

Die dritte Dimension für Lautsprecher-Stereofonie

Günther Theile und Helmut Wittek, Mai 2011

Einleitung	1
1. Surround Sound im Wandel.....	1
2. 3D-Ton für das 3D-Bild	4
2.1. Lautsprecherwiedergabe.....	5
2.2. Kopfhörerwiedergabe	5
3. Mehrkanalton - psychoakustische Anforderungen	6
3.1. Reflexionen und Nachhall	7
3.2. Diffuser nicht reflektierter Schall (Atmo)	7
4. Dreidimensionaler Ton mit Auro-3D	8
4.1. Darstellungsbereich zwischen unten und oben	8
4.2. Reflexionen und Diffusschall im 3D-Surround	10
5. Zur Aufnahmetechnik für Auro-3D	10
5.1. Die Rolle der Psychoakustik bei der Suche nach der Aufnahmetechnik	10
5.2. Kanaltrennung	11
5.3. Nutzung von künstlichem Hall bzw. Faltungshall	12
5.4. Diffuser Schall.....	13
5.5. Design eines Auro3D-Hauptmikrofons.....	14
6. Referenzen:	15

Einleitung

Die dritte Dimension wird für Mikrofonaufnahmen erschlossen, indem neue Mehrkanalformate zum 5.1 Surround-Format Höhenlautsprecher hinzufügen. Sie erlauben eine weit größere Palette an räumlichen Effekten und eine natürlichere räumliche Wiedergabe in Bezug auf Direktschall, frühe und späte Reflexionen, Hall, Diffusschall und Atmo. Am Beispiel des Auro3D 9.1-Verfahren (5.1 + 4) werden die psychoakustischen Prinzipien bei der Wahrnehmung von elevierten Phantomschallquellen, räumlicher Tiefe, Räumlichkeit, Umhüllung, Atmo sowie Richtungswiedergabe und –stabilität innerhalb der Hörzone erläutert. Konkrete Vorschläge für Mikrofonanordnungen können aus diesen Überlegungen entstehen.

1. Surround Sound im Wandel

Seit Verabschiedung des internationalen Standards ITU-R BS. 775-1 im Jahre 1992 verging viel Zeit bis zur Etablierung entsprechender Übertragungstechnik in wichtigen Medien und bis zur Erlangung ausreichender Erfahrung bei der Produktion. Die Umstellung der Aufnahmetechnik von 2.0 auf 5.1 war ein erster wichtiger Schritt, weg von „reiner Stereofonie“ zwischen zwei Lautsprechern vor dem Hörer, hin zur Reproduktion einer real wirkenden akustischen Umgebung.

Der 5.1-Standard ist ein Kompromiss, er war damals notwendig wegen gegebener Rahmenbedingungen (Kompatibilität mit 2.0-Stereo und mit Kino-Formaten, maximal 6 Übertragungskanäle). Die Verbesserungen beschränken sich deshalb im Wesentlichen auf zwei Punkte [1]:

- Vergrößerte Hörzone sowie verbesserte Stabilität und Qualität des stereofonen Klangbildes vor dem Hörer durch Aufteilung der Lautsprecherbasis L-R mit 60° in zwei stereofone Teilbereiche L-C und C-R mit je 30°.

- Einsatz der zwei zusätzlichen Surround-Lautsprecher im seitlichen / hinteren Bereich des Hörers, so dass sich das akustische Umfeld des Hörers in bestimmten Grenzen real gestalten lässt.

Seit einigen Jahren stellen wir fest, dass das technische Umfeld für 5.1 bei Produktion, Distribution und Endgeräten nahezu durchgehend eingeführt ist. Auch akzeptiert der Konsument weitgehend eine höhere Zahl der Lautsprecher, zumindest in Verbindung mit der Fernsehanlage („Heimkino“). Wir stellen aber auch fest, dass tatsächlich nur wenige Hörer die mit Surround Sound erreichbare bzw. erträumte Wiedergabequalität erhalten. Dies hat verschiedene Ursachen:

- Beim Hörer: Ungünstige geometrische und akustische Raumbedingungen, vom Standard zu weit abweichende Anordnung der Lautsprecher, unsachgemäße Einstellung der Geräte.
- Schlechte Aufnahmen. Einerseits bedingt durch ökonomische Zwänge in der Produktion, andererseits verursacht durch unpassende Mikrofon- und Mischtechnik
- Kleine Hörzone des 5.1-Formats. Häufig setzen Hörort-sensible Aufnahmen das Hören im Sweetspot voraus.
- Darstellungsgrenzen des 5.1-Formats: Mangelhafte Rundumabbildung, fehlende Einbeziehung der Höhe und der Kopfnähe, eingeschränkte Möglichkeiten für die Entfernungsdarstellung.

Die Reihenfolge soll nicht unbedingt eine Rangfolge der Bedeutung zum Ausdruck bringen. Die Aufzählung macht aber deutlich, dass vorwiegend die praktische Handhabung Schwierigkeiten bereitet, sowohl produktionsseitig als auch auf Seite des Hörers. Sie lassen sich nicht einfach mit einer Steigerung der Zahl der Übertragungskanäle und Lautsprecher beseitigen. Im Gegenteil, die in letzter Zeit vorgestellten Erweiterungen oder neuen Systeme, angefangen bei verschiedenen 7.1-Versionen bis hin zu High-Order-Ambisonics (HOA) und Wellenfeldsynthese (WFS), erfordern auf der Aufnahmeseite Umdenken, Umrüstung und besondere Sorgfalt, auf der Wiedergabeseite vom Konsumenten Akzeptanz für ein Wohnzimmer mit Heimkino-Outfit. Dabei sind die momentane Formatvielfalt und die fehlenden neuen Format-Standards ein zusätzliches Hindernis. Die aktuelle DCI-Spezifikation (bzw. SMPTE 428M) sieht Channel Mapping vor und lässt bewusst die beliebige Nutzung von 16 Kanälen zu.

Bereits der Standard ITU-R BS. 775-1 enthält optionale Lautsprecher LL und RR zwischen Front- und Surround-Lautsprechern. Sie bieten die Möglichkeit, die stereofone Qualität bei seitlicher Abbildung zu verbessern, die Hörzone zu vergrößern und das „Loch“ zwischen frontseitiger und seitlicher Abbildung zu schließen. Insgesamt ergibt sich eine größere Freiheit für die räumliche Darstellung stationärer seitlicher Hörereignisse oder die wichtigen seitlichen Reflexionen. In Anlehnung daran und in Verbindung mit Filmtone-Entwicklungen propagieren Firmen wie DTS oder Dolby verschiedene 7.1-Formate, die in ähnlicher Weise vier Surround-Kanäle seitlich und hinter der Hörzone einsetzen, bei unveränderter Anordnung der Front-Lautsprecher L-C-R. Es sind einige hundert Home Cinema Blu-ray Discs mit 7.1-Ton auf dem Markt, die oftmals innerhalb einer großen Hörzone mit klar definierter und stabiler Richtungsdarstellung seitlich und hinter den Hörern überzeugen, jedoch kaum mit Musikaufnahmen [2].

All diese Surround-Formate basieren im Prinzip auf Stereophonie, d.h. die Quellenabbildung geschieht mit Phantomschallquellen zwischen zwei benachbarten Lautsprechern. Im Surround-Bereich ist die Phantomschallquellenrichtung extrem vom Hörplatz abhängig und instabil, weshalb hier praktisch nur die Orte der Lautsprecher die Richtungsdarstellung übernehmen. Auch die Lautstärke-Balancen sind platzabhängig, besonders auffallend für das Verhältnis Front-Surround. Daher bezwecken zusätzliche Kanäle in der Horizontalebene eine größere Hörzone bzw. eine homogenere und stabilere Richtungsauflösung.

In der alternativen Anwendung zusätzlicher Kanäle wird die Horizontalebene verlassen. Lautsprecheranordnungen in der Ebene oberhalb des Hörers vervollständigen den räumlichen Gestaltungsbereich und ermöglichen in bestimmten Grenzen den Aufbau eines

dreidimensionalen Klangbildes. Vor fast 10 Jahren hatte *Werner Dabringhaus* mit seiner „2+2+2-recording“-Technik die ersten Musikaufnahmen auf den Markt gebracht (5.1 ohne Center und Subwoofer, dafür 2 Lautsprecher oberhalb L und R) [3]. Das Konzept war für die Audio-DVD ausgelegt und darauf, ein möglichst authentisches Klangbild aus dem Konzertsaal zu reproduzieren und deshalb zugunsten der Lautsprecher für die Höheninformationen auf Center und Subwoofer zu verzichten. In ähnlicher Weise berücksichtigt *Tom Holman* mit zwei hoch angeordneten Lautsprechern schräg rechts und links vor dem Hörer die dritte Dimension. Seine „10.2 channel surround sound“ Konfiguration beansprucht allerdings 8 Kanäle in der Horizontalebene und ist ursprünglich für Kino- und Heimkino-Anwendung gedacht [4].

Im Jahr 2006 hat *Wilfried Van Baelen* in Paris das „Auro-3D“ Format vorgestellt, das die Höhe konsequent mit vier zusätzlichen Kanälen einbezieht. In der Basisversion „Auro-3D 9.1“ ergänzen die Höhenlautsprecher das 5.1-Format, sie befinden sich über den Lautsprechern L, R, RH, LH (Abb. 1). Ebenso lassen sich natürlich andere Formate wie „7.1 Surround“ mit vier etwa „quadrofon“ angeordneten Höhenlautsprechern ergänzen. Über den Stand der Entwicklungen und Erfahrungen wurde auf der vergangenen Tonmeistertagung im Rahmen der Session „Digital Cinema“ berichtet [5].

Das wesentliche Merkmal dieses Formats ist die kubusähnliche Anordnung von 8 Lautsprechern, so dass für die Wiedergabe der (frühen) Reflexionen der ganze obere Halbraum einbezogen werden kann und sich die subjektive räumliche Diffusität des Nachhalls in ausreichendem Maße reproduzieren lässt. Damit sind wiedergabeseitig gute Voraussetzungen für die Darstellung von Umhüllung, Räumlichkeit und räumlicher Tiefe geschaffen. Zusätzlich bieten die oberen Lautsprecher natürlich die gleichen Möglichkeiten der stereofonen Quellenabbildung wie die ITU-Anordnung ohne Center. Phantomschallquellen zwischen den unteren und oberen Lautsprechern, also stabile stationäre Hörereignisrichtungen mit Elevation zwischen 0° und 30° sowie direkt über dem Hörer sind allerdings praktisch nicht realisierbar. Hierauf wird später näher eingegangen.

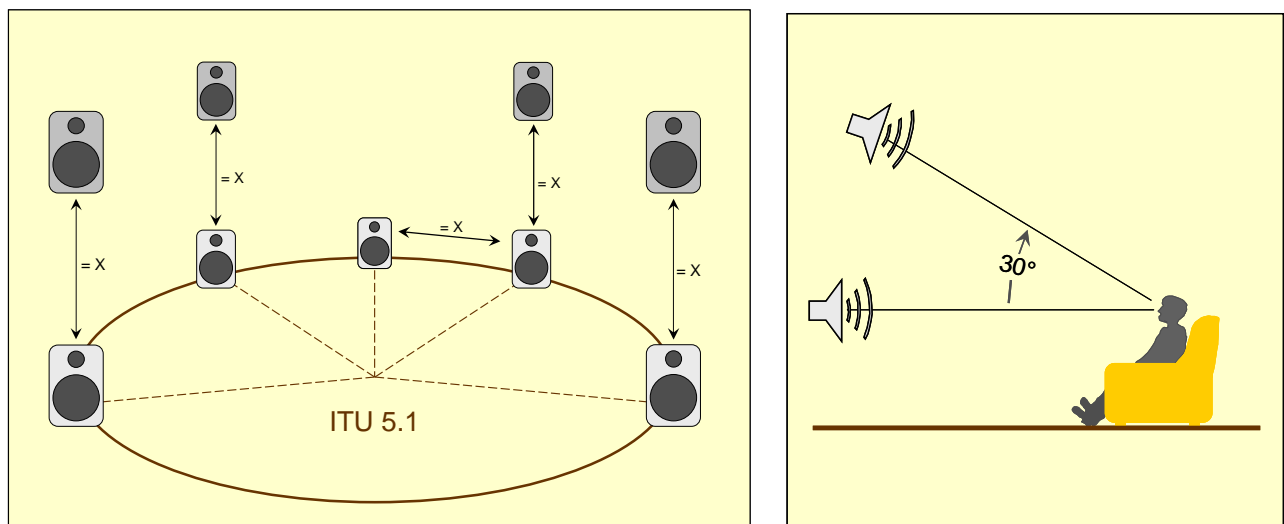


Abb. 1: Basis-Konfiguration „Auro-3D 9.1“ (nach [5]), abwärtskompatibel mit ITU-R BS. 775-1

Bestimmte Schwächen des 5.1 Surround Formats können mit Auro-3D 9.1 beseitigt oder gemildert werden, andere nicht. In Tabelle 1 sind verschiedene Attribute des reproduzierten Klangbildes aufgeführt. Die ersten vier betreffen den Direktschall (meistens mit Panning gestaltet), die nächsten vier betreffen Wirkungen des indirekten Schalls (mit Mikrofontechniken und Processing gestaltet). Mit den Attributen lassen sich die Leistungsprofile unterschiedlicher Verfahren einigermaßen sachgerecht charakterisieren und vergleichen. Dies unter dem

Vorbehalt, dass wiedergabeseitig die Empfehlungen korrekt realisiert und aufnahmeseitig die passenden Mikrofon- und Mischtechniken eingesetzt werden.

Attribute des reproduzierten Klangbilds	2.0 Stereo	5.1 Surround	Auro-3D 9.1	WFS Verfahren /*	Binaurale Verfahren
Richtung vorne	•	••	••	••	•
Richtung rundum		•	•	••	••
Elevation			(•) /***		••
Höhe			•		••
Entfernung, Tiefe	(•) / **	•	••	••	••
Nähe am Kopf				•	••
Intra-aktive Perspektive / ¹				••	
Räumlichkeit	(•) /**	•	••	•	••
Umhüllung		•	••	•	••
Klangfarbe	••	••	••	•	••

Tab. 1: Leistungsprofile der Verfahren im Vergleich, geeignete Aufnahme- und Wiedergabetechnik vorausgesetzt.

/* horizontale Arrays, /** simulierte Tiefe bzw. Räumlichkeit, /*** nur im Sweetspot, instabil

Der Tabelle kann entnommen werden, dass Auro-3D 9.1 verglichen mit den anderen Lautsprecherverfahren bezüglich einiger Attribute prinzipielle Vorteile aufweist. Dies lässt sich auf andere Formate übertragen, die 2D Surround Wiedergabesysteme mit Quadrofonie-ähnlichen Lautsprecheranordnungen in der Ebene über dem Hörer ergänzen. Im Abschnitt 4 soll auf die gestalterischen Möglichkeiten und Grenzen näher eingegangen werden, insbesondere auf den zweckdienlichen Einsatz der Mikrofontchnik.

2. 3D-Ton für das 3D-Bild

Spätestens mit der Weiterentwicklung der Kunstkopftechnik (z.B. [6], [7]) begannen ernsthafte und teilweise erfolgreiche Bemühungen, eine dreidimensionale Tonübertragung für Rundfunk und Tonträger zu etablieren. Die binaurale Reproduktion der Ohrsignale ist die ursprüngliche Methode für 3D-Audio. Im Idealfall sind die reproduzierten Kunstkopfsignale identisch mit den individuellen Ohrsignalen, die der Hörer im Aufnahmerraum am Ort des Kunstkopfs empfangen würde. Das virtuelle Hörereignis entspricht dann dem realen Schallereignis im Aufnahmerraum. Binaurale Verfahren bleiben jedoch aus diversen praktischen Gründen [8] speziellen Anwendungen vorbehalten, sie sind nicht kompatibel mit Lautsprecherwiedergabe, d. h. die zweikanaligen Signale lassen sich nicht überführen in entsprechend wirkende mehrkanalige Lautsprechersignale. Jedoch kann man die mit binauraler Technik erzielbare Qualität der 3D-Raumabbildung als eine Art Referenz ansehen, denn der gesamte obere Halbraum bildet den Abbildungsbereich, Hörereignisse mit beliebiger Elevation und Entfernung sind darstellbar.

¹ Eine intra-aktive Perspektive kennzeichnet eine Szene, in der die Wahrnehmung der Bewegung der agierenden Person folgt. Dieses Merkmal unterstützt die Wahrnehmung der räumlichen Tiefe. Im Gegensatz dazu folgt die *inter*-aktive Szene bestimmten Aktionen der agierenden Person.

Der Vollständigkeit halber sei die „intra-aktive Perspektive“ erwähnt, ein Merkmal des natürlichen Hörens in der auditiven Szene. Die wahrgenommenen Richtungsverhältnisse ändern sich abhängig von den Quellenentfernungen (nahe Quellen wandern bei Ortswechsel weiter aus als ferne). Derartige Nachbildungen sind mit Wellenfeldsynthese-Systemen realisierbar [9] und beispielsweise für zukünftige Entwicklungen im Game-Bereich interessant. Sie sollen hier nicht weiter betrachtet werden.

2.1. Lautsprecherwiedergabe

Von Interesse dagegen ist die Frage: Welche Merkmale der dreidimensionalen Tondarstellung passen zum 3D-Bild? Zunächst ist festzustellen, dass die Ausgangssituation bei der Bilddarstellung anders ist als beim Ton. Aus dem flächigen, zweidimensionalen Bild entsteht ein „3D-Bild“, indem im gegebenen Bildrahmen mit Mitteln der Stereoskopie der Eindruck von Tiefe erzeugt wird / ². Im Gegensatz dazu ist die dritte Dimension beim Ton die Höhe, die zwei ersten Dimensionen sind „Richtung“ und „Entfernung“. 2D-Verfahren sind 2.0 Stereo, 5.1 Surround und WFS, unabhängig davon, wie weit die Gestaltungsmöglichkeiten eingeschränkt sind. Besonders stark natürlich bei 2.0 Stereo (Entfernung/Tiefe nur simuliert, Richtungen nur im Bereich 60°), deutlich weniger bei 5.1 Surround und WFS [1], [9] (siehe Tabelle 1).

Das 3D-Bild ist vorstellbar als Fenster zu einer dreidimensionalen Szene, wobei Objekte in das Fenster hinein ragen können. Entsprechende auditive Objekte vor dem Lautsprecher-Display lassen sich grundsätzlich mit Wellenfeldsynthese darstellen. Wollte man unter allen Umständen räumliche Bild-/Ton-Kongruenz schaffen, so wären aus dieser Sicht im Prinzip lediglich WFS- und Binaural-Techniken geeignet. Jedoch wäre dieser Aufwand aus vielen Gründen weder praktikabel noch sinnvoll, zumal WFS nicht die Höhe einbezieht und binaurale Signale nur über Kopfhörer ausreichend fehlerfrei an die Ohren einer Anzahl Hörer gebracht werden können.

Die erste Wahl ist also Auro-3D (9.1 und höher). Die Technik wird vielen Anforderungen gerecht, die heute an einen zukünftigen universellen und kompatiblen Standard für Digital Cinema, Games, Rundfunk und Musikindustrie zu stellen sind [5]. Wie in den nachfolgenden Abschnitten deutlich wird, erfordert die aufnahmeseitige Handhabung dieser Lautsprecherkonstellation sorgfältige Berücksichtigung psychoakustischer Phänomene, um bestimmte gestalterische Ideen mit guten Ergebnissen umsetzen zu können. Die Einbeziehung der Höhe ist nach Einführung der 5.1-Surround-Kanäle der zweite Schritt zu dem Ziel, den Gestaltungsspielraum der Lautsprecher-Stereofonie zu erweitern. Eine der anspruchvollsten Aufgaben ist die „natürliche“ Musikaufnahme, denn sie erfordert eine besondere Mikrofontechnik, um alle vier Aufgaben der räumlichen Aufnahme gleichzeitig zu meistern: Richtungen und Ausdehnungen der Quellen, Tiefe der Szene, Raumeindruck, Umhüllung. Die neuen Gestaltungsmöglichkeiten werden deshalb in den folgenden Abschnitten anhand dieser Aufnahmesituation diskutiert.

2.2. Kopfhörerwiedergabe

Moderne Faltungstechnik ermöglicht die realitätsgetreue Darstellung eines virtuellen Auro-3D-Studios mit Kopfhörern. Die derzeit kommerziell vorhandenen „Binaural Room Synthesis“-Systeme (BRS) gewährleisten virtuelle 5.0-Lautsprecherwiedergabe in professioneller Qualität, sie lassen sich problemlos modifizieren für die zusätzlichen Kanäle oberhalb des Hörers. Die Surround Sound Signale werden mit gemessenen binauralen Impulsantworten eines hochwertigen Studios gefaltet. Die Datenauswahl für die Faltung geschieht mittels Headtracking in Abhängigkeit von der aktuellen Kopfausrichtung, so dass der Hörer die virtuellen Lautsprecher unabhängig von der Kopfhaltung (raumbezogen) lokalisiert [10]. Im IRT wurden 2007 die Entwicklungsarbeiten mit dem BRS-Plugin für VST-Host Applikationen [11] abgeschlossen. Inzwischen ist ein kostengünstiges eigenständiges BRS-Gerät auf dem Markt, dass mit Hilfe einer Individualentzerrung eine exzellente Nachbildung der Studiosituation erzielt [12].

² Man unterscheidet zwischen 2½D-Wiedergabe (Tiefe wird wahrgenommen durch Bewegung des Beobachters) und 3D-Wiedergabe (intuitiv wahrgenommene Tiefe durch Stereoskopie).

Diese Technologie erlaubt es, selbst im Ü-Wagen oder unter anderen ungünstigen Abhörbedingungen Auro-3D-Aufnahmen zu produzieren. Der Tonmeister kann seine gewohnte Abhörumgebung überall hin mitnehmen. Er kann per Knopfdruck zwischen verschiedenen Abhörsituationen wählen um beispielsweise das Klangbild außerhalb des Sweetspots zu überprüfen oder verschiedene Lautsprecher oder Wiedergaberäume zu vergleichen. Beim Konsumenten wird BRS fast immer die weit bessere Auro-3D-Wiedergabequalität erlauben als mit Lautsprechern im Wohnzimmer überhaupt möglich ist. Der Hörer ist zudem völlig unabhängig von der Wahl der Surround Sound Lautsprecherkonfiguration, ein BRS-Prozessor kann bei geeigneter Signalisierung im Prinzip für jedes Mehrkanal-Format die passende Wiedergabeordnung zur Verfügung stellen, dies ohne die vielen praktischen Probleme bei der korrekten Lautsprecher-Installation in der Wohnung.

BRS-Geräte werden dazu beitragen, die Markt-Akzeptanz von Produktionsqualität, Mehrkanalton und speziell der Auro-3D-Technik zu beschleunigen.

3. Mehrkanalton - psychoakustische Anforderungen

Das Gehör wertet die verschiedenen Schallfeldparameter zu entsprechenden räumlichen Hörereigniseigenschaften aus. In Tab. 2 ist grob dargestellt, welche Bedeutung der Direktschall, die frühen Reflexionen, der Nachhall und der umhüllende Schall für die einzelnen oben genannten Klangbildattribute sowie für die Klangfarbe besitzen. Der umhüllende Schall kann sowohl umgebenden diffusen Schall („Atmo“); als auch den hörbar ausklingenden Nachhall enthalten.

Attribute des Klangbildes im Saal	Direktschall	Frühe Reflexionen	Nachhall	Atmo
Richtung, Elevation	••	•		
Entfernung, Tiefe		••	•	
Räumlichkeit		••		
Umhüllung			•	••
Klangfarbe	••	•	••	

Tab. 2: Zusammenhang von Klangbild-Attributen und Schallfeldtypen.

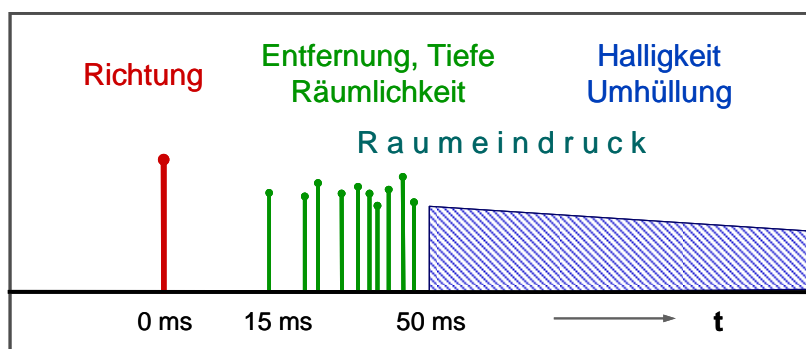


Abb. 2: Zuordnung der Klangbild-Attribute zum zeitlichen Ablauf des Raumschalls.

Das Gehör ist in der Lage, zwischen diesen drei Anteilen im natürlichen Schallfeld intuitiv („spontan“) zu trennen. Je mehr allerdings die örtlichen und zeitlichen Verhältnisse durch eine ungenügende Wiedergabe verschlechtert werden, desto schwieriger gelingt die intuitive Trennung. Ein einfaches Beispiel ist die Monoaufnahme, bei der sich Direktschall, frühe Reflexionen und Nachhall zu einem stark klangverfärbten Mulm mischen. Die räumliche Wahrnehmung erfolgt dann durch kognitive Aspekte, also durch eine bewusste Zuordnung z.B. „langer Hall“ → „großer Raum“ oder „leises Direktsignal“ → „große Entfernung“.

3.1. Reflexionen und Nachhall

Die indirekten Schallanteile dienen der Darstellung des Aufnahme-raums. Ihre Eigenschaften in Relation zum Direktschall bestimmen die räumlichen Attribute des Ereignisses. Der Zusammenhang ist in Abb. 2 schematisch dargestellt. Das natürliche Muster der frühen Reflexionen mit einer Verzögerung von 15 bis 50 ms spielt für das räumliche Hören eine besonders wichtige Rolle. Dieser Teil des reflektierten Schalls erfordert für die Aufnahme besondere Aufmerksamkeit. Die Attribute Entfernung, räumliche Tiefe, Räumlichkeit sind davon entscheidend beeinflusst. Das Gehör entnimmt den frühen Reflexionen die räumlichen Informationen und verwertet sie zu einem räumlichen Ereignis. Dies gelingt beim natürlichen Hören spontan und erstaunlich robust, da hier die Merkmale im Reflexionsmuster unverfälscht enthalten sind. Die wichtigsten sind:

- Die zeitliche Struktur in Relation zum Direktschall
- Ihre Pegel und Spektren
- Ihre Einfallsrichtungen horizontal und vertikal

Die Darstellung eines Raumes gelingt überzeugend, wenn das Gehör die Merkmale des reflektierten Schalls erkennen und interpretieren kann, das Reflexionsmuster „versteht“. Die Reproduktion muss deshalb entsprechend genau einer realen Raumsituation entsprechen. Das betrifft auch die räumliche Verteilung der Einfallsrichtungen für die frühen Reflexionen. Diese Forderung lässt sich mit Raummikrofonen nur schwer erfüllen (siehe Abschnitt 4), denn akustisches Übersprechen in den Kanälen einer Raummikrofon-Anordnung muss ausreichend klein gehalten werden (ca. 10 dB). Eine Reflexion beispielsweise von rechts oben vorne im Saal sollte ebenso reproduziert werden, nicht „in falsche Kanäle geraten“. Das ist z. B. der Fall, wenn für die Raummikrofon-Anordnung Kugelmikrofone eingesetzt werden.

Die Wahrnehmung von Entfernungen und räumlicher Tiefe wird entscheidend von den frühen Reflexionen geprägt. Man kann dies wunderbar nachvollziehen, wenn einer trocken aufgenommenen Quelle ausschließlich frühe Reflexionen (ohne Nachhall) zugemischt werden, die in Ihrer Struktur einem realen Raum zugeordnet werden können. Die Quelle wird entfernt wahrgenommen, ganz entsprechend dem Reflexionsmuster. Dies geschieht besonders stabil, wenn die Reflexionen aus den ursprünglichen Richtungen des oberen Halbraums einfallen. Die Reproduktion der Tiefenstaffelung erfordert sorgfältige Handhabung der frühen Reflexionen.

Fügt man den adäquaten Nachhall im passenden Pegelverhältnis hinzu, so wird die Wahrnehmung räumlicher Tiefe natürlicher und es kommt zur Wahrnehmung eines Raumeindrucks / ³. Die Nachbildung dieser zwei Attribute erzeugt bereits das Erlebnis eines virtuellen, natürlich wirkenden Raumes, wenn die Nachhalldauer kurz ist. Längerer Nachhall, z. B. der eines Konzertsaals oder einer Kirche verursacht als weiteres Attribut des räumlichen Hörens die Empfindung der Umhüllung.

3.2. Diffuser nicht reflektierter Schall (Atmo)

Das akustische Ambiente wird gebildet aus vielen einzelnen, in ihrer Gesamtheit nicht individuell lokalisierbaren, aber räumlich verteilten Schallquellen. Beispiele dafür sind das Blätterrauschen

³ Unter Raumeindruck versteht man die Wirkung von frühen Reflexionen und frühem Nachhall auf die Lokalisation. Das Hörereignis erscheint räumlich ausgedehnter („apparent source width“, ASW) und zeitlich verwischt aufgrund der Halligkeit im Raum.

des Walds, Publikumsgeräusche und -reaktionen und Applaus während eines Konzerts oder einer Veranstaltung. Im Gegensatz zum indirekten Schall kann dieser Anteil des umgebenden Schalls nicht mit Effektgeräten hergestellt werden, so dass eine geeignete Mikrofonierung erforderlich ist.

Für Aufnahmen im Saal liegt es nahe, dass die für die Aufnahme des Nachhalls eingesetzte Mikrofon-Konfiguration auch für die Erfassung der Atmo dienen kann. Durch sorgfältiges Ausprobieren und mit einiger Erfahrung lässt sich mit der Wahl der Kapsel-Richtcharakteristiken sowie Ihrer Positionierungen die Balance (beispielsweise unten/oben) zwischen Nachhall und Applaus / Publikumsgeräusch realitätsgetreu gestalten, aber nicht in allen Situationen und nicht während der Aufnahme. Mehr Flexibilität bietet der Einsatz eines 8-Kanal-Nachhallgerätes: Applaus / Publikumsreaktionen können damit nur auf die unteren Lautsprecher geroutet werden, der Nachhall auf alle acht.

4. Dreidimensionaler Ton mit Auro-3D

Die Lautsprecher in der oberen Ebene weisen natürlich dieselben Abbildungseigenschaften auf wie die in der Horizontalebene ohne den Center-Lautsprecher. Die stereofone Darstellung im Bereich L-C-R wird durch Zweikanal-Stereo in der Basis L_h - R_h der oberen Etage ergänzt, und ebenso können die oberen zusätzlichen Surround-Lautsprecher ebenso genutzt werden wie die unteren. Alleine daraus ergibt sich eine beachtliche Erweiterung des Gestaltungsspielraums. Interessant sind die Möglichkeiten, die sich aus dem Zusammenwirken beider Ebenen ergeben, einerseits hinsichtlich Quellendarstellung in der von den 5 Lautsprechern vor dem Hörer aufgespannten Fläche, andererseits hinsichtlich der Reproduktion von Reflexionen und diffusem Schall im dreidimensionalen Surround.

4.1. Darstellungsbereich zwischen unten und oben

Die Elevation von Quellen

Die Fläche vor dem Hörer zwischen L-R und L_h - R_h lässt sich leider nur am unteren und oberen Rand zur stereofonen Darstellung lokalisierbarer Quellen wie gewohnt nutzen. Die Lokalisation von Phantomschallquellen zwischen den unteren und oberen Lautsprechern ist extrem instabil bezüglich Laufzeitdifferenzen und auch abhängig vom Spektrum. Mit Panning zwischen oben und unten lässt sich eine Elevation nicht herstellen, es stellen sich unkontrollierbare klangliche und räumliche Effekten ein. Abbildung 5 zeigt eine praktische Untersuchung zur Pegelstereofonie zwischen übereinander angeordneten Lautsprechern (0° und 45°) vor dem Hörer [16]. Man erkennt, dass selbst im Sweetspot (bei korrekten Delay-Verhältnissen) keine robuste Lokalisation möglich ist, ganz ähnlich wie bei seitlichen Phantomschallquellen. Eine Elevation stationärer Quellen ist praktisch nicht realisierbar.

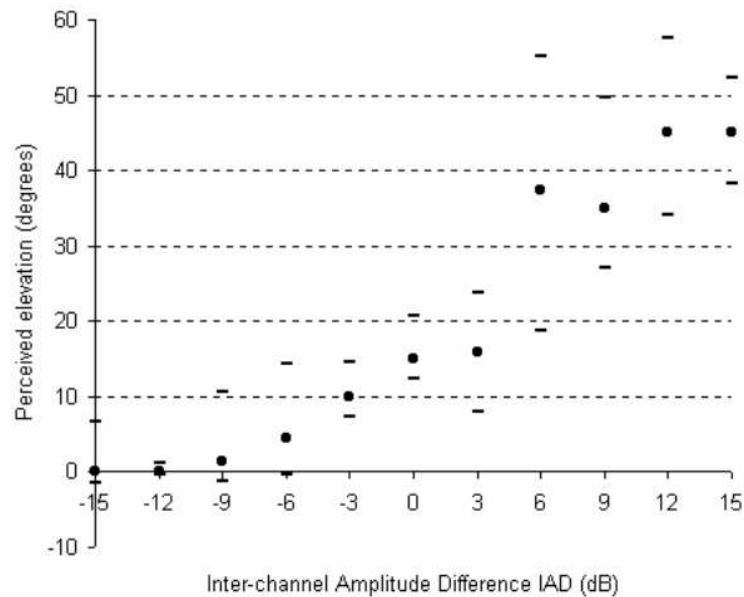


Abb. 5: Stereofone Abbildung in der Medianebene durch Pegelunterschiede (Lautsprecher 0° und 45°), aus [16]

Schon sehr kleine Laufzeitunterschiede führen darüber hinaus zu einer Auswanderung der Phantomschallquelle nach oben oder unten. Ein Delay von 0,5 ms reicht aus, um das Hörereignis in den einen oder anderen Lautsprecher zu verschieben, verbunden mit Klangverfärbung. Diesbezüglich ist die Hörzone in der Tiefe und in der Höhe stark eingeschränkt, Abbildung 6 illustriert Delay-Verhältnisse in einer Auro-3D Heimkino-Lautsprecherkonstellation.

Die Elevation einer stationären Quelle oder ihre Ausdehnung nach oben ist mit Hilfe der oberen Lautsprecher praktisch nicht erreichbar, besonders bei Berücksichtigung der Forderung einer großen Hörzone. Panning zu diesem Zweck funktioniert nicht und birgt nur die Gefahr der Klangverfärbung (die in vielen Fällen allerdings durch Diffusschall mehr oder weniger verdeckt wird). Die Verhältnisse stellen sich ähnlich dar wie bei den seitlichen Lautsprecherpaaren L-LS bzw. R-RS: Stabile Quellenpositionen sind allein die Lautsprecher. In bestimmten Grenzen sind jedoch bewegte Quellen darstellbar.

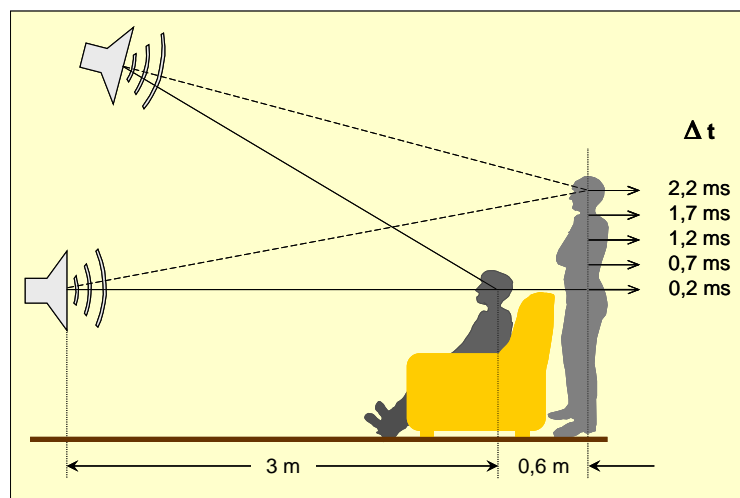


Abb. 6: Laufzeitdifferenzen bei Hörpositionen abweichend vom Sweetspot

Das Füllen der Flächen

Wesentlich günstigere Verhältnisse findet man bei der Reproduktion vieler einzelner, in ihrer Gesamtheit nicht individuell lokalisierbarer, aber räumlich verteilter Schallquellen (und Spiegelschallquellen = Reflexionen). Sie ähneln den Eigenschaften einer Groß-A-B-Konfiguration oder eines Decca-Trees: Obwohl die Richtungsabbildung wegen viel zu steiler Abbildungskurven [8] nicht praktikabel ist, ermöglichen sie dennoch eine platzunabhängige, klanglich ausgewogene Darstellung beispielsweise eines großen Klangkörpers und des reflektierten Schalls. Die Gefahr des „Lochs in der Mitte“ ist in vielen Aufnahmesituationen nicht gravierend, vor allem wenn der diffuse Schallanteil das Klangbild dominiert. Das Füllen der Flächen in der Höhe ist also möglich und ein wichtiges Gestaltungselement.

4.2. Reflexionen und Diffusschall im 3D-Surround

Besonders die frühen Reflexionen lassen sich auf diese Weise vorteilhaft in der Höhe verteilen. Die vielen unterschiedlichen Laufzeitdifferenzen individueller Reflexionen an den Kapseln bewirken dies. Die Reflexionen fallen in natürlicher Weise auch aus oberen Richtungen ein. Die bessere Verteilung der Reflexionen vermindert ihre räumliche Dichte, somit kann das Gehör die räumlichen Informationen besser erkennen. Abb. 4 veranschaulicht den Effekt beim Übergang von 2.0 über 5.1 nach Auro-3D. Wichtig ist dabei auch eine günstige Auswirkung auf die Klangfarbe, die mit einer verbesserten Wahrnehmung der Reflexionen verbunden ist.

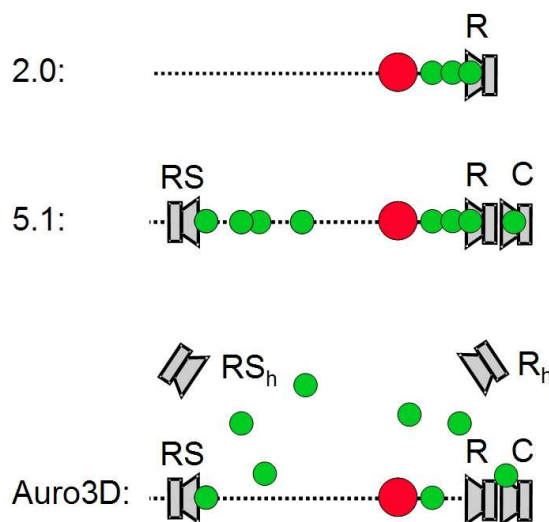


Abb. 4: Skizze zur räumlichen Verteilung der Reflexionsmuster in 2.0, 5.1 und Auro-3D

5. Zur Aufnahmetechnik für Auro-3D

5.1. Die Rolle der Psychoakustik bei der Suche nach der Aufnahmetechnik

Die Grundlagen des räumlichen Hörens helfen bei der gezielten Suche nach einer geeigneten Aufnahmetechnik für Auro-3D. Bei der Komplexität der Materie könnte man meinen, man komme durch Probieren schneller zum Ziel. Schließlich klingt es nur selten wirklich schlecht, wenn man einfach irgendetwas auf die Höhenlautsprecher gibt. Dass das nicht der Sinn der Sache ist, wird schnell klar, denn wir kennen bereits aus den Erfahrungen der Aufnahmetechnik für 5.1, dass zur Akzeptanz der diskreten Aufnahme diese wesentlich besser sein muss als ein automatischer Upmix (und der existiert bereits jetzt für Auro-3D!).

Es sind jedoch noch einige praktische Untersuchungen nötig, um die Hinweise, die direkt aus den besprochenen psychoakustischen Kenntnissen folgen, zu verifizieren, zu verfeinern und

schließlich in die Praxis umzusetzen. Deshalb kann jetzt noch kein konkreter Vorschlag gemacht werden. Der wissenschaftliche Ansatz wird dabei auch in der anderen Richtung erfolgen: Anhand gelungener Aufnahmen können allgemeingültige Regeln aufgestellt werden, und es ist nicht von Nachteil, wenn man auch weiß, warum die Aufnahme gut klingt und warum eine Regel besteht. Hat man nämlich vergessen, warum eine Regel besteht, wird aus ihr schnell ein alter Zopf.

Die Zielsetzung spielt eine erhebliche Rolle bei der Entscheidung für eine Aufnahmetechnik. Auch in 5.1 gibt es Techniken, die besser dazu geeignet sind, eine räumlich überzeugende Abbildung zu liefern und andere, die sich mehr dafür eignen, mit Stützmikrofonen zu arbeiten.

5.2. Kanaltrennung

Will man die oben geforderte räumliche Auflösung der Direktschallkomponenten, der Atmo, des Diffusschalls und/oder der Reflexionen erzeugen, so muss die Mikrofon-Anordnung eine ausreichende akustische Kanaltrennung gewährleisten. Andernfalls wäre die räumliche Anordnung mehrerer Lautsprecher wie bei Auro-3D nur wenig nützlich.

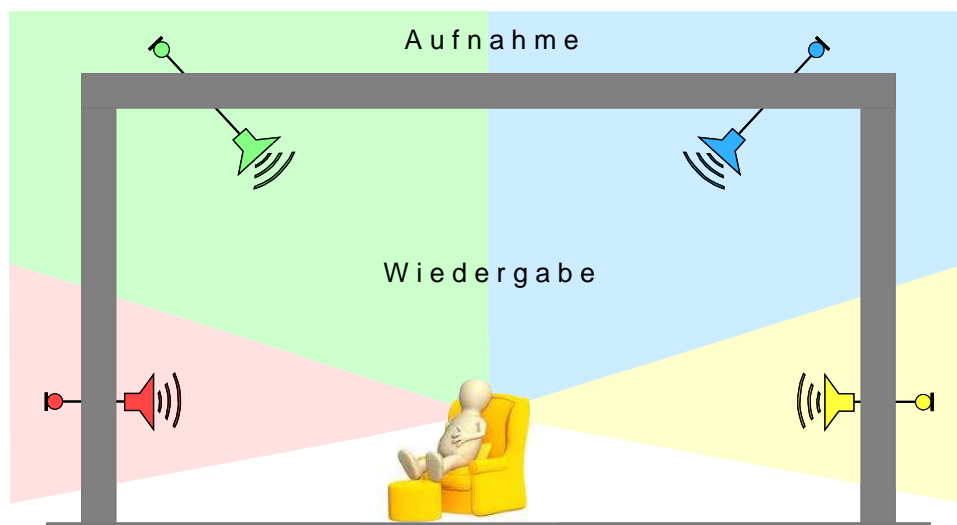


Abb. 7: Die Reproduktion der ursprünglichen Einfallsrichtungen erfordert eine entsprechende räumliche Trennung bei der Aufnahme

Ohne Frage ist die Realisierbarkeit der notwendigen akustischen Kanaltrennung für die Raummikrofon-Verfahren umso schwieriger, je mehr Kanäle für die Wiedergabe vorgesehen sind. Die Gefahr von unerwünschtem Crosstalk, also korrelierten Inhalten auf drei oder mehr Lautsprechern, steigt. Die Folge ist eine deutliche Klangverfärbung, die auch noch von der Hörerposition im Abhörraum abhängt. Eine Mikrofonanordnung ohne störenden Crosstalk ist aber bei 9 Kanälen äußerst schwierig zu bewerkstelligen! Es gibt zwei Lösungen, die auch in 5.1 funktionieren, sind: entweder Crosstalk durch optimierte Verfahren wie OCT Surround vermeiden oder durch größere Abstände und damit Laufzeiten Crosstalk in seiner Wirkung abschwächen.

Abbildung 8 zeigt beispielhaft für zwei Anordnungen, wie unterschiedlich deren erzeugtes Reflexionsmuster ist. Es wurde eine Quelle in einem Quaderförmigen Raum (mit Speigelschallquellen 1.Ordnung) simuliert, die eine Diraq-Stoss wiedergibt. Die ersten 50ms des resultierenden Signals im Sweet Spot einer Auro-3D-Lautsprecheranordnung sind dargestellt.

Im oberen Bild sieht man das Reflexionsmuster, das eine OCT-ähnliche 9-Kanal-anordnung erzeugt (OCT70 + 4 nach oben gerichtete Supernieren). Der Direktschall (schwarze Peaks) als auch die im Aufnahme-raum erzeugten Reflexionen werden sehr klar und ohne Crosstalk aus einer dem Aufnahme-raum entsprechenden Richtung wiedergegeben. Anders im unteren Bild: hier ist eine 9-kanalige Gross-AB Anordnung gezeigt. Es ist deutlich zu sehen, dass kaum verwertbare diskrete Reflexionen existieren und sich sehr schnell diffuser Nachhall bildet. Auch

das Direktsignal wird breit und diffus wiedergegeben. Dies kann allerdings durchaus gewünscht sein. In Räumen mit langer Nachhallzeit, in denen das Hörerlebnis vom diffusen Schall (von der Umhüllung) dominiert wird - beispielsweise in einer Kirche – entsteht so ein sehr guter Raumklang, der durch die Hinzumischung von Stützmikrofonen die nötige Präsenz sowie Stabilität der Abbildung erhält. Eine dem Aufnahmeraum entsprechende Abbildung sowie Tiefenstaffelung und Entfernungswahrnehmung wird nicht erreicht.

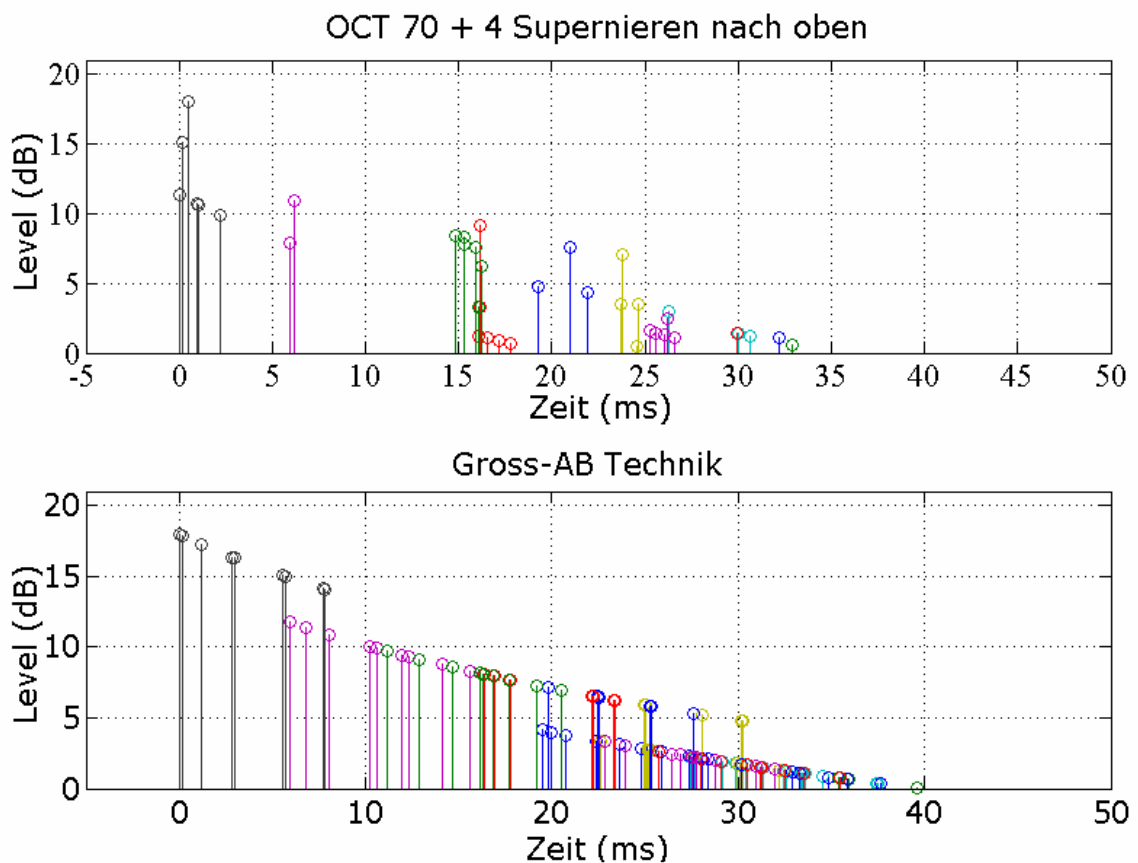


Abb. 8: Reflexionsmuster, erzeugt durch 2 verschiedene Mikrofon-Anordnungen im Sweet Spot einer Auro-3D-Lautsprecheranordnung. Beide Mikrofonanordnungen nehmen dieselbe Quelle im Aufnahmeraum auf. Zur Simulation wurde ein quaderförmiger Aufnahme-Raum erzeugt. Die Quelle gibt eine Dirac-Stoß wieder. Jeder Spiegelschallquellen (1.Ordnung) entspricht eine eigene Farbe der Peaks im Diagramm.

5.3. Nutzung von künstlichem Hall bzw. Faltungshall

Mit heutigen Technologien sind auch alternative Lösungen denkbar, die auf Faltungstechniken basieren, wobei die Rauminformationen entweder aus Messungen im aktuellen Aufnahmeraum, oder in bestehenden, akustisch anerkannten Räumen, oder aus Modellrechnungen gewonnen werden. Das prinzipielle Konzept beinhaltet für verschiedene Orte im Bereich der abzubildenden Quellen (z. B. Bühne) einen Faltungsalgorithmus, so dass einzelne Mikrofonensignale oder Mikrofongruppen mit den Raumimpulsantworten aus spezifischen Raumrichtungen gefaltet werden können. Dies sind für die Auro-3D 9.1 Wiedergabe je Quellensignal 8 Faltungen (mit den Impulsantworten aus 8 Raumecken). Abb. 4 zeigt das Prinzip für einen bestimmten Bühnenbereich (Mikrofongruppe A).

Werden die Raumimpulsantworten zugunsten einer realitätsnahen Raumdarstellung nicht im Modell gerechnet, können sie vorher im Aufnahmeraum mit geeigneten, richtenden Mikrofonen gemessen und abgespeichert werden. Zusätzlich können bei nicht ausreichender Richtwirkung des Mikrofons während der Messung störende, „grenzüberschreitende“ Schalleinfallrichtungen abgeschattet werden, z.B. der Direktschall. Die gemessenen Impulsantworten stehen dann für

alle zukünftigen Aufnahmen in dem Saal zur Verfügung. Gegebenenfalls kann in der Mischung auf die Einbeziehung des gefalteten Nachhalls verzichtet und der Diffusschall inklusive Atmo mit Raummikrofonen aufgenommen werden, so dass sich die Balance zwischen Nachhall und Applaus/Publikumsgeräusch realitätsgetreu gestalten lässt. Der Einsatz der Faltungstechnik befreit von einigen praktischen Problemen der Aufnahme und erweitert gleichzeitig den Gestaltungsspielraum.

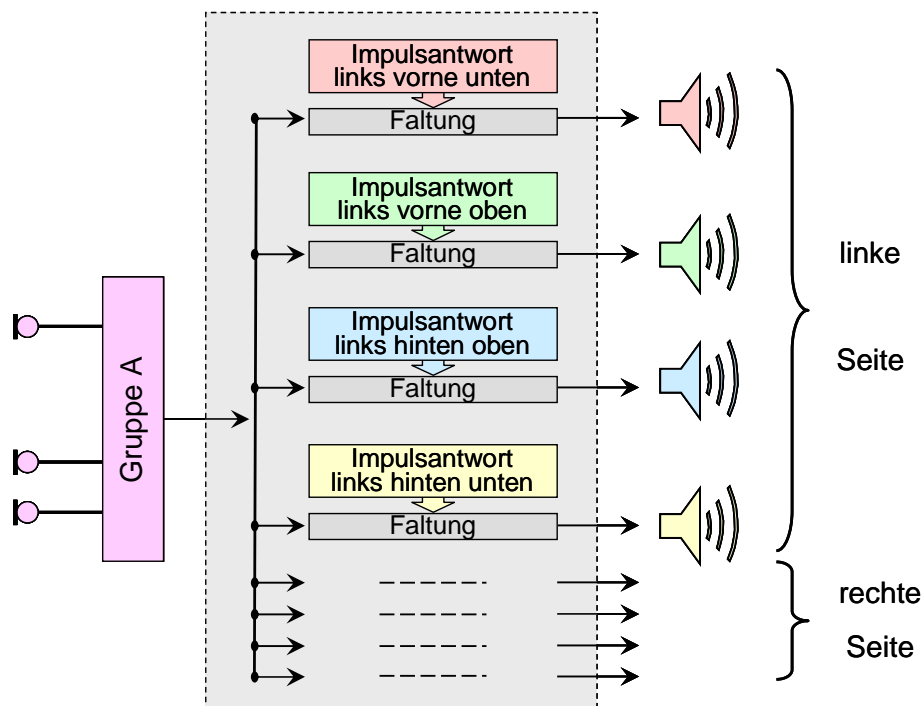


Abb. 9: Konzept eines Faltungsprozessors für achtkanalige Nachbildung der frühen Reflexionen

5.4. Diffuser Schall

Der diffuse Schall, also Nachhall oder Atmo, muss auch diffus beim Hörer ankommen. Dies kann mit Auro-3D gelingen, aber nur, wenn die zusätzlichen Lautsprecher auch geeignet beschriftet werden. Es ist notwendig, dass die diffusen Signale auf allen Lautsprechern ausreichend verschieden sind, also im gesamten Frequenzbereich dekorreliert. Gerade bei tiefen Frequenzen ist ausreichende Unabhängigkeit wichtig, da davon die Wahrnehmung der Umhüllung abhängt (siehe z.B. [14]). Es ist jedoch nun aufnahmeseitig durch die größere Anzahl der Kanäle, die unabhängig sein sollen, wesentlich schwieriger geworden. Da man mit Mikrofonen erster Ordnung in Bezug auf die Erzeugung dekorrelierter Signale sehr schnell an seine Grenzen kommt (z.B. lassen sich aus einem koinzidenten Aufbau wie Doppel-MS oder dem Soundfield-Mikrofon mit besten Willen nur vier ausreichend unabhängige Signale gewinnen [17]), führt dies unweigerlich dazu, dass der Mikrofonaufbau größer wird, da nur dadurch eine Dekorrelation durch Pegel- und Laufzeitunterschiede erreicht wird.

Zu beachten ist: die Korrelation des Diffusfelds ist aus zwei Gründen keiner sehr einfach messbare Größe. Erstens kann sie nur gemessen werden, wenn der Pegel des Diffusschalls deutlich höher ist als der von Direktschall und frühen Reflexionen - das ist nur in größerer Entfernung zur Quelle der Fall. Zweitens bringt es wenig, nur den Korrelationsgrad zu betrachten, denn dieser berücksichtigt nicht, dass gerade die (De-)Korrelation bei tiefen Frequenzen wichtig ist (siehe [13]).

Zur Ermittlung der minimal notwendigen Abstände und Winkel der Mikrofonpaare kann auf eine Studie zur Wirkung der Diffusfeldkorrelation zurückgegriffen werden. Wie in [13] gezeigt, können sowohl koinzidente, äquivalente sowie Laufzeitverfahren im Sinne der Vermeidung von Diffusfeldkorrelation geeignet sein. Abbildung 10 zeigt den Zusammenhang zwischen dem so

genannten DFI-Prädiktor (einem frequenzmäßig gewichteten Kohärenzgrad) und der im Hörversuch wahrgenommenen stereofonen Breite. Ein Monoanteil macht sich im Diffusfeld immer besonders störend bemerkbar, sowohl durch räumliche Enge als auch durch Klangverfärbung. Es wurden verschiedene koinzidente, äquivalente sowie reine Laufzeitanordnungen simuliert. Die Annahme ist: Nur diejenigen Anordnungen sind akzeptabel, die durch ihre geringe Diffusfeldkorrelation zu keiner eingeschränkten Wahrnehmung der räumlichen Breite (Wert auf der y-Achse > 2) führen. Sechs Anordnungen erfüllten diese Forderung: die Blumlein-Anordnung (2 koinzidente Achten +/- 45°), 2 Äquivalenzanordnungen mit Nieren und 3 Kugelarrangements mit Abständen größer 35 cm.

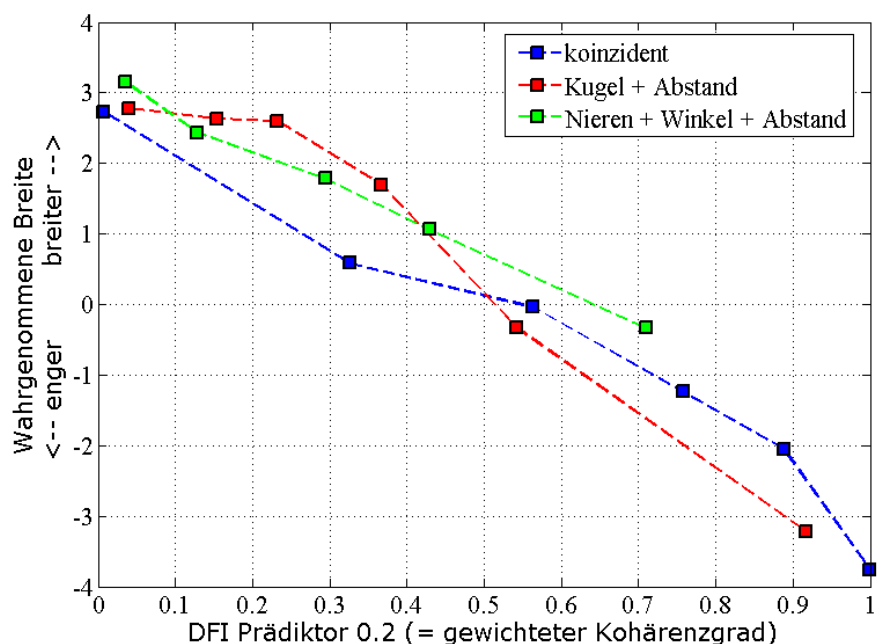


Abb. 10: Beziehung zwischen dem "DFI-Prädiktor" und der wahrgenommenen räumlicher Breite, aus [13]. Anordnungen von links nach rechts: Koinzident (blau): $r=0$ cm, +/-45°, Kugelanteil jeweils 0 (Blumlein); 0,4; 0,5 (Nieren-XY); 0,6; 0,7; 1 (Mono)
Äquivalent (grün): Nieren, +/-30°, Abstand jeweils 1m, 50cm, 20cm, 10cm, 2cm;
Laufzeit (rot): Kugeln: Abstand jeweils 1m, 50cm, 35cm, 20cm, 10cm, 2cm;

5.5. Design eines Auro3D-Hauptmikrofons

Ein Auro3D-Hauptmikrofon oder Atmomikrofon, das auch ohne Stützmikrofone funktioniert, sollte eine Reihe von Aspekte berücksichtigen, die aus den Überlegungen der vorangegangenen Abschnitte folgen. Viele der grundlegenden Regeln sind bereits bekannt, so die Gesetze der Richtungsabbildung (siehe „Image Assistant“ [15] und [16]) und der räumlichen Wahrnehmung ([1]). Neu ist, dass es durch die große Anzahl an Lautsprechern und dadurch Beziehungen zwischen Lautsprechern schwerer geworden ist, eine geeignete Mikrofonie zu finden, die allen Anforderungen gerecht wird. Andererseits ist Experimentieren erlaubt und führt oft zu einem gewissen Erfolg, denn schon alleine die akustische Anregung des oberen Teils des Wiedergaberaums kann vorteilhaft wirken.

Man sollte allerdings nicht den Fehler machen und zum Vergleich zwischen 5.1 und Auro3D einfach die oberen Lautsprecher ein- und ausschalten oder downmixen. Diese Art von Vergleich ist irreführend. Schließlich geht es auch bei dieser neuen Wiedergabetechnik darum, die Hörer/Kunden vom echten Mehrwert einer diskreten, tonmeisterlichen Mischung zu überzeugen. Dazu werden neben der verbesserten räumlichen Wiedergabe auch neue Ideen zur ästhetischen Nutzung der Höhenkanäle nötig sein.

6. Referenzen:

- [1] Theile, G.: "Natural 5.1 Music Recording Based on Psychoacoustic Principles". Nordic Sound Symposium XX, BOLKESJØ, 2001.
www.hauptmikrofon.de/theile/Multich_Recording_30.Oct.2001_.PDF
- [2] Wikipedia: "7.1 surround sound". Website http://en.wikipedia.org/wiki/7.1_surround_sound
- [3] Dabringhaus, W.: "2+2+2 Aufnahme - Verfahren". www.mdg.de/frame2.htm
- [4] Holman, T.: "10.2 channel surround sound". <http://en.wikipedia.org/wiki/10.2>
- [5] Van Baelen, W.: "Challenges for Spatial Audio Formats in the near Future", 26. Tonmeistertagung 2010, Tagungsbericht, ISBN 978-3-9812830-1-3, S. 196-205
- [6] Theile, G.: "Zur Kompatibilität von Kopfsignalen mit intensitätsstereofonen Signalen bei Lautsprecherwiedergabe: Die Klangfarbe". Rundfunktechn. Mitteilungen 4/1981, S. 146-154
- [7] Steickart, H.: "Kopfbezogene Stereophonie - neuere Erfahrungen bei Produktion und Rezeption". 15. Tonmeistertagung 1988, Tagungsbericht, S. 316-331
- [8] Dickreiter, M.; Dittel, V.; Hoeg, W.; Wöhr, M.: Handbuch der Tonstudioteknik, Band 1, Kapitel 5. K. G. Saur Verlag München, 2008, ISBN 978-3-598-11765-7
- [9] Wittek, H.: "Räumliche Wahrnehmung von virtuellen Quellen bei Wellenfeldsynthese". 23. Tonmeistertagung 2004, Tagungsbericht, S. 268-297.
http://hauptmikrofon.de/HW/Wittek_TMT2004_Paper_final.pdf
- [10] Horbach, U.; Pellegrini, R.; Felderhoff, U.; Theile, G.: "'Ein virtueller Surround Sound Abhörraum im Ü-Wagen". 20. Tonmeistertagung 1988, Tagungsbericht, S. 238-245
- [11] IRT-Website: "Binaural Room Synthesis BRS".
www.irt.de/de/produkte/produktion/binaural-room-synthesis-brs.html
- [12] Fey, F.: "Ohrenbetörend. Smyth Research SVS Realiser A-8". Studio Magazin 12, 2009, S. 24-34
- [13] Riekehof-Böhmer, H., Wittek, H., Mores, R.: "Voraussage der wahrgenommenen räumlichen Breite einer beliebigen stereofonen Mikrofonanordnung", 26. Tonmeistertagung 2010
- [14] Griesinger, D., "General overview of spatial impression, envelopment, localization, and externalization" in Proceedings of the 15th International AES Conference, Copenhagen, 1998, pp.136-149.
- [15] Wittek, H., "Image Assistant", JAVA-Applet auf www.hauptmikrofon.de, zuletzt besucht 23.2.2011
- [16] Barbour, J., "Elevation Perception: Phantom Images in the Vertical Hemi-sphere" in Proceedings of the 24th AES Conference on Multichannel Audio, The New Reality, June 2003
- [17] Wittek, H., Haut, C., Keinath, D. "Doppel-MS – eine Surround-Aufnahmetechnik unter der Lupe", 24. Tonmeistertagung 2006, Leipzig