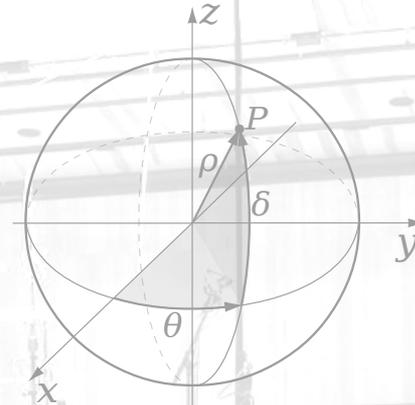


Psychoakustik des räumlichen Hörens

Was bringt es uns, Klänge orten zu können?

Wie wir Klänge orten:

- vor uns
- hinter uns
- in der Höhe
- Entfernung



Themenübersicht:

Klänge im Raum erzeugen:

- einfach aber wirkungsvoll: **“Stell einen Lautsprecher hin!”**
- psychoakustische Reize erzeugen: **Stereophonie/VBAP**
- Ohrsignale nachbilden: **Binaural-Synthese**
- ein Schallfeld im Raum nachbilden: **Wellenfeldsynthese, Higher-order Ambisonics**

**...und was uns das
in der Praxis bringt.**



Jörn Nettingsmeier

<nettingsmeier@tonmeister.de>

freiberuflicher Tontechniker und Berater aus Essen

Spezialgebiet: Ambisonics

Meister für Veranstaltungstechnik (Bühne/Studio)
Mitglied des Vorstands und Referatsleiter "Beschallung"
im Verband deutscher Tonmeister
AES-Mitglied

Was bringt es uns, Klänge zu orten?



Was bringt es uns, Klänge zu orten?





Überleben.

Das Gehör erstreckt sich rings um uns herum.

Es warnt uns vor dem, was wir **noch nicht** sehen können:

- ein Raubtier?
- ein Auto an einer Kreuzung?

Was bringt es uns, Klänge zu orten?





Orientierung.

Wir hören

- die Beschaffenheit des Bodens unter uns,
- die Größe des Raums, in dem wir sind,
- die Position von Wänden und Hindernissen.

Was bringt es uns, Klänge zu orten?





Sozialleben.

Wir können unsere Freunde rufen, bevor sie uns sehen können.

Wir verfolgen ein Gespräch in lauter Umgebung, indem wir gezielt in eine bestimmte Richtung hören.

Was bringt es uns, Klänge zu orten?



**Bei Verstand
bleiben.**

Wir können
unsere Ohren
nicht schließen.
Wir können nicht
"nicht hören".

Richtungshören
erlaubt uns
Konzentration
auf Wichtiges,
auch bei Lärm.

Was bringt es uns, Klänge zu orten?



Kreativität.

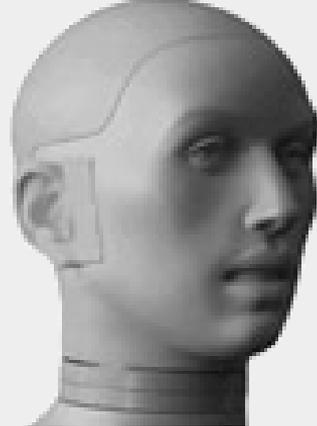
Gute **Ortung** hält den Fokus auf der Bühne.

Einhüllung erzeugt Emotionen, die die Grenzen der Bühne überwinden.

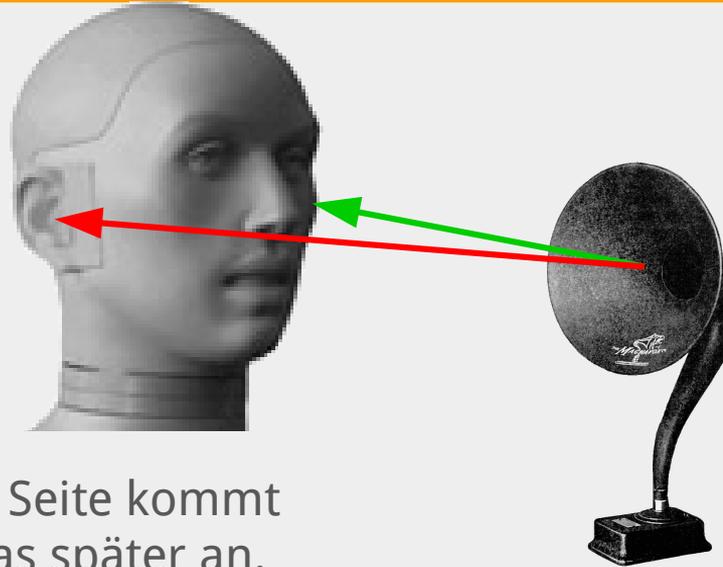
Surround-Effekte versetzen uns mitten in die Handlung.

Komplexe Klangszenen verlangen klare **Lokalisation**.

Wie wir Klänge orten:



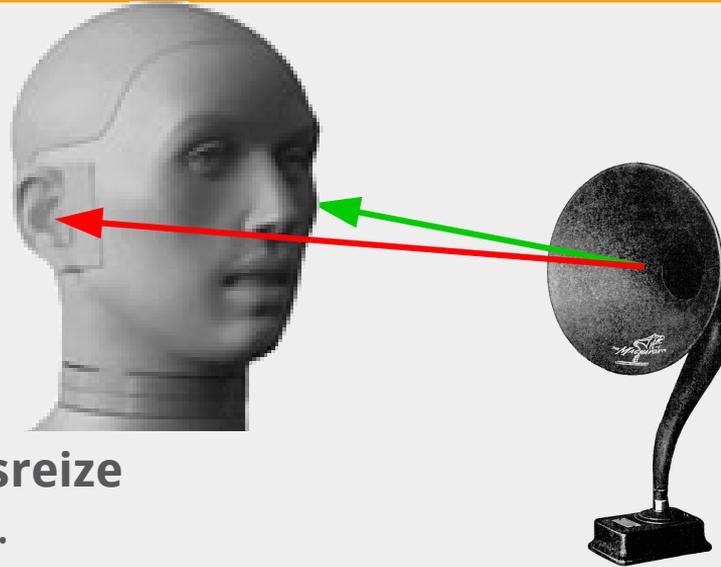
Ein Klangereignis direkt von vorn erreicht die Ohren mit dem gleichen Pegel und zur gleichen Zeit.



Ein Klangereignis von der Seite kommt am entfernteren Ohr etwas später an. Diese Laufzeitdifferenz nennt man **inter-aural time difference (ITD)**.

Das entferntere Ohr liegt im akustischen Schatten des Kopfes, wo der Pegel geringer ist. Dieser Pegelunterschied heißt **inter-aural level difference (ILD)**.

Laufzeitdifferenz (**ITD**)
und Pegeldifferenz (**ILD**)
sind die wichtigsten und
genauesten **Lokalisationsreize**
des menschlichen Gehörs.



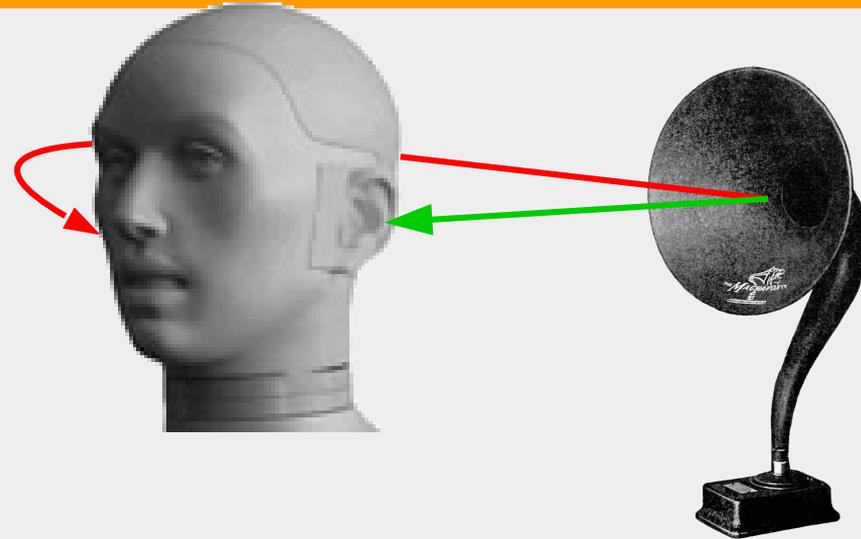
Mit ihrer Hilfe können wir Klänge von vorn sehr exakt orton,
im günstigsten Fall auf bis zu 2° genau.

Klänge hinter uns erzeugen dieselben Laufzeit- und Pegeldifferenzen (**ILD** und **ITD**) wie vorn.

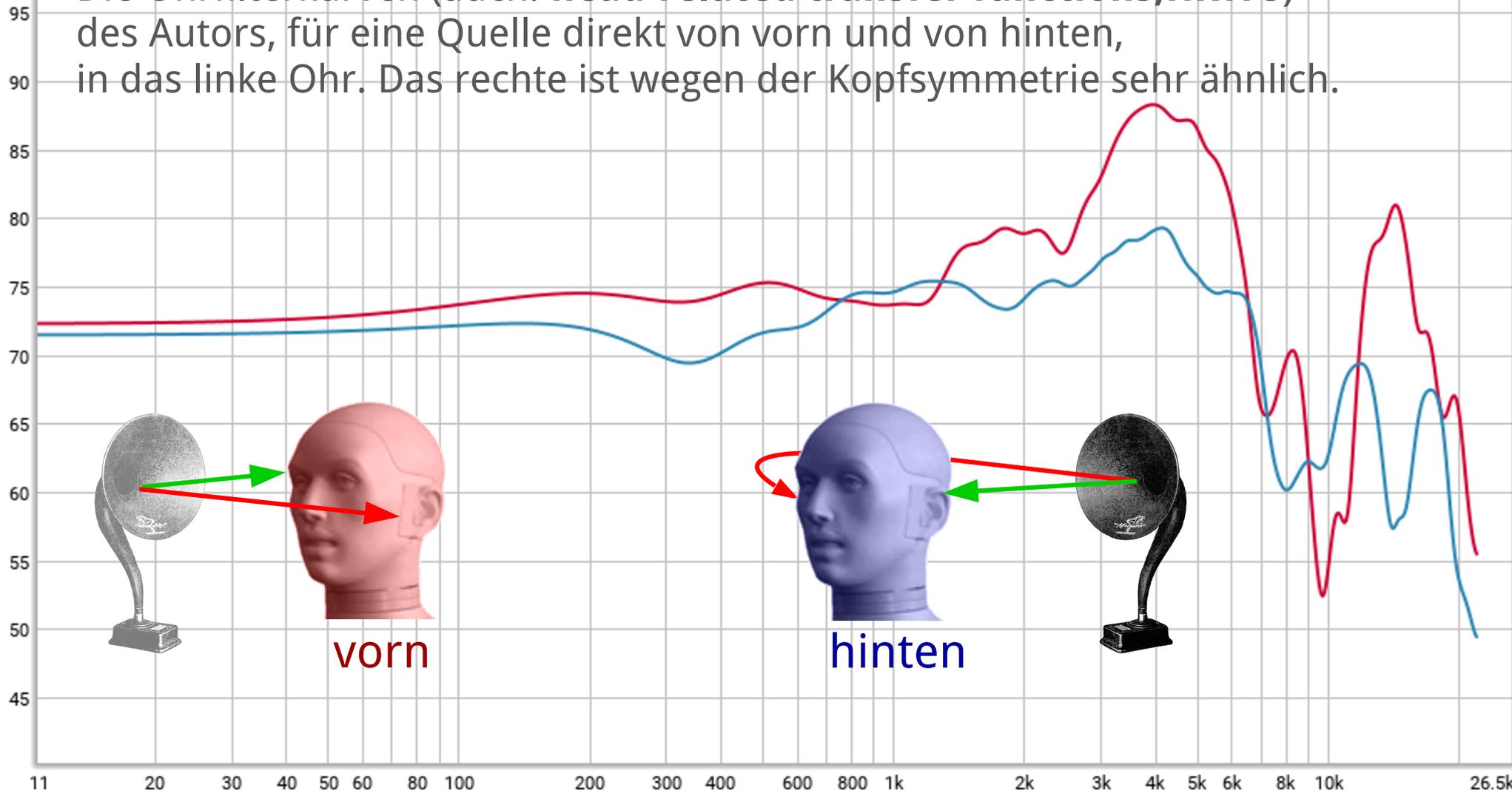
Wir können sie unterscheiden, weil unser Aussenohr wie ein richtungsabhängiger Equalizer wirkt. Es verändert das Klang**spektrum**.

Unser Gehirn nimmt diese Verfärbungen (**spectral cues**) als Richtungsinformation wahr.

Diese krasse Klangverfärbung hören wir nicht bewusst, können sie aber mit einem Mikrofon im Gehörgang messen:



Die Ohrfilterkurven (auch: **head-related transfer functions, HRTFs**) des Autors, für eine Quelle direkt von vorn und von hinten, in das linke Ohr. Das rechte ist wegen der Kopfsymmetrie sehr ähnlich.



vorn

hinten

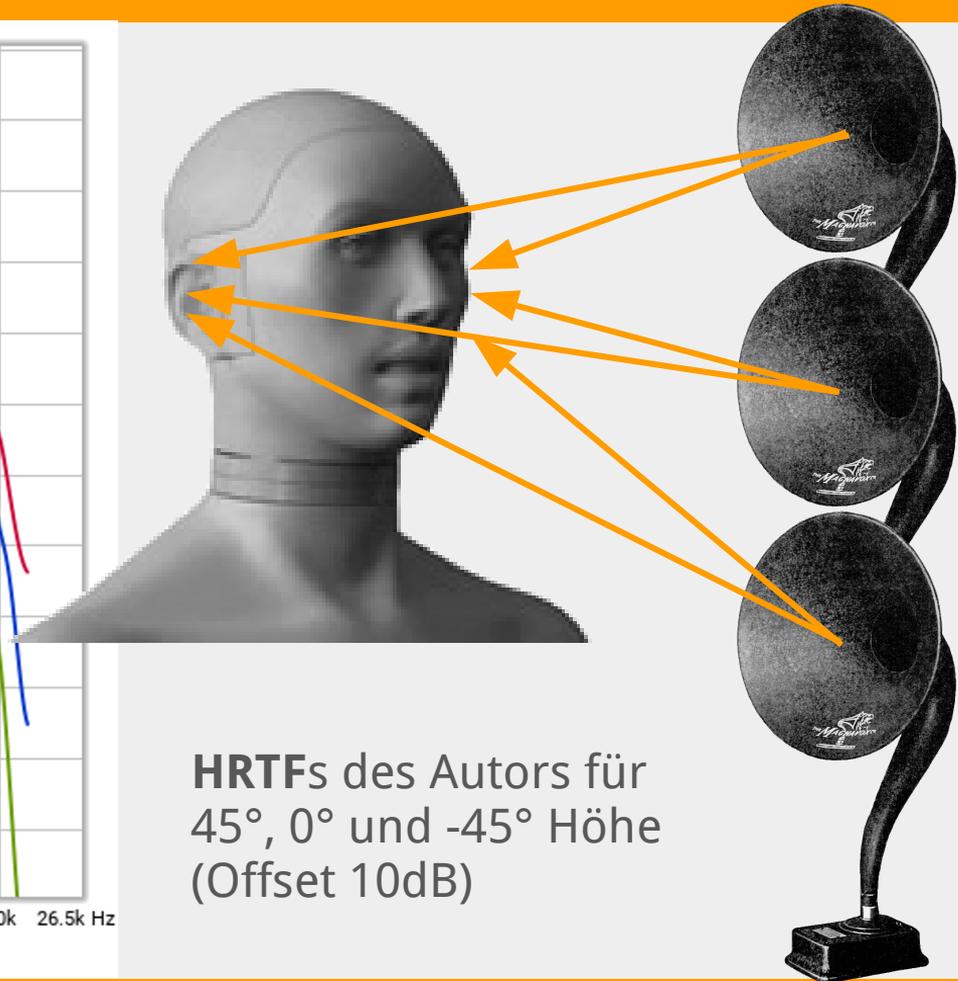
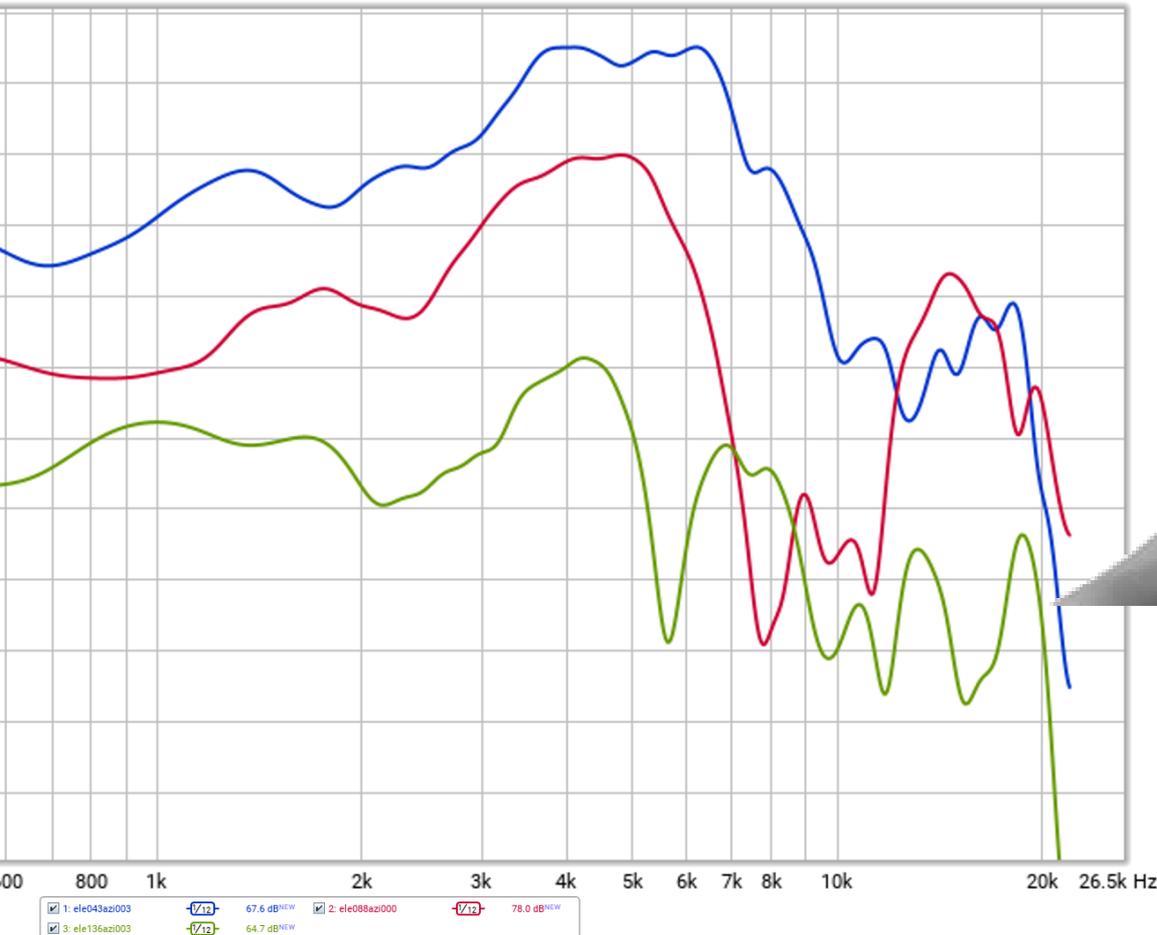
Wenn eine Quelle nach oben oder unten wandert, ändern sich **ILD** und **ITD** nicht.

Wieder müssen wir uns auf den richtungsabhängigen Equalizer unserer Ohrmuscheln verlassen.

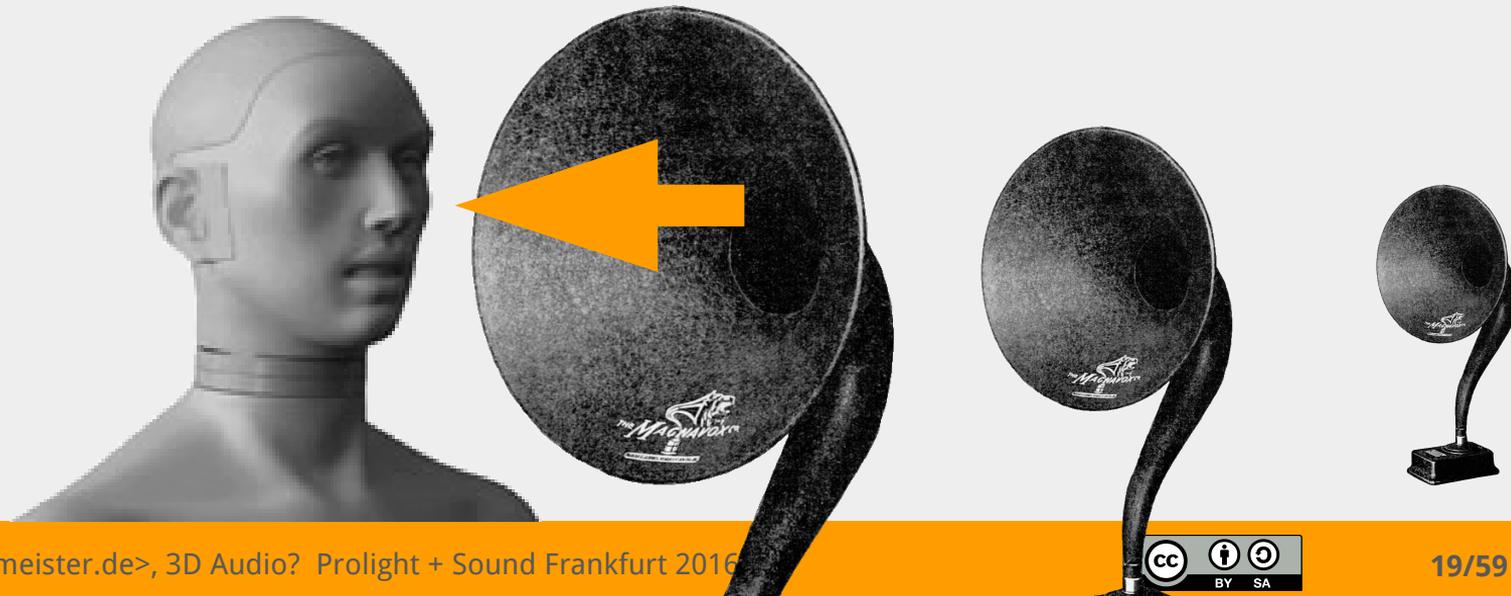
Dieser **“spectral cue”** ist nicht so eindeutig, und deshalb ist unsere Ortung in der Höhe weniger präzise.



Wie wir Klänge orton: Höhe

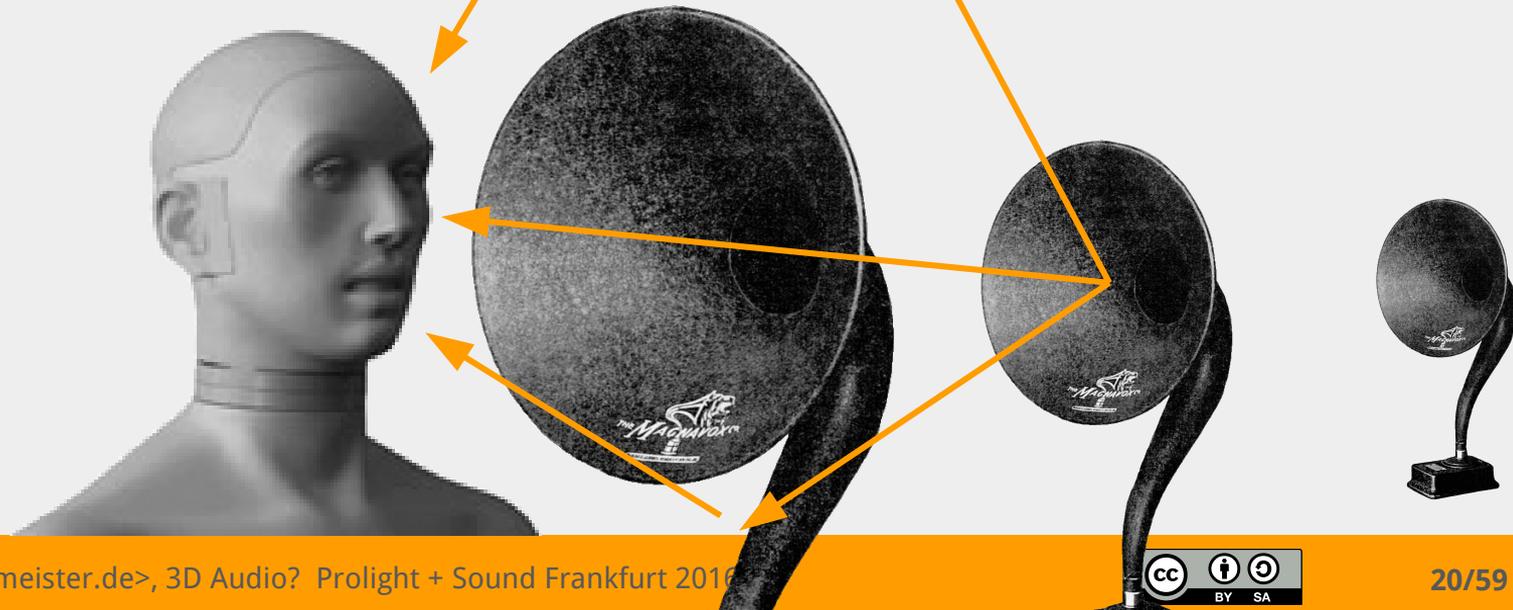


Entfernung können wir nur im Nahbereich direkt wahrnehmen:
unsere Ohren haben, wie jedes Richtmikrofon, einen **Nahbesprechungseffekt**.
Nahe Geräusche haben etwas mehr Bass.



Entfernung können wir nur im Nahbereich direkt wahrnehmen: unsere Ohren haben, wie jedes Richtmikrofon, einen **Nahbesprechungseffekt**. Nahe Geräusche haben etwas mehr Bass.

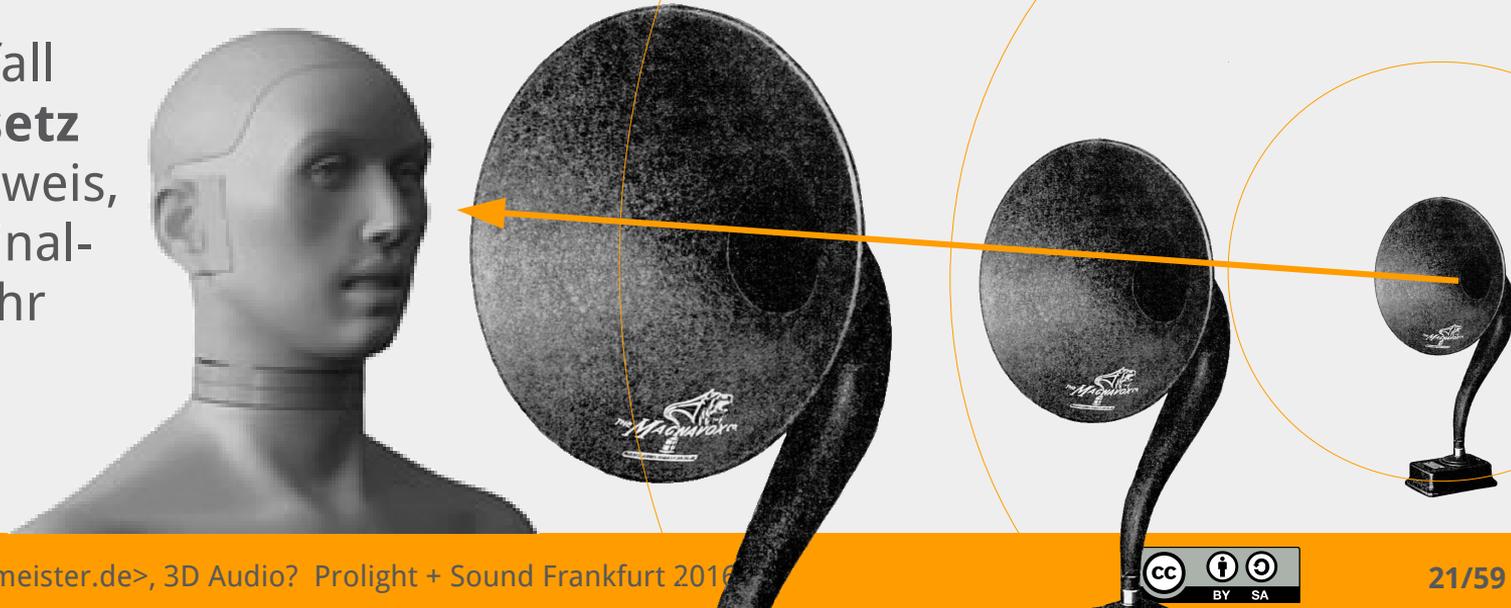
In Räumen können wir Entfernungen durch das **Muster der frühen Reflexionen** und der **Balance von Direktschall und Nachhall** schätzen.



Entfernung können wir nur im Nahbereich direkt wahrnehmen: unsere Ohren haben, wie jedes Richtmikrofon, einen **Nahbesprechungseffekt**. Nahe Geräusche haben etwas mehr Bass.

In Räumen können wir Entfernungen durch das **Muster der frühen Reflexionen** und der **Balance von Direktschall und Nachhall** schätzen.

Auch der Pegelabfall nach dem **1/r-Gesetz** gibt uns einen Hinweis, wenn wir die Original-Lautstärke ungefähr kennen.



Entfernung können wir nur im Nahbereich direkt wahrnehmen: unsere Ohren haben, wie jedes Richtmikrofon, einen **Nahbesprechungseffekt**. Nahe Geräusche haben etwas mehr Bass.

In Räumen können wir Entfernungen durch das **Muster der frühen Reflexionen** und der **Balance von Direktschall und Nachhall** schätzen.

Auch der Pegelabfall nach dem **1/r-Gesetz** gibt uns einen Hinweis, wenn wir die Original-Lautstärke ungefähr kennen.



Am besten hören wir in der Horizontalebene:

ILD und **ITD** (Zeit- und Pegeldifferenz an den Ohren) sind auf bis zu 2° genau.

Vorn und hinten unterscheiden wir durch unsere **Ohrfilter (HRTFs)**.



Auch für die Höhe haben wir nur die **Klangfarbenänderung** der Ohrfilter.

Deshalb orten wir Höhe und vorn/hinten weniger gut.

Um Entfernung einzuschätzen, brauchen wir **Raumhall**.

Klänge im Raum erzeugen:

*“If you want sound to come from a particular place,
put a loudspeaker there!”*

- Curtis Roads

*“If you want sound to come from a particular place,
put a loudspeaker there!”*

- Curtis Roads

- Einfach aber effektiv.
- Alle psychoakustischen Reize sind auf jeden Fall korrekt.
- Wir orten den Lautsprecher perfekt.
- Im Theater oft das Mittel der Wahl.

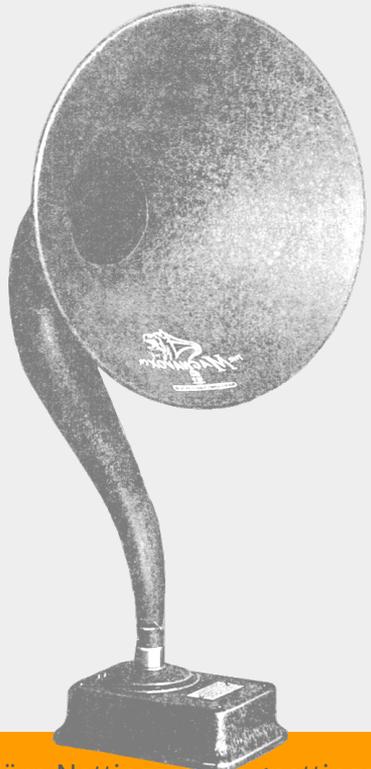
*“If you want sound to come from a particular place,
put a loudspeaker there!”*

- Curtis Roads

Aber:

- Wer will Lautsprecher hören?
- Was tun, wenn wir Klänge aus vielen Richtungen brauchen?
- Was, wenn ein Klang sich bewegen soll?

Klänge im Raum erzeugen: Stereophonie

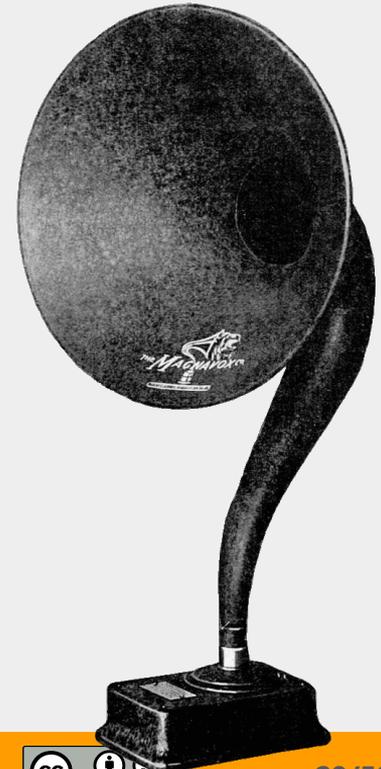
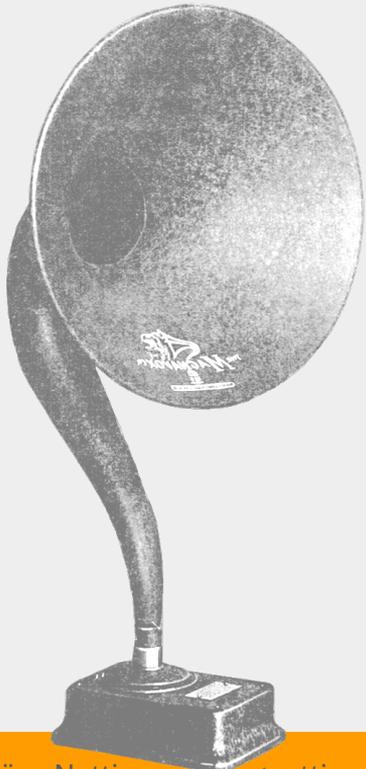


Stereophonie erzeugt die Lokalisationsreize **ILD** und **ITD** zwischen **Lautsprecherpaaren**.

Sie kann **Phantomschallquellen** bilden: Hörereignisse, die scheinbar zwischen den Lautsprechern entstehen.

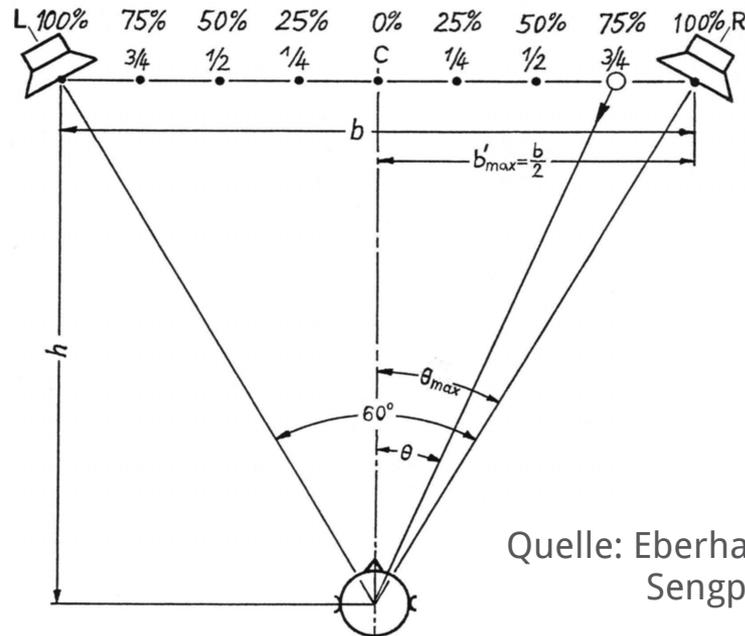
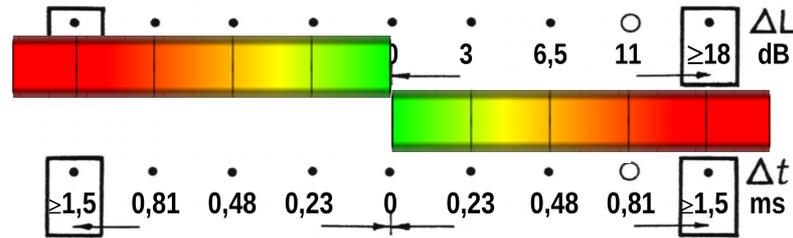
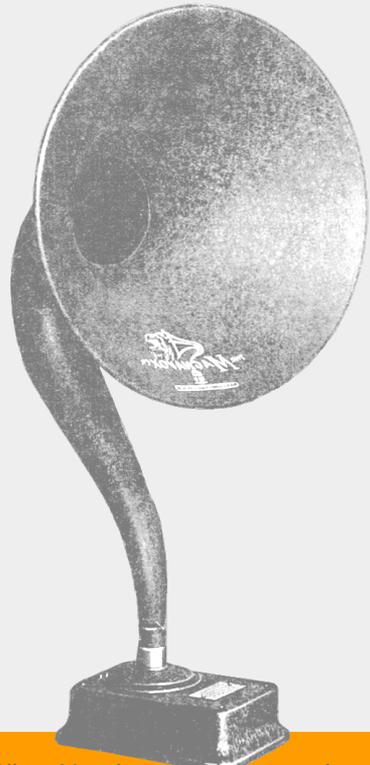
Stereophonie ist nicht auf 2-Kanal-Systeme beschränkt!

Die meisten kanalbasierten Systeme (5.1, 7.1, 22.2, Auro-3D u.a.) benutzen stereophone Lokalisation, zumindest im vorderen Bereich.



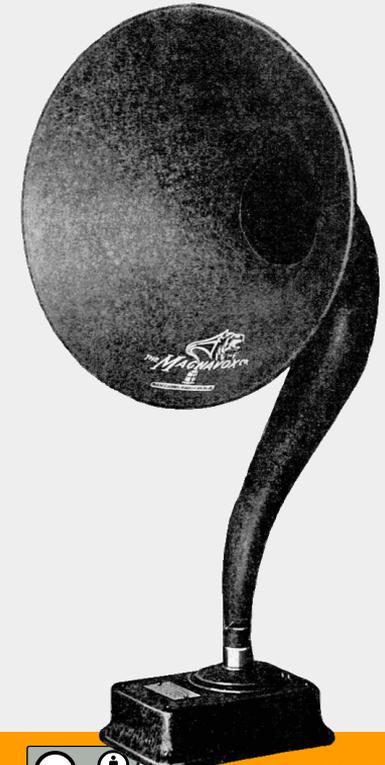
Klänge im Raum erzeugen: Stereophonie

Das Zeitfenster
für ITD ist 1,5 ms.

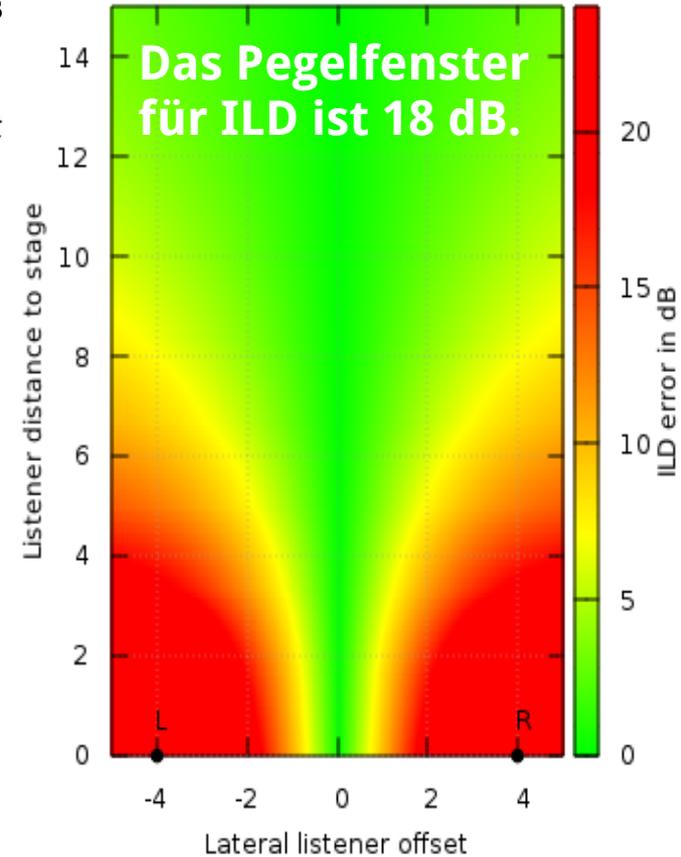
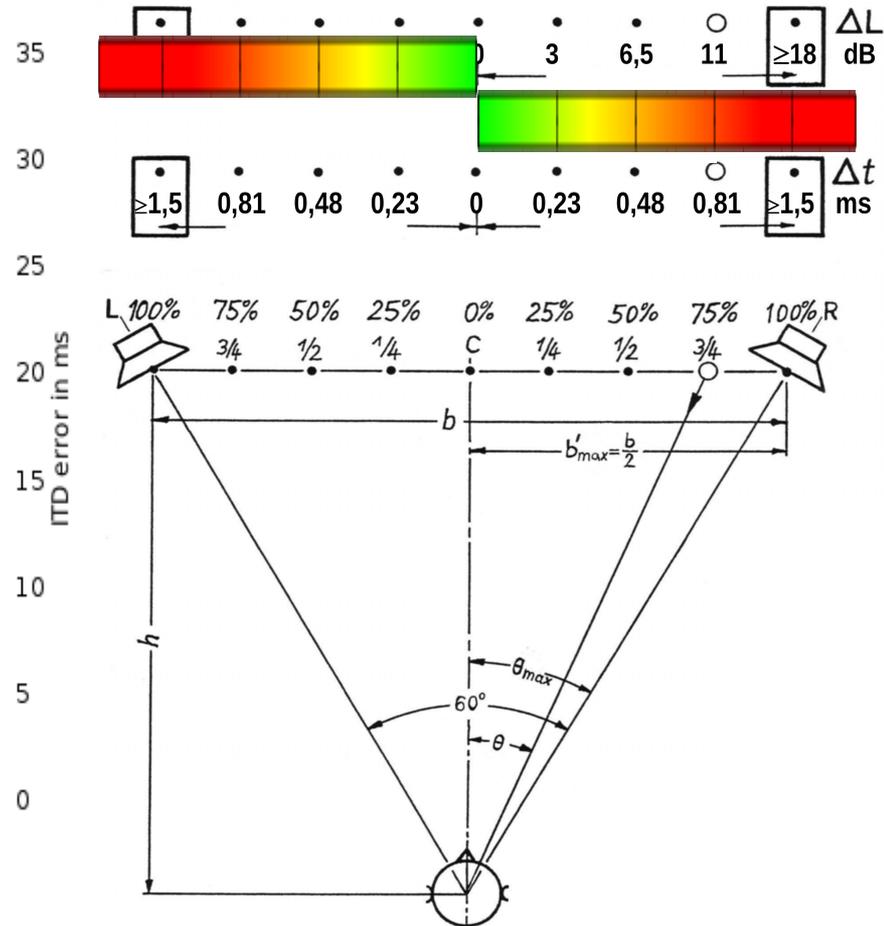
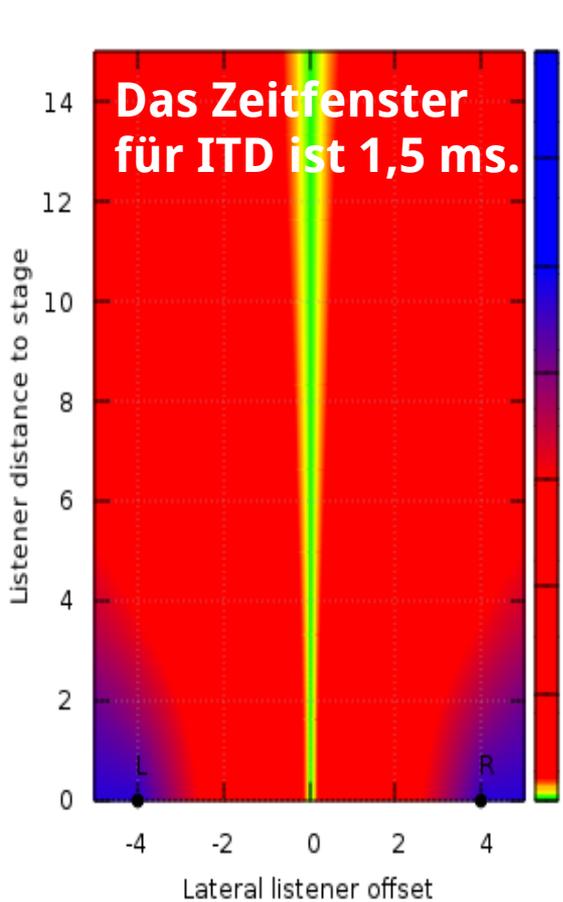


Quelle: Eberhard
Sengpiel.

Das Pegelfenster
für ILD ist 18 dB.



Klänge im Raum erzeugen: Stereophonie

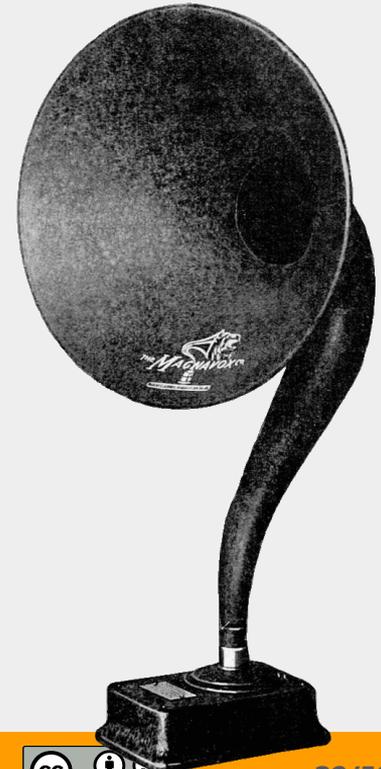
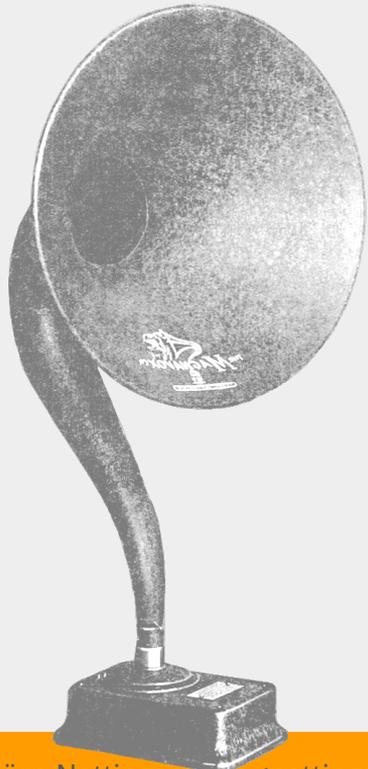


Vorteile:

- sehr gute Klangfarbenwiedergabe (keine Verfärbung)
 - gute Effizienz: geringe Bandbreiten- und Hardwareanforderungen

Nachteile:

- instabile Ortung an den Seiten, keine Höhe
- bewegte Quellen "kleben" an Lautsprechern, ändern ihre Klangfarbe und Schärfe
- sehr kleiner Hörbereich, besonders problematisch in der Beschallung



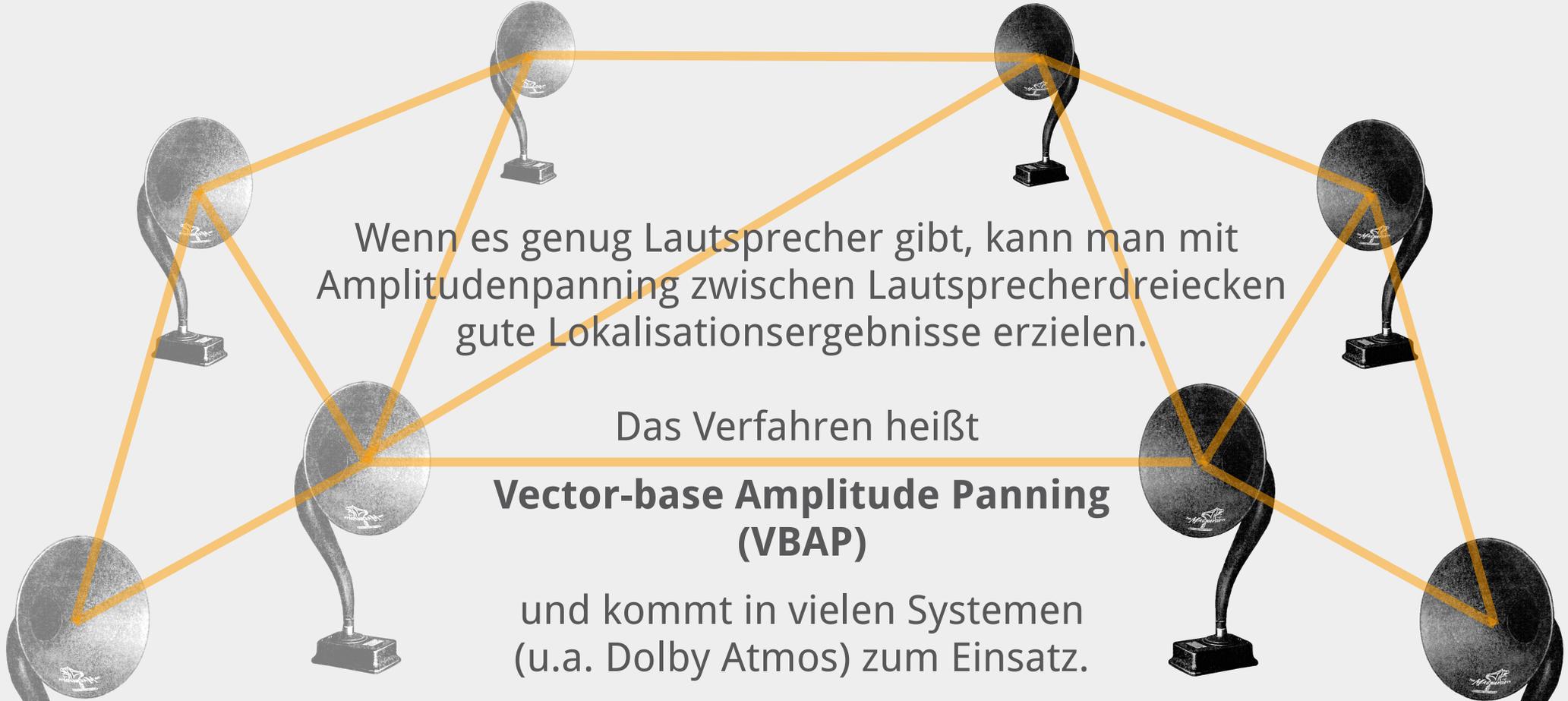


Horizontale Stereophonie kann durch diskrete **Höhenkanäle** erweitert werden (NHK 22.2, Auro-3D).

Die vertikale Lokalisation ist schlecht - keine **ILD** und **ITD**, falsche **spectral cues**: keine vertikalen Phantomschallquellen!

Dafür vergrößert sich der Hörbereich, Klangfarbe und Raumeindruck werden als natürlicher empfunden.

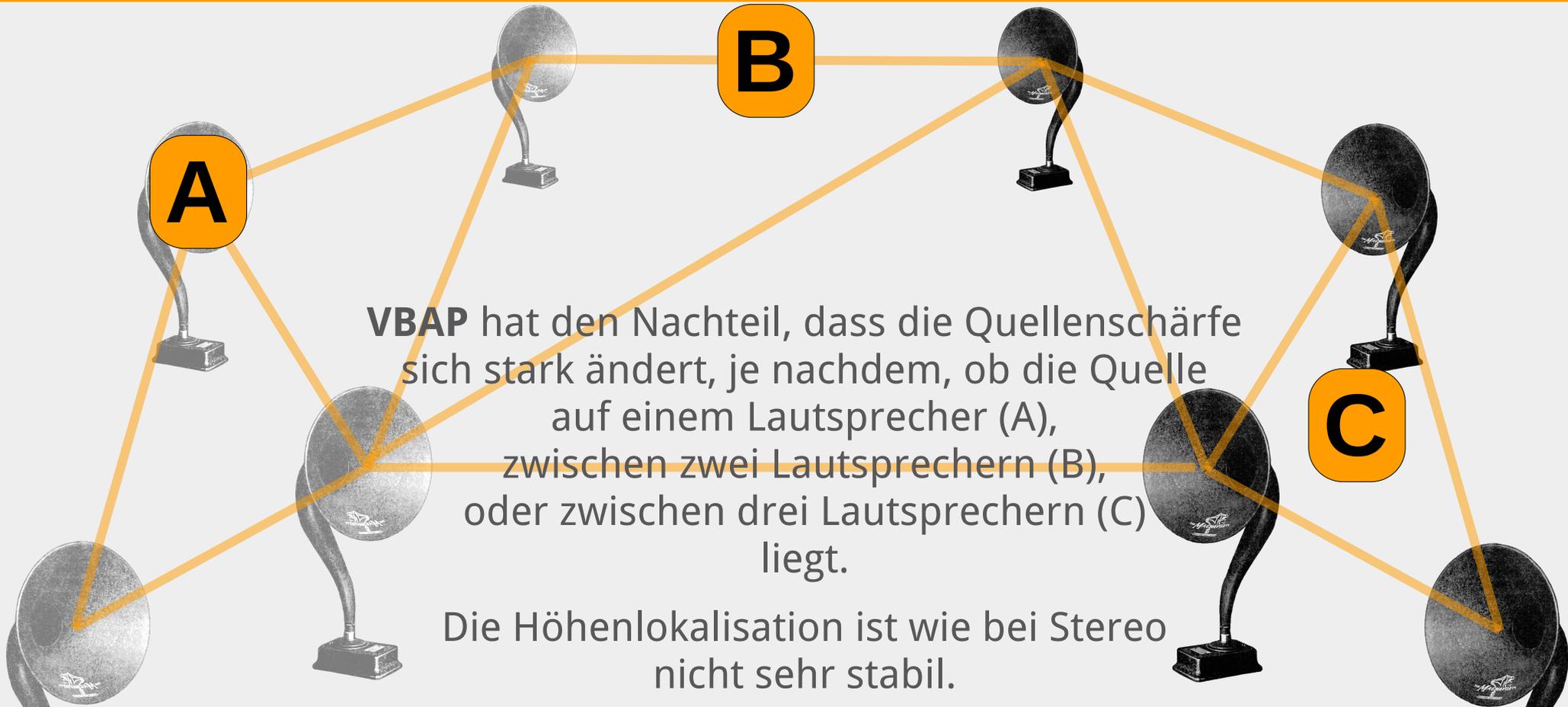


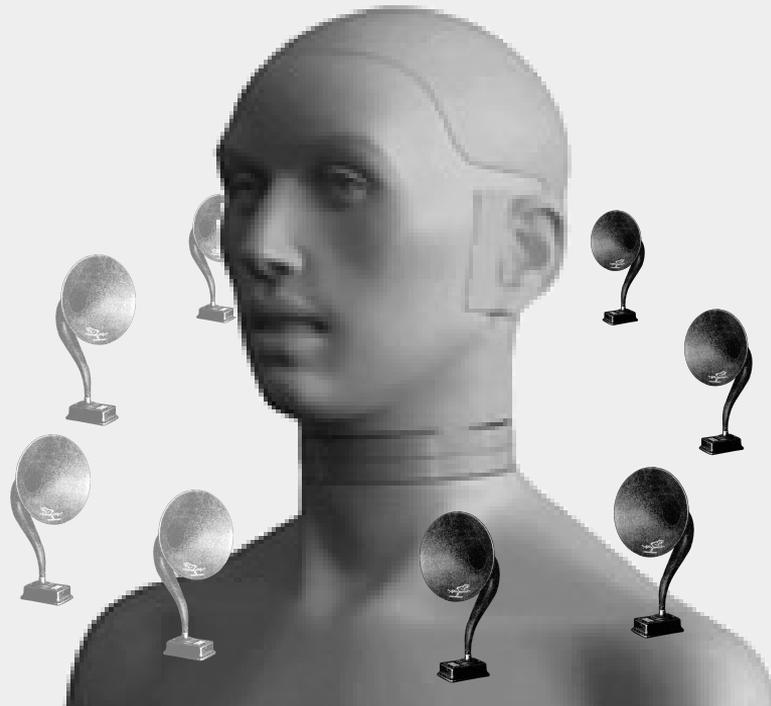


Wenn es genug Lautsprecher gibt, kann man mit Amplitudenpanning zwischen Lautsprecherdreiecken gute Lokalisationsergebnisse erzielen.

Das Verfahren heißt
**Vector-base Amplitude Panning
(VBAP)**

und kommt in vielen Systemen
(u.a. Dolby Atmos) zum Einsatz.



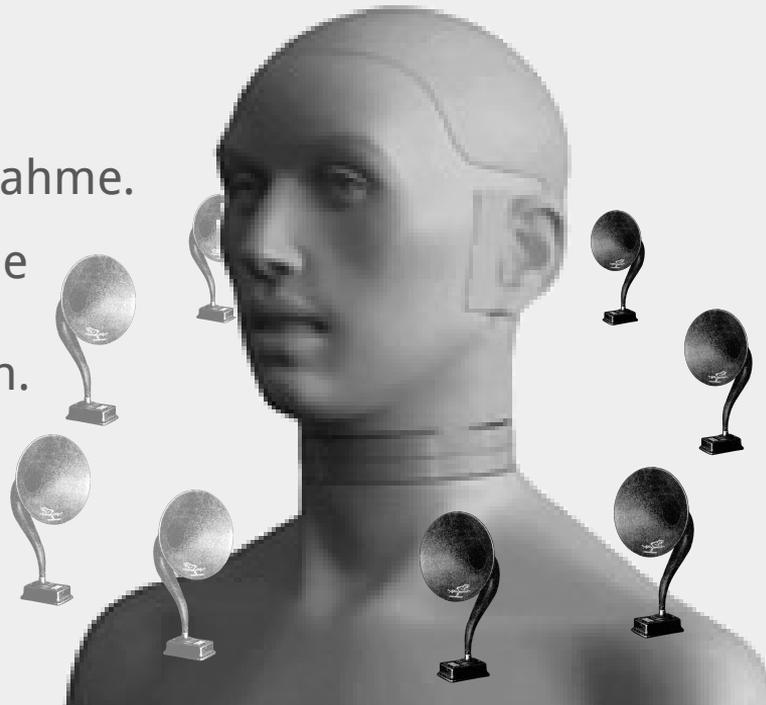


An den Trommelfellen ist die gesamte räumliche Information auf zwei Kanäle reduziert.

Ein **Kunstkopf** mit künstlichen Ohrmuscheln kann diese Information aufzeichnen.

Das Resultat ist eine **kopfbezogene** Aufnahme.

Theoretisch liefert sie perfekten 3D-Sound mit nur zwei Kanälen.

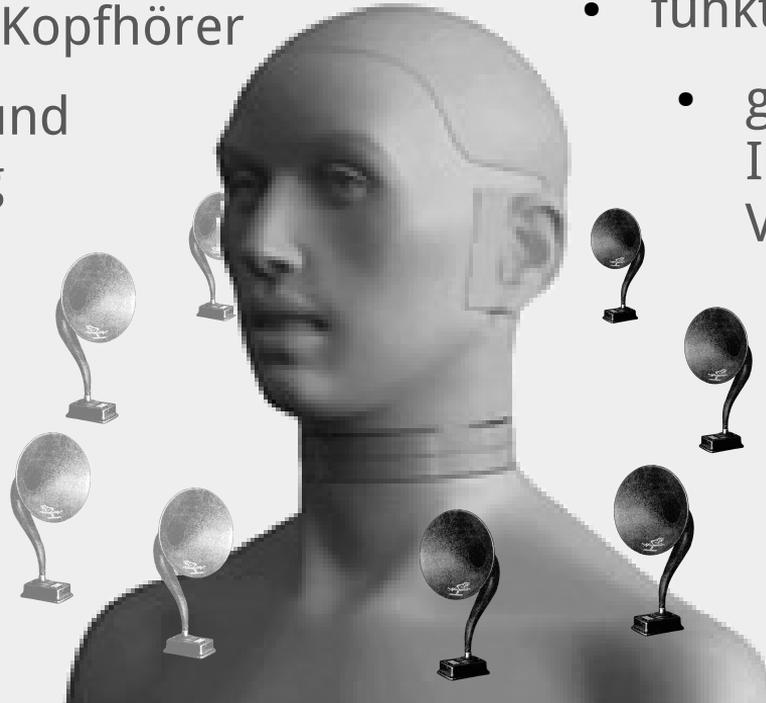


Alternativ kann man ein “trockenes” Signal mit den **HRTFs** für die gewünschte Richtung **falten**:

das Ergebnis klingt genau wie ein echter Klang aus dieser Richtung.

Vorteile:

- nur zwei Übertragungskanäle
- funktioniert über Kopfhörer
- **HRTFs** sind kurz und können mit wenig Rechenaufwand gefaltet werden.



Nachteile:

- funktioniert nur über Kopfhörer
- generische HRTFs erzeugen Im-Kopf-Lokalisation und Verfärbungen
- stabile Ortung braucht **head tracking** oder fixierten Kopf.

Vorteile:

- nur zwei Übertragungskanäle
- funktioniert über Kopfhörer
- **HRTFs** sind klein und können mit wenig Rechenaufwand gefaltet werden

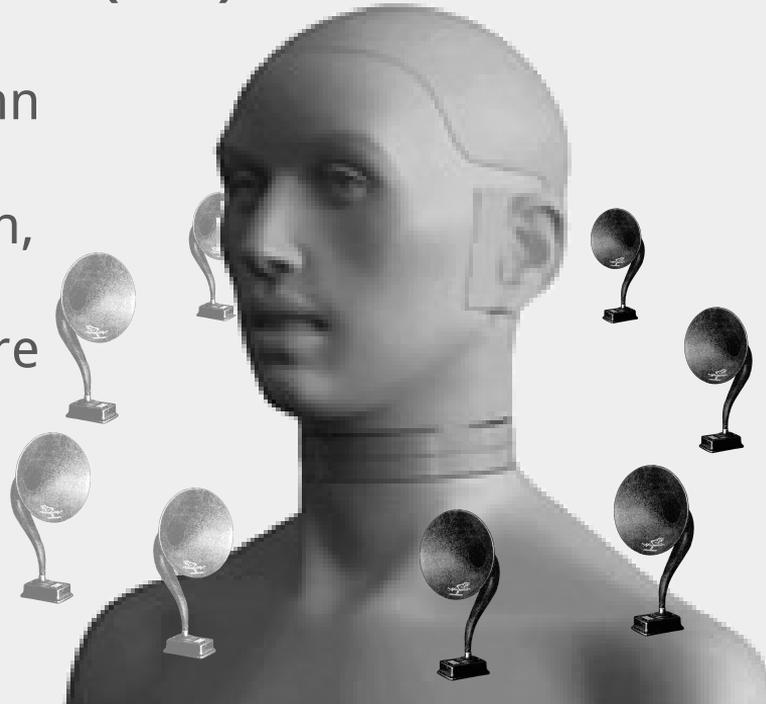
Nachteile:

- funktioniert nur über Kopfhörer
- generische HRTFs erzeugen In-ropf-Lokalisation und Verfälschungen
- stabile Ortung braucht **head tracking** oder fixierten Kopf.

JEDER SMARTPHONE IST EIN 3D-RENDERER

Binauralsignale können über zwei Lautsprecher wiedergegeben werden. Aber weil jeder Lautsprecher beide Ohren erreicht, bracht man dazu **crosstalk cancellation (CTC)**.

CTC-Binaural kann einen räumlichen Eindruck erzeugen, aber der Prozess verursacht hörbare Verfärbungen.

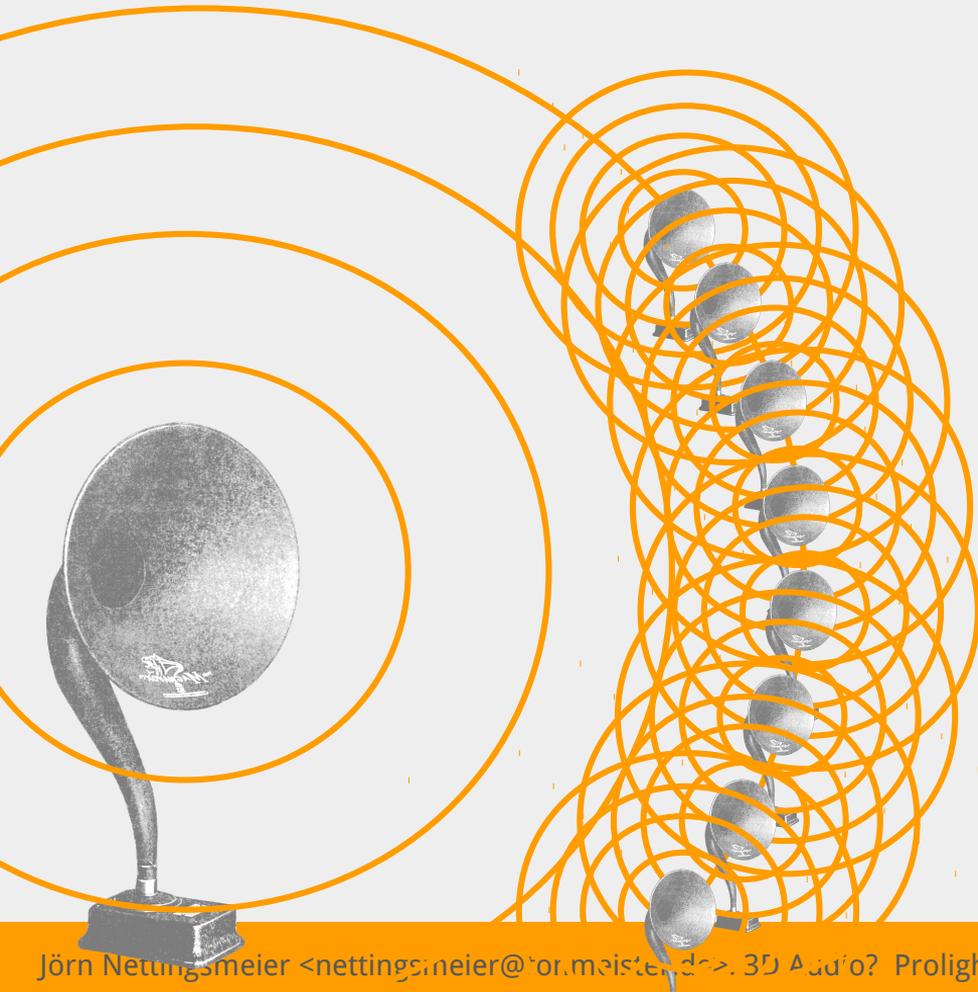


CTC funktioniert nur genau auf der Mittelachse zwischen zwei Lautsprechern.



Das Huygens'sche Prinzip:

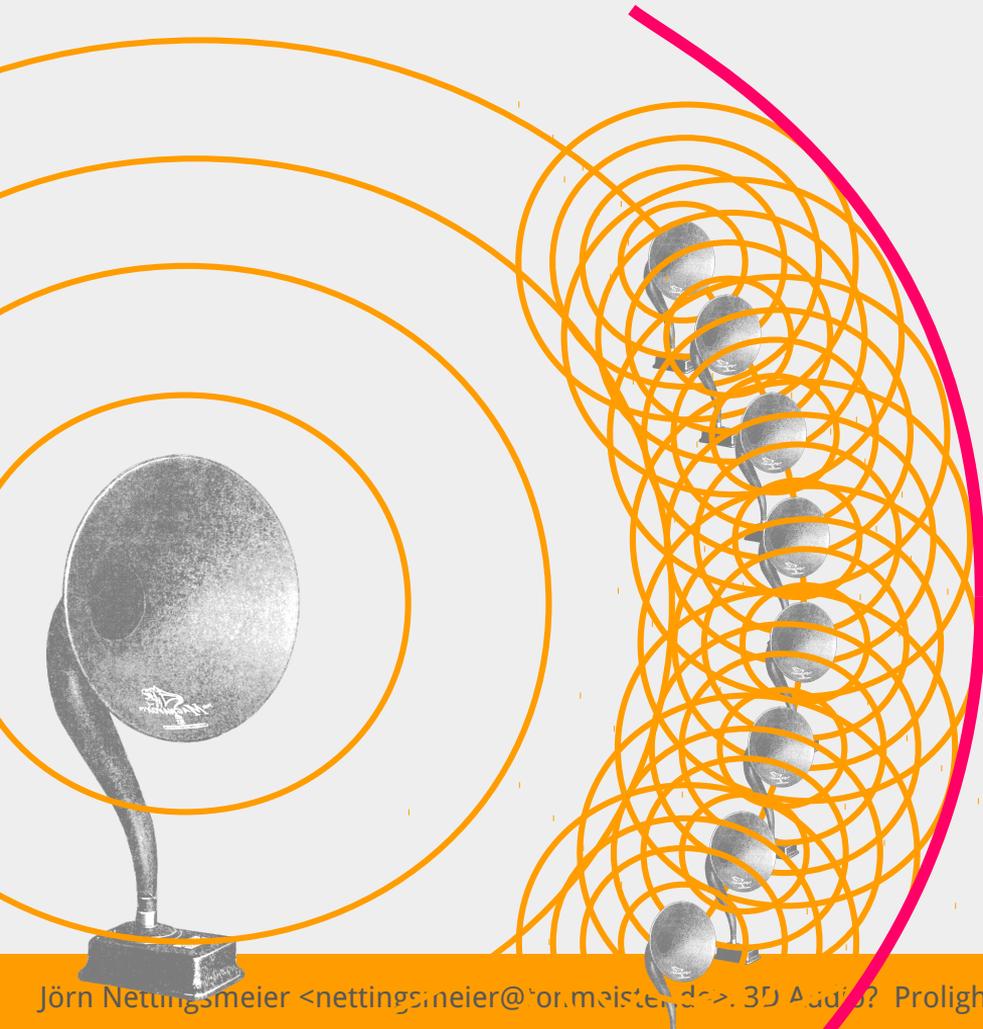
*Jede Wellenfront kann als Überlagerung von vielen **Elementarwellen** betrachtet werden.*



Das Huygens'sche Prinzip:

*Jede Wellenfront kann als Überlagerung von vielen **Elementarwellen** betrachtet werden.*

Oder: jede Klangquelle, egal in welcher Entfernung, kann durch eine Reihe von Lautsprechern reproduziert werden.

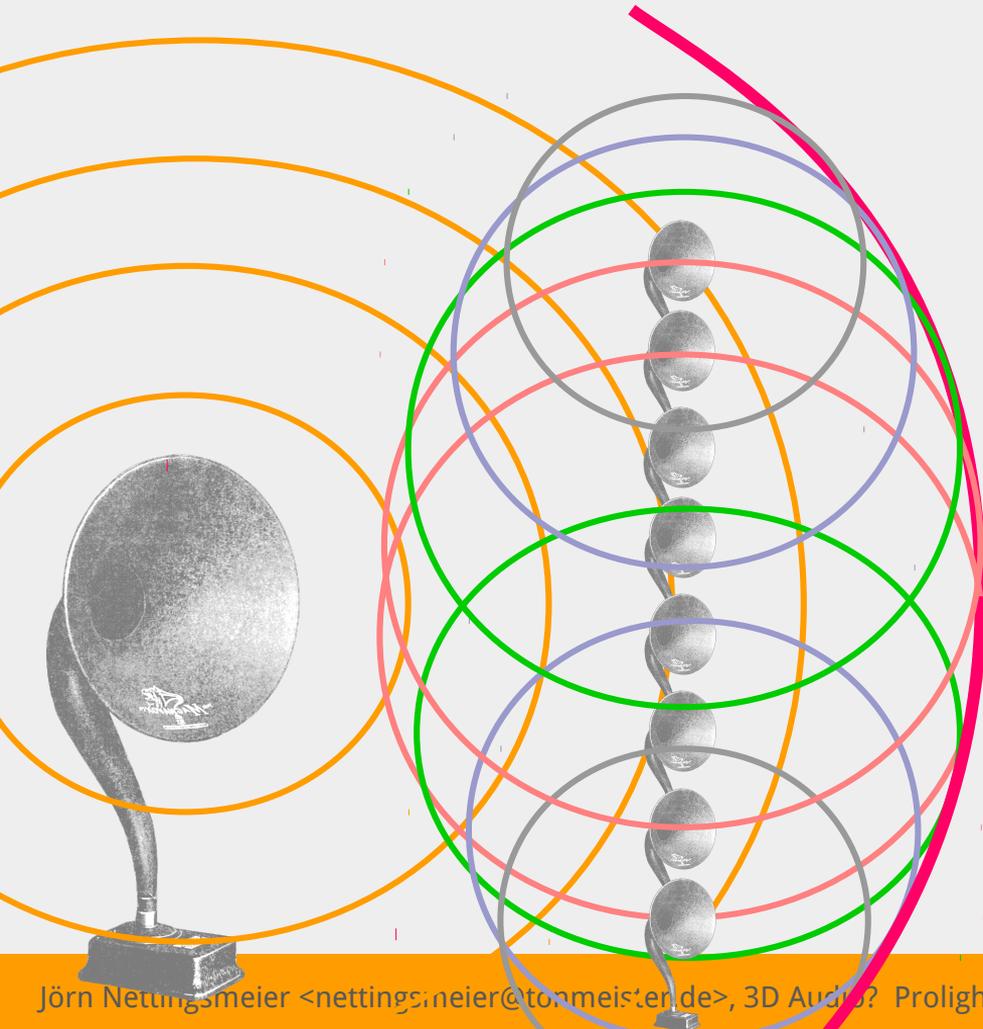


Das Huygens'sche Prinzip:

*Jede Wellenfront kann als Überlagerung von vielen **Elementarwellen** betrachtet werden.*

Eine solche Reihe kann das Schallfeld einer dahinter gelegenen Quelle bis zur **Alias-Frequenz** perfekt nachbilden.

Je nach Lautsprecher-Abstand liegt diese bei 1 – 2 kHz.



Das Huygens'sche Prinzip:

*Jede Wellenfront kann als Überlagerung von vielen **Elementarwellen** betrachtet werden.*

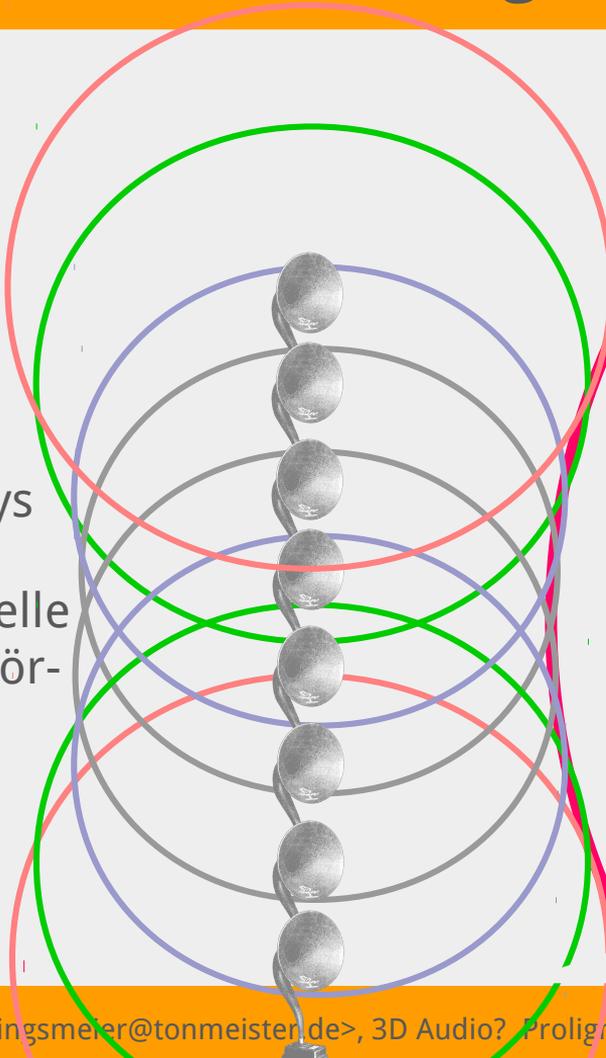
Die Lautsprecher müssen nicht entlang der Wellenfront angeordnet sein.

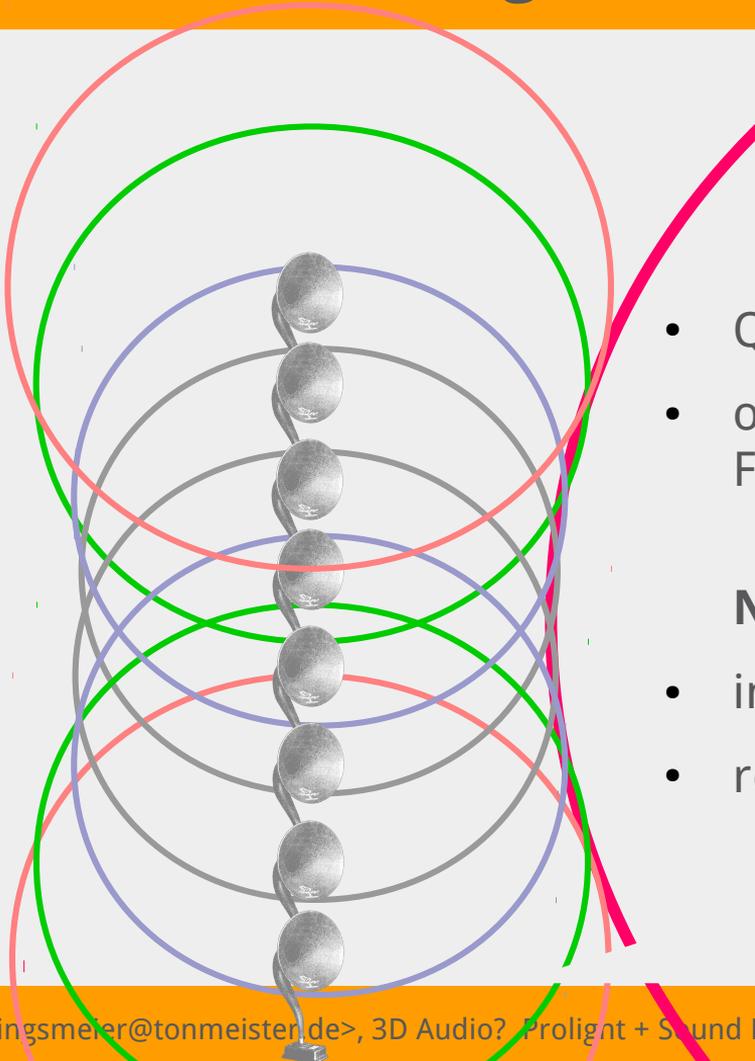
Mit den passenden Delays kann man beliebig geformte Wellenfronten erzeugen.

Sehr beliebt in Vergnügungsparks, aber auch spannend im Tanzbereich. Man muss mal "mitten in" einer Kickdrum gestanden haben, um es zu glauben.



Ein netter Trick:
invertierte Delays
erzeugen eine
fokussierte Quelle
innerhalb des Hör-
bereichs.





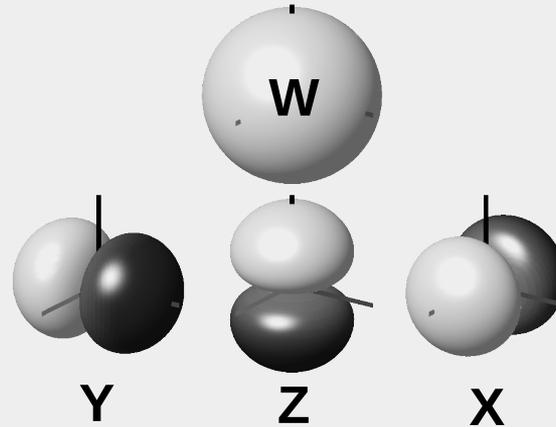
Vorteile:

- Großer Hörbereich, kein “sweet spot”
- sehr stabile Ortung
- Quellen innerhalb der Hörzone
- objektbasiertes Mixing, Content in anderen Formaten kann leicht abgebildet werden.

Nachteile:

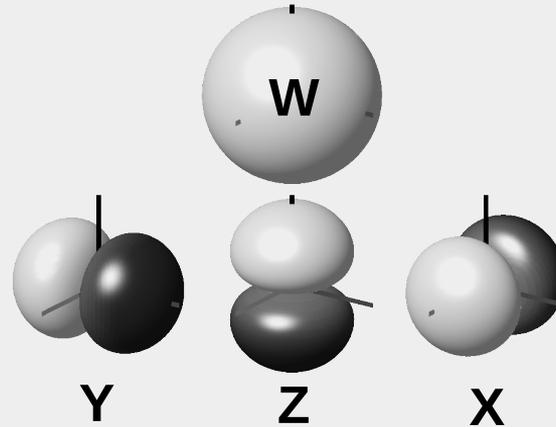
- immenser Hardware-Aufwand
- rechenintensiv
 - Aliasing-Artefakte hörbar
 - Keine Höhenlokalisierung





Ambisonics zerlegt das Schallfeld in richtungsabhängige Komponenten. Im einfachsten Fall kann man das mit Mikrofonen erreichen, die man so nah wie möglich zusammenbringt (**Koinzidenz**):

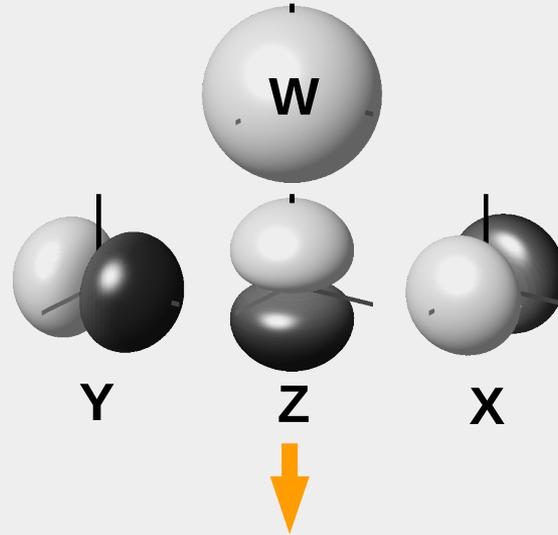
- eine Kugel nimmt den Schalldruck **W** auf
- 3 Achter nehmen die Schallschnelle in **X**-, **Y**- und **Z**-Richtung auf.



Dabei bedeutet das **X**-Signal “vorn minus hinten”, das **Y**-Signal “links minus rechts”, und das **Z**-Signal “oben minus unten”.

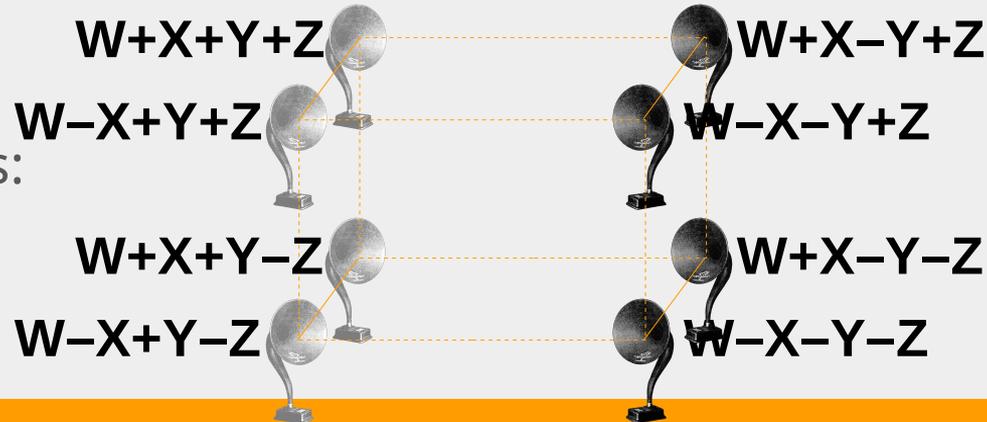
Diese 4 Komponenten nennt man **B-Format**. Mit ihnen kann man nun flexibel verschiedene Lautsprecher-Anordnungen antreiben:

Die Lautsprechersignale ergeben sich durch eine lineare Kombination der B-Format-Komponenten.



Für optimale Wiedergabe werden der Matrix noch einige psychoakustische Korrekturen hinzugefügt.

Hier eine vereinfachte Matrix für einen Kubus:

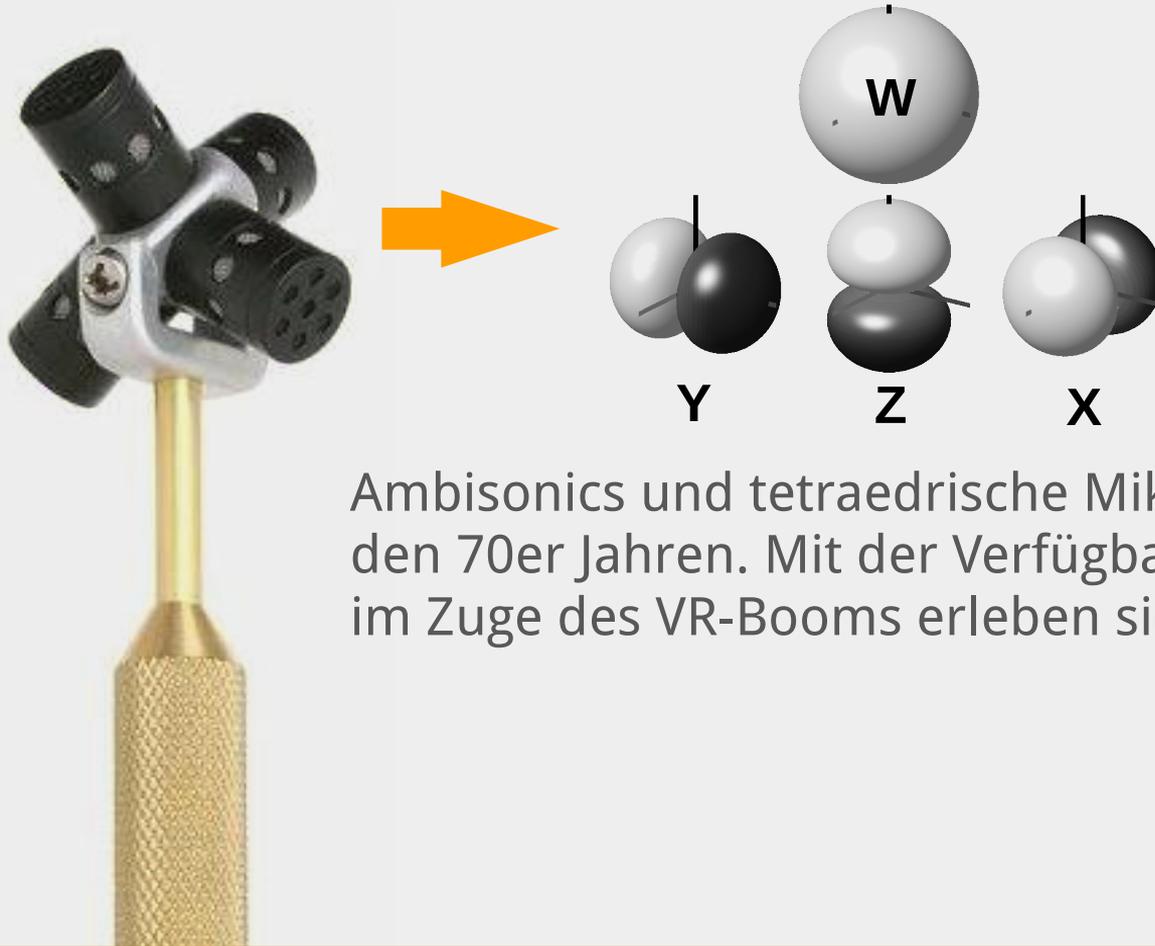


Andere Lautsprecher-Anordnungen benötigen zusätzlich winkelabhängige Gain-Faktoren.

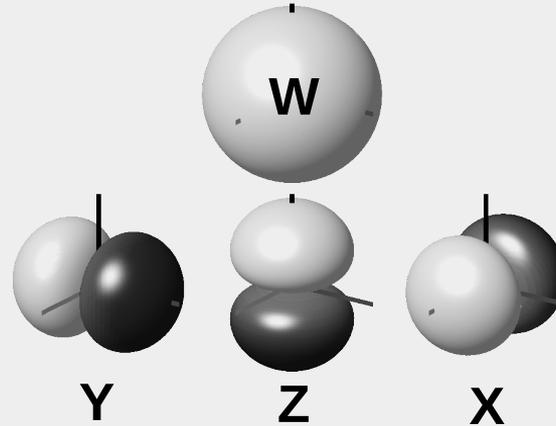


In der Praxis baut man Ambisonics-Mikrofone 1. Ordnung mit vier Nieren in Form eines Tetraeders. Wenn man alle addiert, erhält man die Kugel **W**. Die Achten **XYZ** entstehen, indem man jeweils zwei Nieren addiert und die anderen beiden abzieht.

Auch hier sind in der Praxis noch weitere Kniffe nötig, z.B. Entzerrungen, um den Koinzidenzfehler auszugleichen.



Ambisonics und tetraedrische Mikrofone sind Technologien aus den 70er Jahren. Mit der Verfügbarkeit von Digitaltechnik und im Zuge des VR-Booms erleben sie gerade eine Renaissance.



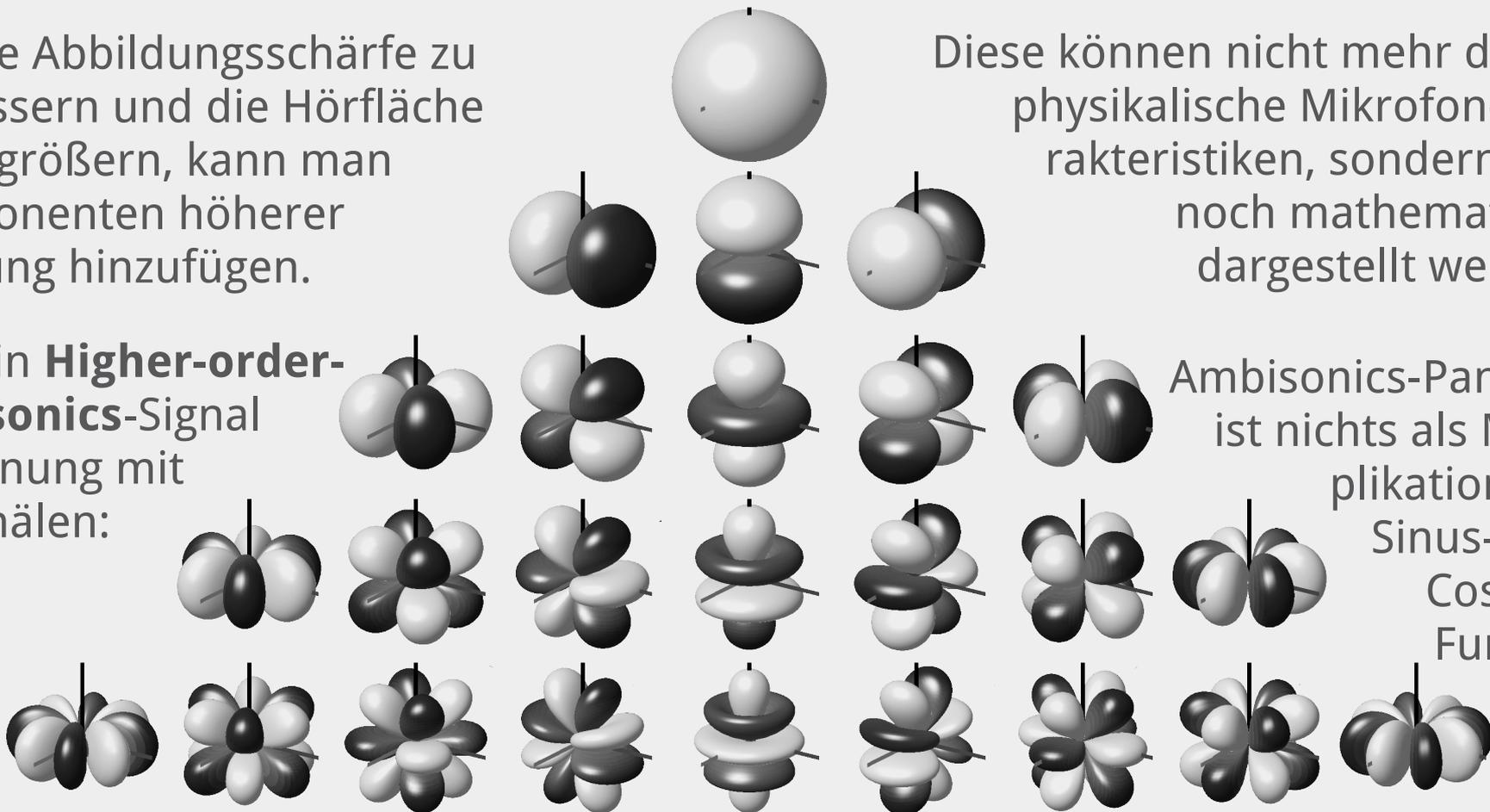
Auch diskrete Mono-Signale, z.B. aus einer Pop-Produktion, kann man mit Ambisonics-Pannern sehr einfach frei im Raum anordnen.

Solche "künstlichen" Ambisonics-Signale haben einen großen Vorteil:

Klänge im Raum erzeugen: Ambisonics

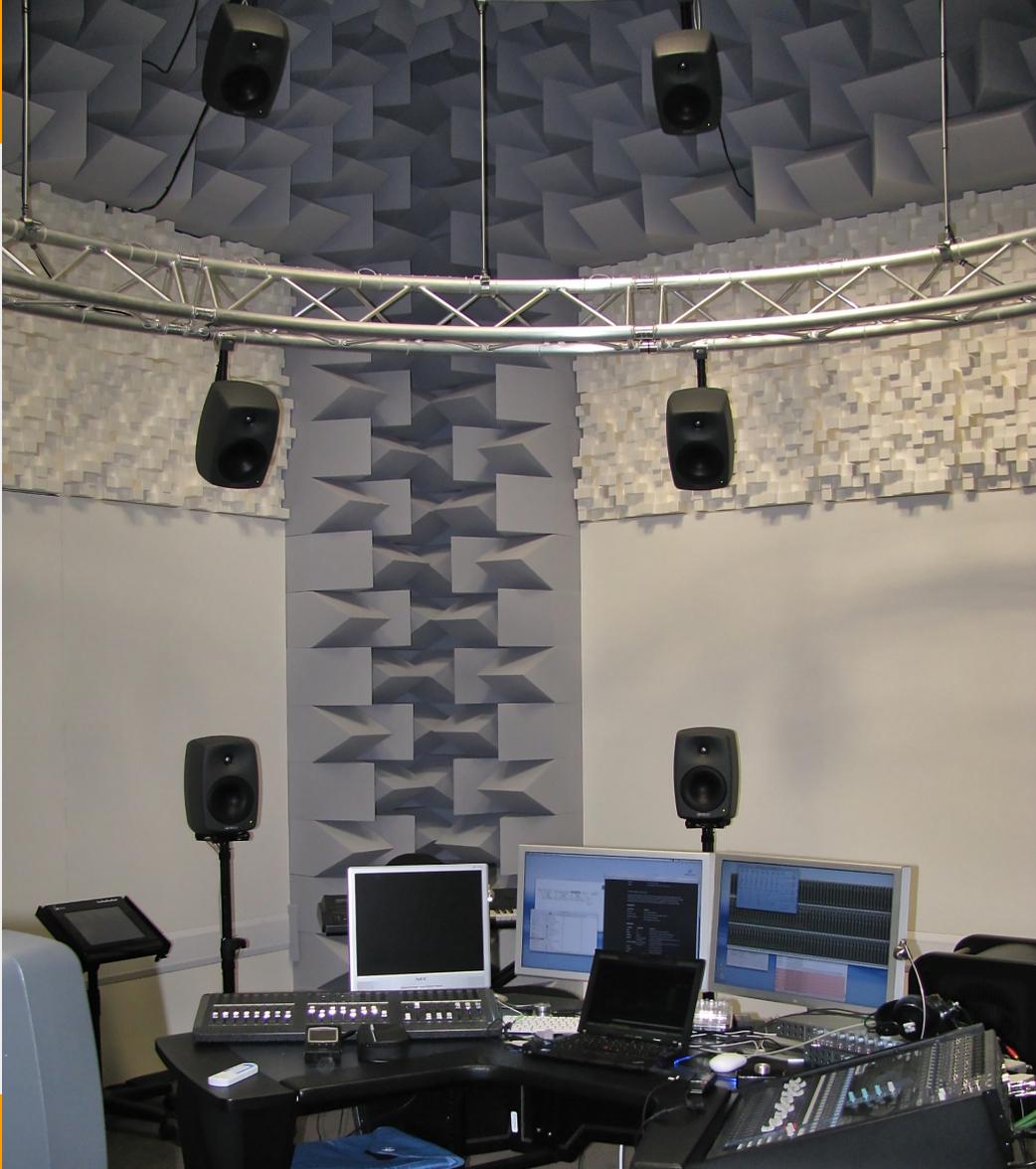
Um die Abbildungsschärfe zu verbessern und die Hörfläche zu vergrößern, kann man Komponenten höherer Ordnung hinzufügen.

Hier ein **Higher-order-Ambisonics**-Signal 4. Ordnung mit 25 Kanälen:



Diese können nicht mehr durch physikalische Mikrofon-Charakteristiken, sondern nur noch mathematisch dargestellt werden.

Ambisonics-Panning ist nichts als Multiplikation mit Sinus- und Cosinus-Funktionen.

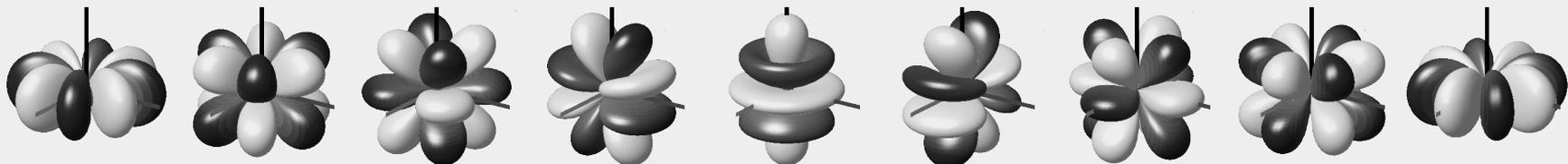


Vorteile:

- einfaches 3D ab 4 Kanälen/8 Lautsprechern
- bei Binauralwiedergabe sehr gute Stabilität bereits bei 1. Ordnung
- gute Skalierbarkeit, 3. Ordnung reicht bereits für größere Hörflächen
- Produktion ist unabhängig vom Lautsprecher-Layout
- flexible Wiedergabemöglichkeiten
- einfache Rotation, dadurch vorteilhaft für Head-Tracking in Kombination mit Binauralsynthese

Nachteile:

- in niedrigen Ordnungen hohe Korrelation zwischen den Lautsprechern, dadurch Verfärbungen und Phasing
- in Theatern schwierig umzusetzen
- derzeit noch nicht von Konsumenten-Geräten unterstützt (aber Bestandteil von MPEG-H)
- etwas schwierig zu verstehen
- Fehlerbehebung in Wiedergabesystemen ist nicht-trivial.

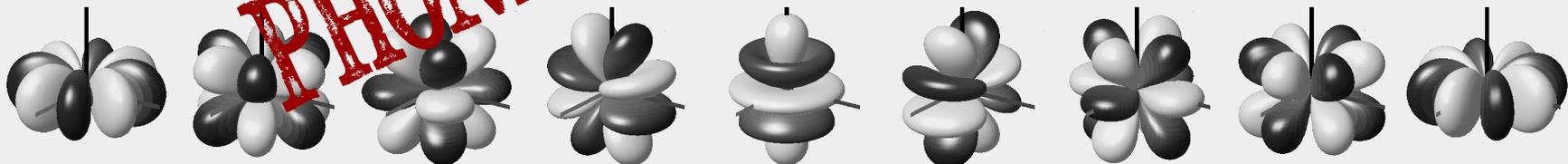


Vorteile:

- einfaches 3D ab 4 Kanälen/8 Lautsprechern
- bei Binauralwiedergabe sehr gute Stabilität bereits bei 1. Ordnung
- gute Skalierbarkeit, 3. Ordnung reicht bereits für größere Hörflächen
- Produktion ist unabhängig vom Lautsprecher-Layout
- flexible Wiedergabemöglichkeiten
- einfache Rotation und durch vorteilhaft für Head-Tracking in Kombination mit Binauralsynthese

Nachteile:

- in niedrigen Ordnungen hohe Korrelation zwischen den Lautsprechern, dadurch Verfärbungen und Phasing
- in Theatern schwierig umzusetzen
- derzeit noch nicht von Konsumenten-Geräten unterstützt (aber Bestandteil von MPEG-H)
- etwas schwierig zu verstehen
- Fehlerbehebung in Wiedergabesystemen ist nicht-trivial.



Vielen Dank
für
Zuhören

Ich freue
mich auf Eure
Fragen

- https://commons.wikimedia.org/wiki/File:Smilodon_Fatalis_by_Salvatore_RabitoNBG.png
- <https://commons.wikimedia.org/wiki/File:Maniki.jpg>
- https://commons.wikimedia.org/wiki/File:Shanghai_Hongqiao_Railway_Station_-_platform_-_P1080122.JPG
- <https://commons.wikimedia.org/wiki/File:Alphah.jpg>
- <http://core-sound.com/TetraMic/1.php>