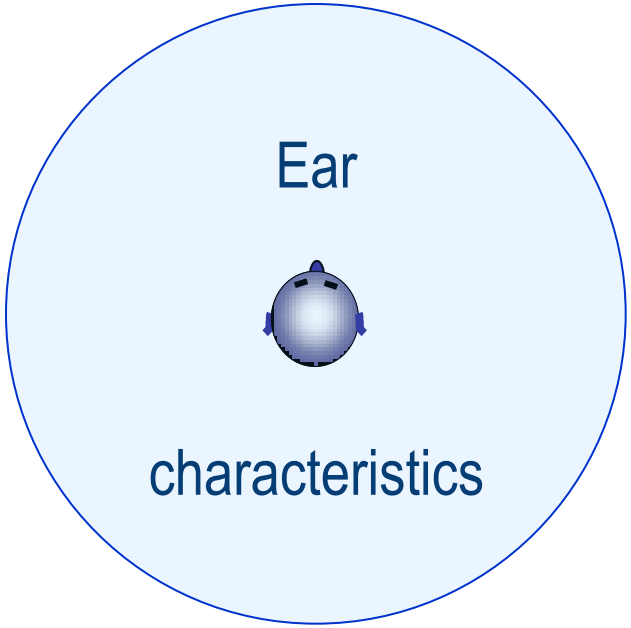
Günther Theile

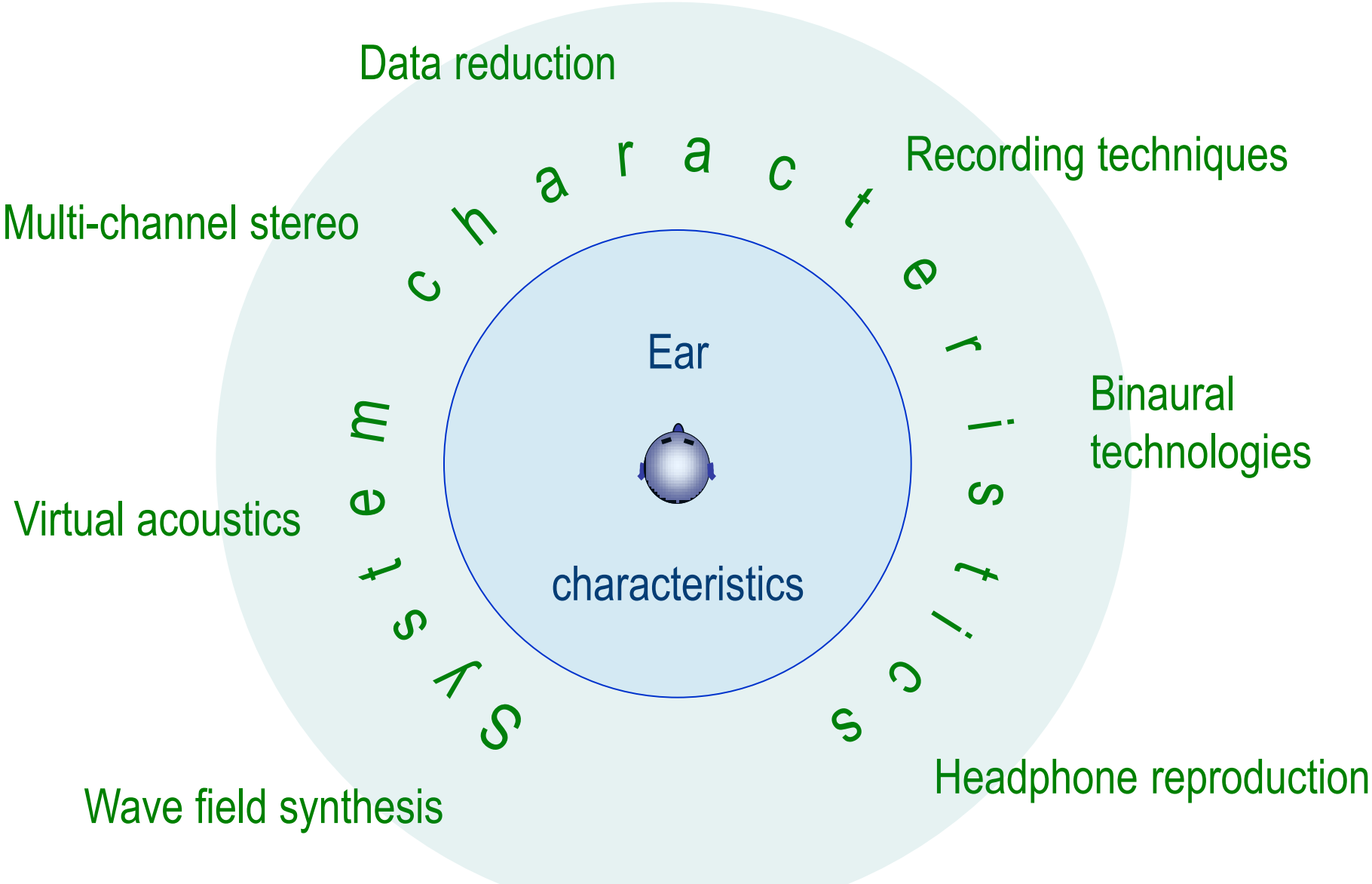
theile@tonmeister.de

Audio developments based on psychoacoustics

- Data reduction
- Headphone reproduction
- Binaural technologies
- Sound field synthesis (WFS / HOA)
- Stereophony
- 5.1 / 7.1 Surround (ITU BS 775)
- 3D Audio

Audio developments based on psychoacoustics




Data reduction
Recording techniques
Multi-channel stereo
Binaural technologies
Virtual acoustics
Headphone reproduction
Wave field synthesis
Ear characteristics
Ear characteristics system

Psychoacoustics

Das Ohr
als Nachrichtenempfänger

Von

Prof. Dr. Eberhard Zwicker

und

Prof. Dr. Richard Feldtkeller

Zweite neubearbeitete Auflage

Mit 217 Bildern

RÄUMLICHES HÖREN

von

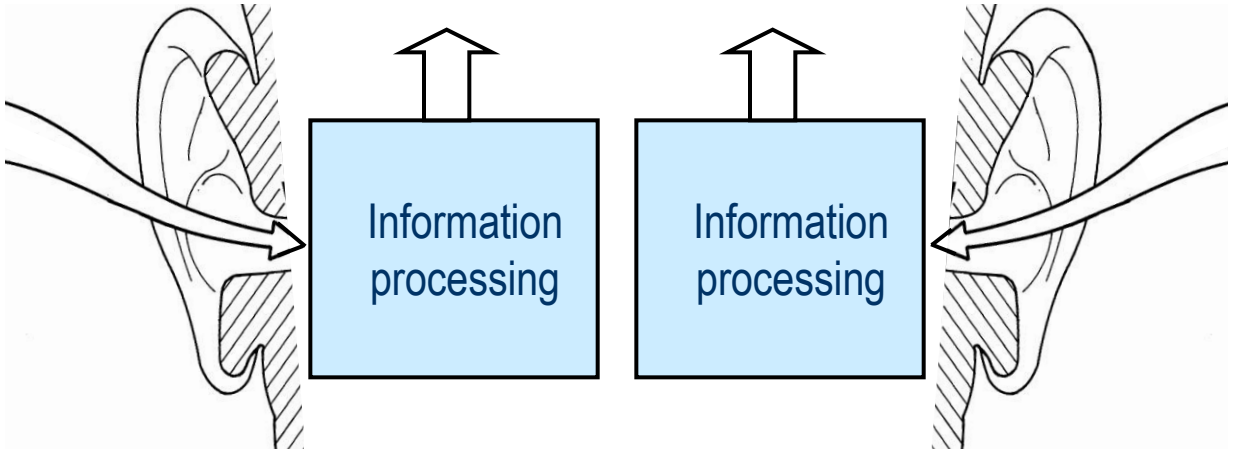
JENS BLAUERT

mit 174 Abbildungen

MONOGRAPHIEN
DER NACHRICHTENTECHNIK

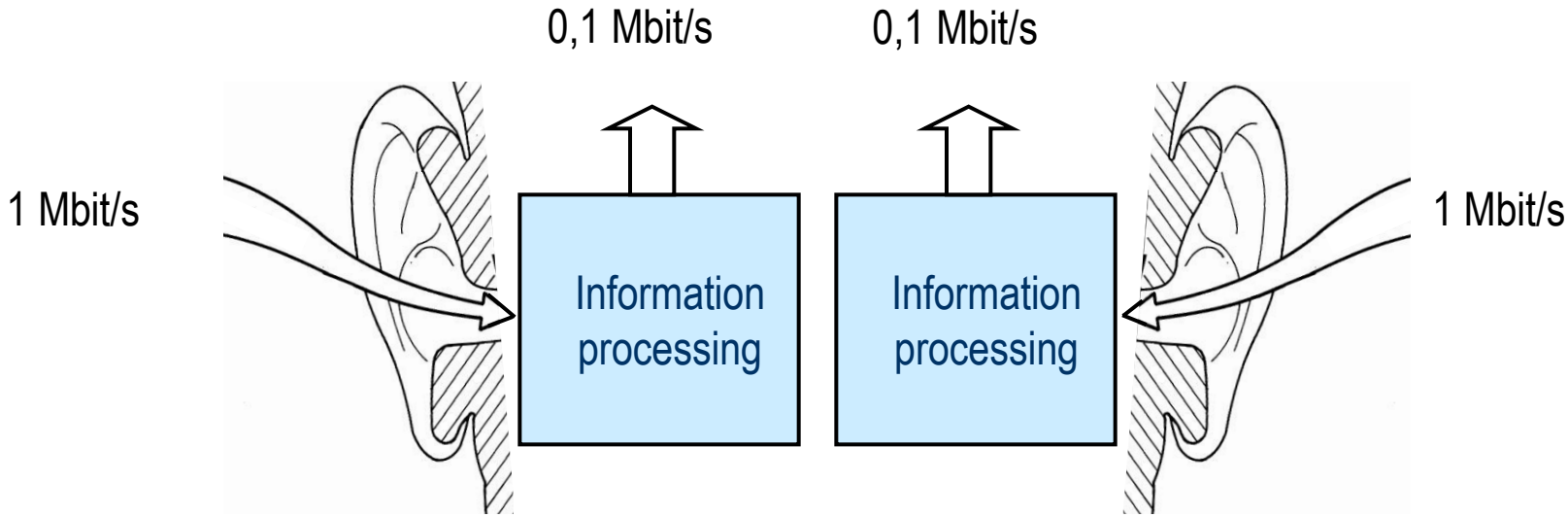
S. HIRZEL VERLAG STUTTGART
1974

Data reduction



Data reduction of auditory information in the peripheral stage

Data reduction

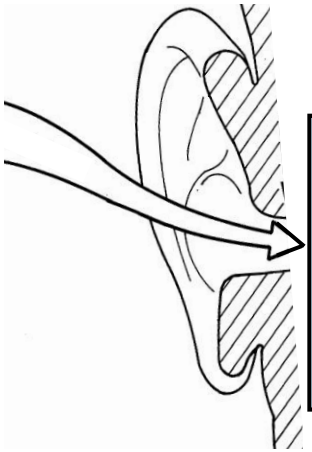


Data reduction of auditory information in the peripheral stage

Perceptual Coding

0,1 Mbit/s

1 Mbit/s



<u>Model:</u>

Internal processing based on „critical bands“ induces frequency dependend and signal dependend noise.

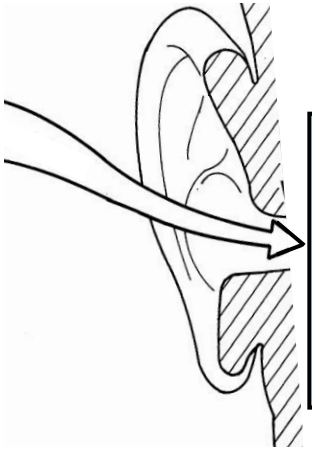
→ Human ear's threshold in quiet and masking threshold

The coding is optimal if the induced noise spectrum corresponds to the hearing threshold

Perceptual Coding

0,1 Mbit/s

1 Mbit/s



<u>Model:</u>

Internal processing based on „critical bands“ induces frequency dependend and signal dependend noise.

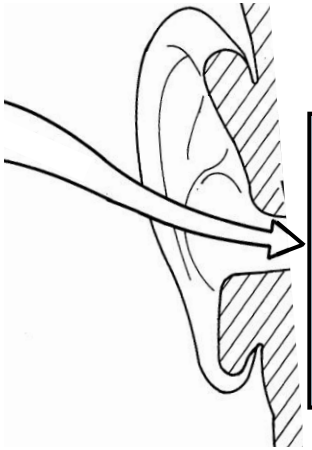
→ Human ear's threshold in quiet and masking threshold

No need to transfer signal components irrelevant to the ear

Perceptual Coding

0,1 Mbit/s

1 Mbit/s



Model:

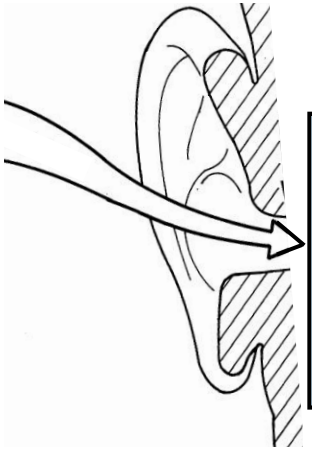
Simultaneous masking:
Some sounds simply vanish in the presence of other sounds with certain characteristics (so called maskers).

No need to transfer signal components irrelevant to the ear

Perceptual Coding

0,1 Mbit/s

1 Mbit/s



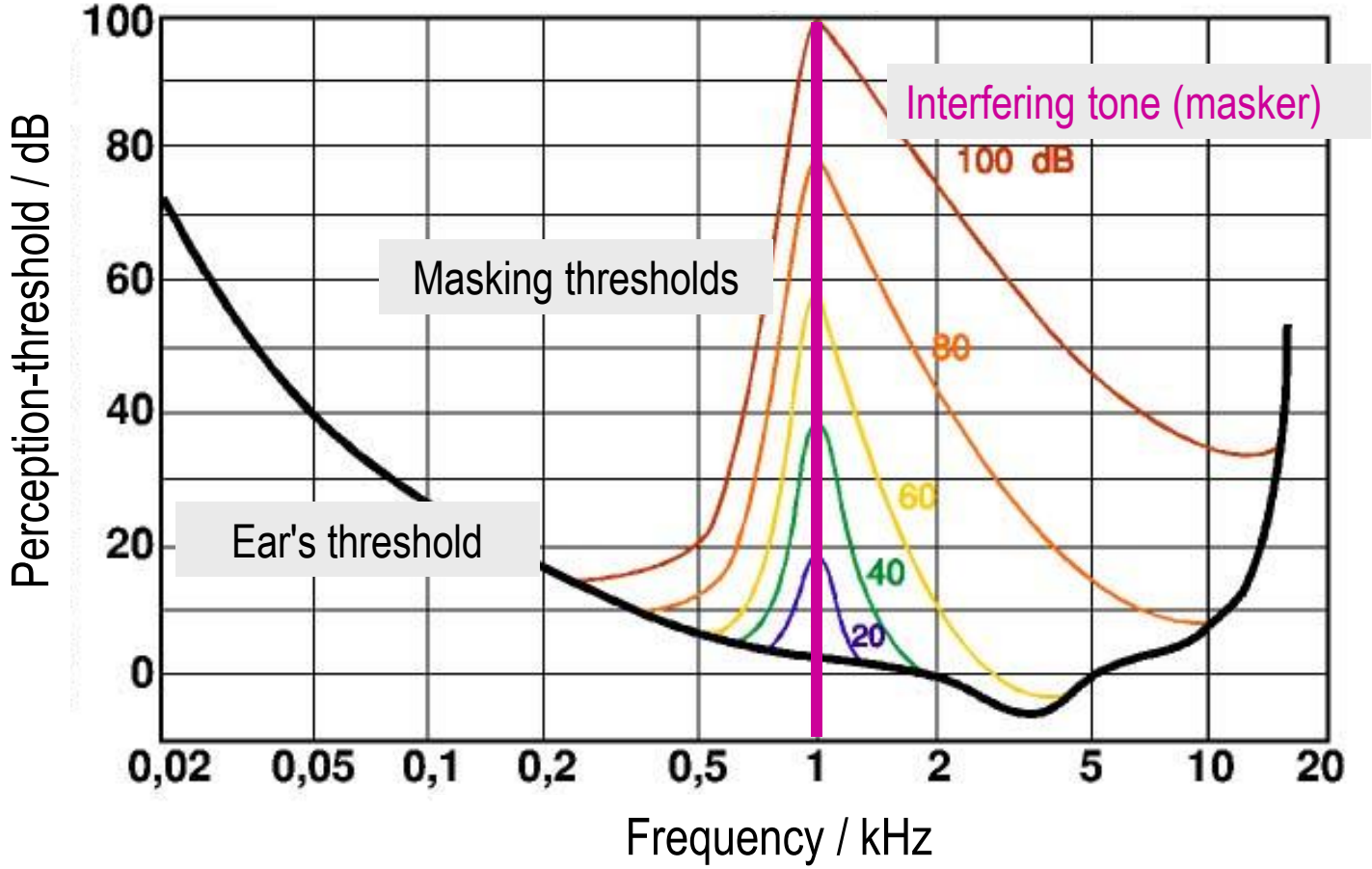
Model:

Temporal masking:
Some sounds are hidden due to maskers which have just disappeared, or even after maskers which are about to appear.

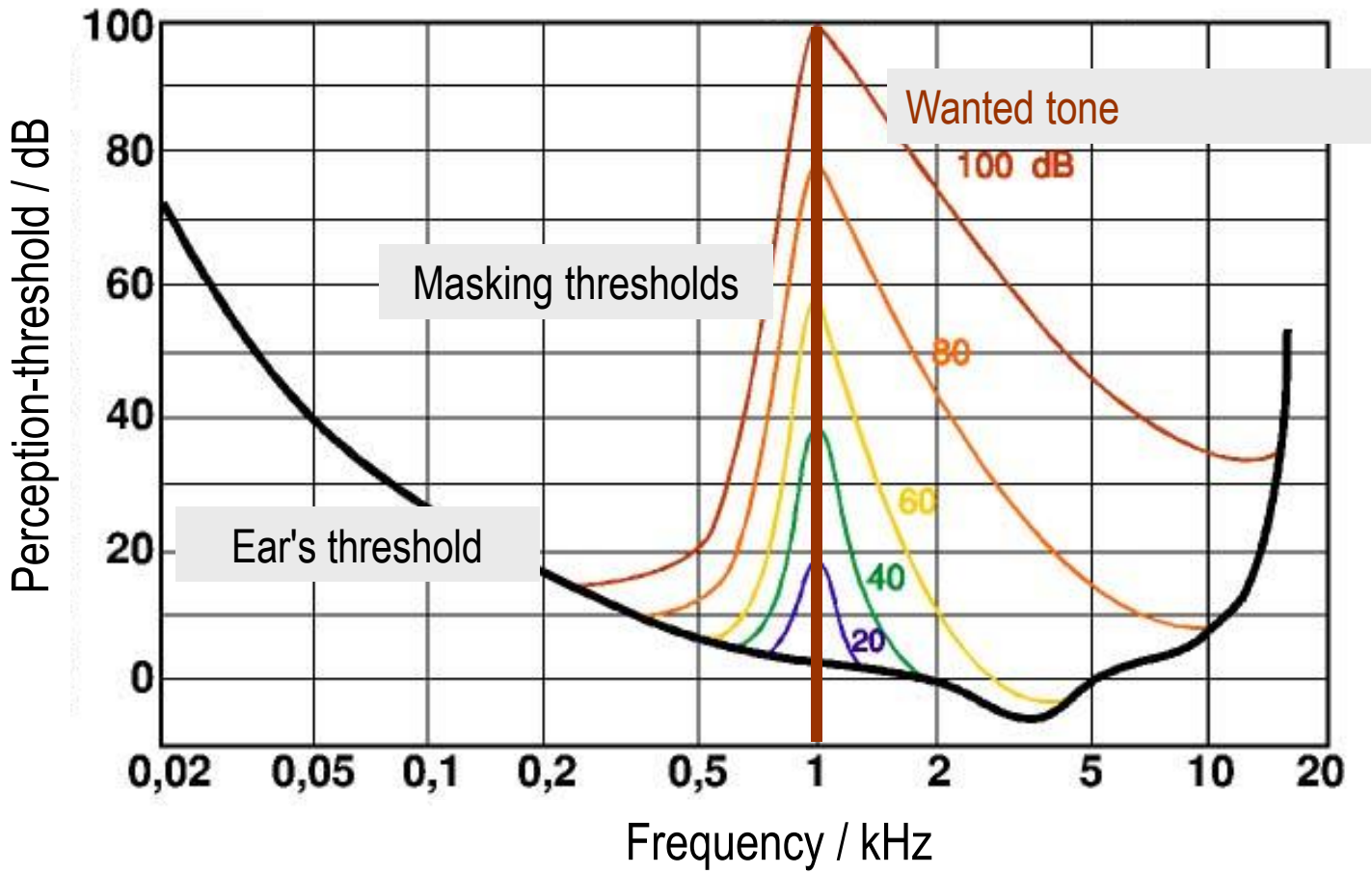
No need to transfer signal components irrelevant to the ear

Simultaneous masking

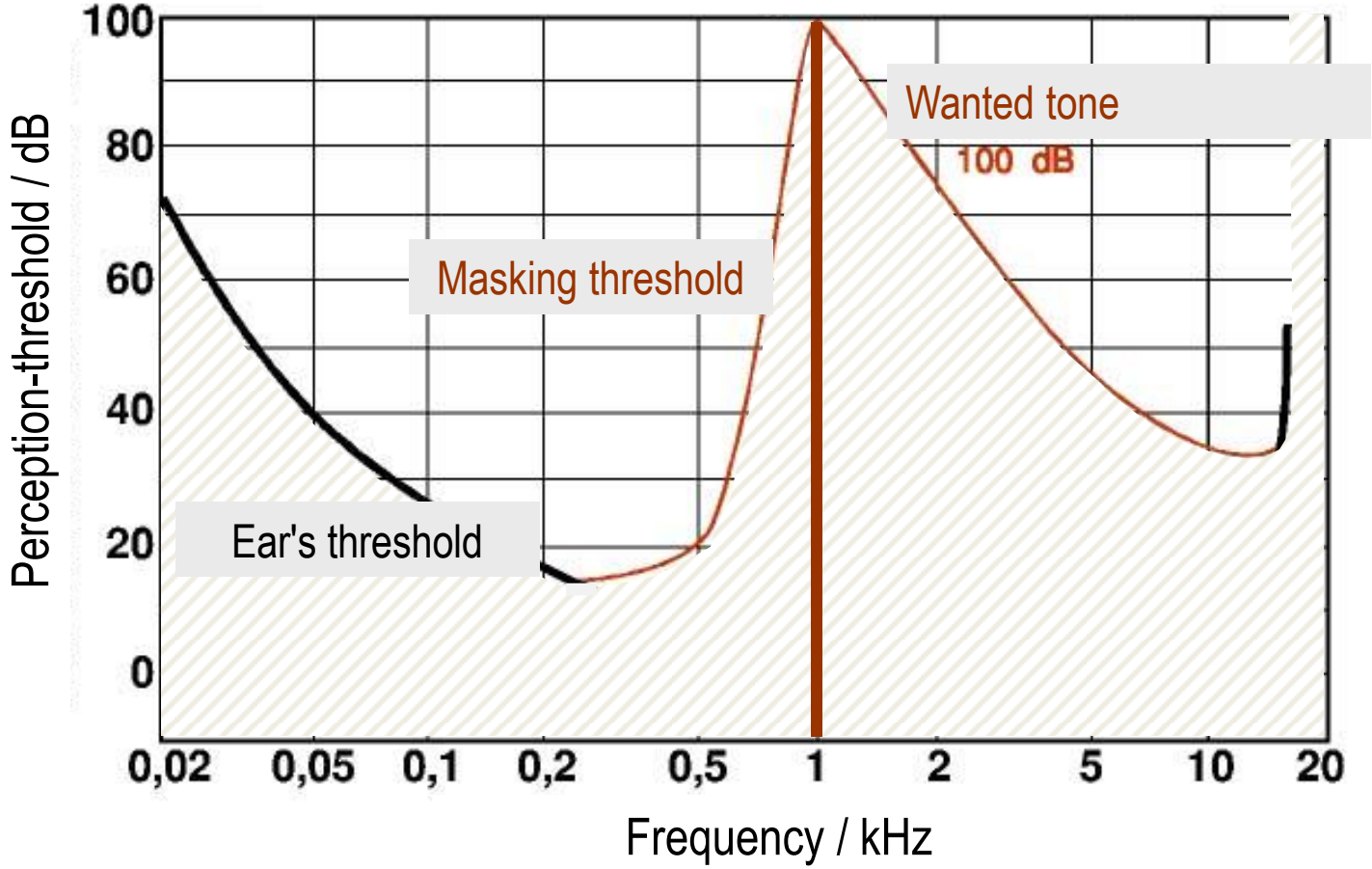
Increase of perception-threshold in the presence of interfering sine wave 1 kHz (*Zwicker / Feldkeller*)



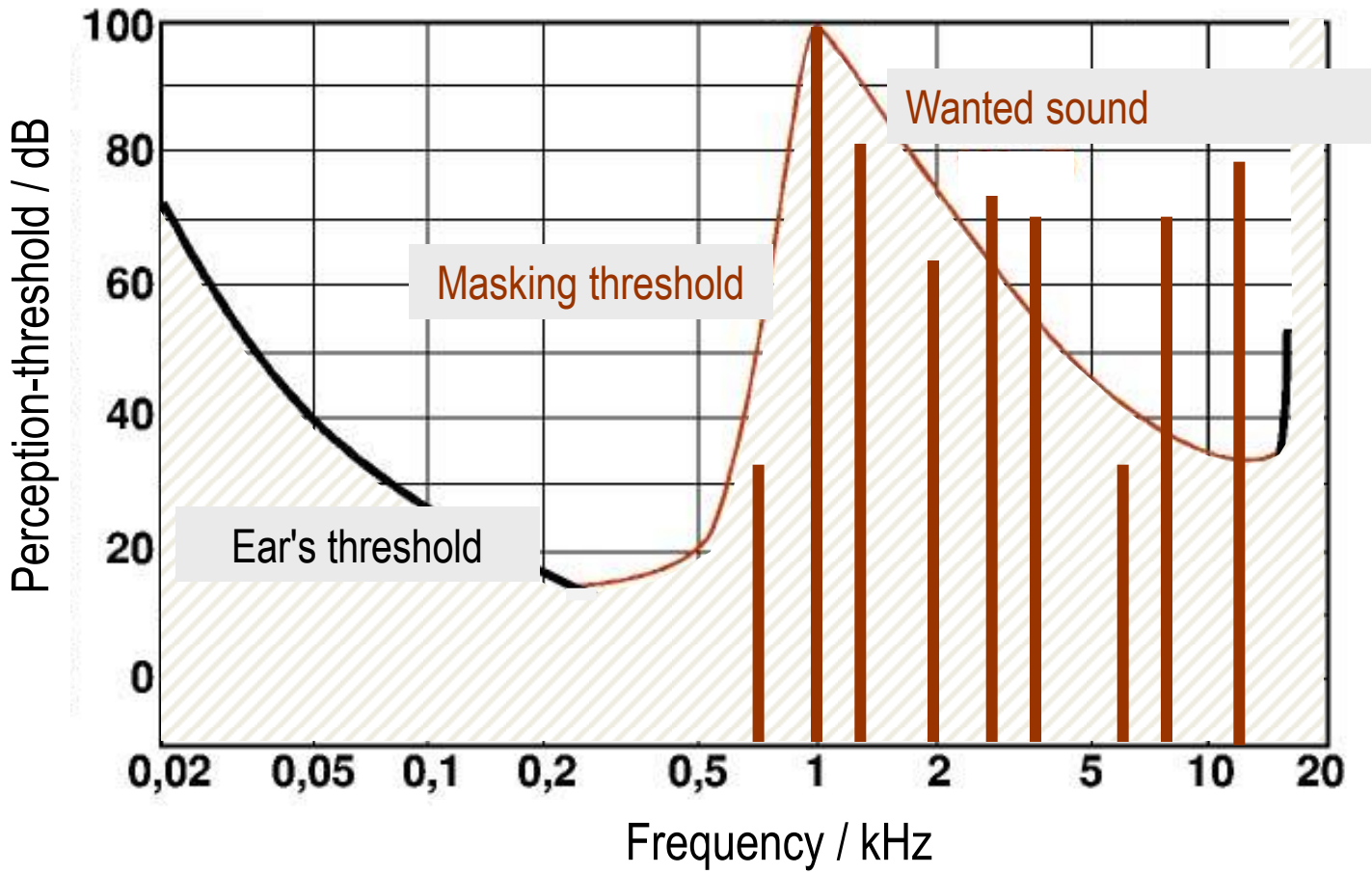
Making use of simultaneous masking



Making use of simultaneous masking

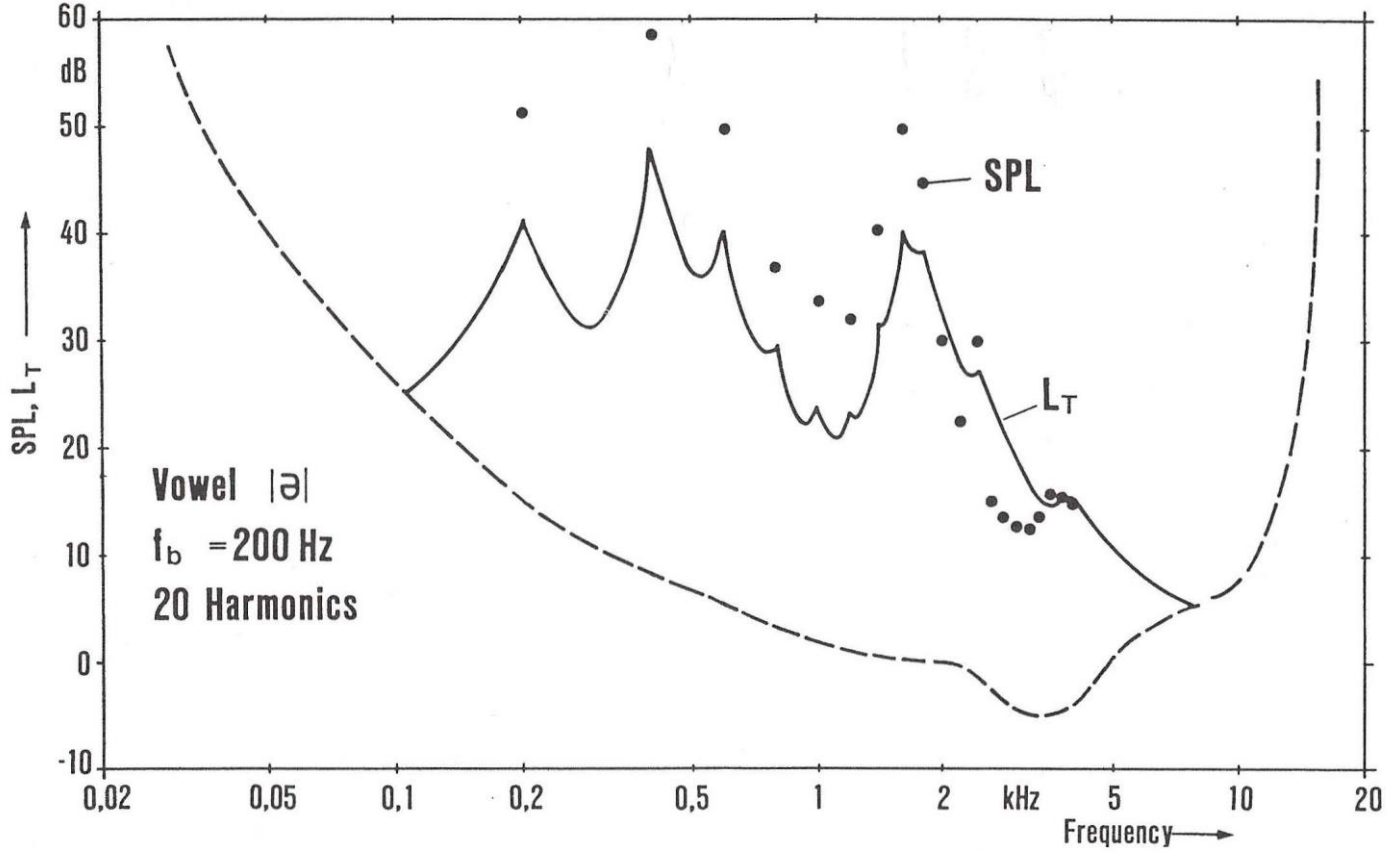


Making use of simultaneous masking

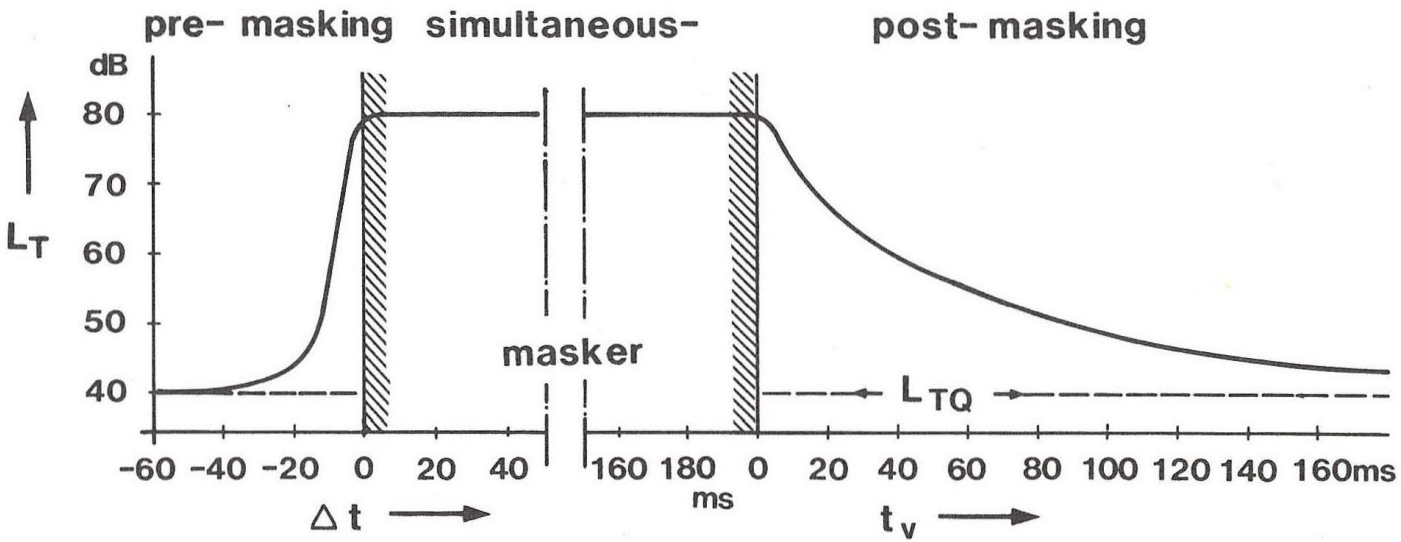


Verband Deutscher Tonmeister e.V.

Making use of simultaneous masking

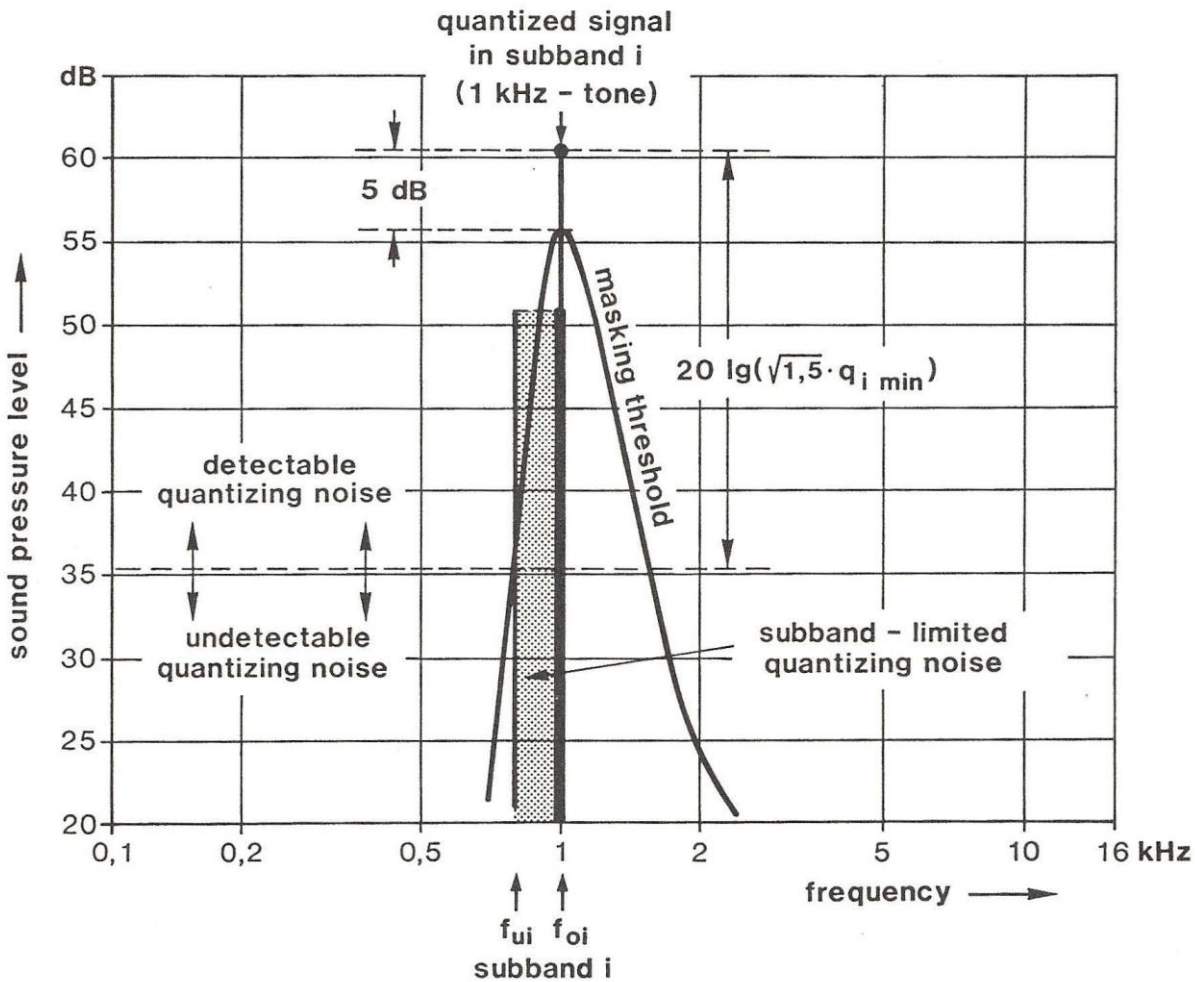


Temporal masking

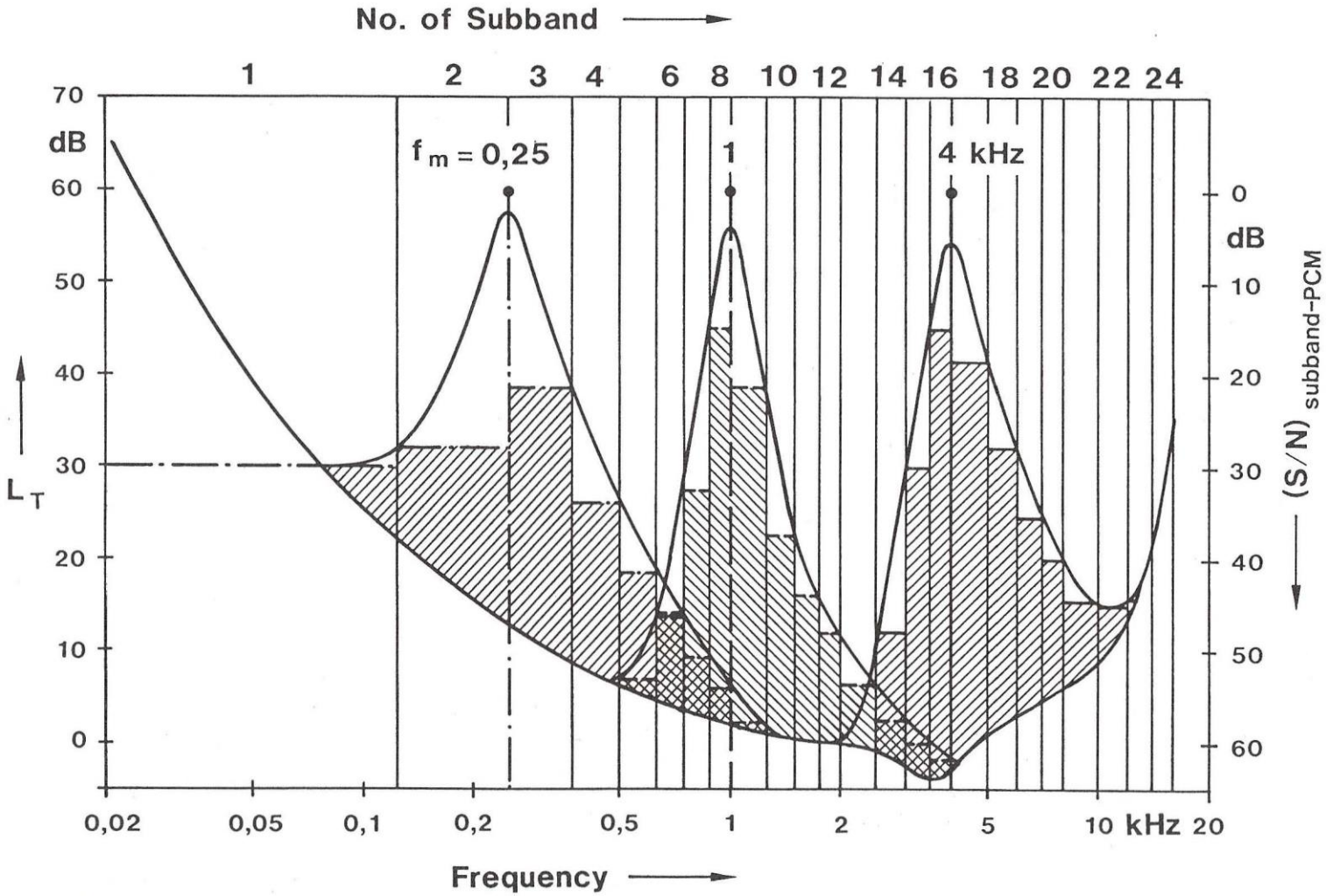


Sub-band Coding

Deriving minimum quantization $q_{i\,min}$ within one subband



Sub-band Coding



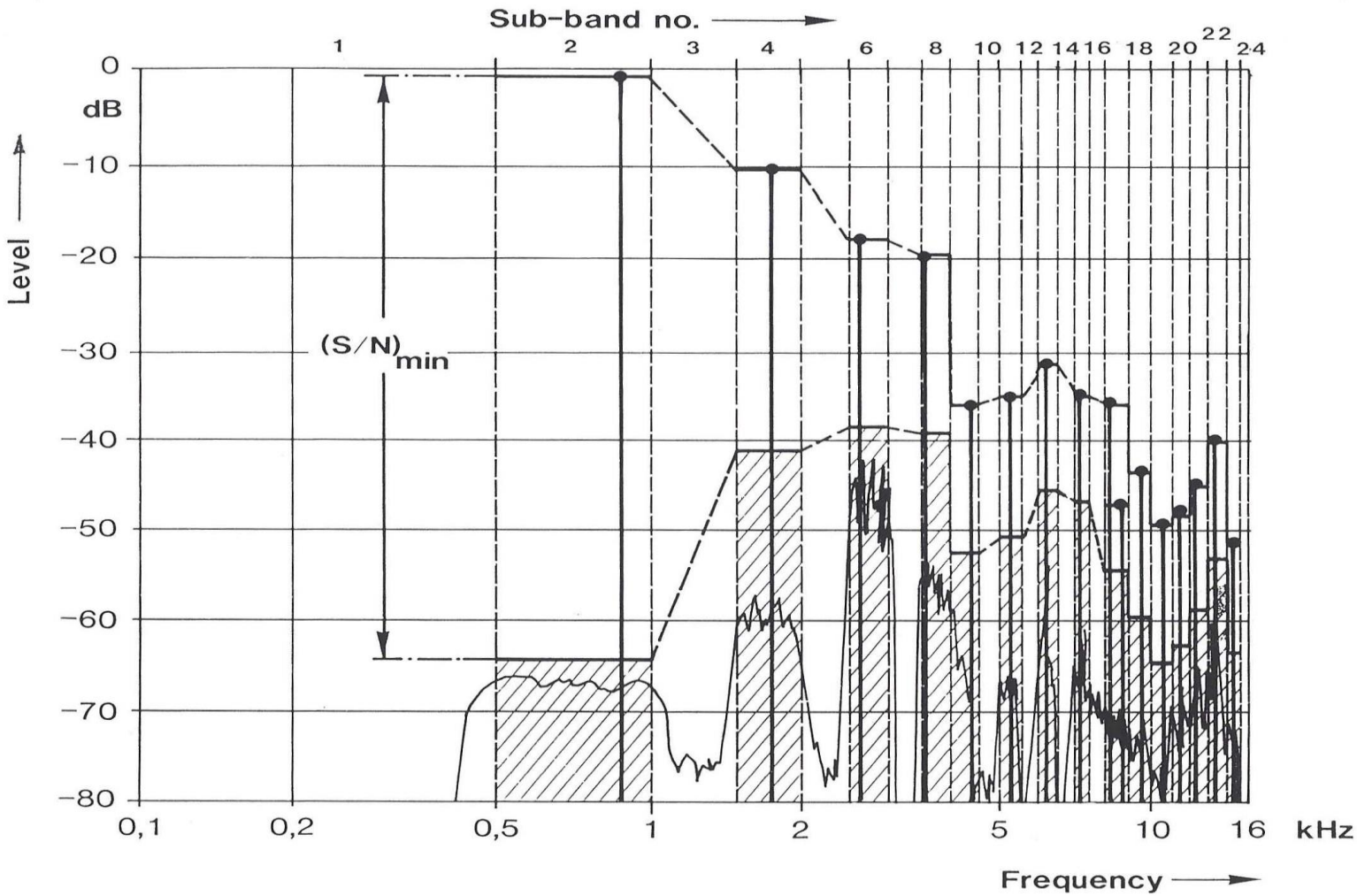
Sub-band Coding

First sub-band coding patent DE 3440613 C 1 (*Theile 1984*)

- Distribution of quantizing noise and bit-error noise across the signal spectrum with respect to masking effects of the human auditory system
- Band-splitting adapted to the psycho-acoustical "critical bands"
- Quantizing resolution (bits/ sample) of each sub-band signal adapted to the "masking pattern“

Sub-band Coding

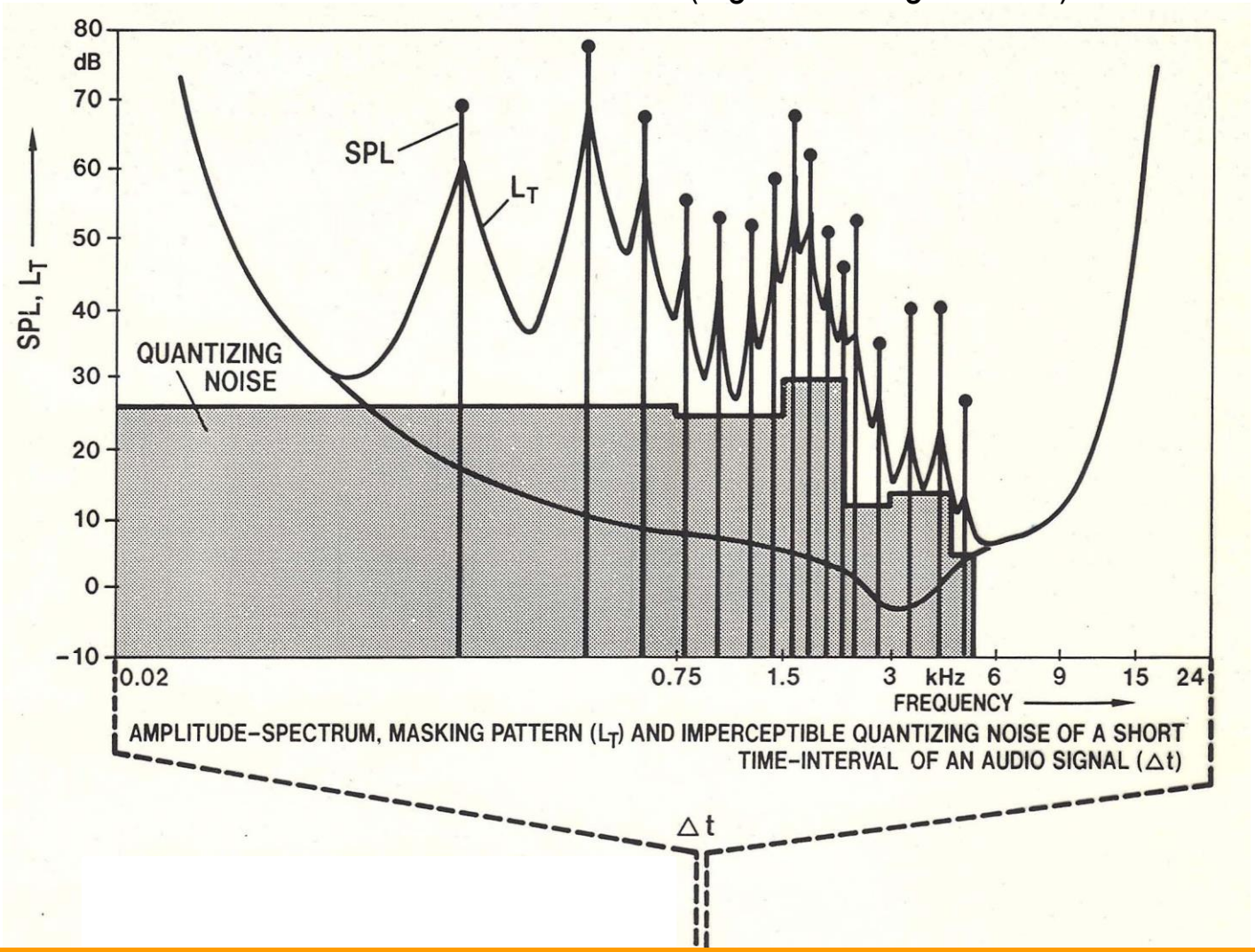
Amplitude spectra of the original signal and quantizing noise



Verband Deutscher Tonmeister e.V.

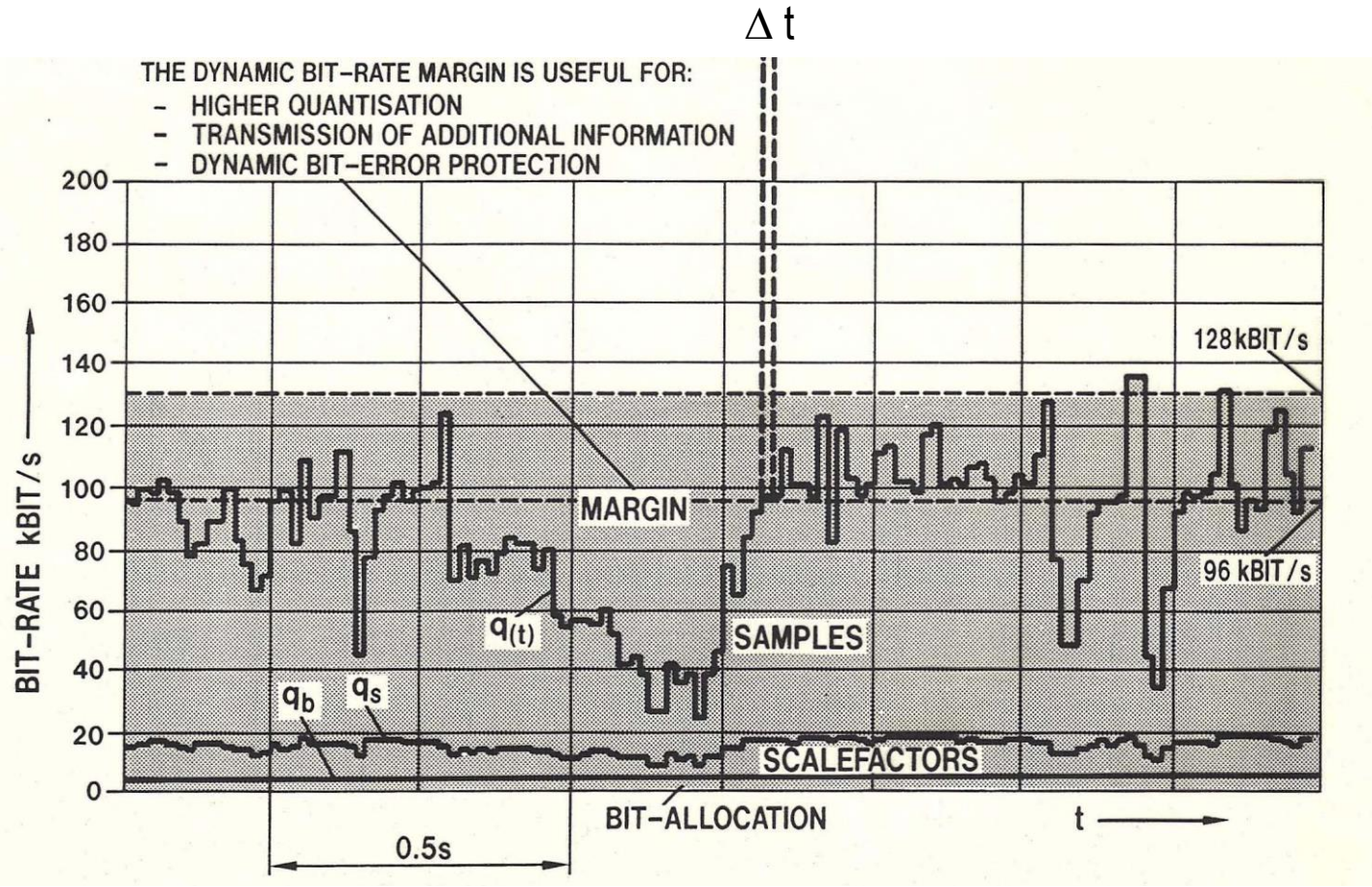
Sub-band Coding

Amplitude spectrum, masking pattern, qunatizing noise of a short time interval Δ t (e.g. block length 24 ms)



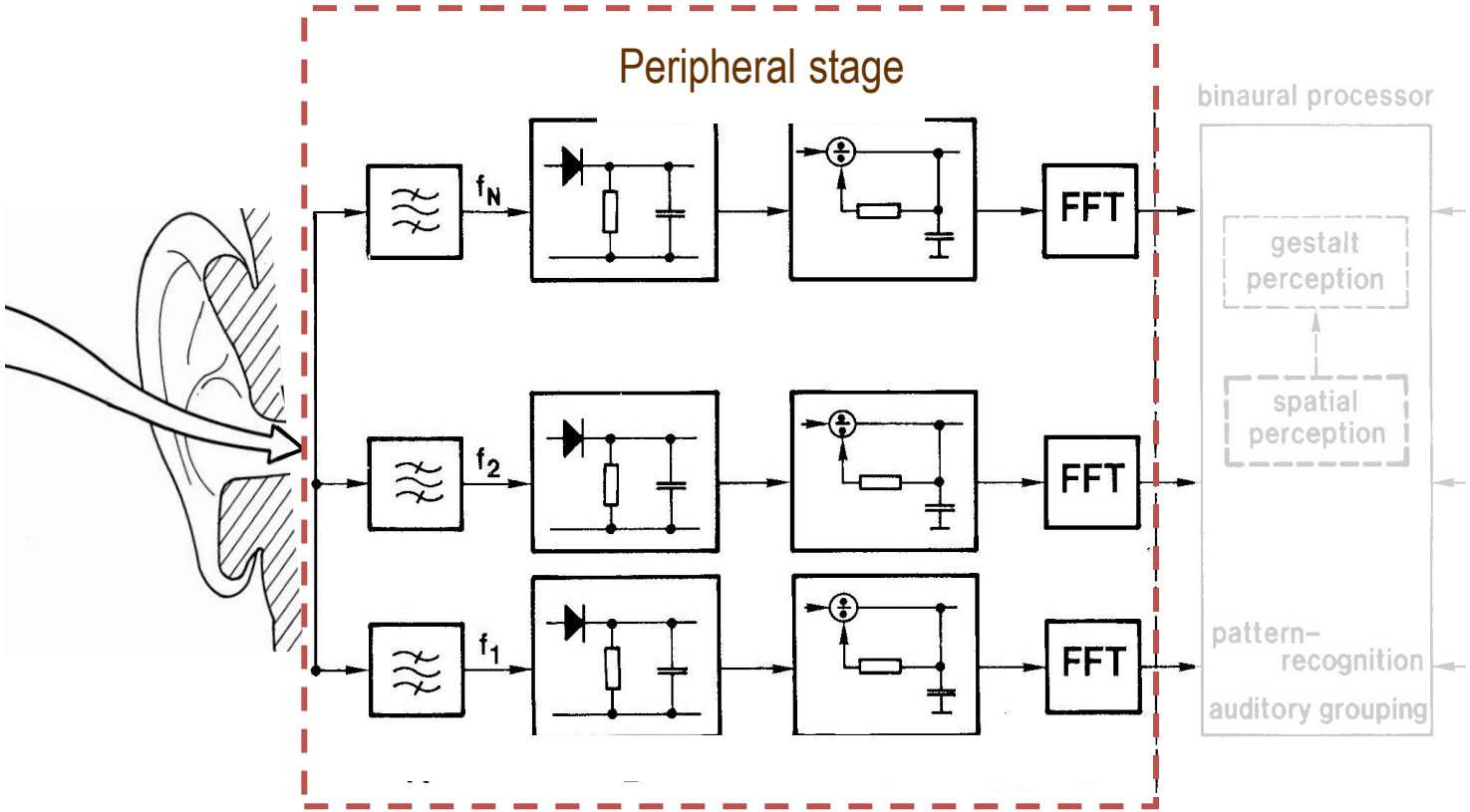
Sub-band Coding

Basic composition of an MPEG Audio Layer 2 (mp2) coded signal
static bitrate 128 kbit/s



Perceptual Coding

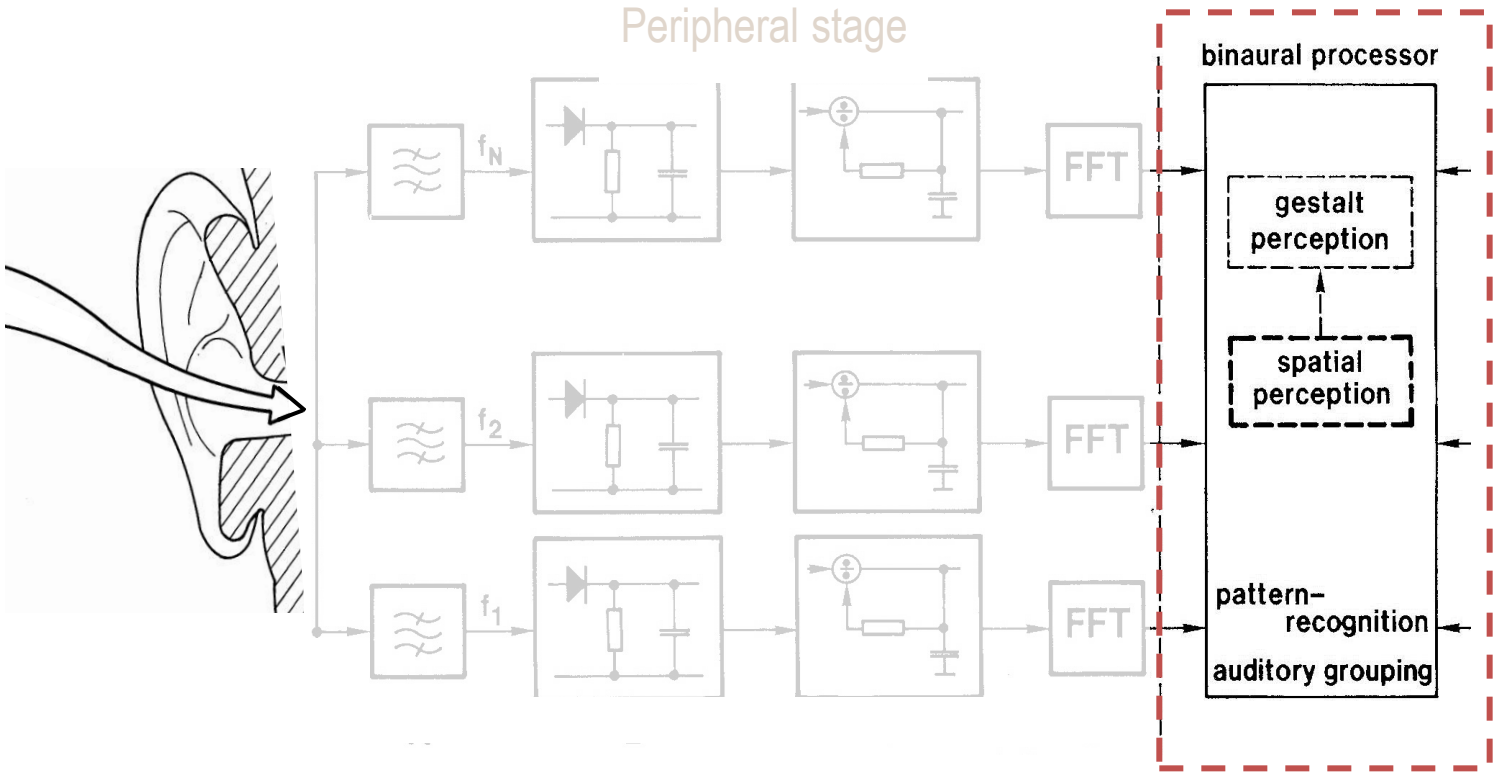
1 Mbit/s



Data reduction of auditory information in the peripheral stage

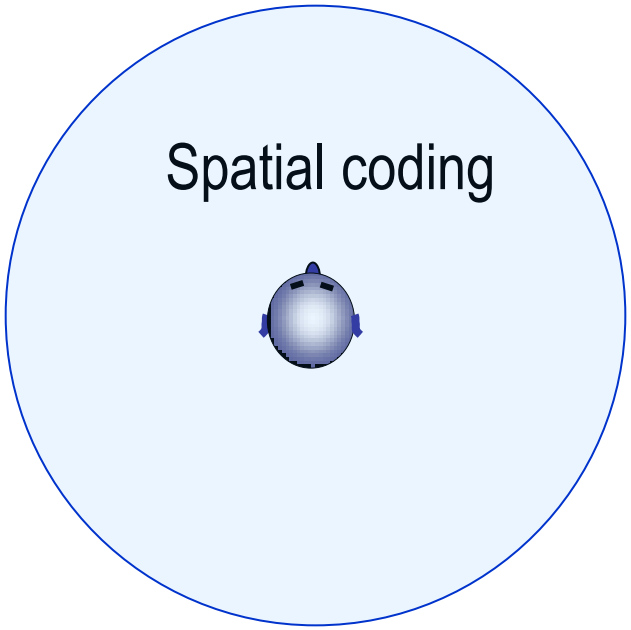
Perceptual Coding

1 Mbit/s



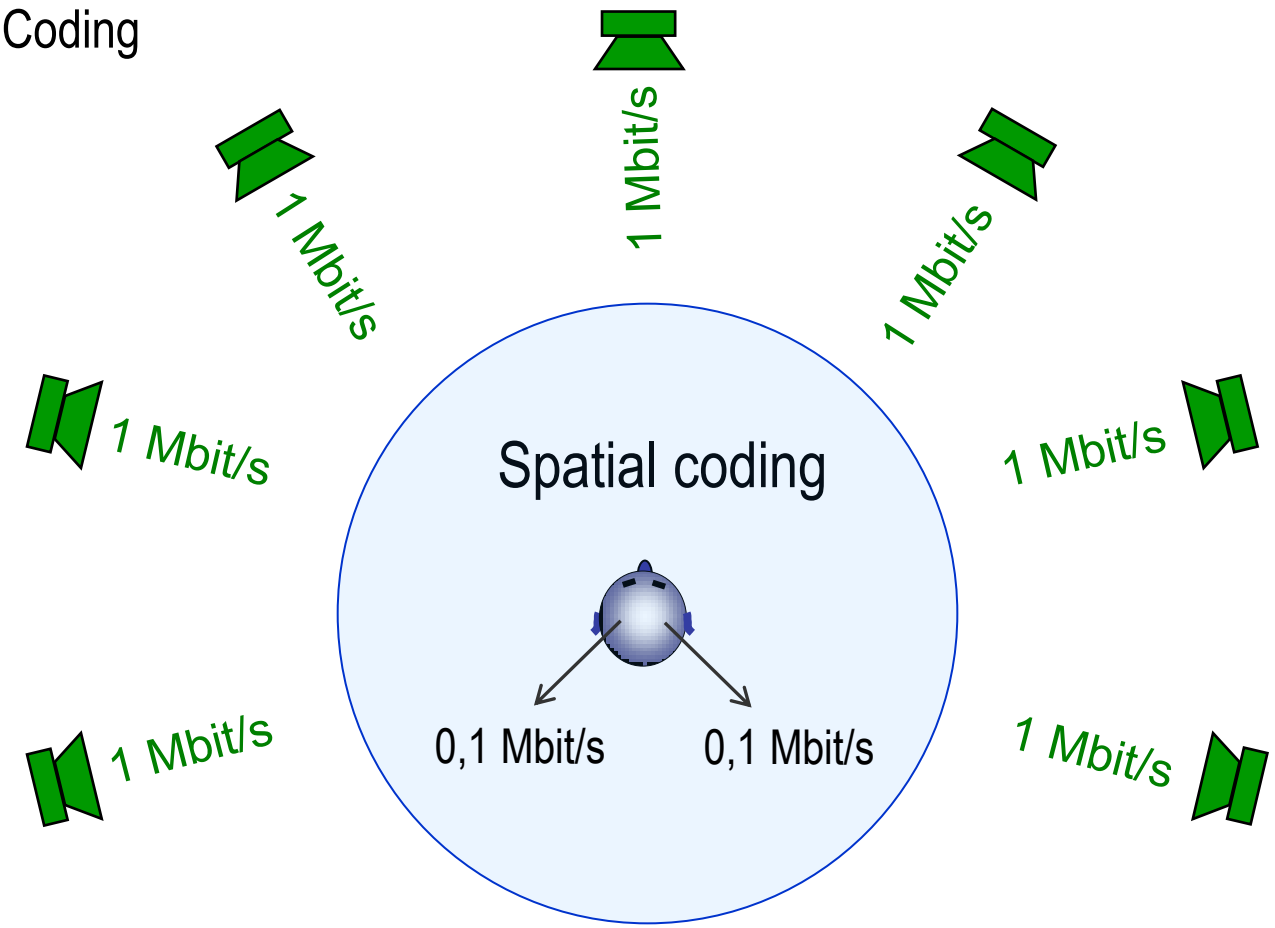
Data reduction of auditory information in the „location association stage“

Perceptual Coding



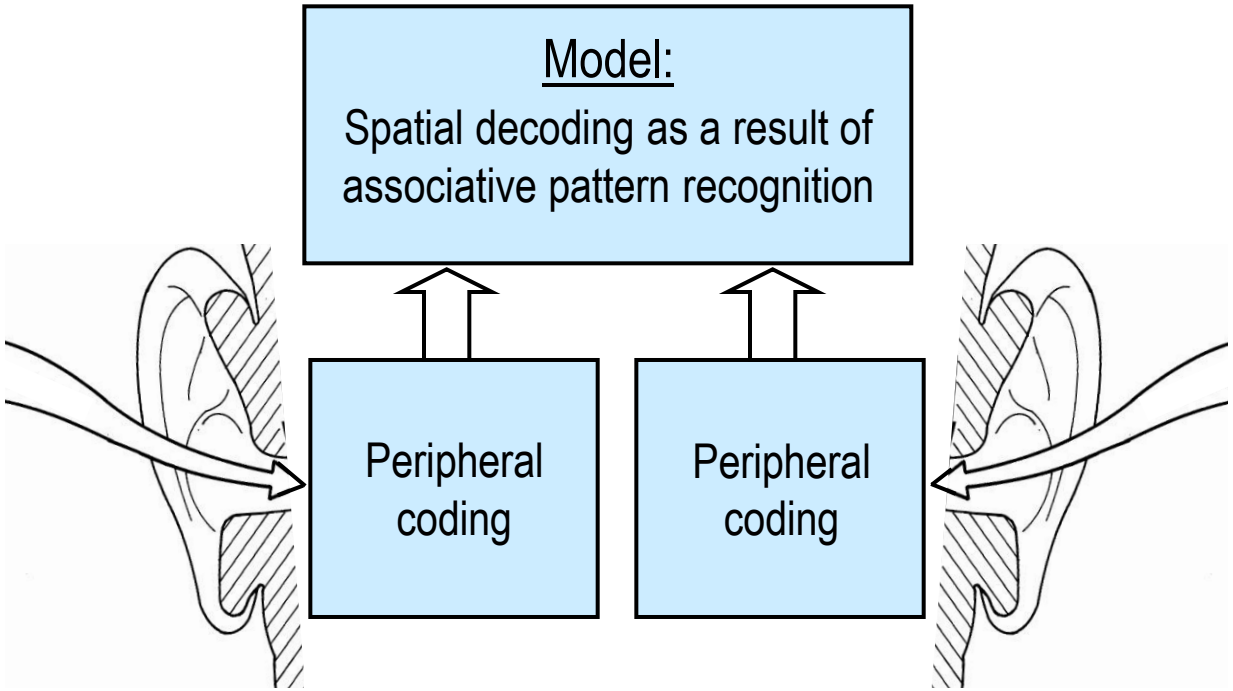
Data reduction of auditory information in the „location association stage“

Perceptual Coding



Data reduction of auditory information in the „location association stage“

Perceptual Coding

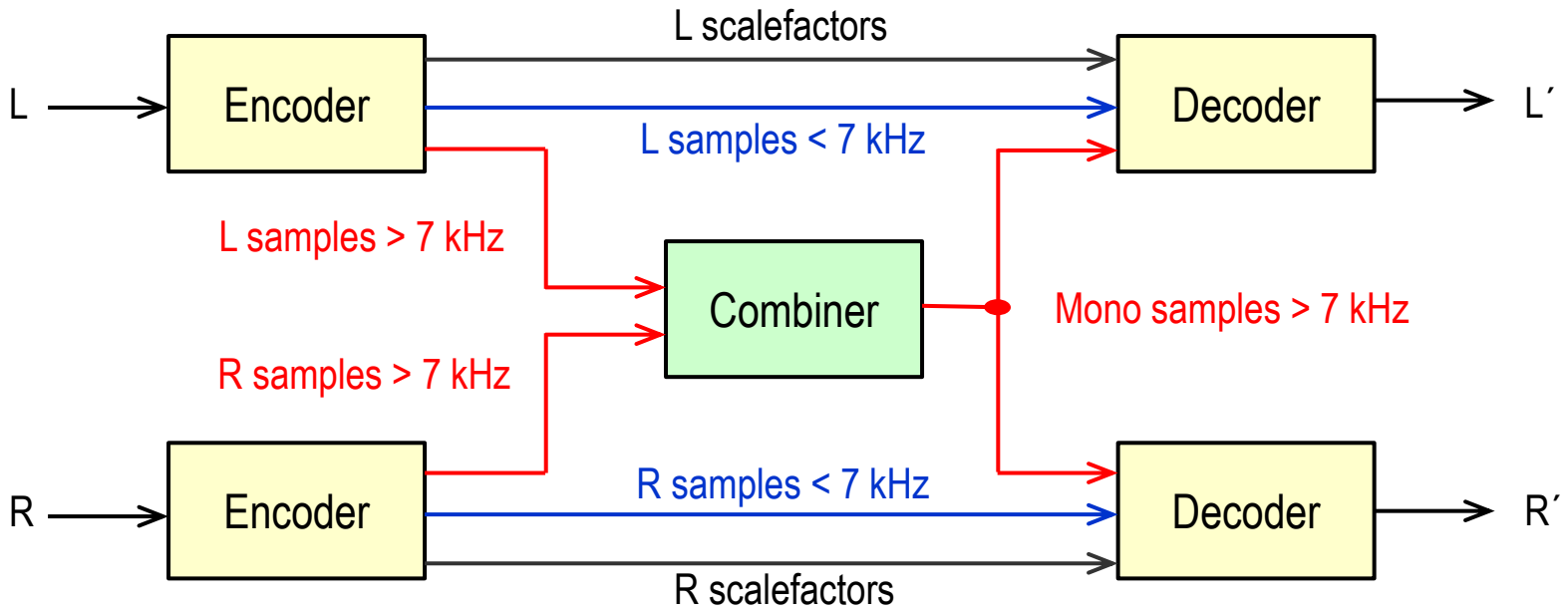


Data reduction of auditory information in the „location association stage“

Spatial Coding

- Portions of the stereophonic signals are irrelevant with respect to the spatial perception of the stereophonic presentation.
- The stereo-irrelevant signal components are not masked, however, they do not contribute to the localization of sound sources, they are ignored in the binaural processor of the auditory system.
- Therefore, stereo-irrelevant components of any stereo signal (L, C, R, Ls or Rs) may be reproduced via any loudspeaker, or via several loudspeakers, without effecting the stereophonic impression.
- This means that for certain time intervals in certain spectrum regions crosstalk is permissible.
- Certain sub-band signals to be identified by means of a signal analysis in the encoder can be set to “mono” and transmitted in only one channel.

Spatial Coding



Example: Dynamic crosstalk → "Intensity stereo mode" *(Theile 1992)*

Multi-channel Coding Standard ISO/IEC 11172-3

Layer 1:
Basic mapping of the digital audio input into 32 subbands, fixed segmentation to format the data into blocks, a psychoacoustic model to determine the adaptive bit allocation, and quantization using block companding and formatting.

Layer 2:
Additional coding of bit allocation, scalefactors and samples. Different framing.
Joint stereo and intensity stereo mode.

Layer 3:
Increased frequency resolution based on a hybrid filter bank. Different (nonuniform) quantizer, adaptive segmentation and entropy coding of the quantized values.

Verband Deutscher Tonmeister e.V.

Associative pattern recognition

- Is a process that associates a current pattern with a stored pattern, even if only parts of the stored pattern are included in the current pattern
- therefore, it acts as a highly effective filter for reduction of information
- allows the spontaneous interpretation of the recceived information and thus the perception of auditory events mapped to the outside world.

Associative pattern recognition

Verband Deutscher Tonmeister e.V.

Definition: :

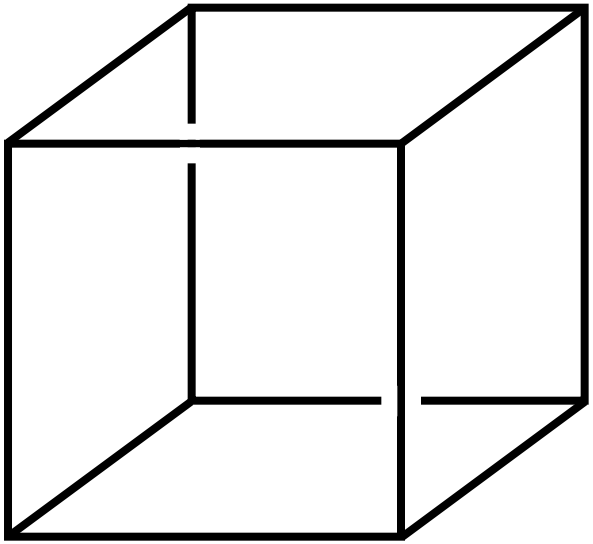
Perceptions are sensory events caused by stimuli that are mapped to the outside world as a result of sensory experience and that can therefore be influenced by learning processes, conscious and subconscious interpretations. They arise as a result of associations.

Definition:

Sensations are sensory events caused by stimuli that cannot be subdivided further and that are not affected by learning processes, conscious or subconscious interpretations.
They do not arise as a result of associations.

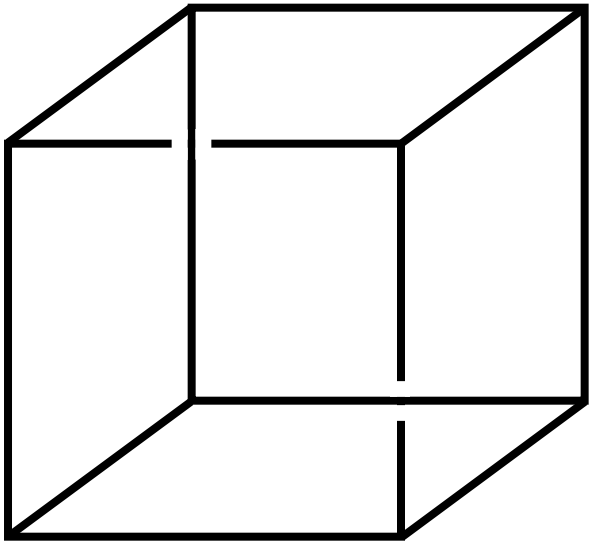
Verband Deutscher Tonmeister e.V.

Associative pattern recognition in visual perception

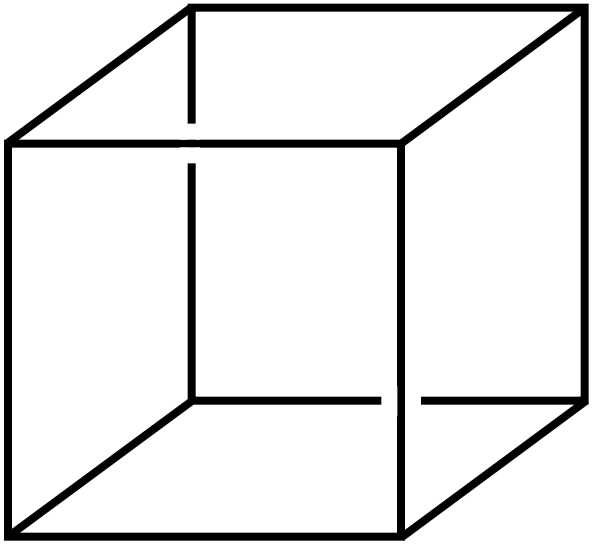


Verband Deutscher Tonmeister e.V.

Associative pattern recognition in visual perception

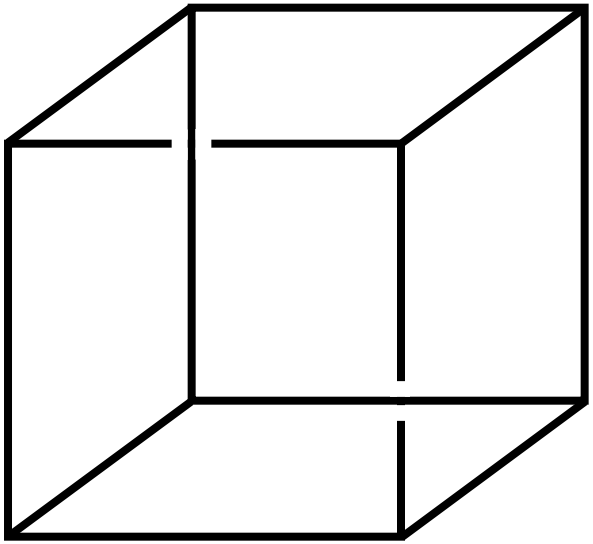


Associative pattern recognition in visual perception

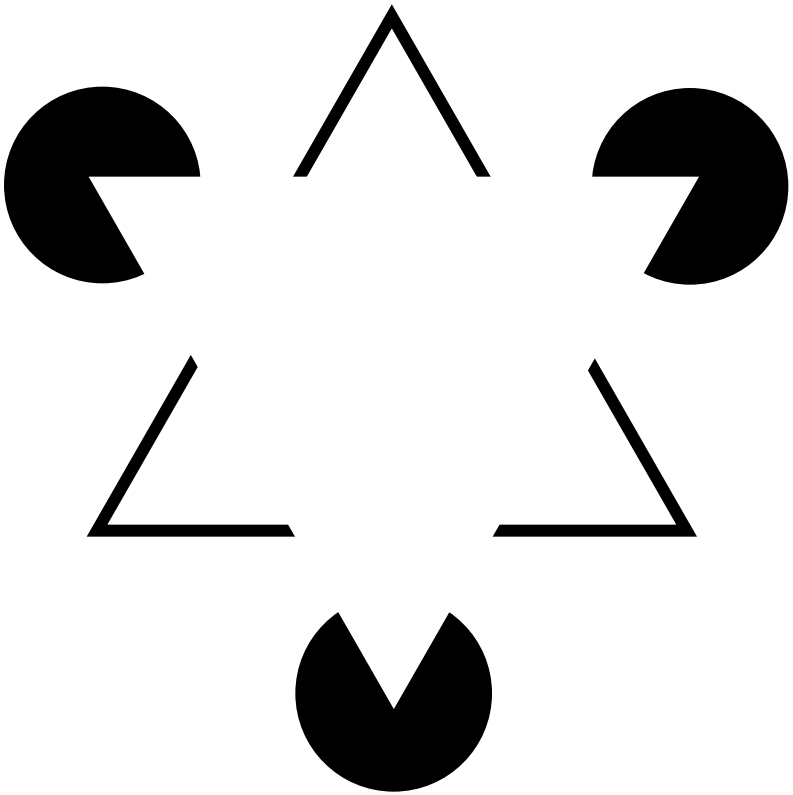


Verband Deutscher Tonmeister e.V.

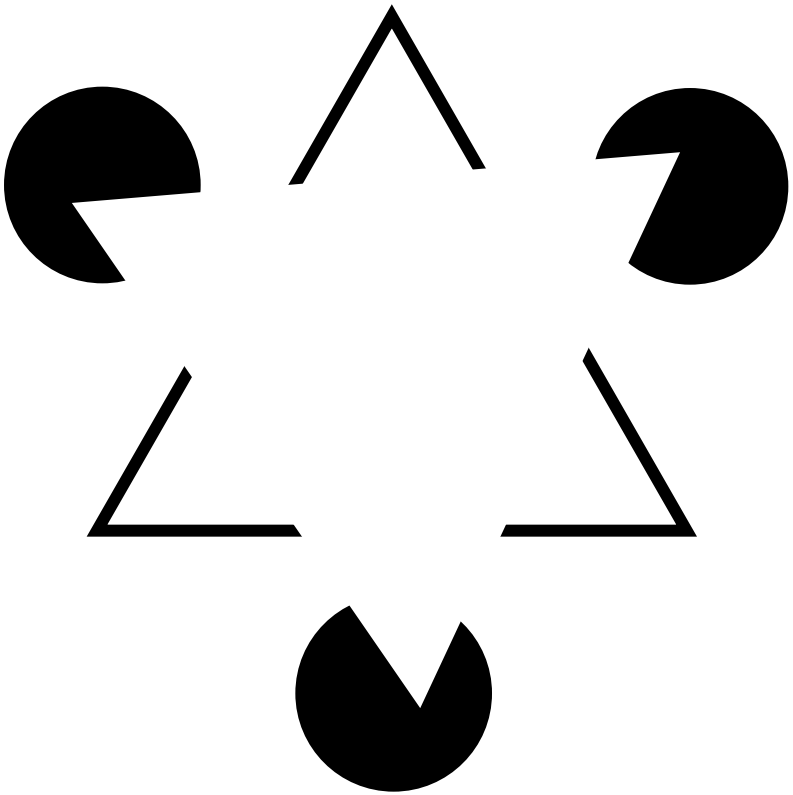
Associative pattern recognition in visual perception



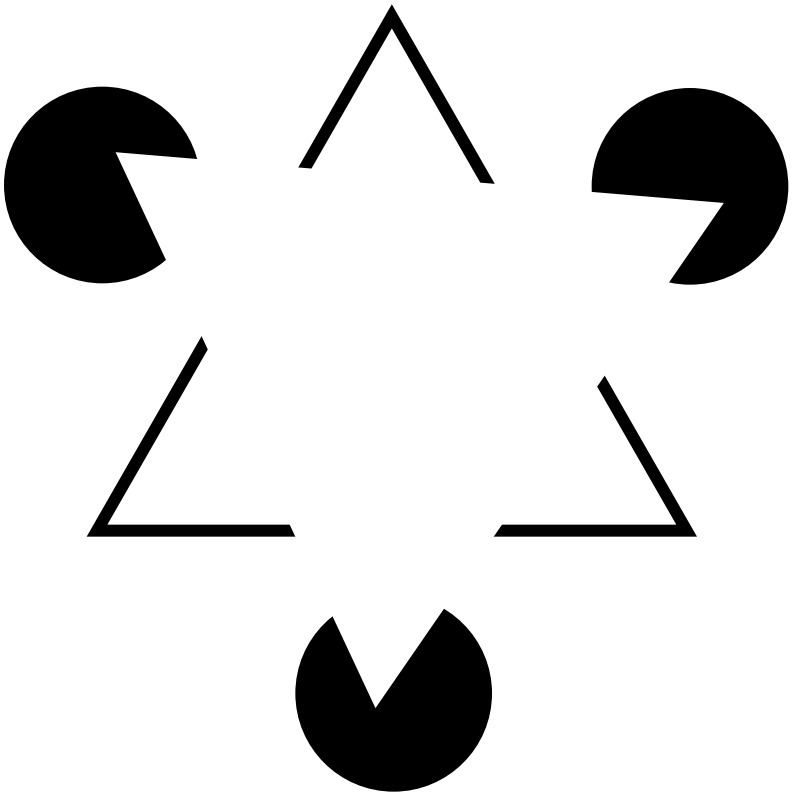
Associative pattern recognition in visual perception



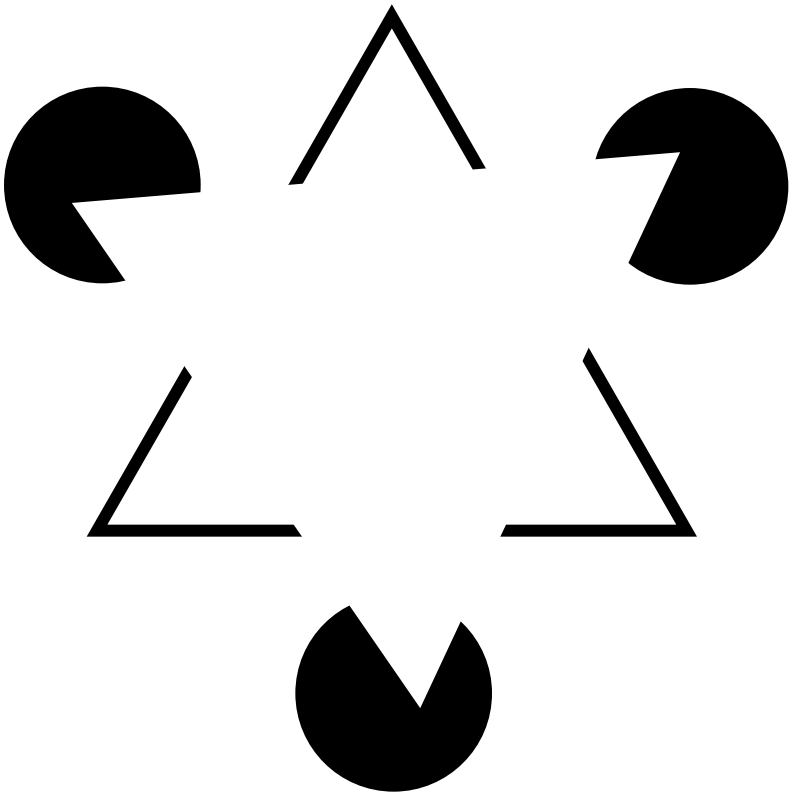
Associative pattern recognition in visual perception



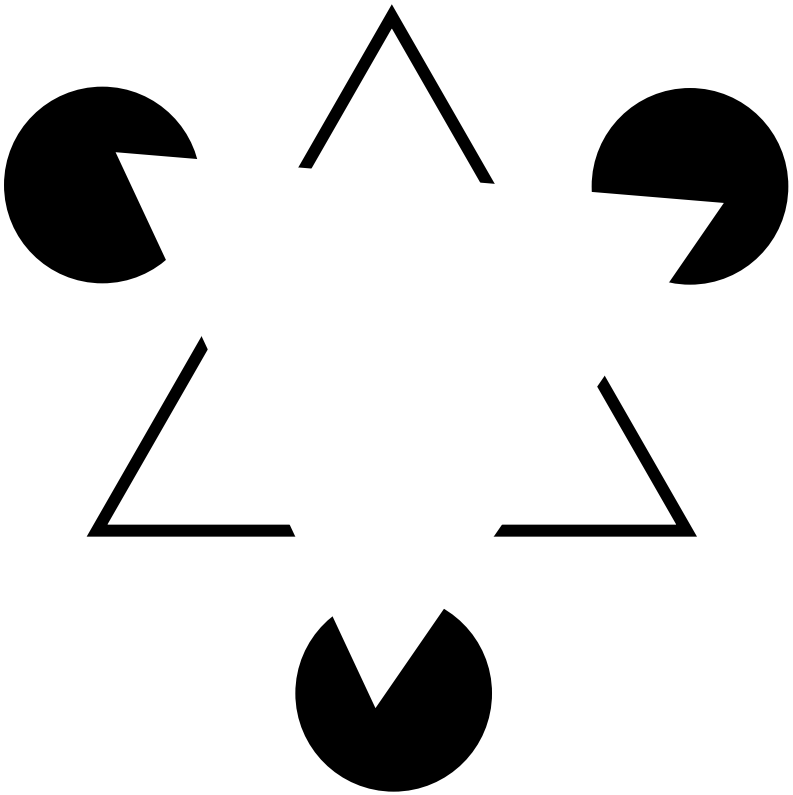
Associative pattern recognition in visual perception



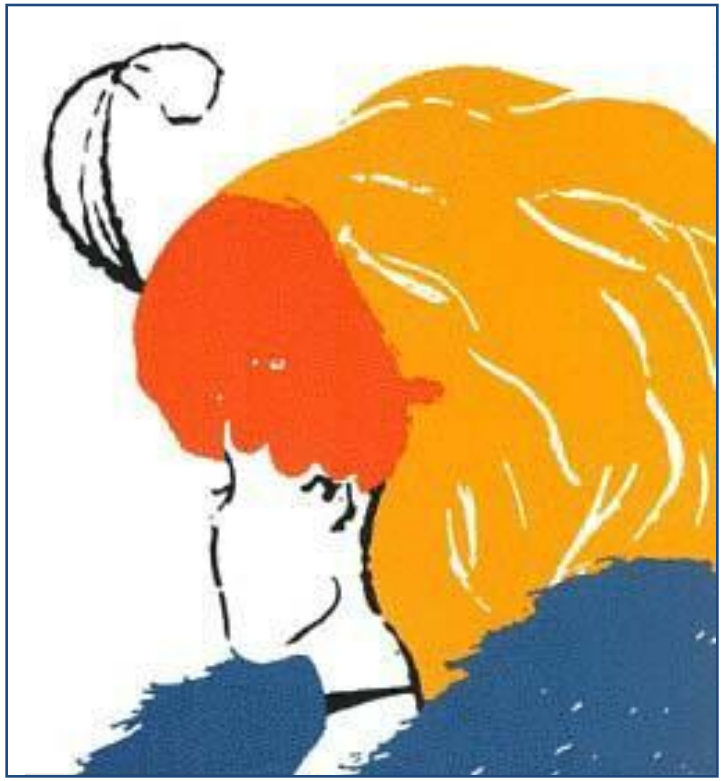
Associative pattern recognition in visual perception



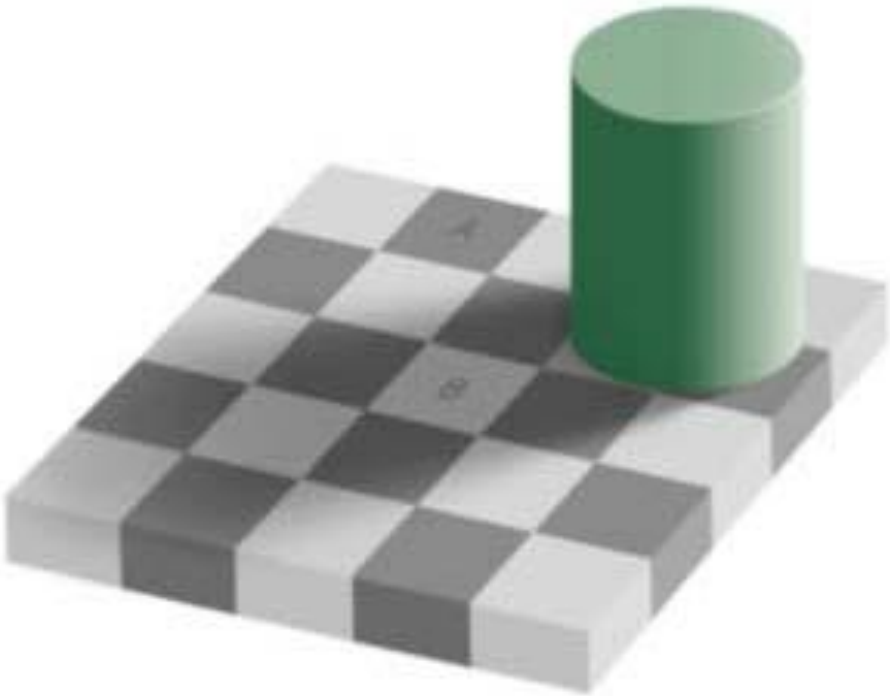
Associative pattern recognition in visual perception



Associative pattern recognition in visual perception

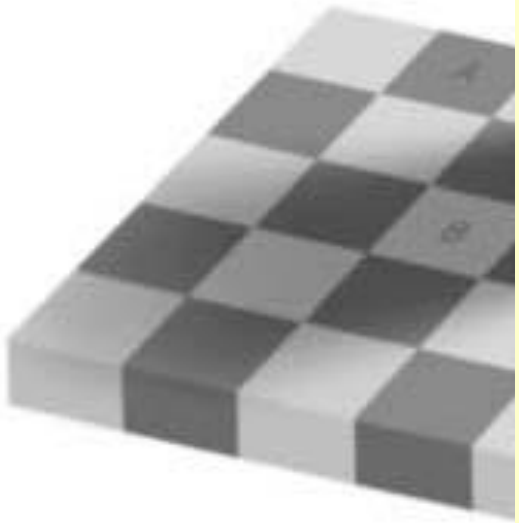


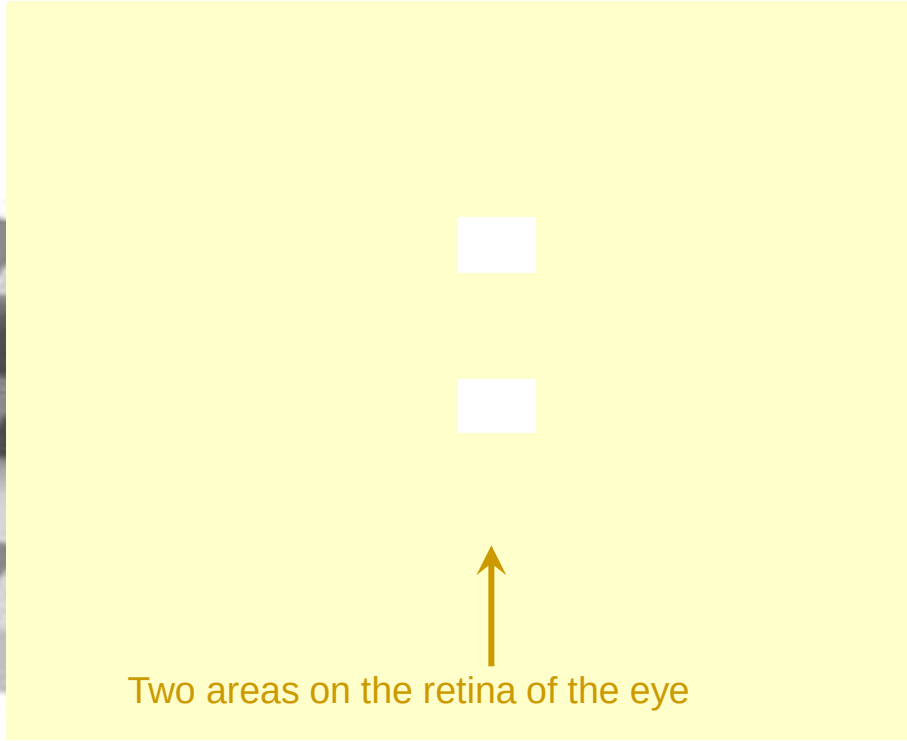
Associative pattern recognition in visual perception



A and B different lightness

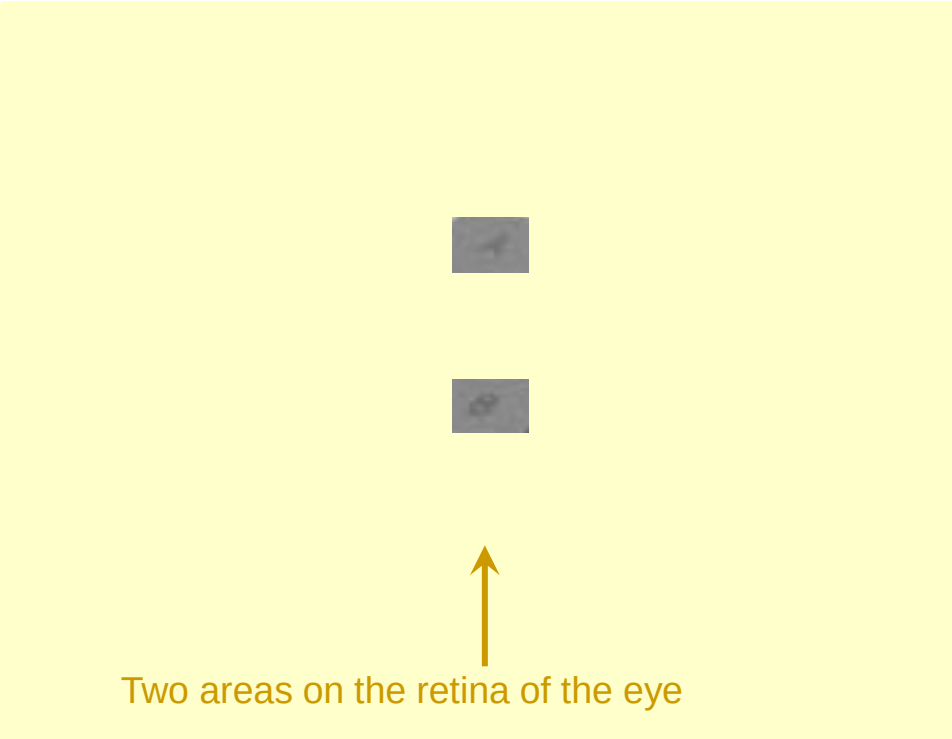
Associative pattern recognition in visual perception





A and B different lightness

Associative pattern recognition in visual perception



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

Associative pattern recognition

Associative pattern recognition in visual perception, more examples

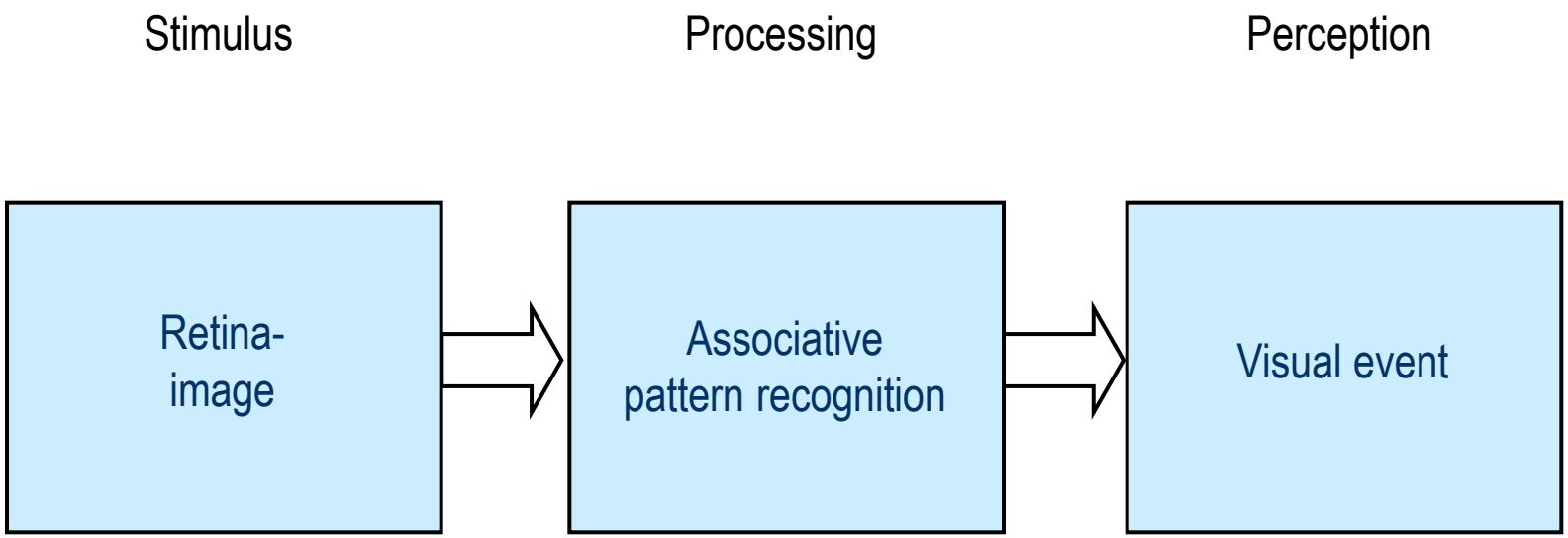
e.g. the "Ames Room" illusion, as a result of spatial pattern recognition and corresponding spontaneous depth perception.

Associative pattern recognition in auditory perception:

- Zwicker Tones (use headphones)
- Effect of visual perception, Tritone Paradox
- Shepard Tone
- Phantom Words (use headphones)
- Cognitive association

Further perceptual audio demos

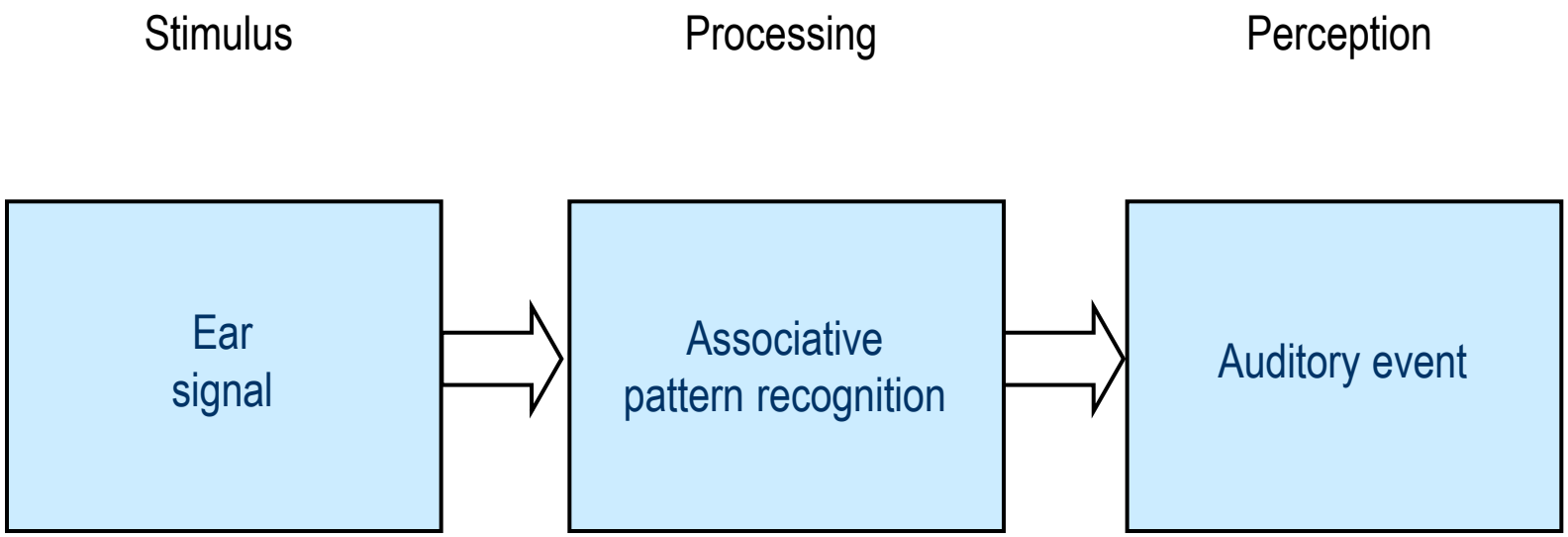
Associative pattern recognition



„Spontanous association“ → the perception can be changed by relearning

„Cognitive association“ → the perception depends on higher levels of processes (intention, attention, situation,….)

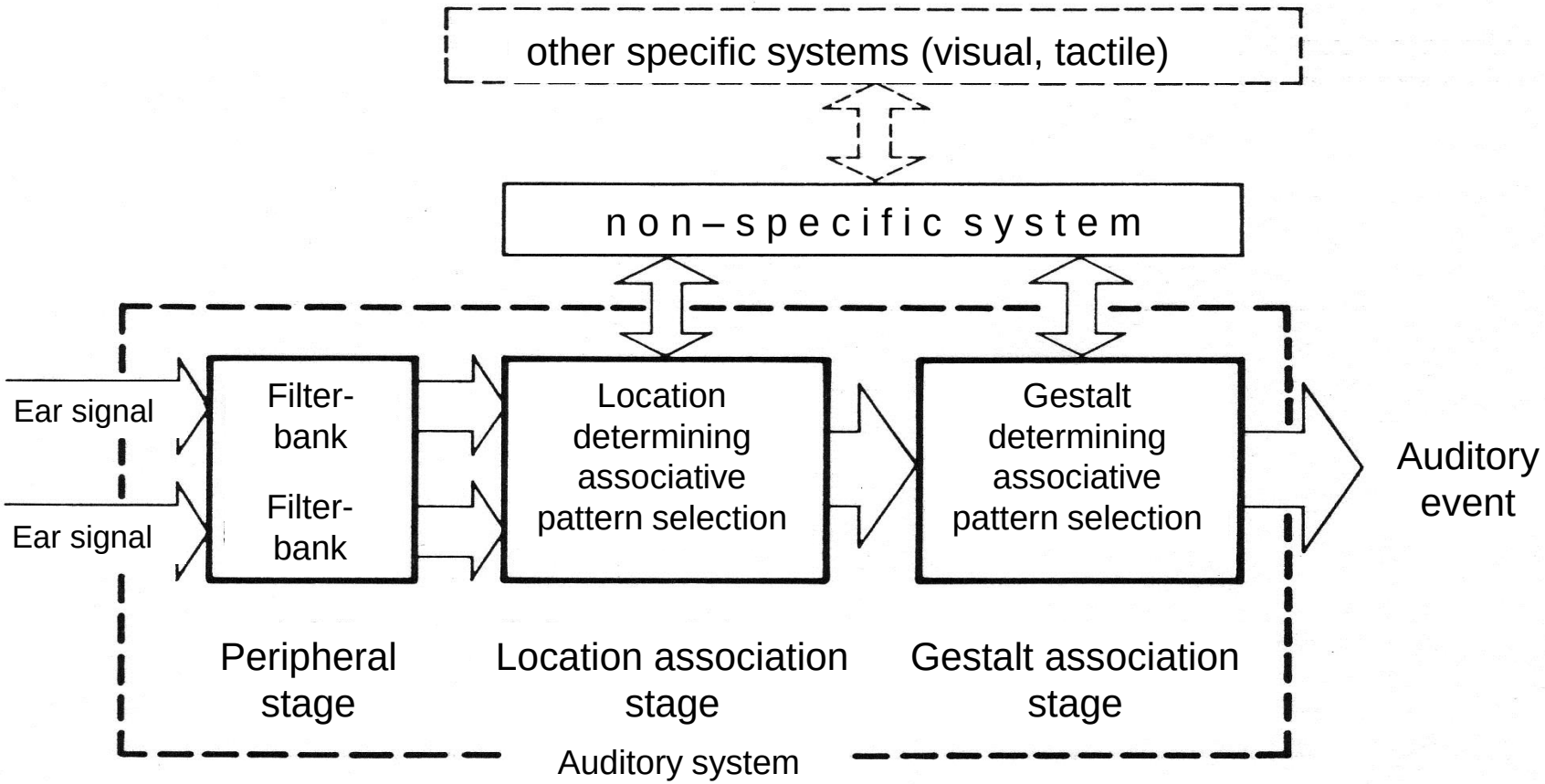
Associative pattern recognition



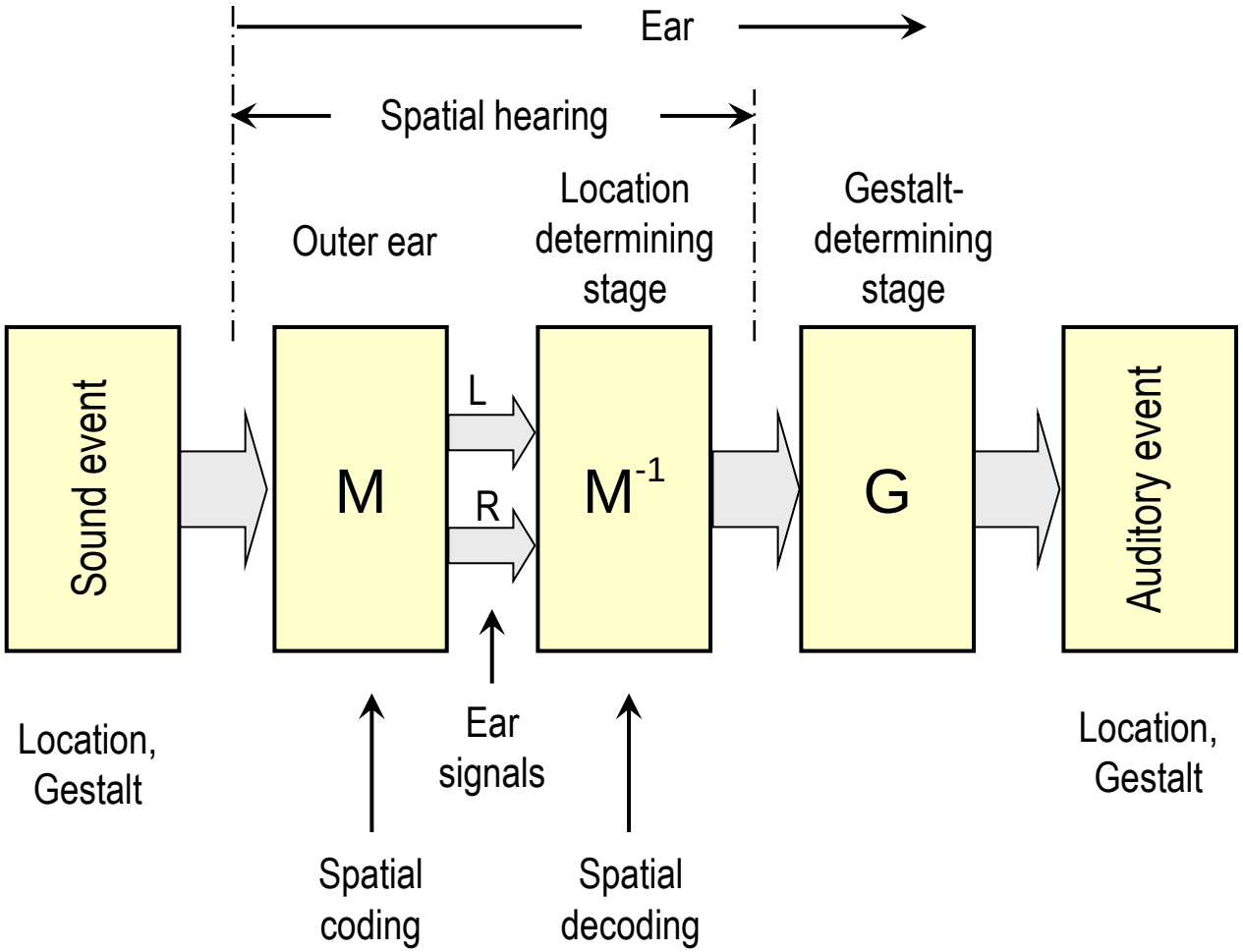
„Spontanous association“ → the perception can be changed by relearning

„Cognitive association“ → the perception depends on higher levels of processes (intention, attention, situation,….)

Association model (*Theile 1980*)



Association model

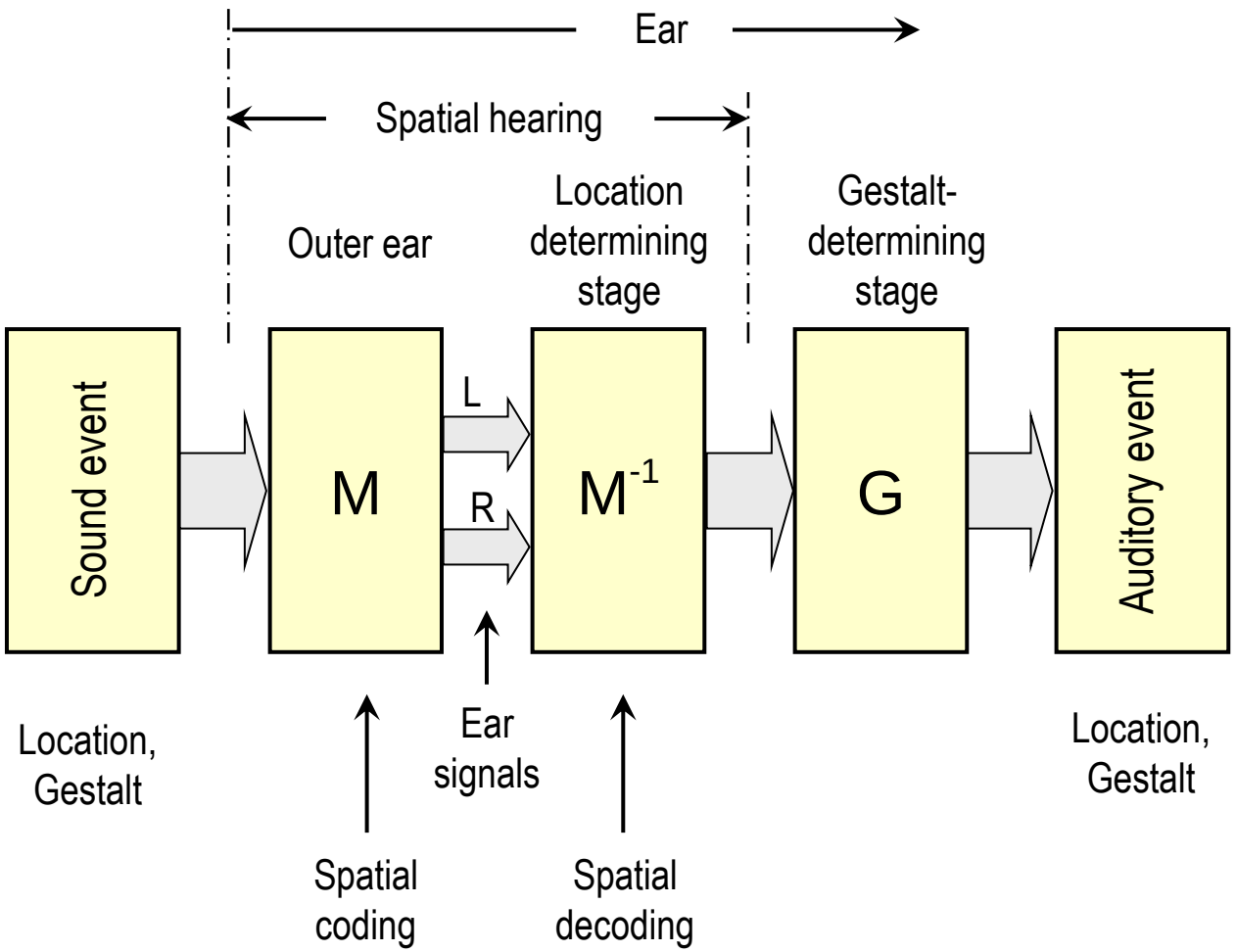


Association model

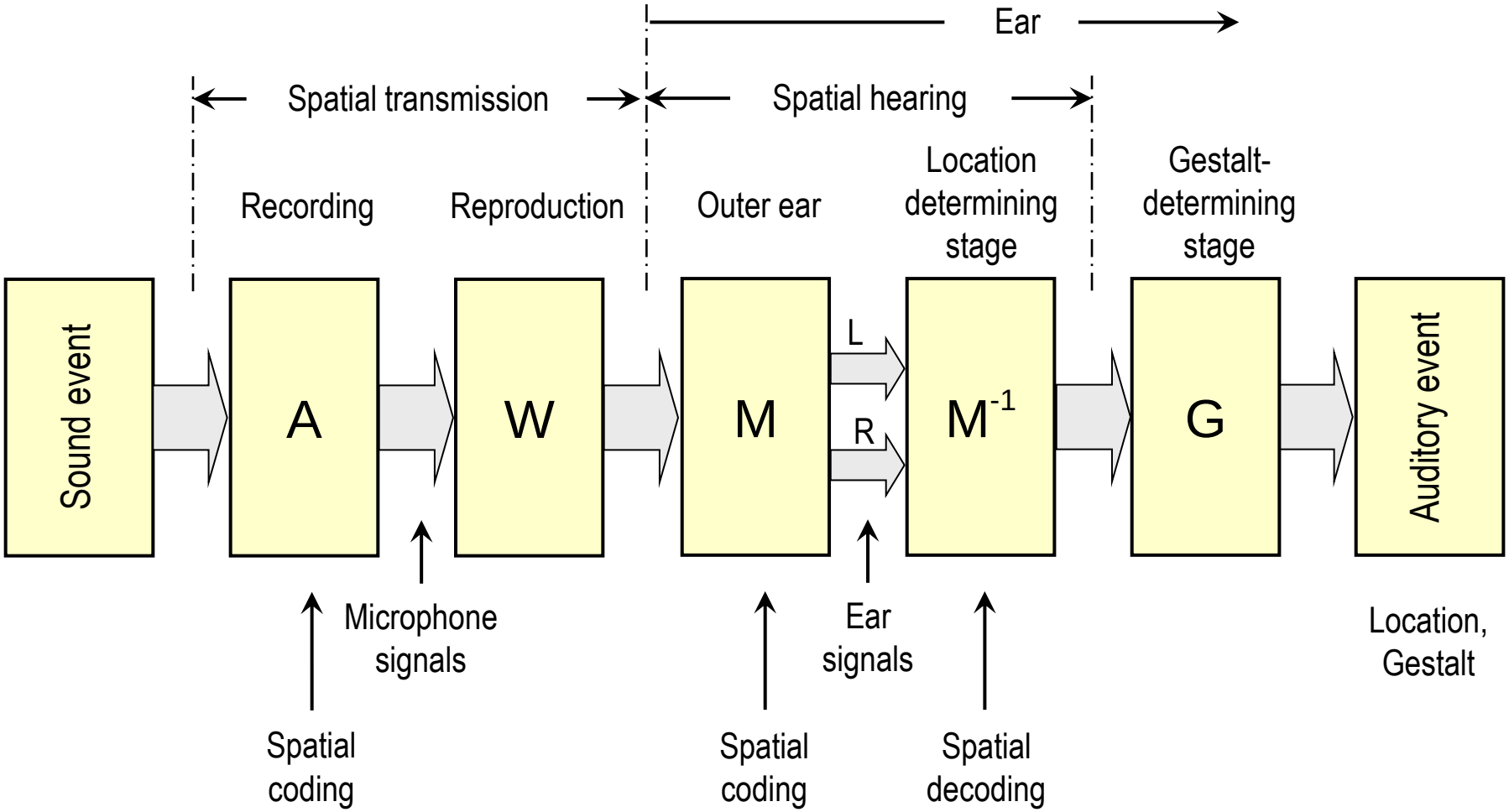
Association model provides explanations of some important auditory phenomena :

- Localisation
  Due to learned pattern
- Phantom source
  Timbre, direction, distance
- Stereophonic loudspeaker reproduction
  Stereophonic sound image caused by inverse filtering
- Lateralisation
  In-head phantom source
- Presedence effect
  Law of the first localisation stimulus

Spatial transmission function chain



Spatial transmission function chain

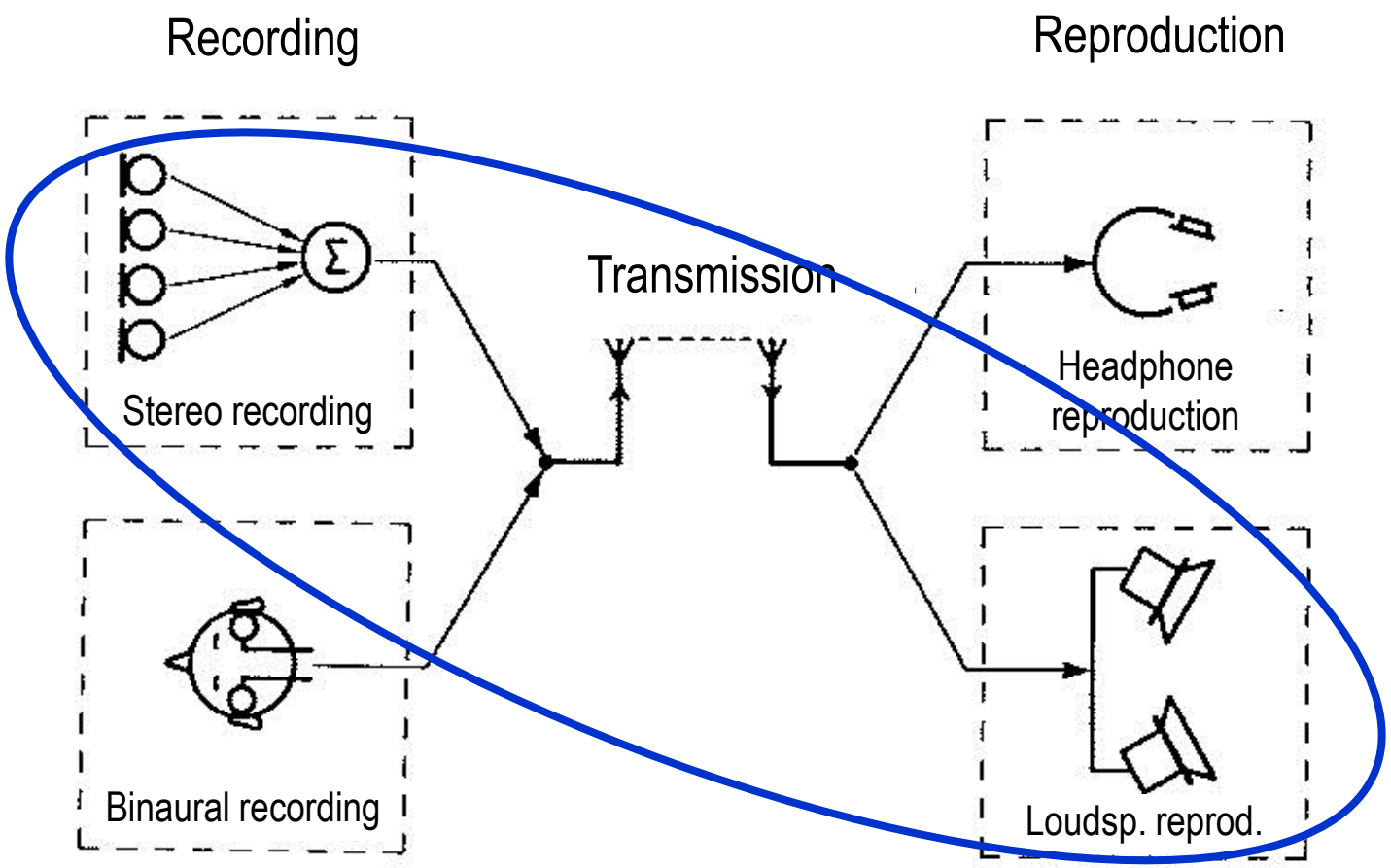


Spatial transmission methods

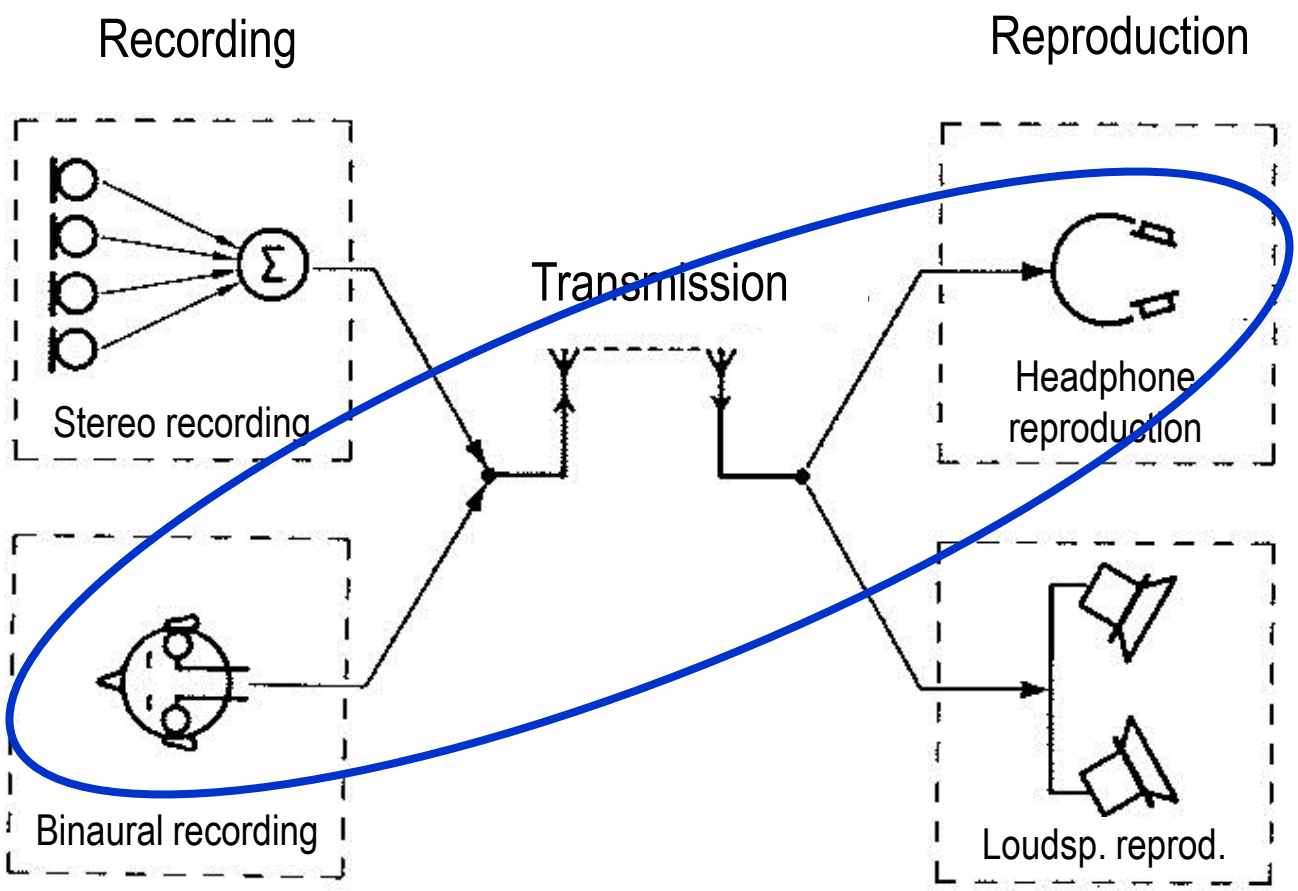
- Binaural reconstruction of the ear input signals
- (Multi-channel) loudspeaker stereophonic representation
- Multi-channel loudspeaker representation of single sources
- Sound field reconstruction of acoustic scenes

Hybrid systems can combine advantages of both methods in practical applications

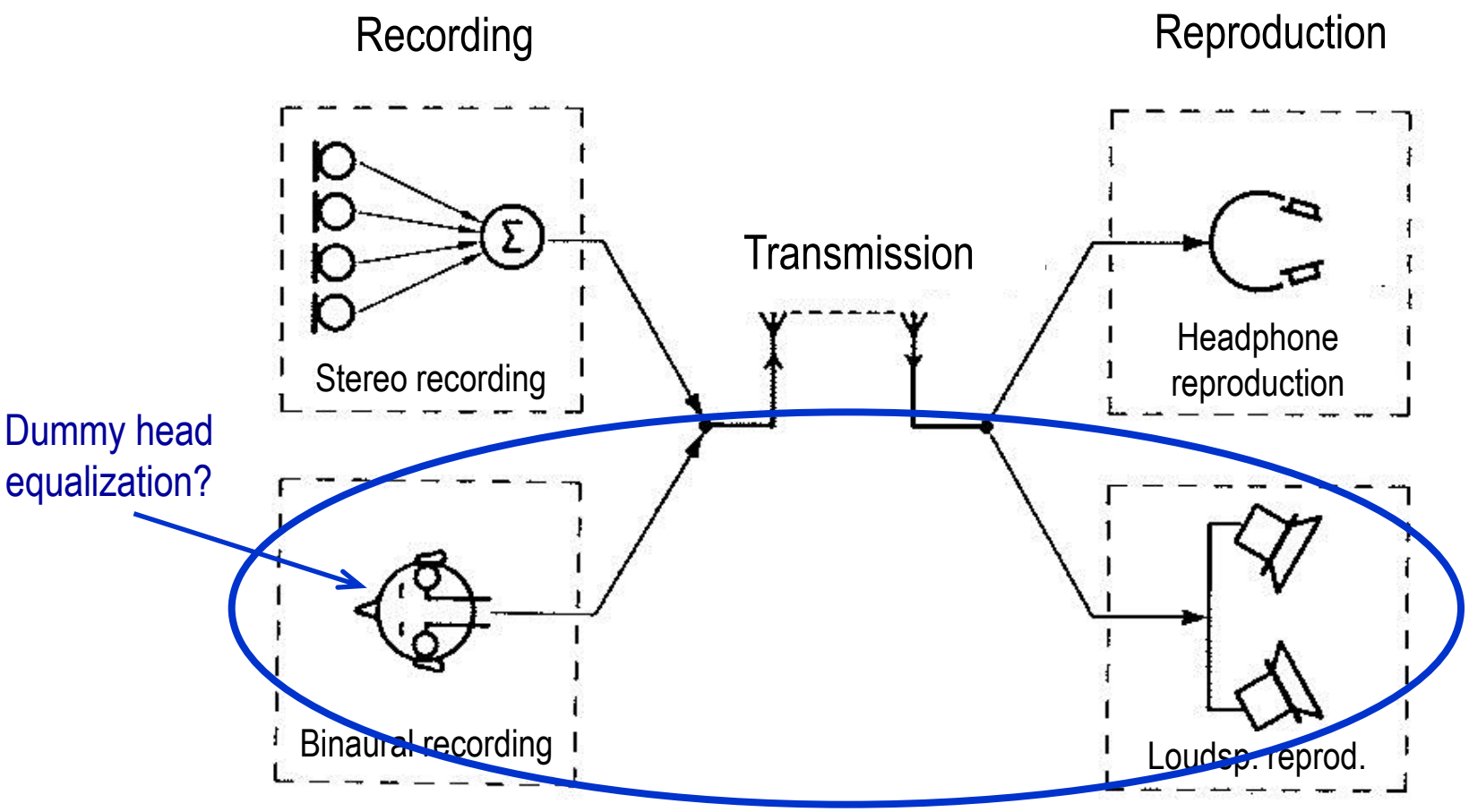
Stereo recording – stereo reproduction



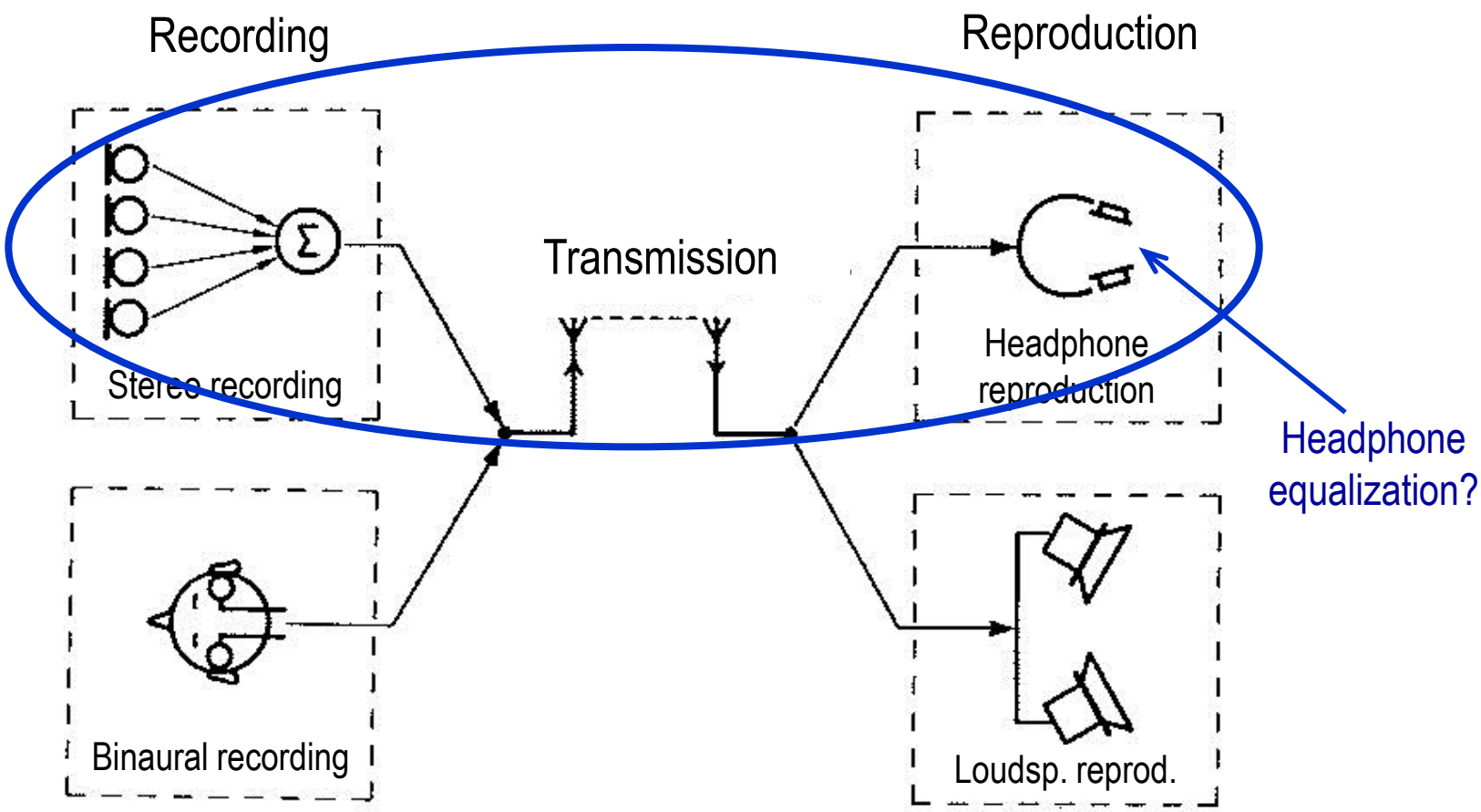
Binaural recording – headphone reproduction



Binaural recording – stereo reproduction



Stereo recording – headphone reproduction



Audio developments based on psychoacoustics

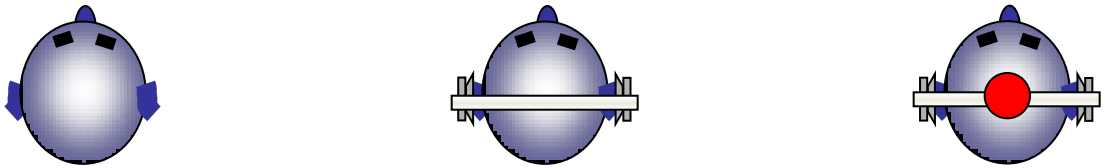
- Data reduction
- Headphone reproduction
- Binaural technologies
- Sound field synthesis (WFS / HOA)
- Stereophony
- 5.1 / 7.1 Surround (ITU BS 775)
- 3D Audio

Headphone diffuse-field equalization

Headphone diffuse field equalization

Loudness comparison (free field equalization)

Auditory event

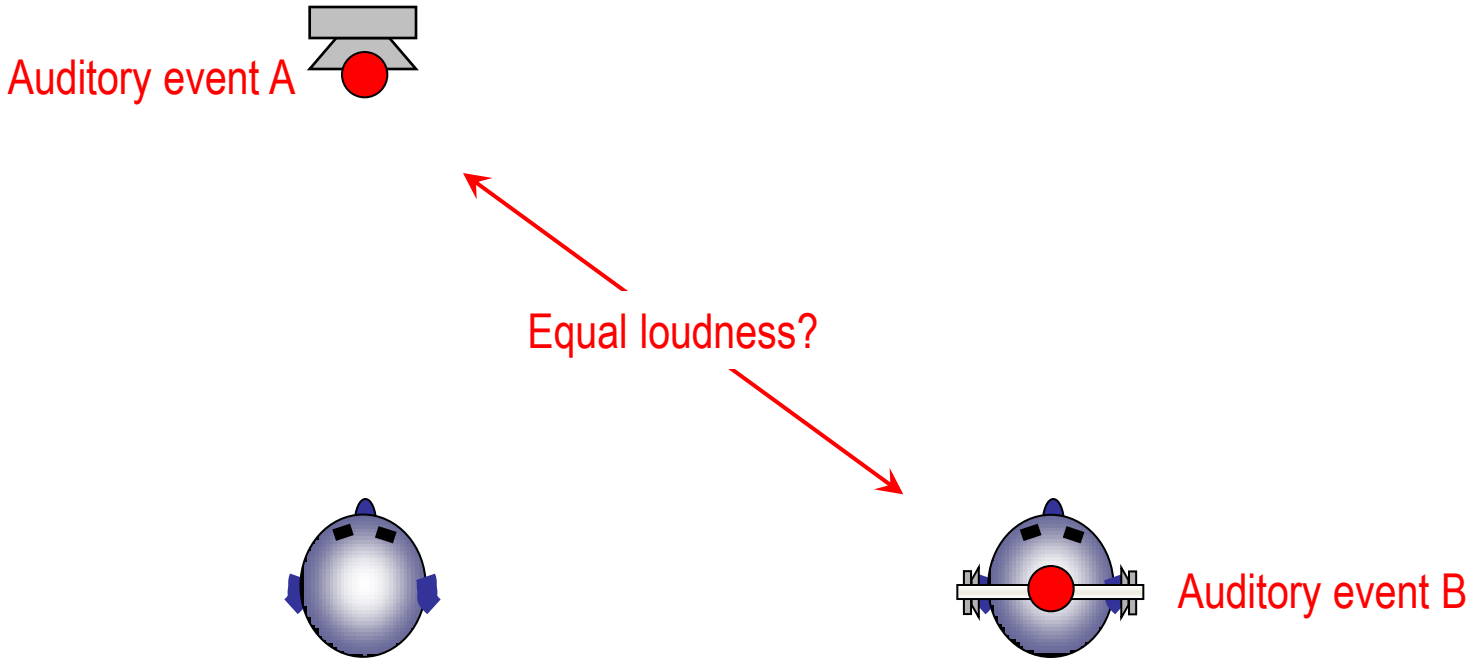


Reference

Required

However!

Loudness comparison (free field equalization)



Headphone diffuse-field equalization

Figures of equal size

Headphone diffuse-field equalization

Figures of equal size

Headphone diffuse-field equalization

Figures of equal size

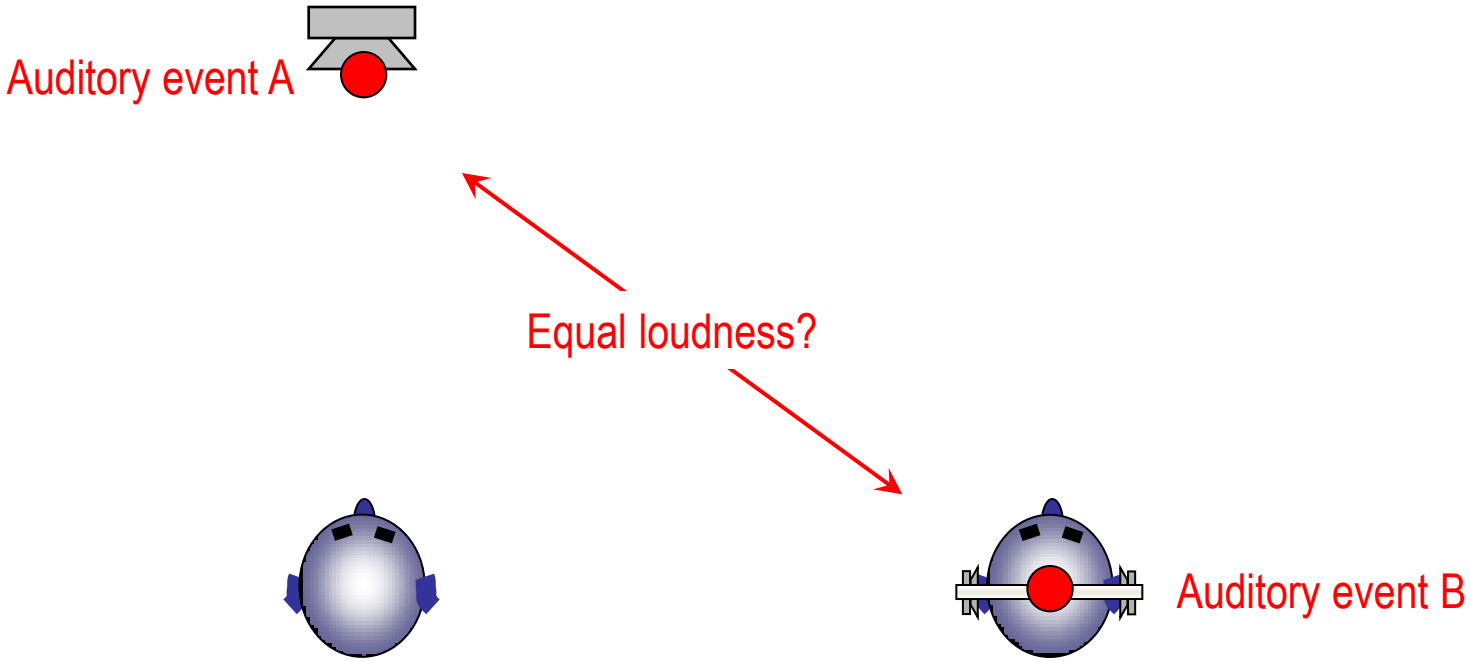
Headphone diffuse-field equalization

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

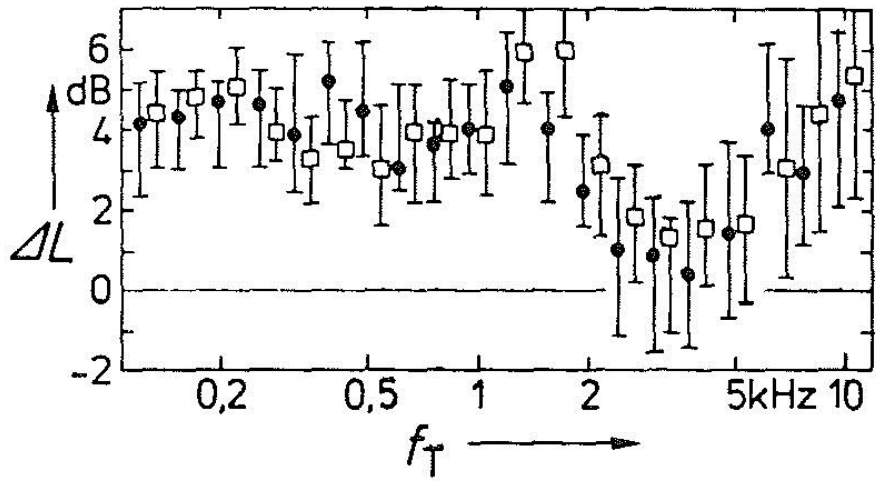


Fig.1: Lautheitsvergleich von Sinustönen, die im Reflexionsarmen Raum aus Kopfhörern bzw. Lautsprechern dargeboten werden. Pegeldifferenz $\Delta L = L_{GKH} - L_{GLS}$ der bei gleicher Lautheit im Gehörgang gemessenen Schallpegel für Kopfhörerbeschallung (L_{GKH}) bzw. Lautsprecherbeschallung (L_{GLS})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

Headphone diffuse-field equalization

Conclusions from the Association model (*Theile 1986*):

- The ear interprets the received information associated with the outside world and deduces the source property on this basis
- The loudness of an auditory event is not clearly defined by the sound pressure in the ear canal
- The free-field adjustment of the headphone is basically wrong equalization with respect to the reproduction of stereophonic signals

Headphone diffuse-field equalization

Solution: Diffuse-field equalization (*Theile 1986*)

- Directional neutral equalization - Integral over all free-field transfer functions (HRTFs)
- Physical probe transfer function measurement

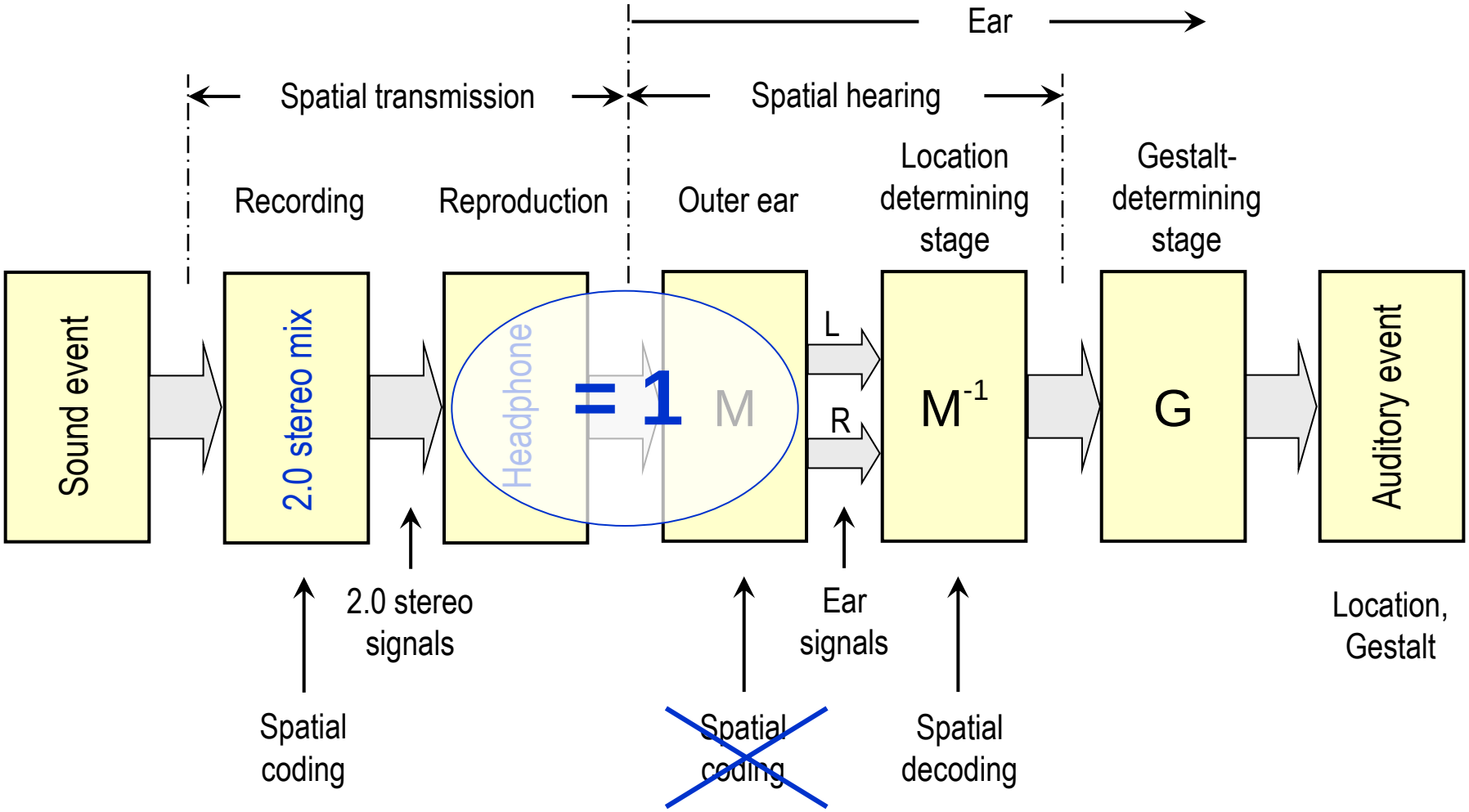
Headphone diffuse-field equalization

Recommendation ITU-R BS.708 (1990)

„Determination of the electro-acoustical properties of studio monitor headphones“

- Directional neutral equalization - Integral over all free-field transfer functions (HRTFs)
- Physical probe transfer function measurement

Stereophony: Headphone reproduction

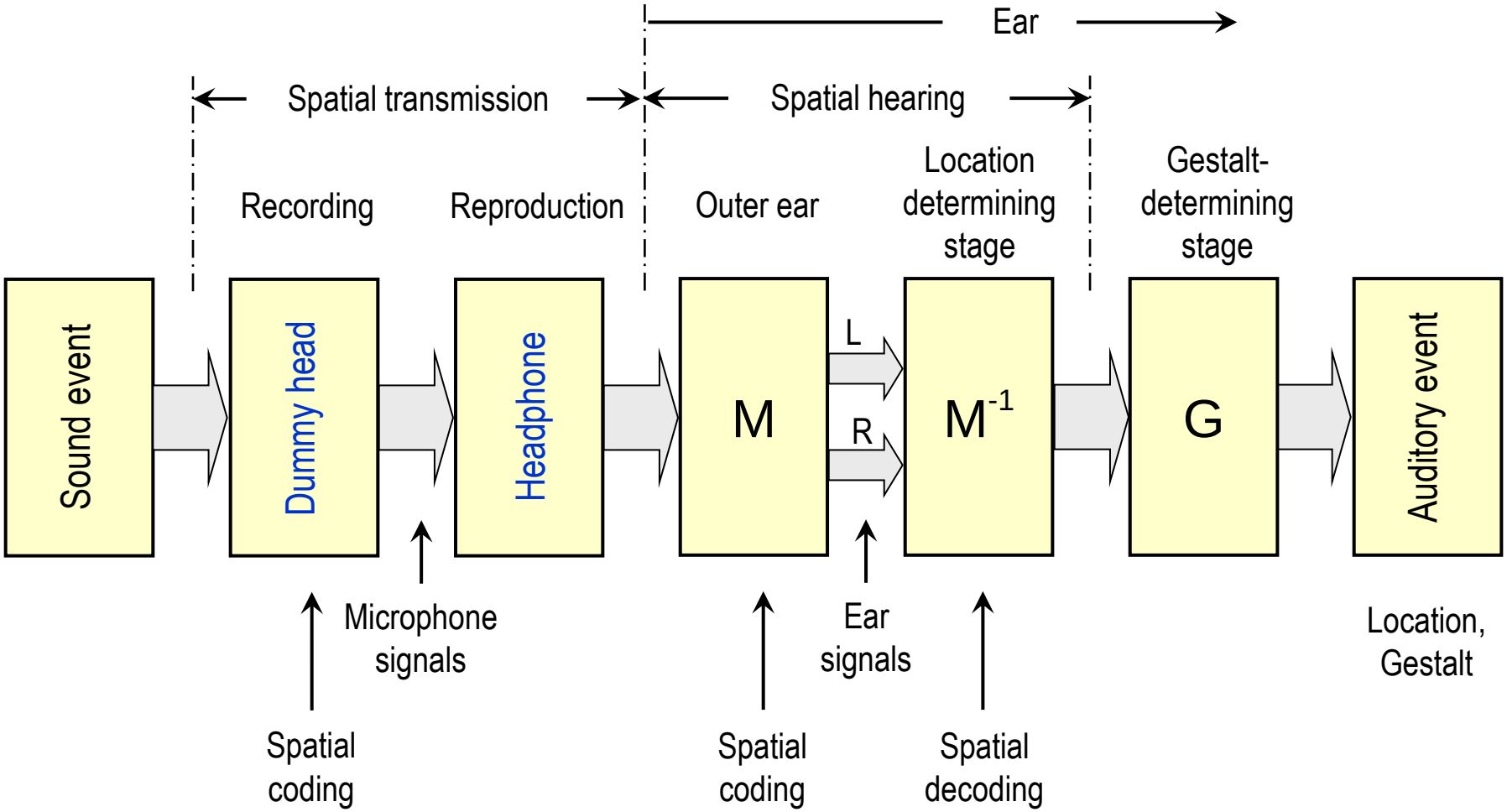


Audio developments based on psychoacoustics

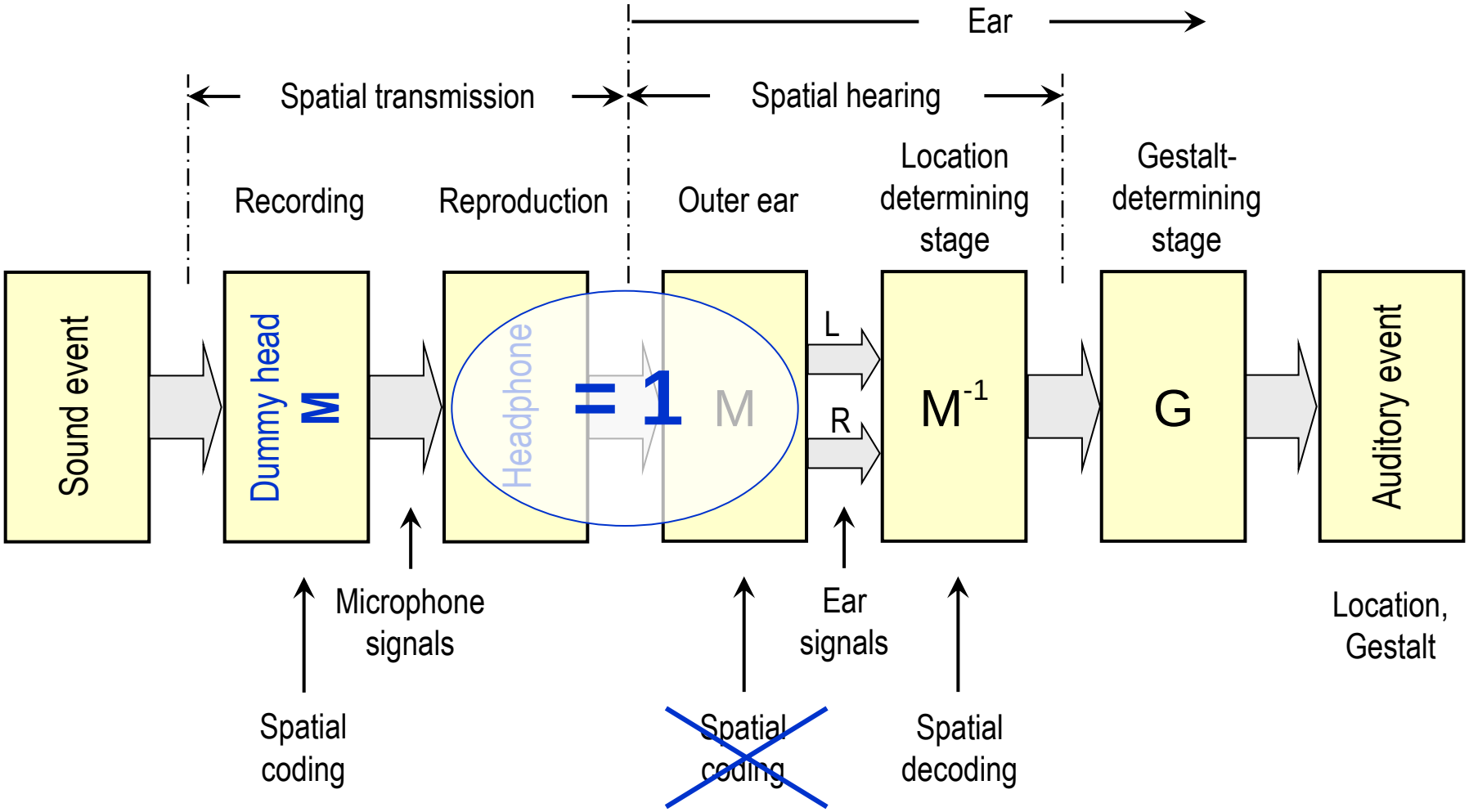
- Data reduction
- Headphone reproduction
- Binaural technologies
- Sound field synthesis (WFS / HOA)
- Stereophony
- 5.1 / 7.1 Surround (ITU BS 775)
- 3D Audio

Verband Deutscher Tonmeister e.V.

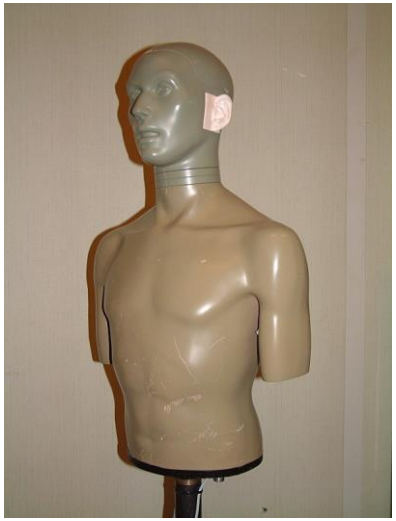
Binaural technology („perfect dummy head“)



Binaural technology („perfect dummy head“)



Dummy heads, actual models



HMS III - HEAD acoustics

HUGO - RWTH Aachen

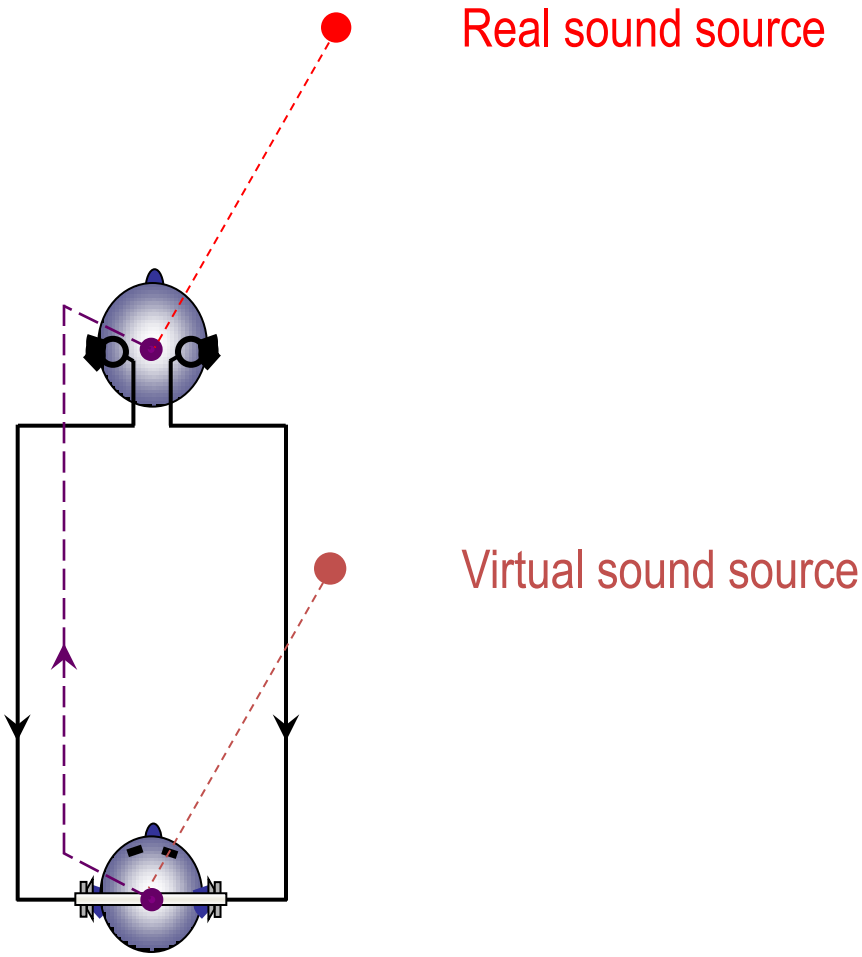
MANIKIN MK1 - Neutrik-Cortex

KEMAR - Knowles

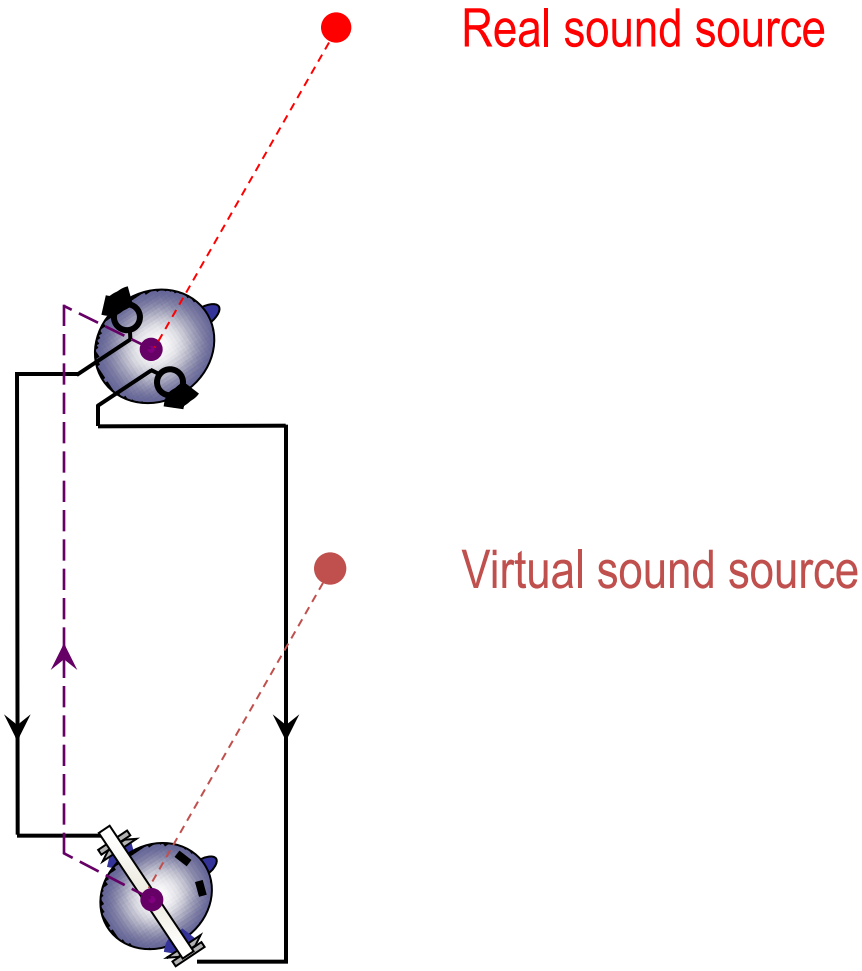
KU 81 - Neumann

KU 100 - Neumann

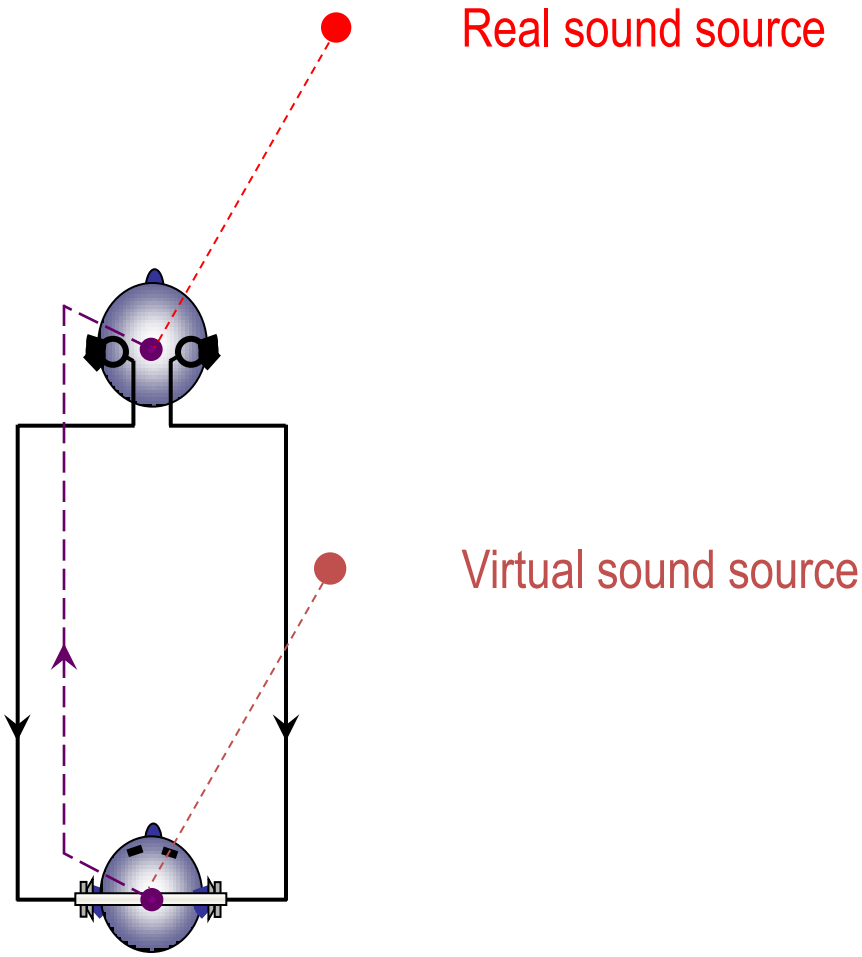
Dummy head (head tracking required!)



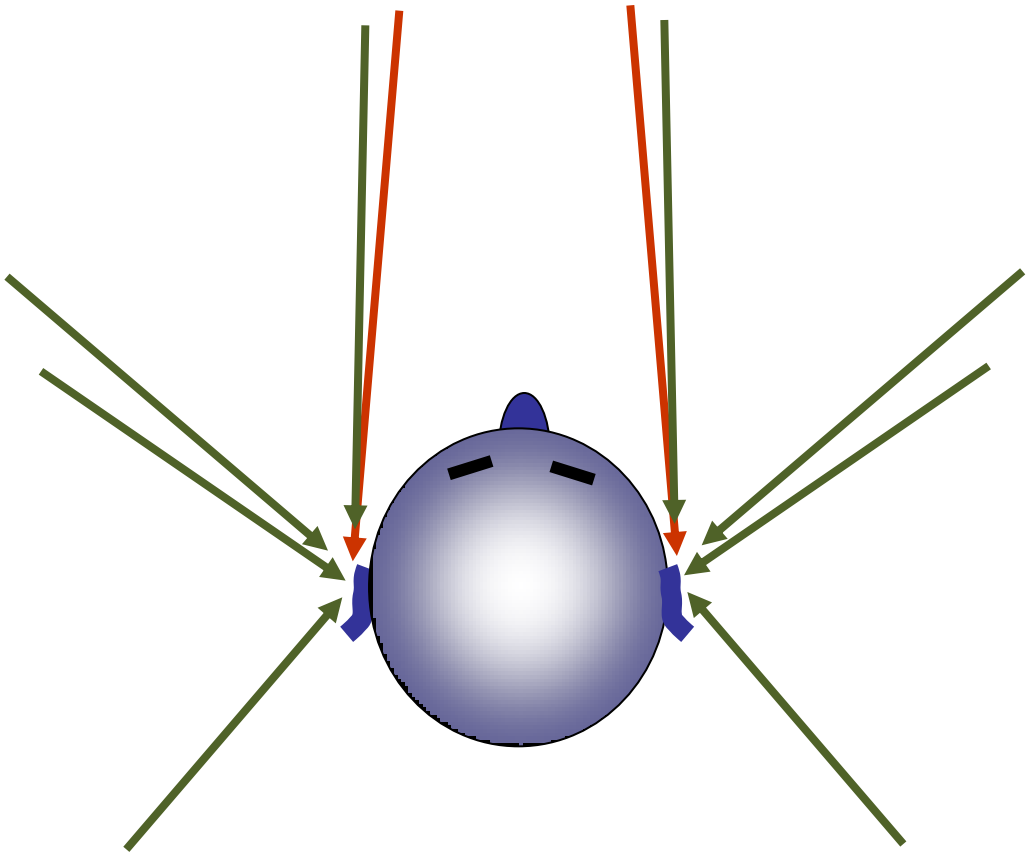
Dummy head (head tracking required!)



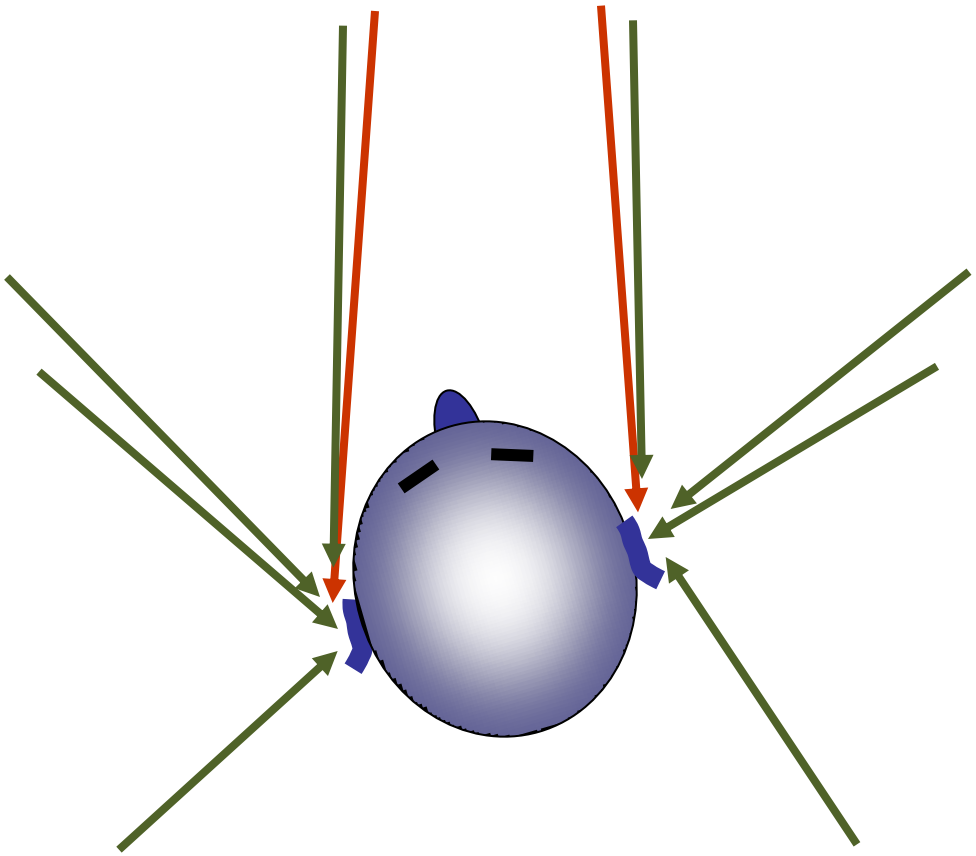
Dummy head (head tracking required!)



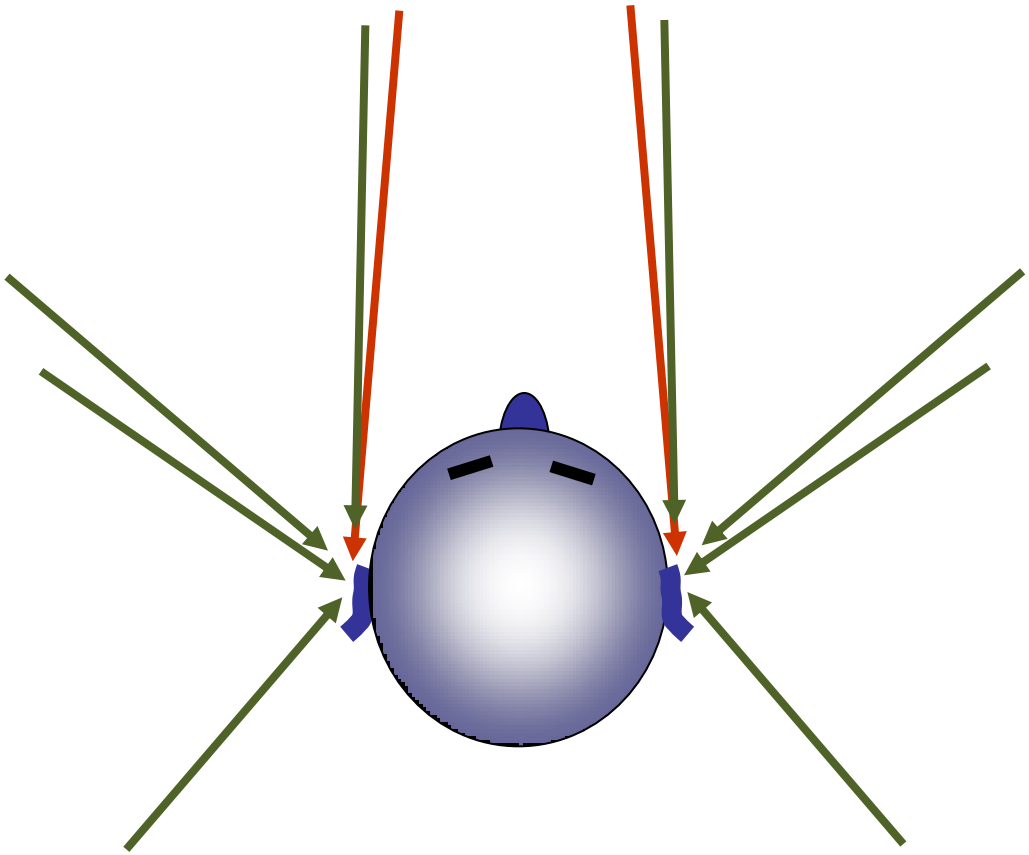
Dummy head (head tracking required!)



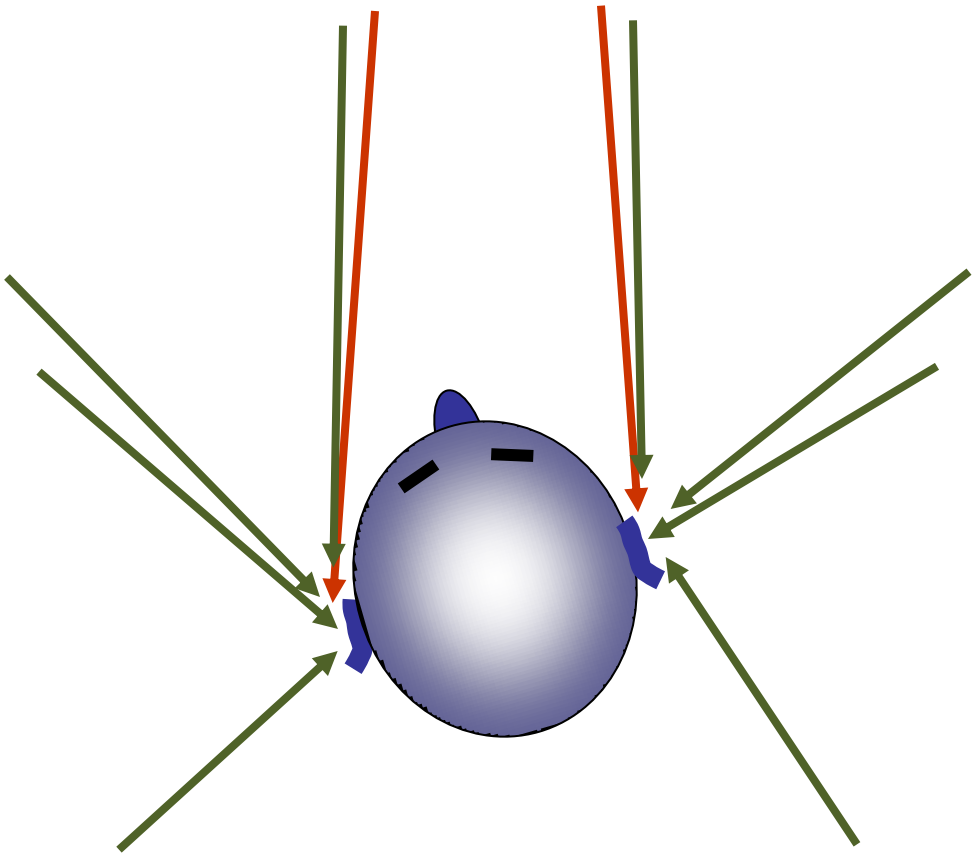
Dummy head (head tracking required!)



Dummy head (head tracking required!)

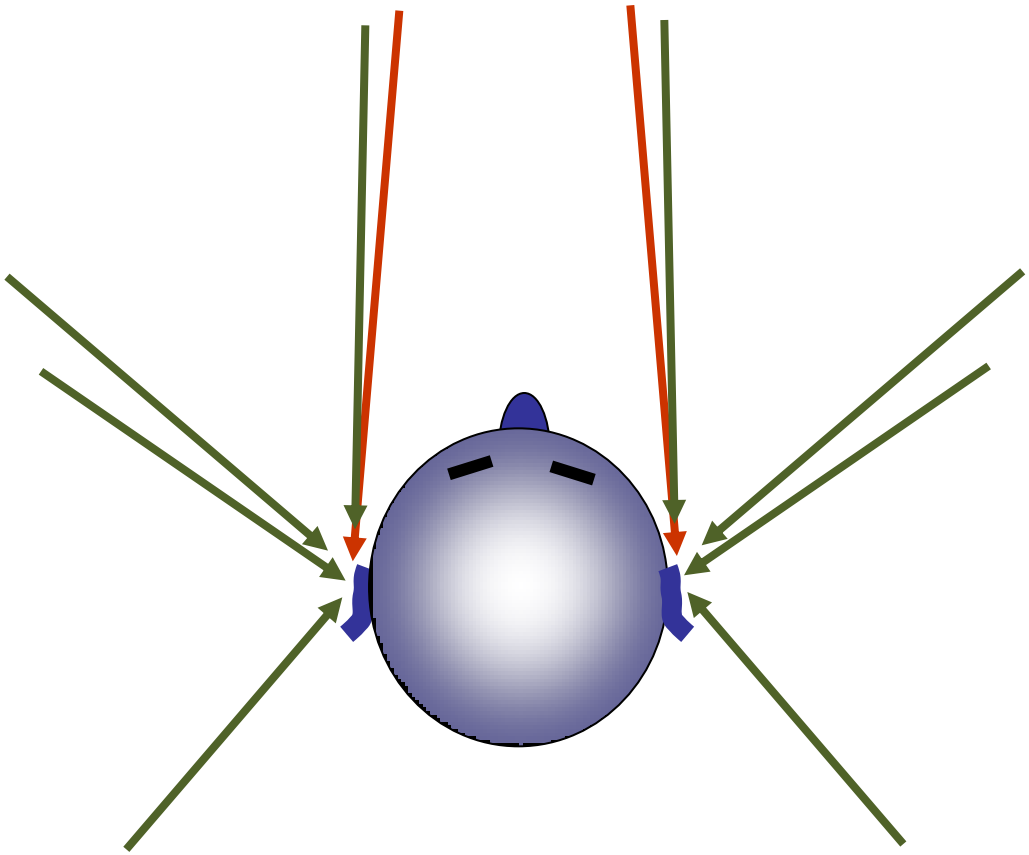


Dummy head (head tracking required!)

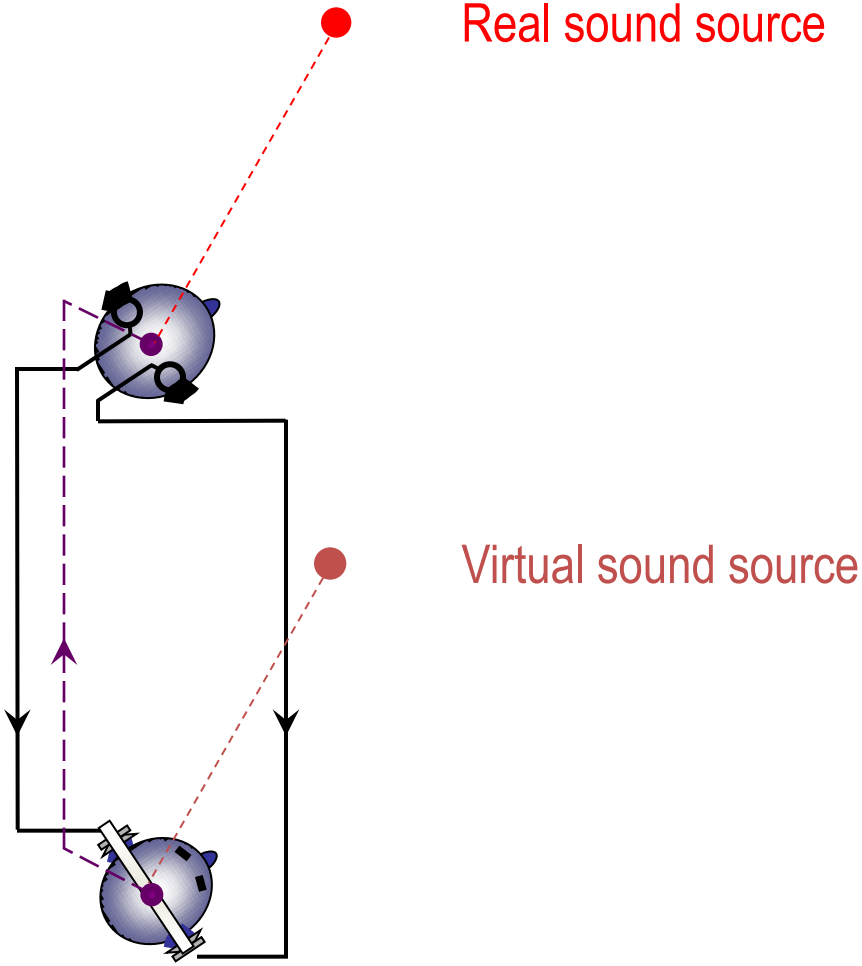


Binaural techniques

Dummy head (head tracking required!)

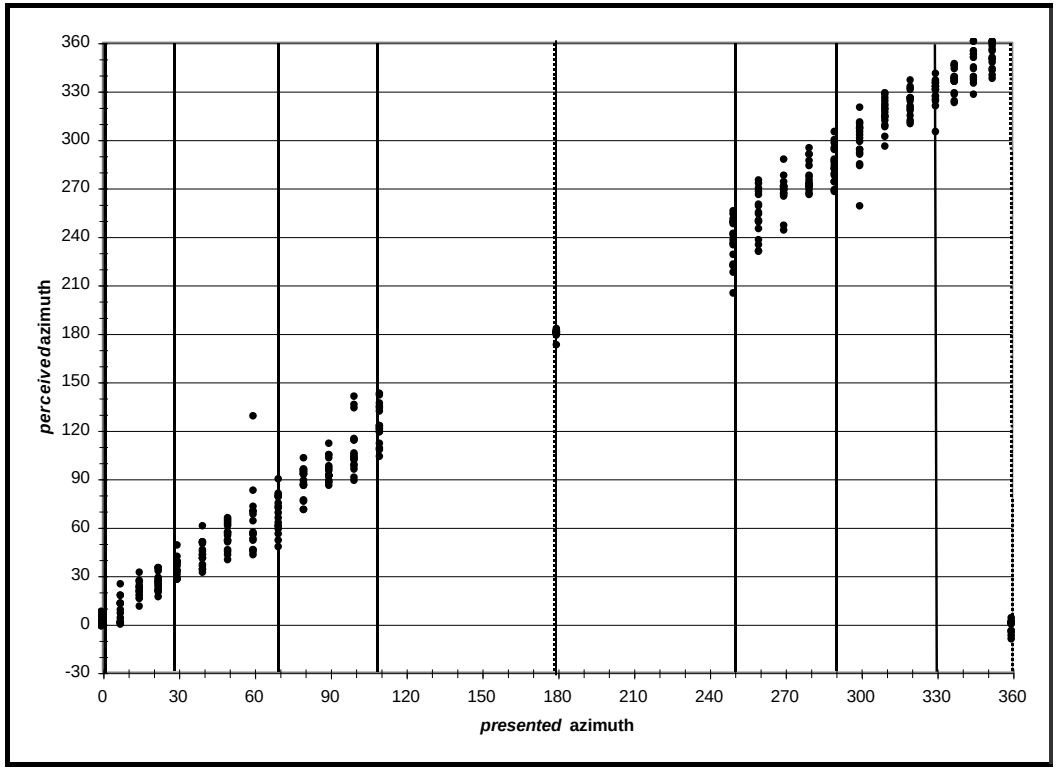


Test: Importance of dynamic cues



Importance of dynamic cues

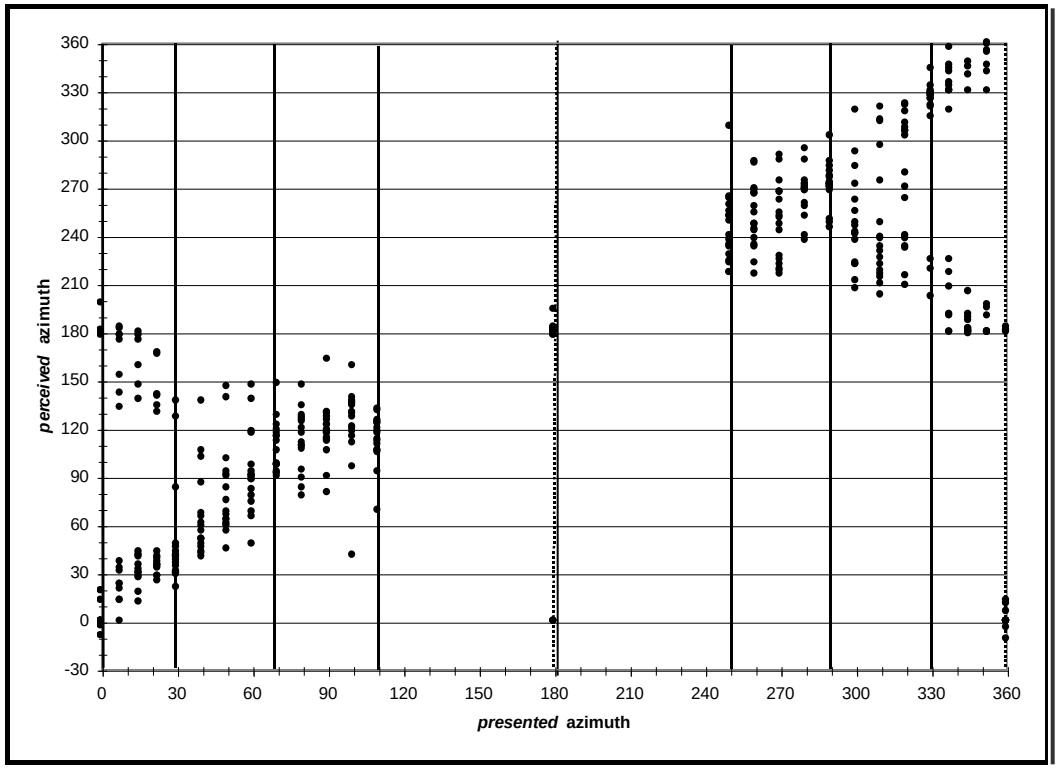
Listening tests horizontal head movements:
Mackensen, Reichenauer, Theile (TMT 1998)



Importance of dynamic cues

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

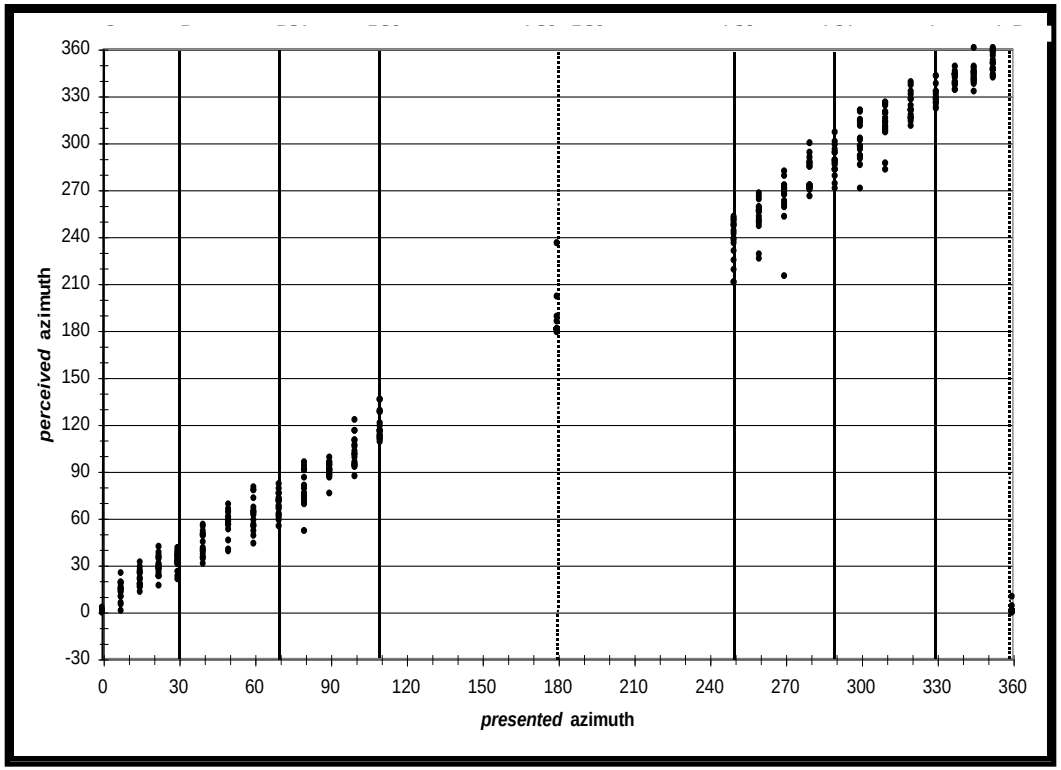
Dummy-head
without
head-tracking



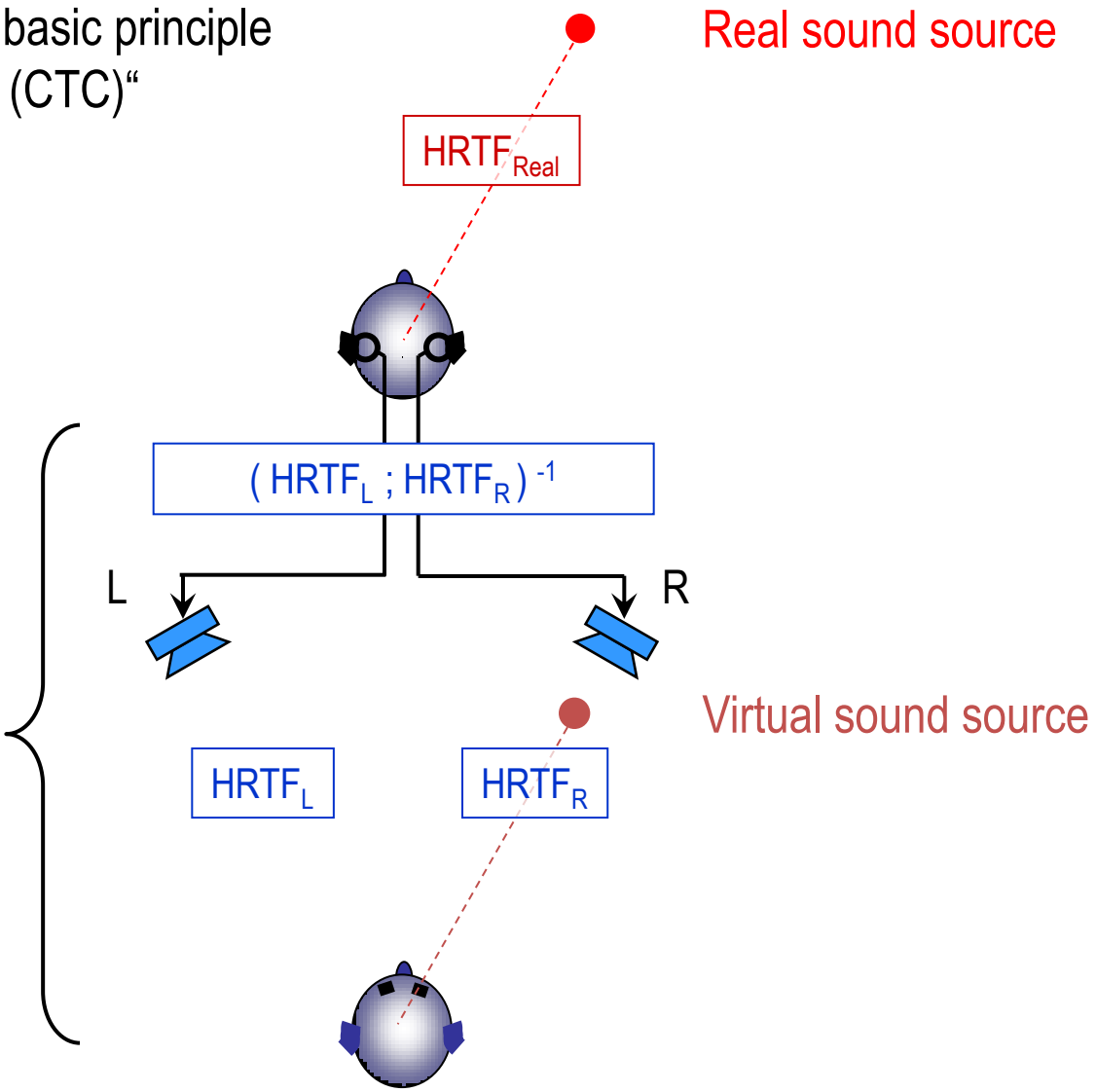
Importance of dynamic cues

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

Natural listening (studio)



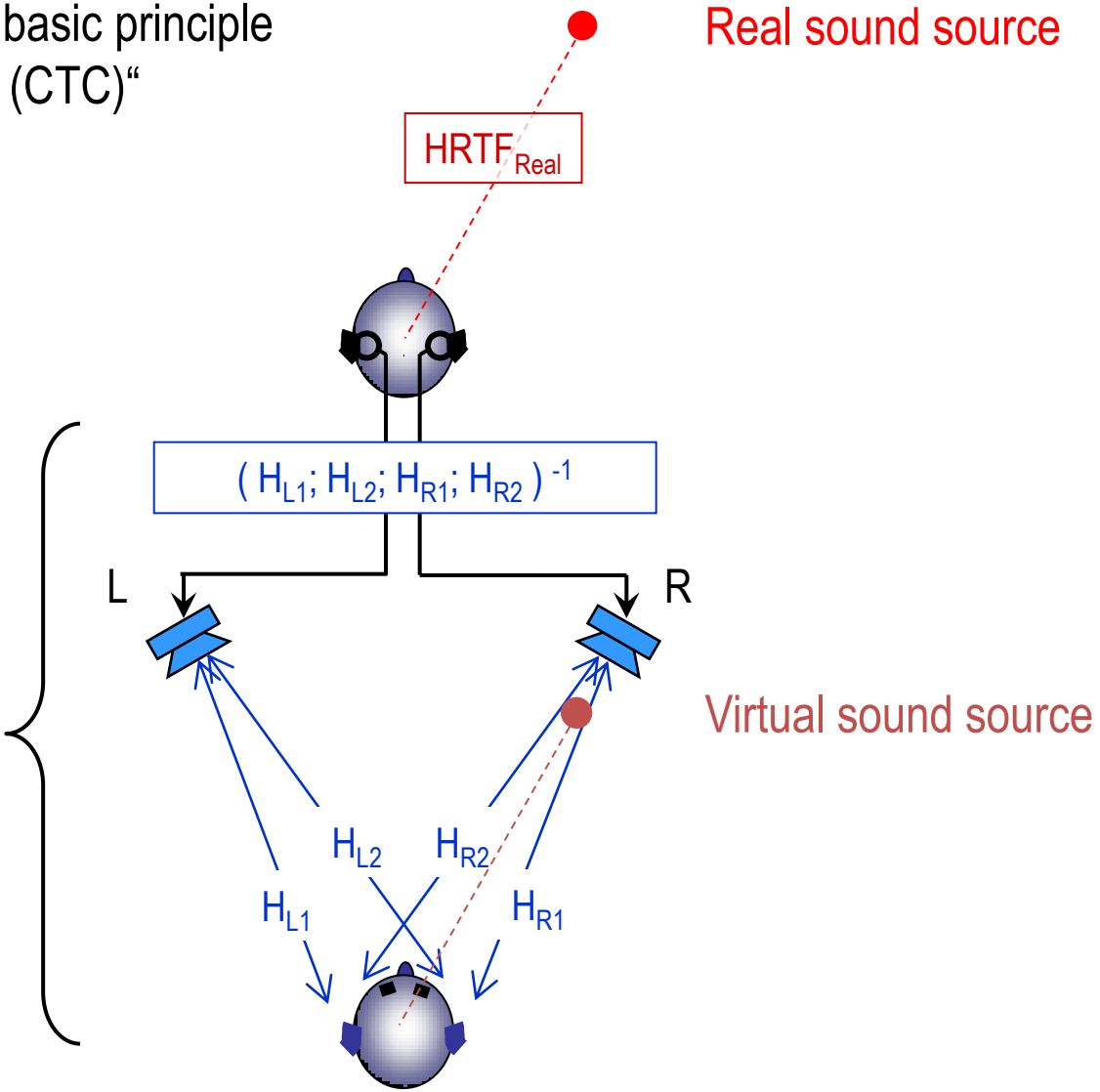
HRTF inverse filtering, basic principle
„Crosstalk Cancelation (CTC)“



Transfer function = 1

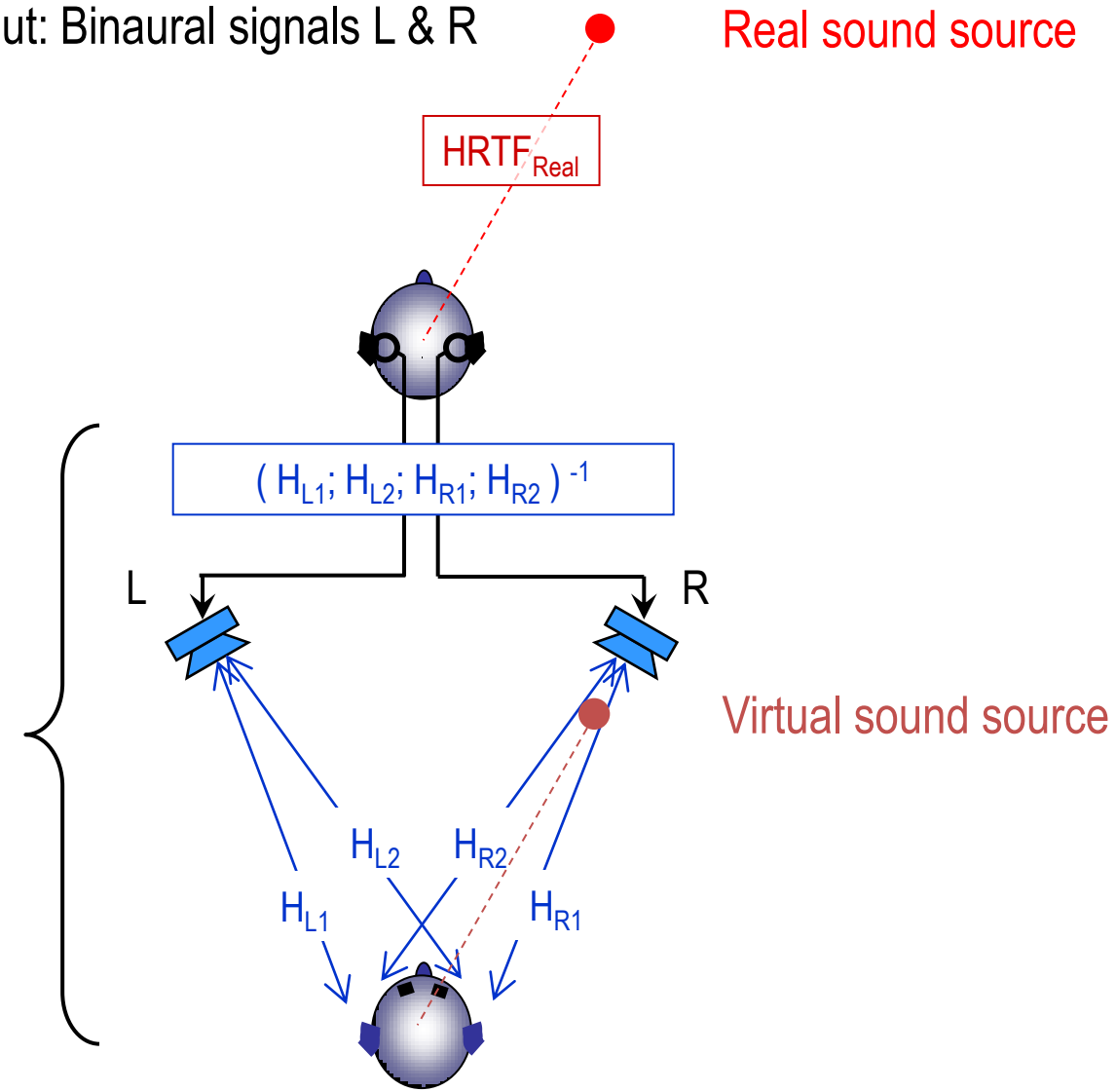
HRTF inverse filtering, basic principle
„Crosstalk Cancelation (CTC)“

Transfer function = 1



Binaural techniques

Virtual headphone input: Binaural signals L & R



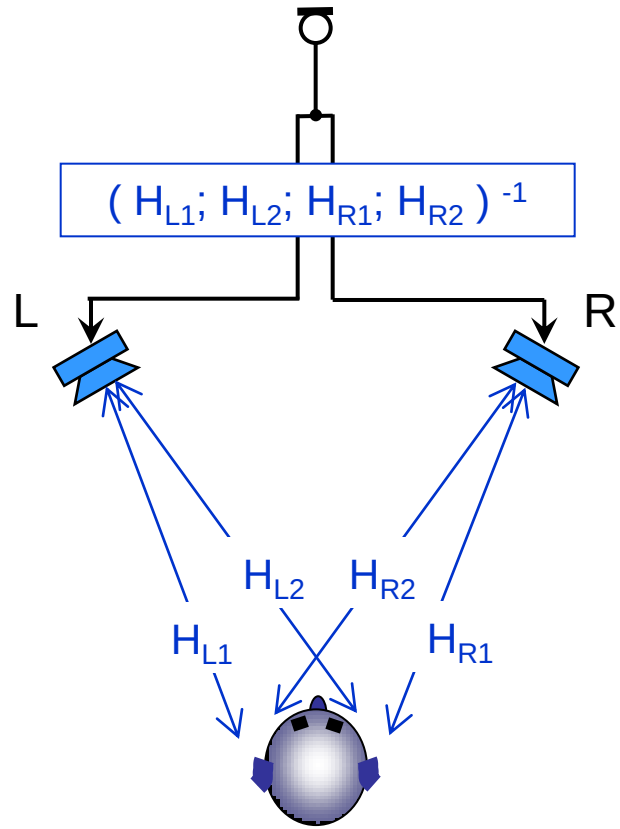
Virtual headphone

Transfer function = 1

Virtual headphone input: Monophonic signals L & R

Virtual headphone

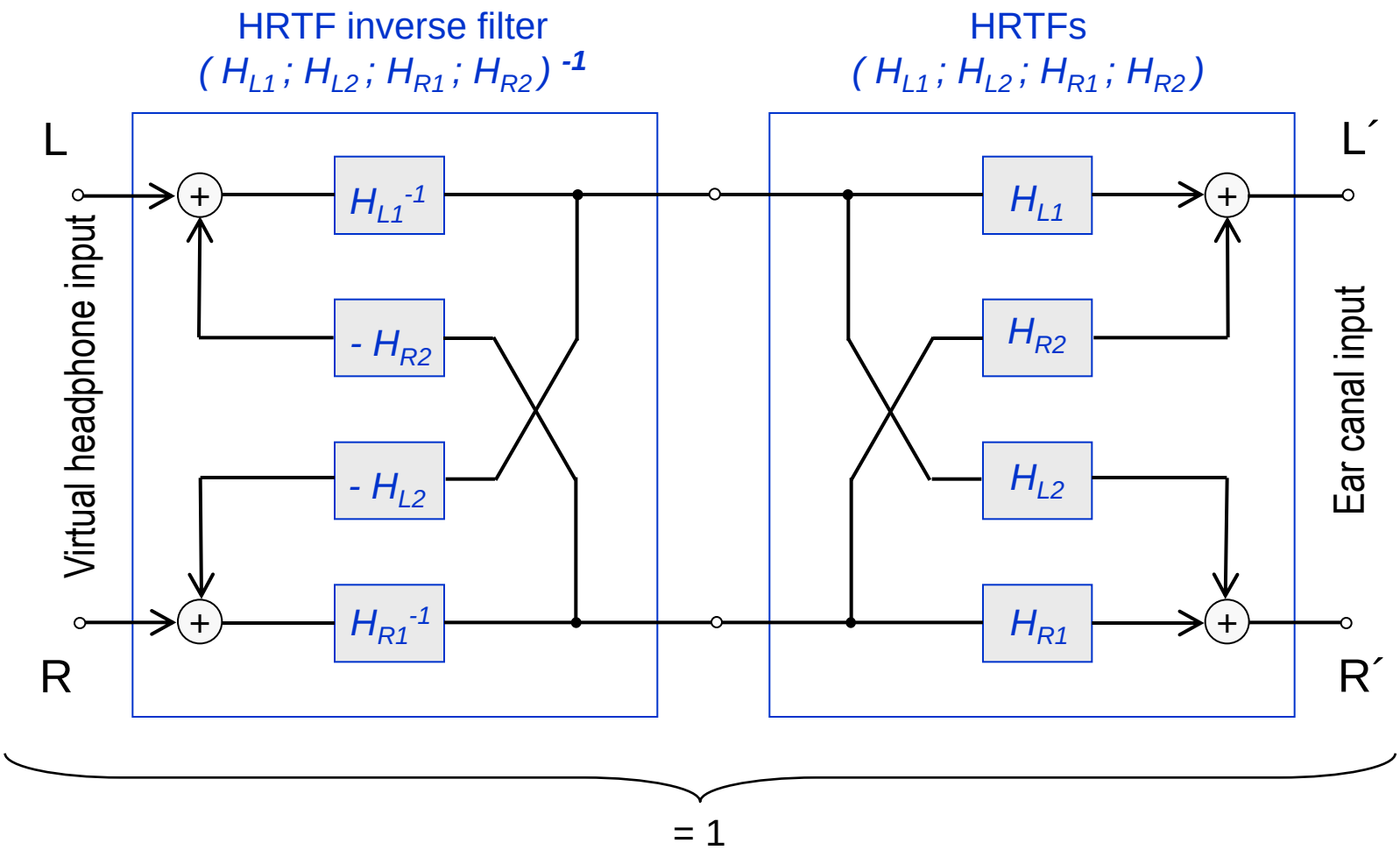
Transfer function = 1



Checking the virtual headphone:

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

HRTF inverse filtering („virtual headphone“)

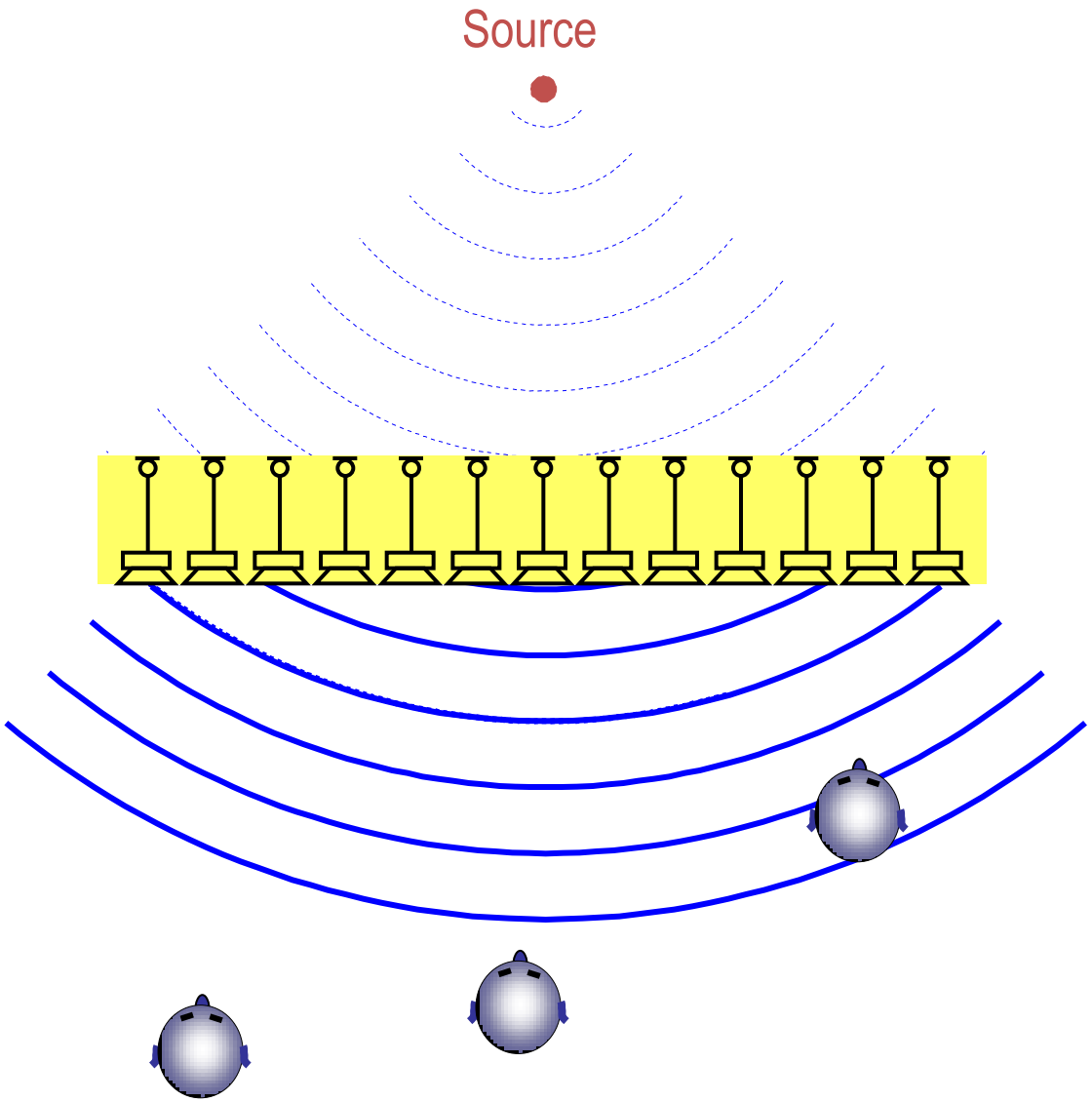


Audio developments based on psychoacoustics

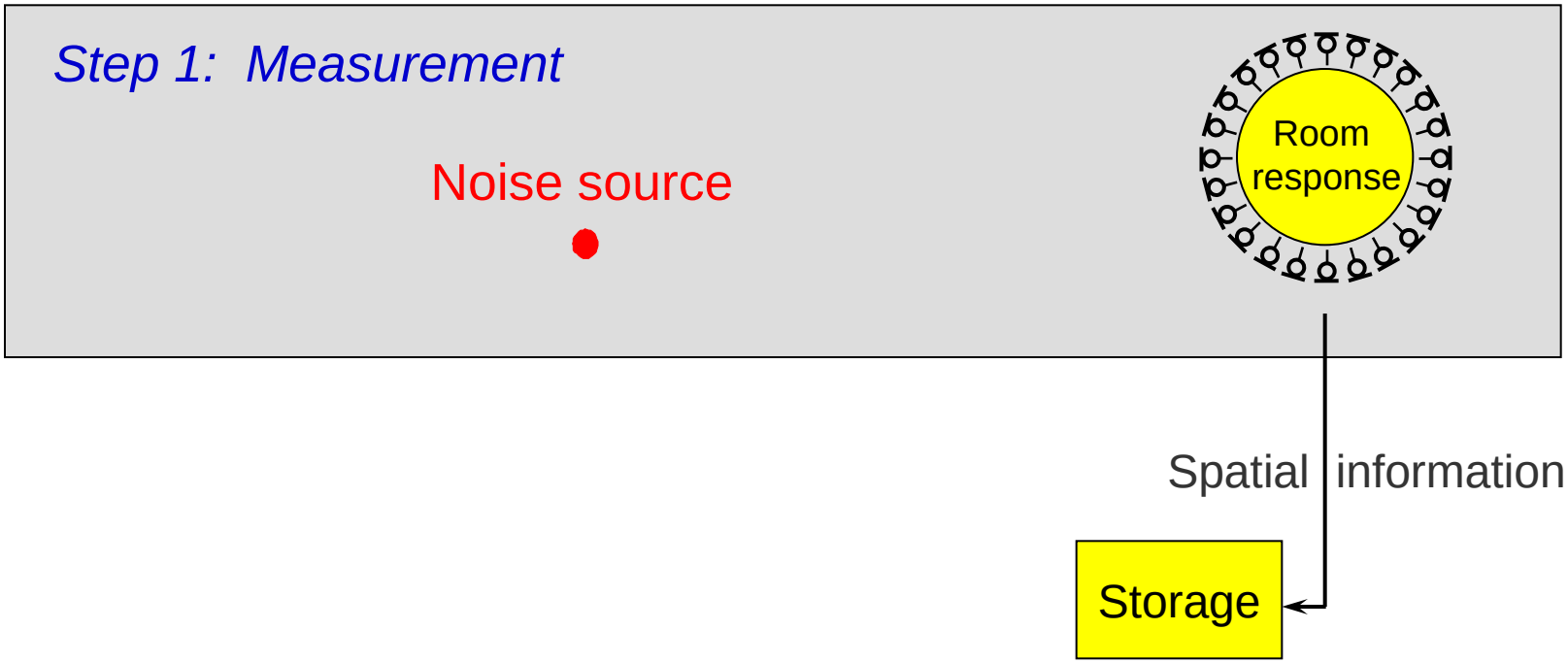
- Data reduction
- Headphone reproduction
- Binaural technologies
- Sound field synthesis (WFS / HOA)
- Stereophony
- 5.1 / 7.1 Surround (ITU BS 775)
- 3D Audio

Virtual sound source, basic principal

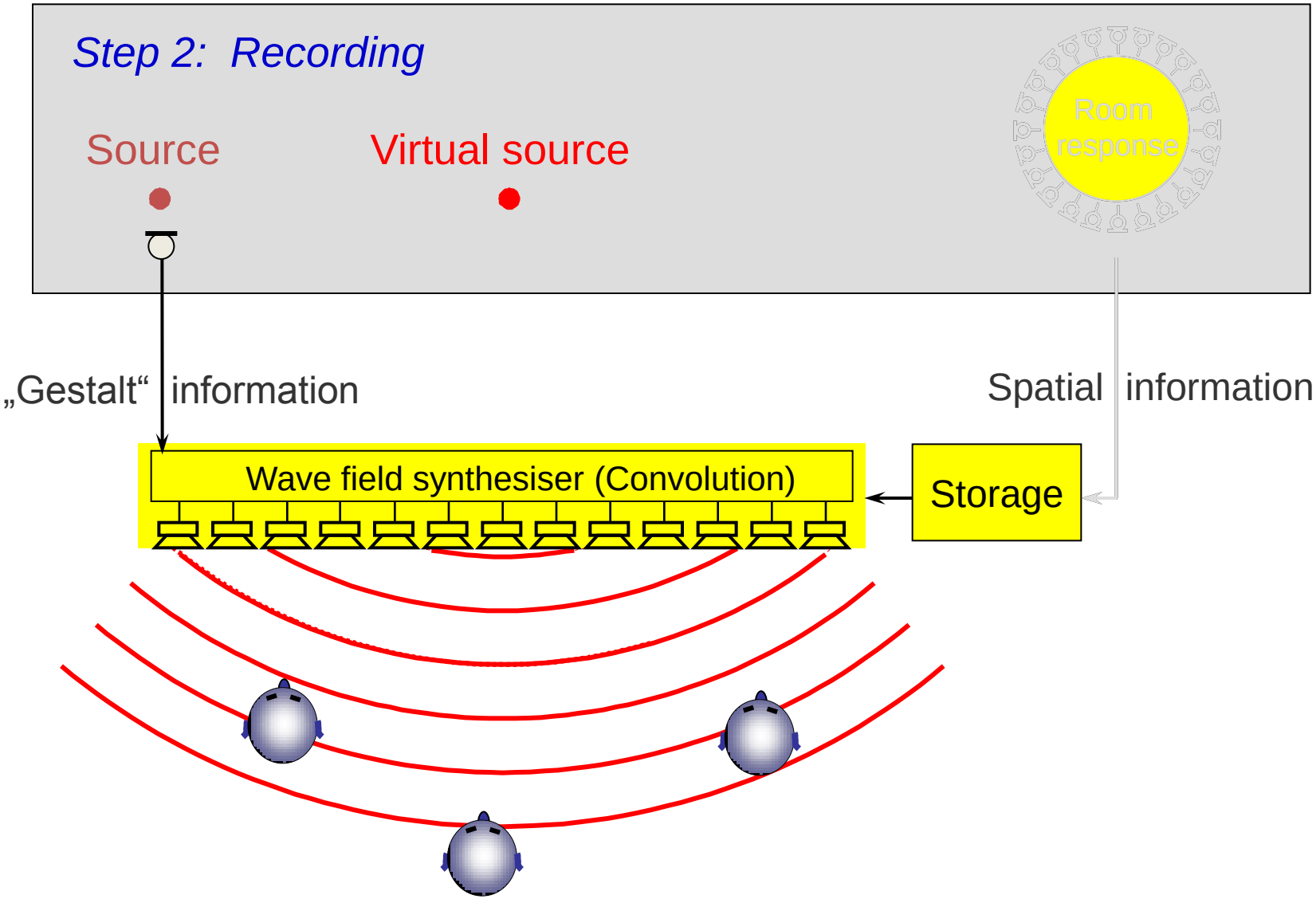
„Acoustic curtain“



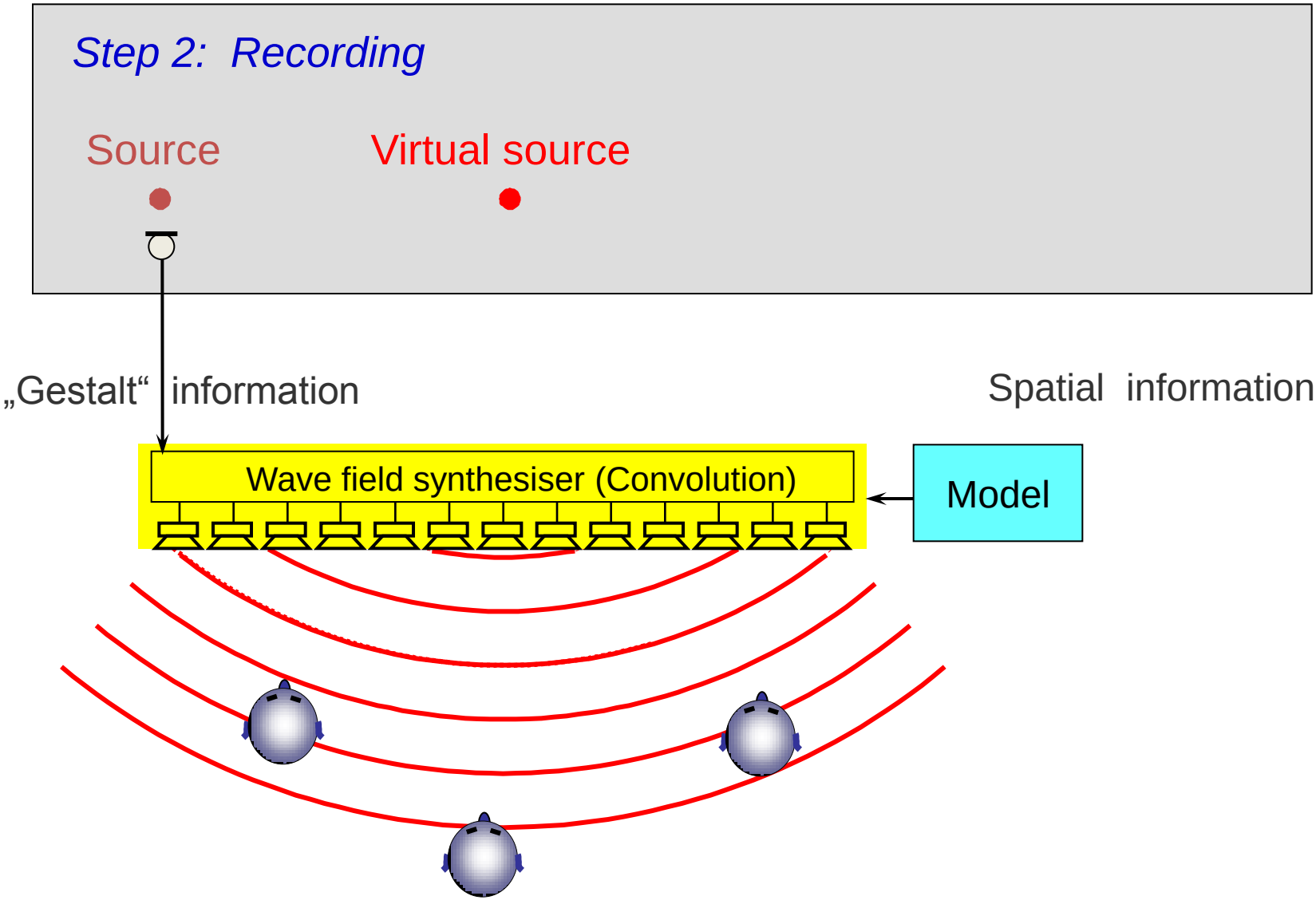
Object-based capturing / rendering (example WFS)



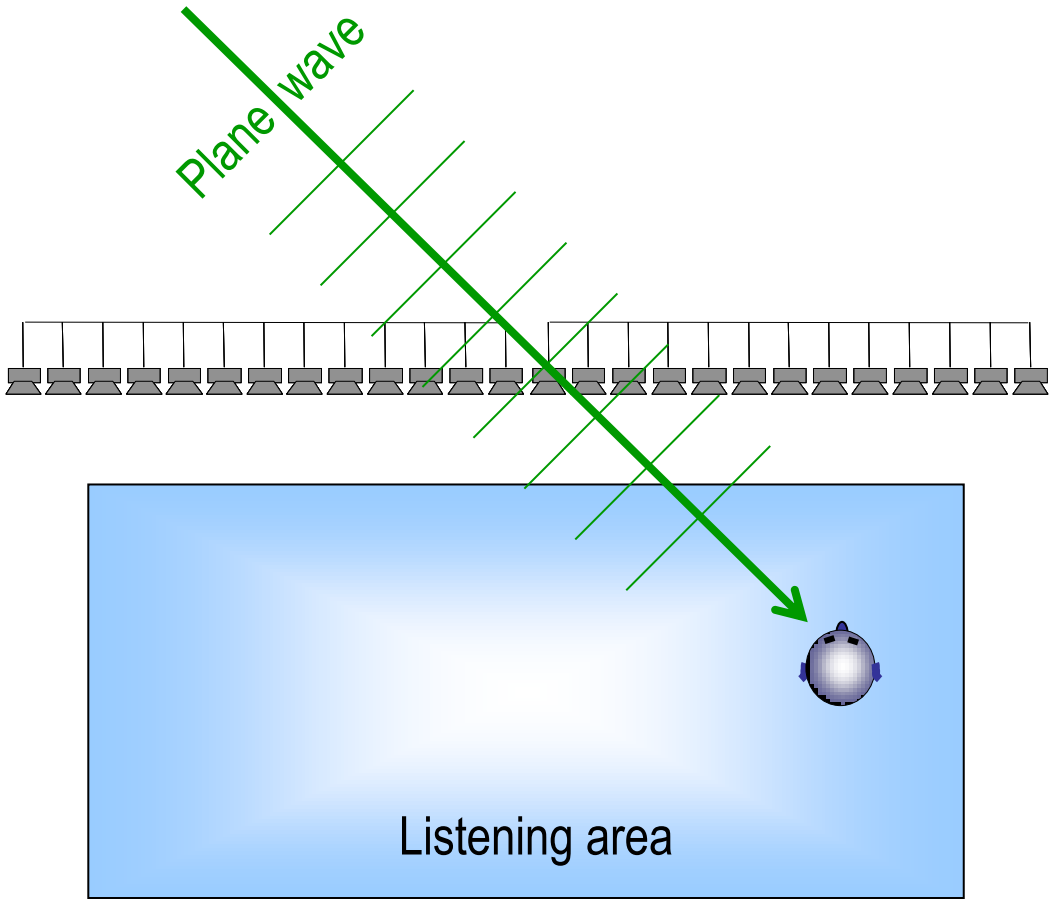
Object-based capturing / rendering (example WFS)



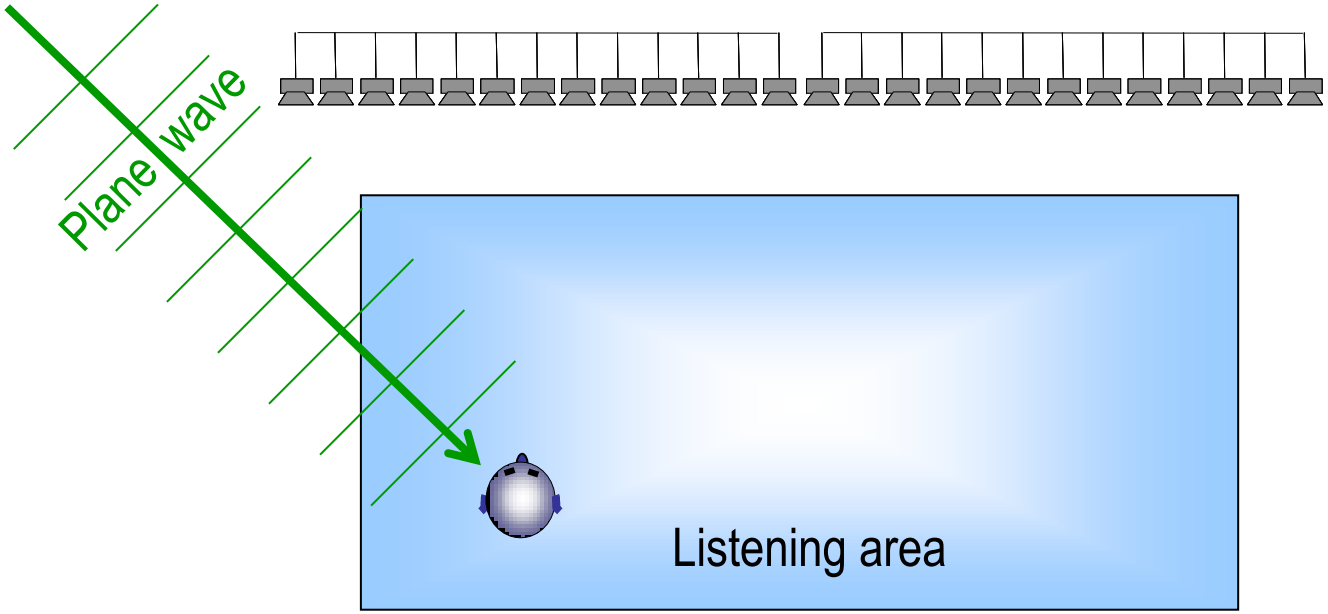
Object-based capturing / rendering (example WFS)



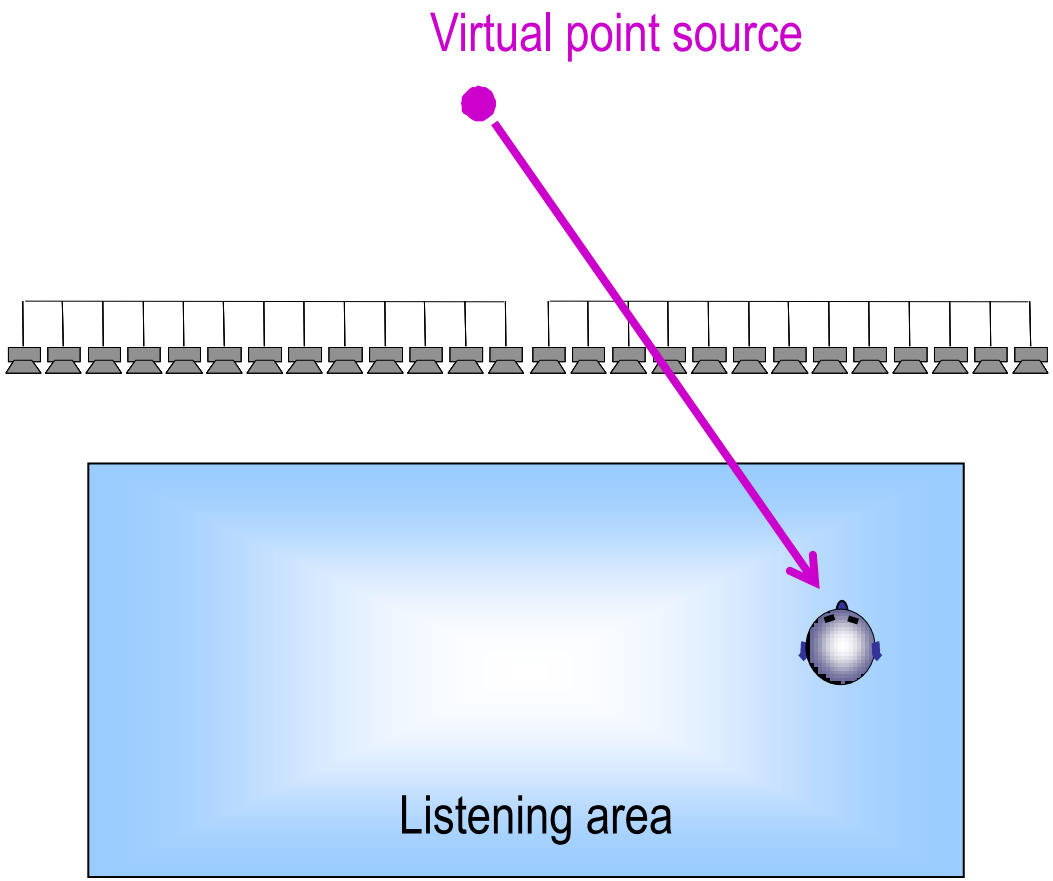
Rendering a perspective acoustic scene



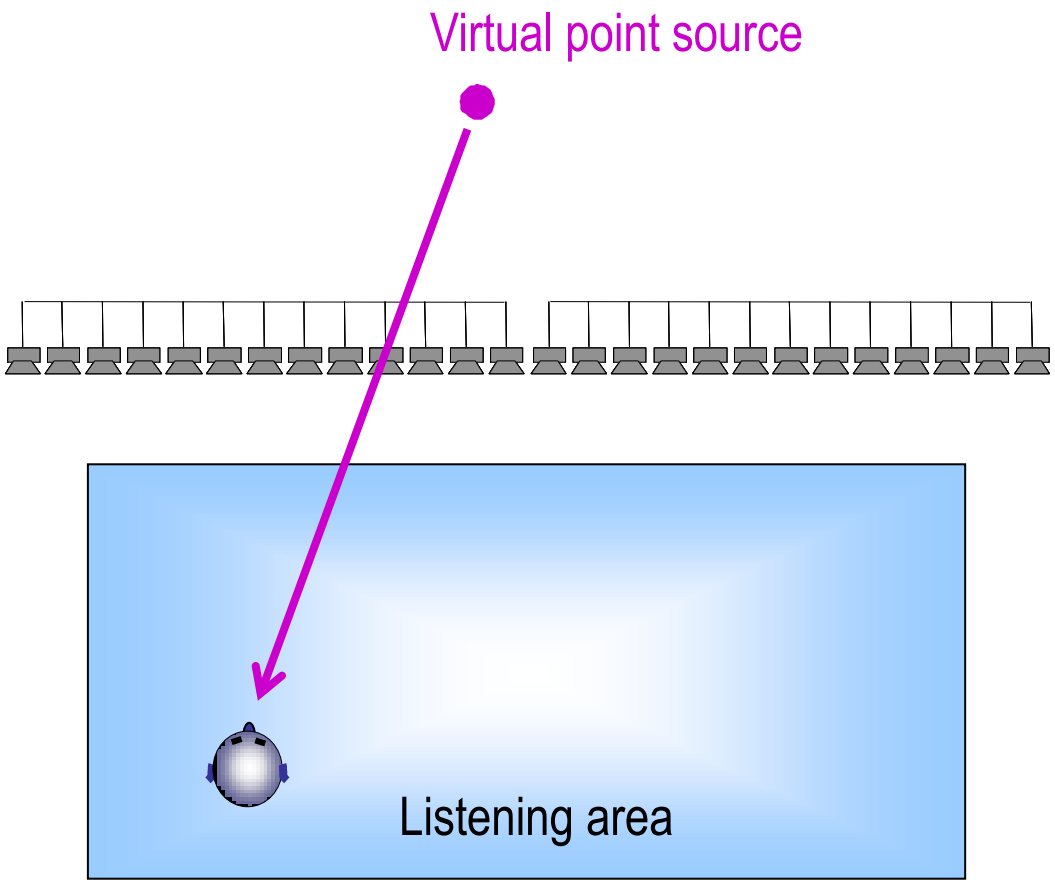
Rendering a perspective acoustic scene



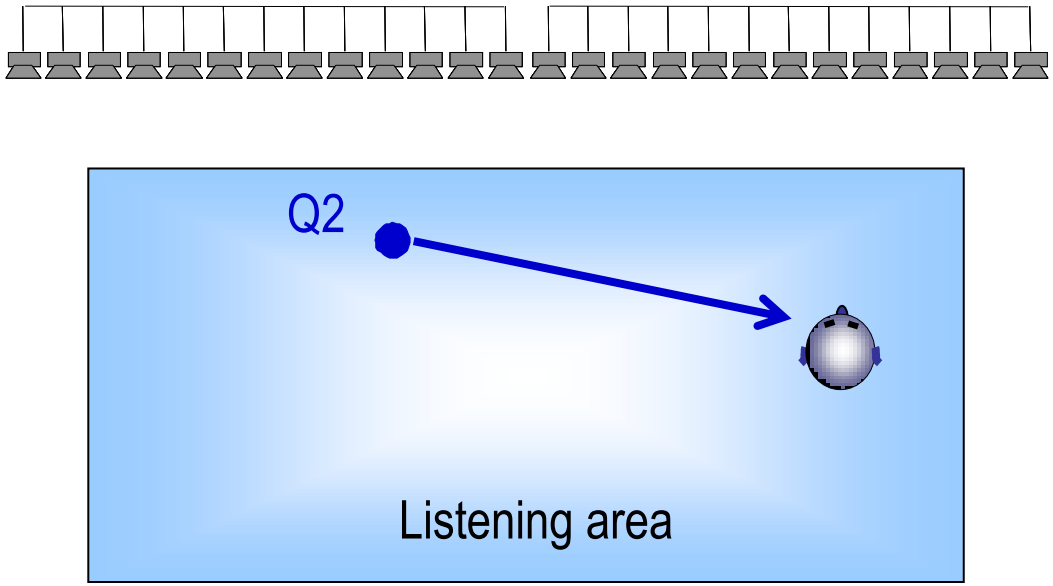
Rendering a perspective acoustic scene



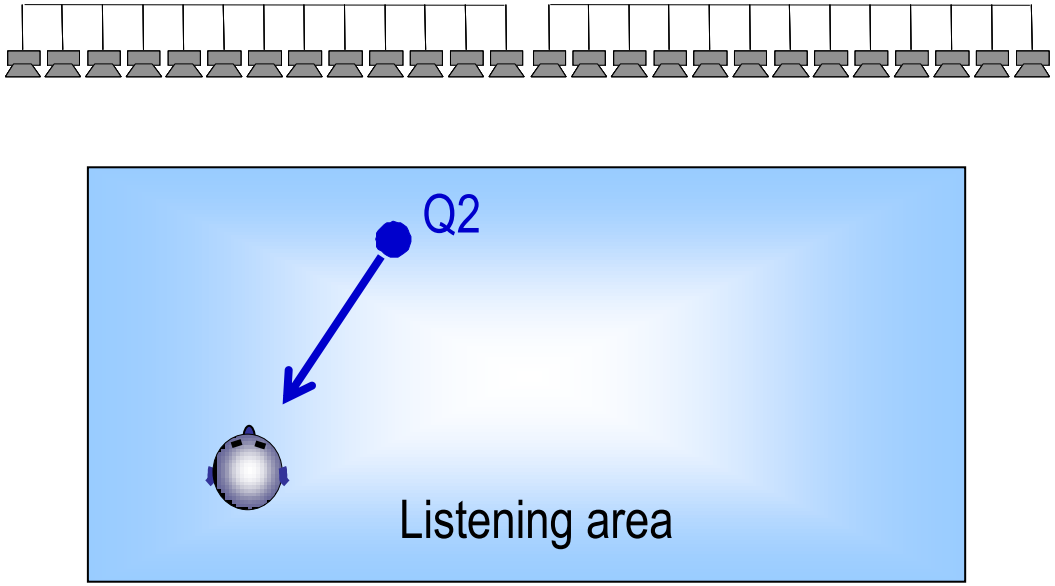
Rendering a perspective acoustic scene



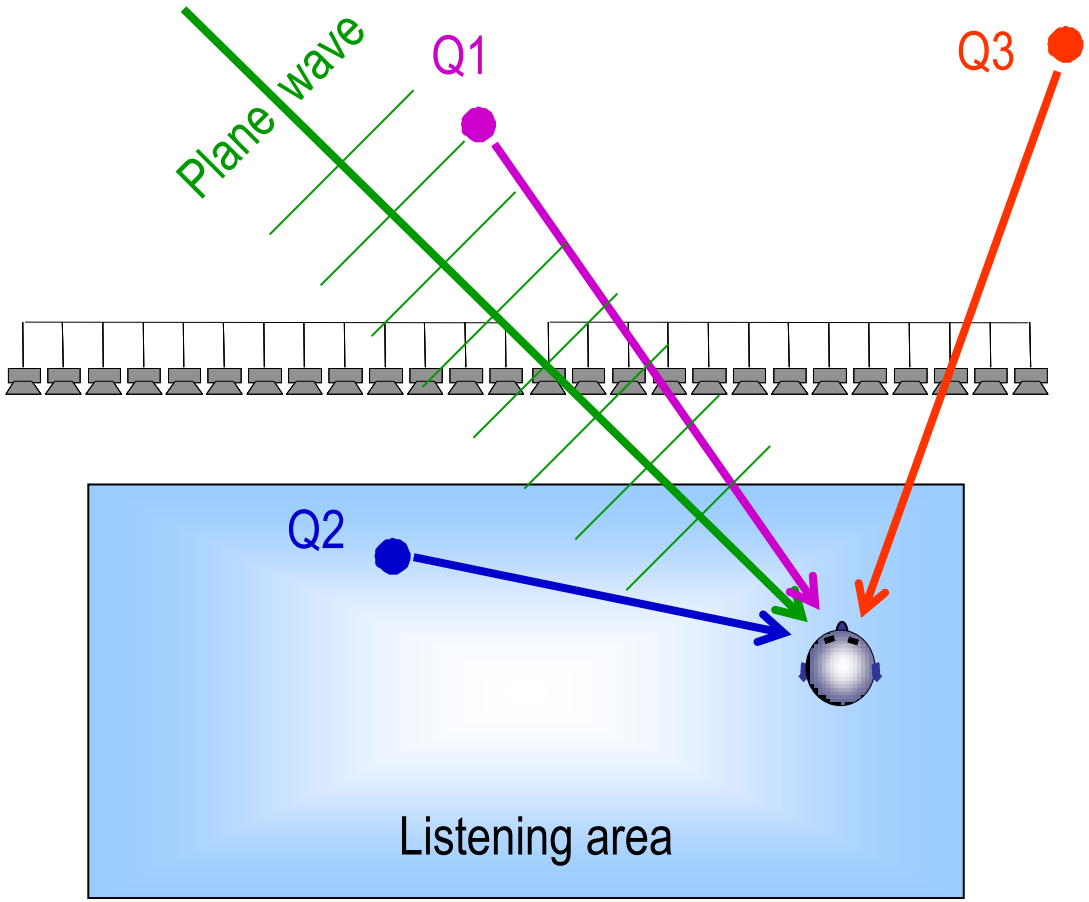
Rendering a perspective acoustic scene



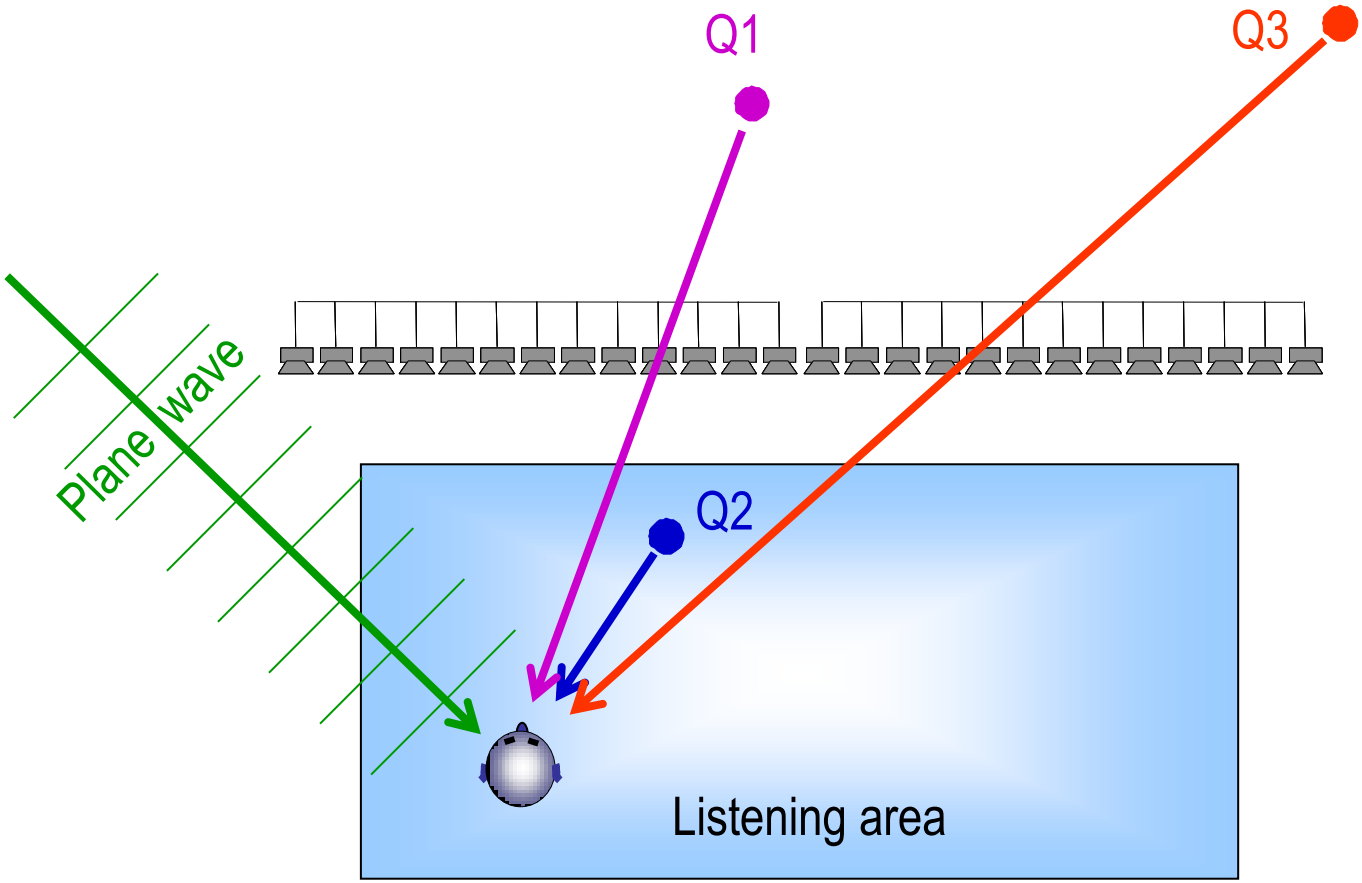
Rendering a perspective acoustic scene



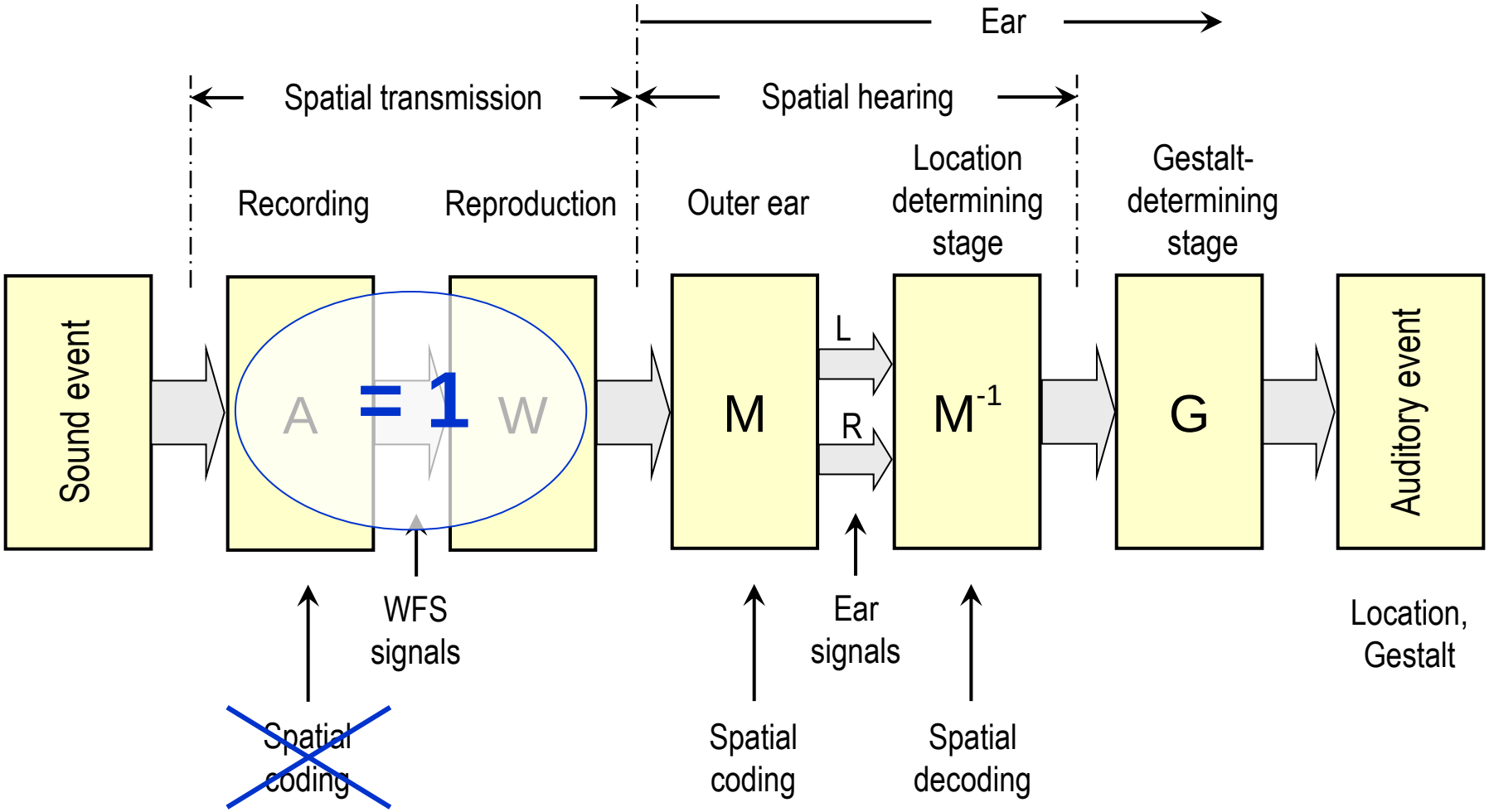
Rendering an intra-active perspective acoustic scene



Rendering an intra-active perspective acoustic scene



Soundfield Synthesis („perfect WFS“)

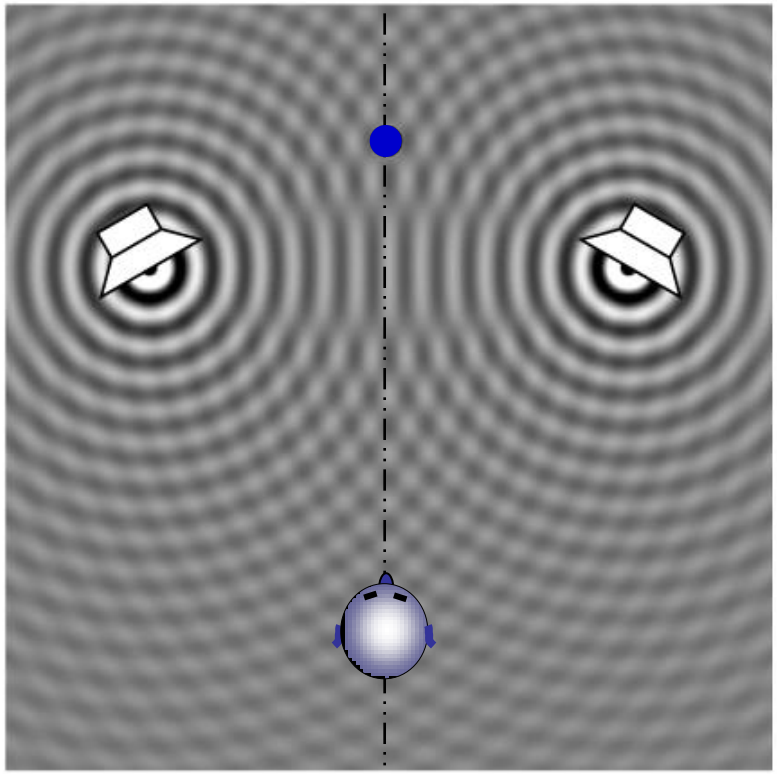


Audio developments based on psychoacoustics

- Data reduction
- Headphone reproduction
- Binaural technologies
- Sound field synthesis (WFS / HOA)
- Stereophony
- 5.1 / 7.1 Surround (ITU BS 775)
- 3D Audio

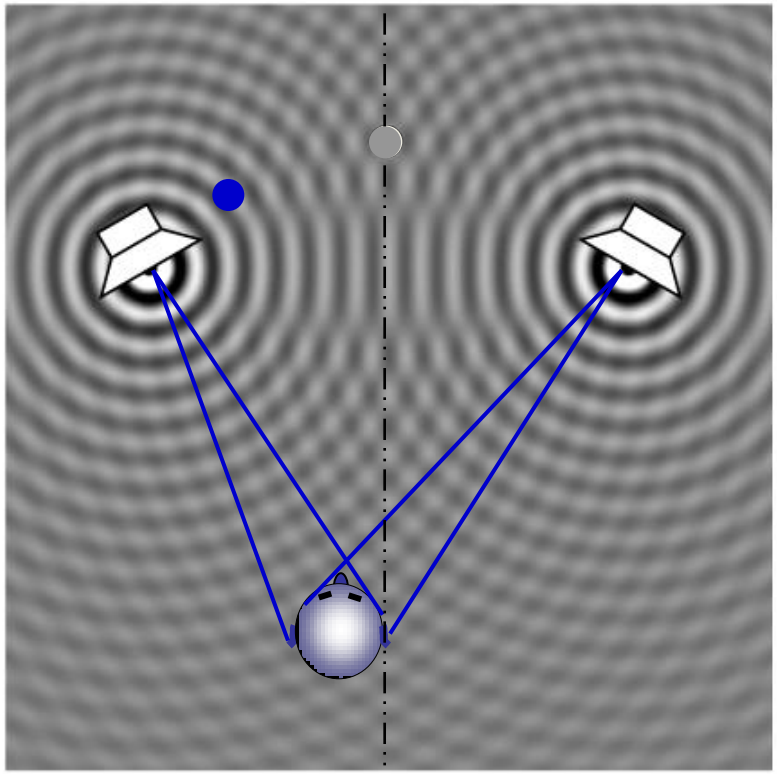
Two-channel stereo, interchannel time difference Δ t = 0 ms

Phantom sound source



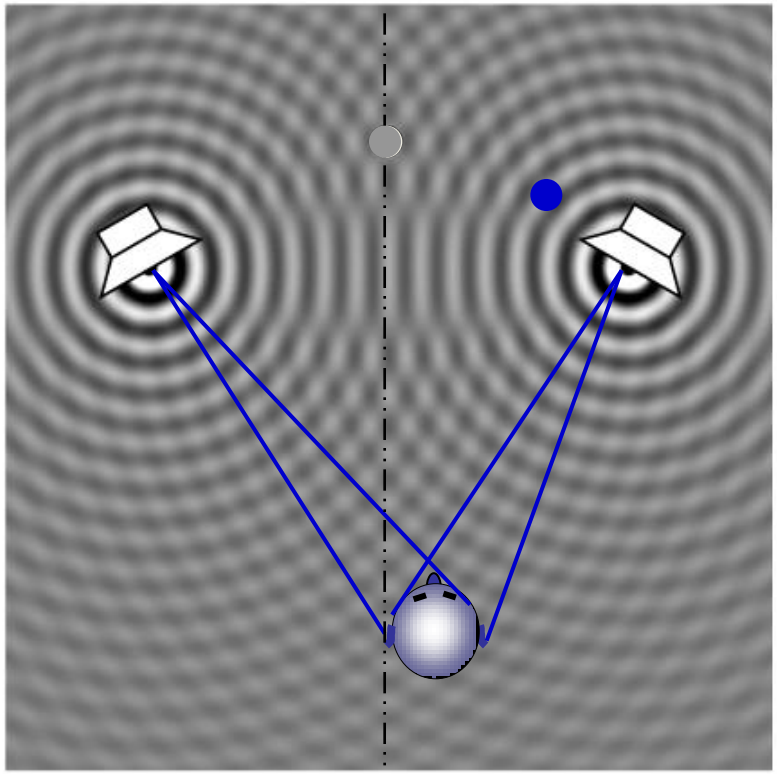
Two-channel stereo, interchannel time difference Δ t = – 0,5 ms

Phantom sound source, – 25°



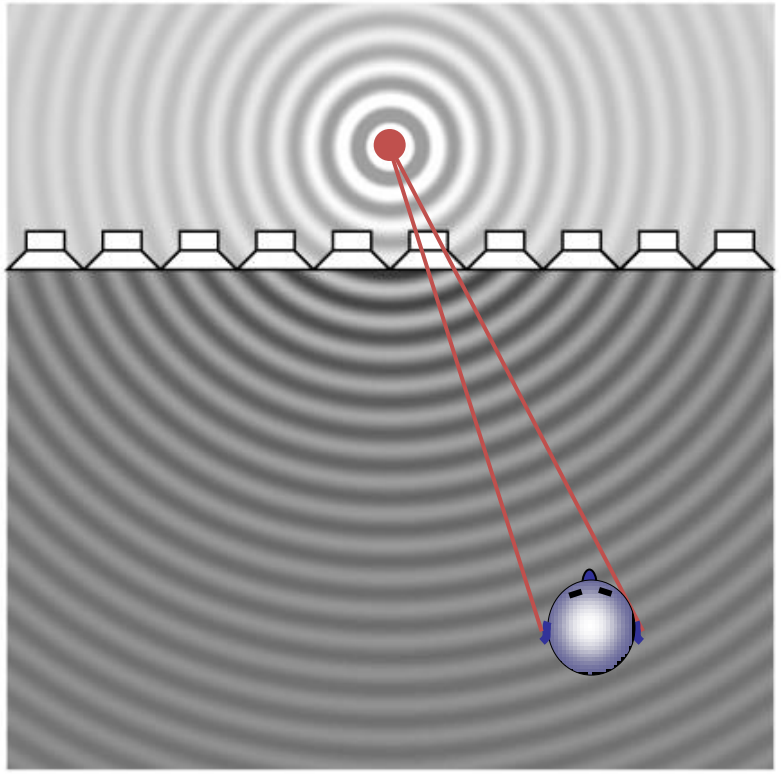
Two-channel stereo, interchannel time difference Δ t = + 0,5 ms

Phantom sound source, + 25°



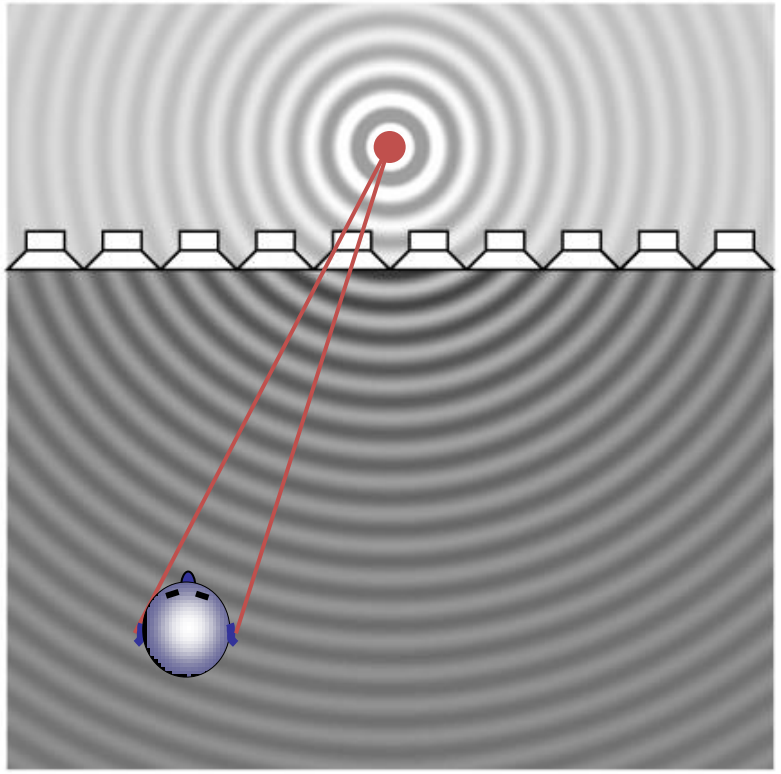
Virtual sound source, basic principal

Virtual sound source, – 25°



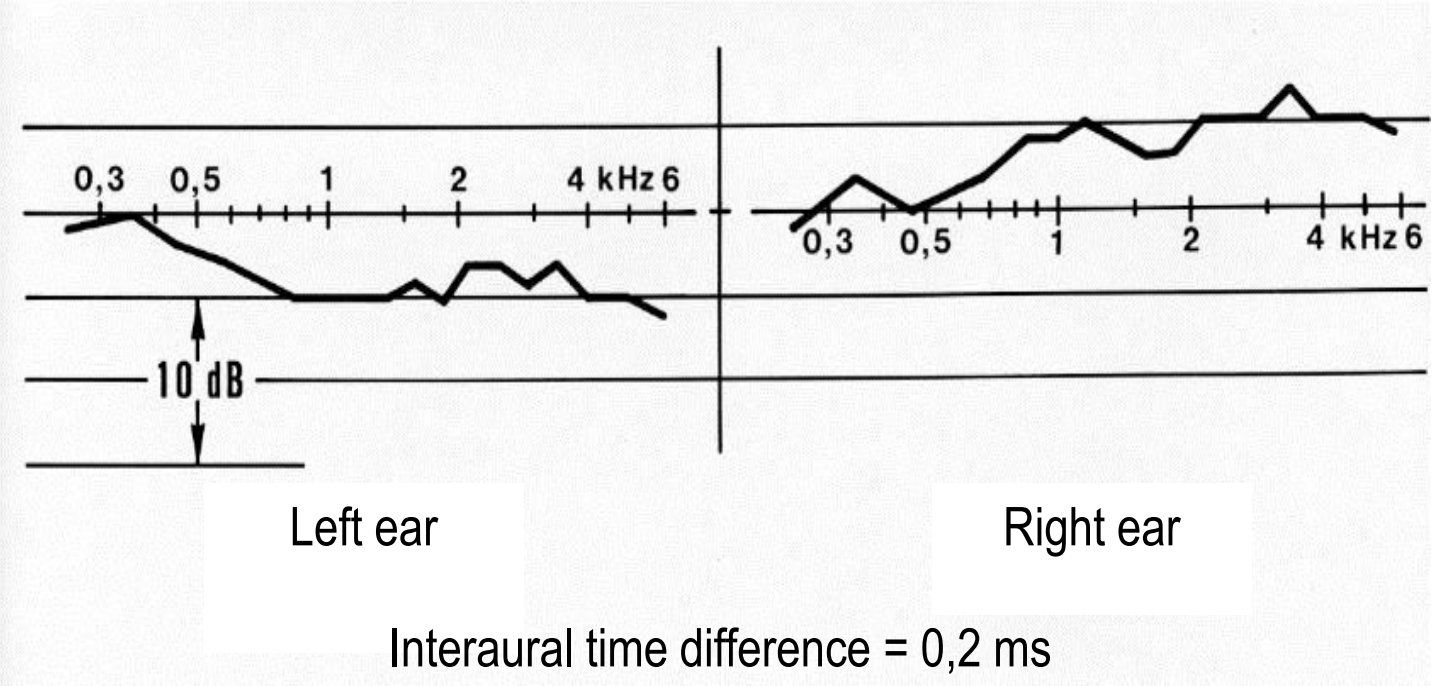
Virtual sound source, basic principal

Virtual sound source, + 25°



Ear input signals

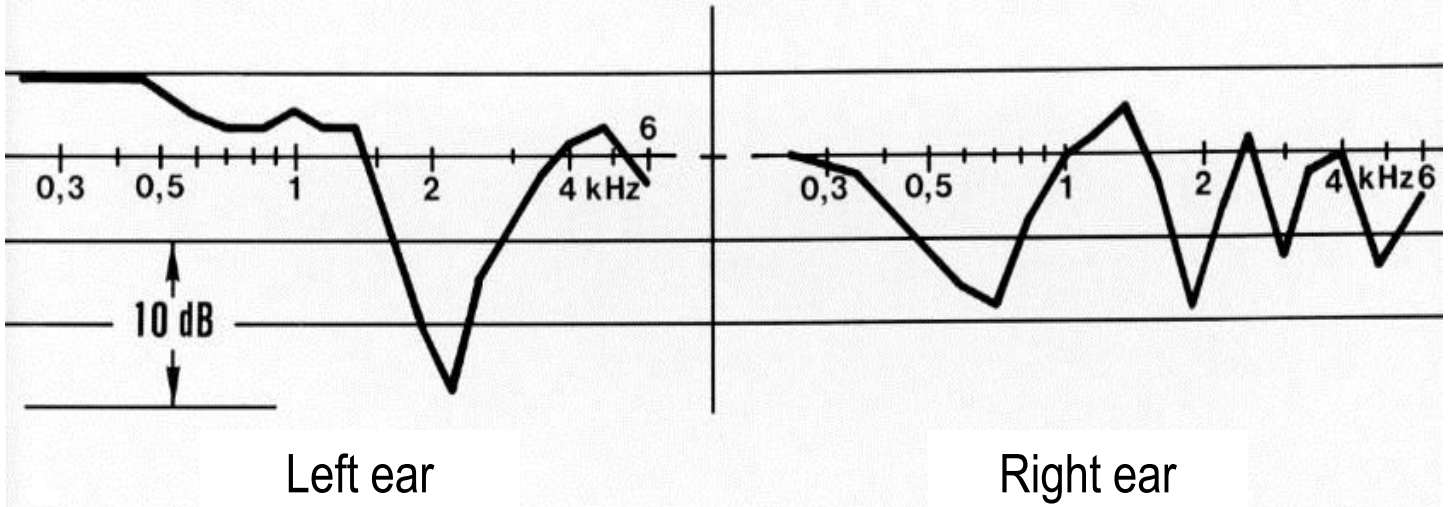
Virtual sound source (sound field synthesis)



→ Perceived direction = + 25°

Ear input signals

Phantom sound source (stereophony)

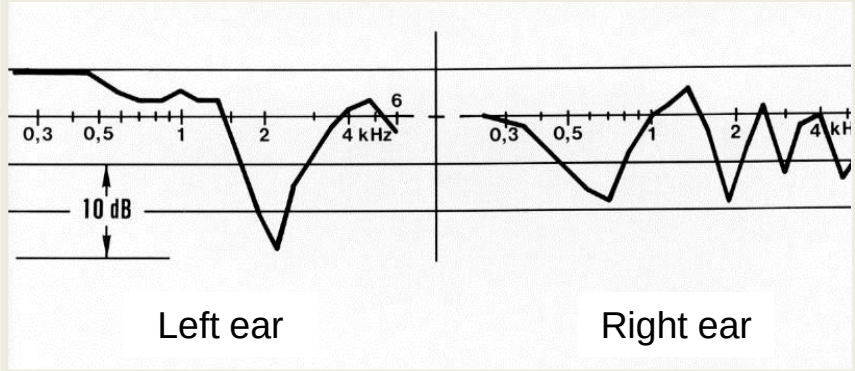


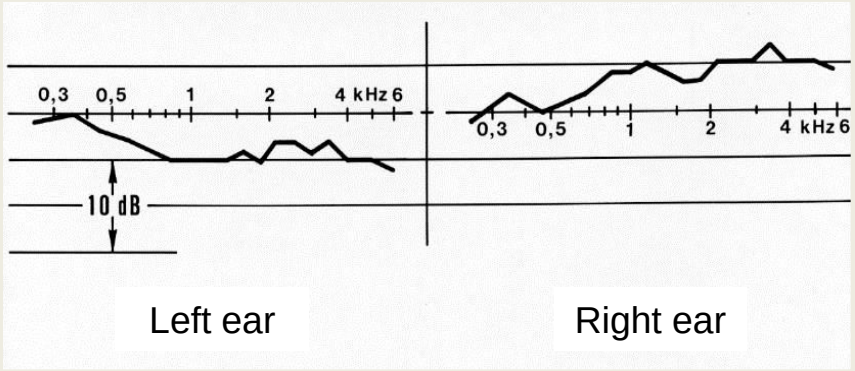
Comb filter effect due to interchannel delay

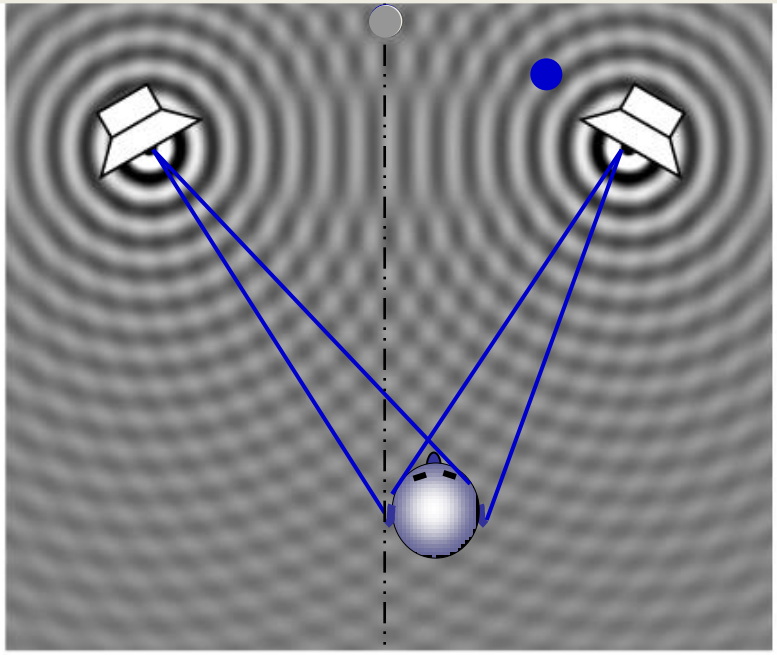
→ Perceived direction = + 25°

Stereophony vs. sound field synthesis

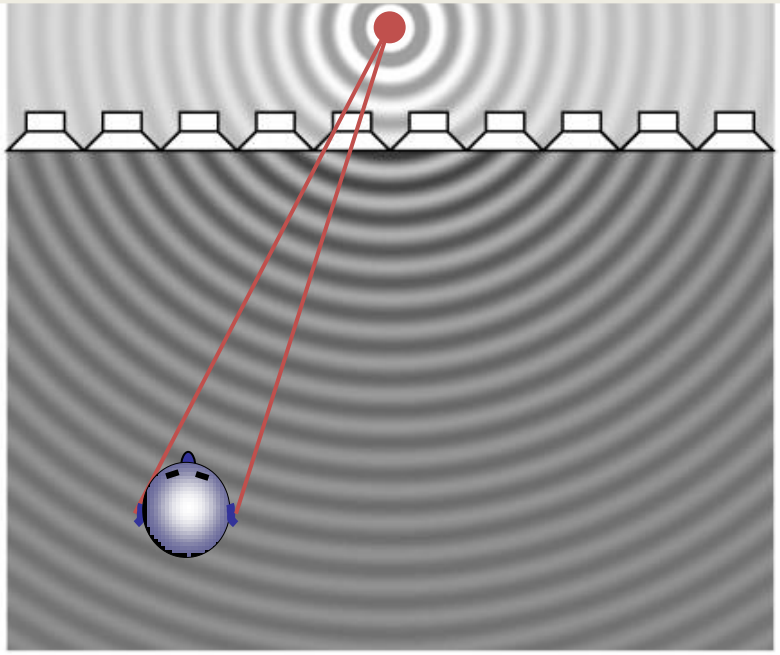
Perceived direction + 25°







Phantom sound source

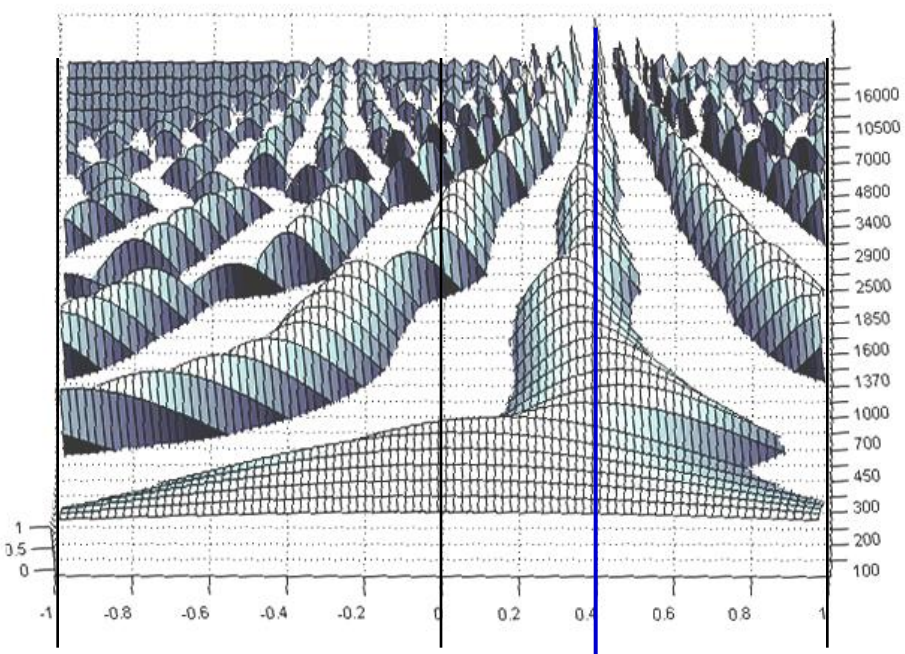


Virtual sound source

Interaural correlation pattern, perceived direction + 15°

Phantom sound source

Interchannel delay 0,4 ms

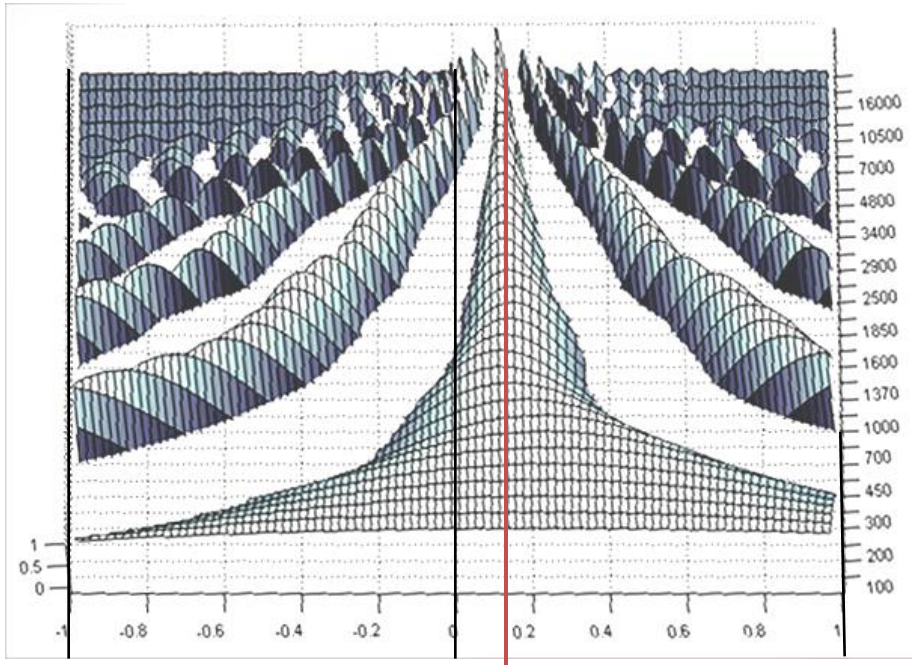


-1 ms 0 +1 ms

0,4 ms

Virtual sound source

Direction + 15°



-1 ms 0 +1 ms

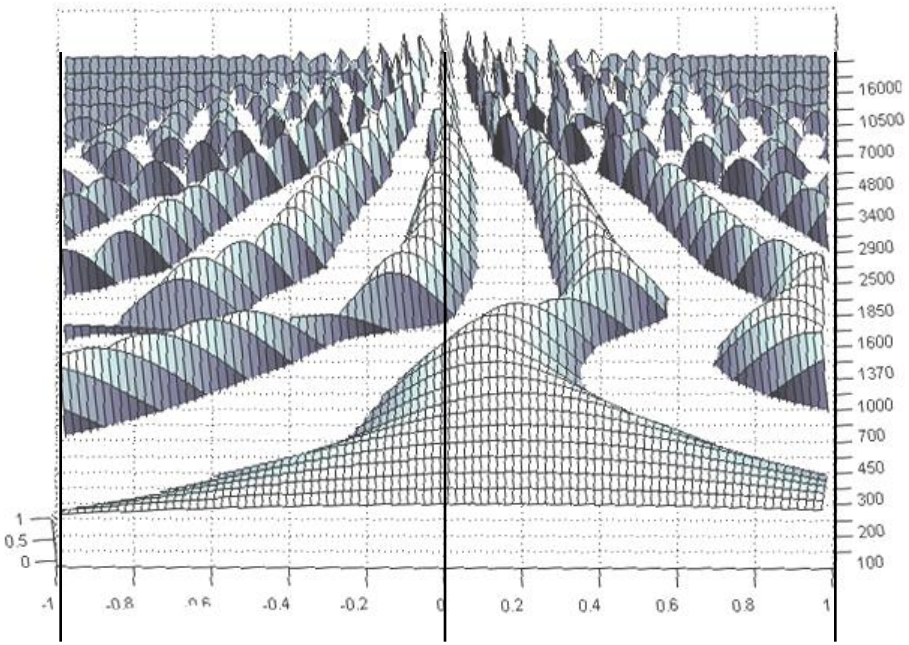
0,13 ms

Plots from Wegmann / Wittek (2005)

Interaural correlation pattern, perceived direction + 15°

Phantom sound source

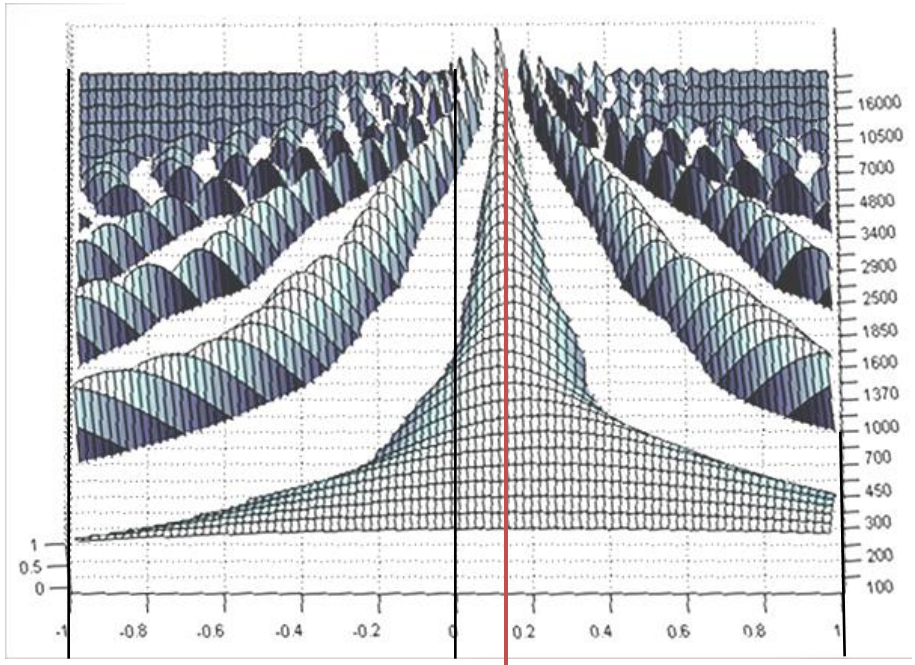
Interchannel level difference 7 dB



-1 ms 0 +1 ms

Real or virtual sound source

Direction + 15°

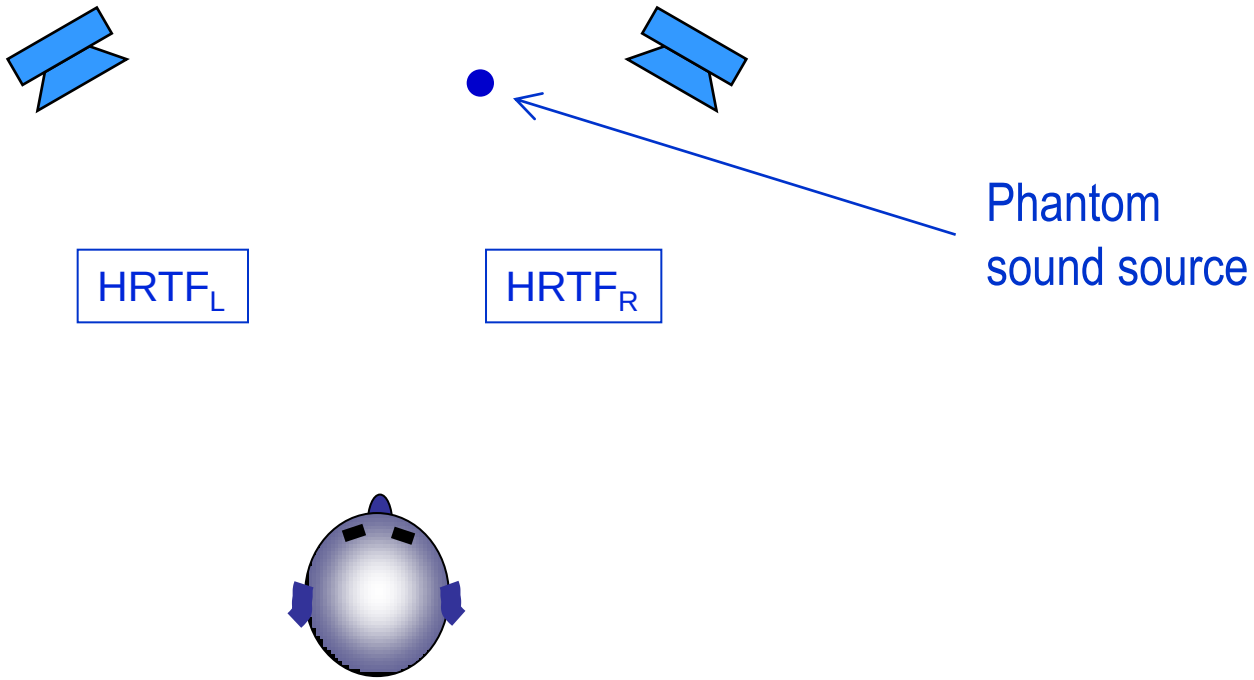


-1 ms 0 +1 ms

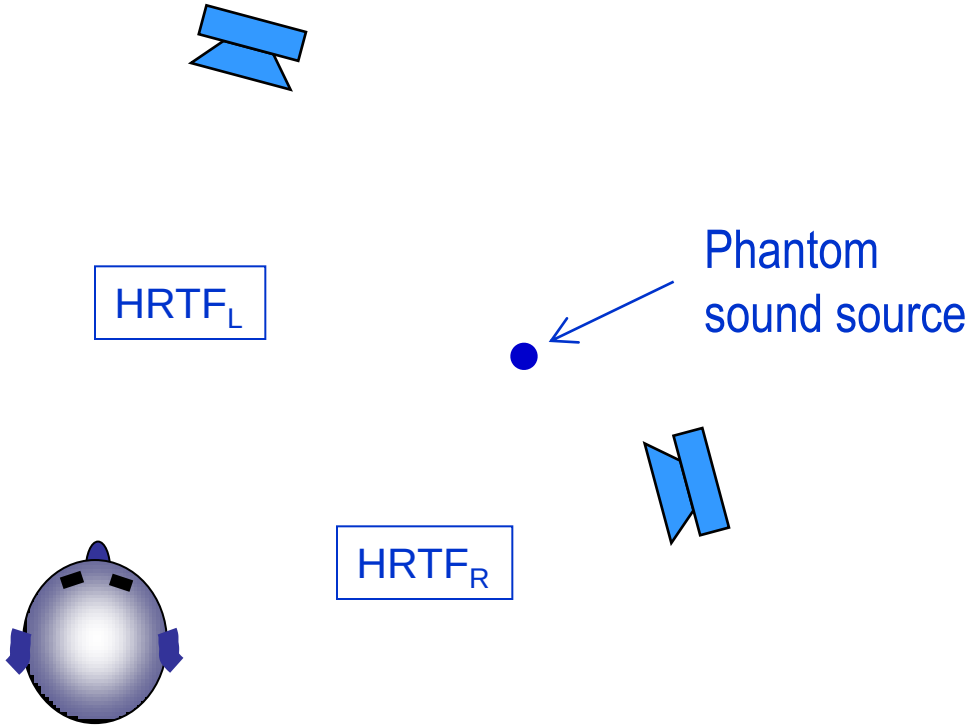
0,13 ms

Plots from Wegmann / Wittek (2005)

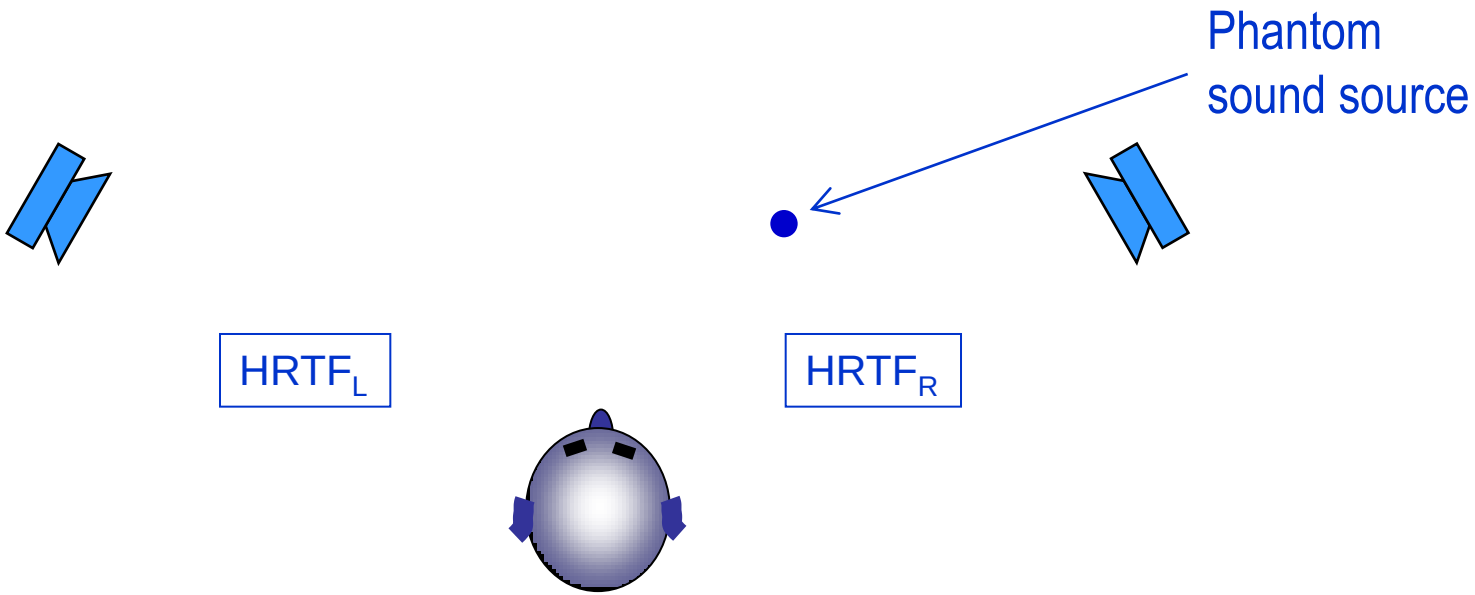
Two-channel stereo, phantom source shift 30 %



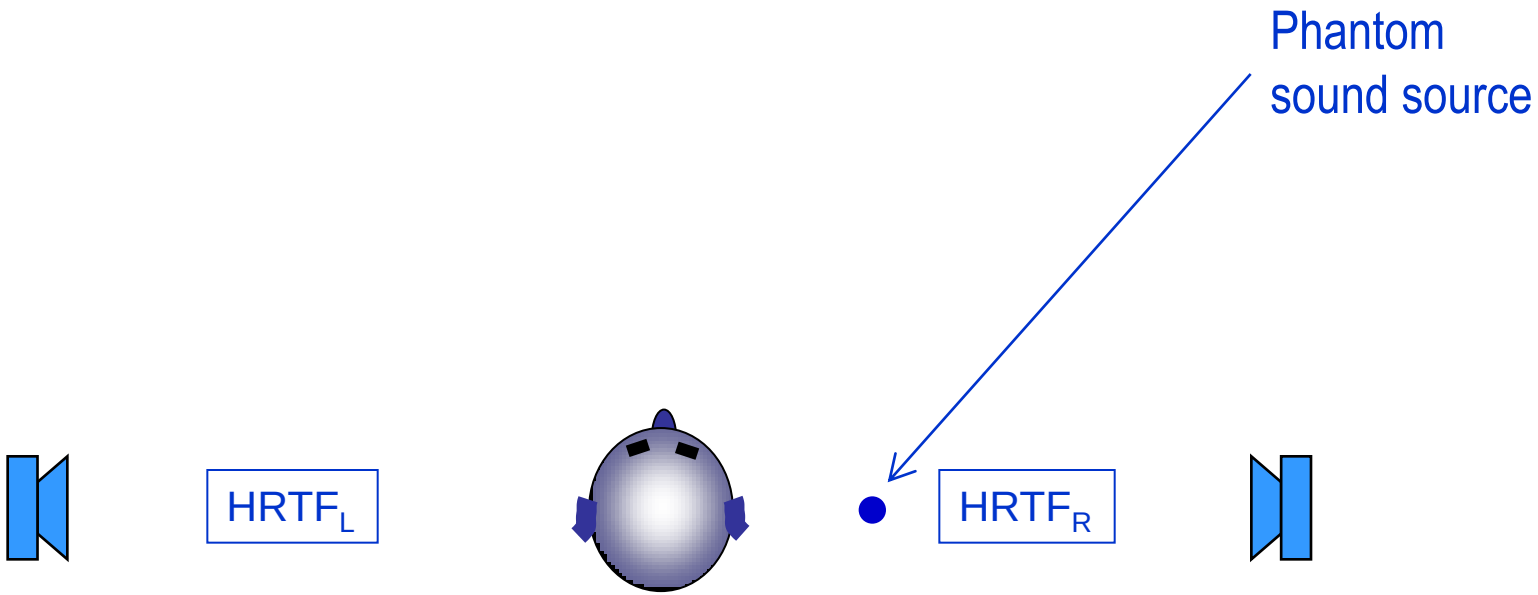
Two-channel stereo, phantom source shift 30 %



Two-channel stereo, phantom source shift 30 %

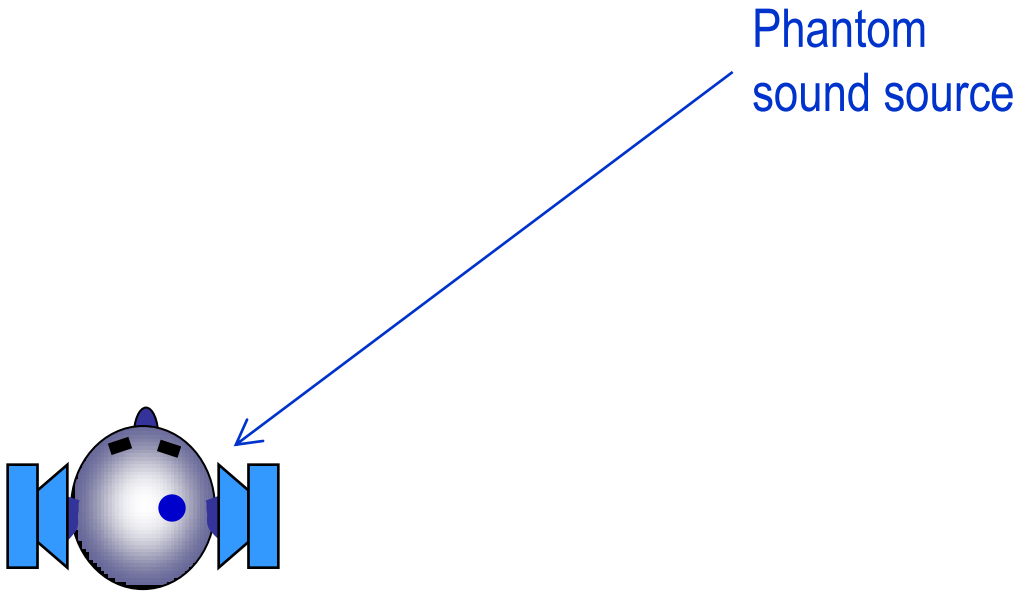


Two-channel stereo, phantom source shift 30 %

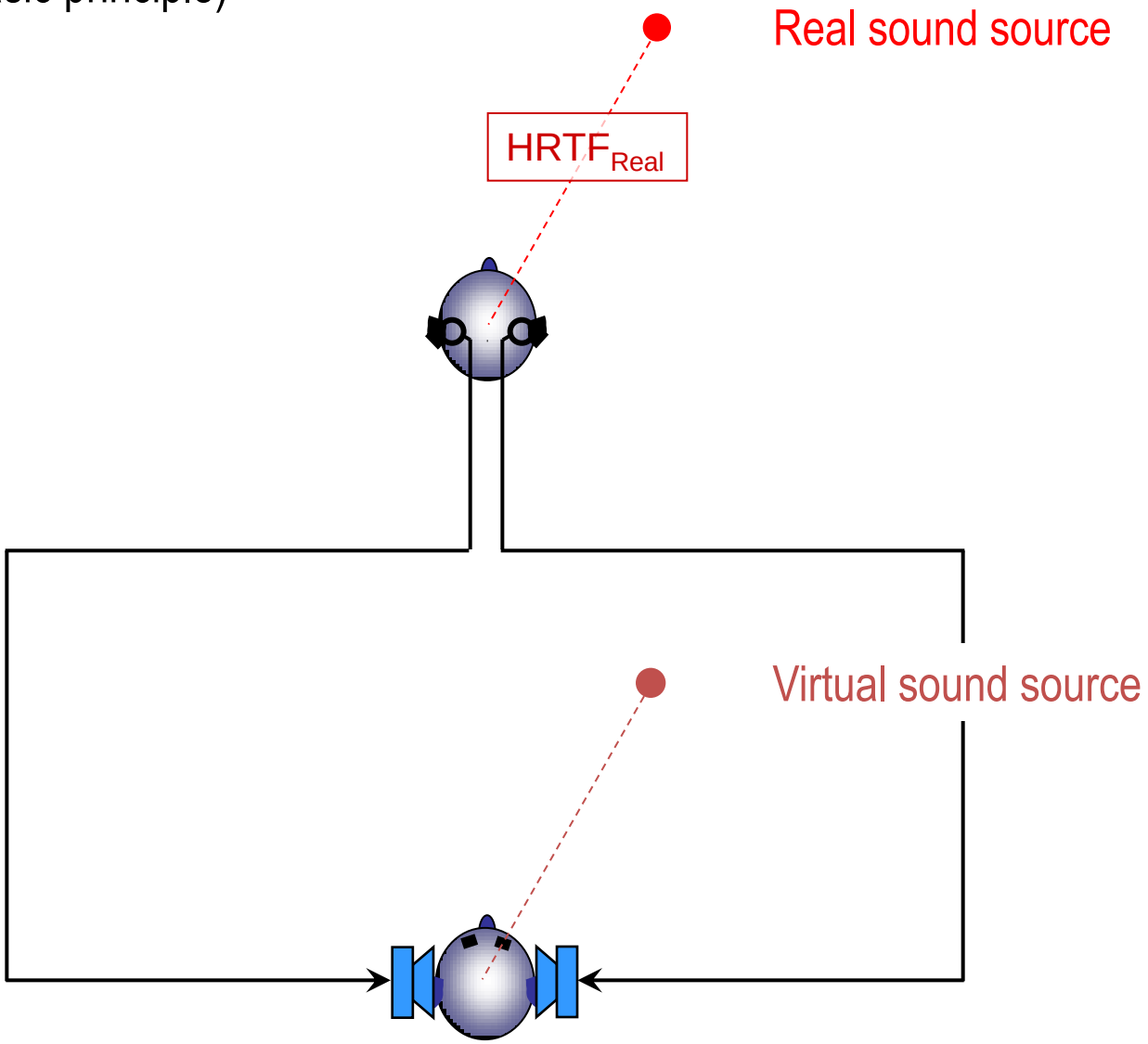


Stereophony

Two-channel stereo, in-head-localization, phantom source shift 30 %

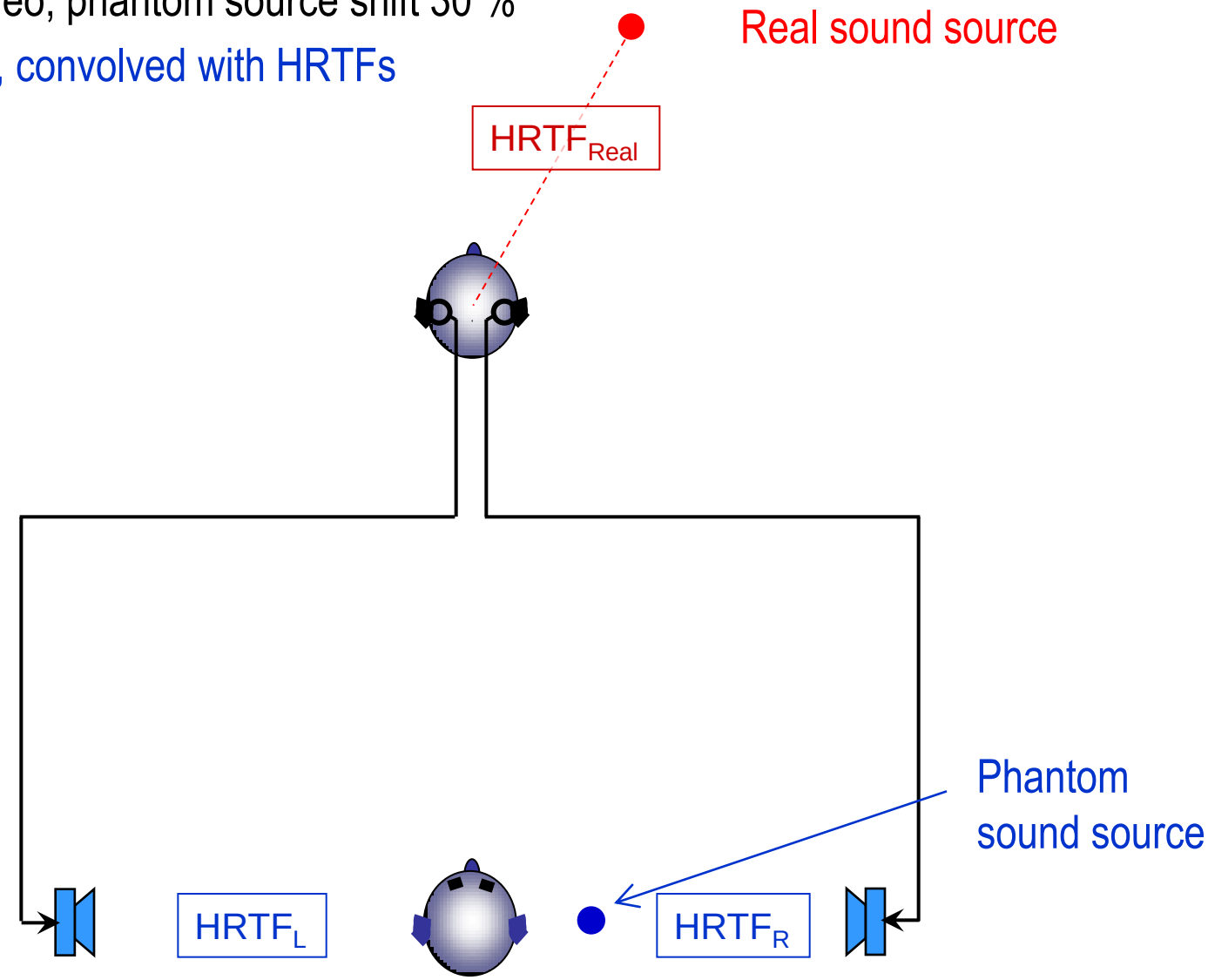


Dummy head (basic principle)



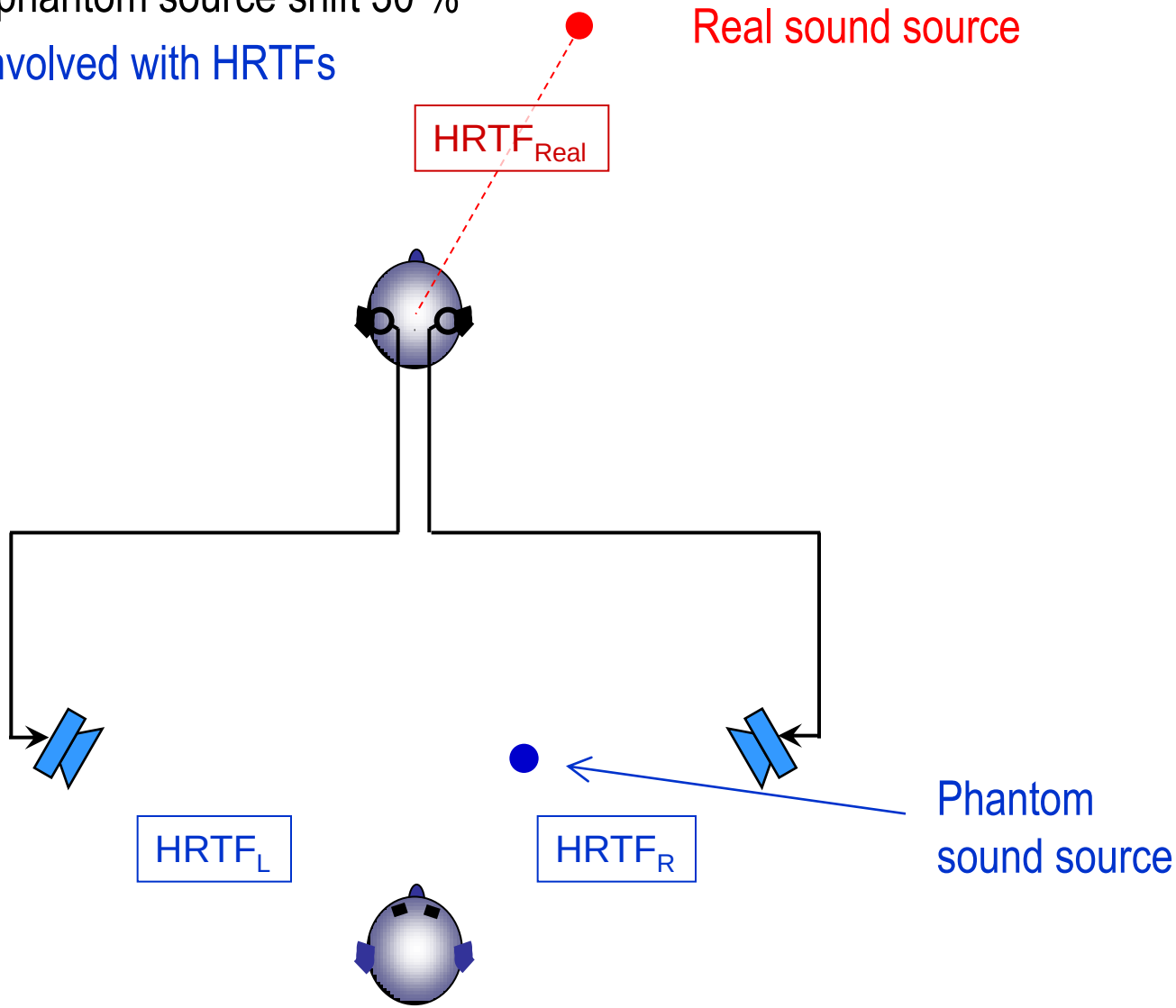
Two-channel stereo, phantom source shift 30 %

Binaural signals, convolved with HRTFs



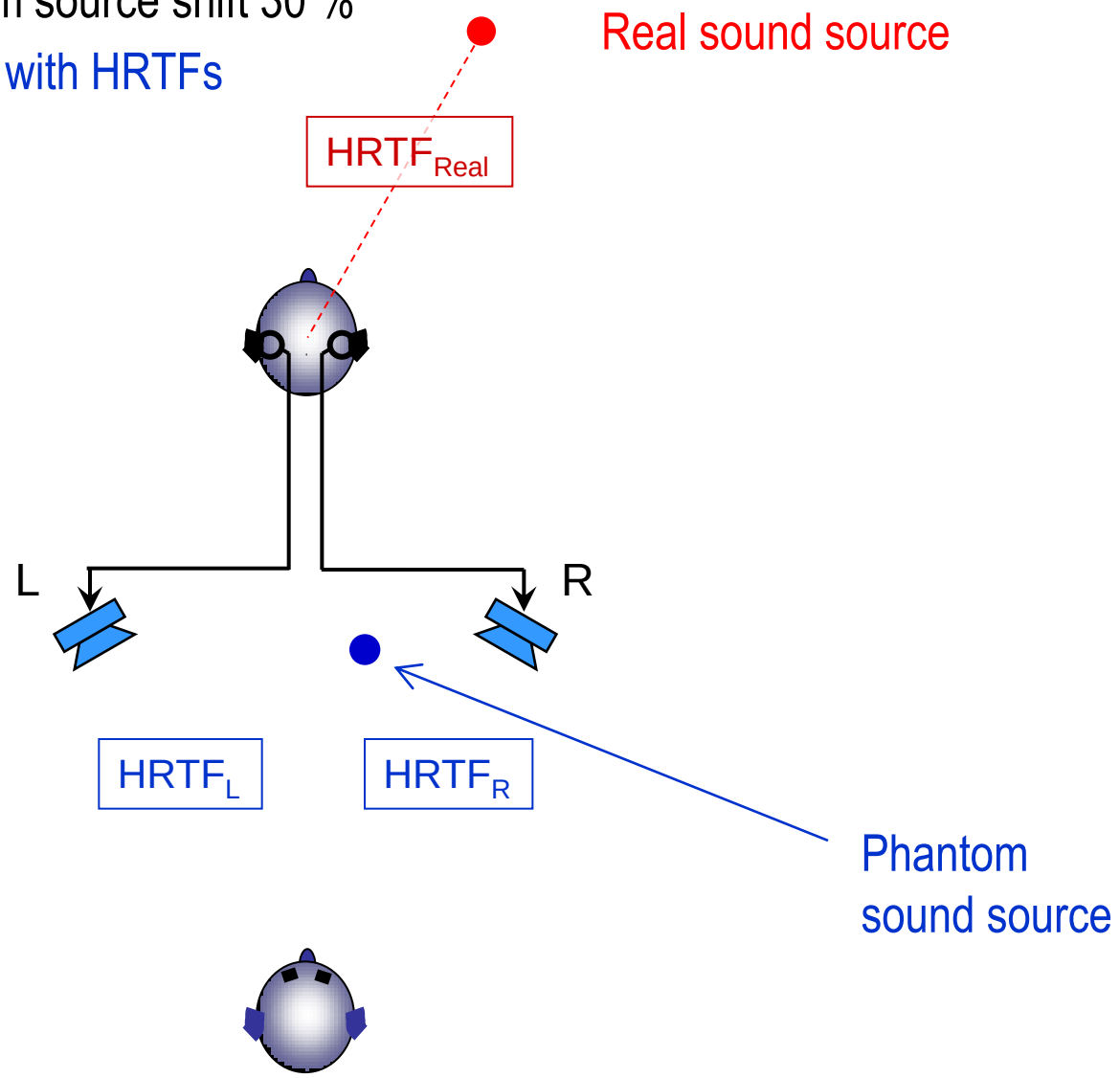
Two-channel stereo, phantom source shift 30 %

Binaural signals, convolved with HRTFs

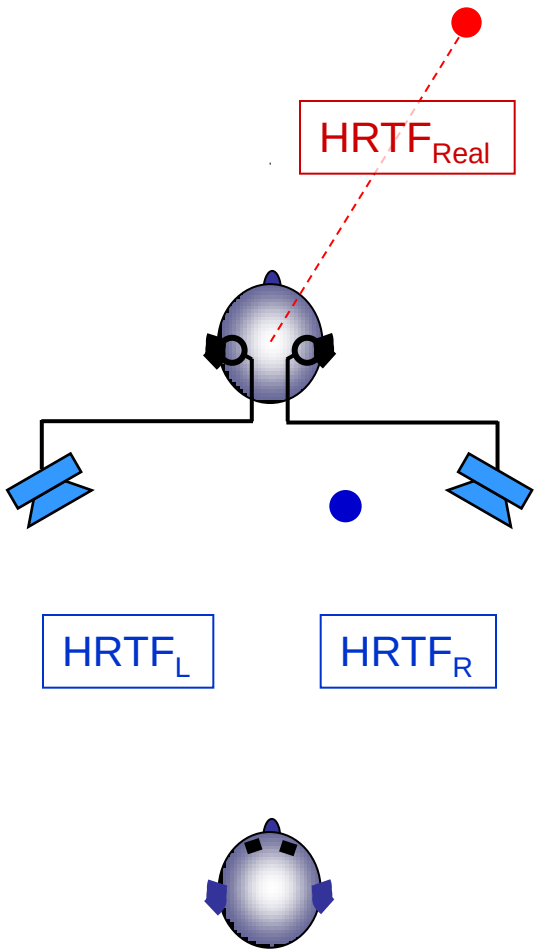


Two-channel stereo, phantom source shift 30 %

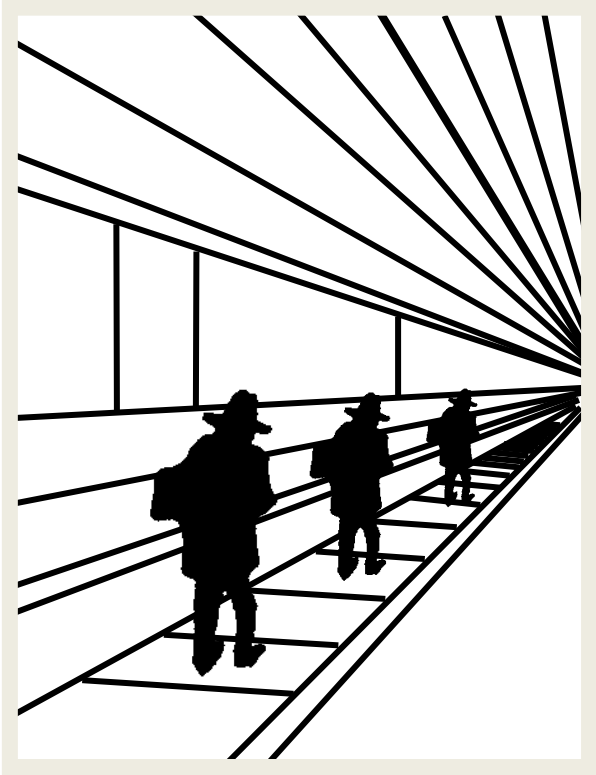
Binaural signals, convolved with HRTFs



Two-channel stereo, phantom source shift 30 %



The distance of the picture can be compared with the distance of stereo loudspeakers.
The stereophonic perspective can be simulated by applying phenomena of spatial hearing.



Two channel stereophonic imaging:

The location association stage recognizes the real distance of the speaker base, comparable to the real distance of the image.

Imaging of depth takes place in the "simulation plane“ by using phenomena of spatial hearing, comparable to the representation of depth in the picture.

The interaural signal differences of the dummy head ensure the optimum depth perception in the loudspeaker base.

The Sphere Microphone provides the interaural signal differences of the dummy head while neutrality of timbre.

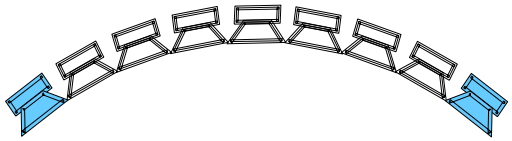
The Sphere Microphone (*Theile 1986*)

- head-related stereo microphone for loudspeaker reproduction
- very good imaging and sense of depth; natural localization, very natural sound
- with qualities of a dummy head when headphones are used for listening
- natural sound with superior low-frequency performance

Schoeps KFM 6
Stereo Sphere Microphone

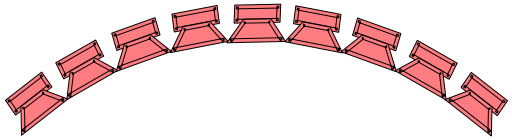
Source representation

Phantom source



Two loudspeakers
produce a stereophonic effect

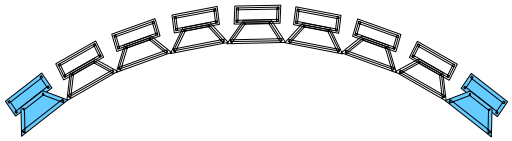
Virtual sound source



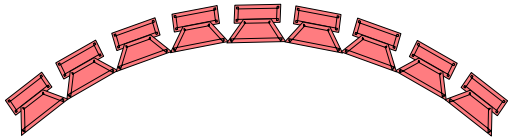
A loudspeaker array
produces a natural sound field

Source rendering

Phantom source



Virtual source



Phsychoacoustic approach

Guideline:

Utilization of stereophonic phenomena as perfect as possible

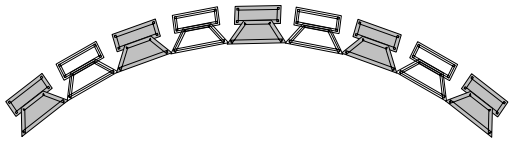
Physical approach

Guideline:

Reproduction of binaural ear signals as perfect as possible

Virtual source rendering

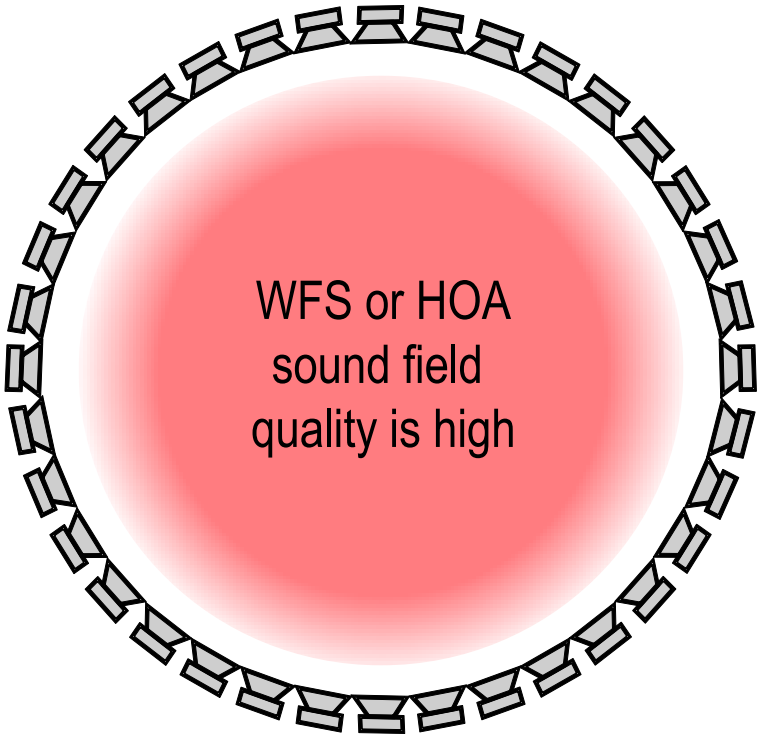
Virtual or phantom sound source?



?

Stereophony vs. sound field synthesis

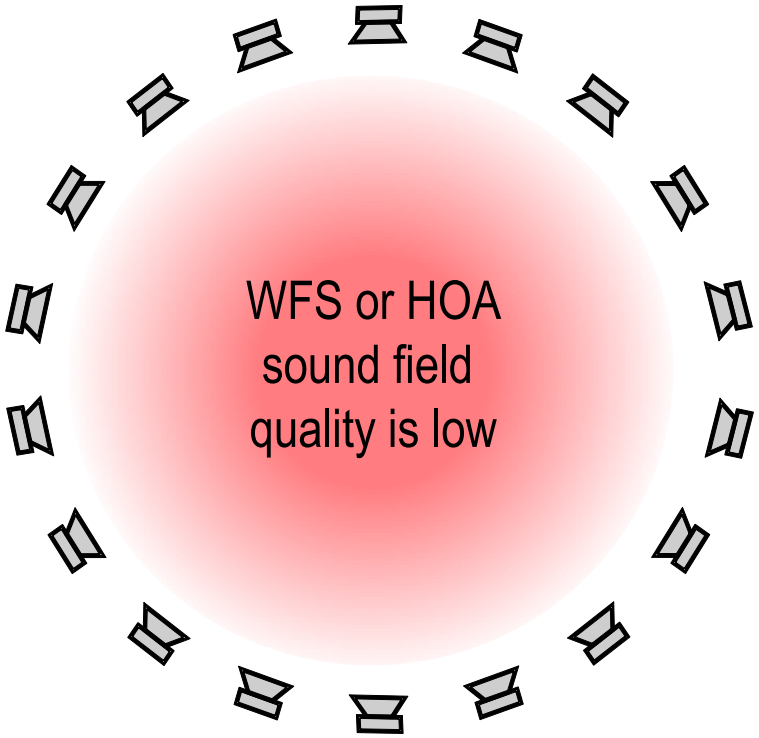
Virtual source rendering



High spatial resolution

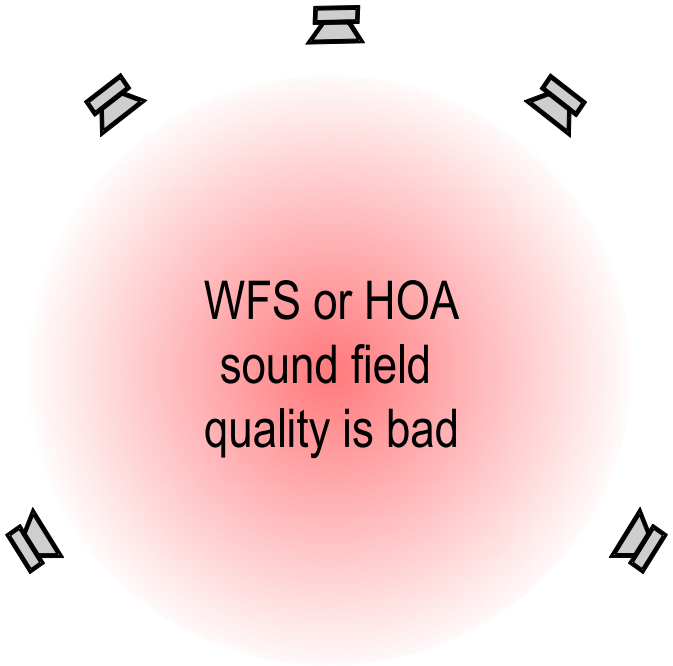
Stereophony vs. sound field synthesis

Virtual source rendering

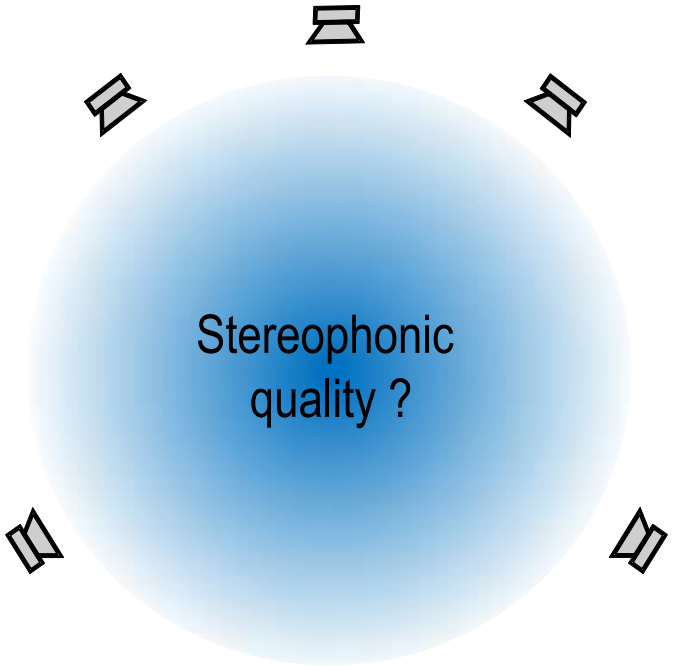


Low spatial resolution

Virtual source rendering

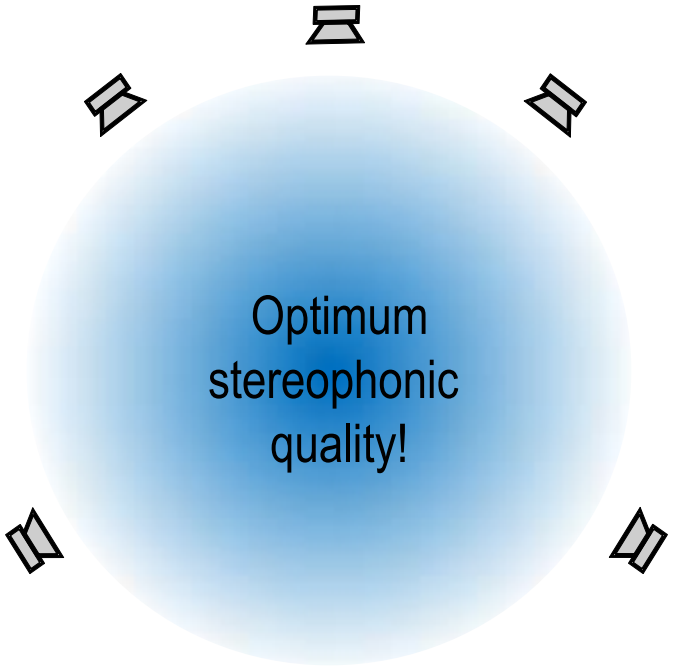


Phantom source imaging



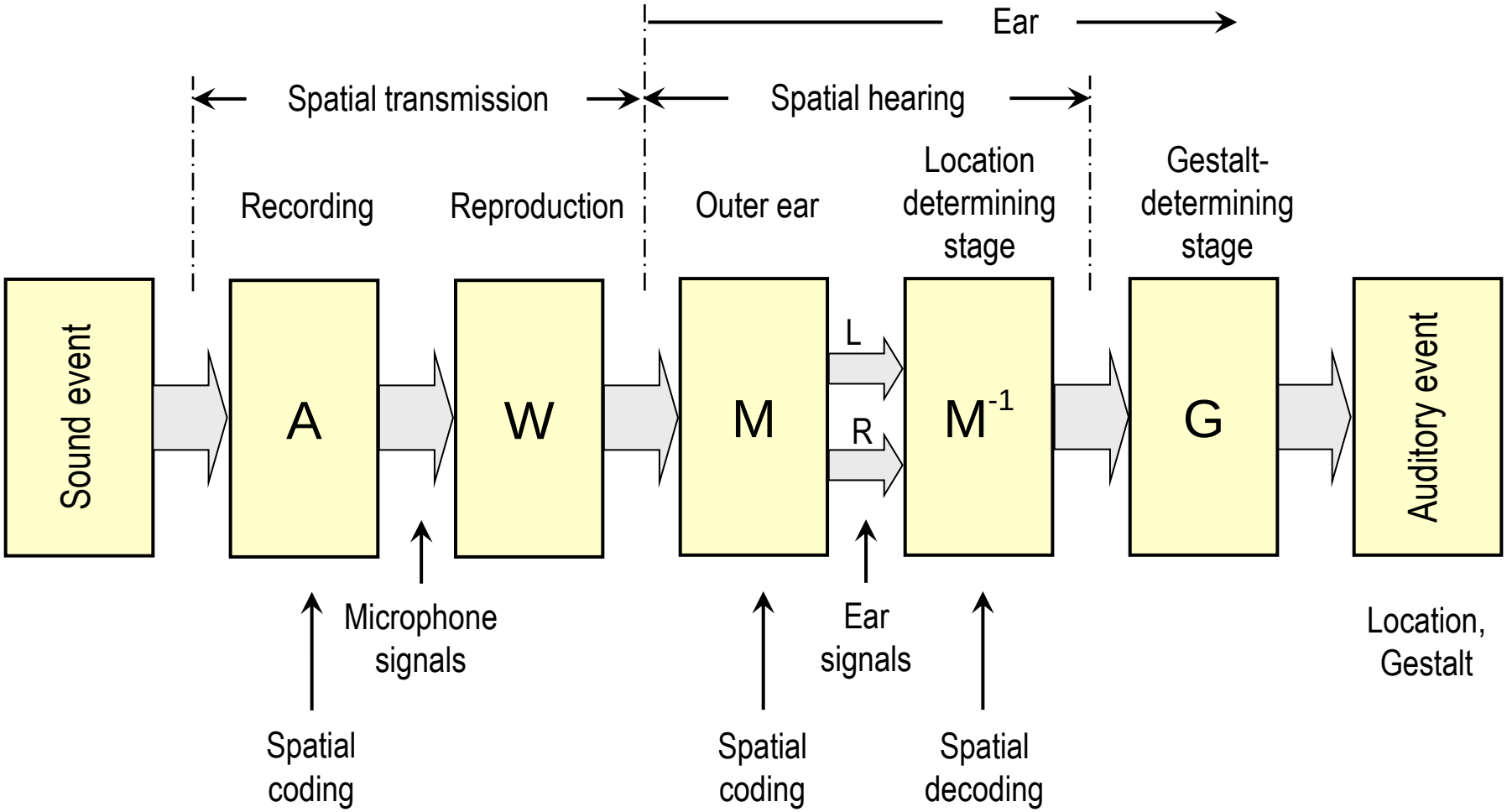
Phantom source imaging

If perfect utilization of
stereophonic phenomena

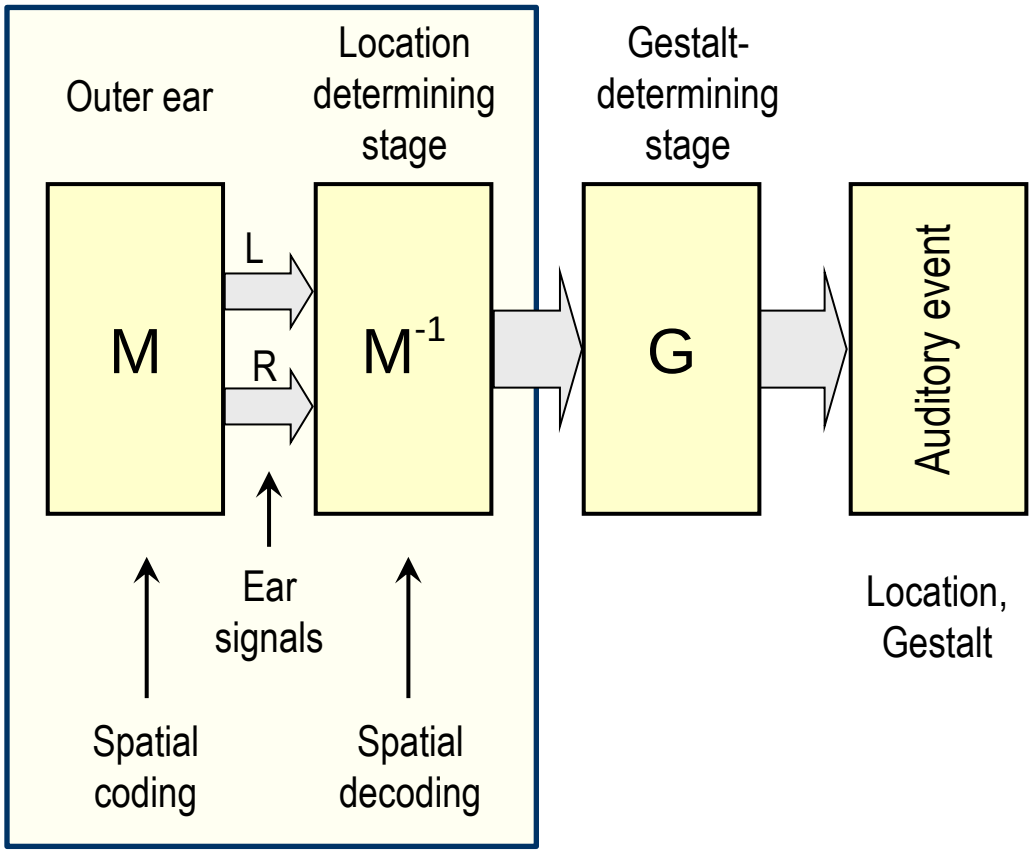


Correlated signals only in two (neighboring) channels („no crosstalk“) !!

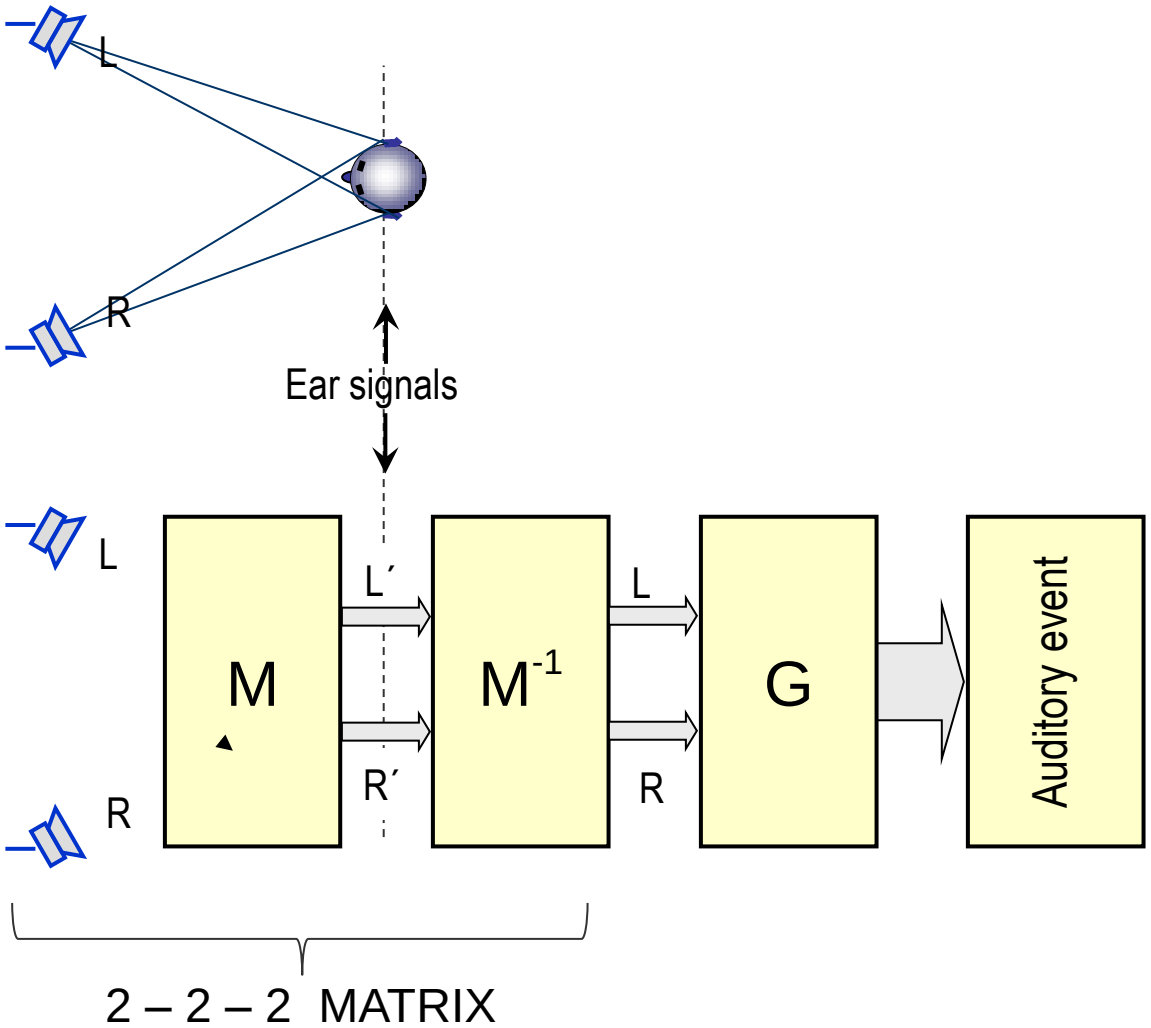
Phantom source imaging



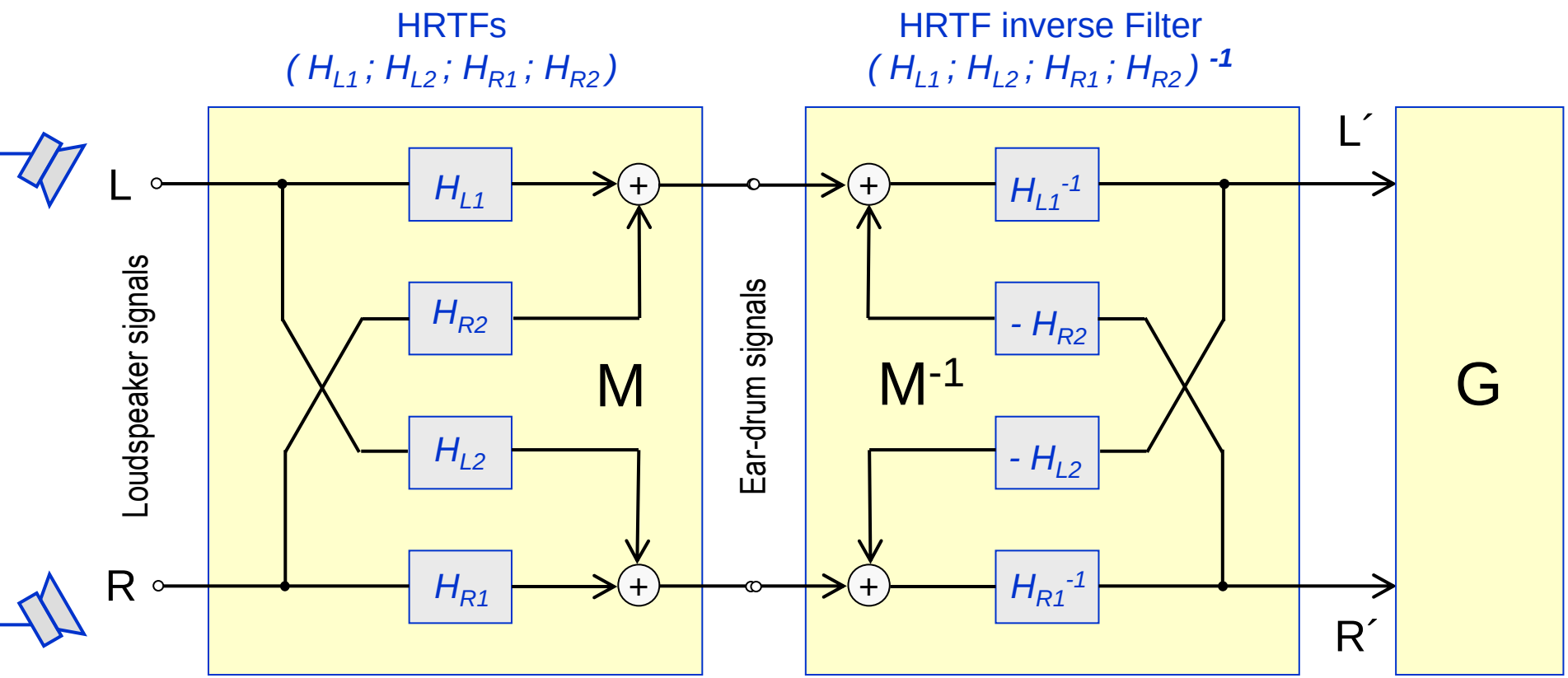
Phantom source imaging



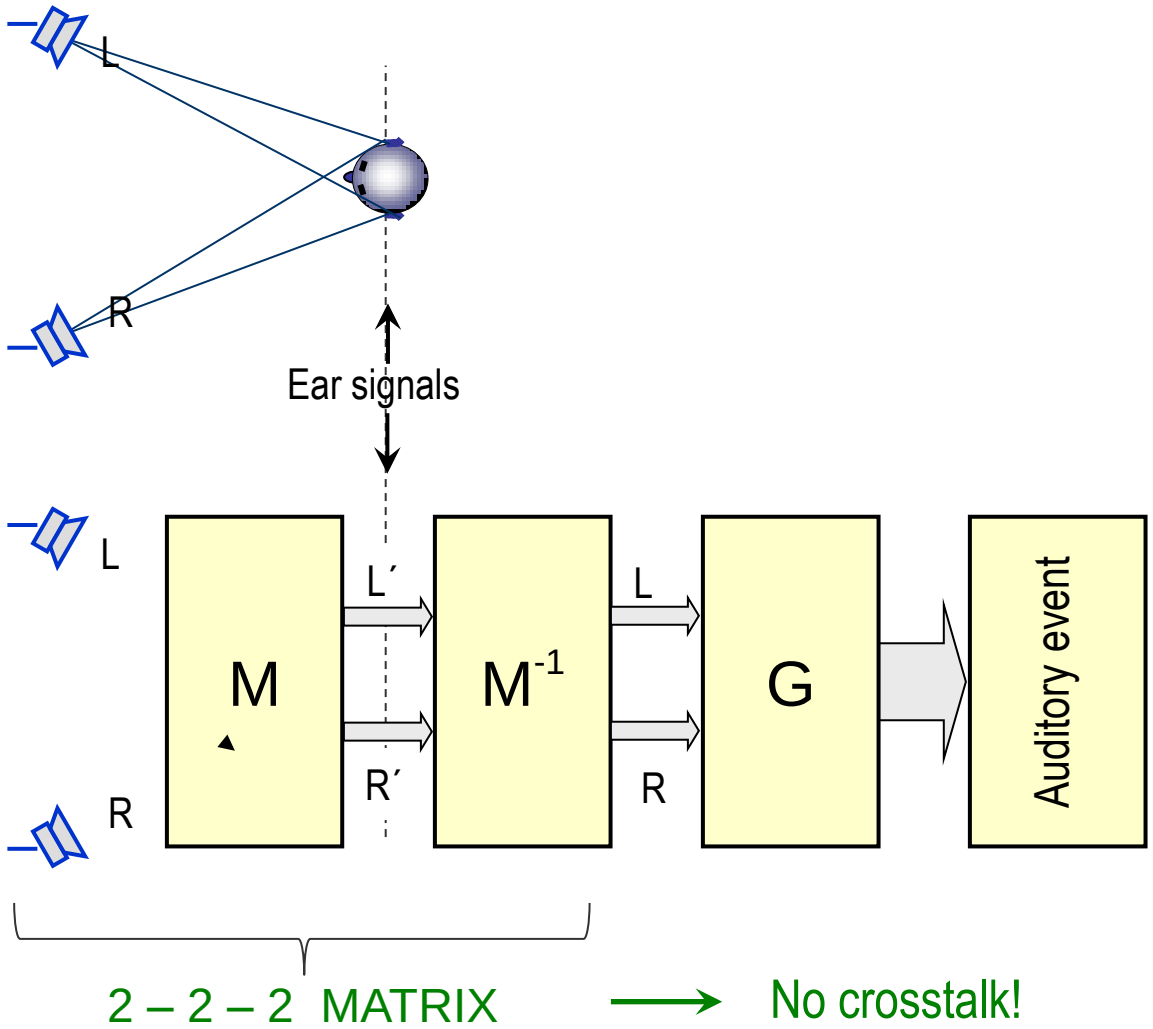
Phantom source imaging



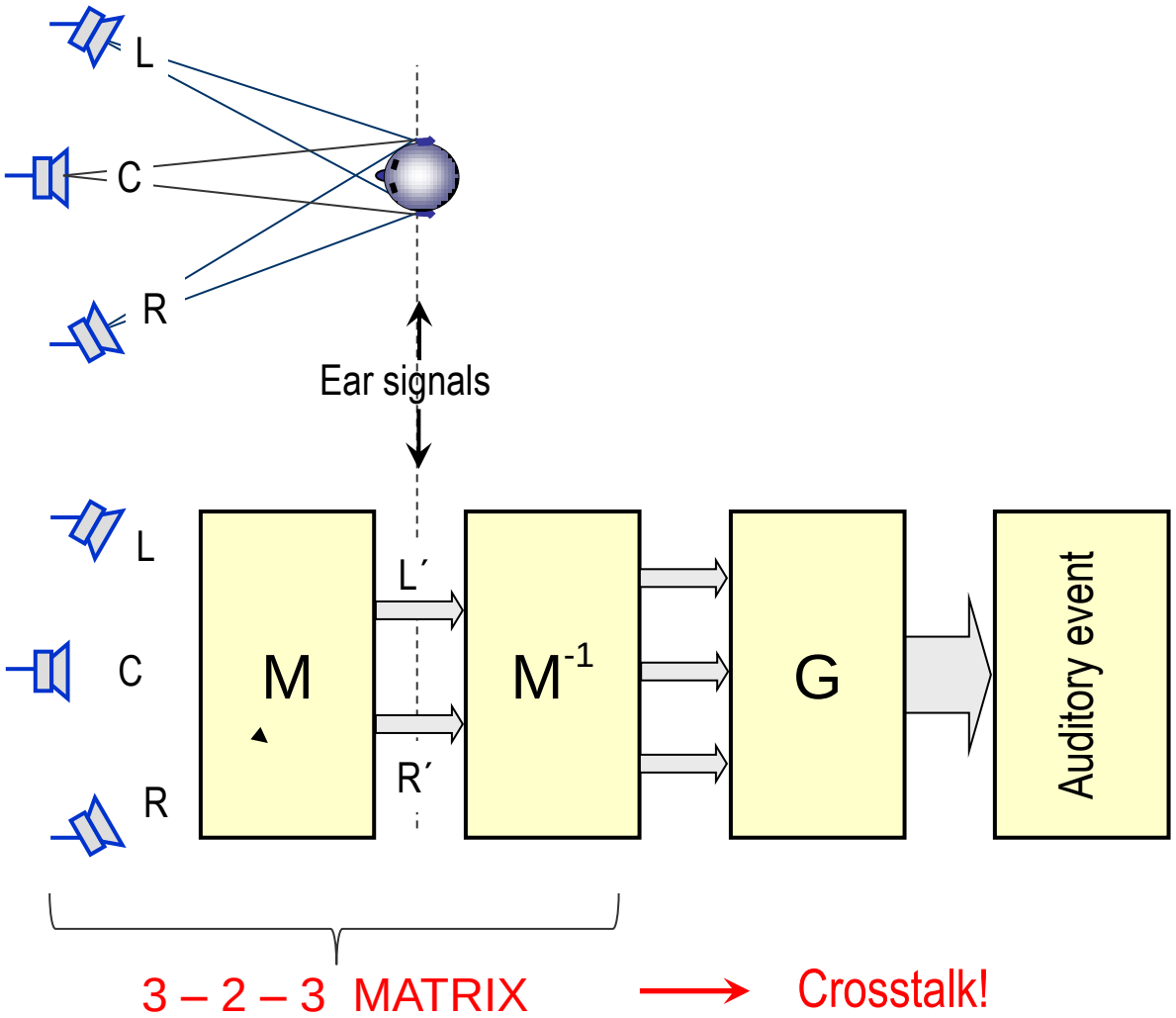
Phantom source imaging



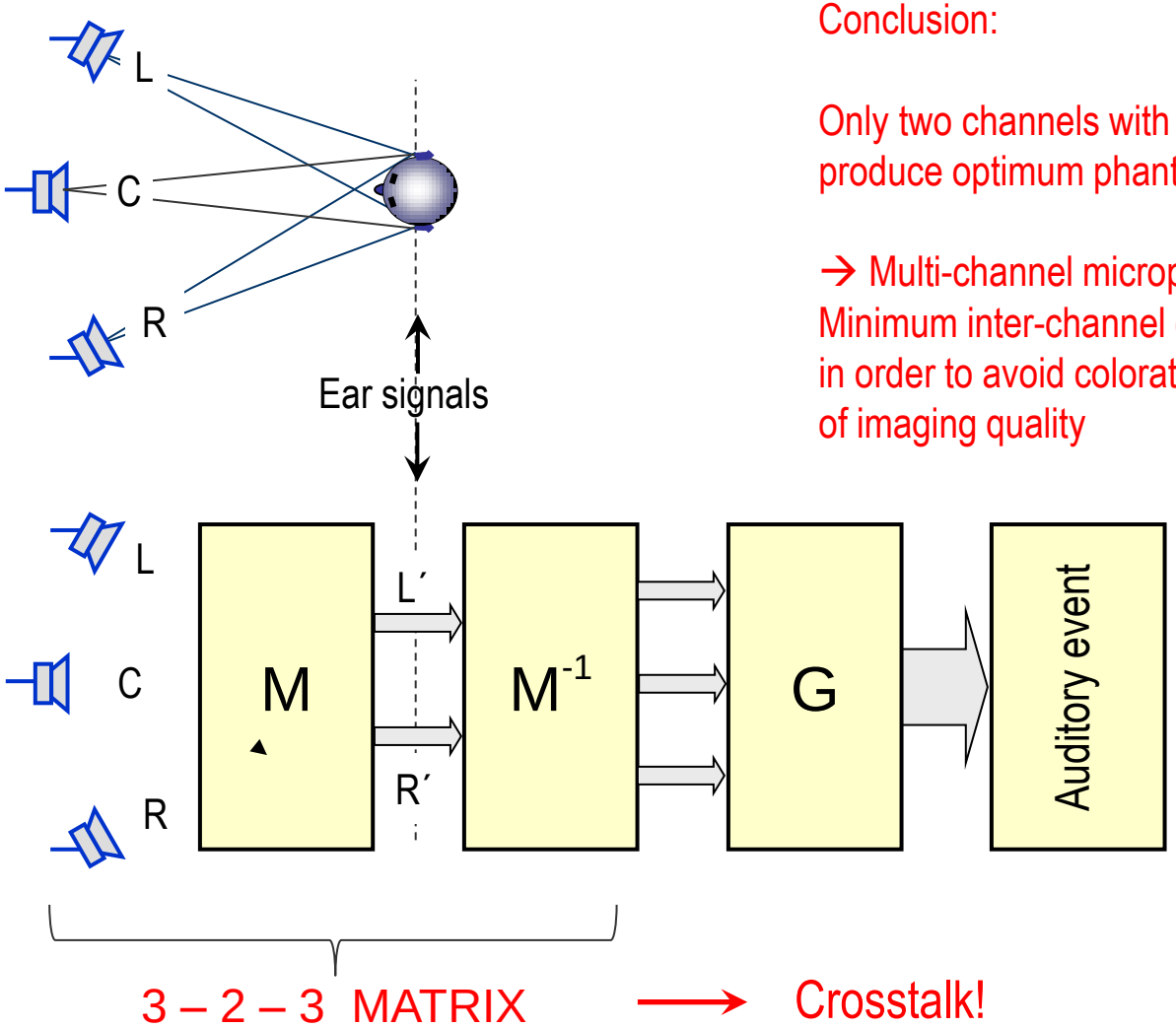
Phantom source imaging



Phantom source imaging



Phantom source imaging



Conclusion:

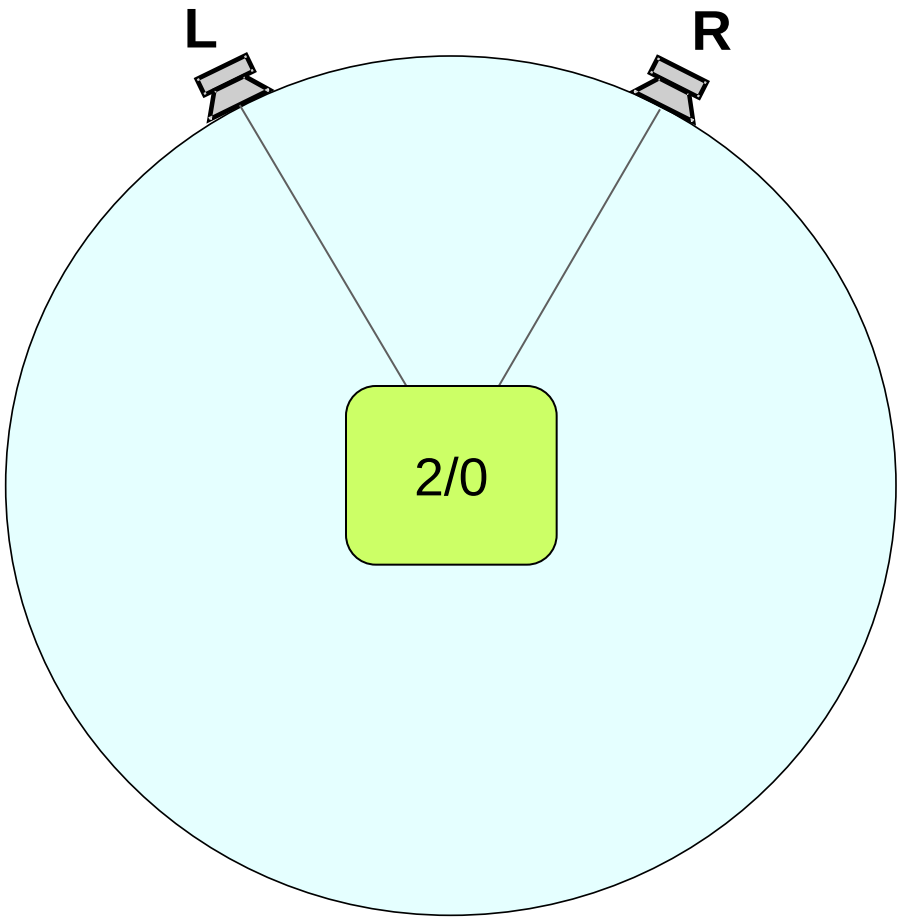
Only two channels with coherent signals can produce optimum phantom source quality.

→ Multi-channel microphone:
Minimum inter-channel crosstalk is required in order to avoid coloration, blurring and lack of imaging quality

Audio developments based on psychoacoustics

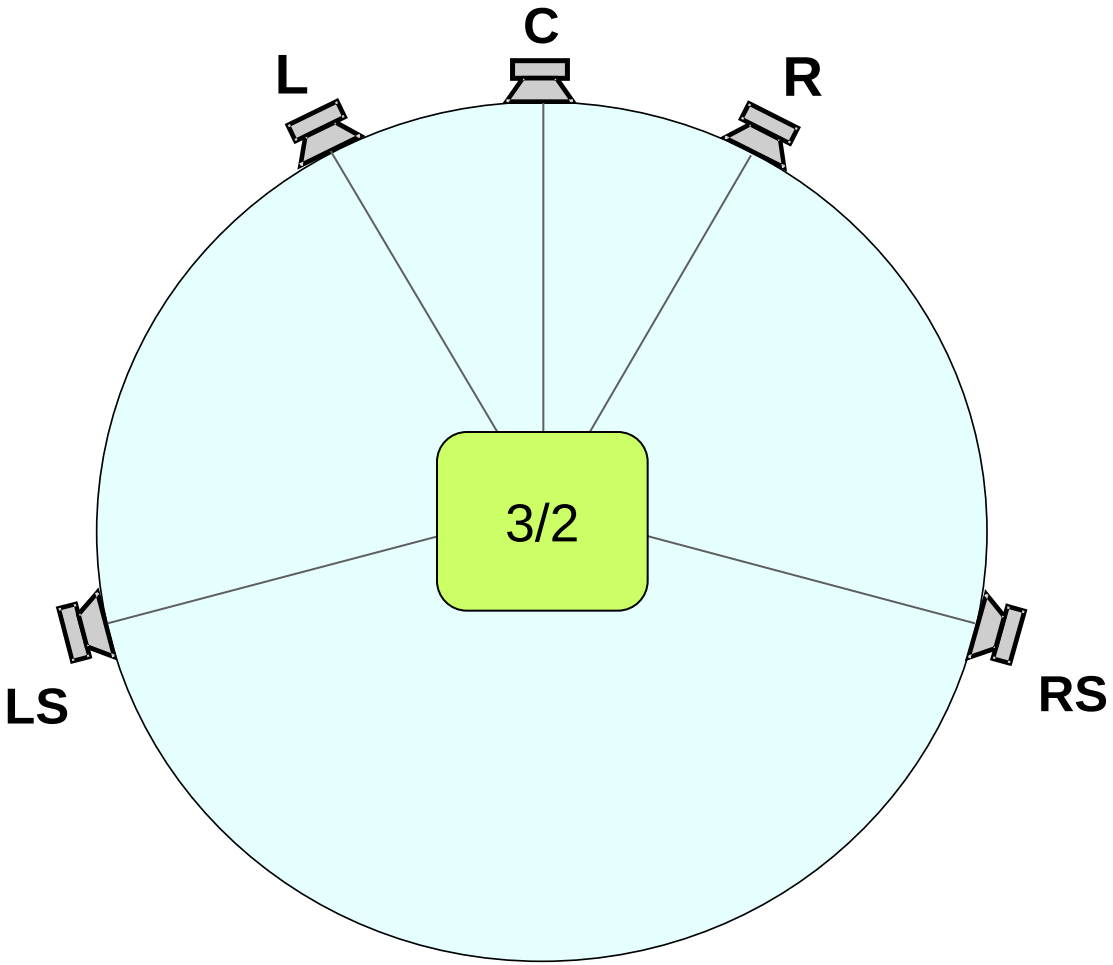
- Data reduction
- Headphone reproduction
- Binaural technologies
- Sound field synthesis (WFS / HOA)
- Stereophony
- 5.1 / 7.1 Surround (ITU BS 775)
- 3D Audio

From 2/0 stereophony to 3D multichannel stereophony



Front direction, robustness	●
Surround direction	-
Elevation	-
Height	-
Distance / depth	(●)
Spatial impression	(●)
Envelopment	-

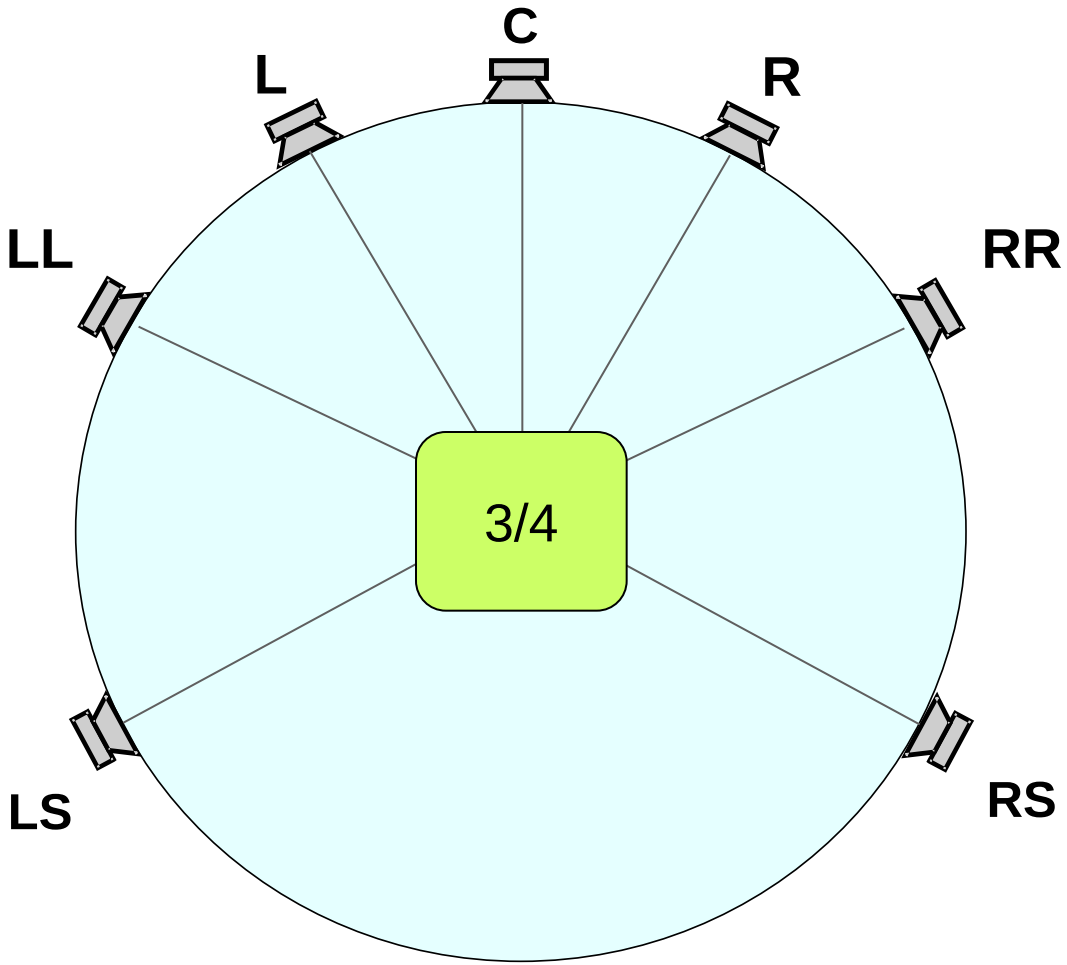
From 2/0 stereophony to 3D multichannel stereophony



“ ITU 5.1” (1992)

Front direction, robustness	●●
Surround direction	●
Elevation	-
Height	-
Distance / depth	●
Spatial impression	●
Envelopment	●

From 2/0 stereophony to 3D multichannel stereophony

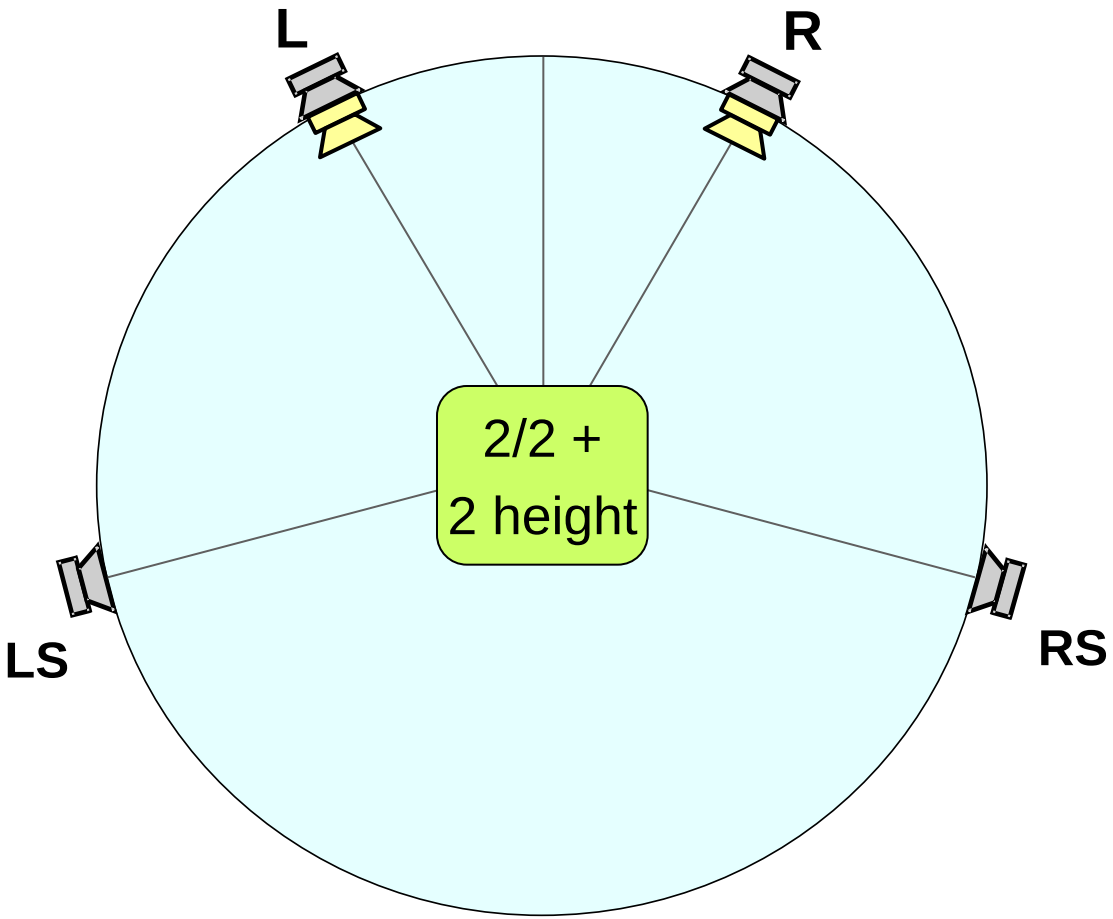


“ ITU 7.1” (1992)

Front direction, robustness	●●
Surround direction	●●
Elevation	-
Height	-
Distance / depth	●
Spatial impression	●
Envelopment	●

From 2/0 stereophony to 3D multichannel stereophony

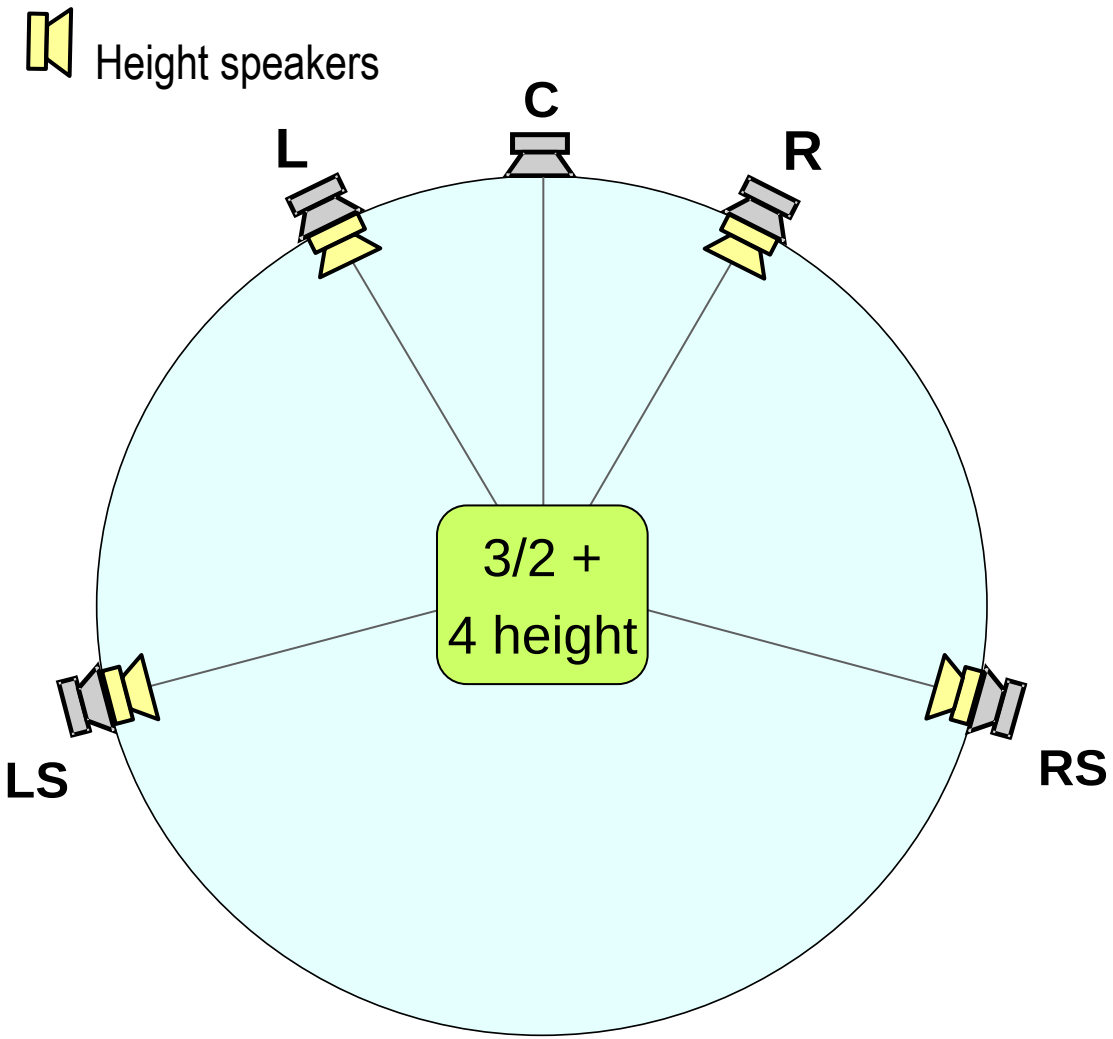
Height speakers



Front direction, robustness	●
Surround direction	●
Elevation	(●)
Height	(●)
Distance / depth	●
Spatial impression	●●
Envelopment	●

“ 2+2+2” (2001)

From 2/0 stereophony to 3D multichannel stereophony



Front direction, robustness	●●
Surround direction	●
Elevation	(●)
Height	●
Distance / depth	●●
Spatial impression	●●
Envelopment	●●

“ 3D Audio 9.1” (Auro 9.1, 2006)

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0
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

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1
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

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1
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

Multichannel stereophony formats – basic characteristics

Verband Deutscher Tonmeister e.V.

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

Binaural techniques - ear input based

Verband Deutscher Tonmeister e.V.

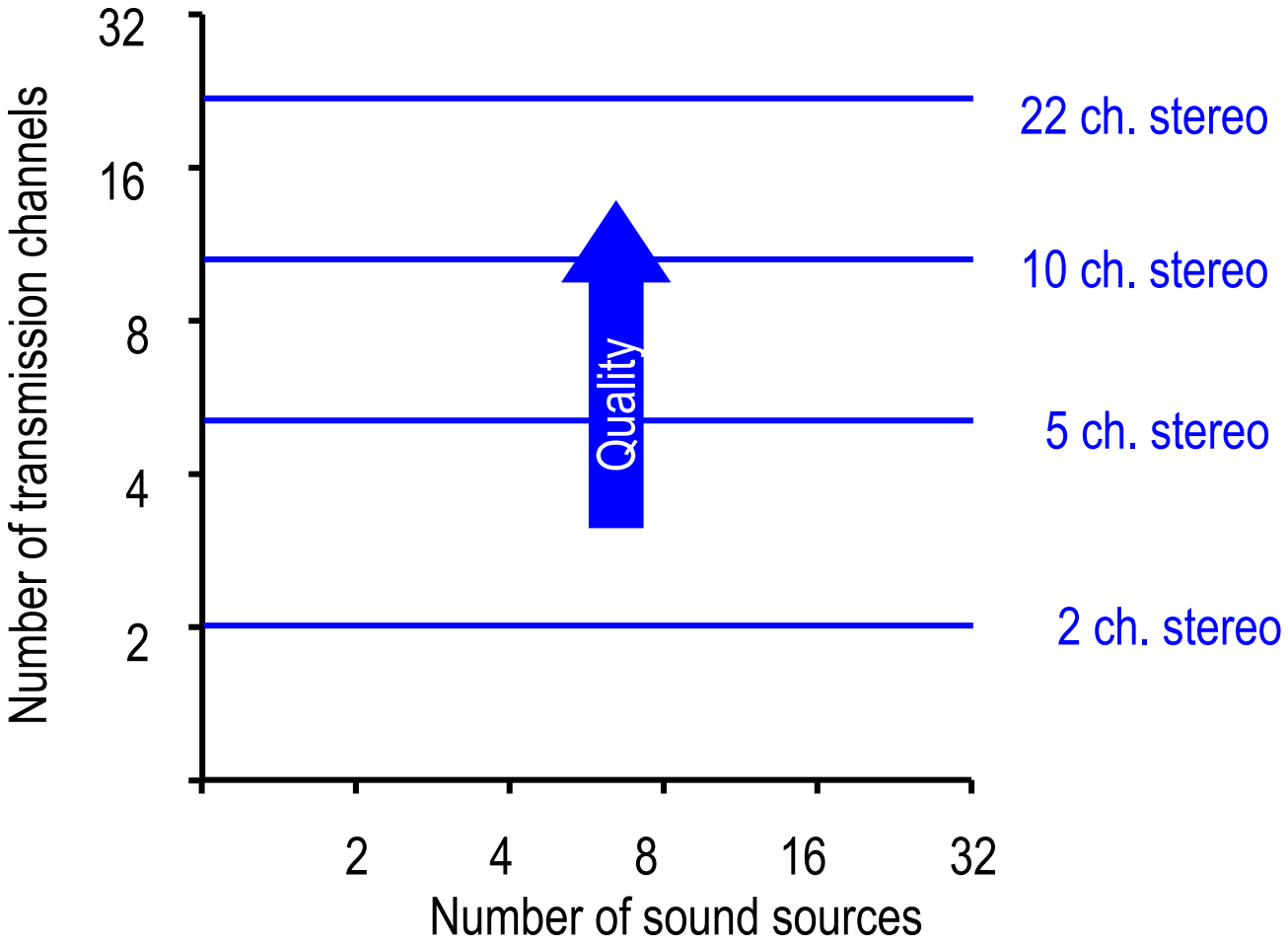
Practical constraints:

- 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

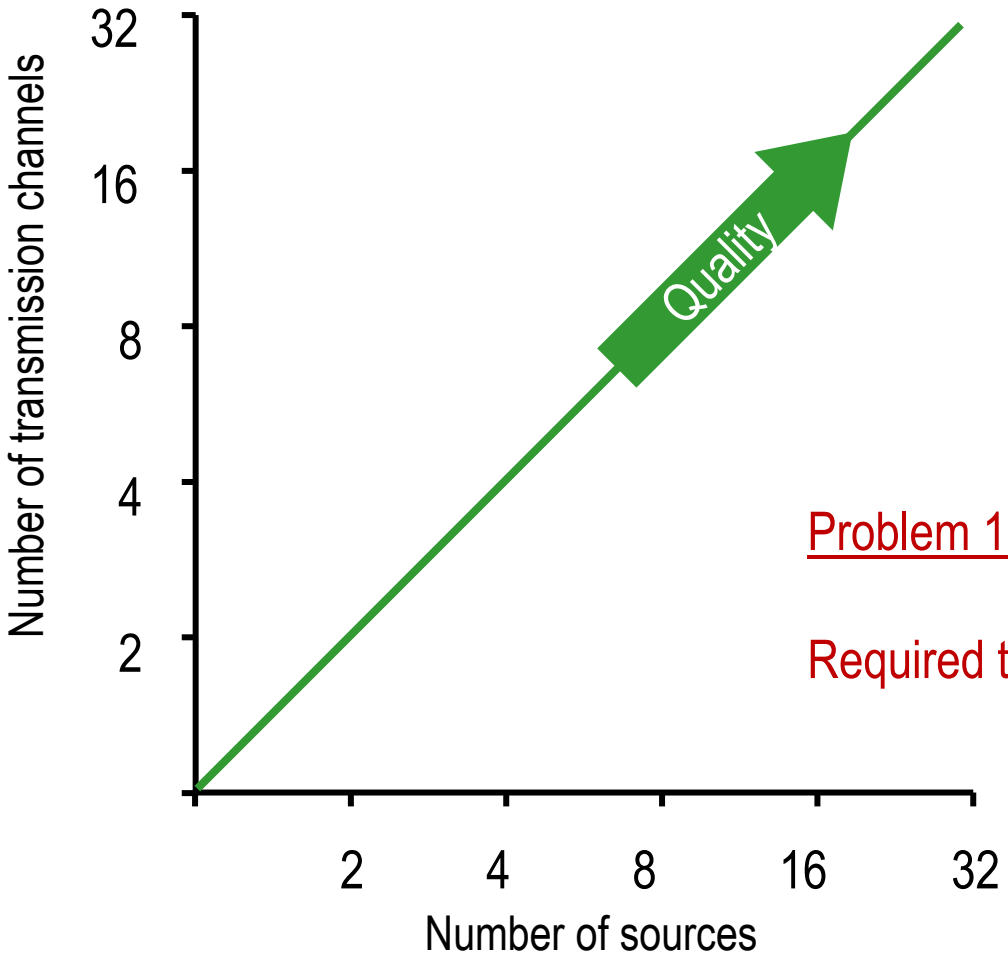
Practical constraints:

- 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

Required transmission capacity with channel-based recording



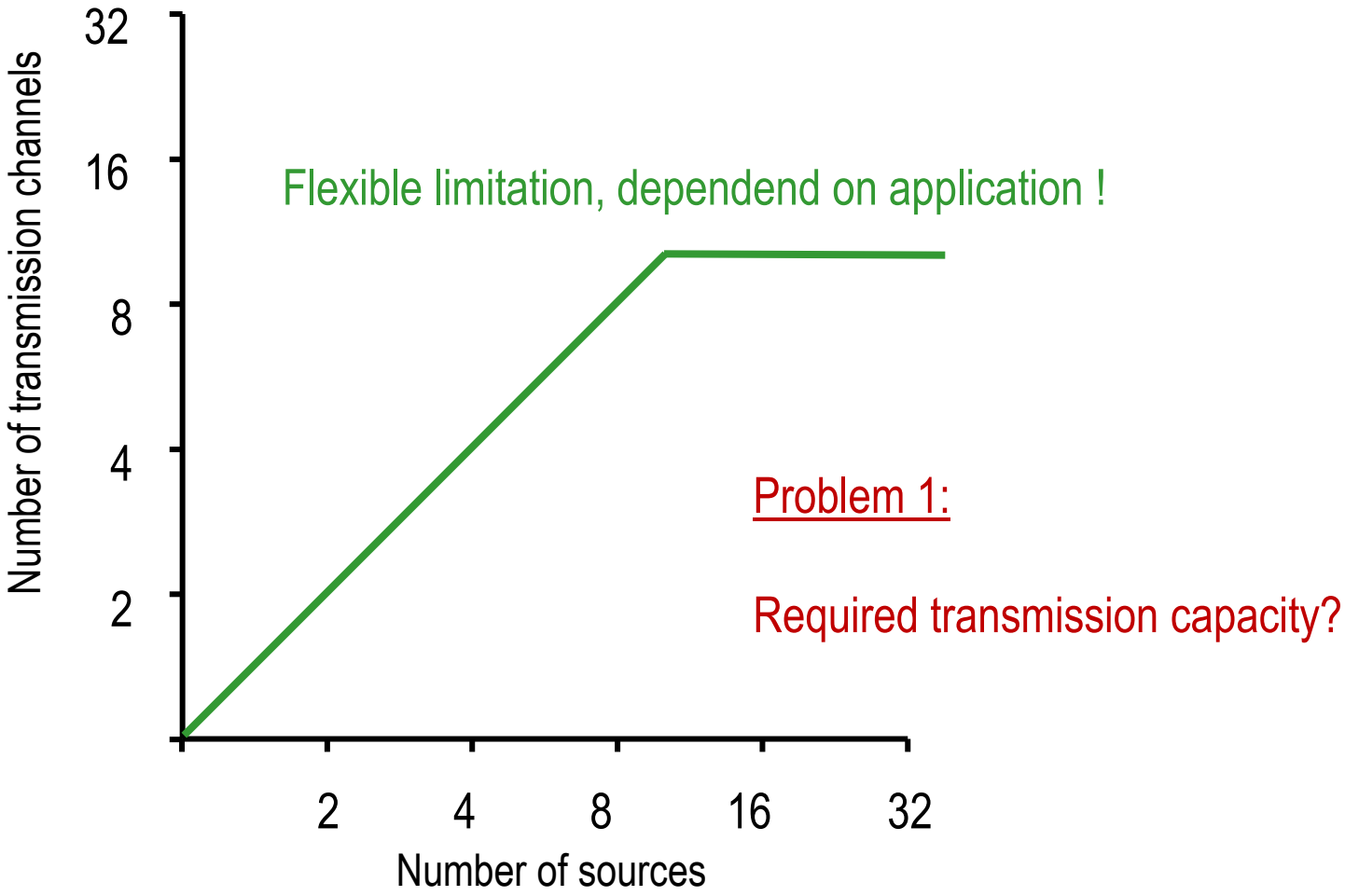
Required transmission capacity with object-based recording



Problem 1:

Required transmission capacity?

Required transmission capacity with object-based recording



WFS: constraints due to object-based recording

Verband Deutscher Tonmeister e.V.

Problem 2:

Imaging of spacious (broad) sound sources?

Problem 3:

Imaging of acoustic environment?

Problem 4:

Imaging of moving sources?

WFS: constraints due to object-based recording

Verband Deutscher Tonmeister e.V.

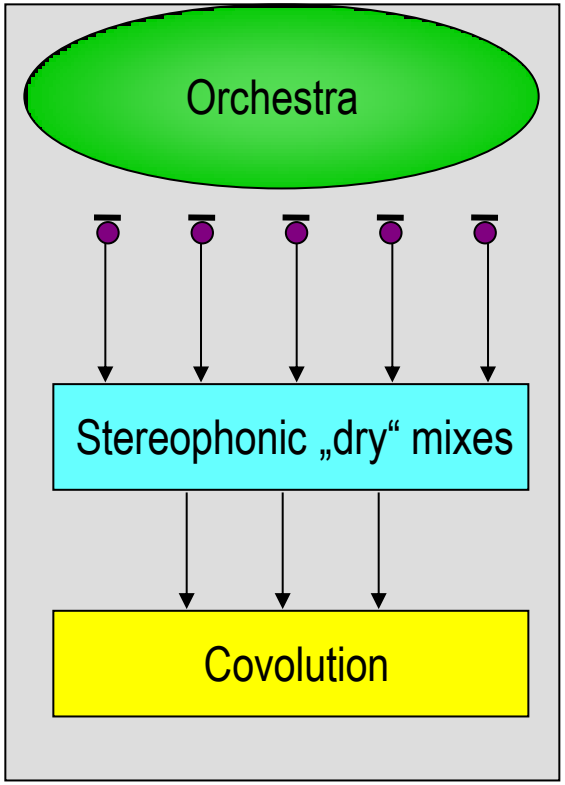
A solution:

Combining the advantages of stereophonic recording methods and WFS techniques by means of „Virtual Panning Spots“ (*Theile et al. 2003*).

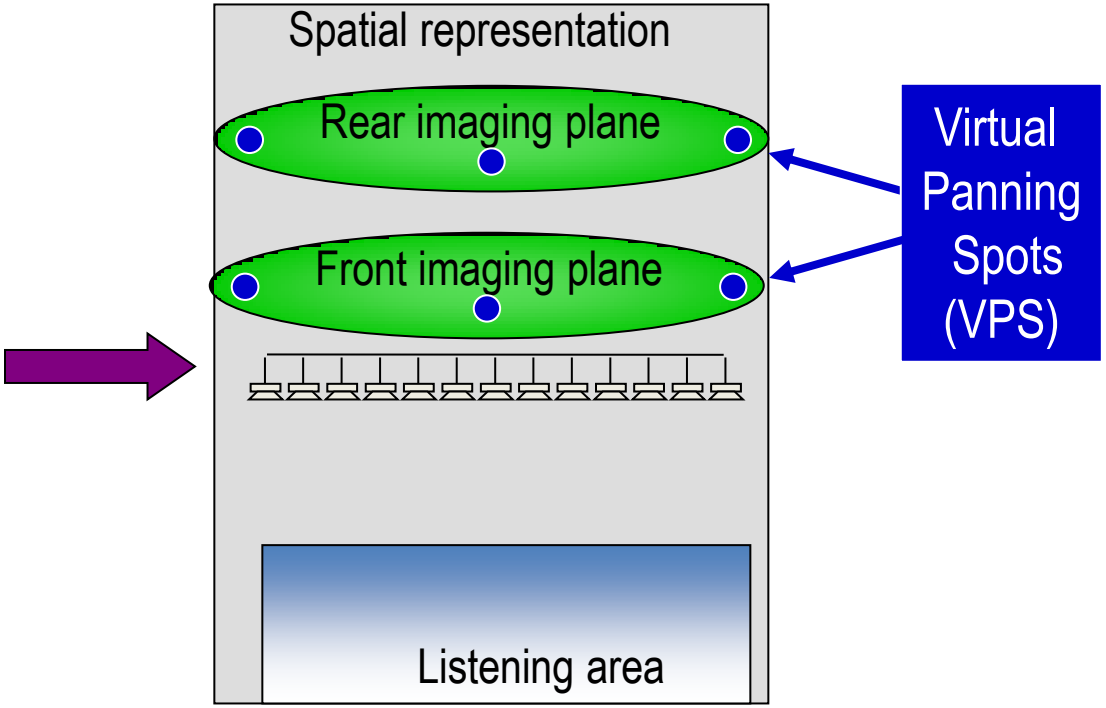
Virtual Panning Spots are selected virtual sources (“virtual loudspeakers”), which produce a stereophonic representation area.

They can be placed at choice in accordance with the recording situation and the intended spatial image.

WFS recording concept: Virtual Panning Spots (VPS)

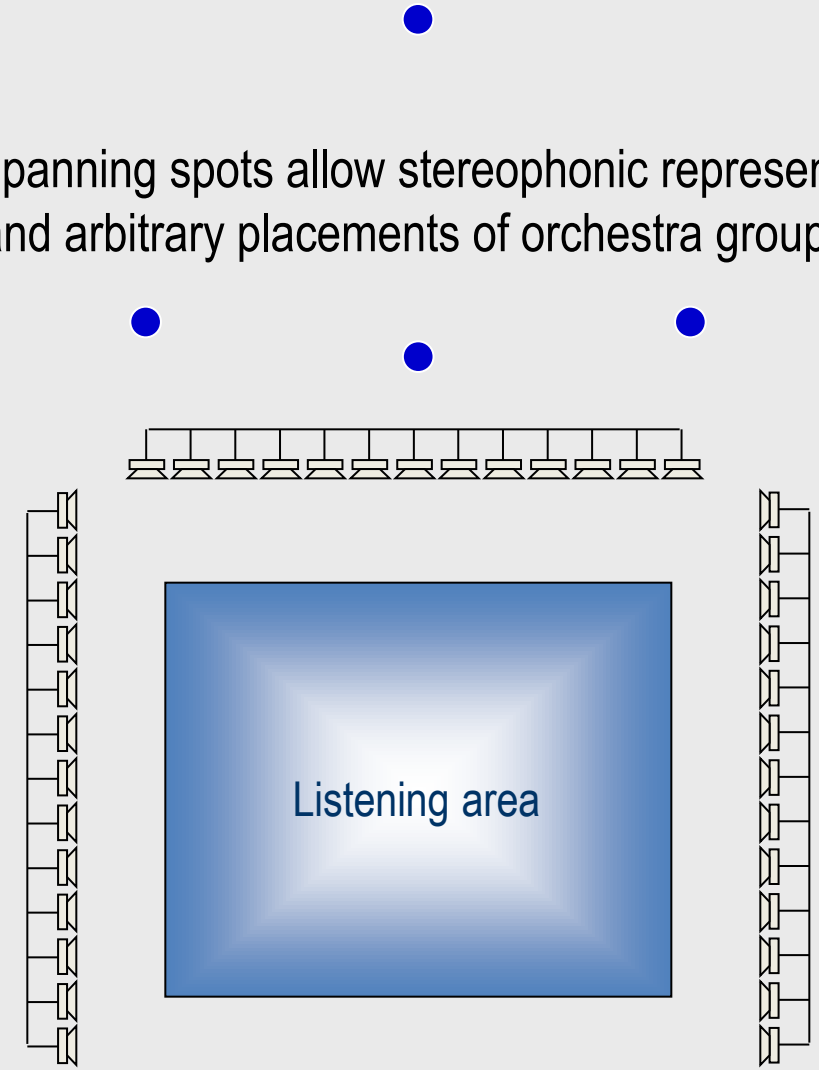


Example:
Creating spatial depth using separate virtual imaging planes

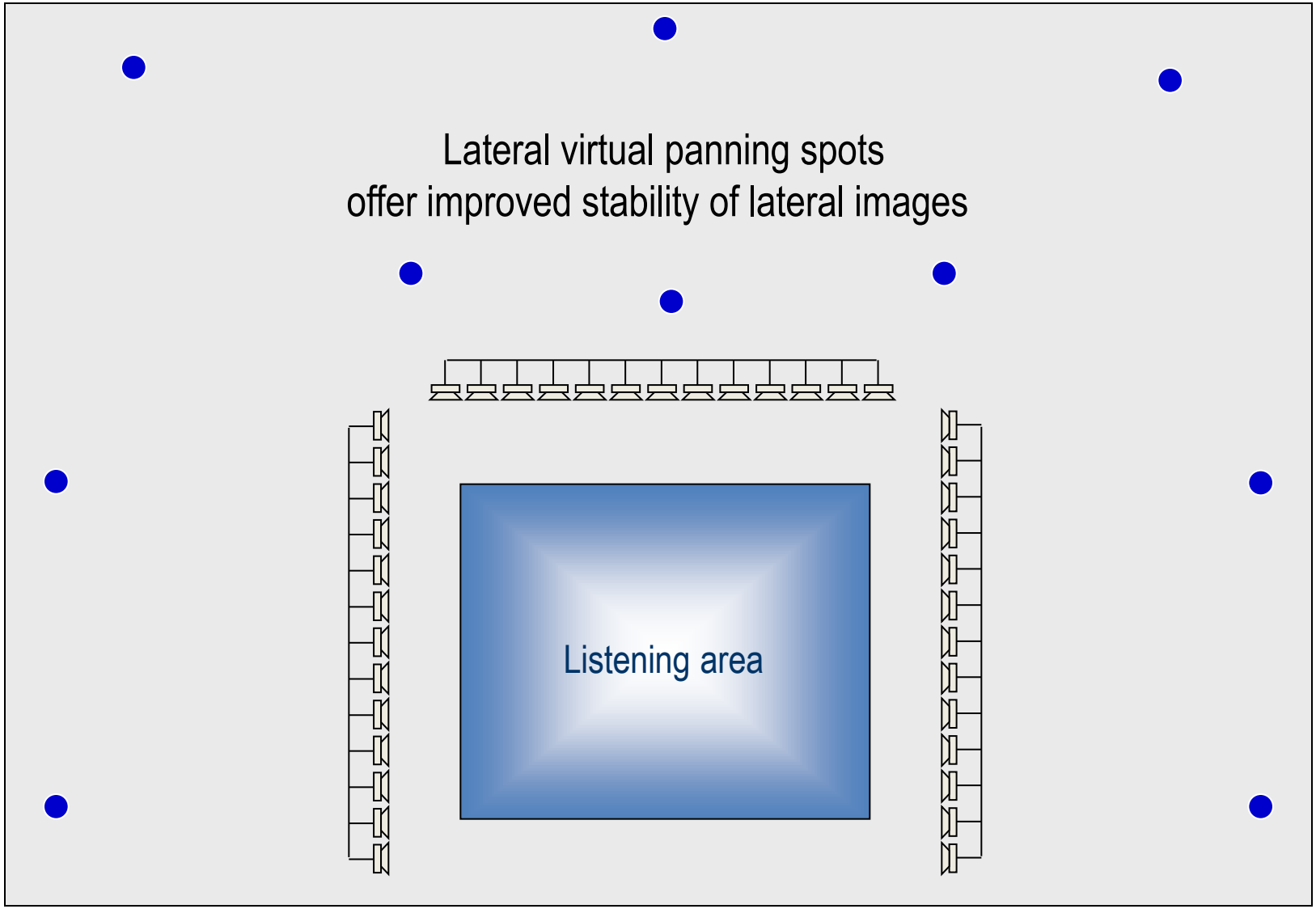


WFS recording concept: Virtual Panning Spots (VPS)

Virtual panning spots allow stereophonic representations
and arbitrary placements of orchestra groups

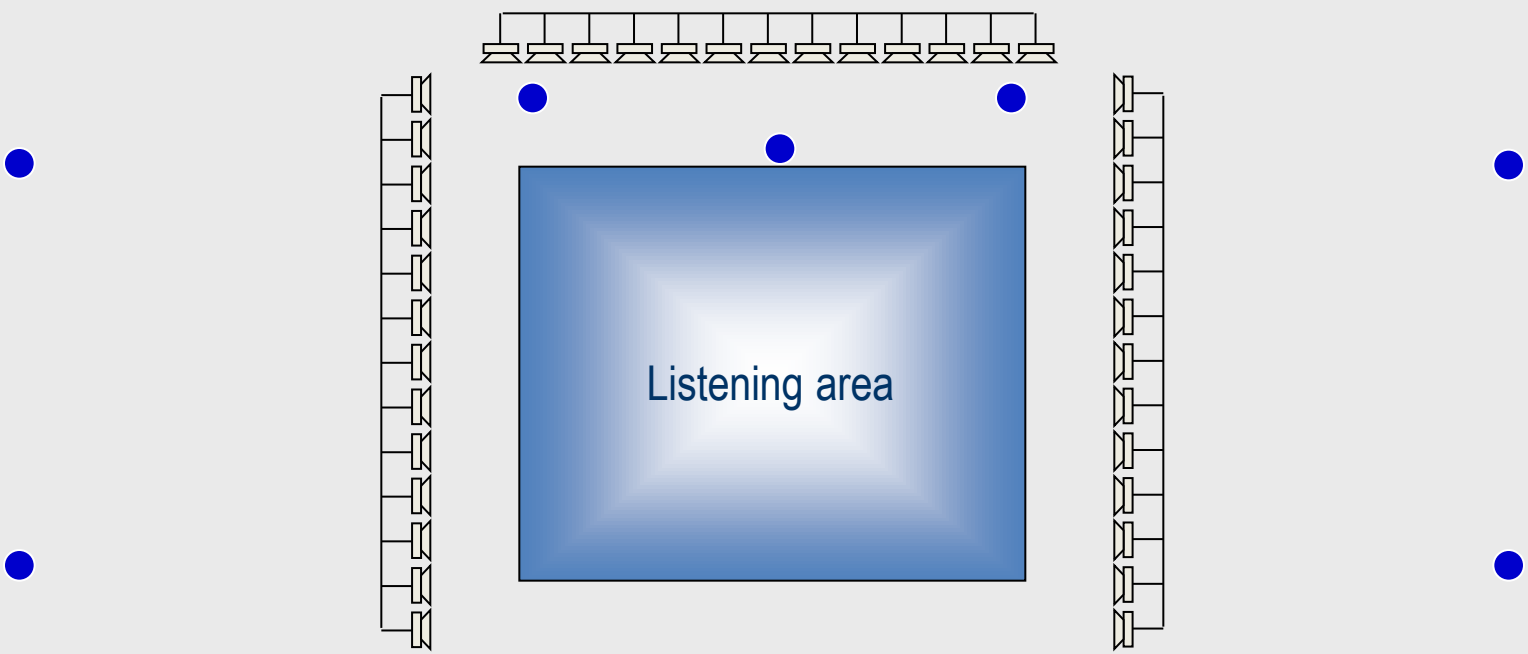


WFS recording concept: Virtual Panning Spots (VPS)



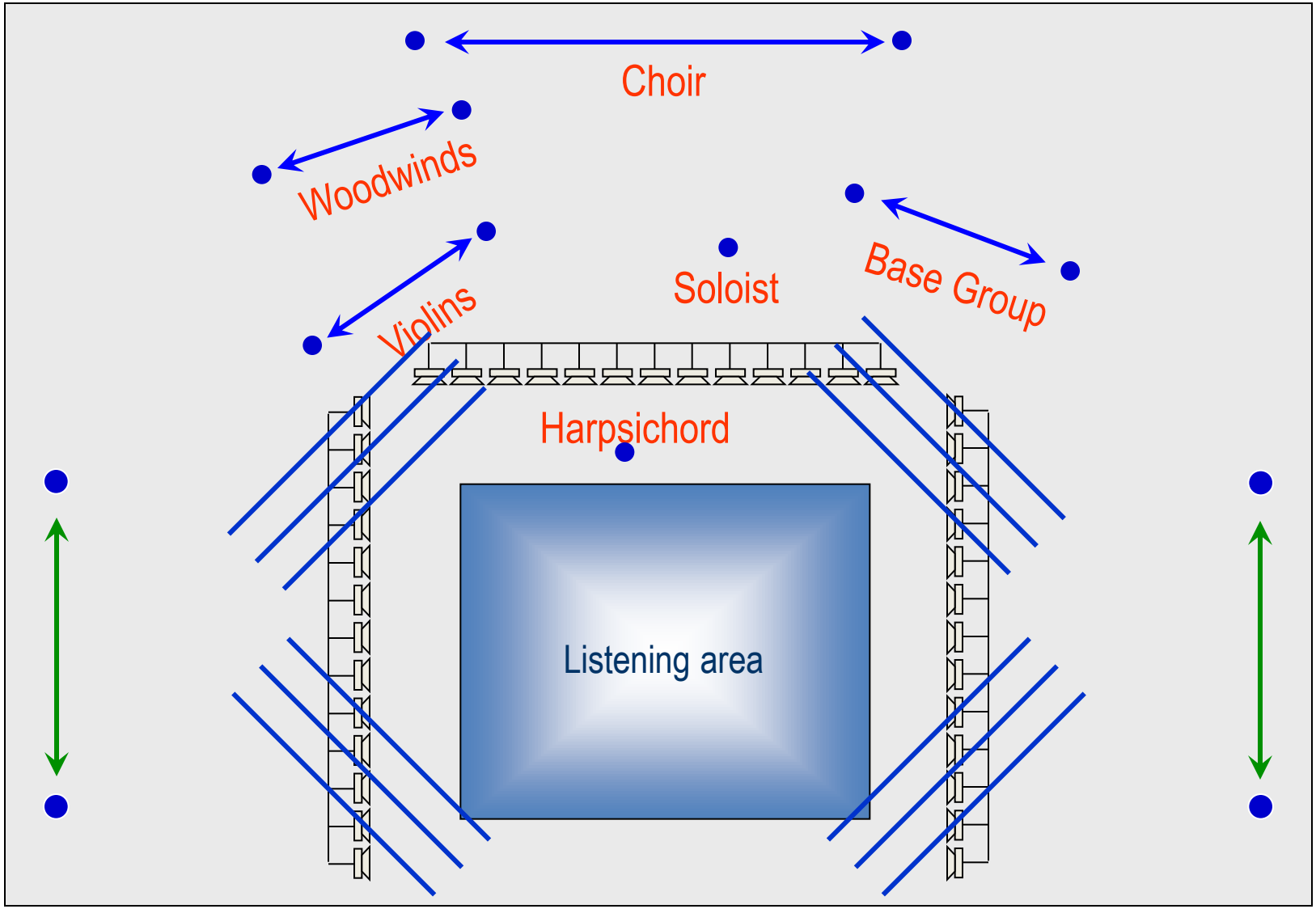
WFS recording concept: Virtual Panning Spots (VPS)

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



Virtual Panning Spots (VPS)
to provide advanced facilities for spatial design

Virtual Panning Spots (VPS)
to provide advanced facilities for spatial design



Channel-based and object-based techniques

Channel-based:

Each produced signal channel is intended to directly drive a loudspeaker in a designated position, e.g. ITU-R775 5.1 & 7.1, Auro 9.1

Object-based:

Each produced signal channel is intended to be rendered at a designated spatial position - independently of the number and location of available loudspeakers, e.g. WFS, HOA, MPEG-4.

Practical constraints:

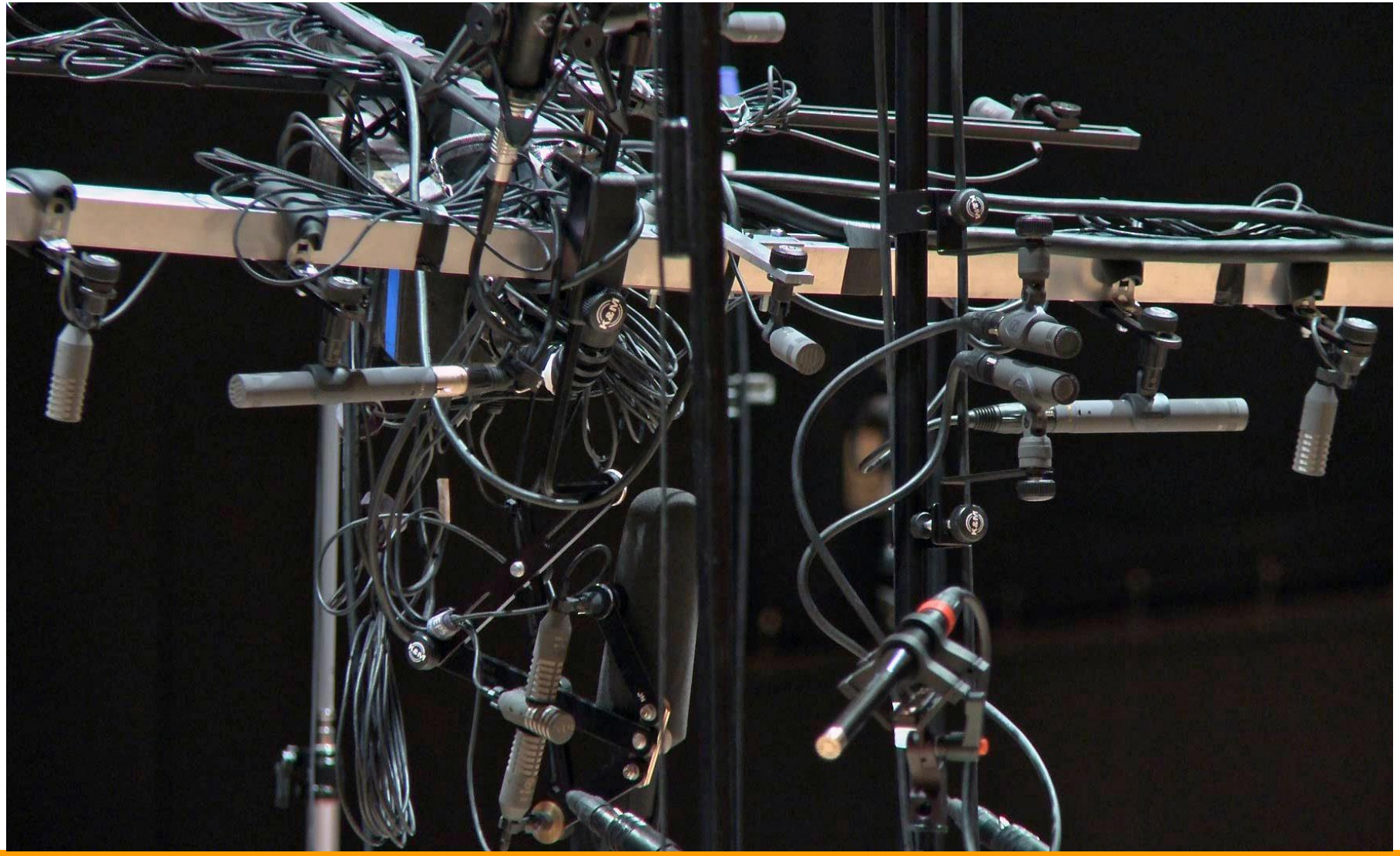
- Spatial aliasing, due to the array discretisation
- Spatial interference, due to limitation of array dimensions
- Sensitivity to room acoustics
- Restriction to the horizontal plane
- Source depending transmission capacity, due to object-based recording
- Extended or moving sources, ambience, due to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

Listening test

Test recordings, ICSA 2011, ETI Detmold

Test recordings, ICSA 2011, ETI Detmold

Detail from the microphone tree carrying a number of configurations…



Test recordings, ICSA 2011, ETI Detmold

Verband Deutscher Tonmeister e.V.

Four 2D systems under test:

1. 2.0 Stereo
2. 7.0 Surround according to ITU-R 775

} channel-based

3. High Order Ambisonics (HOA)
4. Wave Field Synthesis (WFS)

} object-based

Details in *Malte et al., Proceedings ICSA 2011*

Test recordings, ICSA 2011, ETI Detmold

Verband Deutscher Tonmeister e.V.

Test material, selected pieces:

1. Voice and piano
2. Wind instruments, Quintett
3. Choir (moving)
4. Brass ensemble (on stage and behind the listener)
5. Percussion (around the listener)
6. Organ
7. Applause (around the listener)

Test questionnaire, ICSA 2011, ETI Detmold

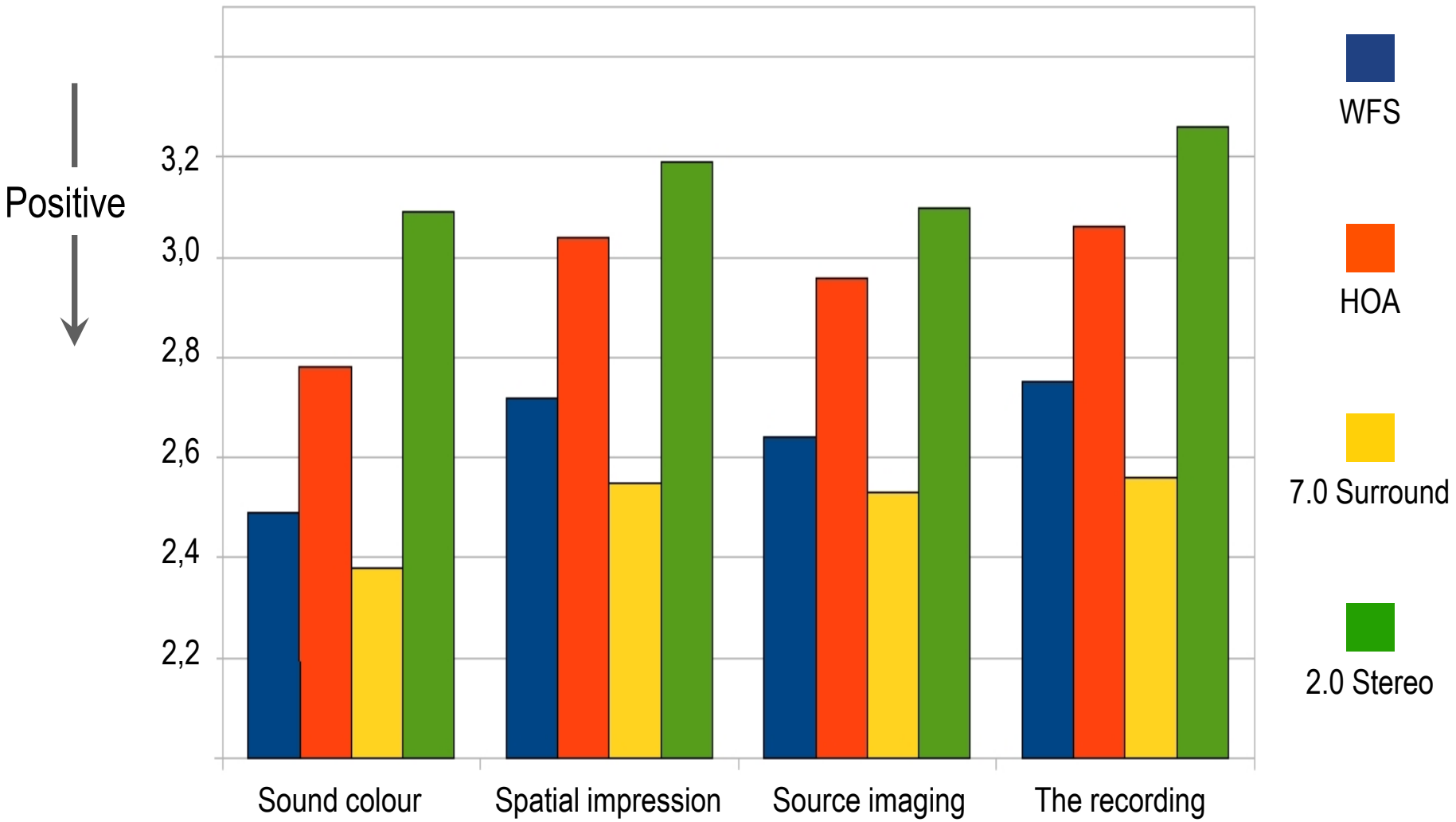
Sound colour							
	1	2	3	4	5		No vote
satisfactory						unsatisfactory	

Spatial presentation							
	1	2	3	4	5		No vote
perfect spatial impression						imperfect spatial impression	
natural envelopment						unnatural envelopment	
no interfering direct sound in surround speakers						interfering direct sound in surround speakers	

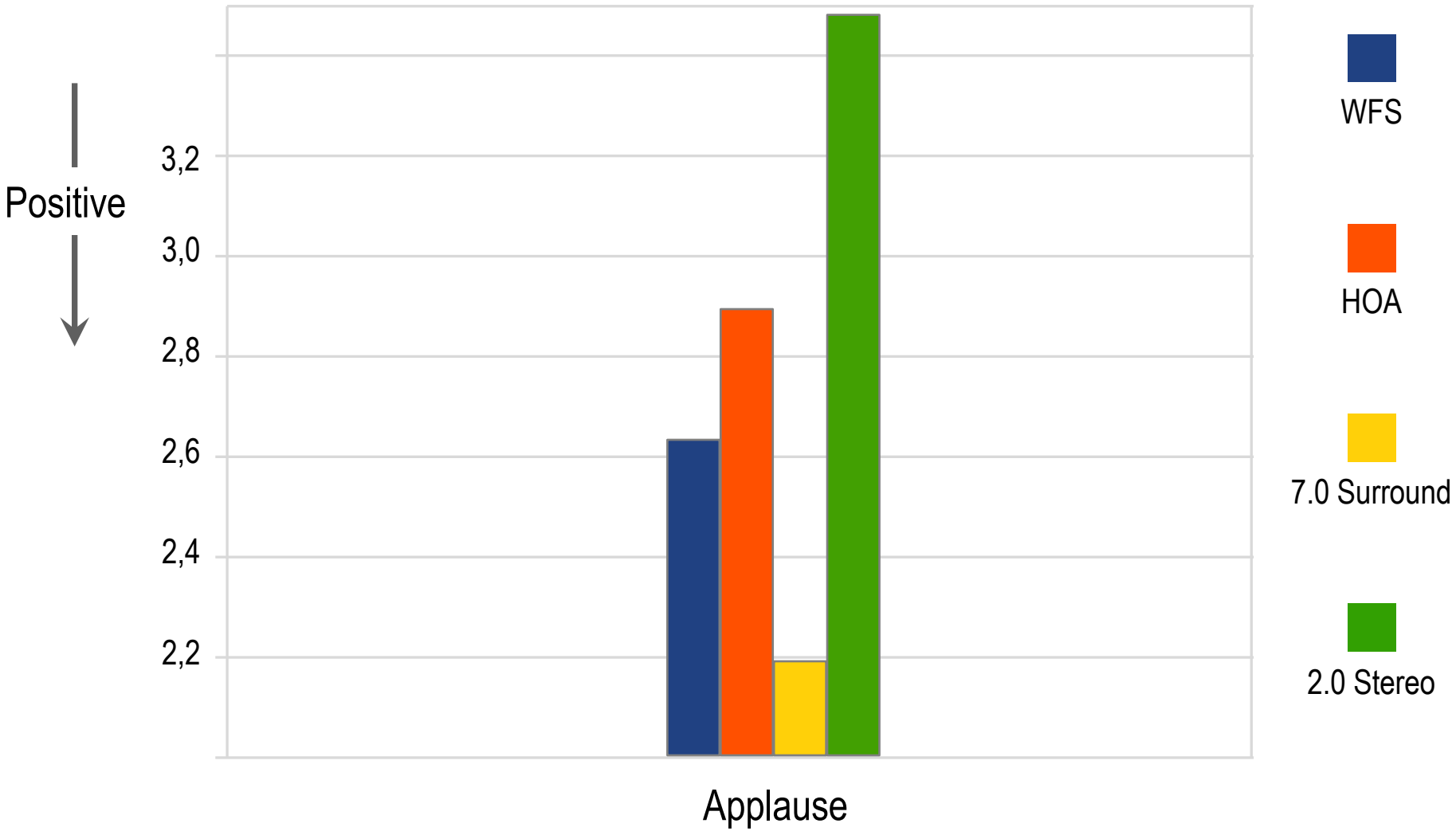
Source imaging							
	1	2	3	4	5		No vote
distant / deep						close / flat	
well-balanced front directions						unbalanced front directions	
well-balanced surround directions						unbalanced surround directions	
precise						blurred	
stable						unstable	

The recording							
	1	2	3	4	5		No vote
sounds natural						sounds unnatural	
sounds good						sounds poor	

Test results, ICSA 2011, ETI Detmold



Test results, ICSA 2011, ETI Detmold



Multichannel stereophony formats – basic characteristics

Verband Deutscher Tonmeister e.V.

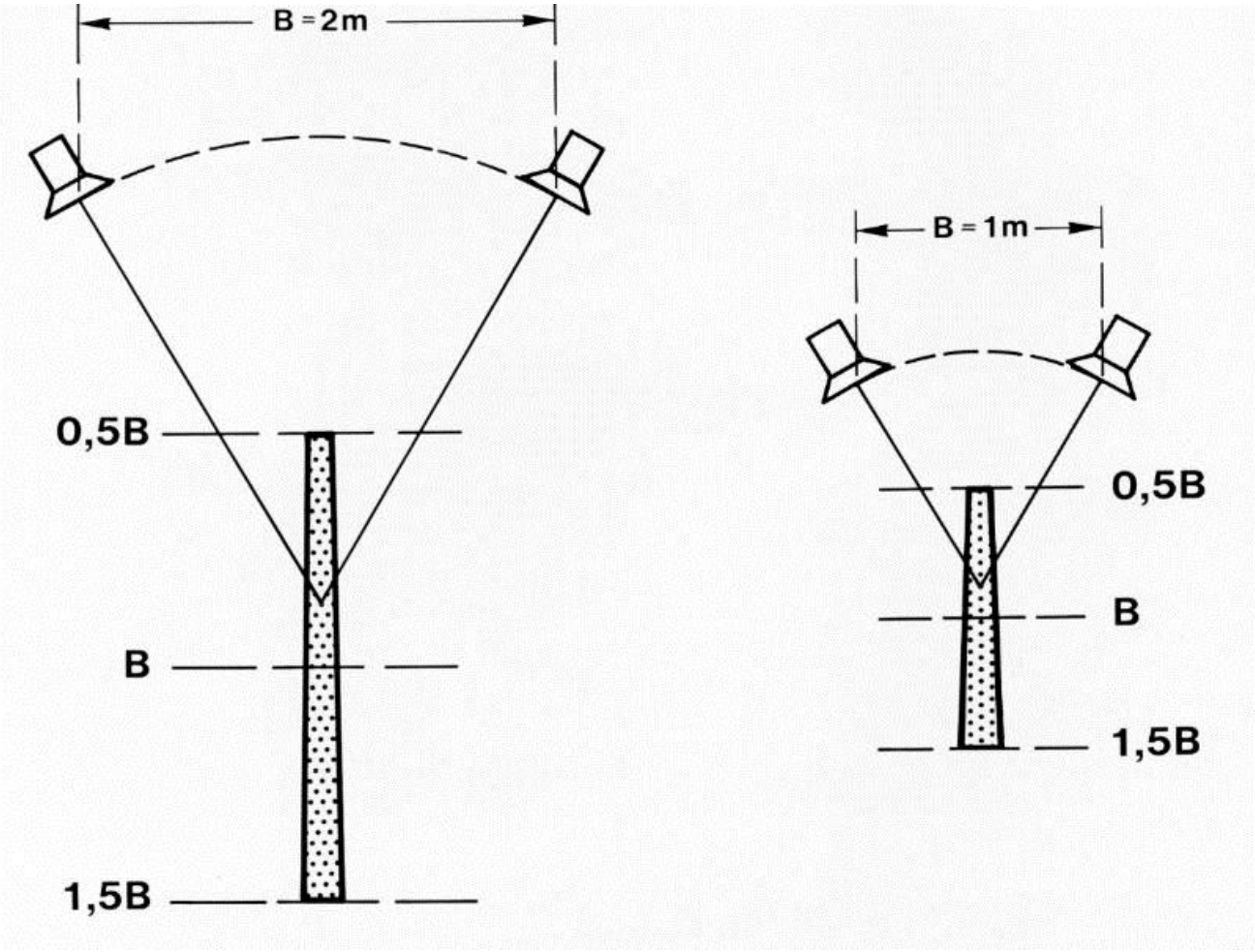
Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

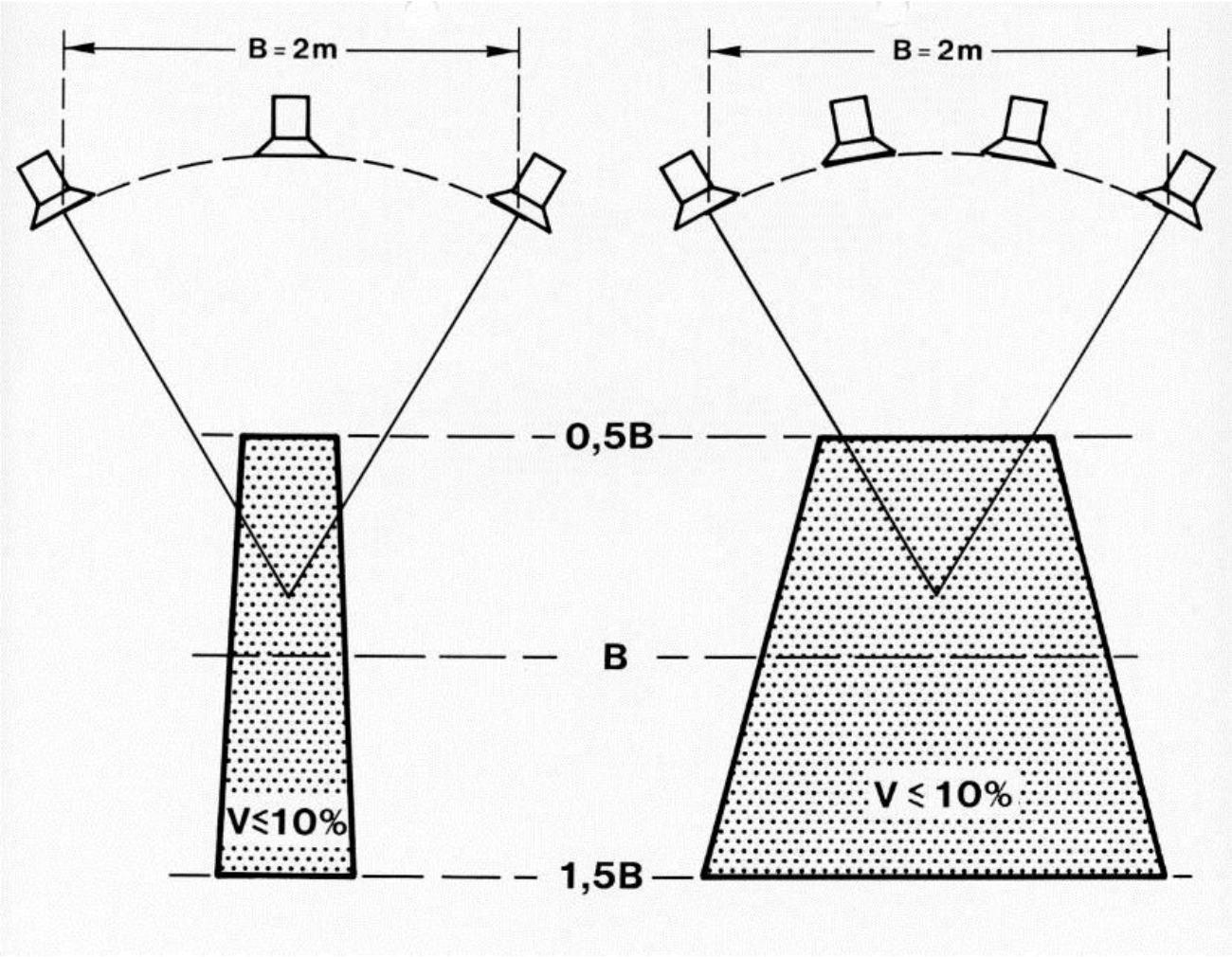
Two-channel stereo

Sweet area

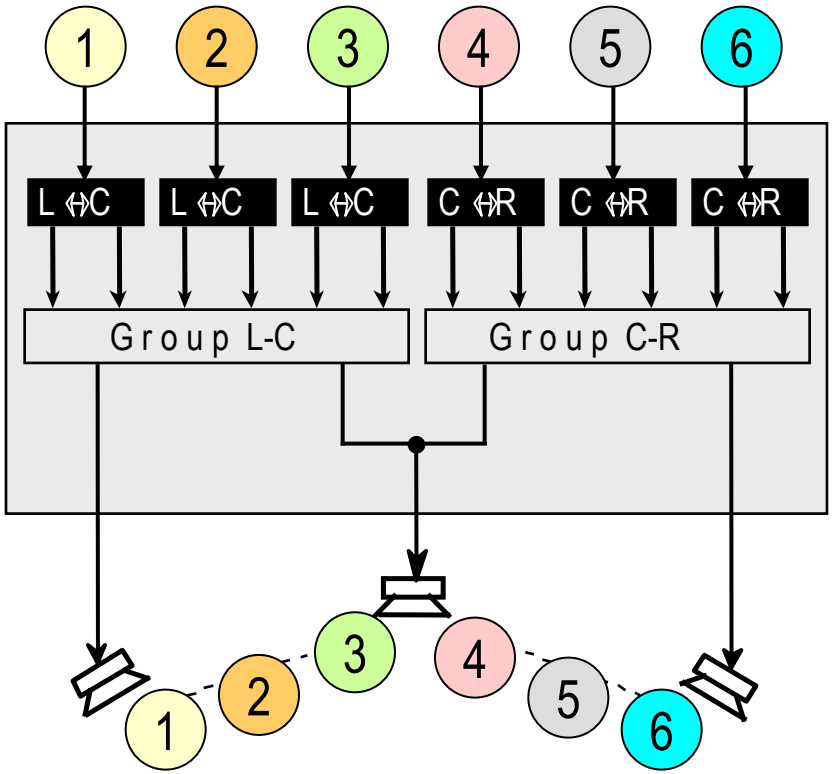


Three- and four-channel stereo

Sweet areas



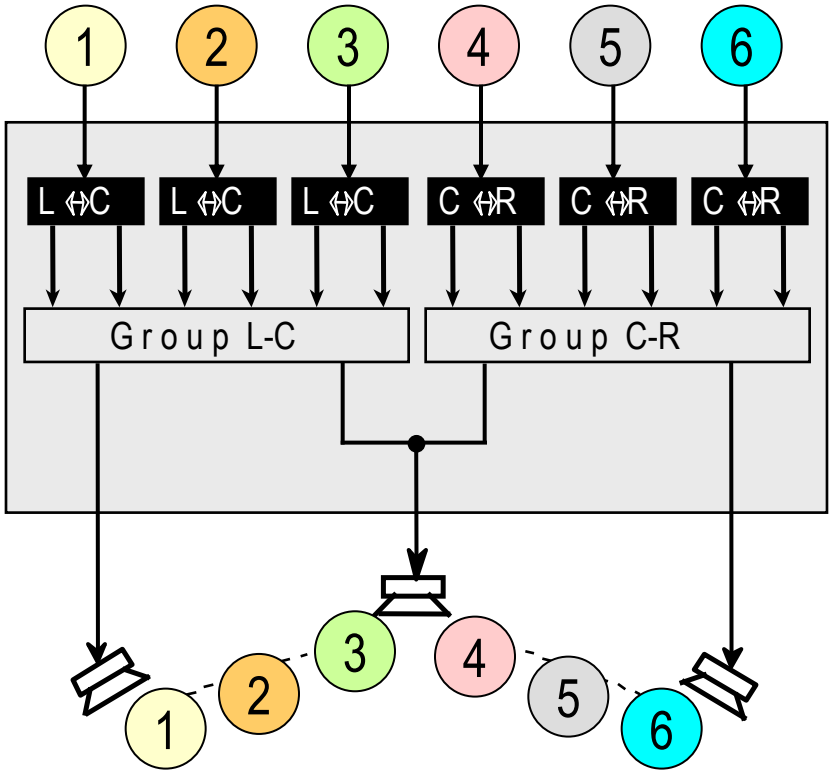
Use of three front channels



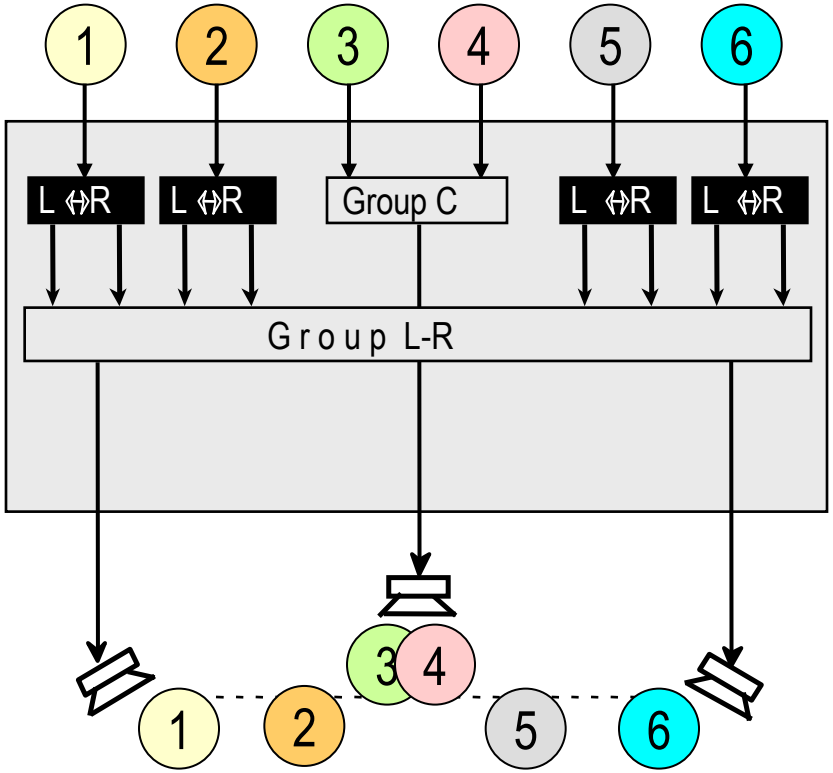
Left-Center-Right Pan

Use of three front channels

Two basic pan methods



Left-Center-Right Pan

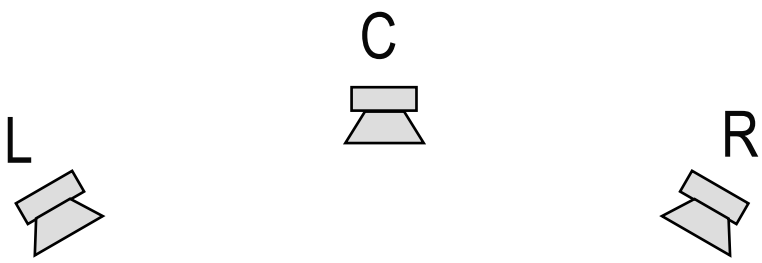


Left-Right-Pan Plus Center

Two vs. three front channels

One stereophonic imaging sector

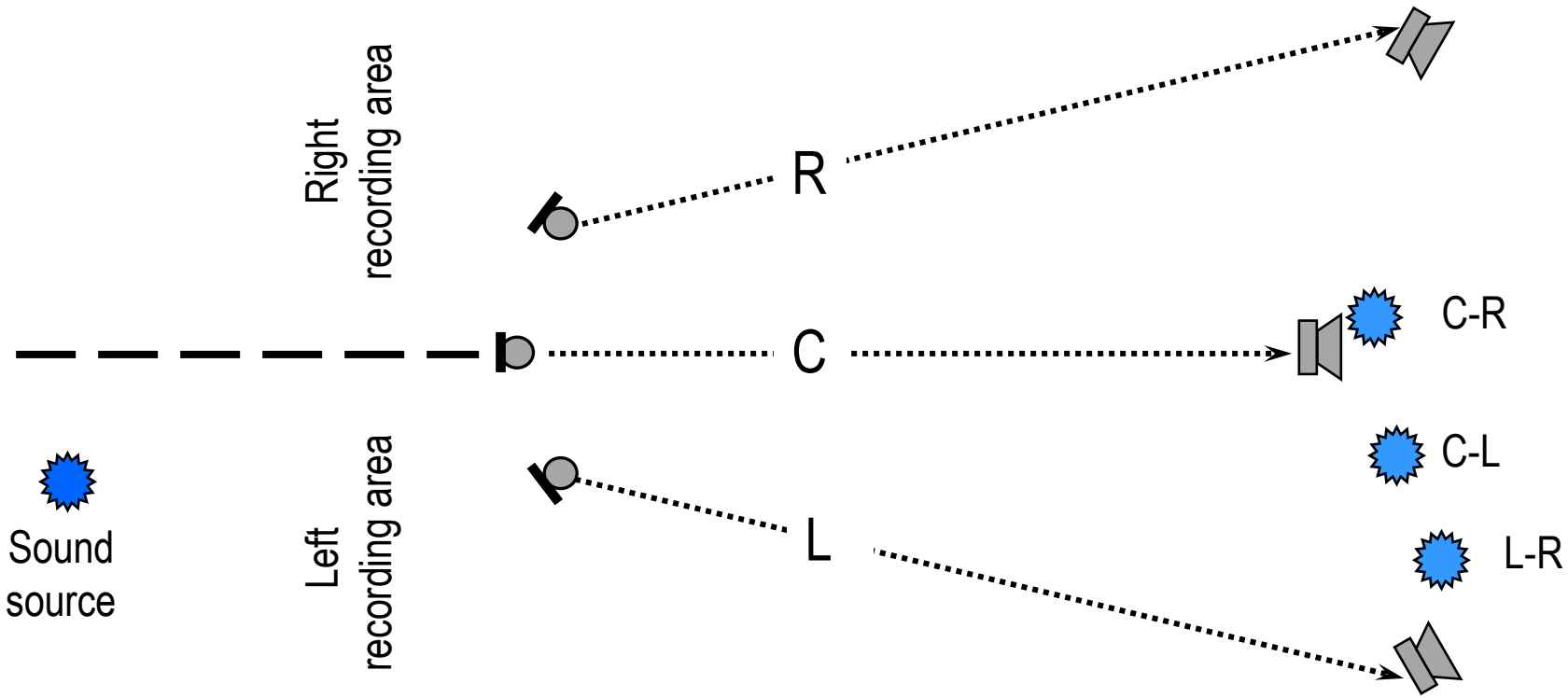
Two independent stereophonic imaging sectors



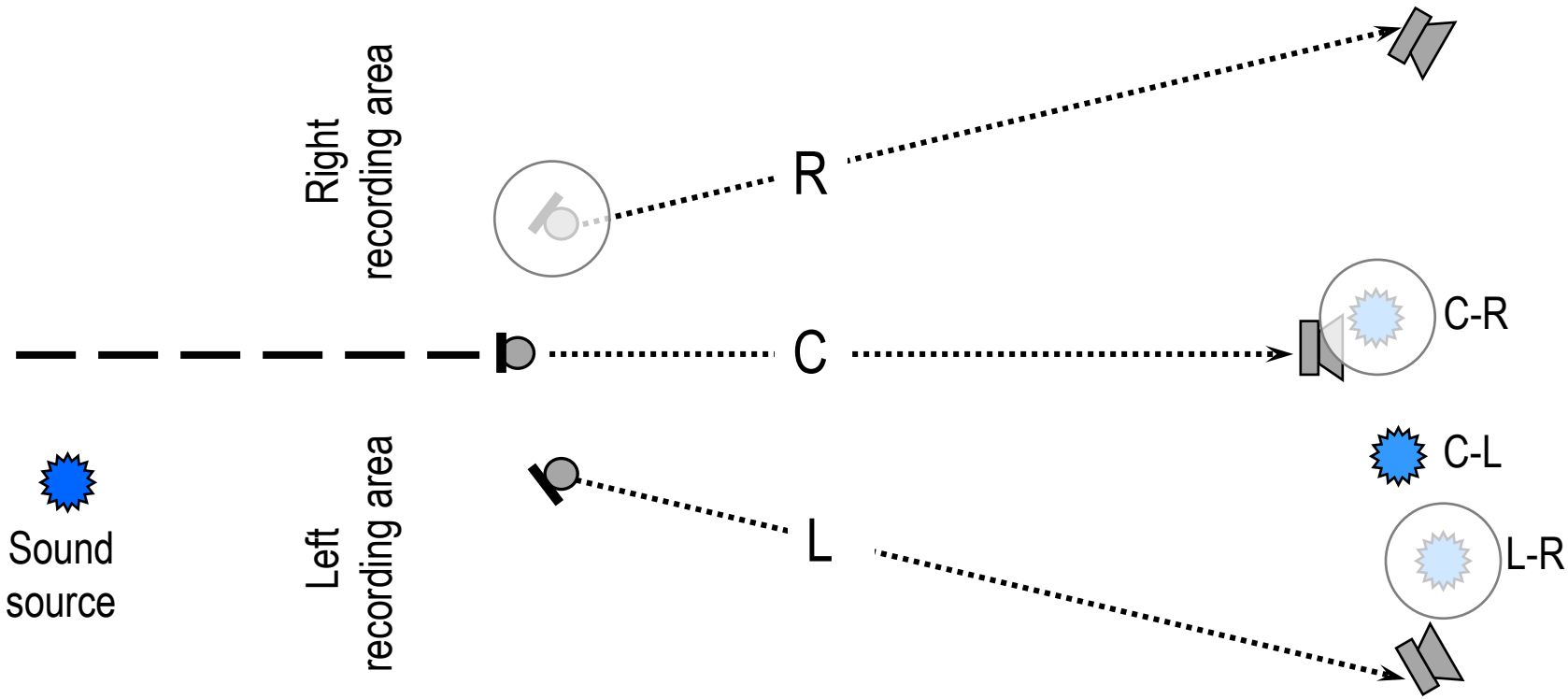
Enlarged listening area

Optimum center image quality (stability, clarity, sound colour)

The interchannel crosstalk problem

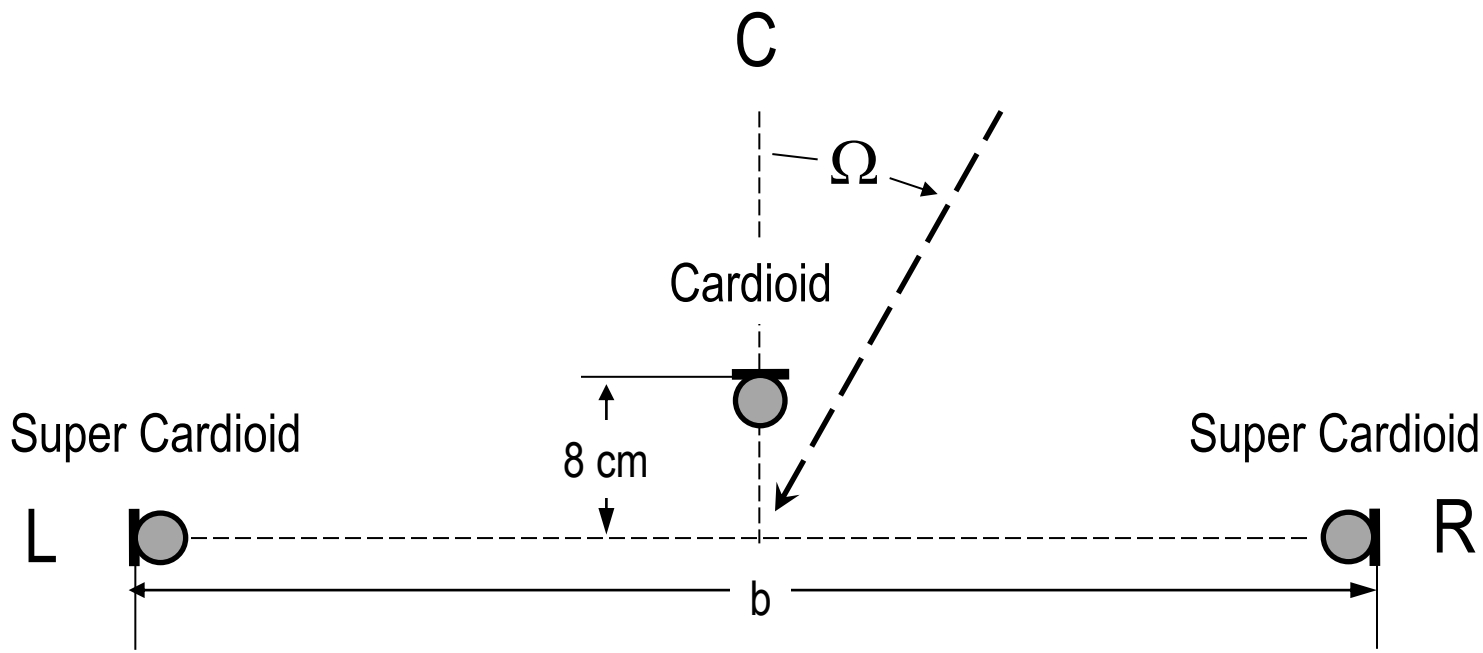


The interchannel crosstalk problem



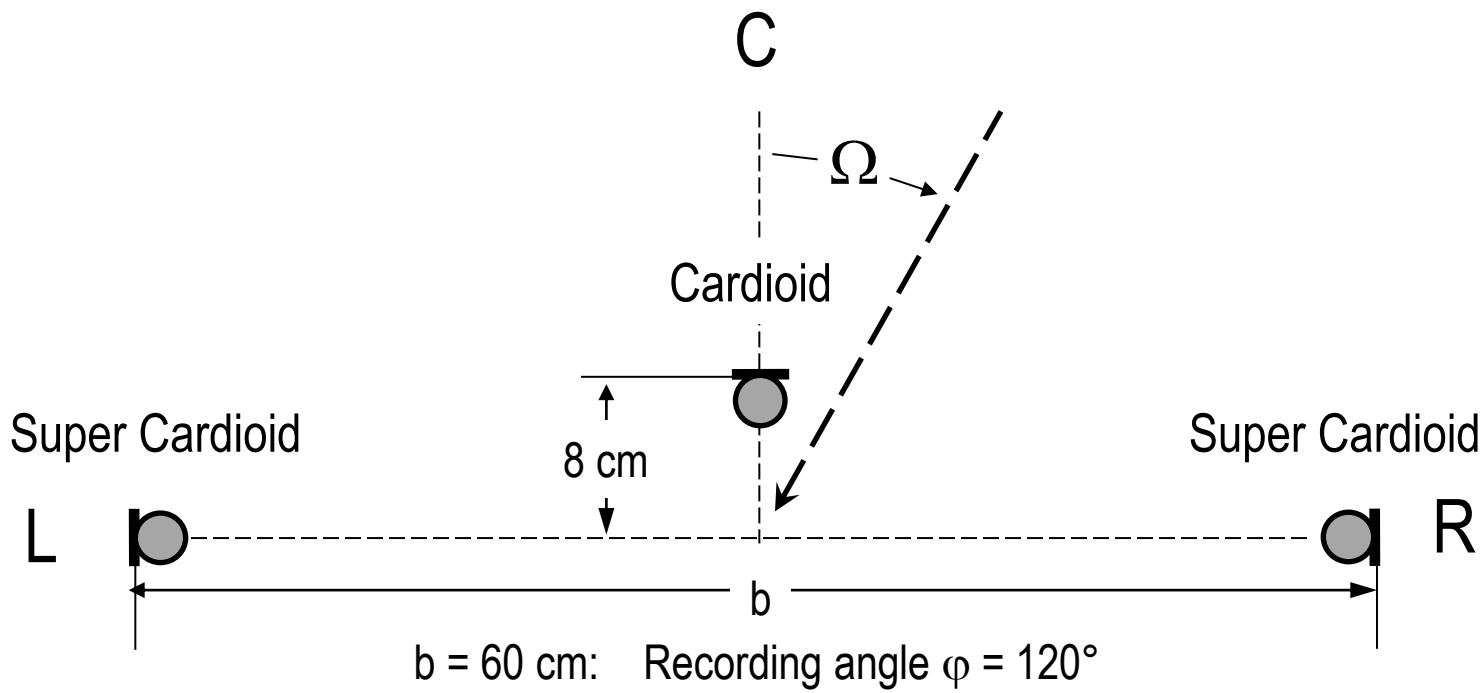
Desirable channel separation > 10 dB

Optimised Cardioid Triangle (OCT)
(*Theile 2001*)



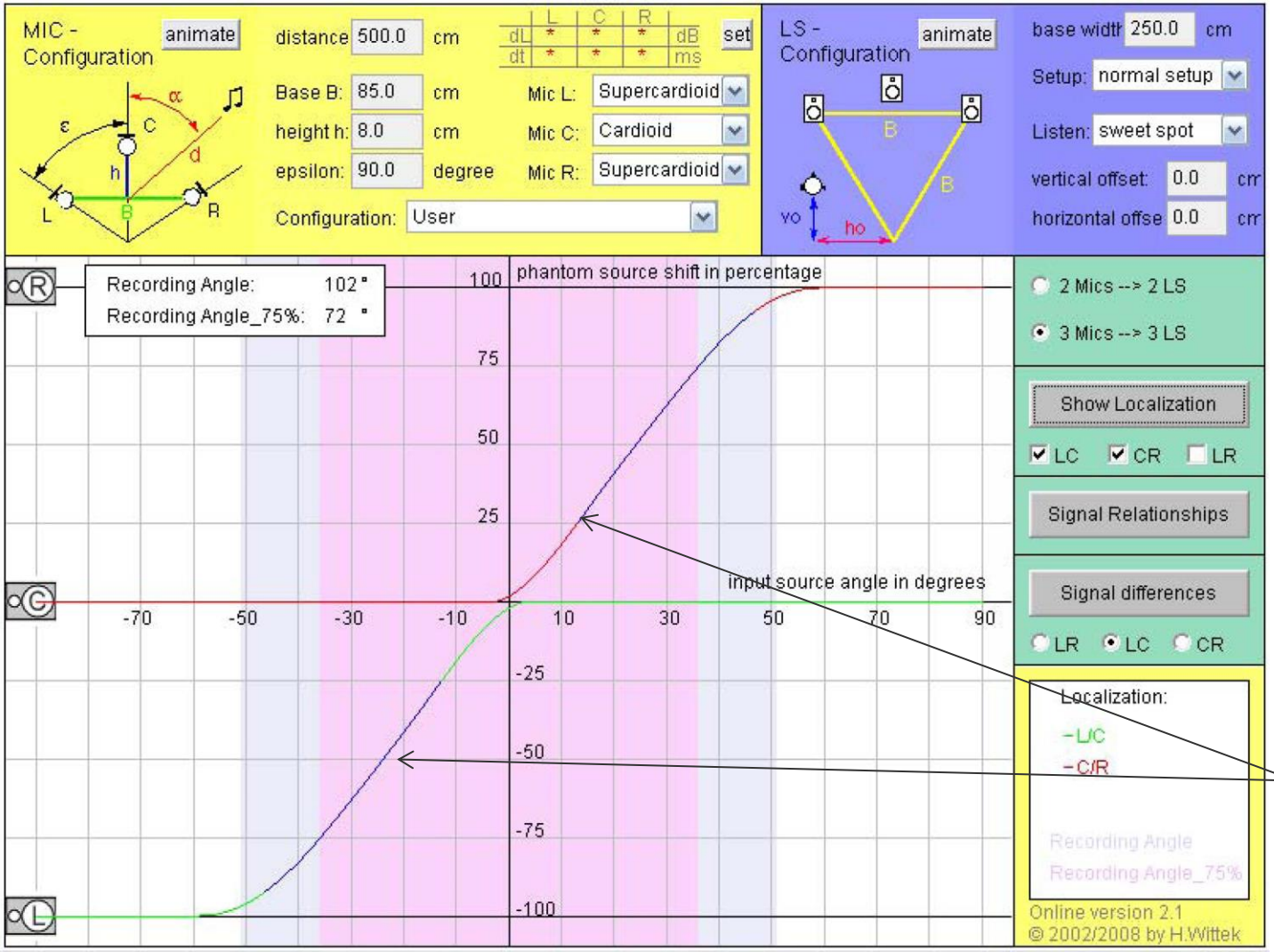
Ω = 0°	- 10 dB	0 dB	- 10 dB
Ω = 30°	- 18 dB	- 1 dB	- 3 dB
Ω = 45°	- 18 dB	- 2 dB	- 2 dB

Optimised Cardioid Triangle (OCT)
(*Theile 2001*)



	Channel crosstalk	Perceived angle
Ω = 0°	- 10 dB	0°
Ω = 30°	- 18 dB	17°
Ω = 45°	- 18 dB	24°

Calculation of localization curves



[Image Assistant](#)
interactive JAVA Applet
[*Wittek*]

L-C-R localization curve:
OCT
b = 85 cm
h = 8 cm

Multichannel stereophony formats – basic characteristics

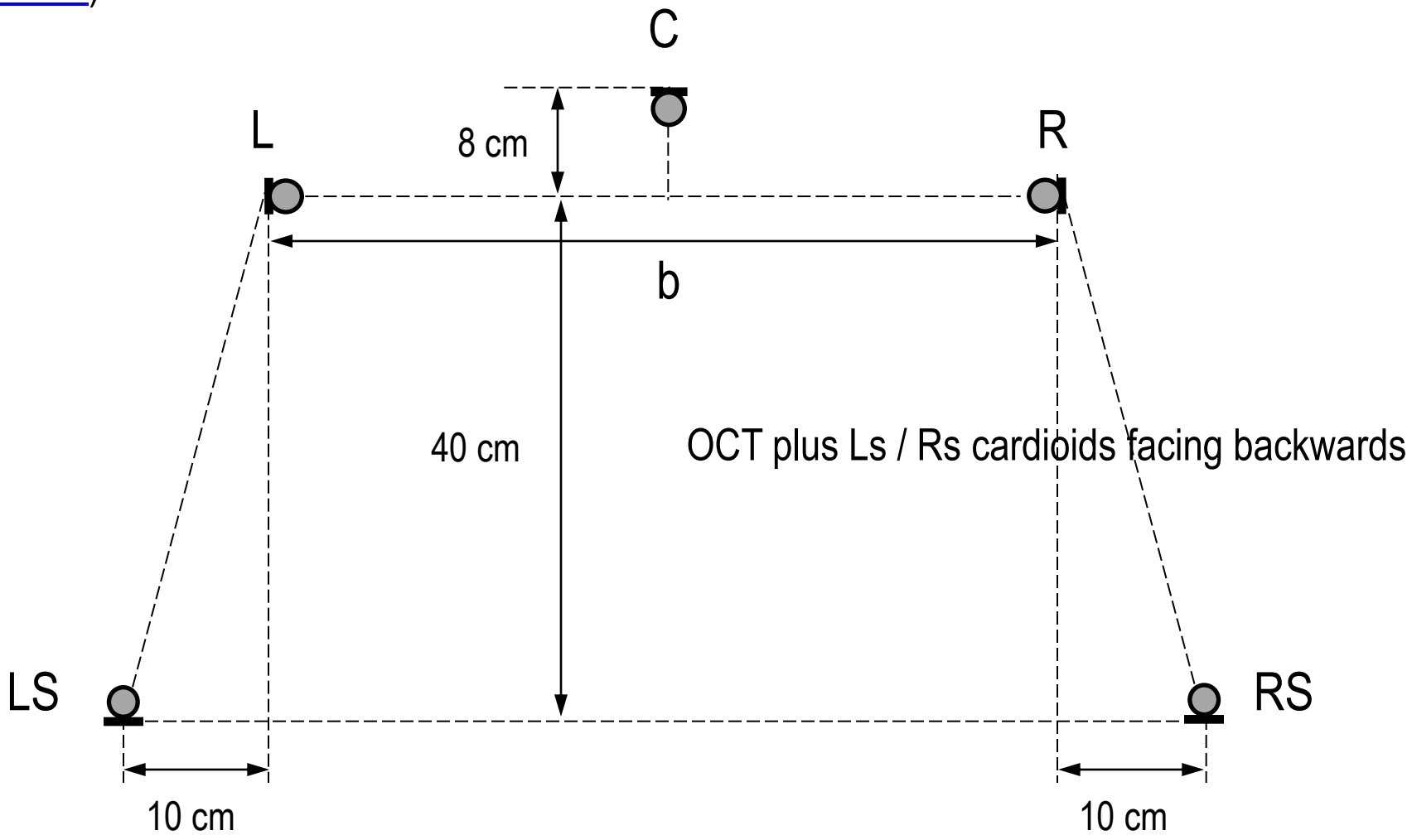
Verband Deutscher Tonmeister e.V.

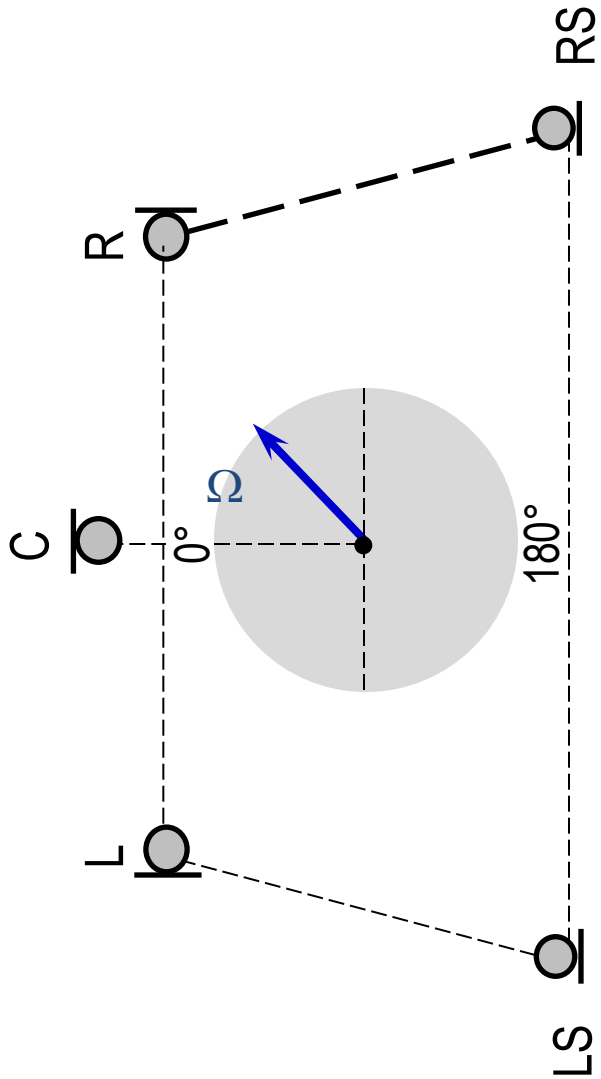
Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

Main microphone OCT Surround (3/2-Stereo setup)
(*Theile 2001*)



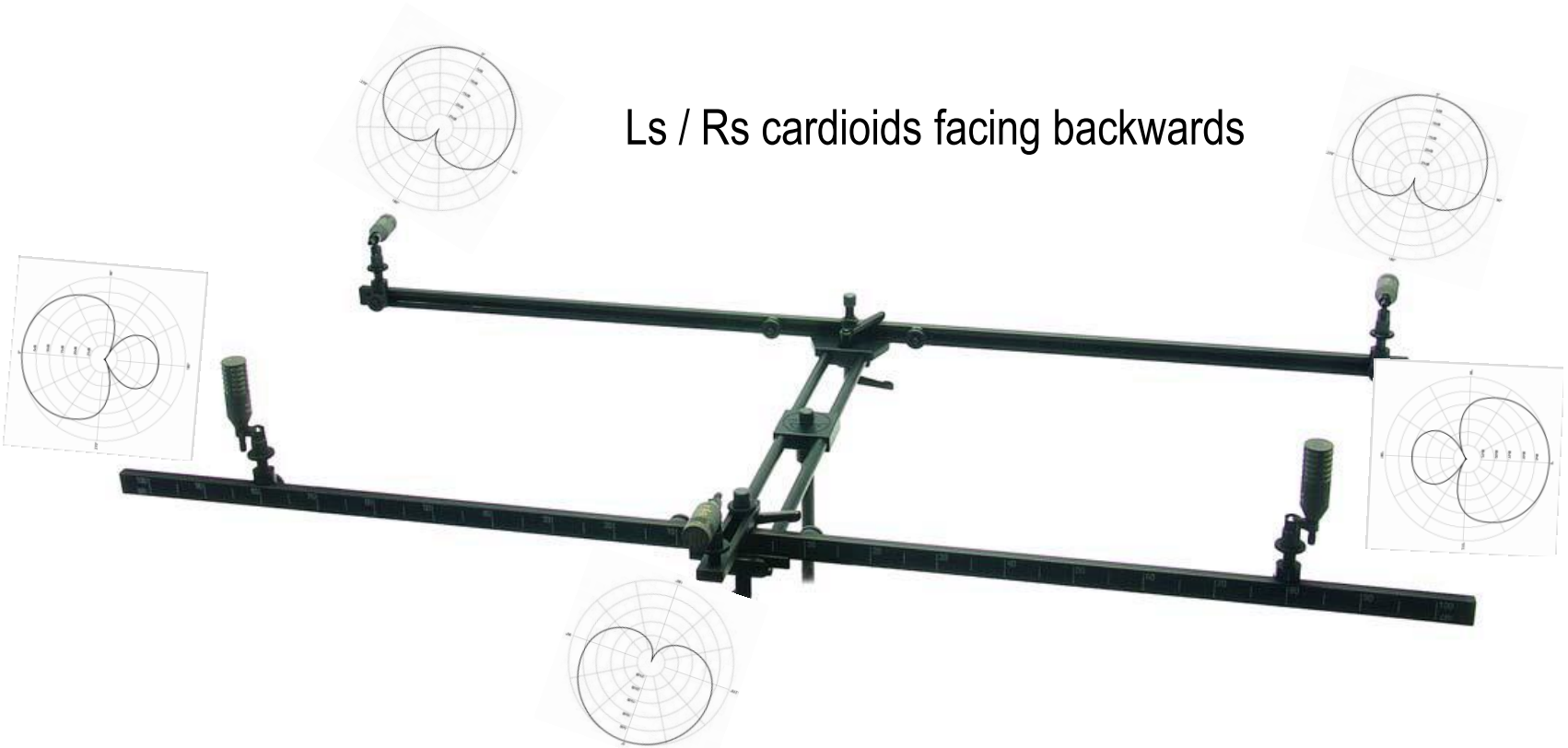


OCT-Surround at the critical distance

Level of direct sound
in relation to the level of indirect sound
(same diffuse sound level in each channel)

Ω	L	C	R	RS	LS
0°	- 4	+ 5	- 4	< -20	< -20
45°	< -20	+ 3	+ 3	- 10	- 10
90°	- 6	- 1	+ 5	- 1	- 1
135°	< -20	- 10	+ 3	+ 3	+ 3
180°	- 4	< -20	- 4	+ 5	+ 5

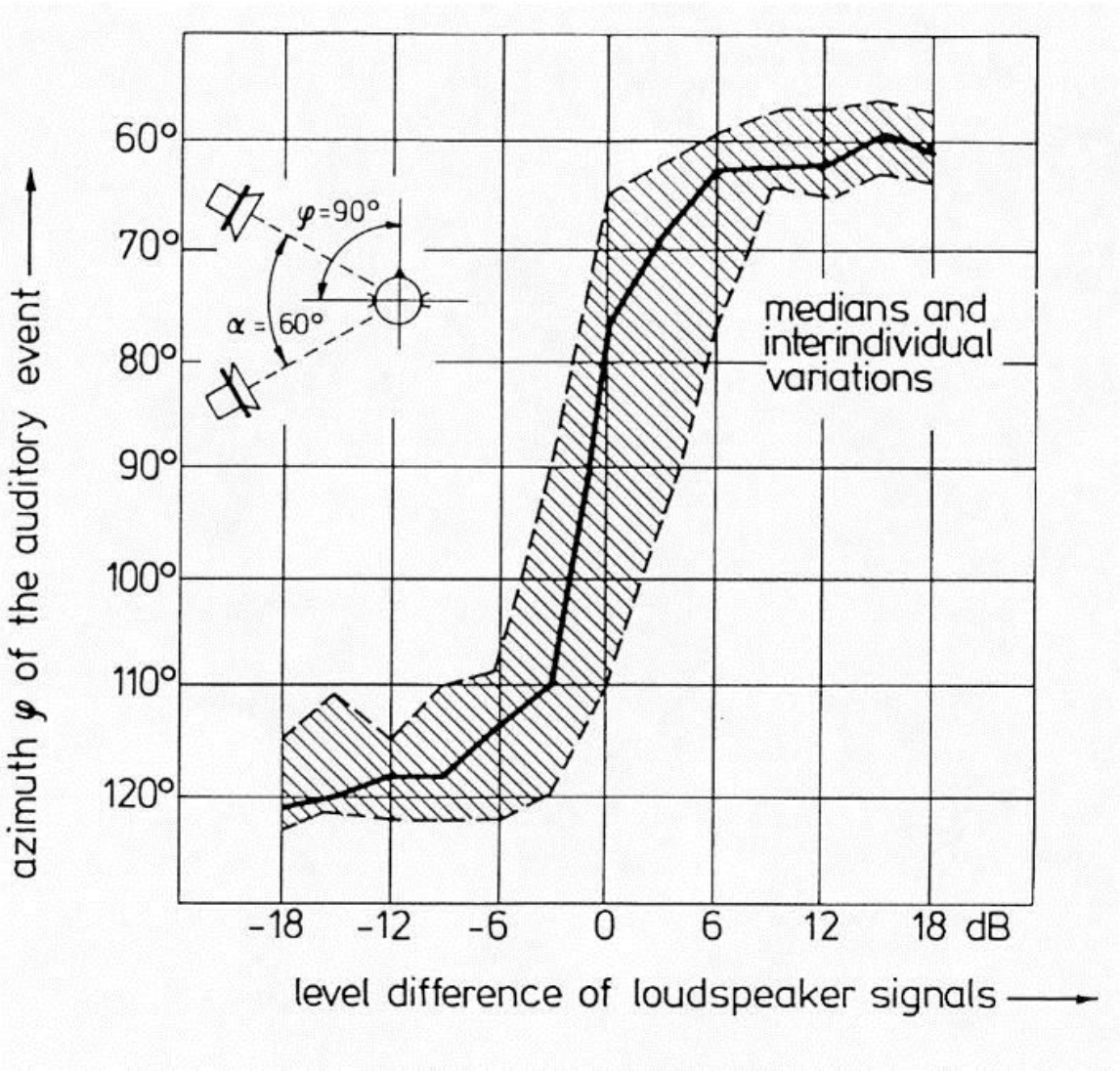
Main microphone OCT Surround (3/2-Stereo setup)
(*Theile 2001*)



Practical setup (example Schoeps)
3 x CCM4, 2 x CCM41, 2 x MAB 1000, 1 x CB/MAB

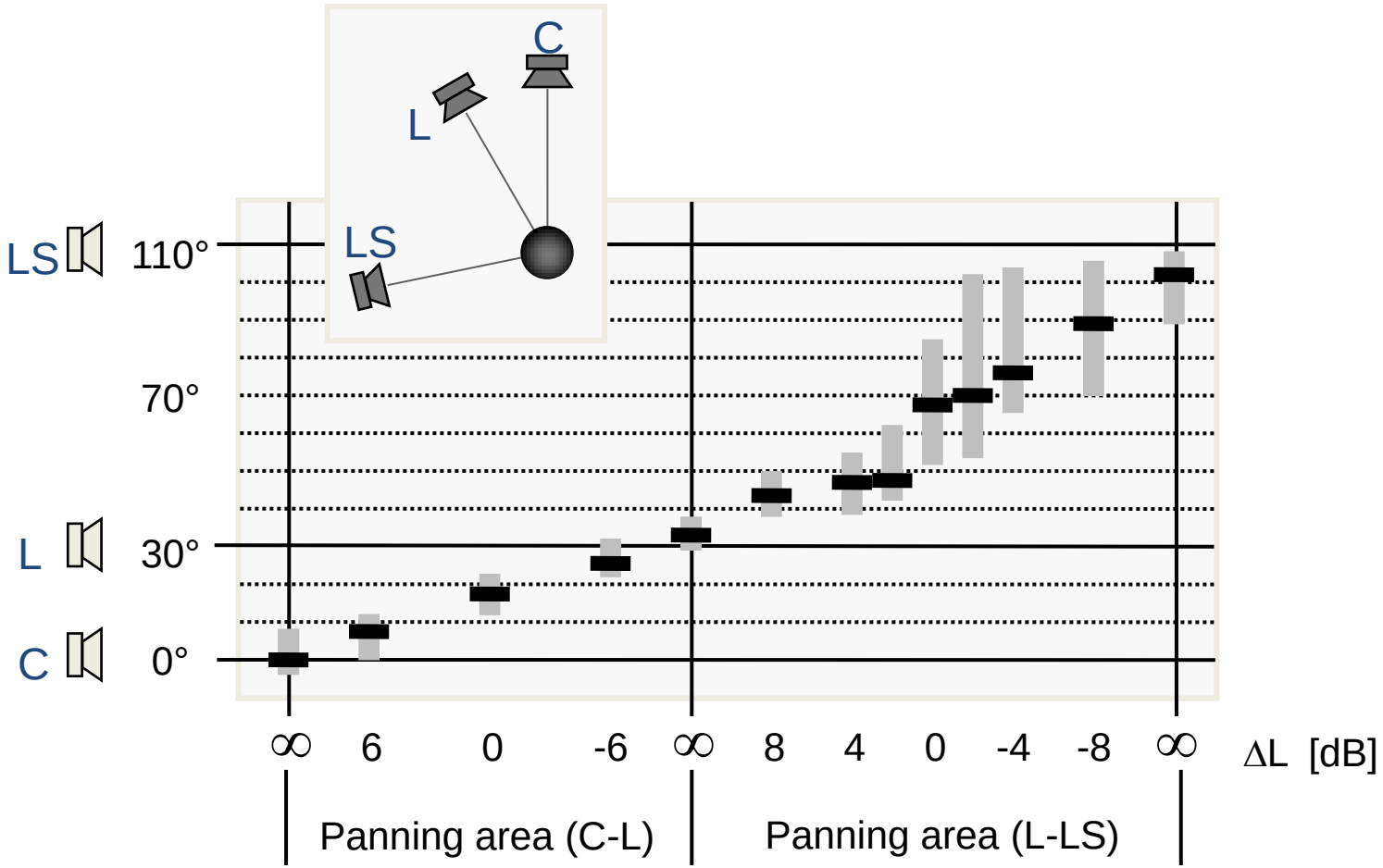
Lateral phantom sources

Listening tests (*Theile et al. 1977*)



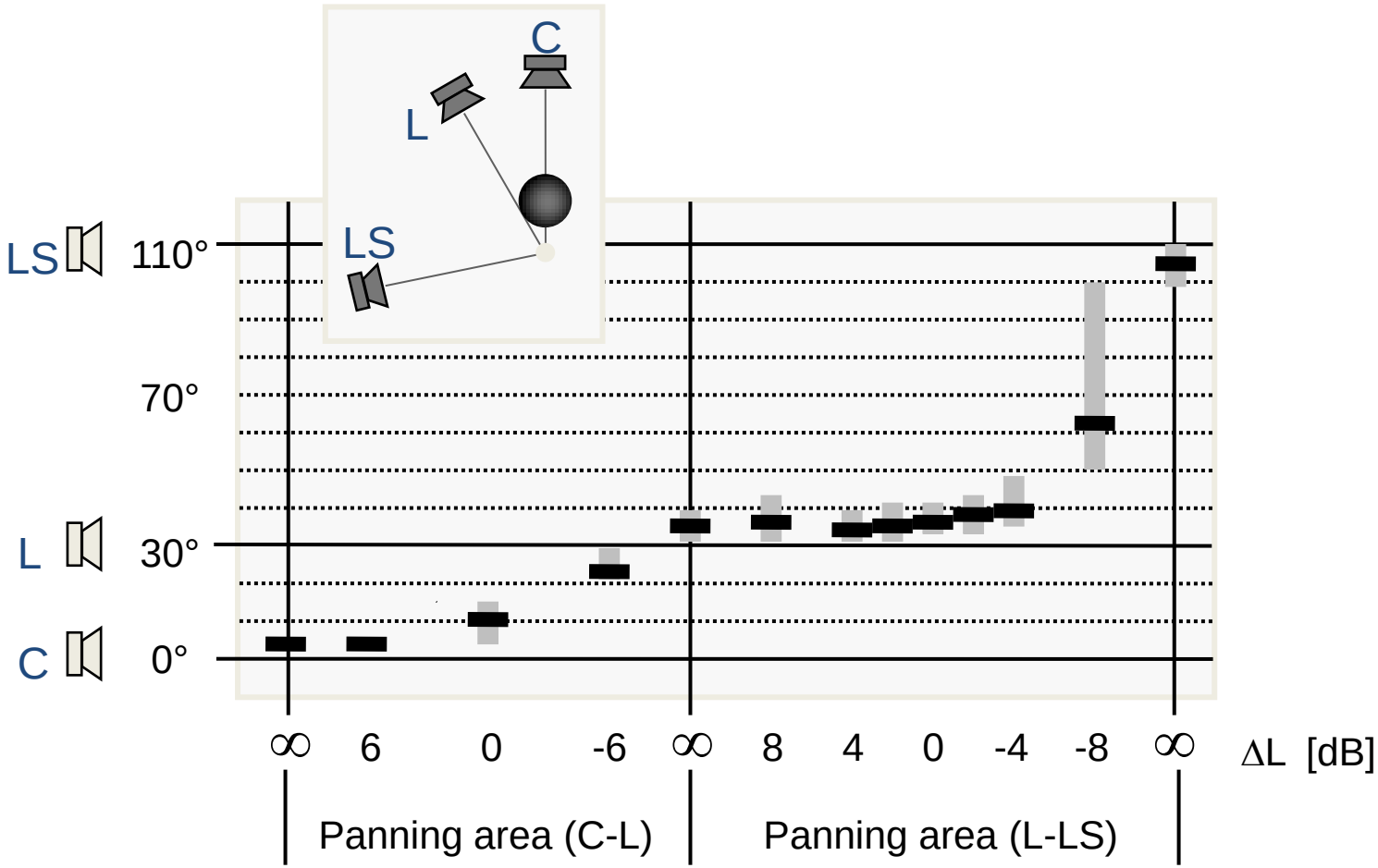
Lateral phantom sources

Listening tests (*Zieglmeier et al. 1996*)



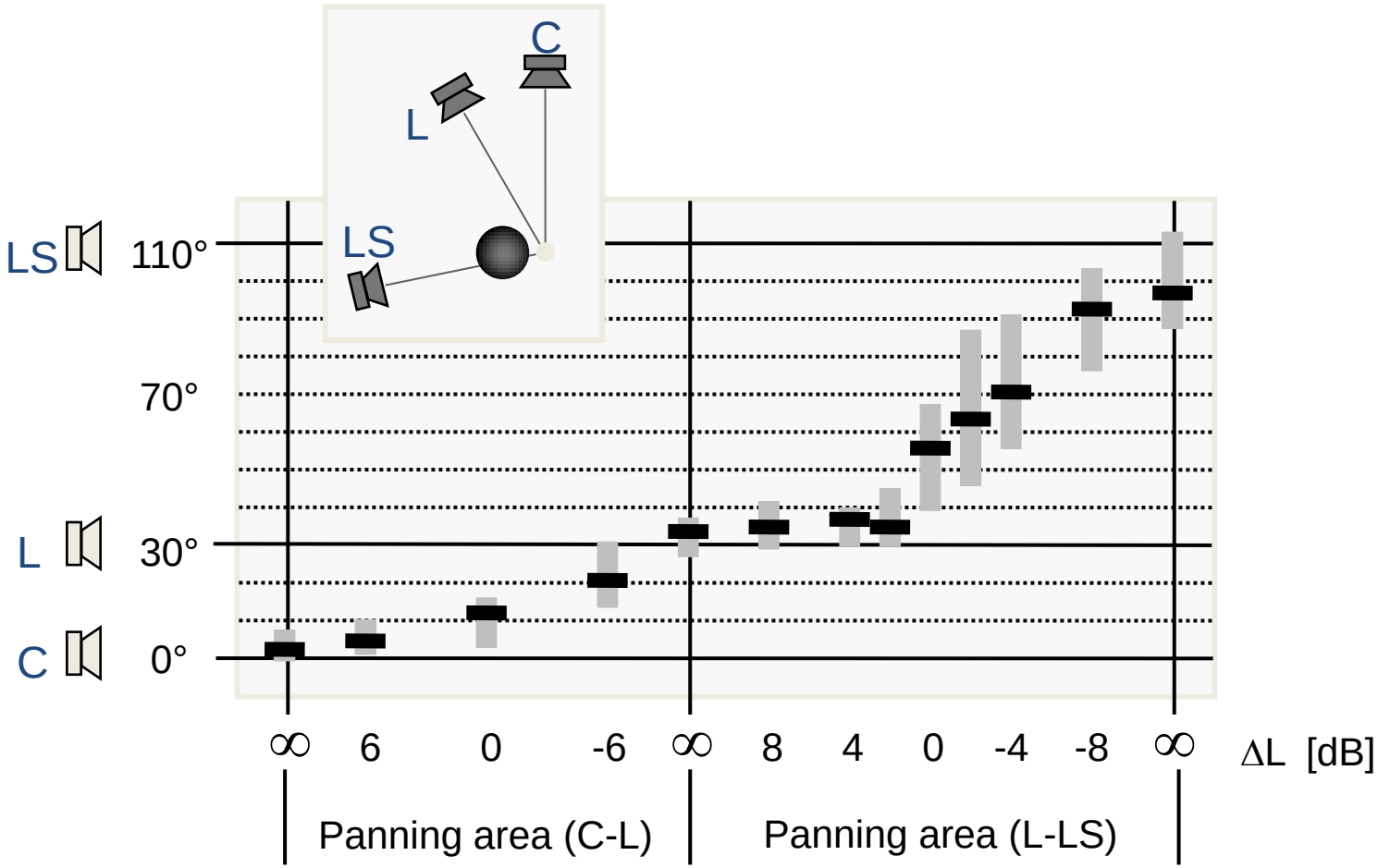
Lateral phantom sources

Listening tests (*Zieglmeier et al. 1996*)

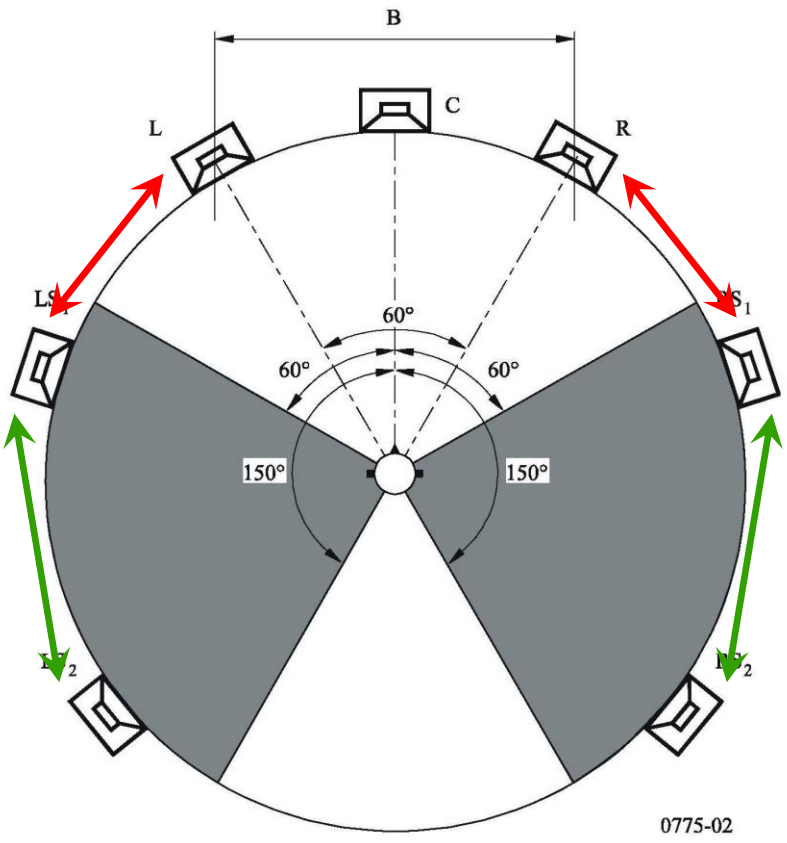


Lateral phantom sources

Listening tests (*Zieglmeier et al. 1996*)



7.1 Surround according to ITU-R 775

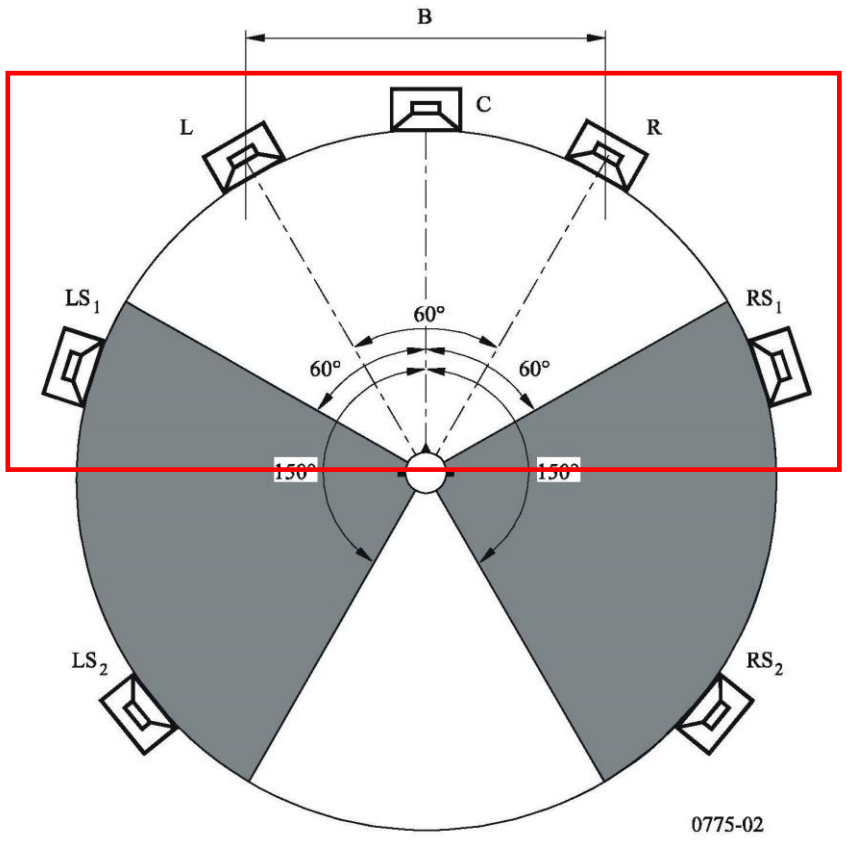


RECOMMENDATION ITU-R BS.775-2
Multichannel stereophonic sound system with and without accompanying picture

NOTE 4 – If more than two rear/side loudspeakers are used, then the loudspeakers should be disposed symmetrically and at equal intervals on the arc which measures from 60° to 150° from the centre front reference

7.1 Surround according to ITU-R 775

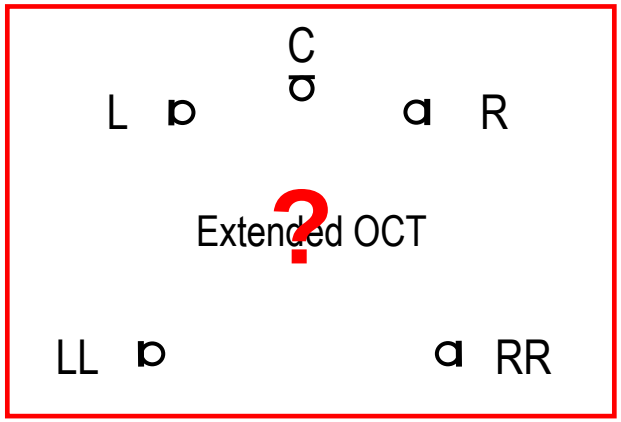
RECOMMENDATION ITU-R BS.775-2
Multichannel stereophonic sound system with and without accompanying picture

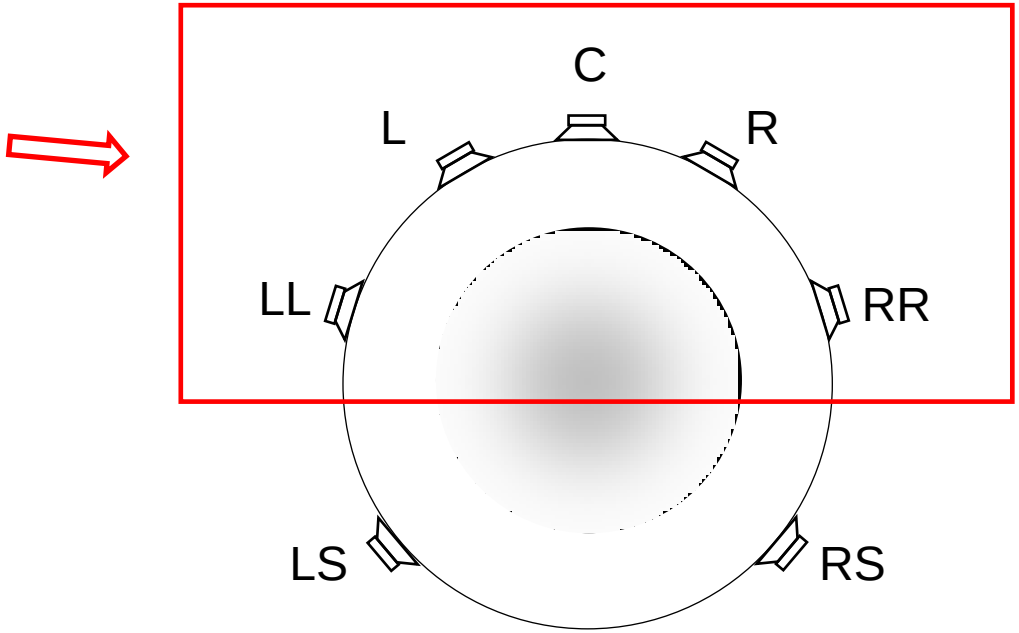


NOTE 4 – If more than two rear/side loudspeakers are used, then the loudspeakers should be disposed symmetrically and at equal intervals on the arc which measures from 60° to 150° from the centre front reference

Verband Deutscher Tonmeister e.V.

Main Microphone: Extended OCT

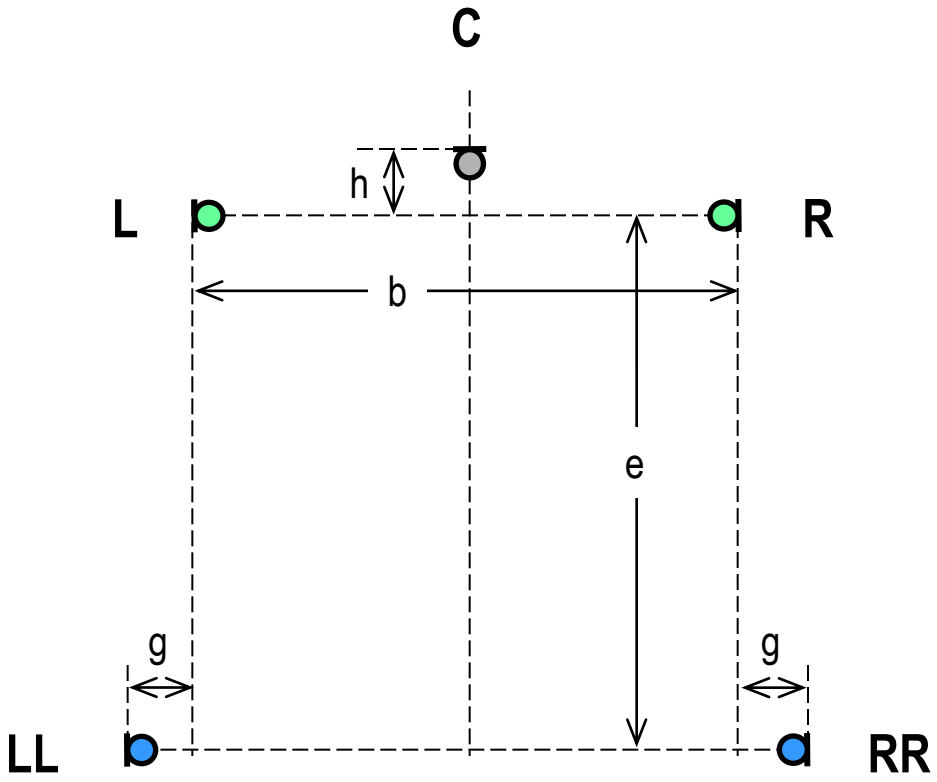




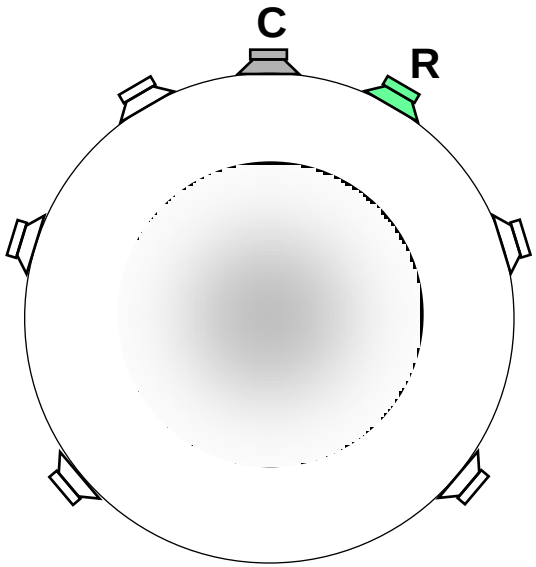
Verband Deutscher Tonmeister e.V.

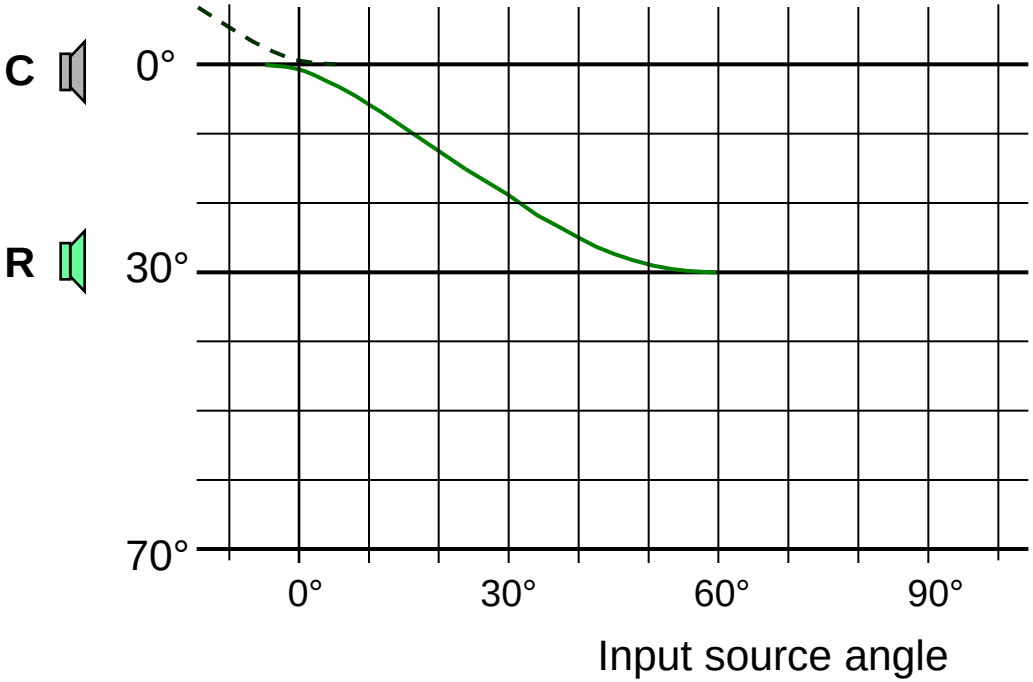
Main Microphone: Extended OCT

L, R, LL, RR:	Supercardioid
C:	Cardioid
h:	8 cm
b:	85 cm
e:	85 cm
g:	8 cm

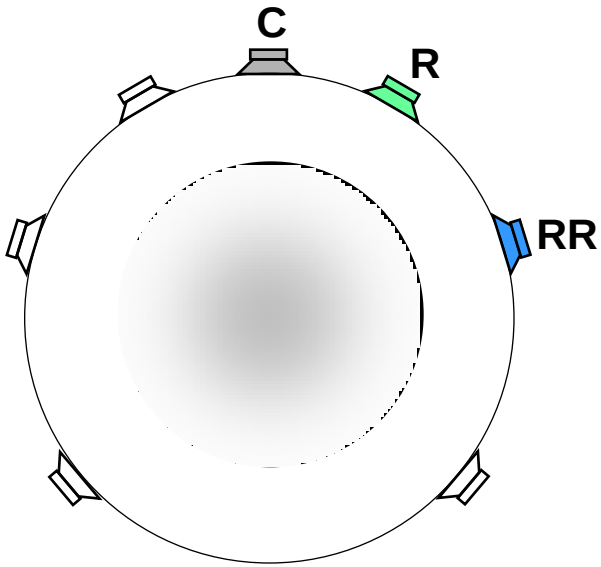


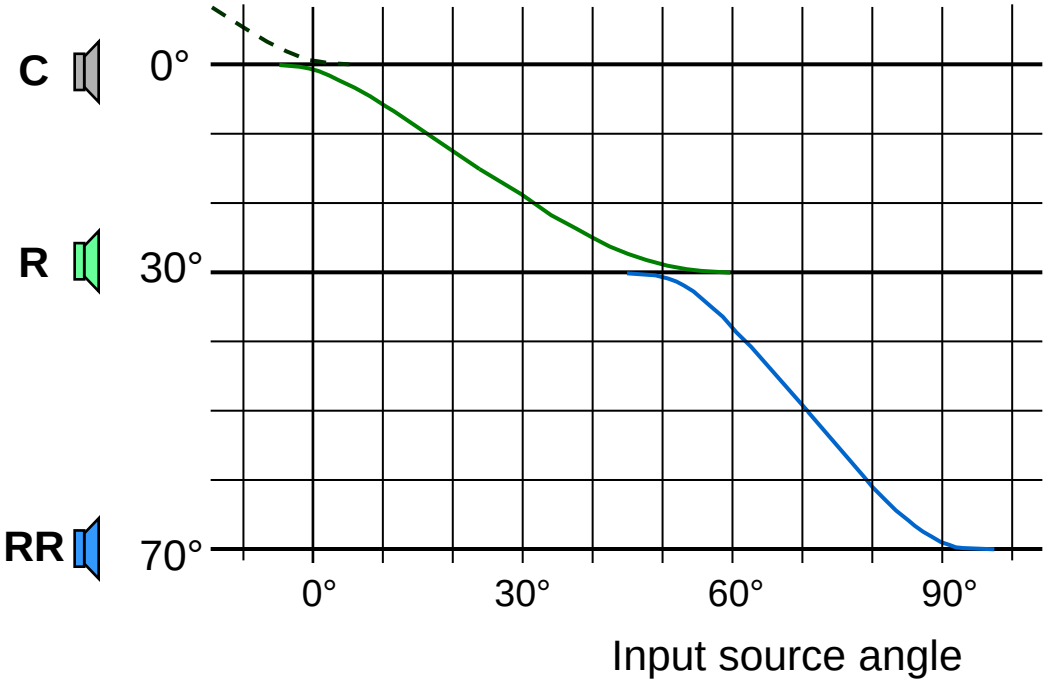
Localization curve of extended OCT



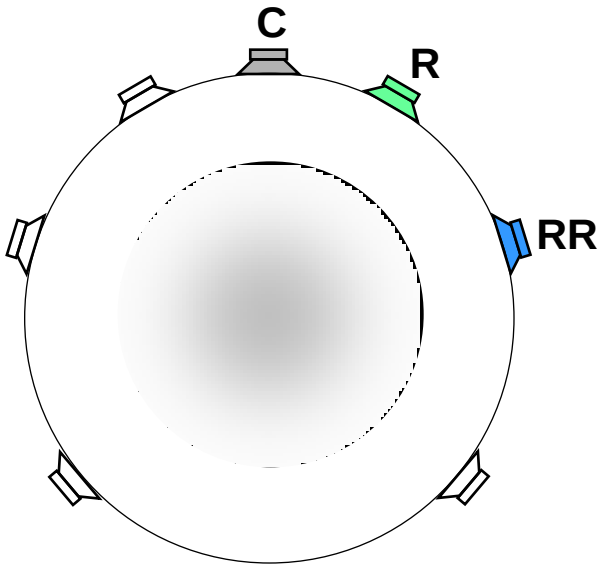


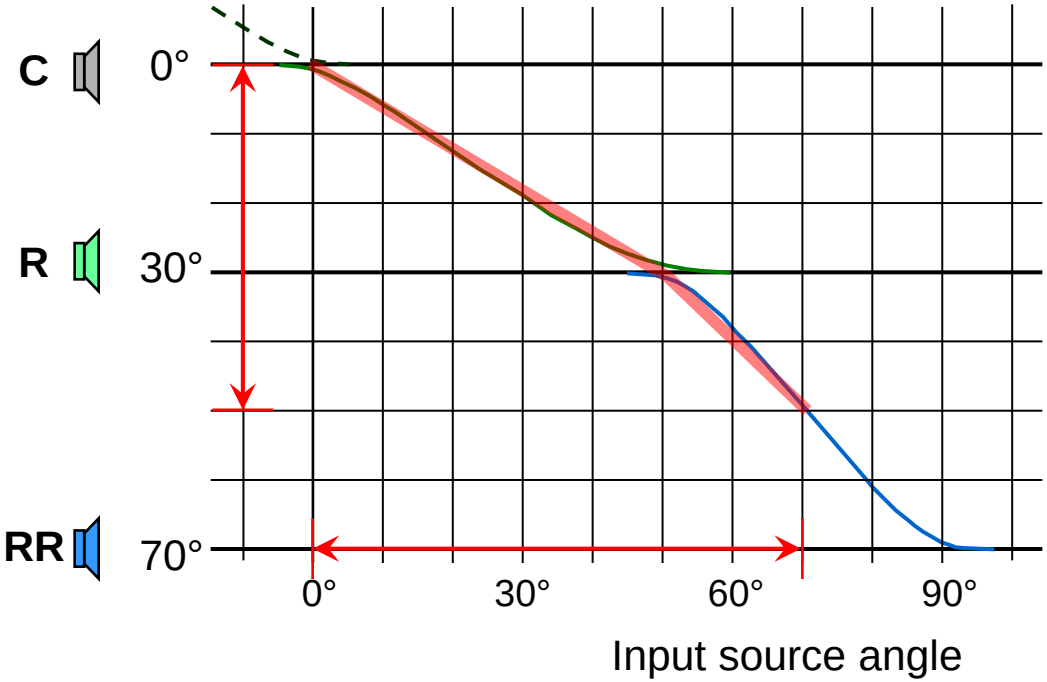
Localization curve of extended OCT



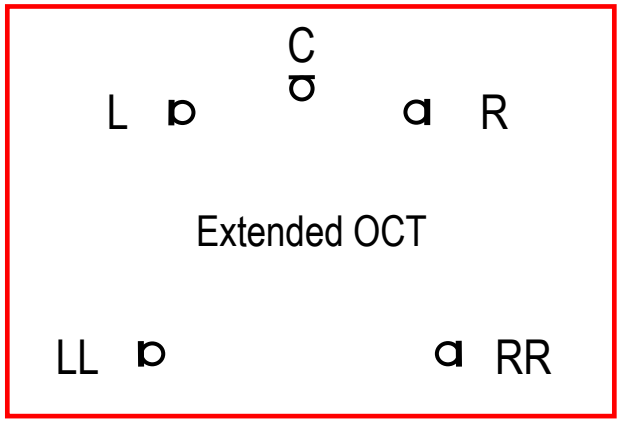


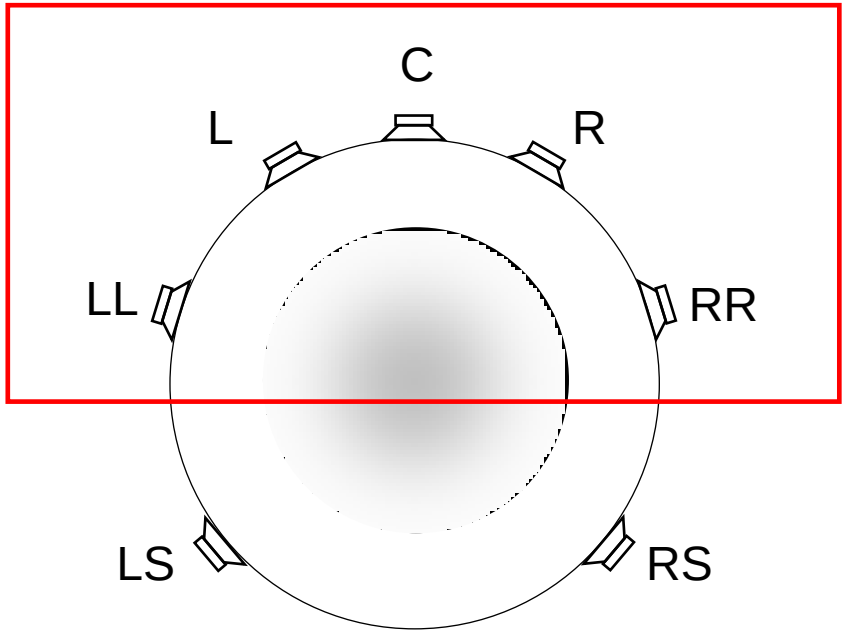
Localization curve of extended OCT



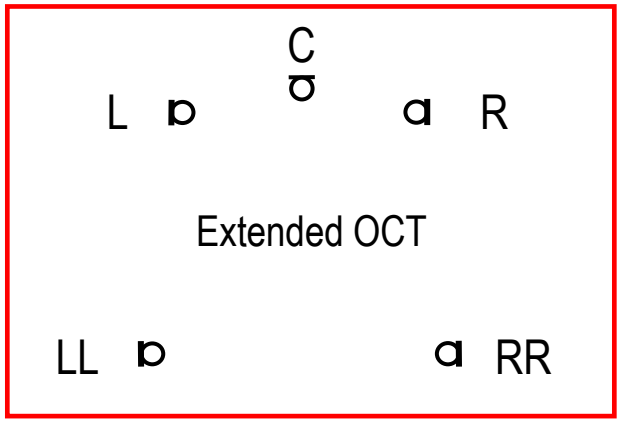


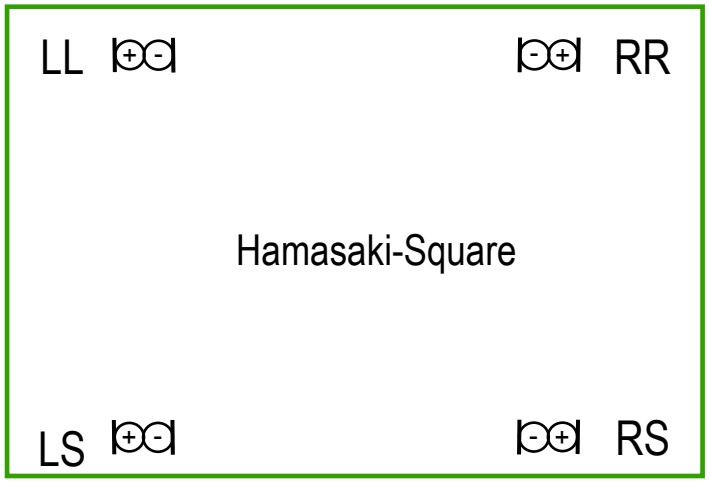
Extended OCT

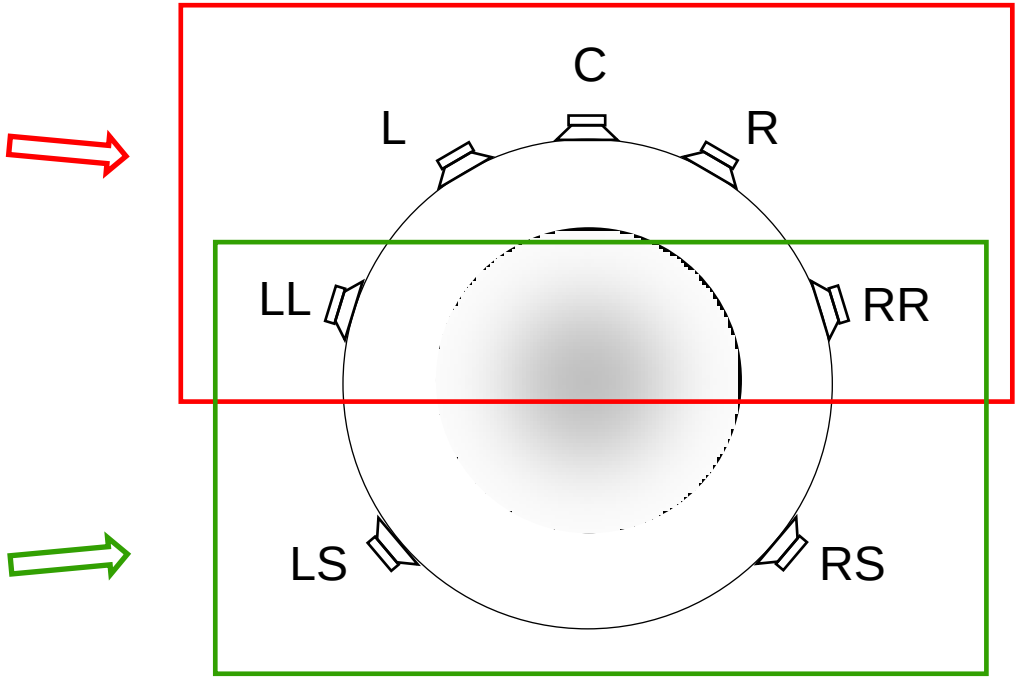




Wide stage recording plus Hamasaki-Square







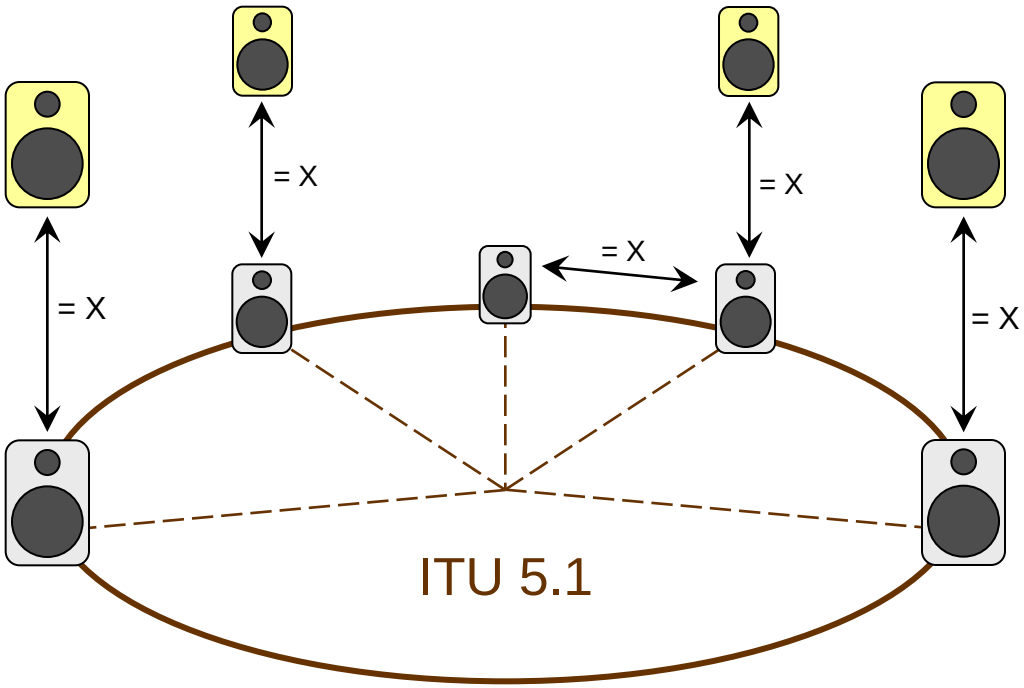
Verband Deutscher Tonmeister e.V.

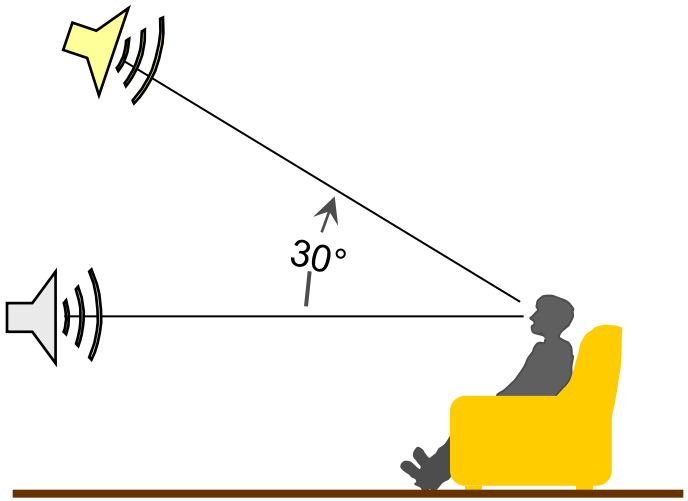
Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

5.1 Surround according to ITU-R 775 plus height: 3D Audio



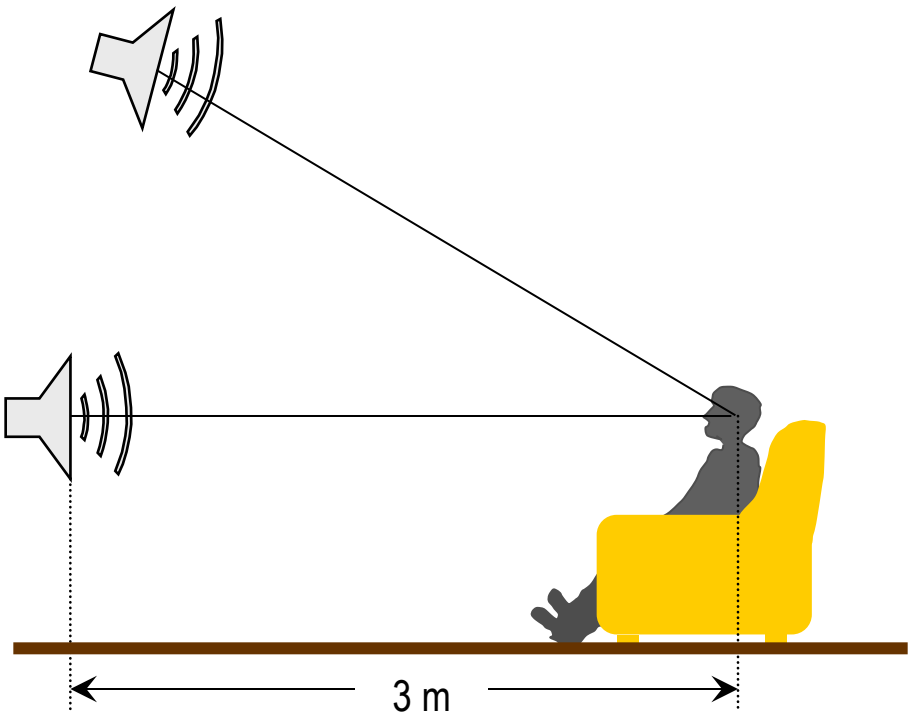


Example:
Auro-3D basic setup (backward compatible with ITU-R BS.775-1)

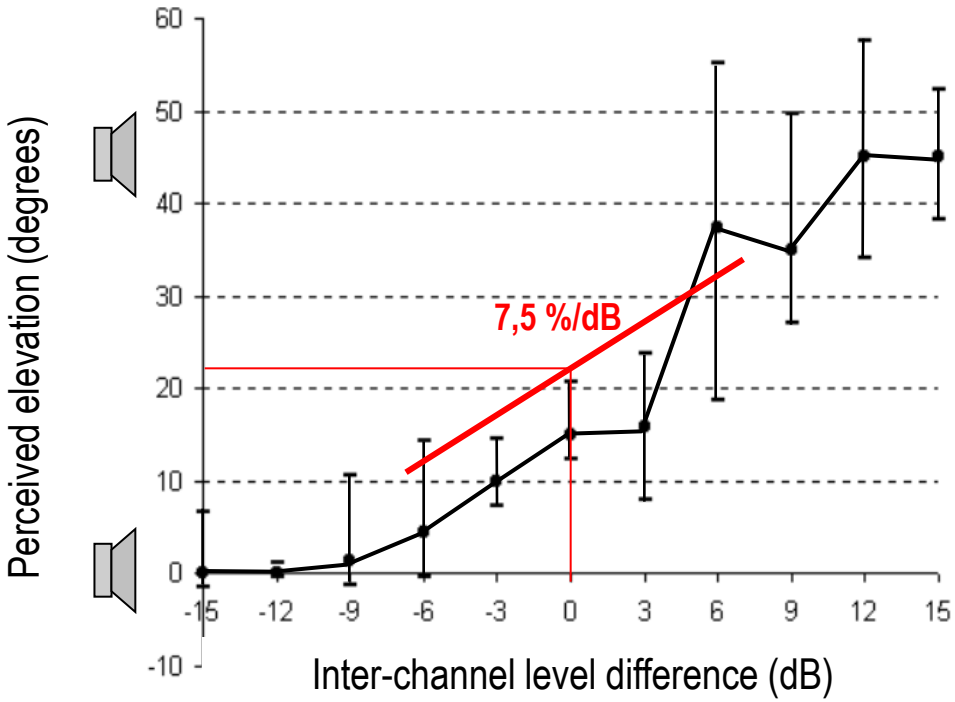
5.1 Surround + 4 height channels

Imaging between surround layer and height layer

Elevation



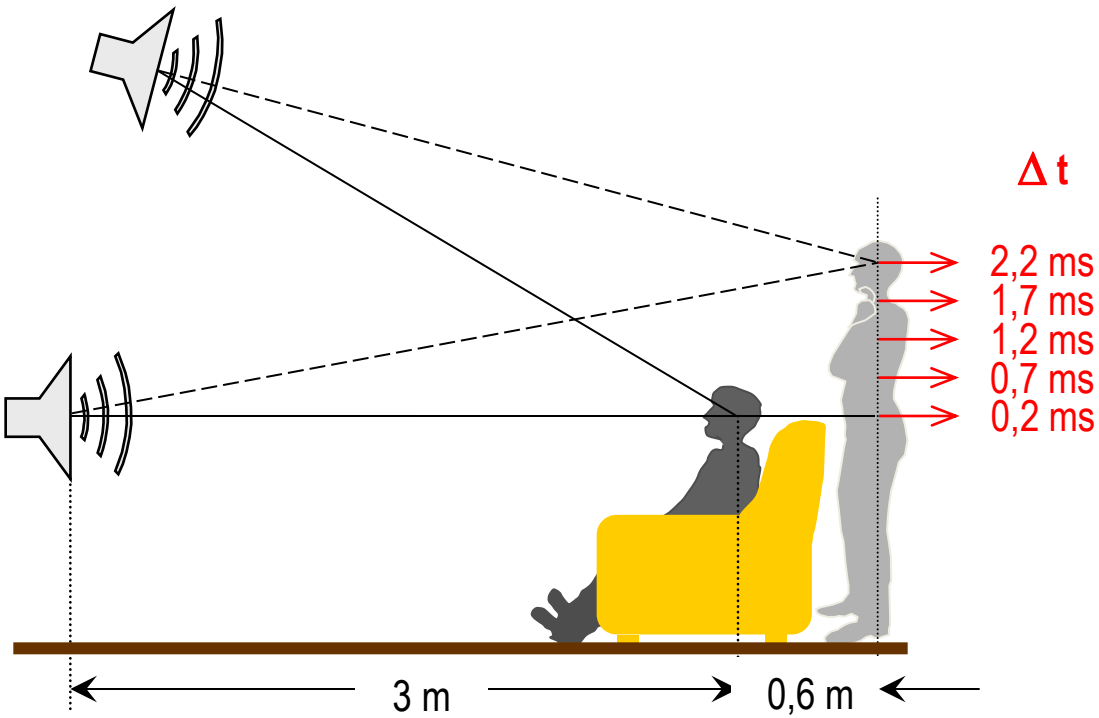
Perceived elevation in the sweet spot



Stereophonic imaging on the median plane affected by level differences (speaker angles: 0° and 45°)
[J. Barbour, 24th AES Conf., 2003]

Elevation

Delay differences occurring in listening positions beyond the sweet spot



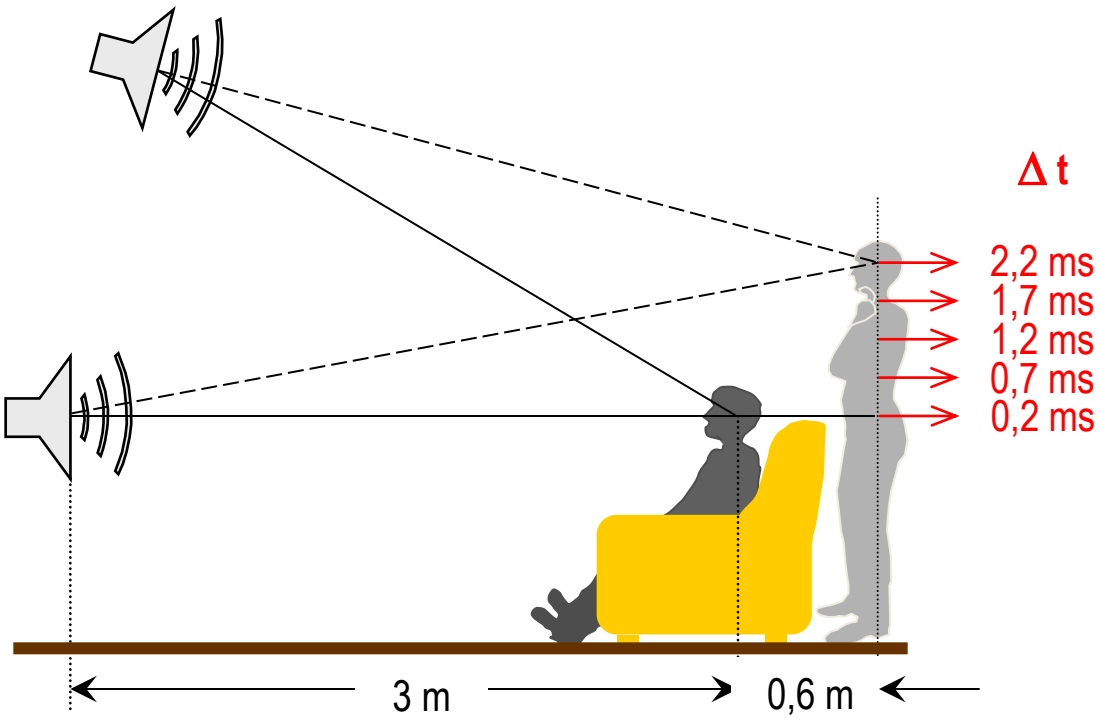
1 m - 3 ms
1 ms - 0,33 m

Directional shift
in case of 2/0 stereo:

15 % / 0,1 ms

Imaging between surround layer and height layer

Elevation



1 m - 3 ms
1 ms - 0,33 m

→ Panning is not recommended!

<u>However:</u>

Filling the areas in height with conveniently decorrelated signals is possible and an important creative element.

Spatially distributed individual acoustic sources (e.g. a choir) recorded with a largely spaced A/B setup can produce many different interchannel delays, resulting in robustness against **Δ t**.

The risk of creating a “hole” in the center is controllable in many recording scenarios.

Verband Deutscher Tonmeister e.V.

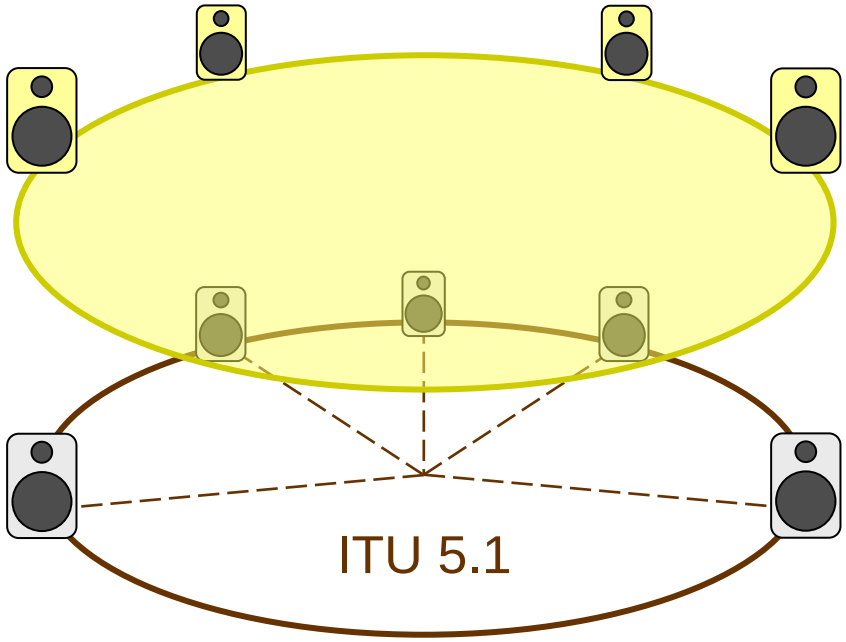
Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

Imaging in the height layer

- Known 5.1 surround directional imaging characteristics in the horizontal area
- The upper layer can be used in similar way
- 4 stable height sources



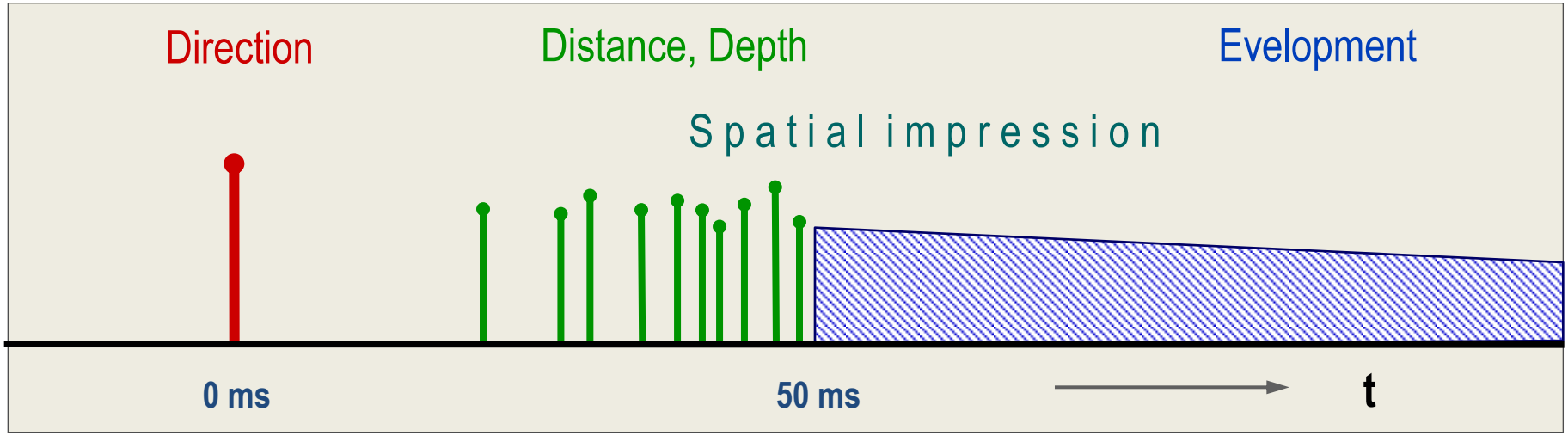
Verband Deutscher Tonmeister e.V.

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	WFS* (HOA*)	BINAURAL TECHNIQUES
Front direction, robustness	●	●●	●●	●●	●
Surround direction		●	●	●●	●●
Elevation			(●)***		●●
Height			●		●●
Distance / depth	(●)**	●	●●	●●	●●
Proximity to the head				●	●●
Intra-active perspective				●●	
Spatial impression	(●)**	●	●●	●	●●
Envelopment		●	●●	●	●●
Timbre	●●	●●	●●	●	●●

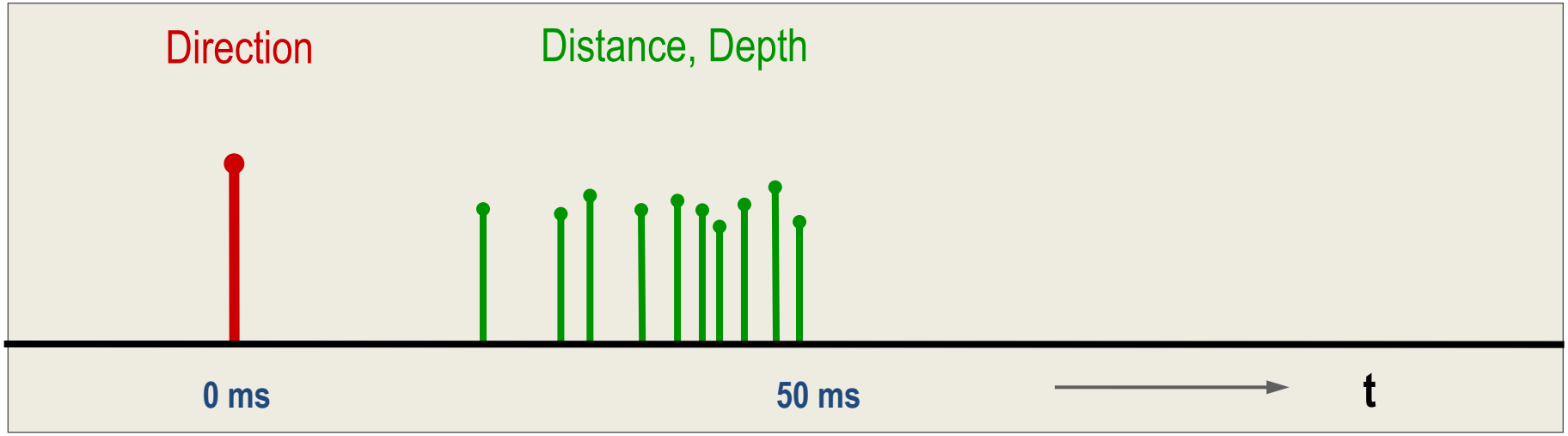
*horizontal arrays **simulated depth/spatial impression ***unstable; at the sweetspot only

Sound field types vs. perceptual attributes



SOUND ATTRIBUTES IN THE HALL	DIRECT SOUND	EARLY REFLECTIONS	REVERB	BACKGROUND NOISE
Timbre	●●	●	●●	
Direction, elevation, height	●●	●		
Distance / depth		●●	●	
Spatial impression		●	●●	
Envelopment			●●	●●

Sound field types vs. perceptual attributes



SOUND ATTRIBUTES IN THE HALL	DIRECT SOUND	EARLY REFLECTIONS		
Timbre	●●	●		
Direction, elevation, height	●●	●		
Distance / depth		●●		
Spatial impression		●		
Envelopment				

Imaging of distance and depth

Equal image size → Equal image loudness

Imaging of distance and depth

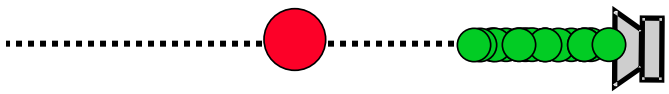
Equal image size → Equal image loudness

Additional room information

Early reflection pattern plus corresponding image loudness

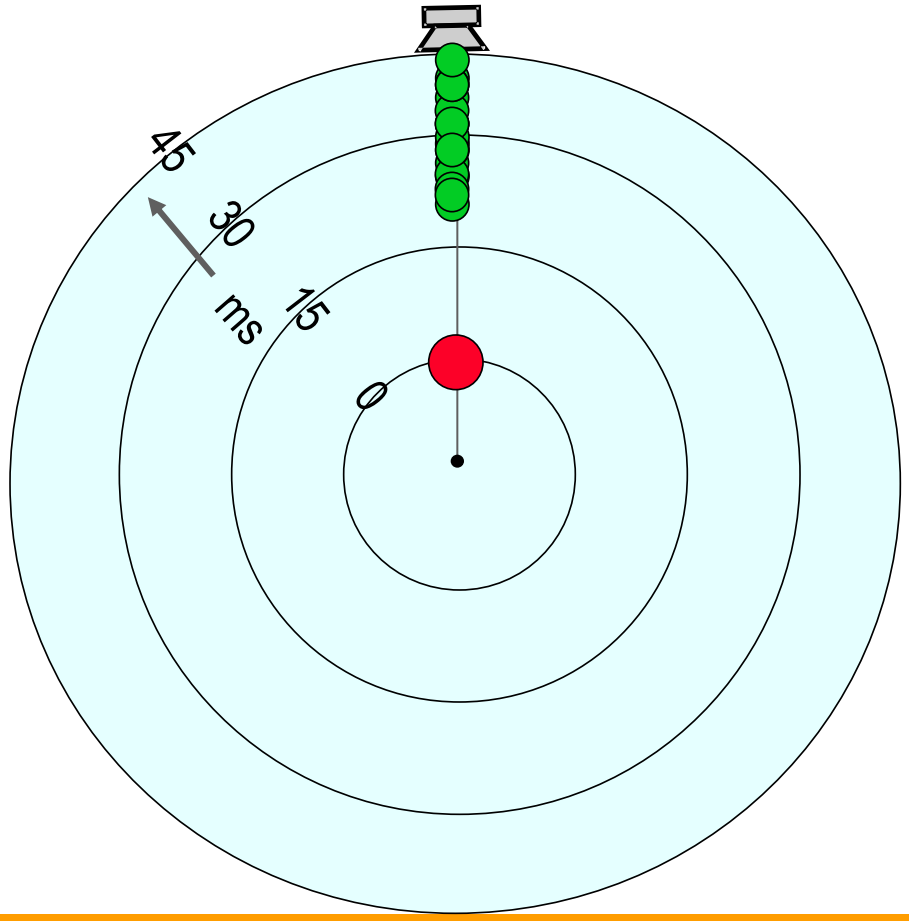
Spatial rendering of reflection pattern

side view



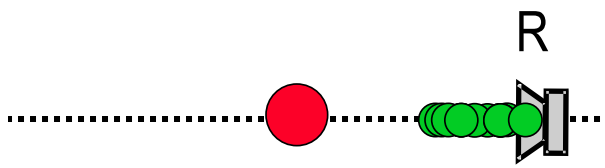
1/0

top view



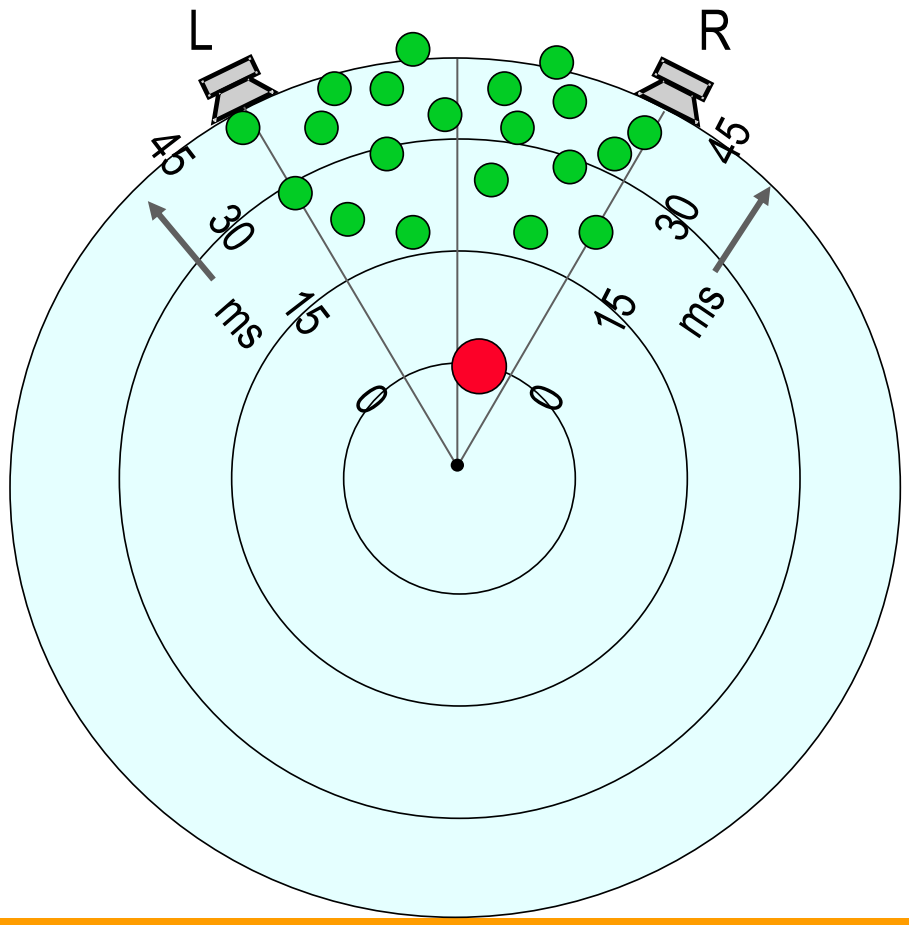
Spatial rendering of reflection pattern

side view



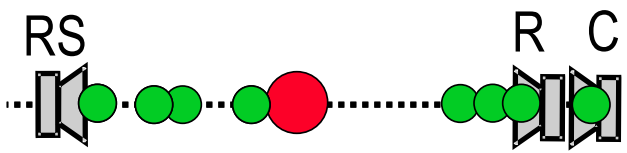
2/0

top view



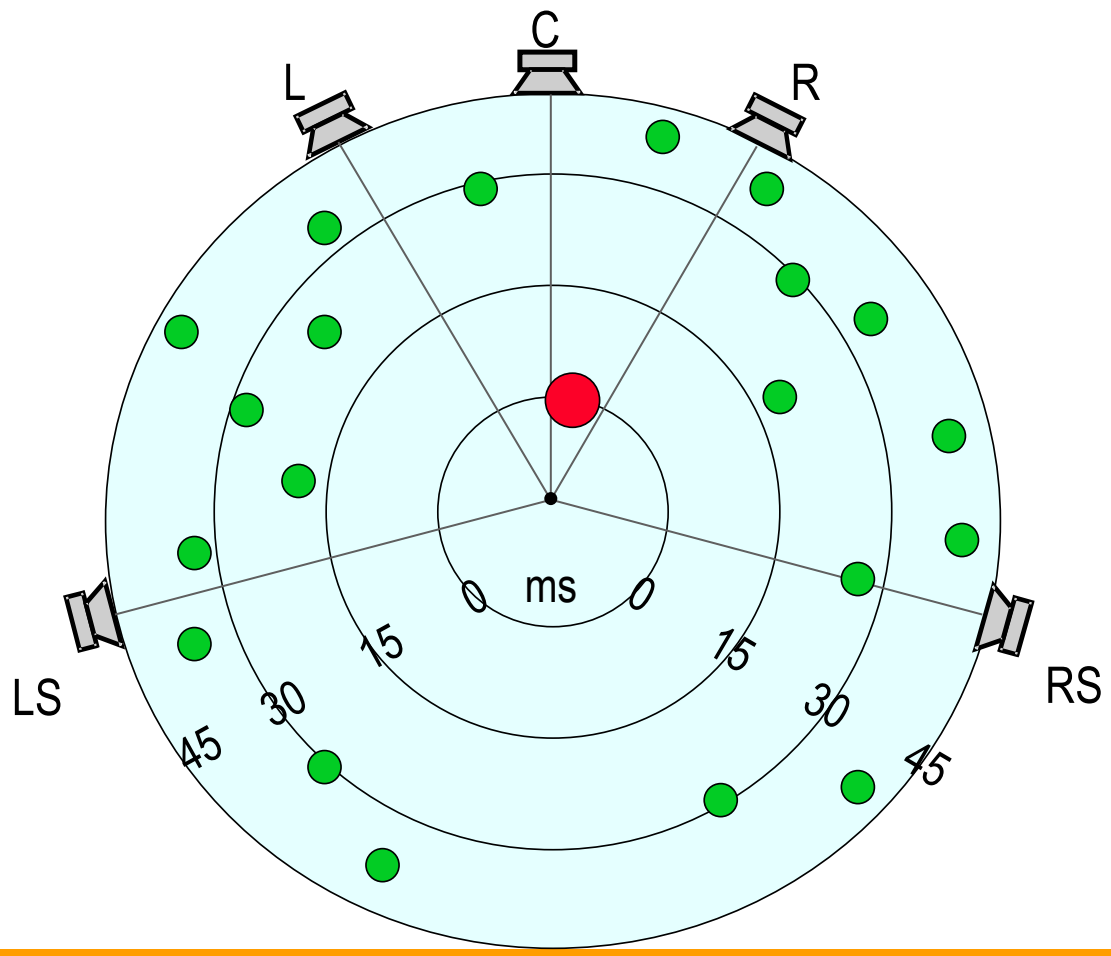
Spatial rendering of reflection pattern

side view



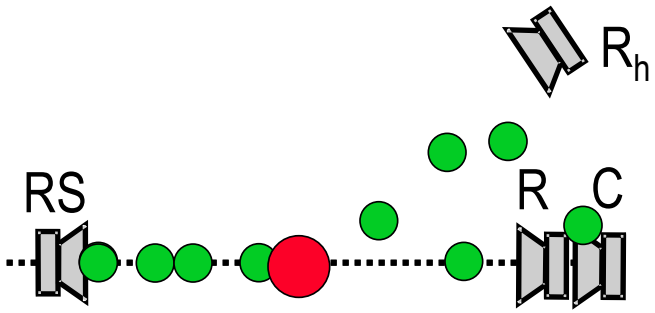
3/2

top view



Spatial rendering of reflection pattern

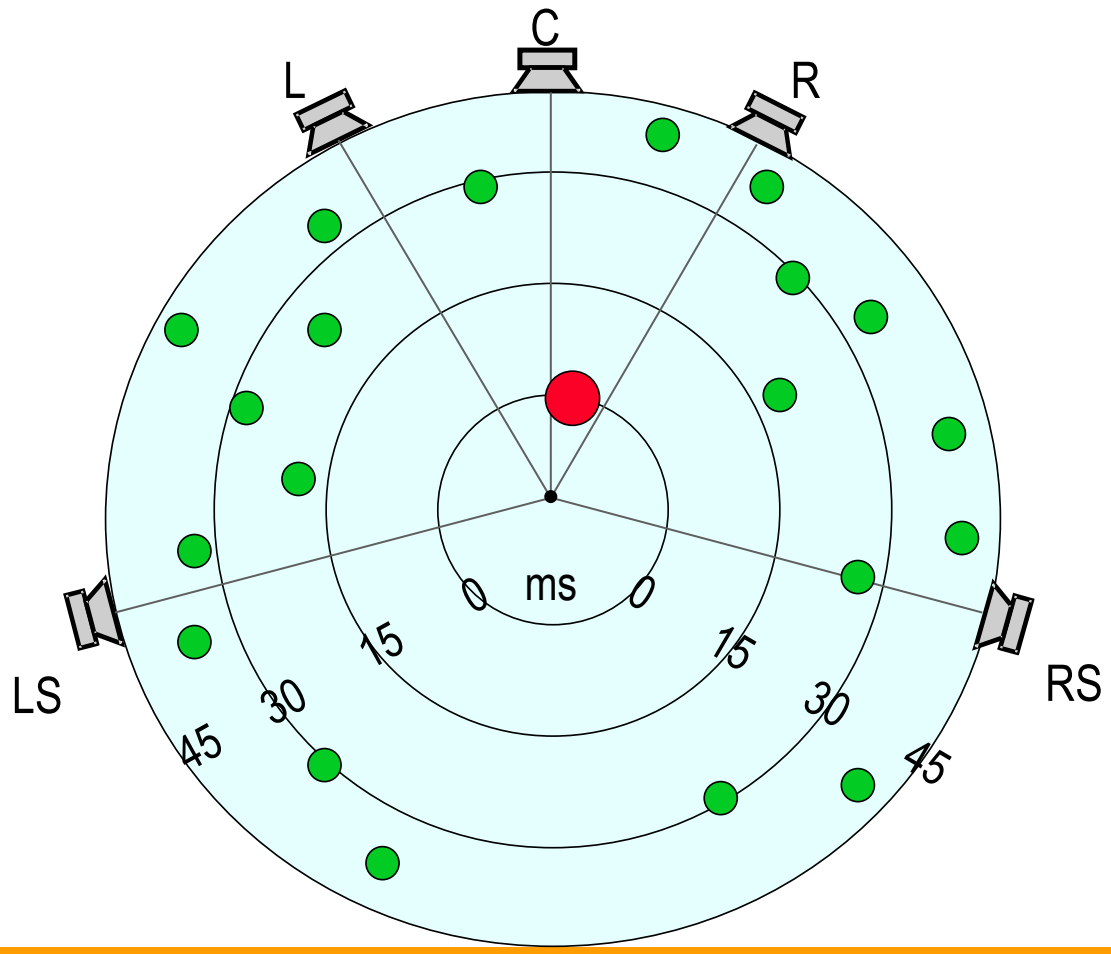
side view



3/2 +

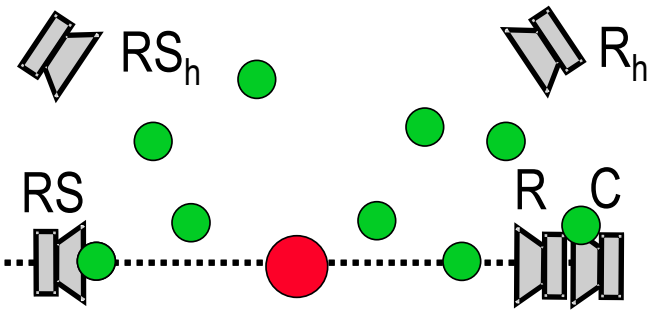
height

top view

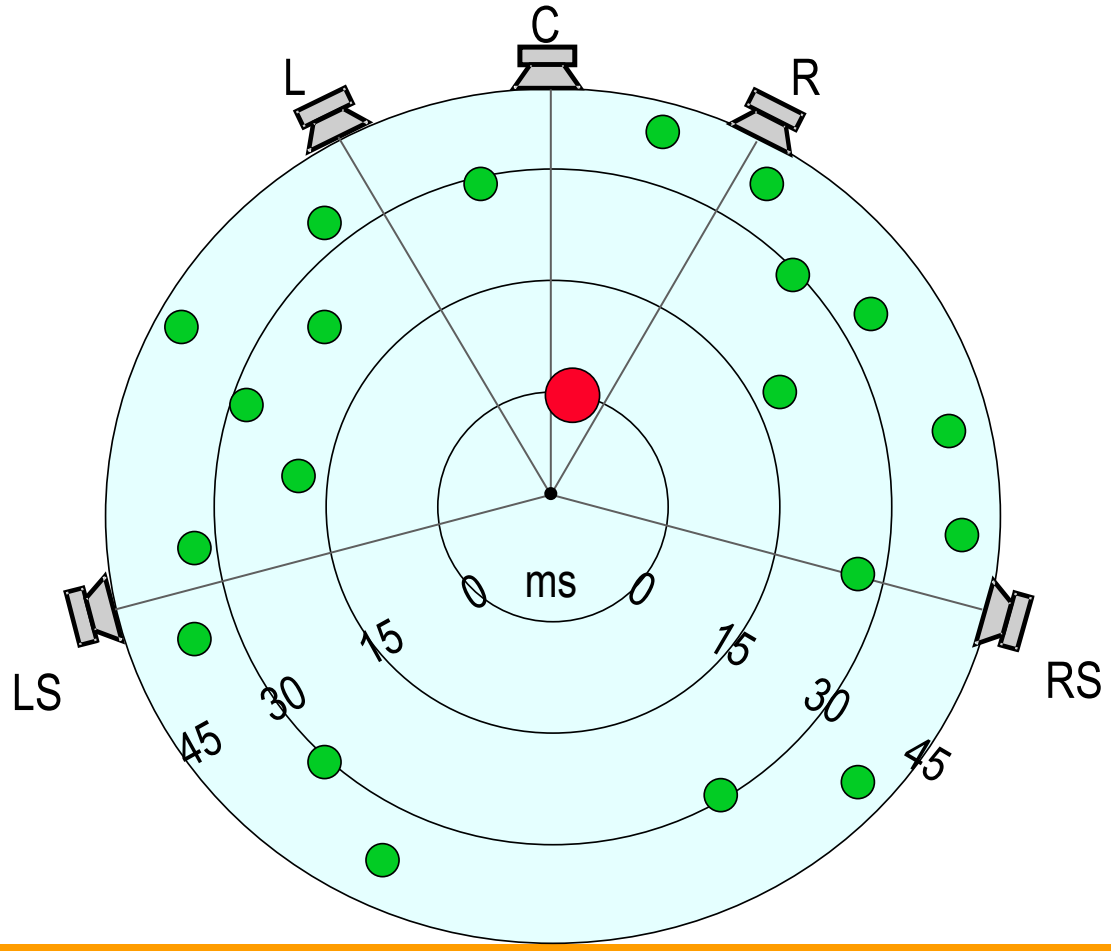


Spatial rendering of reflection pattern

side view



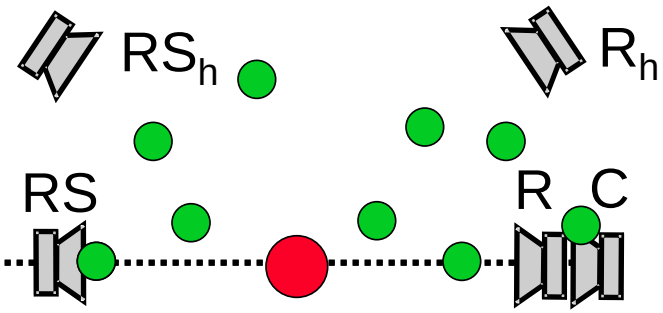
top view



3D Audio
9.1

Spatial rendering of reflection pattern

side view



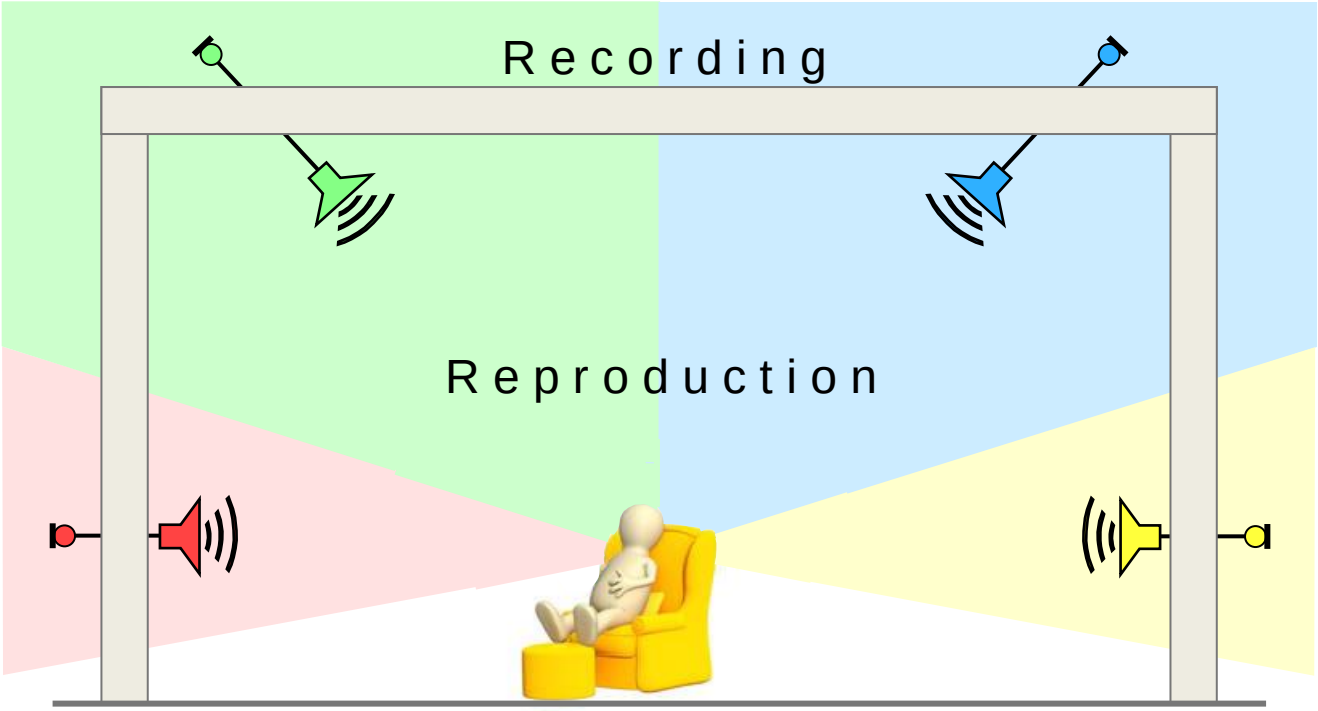
3D Audio
9.1

3D Audio:

- Reflection density is lower
- Spatial distribution of reflections is more natural
- Detection and interpretation of reflections is easier for the brain
  → coloration is minimized
  → distance / depth perception is better

Channel separation is mandatory

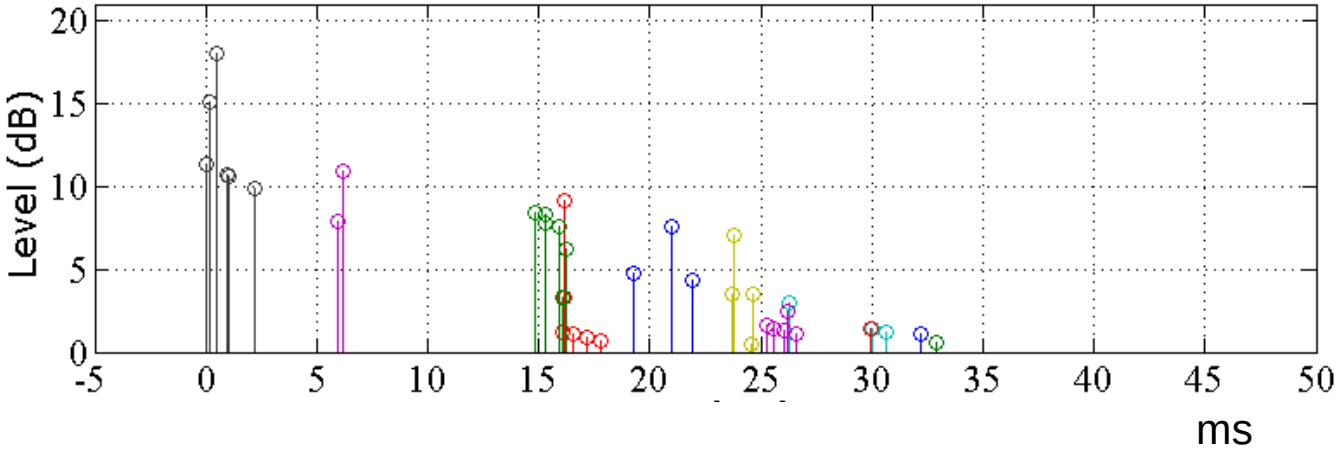
Reproducing the original incident directions requires
strict channel separation during the recording process



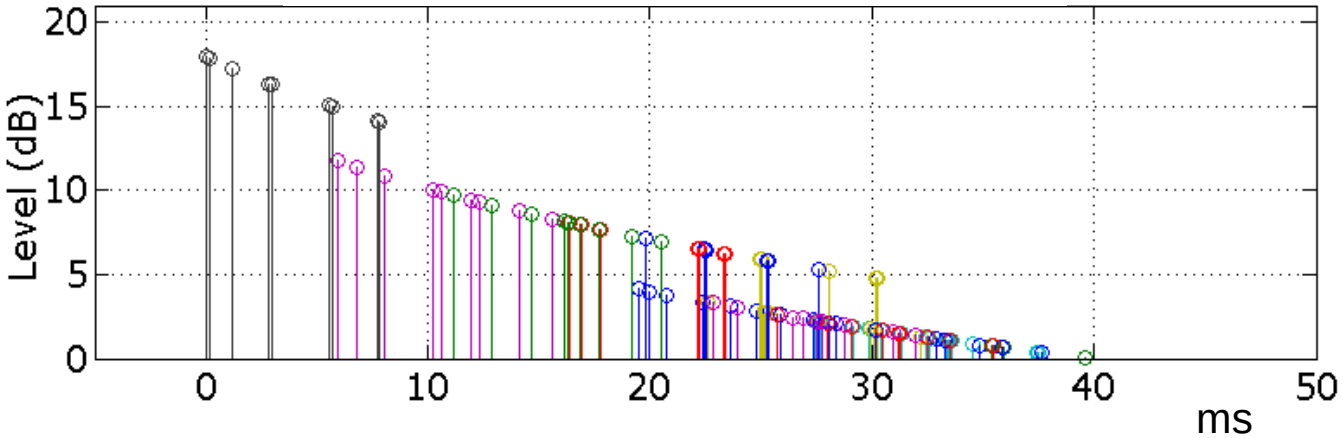
How to realize channel separation for room information?

Reflection patterns in the sweet spot of a 3D Audio 9.1 speaker array [Wittek 2011]

OCT 70 + 4 supercardioids pointing upwards



Largely spaced 9-channel A/B setup



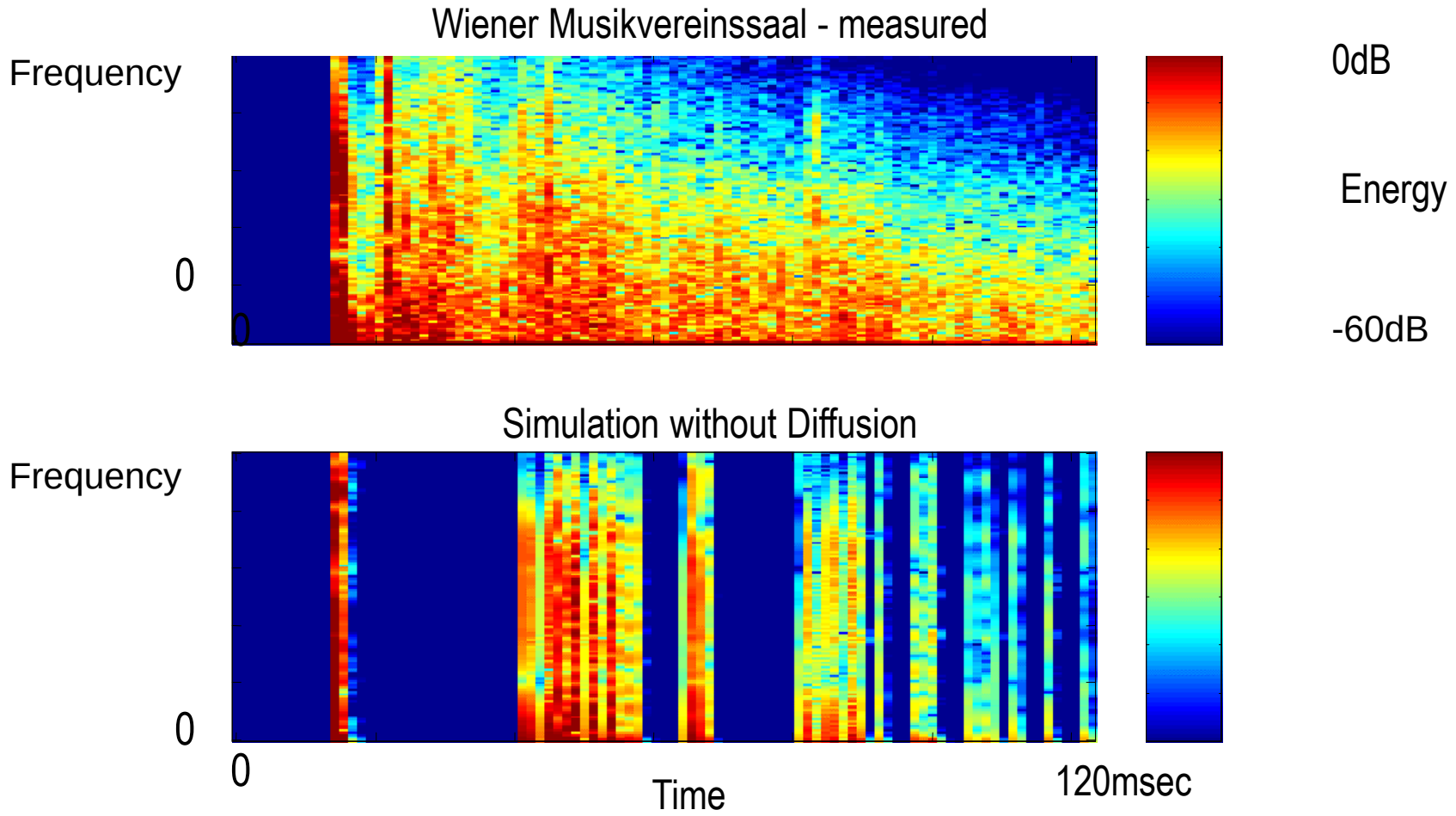
Characteristics of largely-spaced 9-channel A/B setup

Excellent results regarding envelopment and timbre.
Sufficient spatial diffusivity even at low frequencies due to largely spacing.

However:

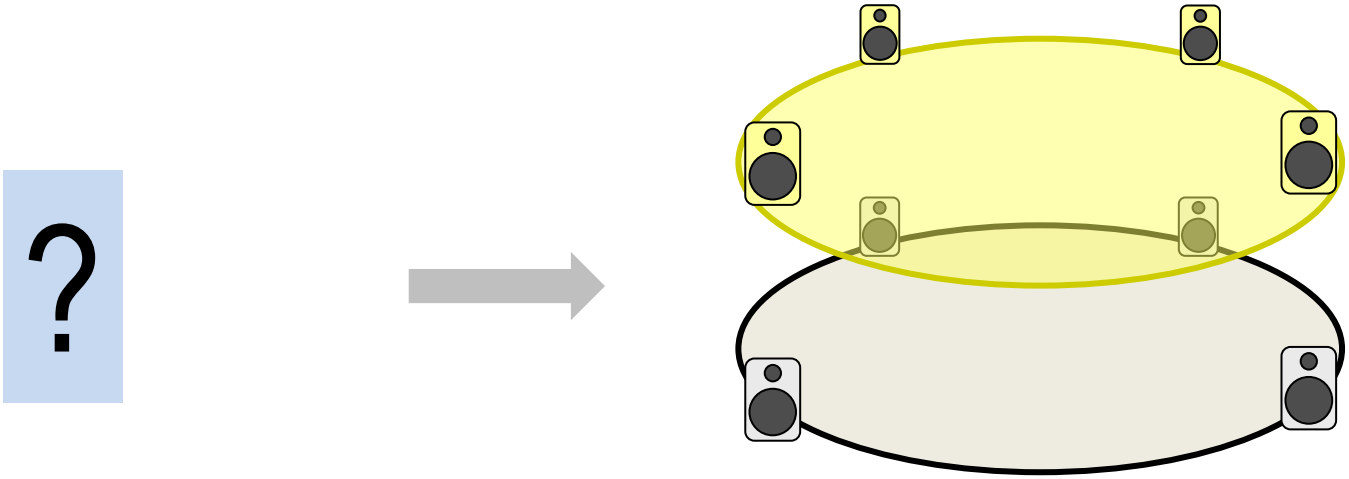
- Danger of perception of direct sound in the surround channels
- No channel separation. Each channel contains identical spatial information
- The original directional characteristics of the reflection pattern are completely lost
  → Distance imaging only by means of direct sound (spot mikes) / indirect sound balancing
- Imaging according to well-known characteristics of large A/B, including the vertical areas
  → Instable localization, dependent on spectrum
  → Elevation balance only for listener in the sweet spot
  → Filling up the areas in height is possible and an important creative element
- Spot miking required

Imaging of distance and depth



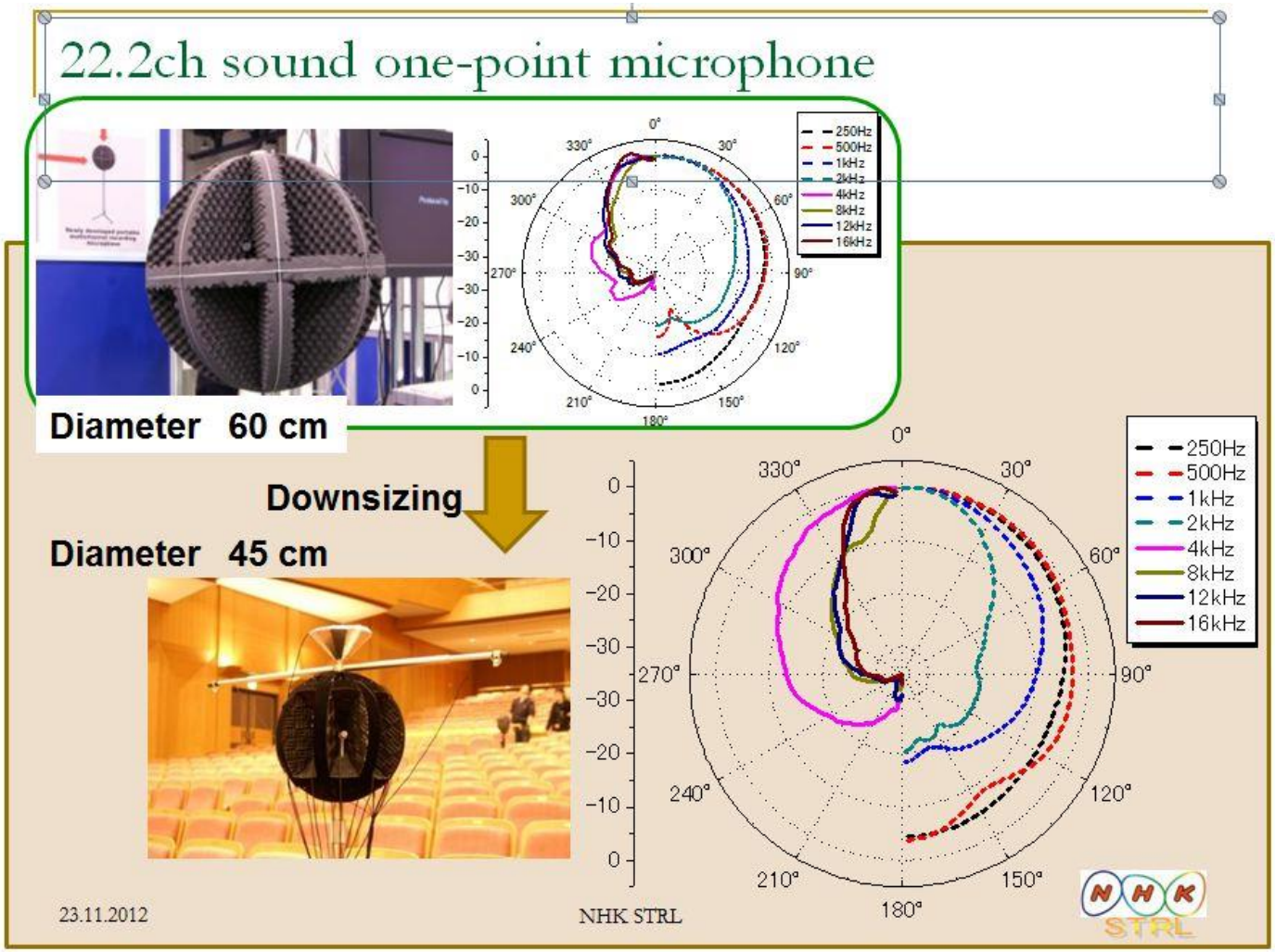
Discrete and characteristic key reflections of the room should not be masked !

How to get channel separation for recording the reflection pattern?



How to get channel separation for recording the reflection pattern?

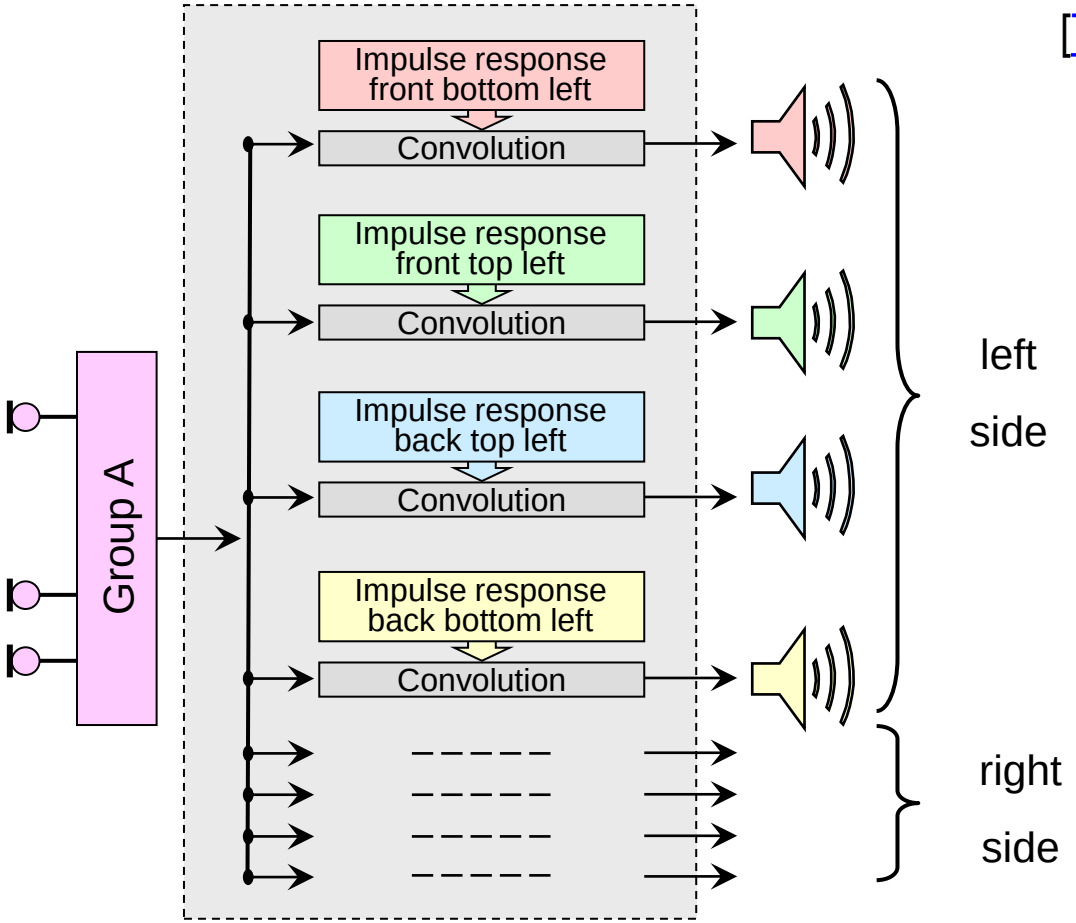
Proposal Hamasaki
(15th FISM 2012)



How to get channel separation for recording the reflection pattern?

Concept of a convolution processor producing 8-channel early reflections (plus reverberation)

[Theile & Wittek 2011]



Multichannel stereophony formats – basic characteristics

Verband Deutscher Tonmeister e.V.

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

Principle weakness of stereophonic formats

Stereophony:
Zero imaging space between loudspeakers and listener

Imaging of stable auditory events between loudspeakers and listener is not possible.
Acoustic environment close to the head cannot be reproduced.

Verband Deutscher Tonmeister e.V.

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

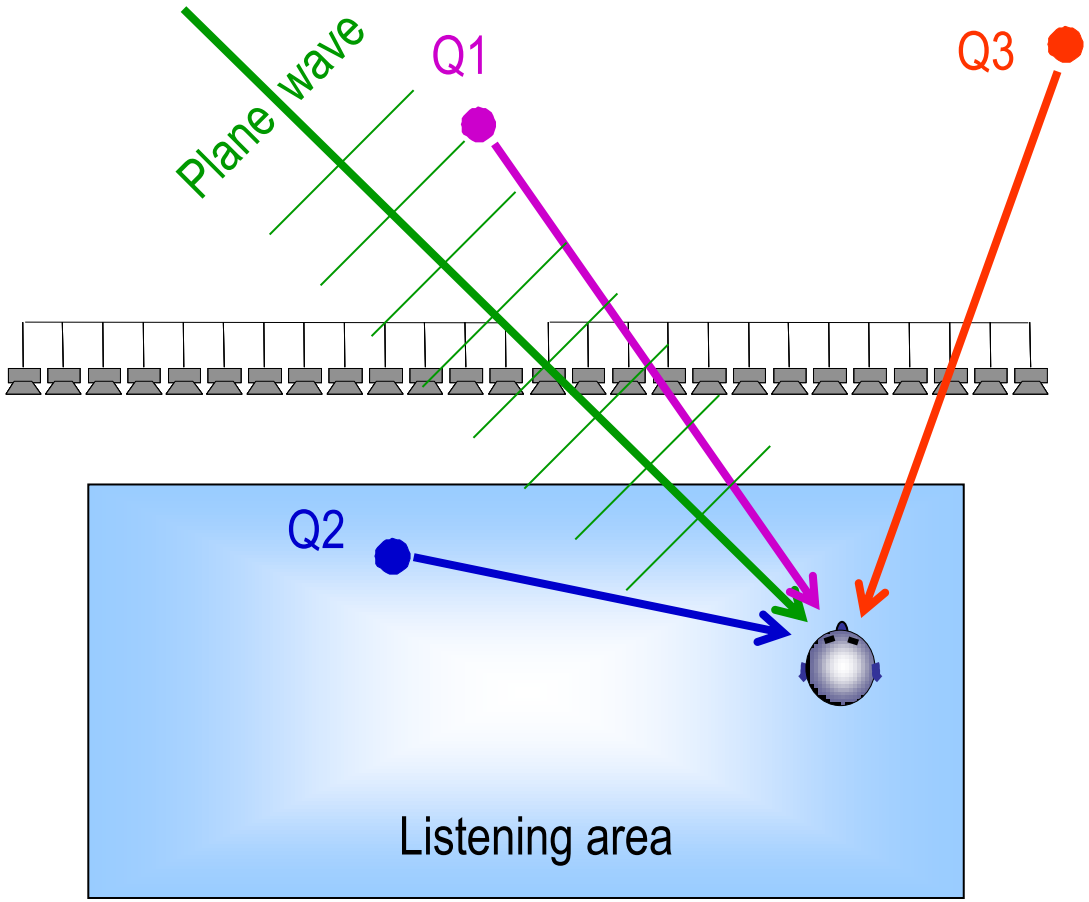
Intra-active perspective acoustic scene

A feature of natural auditory scene analysis.

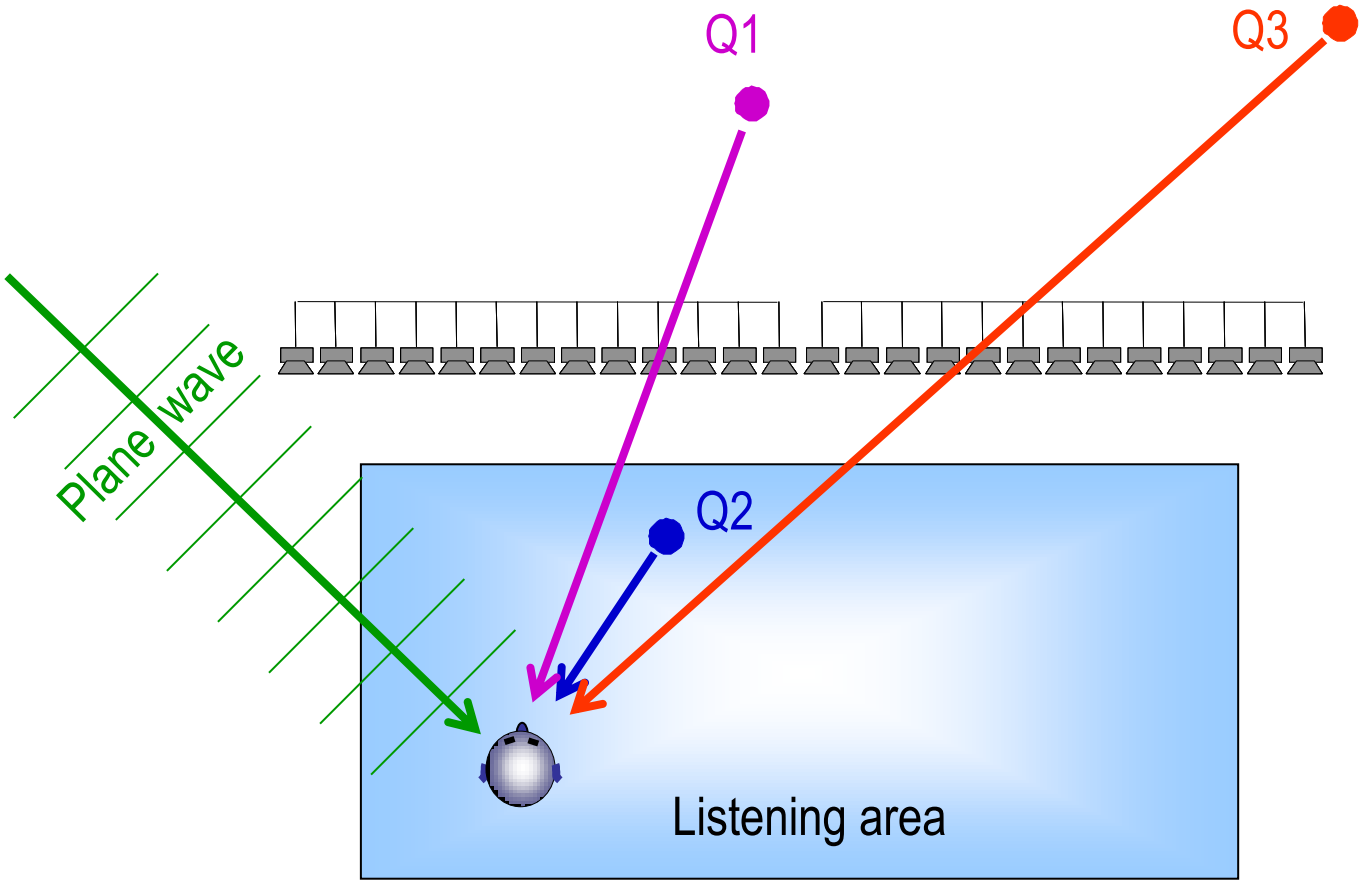
The way directions are perceived changes depending on the distance to the sources: when nearby sources move, they travel “further” than remote sources.

WFS systems allow for reproducing this behavior – a fact of that may be interesting, for example, for future developments in gaming.

WFS:
Rendering an intra-active perspective acoustic scene

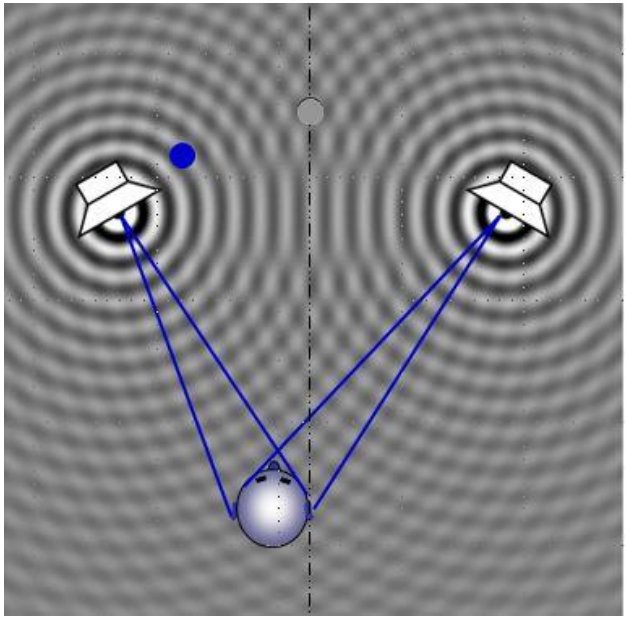


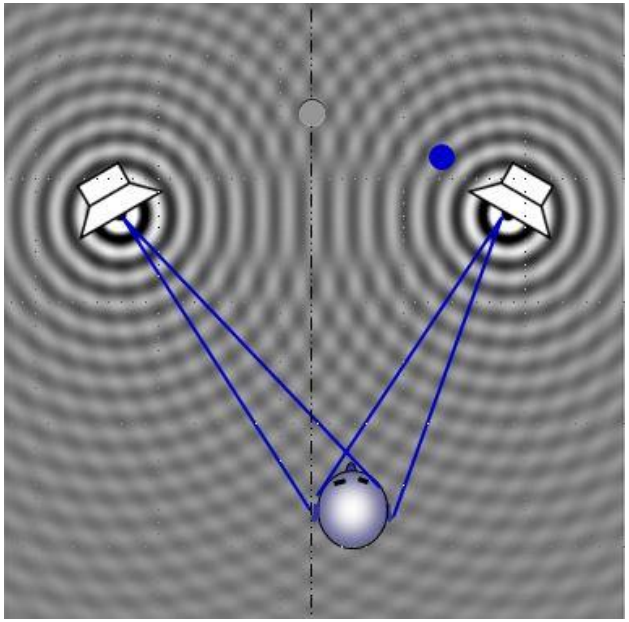
WFS:
Rendering an intra-active perspective acoustic scene



Principle weakness of stereophonic formats

Stereophony:
Zero rendering an intra-active perspective acoustic scene





Multichannel stereophony formats – basic characteristics

Verband Deutscher Tonmeister e.V.

Comparison of stereo / surround format profiles

ATTRIBUTES OF SOUND REPRODUCTION	STEREO 2.0	SURROUND 5.1	3D AUDIO 9.1	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

Atmo

Desirable effects :

- Representation of the auditory scene environment
- Support of spatial perception
- Involvement of the listener in the action

Atmo

Provides the listener an immersive experience

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)

Reverberance / envelopment / ambient noise recording

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

Reverberance / envelopment / ambient noise recording

Note:

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

That´s why:
No sound sources close to an atmo-microphone (minimum distance ≈ 3 m)!

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

Reverberance / envelopment / ambient noise recording

IRT-cross (*Theile 1996*):

4.0 stereo atmo microphone cofiguration:
4 cardioids 25 cm spaced

The symmetrical capsule arrangement ensures equivalent correlation for the signal pairs LR, R-Rs, Rs-Ls, Ls-L in the diffuse sound field as well as a good stereophonic quality of the surrounding environment.

The center channel C is free for commentary, dialogue or other sources (for example, image-related).

Reverberance / envelopment / ambient noise recording

IRT-cross (*Theile 1996*):

4.0 stereo atmo microphone cofiguration:
4 cardioids 25 cm spaced

Use as room microphone in the concert hall is less recommended because the direct sound from the stage is not suppressed!

Reverberance / envelopment / ambient noise recording

IRT-cross (*Theile 1996*):

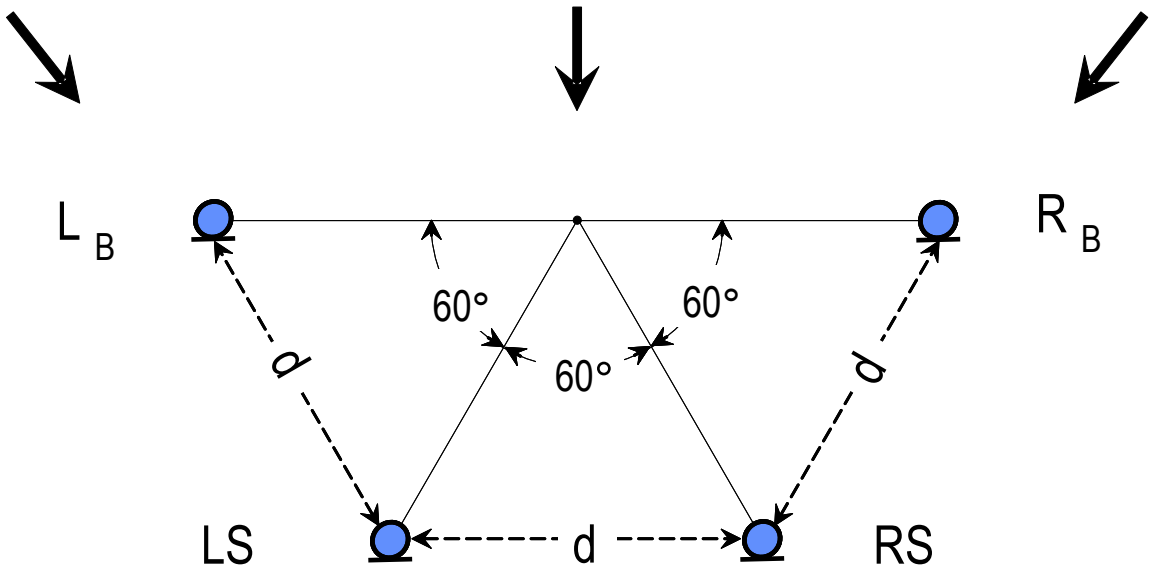
4.0 stereo atmo microphone cofiguration:
4 cardioids 25 cm spaced

Test recordings 20 years ago …

Directional Atmo Microphone (DAM), *Theile, TMT 2012*

Recording the backward and lateral (ambient) sound

Cardioid configuration for suppression of direct sound (e.g. public address live recording)



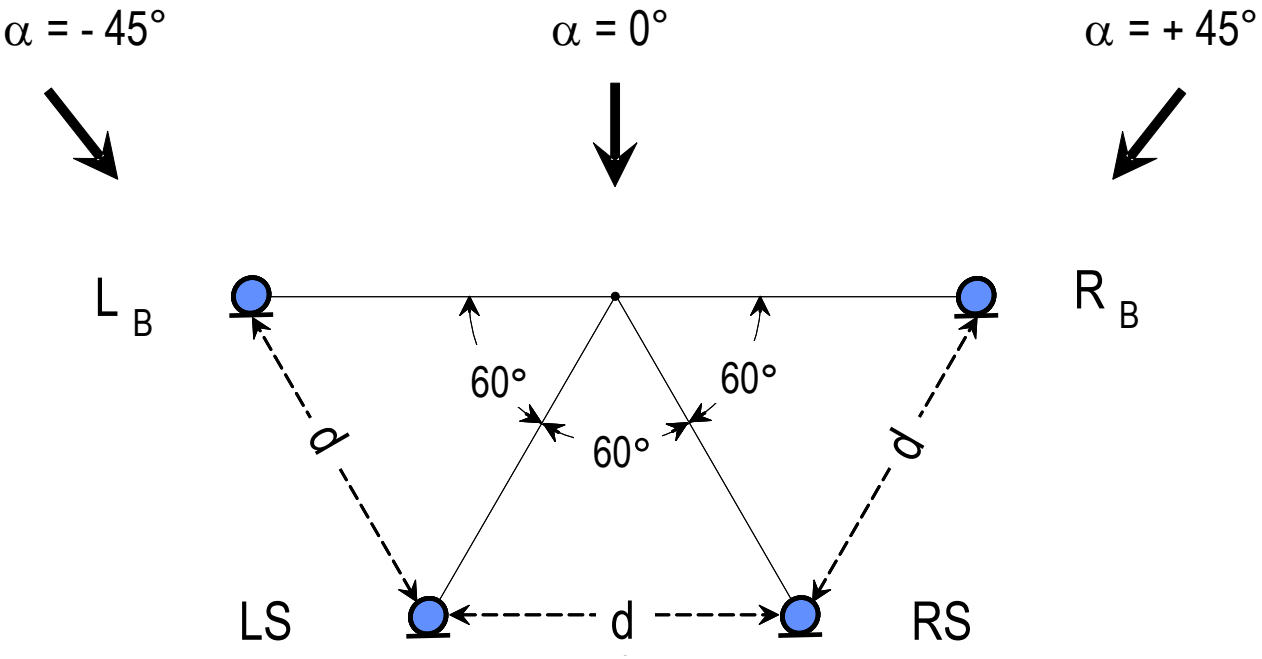
d = 60 cm

All cardioids are looking backwards in order to suppress direct sound from the stage.
The pairs L_B-LS, LS-RS and RS-R_B are working as narrow-spaced a/b ($\Delta L = 0$).

Directional Atmo Microphone (DAM), *Theile, TMT 2012*

Recording the backward and lateral (ambient) sound

Cardioid configuration for suppression of direct sound (e.g. public address live recording)



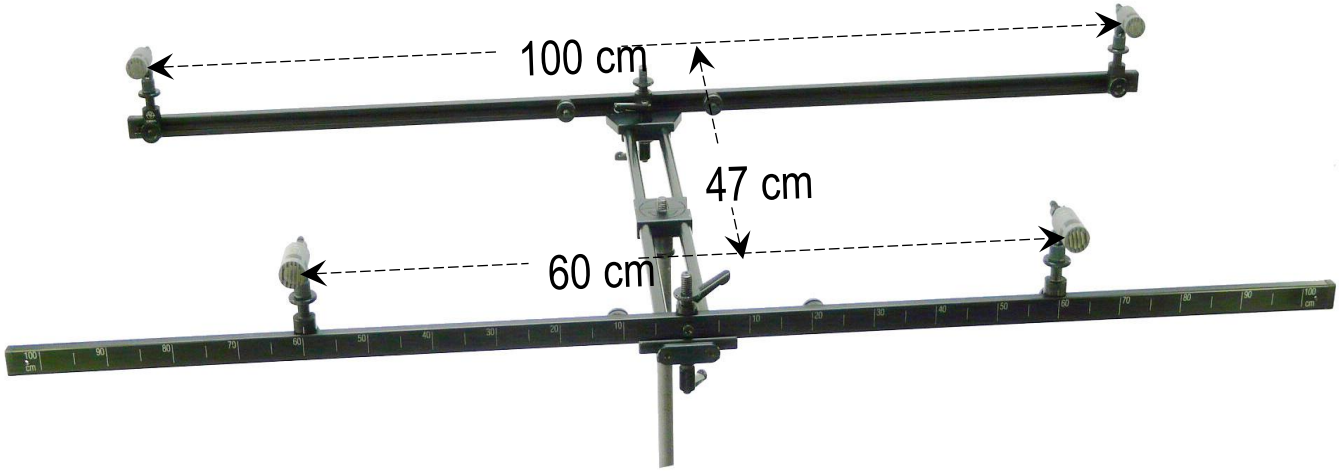
α = ± 30°	→	- 18 dB
α = ± 45°	→	- 15 dB
α = ± 60°	→	- 11 dB
α = ± 90°	→	- 6 dB
α = ± 120°	→	- 3 dB
α = ± 150°	→	- 1dB
α = ± 180°	→	- 0dB

All cardioids are looking backwards in order to suppress direct sound from the stage.
The pairs L_B-LS, LS-RS and RS-R_B are working as narrow-spaced a/b ($\Delta L = 0$).

Directional Atmo Microphone (DAM), *Theile, TMT 2012*

Recording the backward and lateral (ambient) sound

Cardioid configuration for suppression of direct sound (e.g. public address live recording)



Practical setup (example Schoeps)
4 x CCM4, 2 x MAB 1000, 1 x CB/MAB

$\alpha = \pm 30° \rightarrow -18$ dB
$\alpha = \pm 45° \rightarrow -15$ dB
$\alpha = \pm 60° \rightarrow -11$ dB
$\alpha = \pm 90° \rightarrow -6$ dB
$\alpha = \pm 120° \rightarrow -3$ dB
$\alpha = \pm 150° \rightarrow -1$ dB
$\alpha = \pm 180° \rightarrow -0$ dB

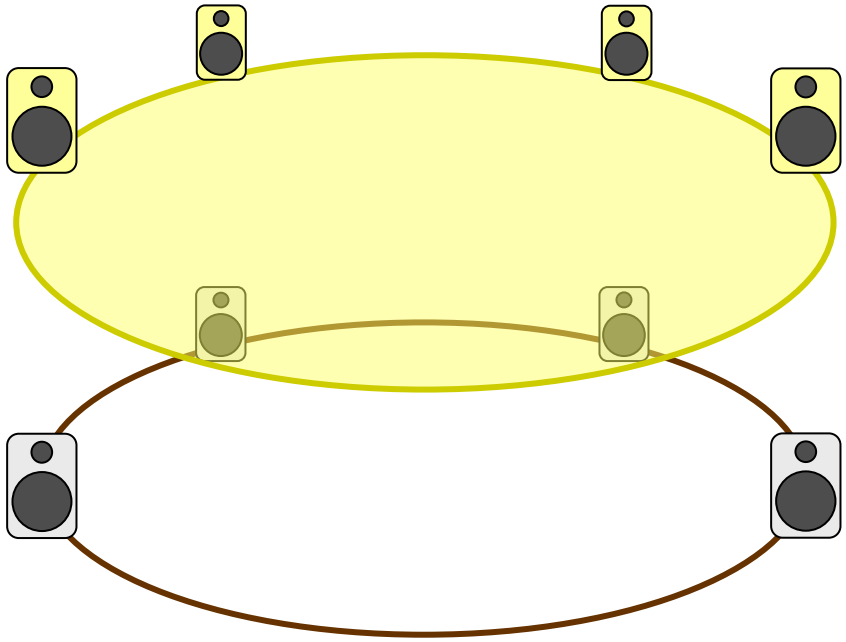
Spatial diffusivity

Spatial diffusivity of reverberation and ambient diffuse noise over the upper hemisphere is possible by means of

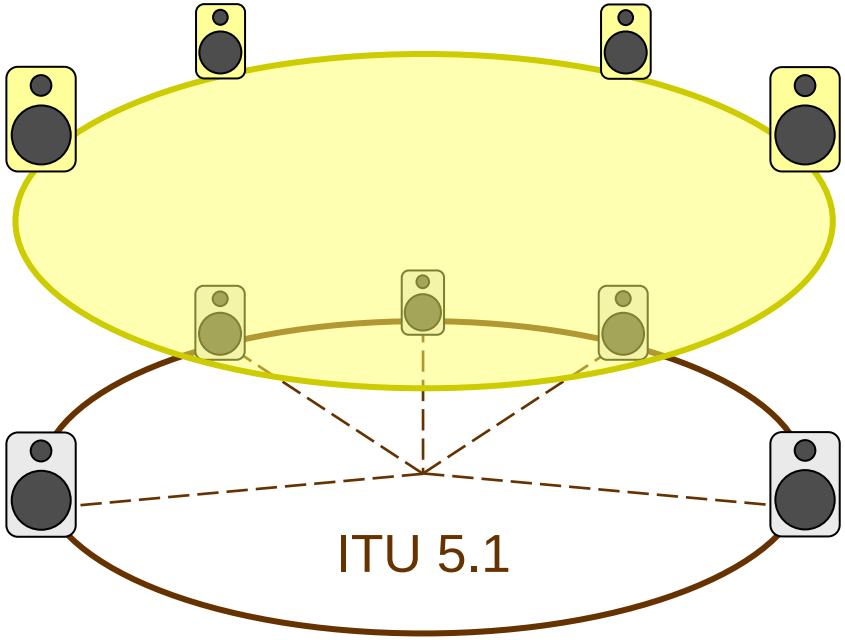
→ eight independent channels, decorrelated signals
→ loudspeakers in each room corner.

→ matches with 3D Audio 9.1 setup

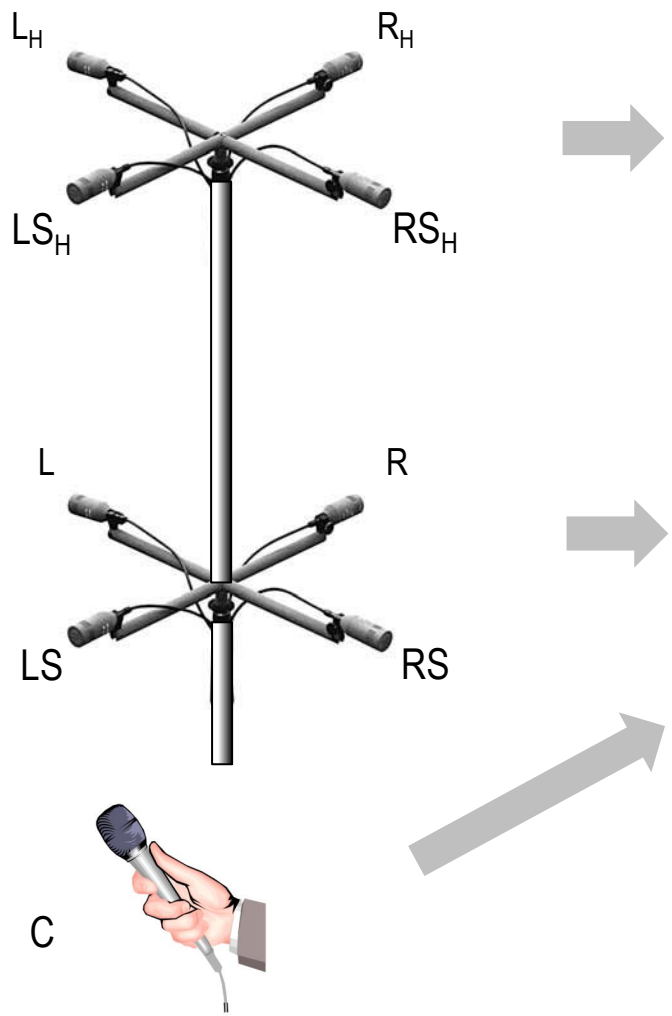
→ adequate microphone technique is required

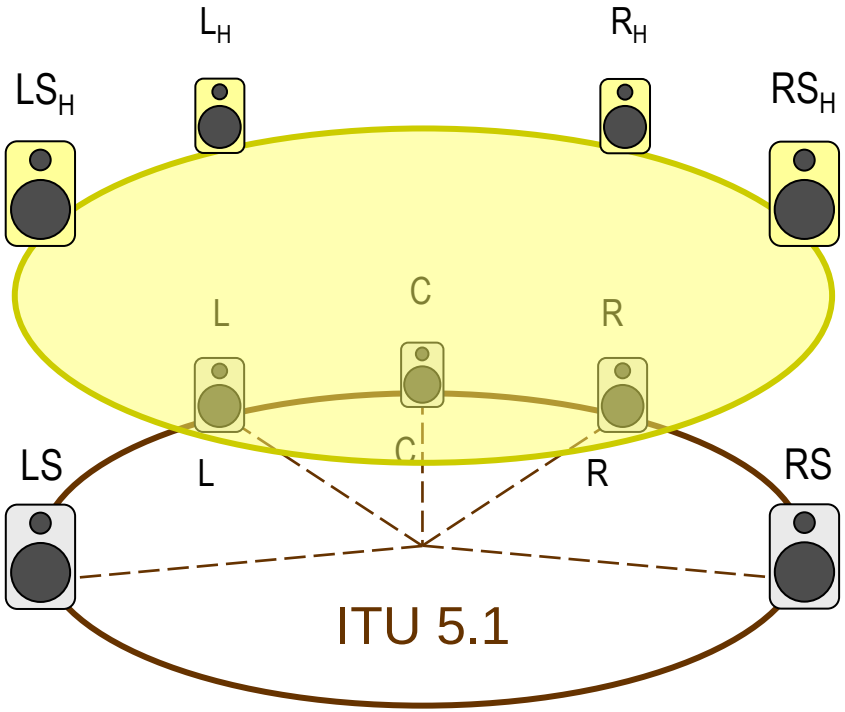


Reverberance / envelopment / ambient noise (atmo)

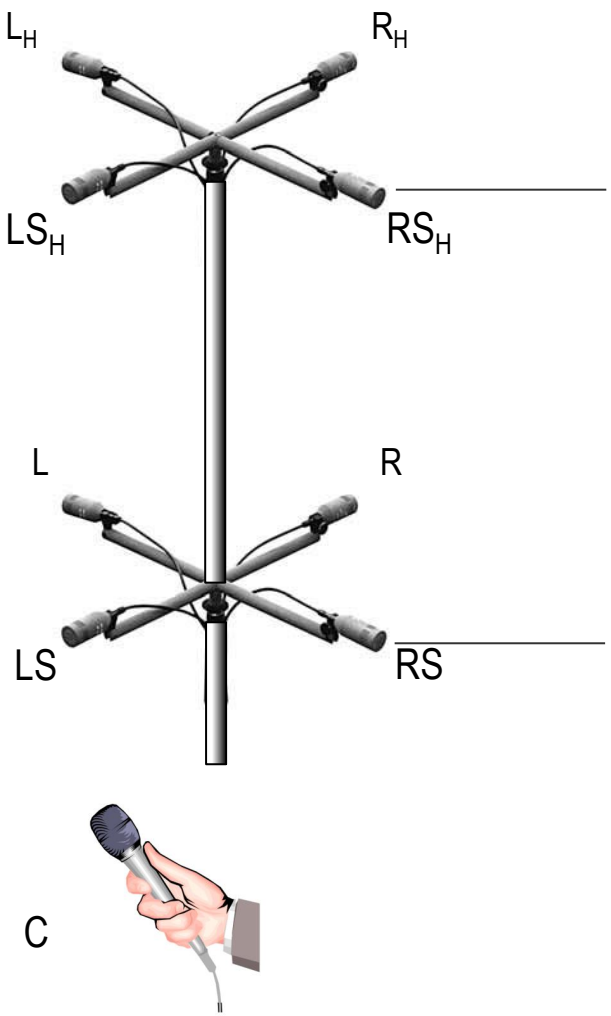


3D IRT cross





3D IRT cross

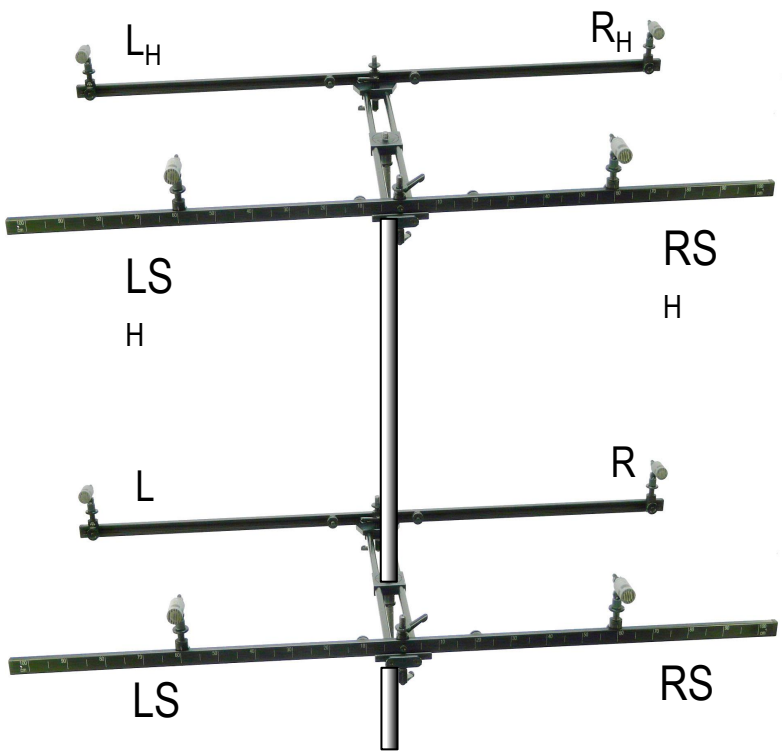


Height layer: IRT cross

Vertical largely-spaced A/B setup

Surround layer: IRT cross

3D Directional Atmo Microphone (3D-DAM)

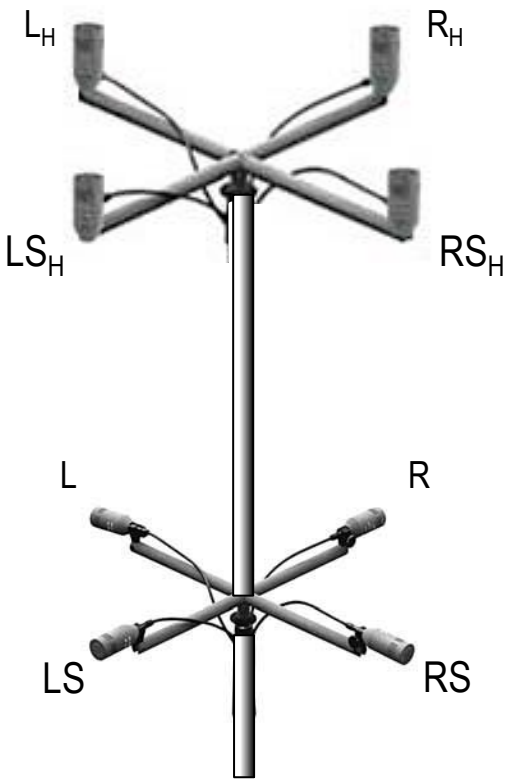


Height layer: DAM

Vertical largely-spaced A/B setup

Surround layer: DAM

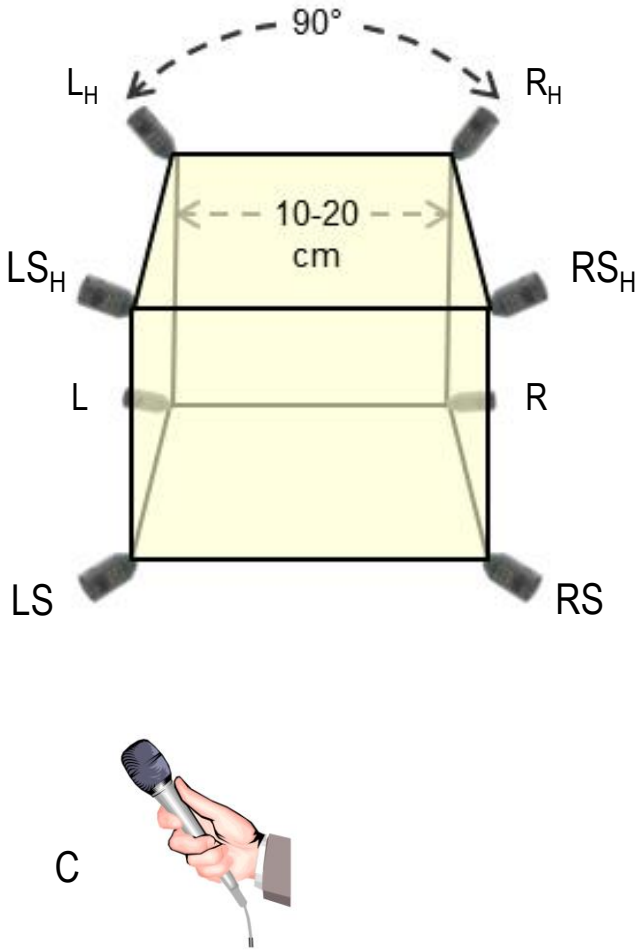
3D IRT cross



Height layer:
Cross of upwards pointing supercardioids
(spacing > 40 cm)

Surround layer: IRT cross

Supercardioid 8 channel hedgehog (*Theile, TMT 2012*)



Each capsule is pointing to one corner of a cuboid

→ Channel separation = 10 dB

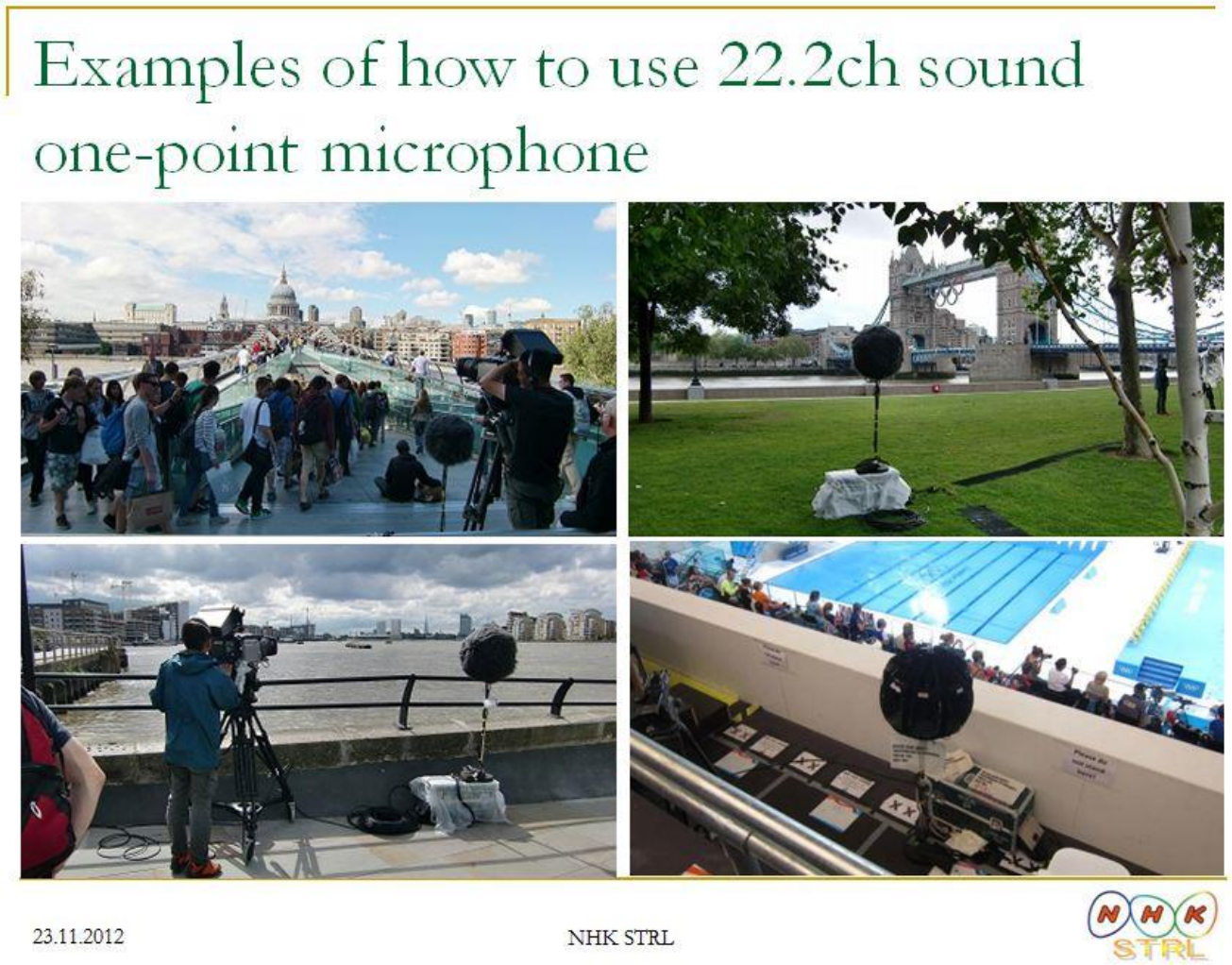
Each of the 12 stereo pairs:
d = 10 – 20 cm
α = 90°

Supercardioid 8 channel hedgehog (*Theile, TMT 2012*)

Schoeps ORTF-3D

Recording the ambience sound

Test recordings
(*Hamasaki*
15th FISM 2012)

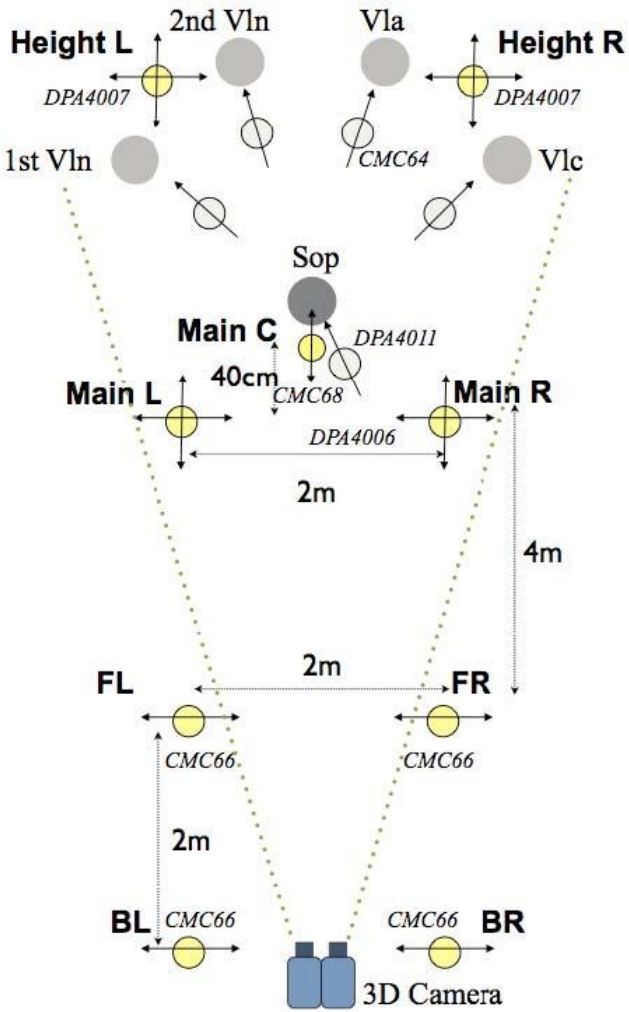


3D suround sound for 3D picture

Matching new requirements for (3D-)digital cinema, game, broadcast, and music productions

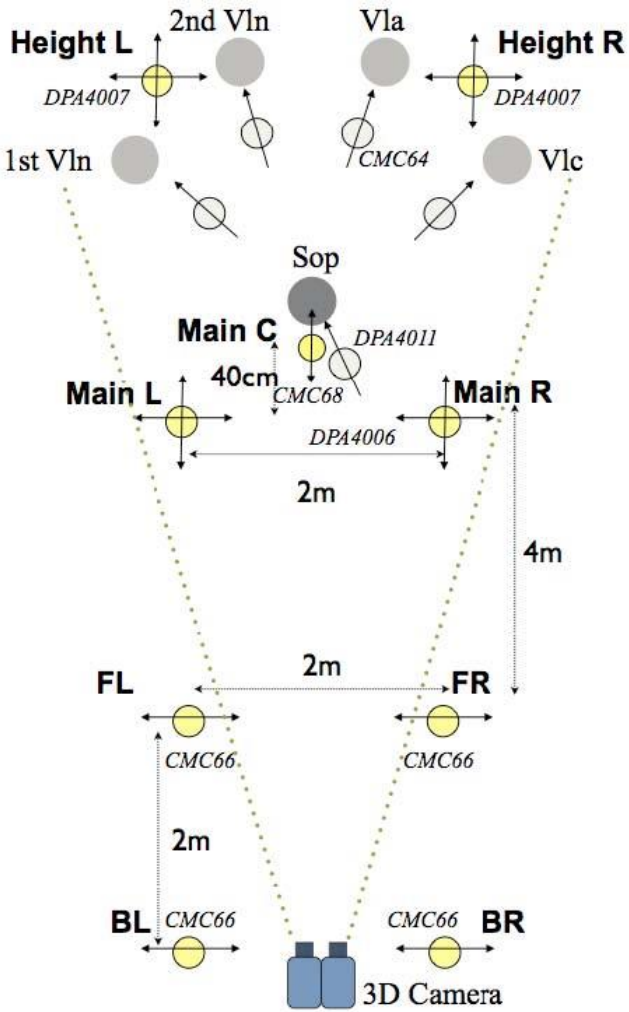
Listening tests [*Kamekawa et al. 2011*]

Microphone setting for soprano with string quartet recording

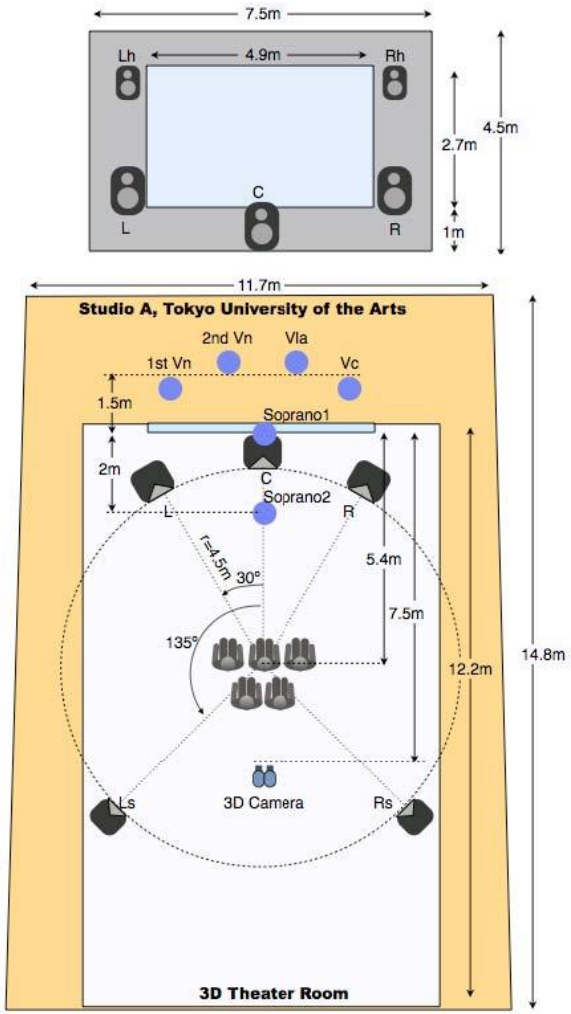


Listening tests [*Kamekawa et al. 2011*]

Microphone setting for soprano with string quartet recording



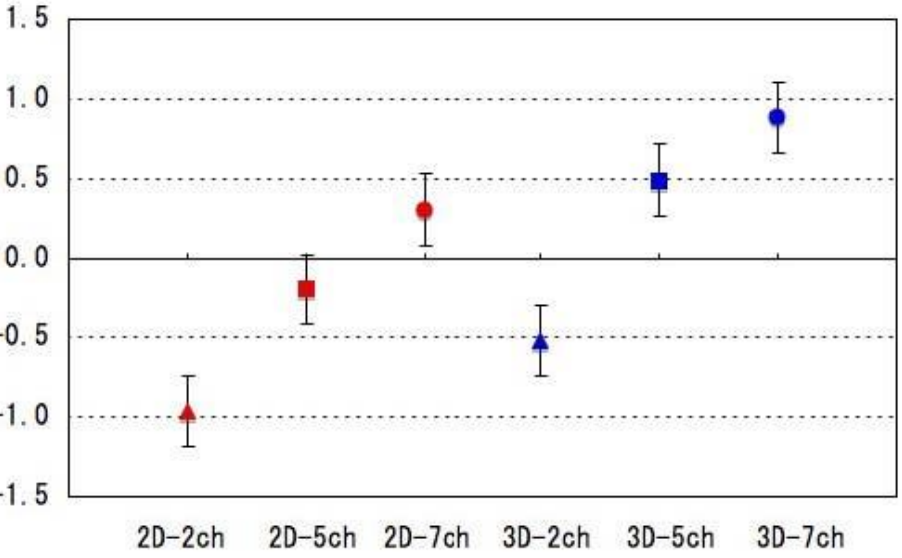
Speaker setting for listening experiment



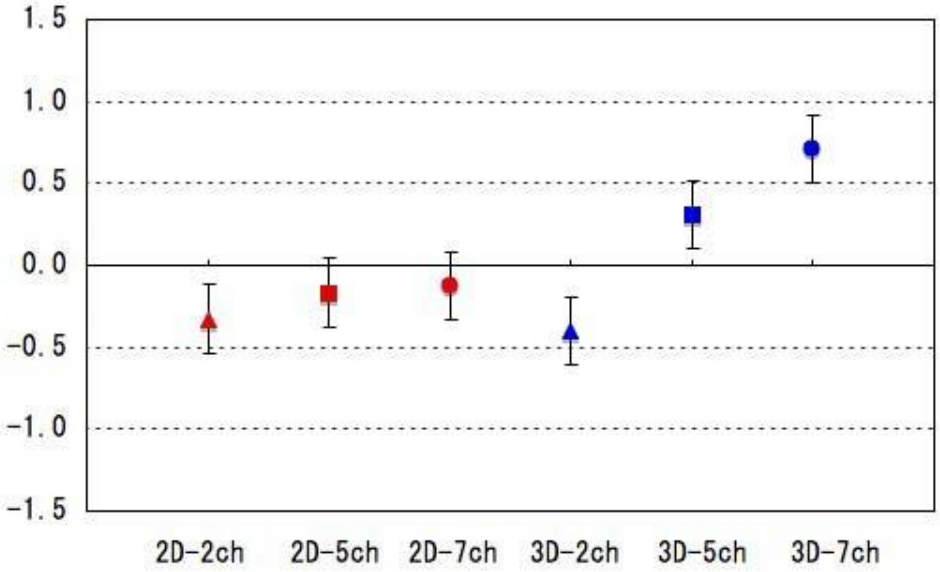
Verband Deutscher Tonmeister e.V.

Results [*Kamekawa et al. 2011*]

Mean score for each combination of all stimuli



Perception of depth



Suitability of visual and audio image

Auro-3D and Dolby Atmos
Two 3D audio standards for immersive audio in the professional cinema

Since 2014 in the home as well:

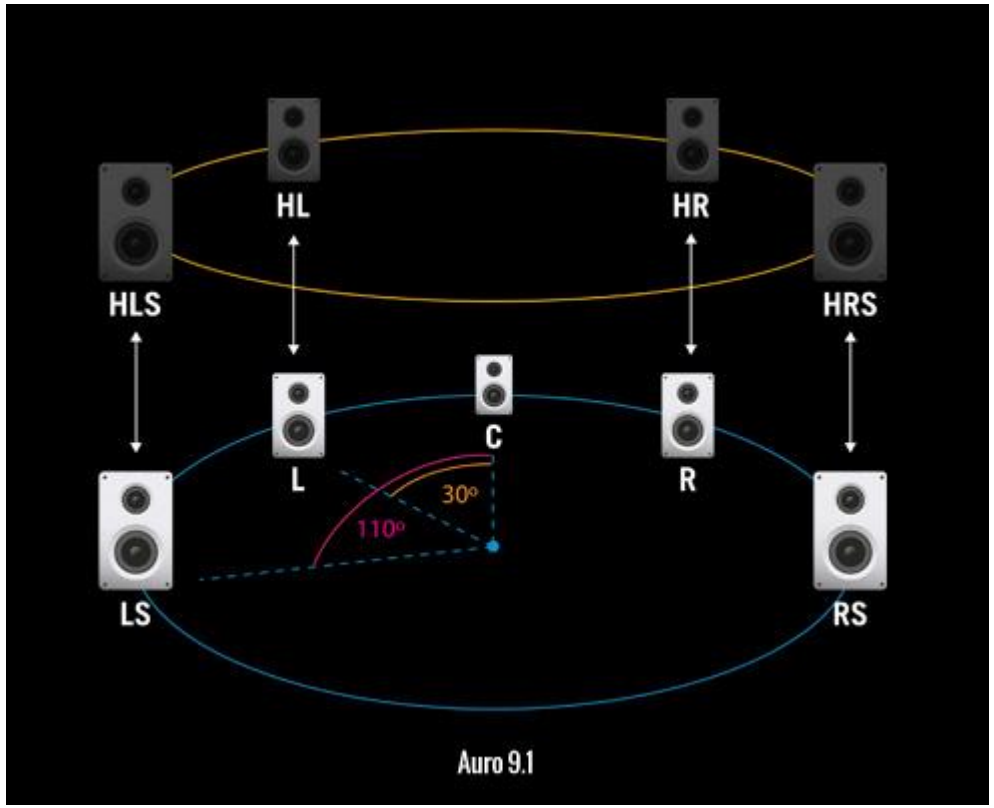
e. g.

Auro-3D® Mensa™ and Crux™ AV Sound Processors

- To feature both Auro-3D and Dolby Atmos
- 3D surround sound to be easily delivered via Blu-ray™ discs
- Playback in 2.0 Stereo, 5.1 / 7.1 Surround, Auro 9.1, 10.1, 11.1, 13.1

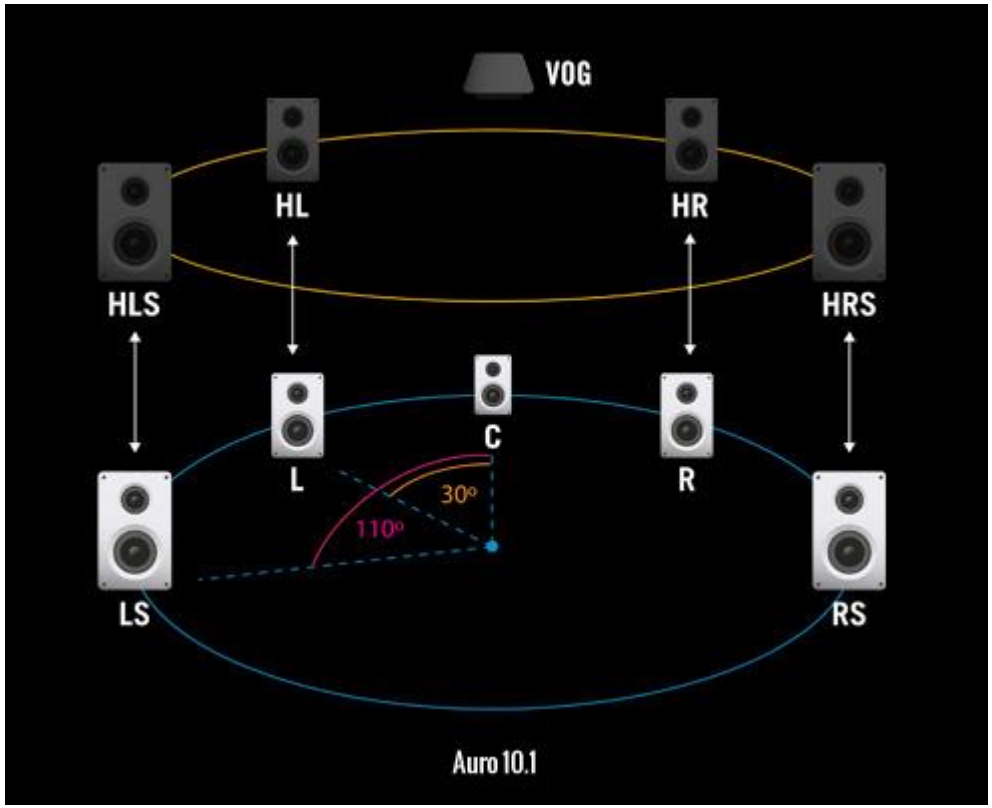
Auro-3D and Dolby Atmos
Two 3D audio standards for immersive audio

Auro 9.1 format



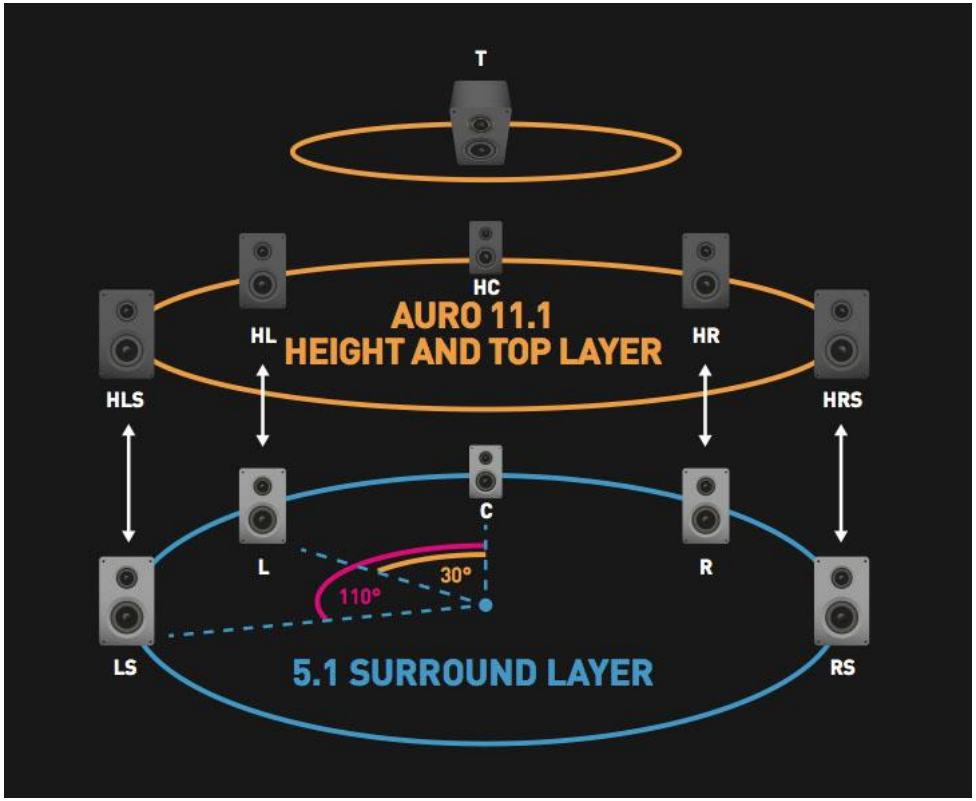
Auro-3D and Dolby Atmos
Two 3D audio standards for immersive audio

Auro 10.1 format



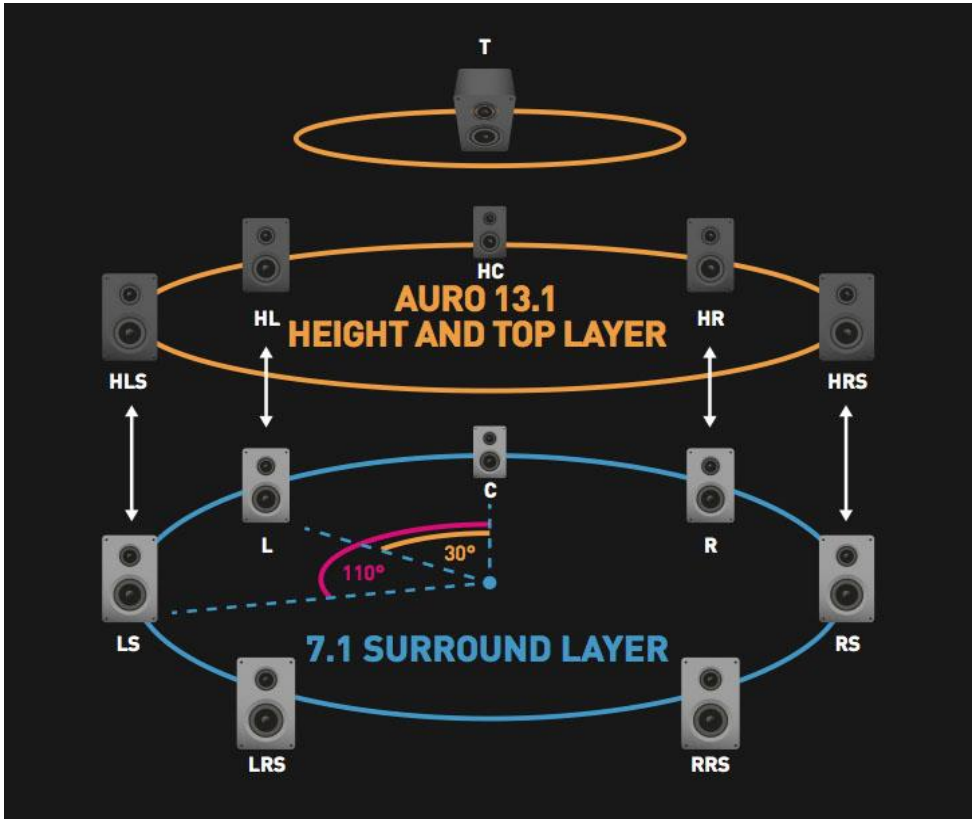
Auro-3D and Dolby Atmos
Two 3D audio standards for immersive audio

Auro 11.1 format



Auro-3D and Dolby Atmos
Two 3D audio standards for immersive audio

Auro 13.1 format



Many sources and more information:

www.hauptmikrofon.de

Handbuch der Tonstudiotechnik

8., völlig überarbeitete und erweiterte Auflage
Erscheinungstermin: Januar 2014

ISBN: 978-3-11-031650-6

1450 Seiten, 2 Bände

Ladenpreis: 99,95 €

Verlag: DE GRUYTER, München

Inhaltsverzeichnis

www.degruyter.com/view/product/183540

Handbuch der Tonstudiotechnik, 8. Auflage

Akustik
Schallquellen
Schallwahrnehmung
Schallwandler
Aufnahme- und Wiedergabeverfahren
Klanggestaltung
Beschallung
Tonsignalspeicherung
Tonmesstechnik

Grundlagen der digitalen Tontechnik
Audiocodierung
Digitale Studioprozesse
Fernsehtontechnik
Digitale Betriebstechnik
Rundfunksysteme
Qualitätssicherung
Arbeitssicherheit und Gesundheitsschutz
Fachwörter und Abkürzungen Englisch - Deutsch
Sachregister

Handbuch der Tonstudiotechnik, 8. Auflage

Grundlagenwerk für Ausbildung und Beruf.
Seit über 40 Jahren ein Standardwerk für den Bereich der professionellen Tonstudiotechnik.

Grundlagen und aktuelle Entwicklungen mit Beiträgen von 36 ausgewiesenen Experten ihres Fachgebiets:

Peter Arasin
Hannah Baumgartner
Stefan Bock
Nikolaus Büttner
Florian Camerer
Michael Dickreiter
Ernst Dohlus
Nikolaus Färber
Bernhard Feiten
Jürgen Goeres-Petry
Wolfram Graul
Bernhard Grill
Klaus Hackbart
Christian Hartmann
Klaus M. Heidrich
Roger Heimann
Siegbert Herla
Wolfgang Hoeg
Thomas Lauterbach
Frank Lott
Hans-Joachim Maempel
Stephan Maniak
Gernot Meyer-Schwarzenberger
Klemens Nicklaus
Heinz Peter Reykers
Götz Romahn
Michael Schäfer
Thomas Schierbaum
Bernhard Schullan
Henrik Schulze
Karl M. Slavik
Gerhard Spikofski
Ralf Steuck
Gerhard Stoll
Günther Theile
Ralf Zuleeg