

---

# Über Methoden zur Qualitätsbeurteilung von virtuellen Kopfhörern

---

Klaus J. Laumann







**Technische Universität München**

Lehrstuhl für Mensch-Maschine Kommunikation

# Über Methoden zur Qualitätsbeurteilung von virtuellen Kopfhörern

**Klaus Josef Laumann**

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines **Doktor-Ingenieurs** genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Norbert Hanik

Prüfer der Dissertation:

1. Prof. Dr.-Ing. habil. Hugo Fastl
2. Prof. Dr.-Ing. Klaus Diepold

Die Dissertation wurde am 12.04.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 14.11.2016 angenommen.



# Inhalt

<b>1 Einleitung</b>	<b>1</b>
1.1 Virtuelle Akustik . . . . .	1
1.2 Motivation und Abgrenzung der behandelten Thematik . . . . .	2
1.3 Aufbau der Dissertation . . . . .	4
<b>2 Binauraltechnik</b>	<b>7</b>
2.1 Grundlagen der Binauraltechnik . . . . .	8
2.2 Wiedergabe über Kopfhörer . . . . .	14
2.3 Binaurale Raumsynthese . . . . .	16
2.4 Wiedergabe über Lautsprecher . . . . .	18
<b>3 Schallfeldrekonstruktion und Schallfokussierung</b>	<b>33</b>
3.1 Wellengleichung für Schallfeldgrößen . . . . .	33
3.2 Kirchhoff-Helmholtz Integral . . . . .	35
3.3 Wellenfeldsynthese . . . . .	37
3.4 Schallfokussierung . . . . .	42
<b>4 Ein virtueller Kopfhörer basierend auf Wellenfeldsynthese</b>	<b>45</b>
4.1 Funktionsprinzip . . . . .	45
4.2 Prototyp . . . . .	47
4.3 Konzeptionelle und technische Weiterentwicklung des Prototyps . . . . .	50
4.4 Geräte und Systemkomponenten . . . . .	52
<b>5 Psychoakustische Untersuchungen am virtuellen Kopfhörer</b>	<b>55</b>
5.1 Lokalisation von binauralen Signalen . . . . .	55
5.2 Lateralisation . . . . .	59
5.3 Wahrnehmbarkeit des interauralen Kohärenzgrads . . . . .	62
<b>6 Experimente mit geänderten geometrischen Parametern</b>	<b>67</b>
6.1 Lokalisation von binauralen Signalen . . . . .	67
6.2 Anordnung der Sekundärquellen . . . . .	71
6.3 Lateralisation . . . . .	74
6.4 Wahrnehmbarkeit des interauralen Kohärenzgrads . . . . .	76
6.5 Konstellation der virtuellen Quellen . . . . .	78
<b>7 Ein neuer Ansatz zur Qualitätsbeurteilung von virtuellen Kopfhörern</b>	<b>85</b>
7.1 Diskussion der Versuchsergebnisse . . . . .	87
<b>8 Ausblick und Zusammenfassung</b>	<b>103</b>
8.1 Anwendung und Erweiterung des virtuellen Kopfhörers . . . . .	103
8.2 Zusammenfassung . . . . .	105
<b>Literatur</b>	<b>109</b>



# Vorwort

Auch ich möchte an dieser Stelle die Gelegenheit nutzen und einige persönliche Worte des Danks aussprechen. Ganz oben auf meiner Liste stehen die zwei Mentoren hinter diesem Promotionsvorhaben, mein Hochschullehrer und Doktorvater Professor Hugo Fastl vom Lehrstuhl für Mensch-Maschine-Kommunikation der Technischen Universität München und Doktor Günther Theile, der damalige Leiter der Abteilung Audiosystemtechnik am Institut für Rundfunktechnik (kurz: IRT). Sie haben mich den ganzen Weg lang mit all seinen Schwierigkeiten verlässlich begleitet. Ihre Hilfe, ihre konstruktiven Anregungen und ihr immer offenes Ohr für meine Fragen rechne ich ihnen sehr hoch an, zumal beide bereits seit 2008 im Ruhestand sind.

Ein Großteil der beschriebenen Experimente wurde während meiner dreijährigen Forschungstätigkeit am IRT durchgeführt. Mehrere von mir betreute Studenten haben im Rahmen von Abschlussarbeiten oder Praktika einen wichtigen Beitrag zu dieser Arbeit geleistet: Tobias Weissgerber und Ulrich Mühlbauer von der Technischen Universität München, Maria Scherr und Michael Meier von der Fachhochschule Deggendorf, Roman Franke von der Fachhochschule Düsseldorf und Herbert Petritsch von der Technischen und der Universität für Musik und darstellende Kunst Graz. Alle haben in dieser Dissertation ihre ganz persönliche Spur hinterlassen, von denen ich jede einzelne sehr zu schätzen weiß! Darüber hinaus danke ich Doktor Helmut Wittek, dass er alle Fragen zu seiner Vorarbeit bereitwillig beantwortet hat und dem IRT, dass es meine Forschungsarbeiten während dieser drei Jahre uneingeschränkt gefördert hat.

Was mich sehr beeindruckt hat, war der Zusammenhalt der Münchner Schule der Psychoakustik. Wenn ich als externer Doktorand auf Beistand von meiner Alma Mater angewiesen war, waren meine Kommilitonen an der Universität immer zur Stelle, insbesondere die Doktoren Florian Völk und Daniel Menzel sowie Jakob Putner, anfangs zusätzlich auch Doktor Stefan Kerber. Vielen Dank!

Als ich 2015 als Gastwissenschaftler an die TU München zurückkehrte, wurde ich zudem vom Fachgebiet Audio-Signalverarbeitung mit offenen Armen empfangen und unterstützt. Hier gilt mein Dank vor allem dem Leiter Professor Bernhard Seeber und Doktor Fritz Menzer.

Mein Manuskript wurde von Constanze Rau sorgfältig kontrolliert und kritisch kommentiert. Sie hat mir damit einen großen Dienst erwiesen, für den ich ihr sehr dankbar bin.

Zum Schluss ist es mir noch ein persönliches Anliegen, an meine beiden ehemaligen IRT-Kollegen Gerhard Stoll und Martin Link zu erinnern, die 2010 beziehungsweise 2012 jeweils viel zu früh aus dem Leben gerissen worden sind.



# 1

## Einleitung

Der technische Fortschritt – vor allem im Bereich der digitalen Signalverarbeitung – und die Leistungsfähigkeit moderner Rechner ermöglichen die Umsetzung von altbekannten Konzepten aus der Forschung zum räumlichen Hören. Daraus ist mit der Virtuellen Akustik inzwischen ein eigenständiges Fachgebiet entstanden, das sich diese Entwicklung zunutze macht und im Wesentlichen die Erzeugung und Anwendung auditiver virtueller Umgebungen behandelt.

### 1.1 Virtuelle Akustik

„Virtuelle Akustik“ bezieht sich auf denjenigen Anteil einer Virtuellen Realität, der das Gehör anspricht. Dabei lassen sich zwei elementare Methoden für die Erzeugung sogenannter „auditiver virtueller Umgebungen“ unterscheiden<sup>vgl. [28]</sup>:

**Kopfbezogener Ansatz:** Der Fokus liegt auf den Schallsignalen im Ohr, die auditive Umgebung wird gewissermaßen vom Standpunkt des Hörers aus betrachtet. Die gesamte räumliche Information liegt „binaural“ vor, das bedeutet, sie ist in den beiden „Ohrsignalen“ enthalten. Vorteilhaft ist beim kopfbezogenen Ansatz die geringe Datenrate, da pro Hörer lediglich zwei Audiosignale verarbeitet werden müssen.

**Raumbezogener Ansatz:** Bei ihm wird die auditive Umgebung als komplettes Schallfeld in einem (begrenzten) Raum betrachtet. Da er – im Gegensatz zum kopfbezogenen Ansatz – völlig unabhängig von den Hörern ist, ermöglicht der raumbezogene Ansatz prinzipiell eine eindeutige und verlustfreie Verarbeitung der gesamten räumlichen Information.

Beide Ansätze sind rein physikalisch motiviert, da sie sich ausschließlich mit Schall beschäftigen. Psychoakustisch begründete, sogenannte „stereofone“ Verfahren, die den Effekt einer beim Hören im überlagerten Schallfeld wahrgenommenen Phantom-schallquelle verwenden, werden hier per definitionem nicht zur Virtuellen Akustik gezählt, auch weil ihre räumlichen Abbildungsmöglichkeiten begrenzt sind.

Üblicherweise wird für den kopfbezogenen Ansatz die Bezeichnung „Binauraltechnik“ verwendet. Pragmatische binauraltechnische Versuche, bei denen zur Aufnahme von Ohrsignalen ein auf einer Schneiderpuppe befestigter Kopf aus Wachs mit eingebauten Mikrofonen zum Einsatz kam, wurden bereits 1931 durchgeführt<sup>[68]</sup>. Ein 1973 anlässlich der Internationalen Funkausstellung in Berlin im Rundfunk gesendetes Kunstkopfhörspiel<sup>[51]</sup> deutet dann schon auf Kommerzialisierungsversuche hin. Aber erst mit den technischen Möglichkeiten, Kopfposition und Kopfausrichtung eines Hörers instantan detektieren und die notwendige Signalverarbeitung in Echtzeit durchführen zu können, sind adaptive Systeme realisierbar, die auf Bewegungen des Hörers reagieren und somit in einer interaktiven virtuellen Umgebung eingesetzt werden können (vgl. Kap. 2).

Auch die theoretische Beschreibung des raumbezogenen Ansatzes lässt sich auf etwa 1930 datieren. Schon damals wurde vorgeschlagen, ein ganzes Schallfeld mit einer jeweils (im Idealfall unendlich) großen Anzahl von (im Idealfall infinitesimal kleinen) Mikrofonen und Lautsprechern aufzunehmen, zu übertragen und wiederzugeben<sup>[56]</sup>. Da diese Forderung nicht annähernd erfüllbar war, musste man sich lange Zeit auf wenige Übertragungskanäle beschränken und konzentrierte sich daher auf die Weiterentwicklung der stereofonen Verfahren. Mit der Verfügbarkeit leistungsfähiger digitaler Signalprozessoren ist man aber mittlerweile in der Lage, eine ausreichend große Anzahl an Audiosignalen zu berechnen, um damit künstliche Schallfelder zu erzeugen (vgl. Kap. 3).

## **1.2 Motivation und Abgrenzung der behandelten Thematik**

Die Generierung Virtueller Realitäten ist heutzutage nicht mehr nur Laborsystemen in der Forschung vorbehalten. Virtuelle Produktentwicklung oder Telepräsenzsysteme sind nur zwei Beispiele dafür, wie diese Technologie in Form von hoch entwickelten Anwendungen in der Industrie eingesetzt wird. Auch im Unterhaltungsbereich finden Konzepte aus dem Bereich der Virtuellen Realitäten zunehmend Einzug: 3D-Filme, die lange Zeit nur speziell ausgerüsteten Filmtheatern vorbehalten waren, faszinieren ein breit gefächertes Kinopublikum und erobern über 3D-Fernsehgeräte mittlerweile den Heimbereich, und bei vielen modernen Videospiele blickt der Spieler aus der Ich-Perspektive in detailreich und realistisch dargestellte Spielwelten.

Authentizität und Plausibilität sind wichtige Faktoren für die Akzeptanz virtueller Umgebungen. Der Benutzer soll dabei ein Gefühl der Präsenz erleben, so dass eine Immersion, das heißt ein Eintauchen in diese Virtuelle Realität, erfolgt<sup>[134]</sup>. An den oben genannten Anwendungsbeispielen lässt sich erkennen, dass dazu das Hauptaugenmerk meist auf der optischen Darstellung liegt. Da der Mensch seine Umwelt

multimodal wahrnimmt, sollten aber auch die weiteren Sinne mit konsistenten Reizen bedient werden<sup>[138]</sup>. Dem Gehör kommt dabei eine erhöhte Bedeutung zu, da Hörsinn und Sehsinn wichtige Informationen zur Orientierung in der Umgebung liefern. Es ist beispielsweise bekannt, dass sich die empfundene Qualität – und damit auch die Akzeptanz – der optischen Darstellung erhöht, wenn sie von passenden auditiven Sinneseindrücken begleitet wird<sup>[16,17]</sup>, oder dass sich, speziell bei audio-visuellen Medien im Unterhaltungsbereich, Bildqualität und Tonqualität in ihrer Wahrnehmung gegenseitig beeinflussen können<sup>[90]</sup>. Eine plausible, bestenfalls authentische Virtuelle Akustik kann also das Gesamterlebnis einer Virtuellen Realität verbessern.

Unabhängig von der Kombination mit einer optischen Darstellung hat sich die Virtuelle Akustik zudem eigenständige Anwendungsgebiete erschlossen: Mit der sogenannten „Auralisierung“ lassen sich in der Architektur Räume bereits während der Planungsphase Probe hören. Sie wird bei Bauvorhaben eingesetzt, bei denen die Raumakustik eine wichtige Rolle spielt, z. B. bei Räumlichkeiten für Musikdarbietung oder Musikbearbeitung wie Konzertsälen und Tonstudios<sup>[25,62]</sup>. Davon abgeleitet liefert das speziell für die Anforderungen im Rundfunk konzipierte BRS-Verfahren<sup>1</sup> sozusagen binaurale Kopien von Produktionsplätzen. Unter akustisch ungünstigen Arbeitsbedingungen – wie bei Außenproduktionen im Übertragungswagen – kann damit eine virtuelle Abhörmöglichkeit geschaffen werden, die sogar das Abmischen von Mehrkanalaufnahmen ermöglicht<sup>[76,77]</sup>. Gegenstand aktueller Forschung ist auch die Realisierung binauraler Telefonkonferenzen mit räumlich verteilten Sprechern, die eine Unterscheidung der Konferenzteilnehmer erleichtert<sup>[144]</sup>.

Viele Anwendungen aus der Virtuellen Akustik, auch die bisher erwähnten, greifen auf den kopfbezogenen Ansatz zurück<sup>[16,70]</sup>. Das liegt unter anderem daran, dass mit Kopfhörern schon lange die Möglichkeit zur Verfügung steht, Ohrsignale relativ unkompliziert wiederzugeben. Allerdings schottet diese Methode den Nutzer stark ab. Das kann einerseits erwünscht sein, da das Umfeld weniger gestört wird oder die Vertraulichkeit persönlicher Gespräche leichter garantiert ist. Andererseits stellt dies auch einen gravierenden Nachteil dar. Der Hörer ist sich der unnatürlichen Situation durch die Abdeckung seiner Ohren bewusst, was den psychologischen Eindruck bei ihm mindert, einer realen externen Schalleinwirkung ausgesetzt zu sein<sup>[49]</sup>. Daneben gibt es bei Kopfhörern offensichtliche Unannehmlichkeiten, wie Druckschmerzen und Wärmeentwicklung im Bereich der Ohren oder die Einschränkung des Bewegungsbereichs durch das Anschlusskabel. Daher existieren für den kopfbezogenen Ansatz auch alternative Verfahren, die eine Wiedergabe von Ohrsignalen über Lautsprecher statt über Kopfhörer ermöglichen. Ihre Ursprünge lassen sich bis in die 1960er-Jahre auf Schaltungen zurückführen, die für das Anhören von Kunstkopfaufnahmen

---

<sup>1</sup>Binaurale Raum-Synthese (kurz: BRS), die ursprünglich geprägte englische Bezeichnung „Binaural Room Scanning“ wurde damit ersetzt

mit einer Stereo-Lautsprecheraufstellung entworfen wurden<sup>[13]</sup>. Dabei besteht die Schwierigkeit darin, beide Ohrsignale individuell zu kontrollieren, obwohl die Ohren nicht mehr unabhängig voneinander beschallt werden können.

Ein solches Verfahren wird im Folgenden vorgestellt, untersucht und weiterentwickelt. Anders als oft üblich, ist es allerdings nicht a priori für die Wiedergabe binauraler Signale konzipiert. Im Fokus steht das allgemeinere Ziel, eine Kopfhörerwiedergabe über Lautsprecher zu emulieren. Dazu werden, um dem Hörer einen angemessenen Bewegungsspielraum einräumen zu können, kopfbezogener und raumbezogener Ansatz kombiniert. Das Verfahren ist dadurch nicht für eine spezielle Anwendung oder bestimmte geometrische Vorgaben optimiert und lässt sich somit flexibel an verschiedene Randbedingungen anpassen.

### 1.3 Aufbau der Dissertation

Nachdem im vorhergehenden Teil von **Kapitel 1** die behandelte Thematik motiviert und abgegrenzt wurde, ist die Arbeit ab dem nächsten Kapitel entsprechend der folgenden Inhaltsübersicht gegliedert:

In **Kapitel 2** werden Grundlagen der Binauraltechnik erläutert. Dieser Oberbegriff fasst alle Verfahren zusammen, die eine Aufnahme, Synthese oder Wiedergabe von Ohrsignalen beinhalten<sup>[70]</sup>. Aufbauend auf der systemtheoretischen Betrachtung des Unterschieds zwischen Kopfhörer- und Lautsprecherwiedergabe werden zwei Verfahren näher erklärt, die auch in den später beschriebenen Experimenten zum Einsatz kamen: erstens, die aus der Kunstkopf-Stereofonie entstandene Technik der „Binauralen Raumsynthese“<sup>[99]</sup>; zweitens, die auch als „transaural“ bezeichnete Wiedergabe von Ohrsignalen über Lautsprecher mit einem Filter zur sogenannten „Übersprechkompensation“<sup>[11]</sup>.

**Kapitel 3** befasst sich mit Möglichkeiten der Schallfeldrekonstruktion, wobei der Ansatz der „Wellenfeldsynthese“<sup>[20,21]</sup> besonders vertieft wird. Dieser leitet sich aus einer allgemeingültigen mathematisch-physikalischen Formulierung ab, die auf einer Auswertung der akustischen Wellengleichung mittels einer Randintegralmethode beruht<sup>[170]</sup>. Mit Wellenfeldsynthese kann Schall auch zu sogenannten „fokussierten Quellen“ gebündelt werden, wodurch sich Schallereignisse nahe am Hörer simulieren lassen<sup>[2,31,215]</sup>. Vervollständigt wird der Abschnitt durch einen kurzen Überblick über alternative Verfahren zur Schallfokussierung, wie dem Ansatz der „Zeitumkehr“<sup>[226]</sup>.

Basierend auf den in den Kapiteln 2 und 3 erwähnten technischen Verfahren wird in **Kapitel 4** ein System zur Reproduktion von Ohrsignalen vorgestellt, das im

weiteren Verlauf als „virtueller Kopfhörer“ bezeichnet wird. Es stützt sich auf Ideen und Vorarbeiten für eine Apparatur namens „Binaural Sky“<sup>[123,124]</sup>, die im Jahr 2009 in Deutschland patentiert<sup>[221]</sup> wurde. Das Kapitel enthält eine Erläuterung des Funktionsprinzips, sowie eine detaillierte Beschreibung des Prototyps und seiner technischen Weiterentwicklung.

In **Kapitel 5** sind die zur Untersuchung dieses Prototyps durchgeführten psychoakustischen Experimente beschrieben. Sie konzentrieren sich im Wesentlichen auf einen Vergleich zwischen realer und virtueller Kopfhörerwiedergabe, zum einen bezüglich der Phänomene Im-Kopf-Lokalisation und Lateralisation, zum anderen bezüglich der Wahrnehmbarkeit des interauralen Kohärenzgrades.

**Kapitel 6** behandelt ähnliche Versuche, die sich vor allem mit Auswirkungen der Abänderung von geometrischen Vorgaben des Prototyps beschäftigen.

Die Auswertung der Hörversuche erfolgt in **Kapitel 7**. Die Ergebnisse werden im Überblick präsentiert und mit Bezug auf das sogenannte Assoziationsmodell<sup>[189]</sup> diskutiert. Daraus wird ein Vorschlag zur Qualitätsbeurteilung virtueller Kopfhörer abgeleitet, der sich an bekannten Lateralisationsexperimenten<sup>vgl. z. B. [23,54]</sup> orientiert.

In **Kapitel 8** werden potentielle Anwendungsfälle und Erweiterungsmöglichkeiten des virtuellen Kopfhörers dargelegt. Im Rahmen einer Zusammenfassung findet sich abschließend ein kurzer Überblick über die wesentlichen Aspekte und die wichtigsten Ergebnisse der Arbeit.



# 2

## Binauraltechnik

Die Binauraltechnik ist ein Teilgebiet der Virtuellen Akustik und umfasst allgemein das gesamte Spektrum aller Techniken, die Ohrsignale verarbeiten<sup>[70]</sup>. Dabei versteht man als *Ohrsignale* die in den Gehörgängen eines Hörers herrschenden und damit auf die beiden Trommelfelle einwirkenden Schalldruckschwankungen.

Die Grundidee der Binauraltechnik beruht auf der Annahme, dass die auditive Wahrnehmung – insbesondere das Erkennen von Richtung und Entfernung einer Schallquelle relativ zum Hörer – vollständig durch eine Auswertung der Ohrsignale bestimmt ist. Daraus folgt direkt, dass sich ein Schallereignis wahrnehmungsgerecht in lediglich zwei voneinander abhängige Schalldruck-Zeit-Funktionen –  $p_l(t)$  für das linke und  $p_r(t)$  für das rechte Ohrsignal – transformieren lässt. Diese Zuordnung gilt als eindeutig, was bedeutet, dass ein bestimmtes *binaurales Signal*, also ein Paar von Ohrsignalen, eine immer gleiche zugehörige auditive Wahrnehmung, also ein bestimmtes *Hörereignis*, hervorrufen wird.

Genau genommen ist diese Annahme nicht haltbar, da sie unter anderem multimodale und kognitive Effekte vernachlässigt<sup>[25,S.12]</sup>: Die wahrgenommene Position einer Schallquelle kann sich beispielsweise zu einem passenden visuellen Stimulus hin verschieben<sup>[205]</sup>. Auch der Umkehrschluss trifft nicht automatisch zu, da die Zuordnung zwischen Ohrsignalen und Hörereignissen keineswegs eineindeutig ist: Verschiedene Ohrsignale können nämlich identische Hörereignisse bewirken, wie am Beispiel der Phantomschallquelle gezeigt werden kann<sup>[189]</sup>.

Trotz aller Einwände hat sich die Annahme aber als allgemein anerkannte und brauchbare Hypothese erwiesen<sup>[25,S.12]</sup>. Damit ist das Ziel der Binauraltechnik, die Wahrnehmung natürlicher Hörsituationen mit bestimmten Schallereignissen nachzubilden, indem sie deren Ohrsignale mit hinreichender Genauigkeit künstlich erzeugt.

## 2.1 Grundlagen der Binauraltechnik

Die *kopfbezogene Übertragungsfunktion*<sup>2</sup> spielt in der theoretischen Betrachtung der Binauraltechnik eine zentrale Rolle. Mit ihrer Hilfe lassen sich die Ohrsignale für bestimmte Hörsituationen berechnen, ohne dass eine direkte Aufnahme mit einem Kunstkopf oder Mikrofonen im Ohr durchgeführt werden muss. Man geht generell davon aus, dass sich alle auftretenden akustischen Vorgänge zumindest näherungsweise durch lineare zeitinvariante Systeme abbilden lassen. Nur dann sind die mathematischen Formulierungen aus der Nachrichtentechnik und die zugehörigen Verfahren der Signalverarbeitung, wie sie im Folgenden verwendet werden, gültig.

Ein lineares zeitinvariantes System ist vollständig durch seine Impulsantwort  $h(t)$  beschrieben. Im Spektralbereich wird diese durch die komplexwertige Übertragungsfunktion  $H(f)$  repräsentiert, die normalerweise als Bode-Diagramm aufgeteilt in Betrag  $|H(f)|$  und Phase  $\arg(H(f))$  dargestellt wird. Das Ausgangssignal  $y(t)$  ergibt sich durch Faltung des Eingangssignals  $x(t)$  mit der Impulsantwort  $h(t)$ ,

$$y(t) = x(t) * h(t) , \quad (2.1)$$

oder, gemäß der Produktregel<sup>[145,S. 313]</sup>, durch Multiplikation des in den Spektralbereich transformierten Eingangssignals  $X(f)$  mit der Übertragungsfunktion  $H(f)$ ,

$$Y(f) = X(f) \cdot H(f) , \quad (2.2)$$

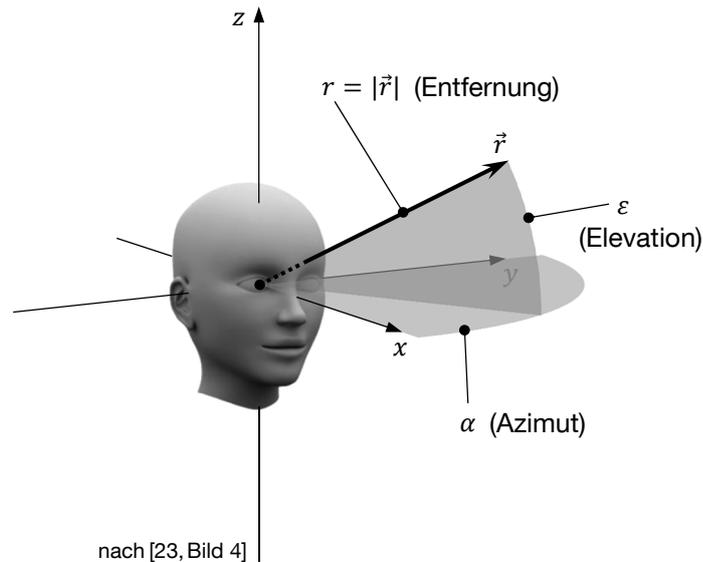
und der anschließenden Rücktransformation des Ausgangsspektrums  $Y(f)$  in den Zeitbereich. Abhängig davon, ob die Signalverarbeitung analog oder digital erfolgt, sind für die Faltung und die Transformation in den Frequenzbereich geeignete zeitkontinuierliche oder zeitdiskrete Berechnungsvorschriften auszuwählen. Wird die unabhängige Variable in eckige Klammern gesetzt ( $[t]$  oder  $[f]$ ), weist dies hier explizit auf eine Zeitdiskretion und damit auf digitale Signale hin, für zeitkontinuierliche bzw. verallgemeinerte Formulierungen werden, wie in den Gln. (2.1) und (2.2), runde Klammern verwendet. Außerdem werden Signale im Zeitbereich mit kleinen, im Frequenzbereich mit großen Buchstaben gekennzeichnet.

Zur Veranschaulichung soll folgende einfache Situation dienen: gegeben sei ein akustischer Monopol, also eine kugelförmig abstrahlende Schallquelle, und ein Zuhörer im freien Schallfeld. In der Praxis lässt sich das in guter Näherung mit einer geschlossenen Lautsprecherbox im reflexionsarmen Raum darstellen. Zuerst wird ein egozentrischer Referenzrahmen festgelegt: Die Position der Schallquelle wird in einem Kugelkoordinatensystem beschrieben, dessen Ursprung im Kopfmittelpunkt

---

<sup>2</sup>engl.: head related transfer function (kurz: HRTF); der ursprüngliche deutsche Begriff „Außenohrübertragungsfunktion“ wurde fast vollständig abgelöst

**Abbildung 2.1** Kopfbezogenes Koordinatensystem mit dem Koordinatentripel Entfernung, Azimut und Elevation  $(r, \alpha, \varepsilon)$ . Allgemein wird die  $x$ - $y$ -Ebene als Horizontalebene, die  $x$ - $z$ -Ebene als Medianebene und die  $y$ - $z$ -Ebene als Frontalebene bezeichnet. In dieser Arbeit werden in Anlehnung an die Anatomie zusätzlich die Begriffe sagittal für die  $x$ -Richtung, transversal für die  $y$ -Richtung und longitudinal für die  $z$ -Richtung benutzt.



des Hörers liegt. Dieses Koordinatensystem bewegt sich mit dem Hörer, das heißt, ändert der Hörer Position oder Ausrichtung des Kopfes, dann ändern sich demnach die Kugelkoordinaten Entfernung  $r$ , Azimut  $\alpha$  und Elevation  $\varepsilon$  der Schallquelle, auch wenn sich die Situation für einen Beobachter in einem Weltkoordinatensystem genau umgekehrt darstellt. Durch diese Festlegung kann mit dem Kugelkoordinatentripel  $(r, \alpha, \varepsilon)$  beziehungsweise dem Vektor  $\vec{r}$  ein Schallquellenort immer in Bezug auf den Hörer, oder genauer in Bezug auf den Kopf und damit auf die Ohren des Hörers, angegeben werden (vgl. Abb. 2.1).

Die kopfbezogene Übertragungsfunktion  $HRTF$  ist dann allgemein gegeben durch das Verhältnis der Schallspektren in den Ohren zu dem eines Referenzpunkts  $(r_0, \alpha_0, \varepsilon_0)$  im Schallfeld<sup>[39,70]</sup>. Gemäß Gl. (2.2) ergibt sich damit:

$$HRTF(f, r, \alpha, \varepsilon) = \begin{bmatrix} P_l(f) \\ P_r(f) \end{bmatrix} \cdot \frac{1}{P_{ref}(f, r_0, \alpha_0, \varepsilon_0)} \quad (2.3)$$

Eine vollständige kopfbezogene Übertragungsfunktion besteht demnach immer aus zwei Komponenten und charakterisiert die beiden Übertragungstrecken zwischen dem gewählten Referenzpunkt und den Ohren. Sie hängt neben der individuellen Anatomie des Hörers vor allem von der Position der Schallquelle im kopfbezogenen Koordinatensystem  $(r, \alpha, \varepsilon)$  ab.

Der Referenzpunkt kann unterschiedlich festgelegt sein. Eine übliche Vorgehensweise ist an die Definition der Freifeldübertragungsfunktion des äußeren Ohres angelehnt<sup>[23, S. 62 ff.]</sup>. Dabei bestimmt man als Referenz das Schallspektrum im Freifeld

an der Stelle, an der auch der Ursprung des kopfbezogenen Koordinatensystems liegt, das heißt  $r_0 = \alpha_0 = \varepsilon_0 = 0$ . Das bedeutet aber, dass dieses Bezugsspektrum separat und in Abwesenheit des Hörers ermittelt werden muss<sup>[34,70]</sup>. Der Vorteil besteht darin, dass dann der Einfluss des Abstands der Schallquelle  $r$  bei ausreichend großen Werten vernachlässigbar wird und die kopfbezogene Übertragungsfunktion nur noch von der Einfallrichtung  $(\alpha, \varepsilon)$  des Schalls abhängig ist<sup>[127]</sup>. Der kritische Abstand beträgt in etwa ein bis zwei Meter. Darunter, also für  $r \leq 1 \dots 2 m$ , lässt sich messtechnisch ein sogenannter „akustischer Parallaxenfehler“<sup>[36]</sup> feststellen, so dass ein hörbarer Unterschied<sup>[103]</sup> auftreten kann. Im Nahbereich des Hörers bleibt die kopfbezogene Übertragungsfunktion daher entfernungsabhängig.

Ermittelt man systematisch durch Verschieben der Schallquelle oder Verändern der Kopfausrichtung die kopfbezogene Übertragungsfunktion für eine Vielzahl unterschiedlicher Einfallrichtungen, lässt sich die Richtcharakteristik des Hörers darstellen. Sie beschreibt, wie der eintreffende Schall durch die Anatomie frequenzabhängig verzerrt wird. Als beeinflussend gelten Abschattung, Beugung und Reflexion vor allem am Kopf und an den Ohrmuscheln, aber auch an den Schultern und am Körper<sup>[49,94]</sup>. Die Definition ist besonders zweckmäßig, wenn Zeitpunkt und Richtung der am Hörerort einfallenden Schallanteile einzeln bekannt sind. Bei Auralisierungsverfahren zum Beispiel werden sie mittels geometrischer oder wellentheoretischer Modelle berechnet<sup>[25, Kap. 2] [62]</sup>. Dem Superpositionsprinzip<sup>[33, S. 704]</sup> gemäß dürfen die Anteile überlagert werden, wodurch man die resultierenden Ohrsignale auch für komplexe Situationen mit Schalleinfall aus unterschiedlichen Richtungen erhält.

Eine zweite mögliche Vorgehensweise besteht darin, als Referenz das Schallspektrum an der Quelle zu wählen, das heißt  $(r_0, \alpha_0, \varepsilon_0) = (r, \alpha, \varepsilon)$ . Die kopfbezogene Übertragungsfunktion beschreibt dann die komplette Schallausbreitung von der Quelle bis zum Hörer und berücksichtigt dabei unter anderem auch die Laufzeit des Schalls. Daher kann der Abstand der Schallquelle nicht vernachlässigt werden. Allerdings ist bei dieser Definition keine Referenzmessung in Abwesenheit des Hörers nötig und es sind keine Freifeldbedingungen gefordert. Eine solche kopfbezogene Übertragungsfunktion kann demnach für beliebige Übertragungstrecken auch in reflektierenden Umgebungen angegeben werden. Sie enthält, neben den anatomischen Übertragungseigenschaften, auch alle weiteren akustischen Parameter der Übertragungstrecke, wie die Richtcharakteristik der Quelle oder den Einfluss der Raumakustik. Damit beschreibt sie eine ganz bestimmte Hörsituation und ist daher nicht verallgemeinert anwendbar. Auf diese alternative Definition wird zur Ermittlung von Ohrsignalen in festgelegten Wiedergabesituationen mit bekannten Quellsignalen zurückgegriffen, wobei auch hier mehrere Übertragungstrecken überlagert werden dürfen. Dies ist zum Beispiel beim BRS-Verfahren (vgl. Abschn. 2.3) oder bei der Bestimmung von Filtern zur Übersprechkompensation (vgl. Abschn. 2.4) relevant. Im Kontext des BRS-Verfahrens übernahm man zur klaren Unterscheidung für eine so definierte und

in den Zeitbereich transformierte kopfbezogene Übertragungsfunktion den Begriff *binaurale Raumimpulsantwort*<sup>3</sup>[112], was auch im Folgenden beibehalten werden soll.

Ob ein binaurales Signal aufgenommen, eine kopfbezogene Übertragungsfunktion oder eine binaurale Raumimpulsantwort ermittelt werden soll, in allen Fällen müssen Ohrsignale gemessen werden. Schon früh setzte man dafür nachgebildete Köpfe mit Mikrofonen im Ohr, zum Beispiel präparierte Puppen<sup>[29,45,68]</sup>, ein. Mittlerweile sind solche Messanordnungen weit entwickelt und werden *Kunstkopf*<sup>4</sup> genannt. Sie sind in verschiedener Ausführung, teilweise zugeschnitten auf ein bestimmtes Einsatzgebiet, kommerziell erhältlich<sup>[49,137]</sup>.

Die Kunstkopf-Stereofonie<sup>5</sup>, also die direkte Aufnahme realer Hörsituationen mit Kunstköpfen, wurde eine Zeit lang enthusiastisch verfolgt. Im Vergleich mit anderen üblichen Aufnahmeverfahren rufen Kunstkopf-Aufnahmen nämlich eine wesentlich natürlichere Wahrnehmung der räumlichen Verhältnisse hervor. Allerdings müssen sie dazu mit Kopfhörern abgehört werden. Für die direkte Lautsprecherwiedergabe sind sie prinzipbedingt ungeeignet, was eine Akzeptanz im Rundfunk und in der Musikindustrie verhinderte<sup>[24,51]</sup>. Zudem zeigten sich damals bei systematischen Untersuchungen noch weitere Mängel: erstens orteten viele Versuchspersonen die eigentlich vor ihnen liegenden Schallquellen oftmals hinter dem Kopf, seltener auch umgekehrt, was als Richtungsinversion oder vorne/hinten-Vertauschung bezeichnet wird. Oder es kam, zweitens, zu einer sogenannten Im-Kopf-Lokalisation<sup>6</sup>, das heißt, die Entfernungswahrnehmung ist nicht korrekt und die gesamte Szene wird nur sehr nahe am oder sogar komplett innerhalb des Kopfes gehört. Drittens, traten vor allem für vorne liegende Schallquellen deutliche Richtungsabweichungen nach oben oder zur Seite hin auf<sup>[18,78]</sup>.

Daher beschäftigte man sich – neben diversen Versuchen, die Kunstköpfe durch konstruktive Maßnahmen zu optimieren und Fehler in der Übertragungskette zu identifizieren – eingehend mit Messungen direkt am Menschen mittels Sonden- und Miniaturmikrofonen, um das Übertragungsverhalten der Ohren genauer zu analysieren. Dabei hat sich die Methode mit blockierten Gehörgängen, also mit außen in Ohrenstöpseln eingesetzten Miniaturmikrofonen, durchgesetzt, weil sie eine höhere Reproduzierbarkeit der Messergebnisse gewährleistet<sup>[71,91,131,152]</sup>. Die weitere Schallausbreitung im Gehörgang bis zum Trommelfell ist als richtungsunabhängig erwiesen und muss nur durch eine wiedergabeseitige Kalibrierung berücksichtigt werden<sup>[70]</sup>. Eine kopfbezogene Übertragungsfunktion berechnet man dabei aus der Systemantwort auf ein geeignetes Anregungssignal, das den hörbaren Spektral-

---

<sup>3</sup>engl.: binaural room impulse response (kurz: BRIR)

<sup>4</sup>engl.: dummy head, artificial head, mannequin, manikin

<sup>5</sup>auch: kopfbezogene/kopfbezügliche Stereofonie

<sup>6</sup>auch: Im-Kopf-Lokalisiertheit, Im-Kopf-Ortung

bereich umfasst. Ideale Impulse finden wegen praktischer Einschränkungen keine Anwendung, lange bewährt haben sich Maximalfolgen<sup>7</sup> und neuerdings gewobbelte Signale in der Form eines mehrmals wiederholten, kontinuierlich exponentiell in der Frequenz ansteigenden Sinustons<sup>8</sup>[132].

In aufwendigen Reihenuntersuchungen stellte man dann in den Beträgen der kopfbezogenen Übertragungsfunktionen deutliche inter-individuelle Unterschiede fest. Zwar lässt sich grundsätzlich ein ähnlicher Verlauf erkennen, aber besonders im Bereich oberhalb von etwa  $7\text{ kHz}$  weisen die Amplitudengänge mehrere markante Spitzen und Einbrüche auf, die von Versuchsperson zu Versuchsperson bei anderen Frequenzen auftreten<sup>[70,130,216,219]</sup>. Nicht einmal die Ohren einer einzelnen Testperson zeigen ein identisches Übertragungsverhalten: Bei zwei symmetrisch zur Medianebene liegenden Schallquellen, das heißt  $(r_1, \alpha_1, \varepsilon_1) = (r_2, -\alpha_2, \varepsilon_2)$ , liefern die beiden Messungen am jeweils zugewandten Ohr im Normalfall zumindest leicht voneinander abweichende Übertragungsfunktionen<sup>[216]</sup>.

Das „Hören mit fremden Ohren“ galt bereits bei der Kunstkopf-Stereofonie als Ursache der erwähnten Diskrepanzen zwischen beabsichtigter und tatsächlicher Wahrnehmung. Wie verschiedene Studien bestätigen, nehmen dadurch insbesondere Richtungsinversionen zu<sup>[131,212]</sup>. Dagegen lässt sich durch die akkurate Aufnahme und Wiedergabe von individuellen Ohrsignalen eine fast realitätstreue Wahrnehmung erreichen<sup>[125,131]</sup>. Man stellte andererseits aber auch fest, dass sich Versuchspersonen an das Hören mit fremden Ohren gewöhnen können<sup>[219]</sup>. Zudem gibt es „typische“ fremde Ohren, mit denen bessere Ergebnisse erzielt werden als mit anderen<sup>[70,125,212]</sup>. Kunstköpfe, die solche typischen – anstatt beliebiger oder „gemittelter“ – Ohren nachbilden, schneiden daher in Vergleichen ebenfalls besser ab. Trotzdem sind am Menschen aufgenommene Ohrsignale, auch fremde, denen von Kunstköpfen noch oft überlegen<sup>[49,69,70,125]</sup>. Daraus resultieren beispielsweise Vorschläge, kopfbezogene Übertragungsfunktionen entweder an einer sorgfältig ausgewählten Versuchsperson zu messen<sup>[125]</sup>, oder an mehreren verschiedenen zu ermitteln und die jeweils am besten passende später individuell auswählen zu lassen<sup>[156]</sup>, oder auch mit anderen Maßnahmen zu individualisieren<sup>[224]</sup>.

Allen bisher erwähnten Erkenntnissen liegen statische Aufnahmen zugrunde. Das bedeutet, dass die in den Hörversuchen verwendeten Ohrsignale – menschliche wie künstliche – für eine fixierte Kopfposition und Kopfausrichtung aufgenommen oder berechnet wurden. Wie bei der Kunstkopf-Stereofonie sind die Szenen bei der Wiedergabe dann aber völlig statisch und bleiben an den Hörer gekoppelt. Bewusste Bewegungen des Hörers in der Szene, wie ein Wegdrehen von der Schallquelle, sind so

---

<sup>7</sup> *engl.*: maximum length sequences (kurz: MLS)

<sup>8</sup> auch: Gleitsinus, *engl.*: chirp, sweep; daher oft rückübersetzt als (logarithmischer/exponentieller) Sinus-Sweep

nicht möglich, da sich die gesamte Szene stattdessen mit dem Hörer bewegen würde. Dieser unnatürliche und damit unerwünschte Effekt kann mit einer dynamischen Binauraltechnik, also einem System, das auf den Hörer reagiert, verhindert werden. Da sich im egozentrischen Referenzrahmen mit jeder Bewegung die kopfbezogenen Koordinaten einer Schallquelle ändern, muss dazu aber bei der Berechnung der Ohrsignale immer die aktuell gültige kopfbezogene Übertragungsfunktion oder binaurale Raumimpulsantwort herangezogen werden (vgl. Abschn. 2.3).

Mit einem pragmatischen Experiment wurde ein wichtiger Vorteil von dynamischen Systemen nachgewiesen: Als Referenz diente eine Lautsprecheraufstellung gemäß dem von der Internationalen Fernmeldeunion<sup>9</sup> empfohlenen Mehrkanal-Stereoformat 3/4<sup>[80]</sup> mit sieben Lautsprechern in zwei Meter Abstand und auf Ohrhöhe einer sitzenden Person. In dieser Referenzszene wurde am optimalen Abhörpunkt im Zentrum ein auf einem Drehmotor montierter Kunstkopf aufgestellt. Im Nebenraum hörte eine Versuchsperson die vom Kunstkopf aufgenommenen Ohrsignale direkt mit einem Kopfhörer ab und konnte dabei die Ausrichtung des Kunstkopfes steuern. Die Mechanik arbeitete so schnell und genau, dass der Kunstkopf den Kopfdrehungen der Versuchsperson nahezu synchron folgte. Mit diesem Aufbau wurde eine präzise Richtungswahrnehmung erreicht, die der beim natürlichen Hören entsprach, als sich anstelle des Kunstkopfes die Versuchsperson selbst in der Referenzszene befand<sup>[111]</sup>.

Wurde der Motor jedoch abgeschaltet, so dass der Kunstkopf fixiert war, traten die bekannten Probleme der Kunstkopf-Stereofonie, deutliche Richtungsabweichungen und häufige Richtungsinvansionen auf<sup>[112]</sup>. Die zusätzlichen Informationen, die ein dynamisches binaurales System dem Hörer liefern kann, kompensieren offensichtlich andere Ungenauigkeiten der Ohrsignale, die hier durch das Hören mit den fremden Ohren des Kunstkopfes entstehen. Zumindest hinsichtlich einer korrekten Richtungswahrnehmung kann man dann auf die Ermittlung individueller kopfbezogener Übertragungsfunktionen verzichten.

Der beschriebene Versuch mit dem ferngesteuerten Kunstkopf wurde sowohl in einem Referenzabhörraum gemäß der Spezifikation der Europäischen Rundfunkunion<sup>10</sup><sup>[53]</sup>, als auch in einem reflexionsarmen Raum durchgeführt. Im Vergleich zeigt sich, dass die Richtungswahrnehmung unter Freifeldbedingungen ungenauer ist<sup>[112,114]</sup>. Hierzu gibt es aber bestätigende<sup>[18]</sup> wie widersprechende<sup>[15]</sup> Ergebnisse. Mehrere Untersuchungen stellten jedoch übereinstimmend fest, dass die Berücksichtigung von Reflexionen die Distanzwahrnehmung verbessert und sich so das Problem der Im-Kopf-Lokalisation verringern lässt<sup>[15,18,201]</sup>.

---

<sup>9</sup>engl.: international telecommunication union (kurz: ITU)

<sup>10</sup>engl.: European broadcasting union (kurz: EBU)

## 2.2 Wiedergabe über Kopfhörer

Zur Wiedergabe von aufgenommenen oder synthetisierten Ohrsignalen greift man in der Binauraltechnik bevorzugt auf Kopfhörer zurück. Der Vorteil liegt auf der Hand: Durch die Platzierung der Schallwandler direkt am Ohr sind die Schallsignale im Ohr relativ einfach kontrollierbar. Jedes Ohr kann individuell beschallt werden und ist von Störeinflüssen weitgehend abgekoppelt. Anders als bei Lautsprecherwiedergabe wird dadurch eine Beeinträchtigung der Schallübertragung durch äußere Einflüsse, wie raumakustisch bedingte Reflexionen oder Störschall, vermieden. Auch ein Übersprechen auf das jeweils andere Ohr ist, wenn es überhaupt auftritt, vernachlässigbar gering, so dass ohne weitere Maßnahmen eine sehr hohe Kanaltrennung zwischen dem linken und rechten binauralen Signal erreicht wird (vgl. Abb. 2.3).

Ähnlich wie die kopfbezogenen Übertragungsfunktionen (vgl. Abschn. 2.1) zeigen auch an mehreren Testpersonen durchgeführte Vergleichsmessungen von Kopfhörerübertragungsfunktionen – zwischen Kopfhörermuschel und einem Messpunkt im Gehörgang – wieder größere individuelle Unterschiede. Zudem führen wiederholte Messungen an derselben Testperson durch einen veränderten Sitz des Kopfhörers zu Abweichungen<sup>[40,128,191,219]</sup>, außerdem können auch mehrere Kopfhörerexemplare gleicher Bauart beziehungsweise mit gleicher Modellbezeichnung eines Herstellers Unterschiede in den Kopfhörerübertragungsfunktionen aufweisen<sup>[219]</sup>.

Für die optimale Wiedergabe individueller Ohrsignale ist also eine für Person und Gerät spezifische Kalibrierung des verwendeten Kopfhörers nötig, die auch die beim mehrfachen Aufsetzen des Kopfhörers auftretende Variabilität berücksichtigt. Am besten benutzt man dazu dieselbe Messanordnung, mit der auch die Ohrsignale selbst aufgenommen werden<sup>[70,131,216]</sup> und erlaubt der Versuchsperson, den Kopfhörer selbst zu platzieren<sup>[70,128]</sup>. Wird dabei mit blockierten Gehörgängen gearbeitet, sollte der Kopfhörer freifeldäquivalent an die Ohren ankoppeln<sup>[128]</sup>, so dass eher solche mit offener Bauart eingesetzt werden. Aus diesem Grund tauchen immer wieder „extra-auraler Kopfhörer“ oder „Kopflautsprecher“ genannte Konstruktionen auf, bei denen die Schallwandler zwar nahe am Kopf sitzen, die Ohren jedoch frei bleiben<sup>[128,155]</sup>. Meist benutzt man aber bewährte, circum-aurale oder supra-aurale, also ohrumschließende oder ohraufliegende, hochwertige Kopfhörermodelle. Die im Endverbrauchermarkt mittlerweile weit verbreiteten Einsteckhörer<sup>11</sup> sind hingegen in der Binauraltechnik bislang wenig beachtet worden und werfen messtechnische Herausforderungen auf<sup>[71]</sup>.

Eine individuelle Kalibrierung von Kopfhörern ist natürlich nicht immer möglich und oft nicht praxistauglich. Werden keine individuellen Ohrsignale wiedergegeben,

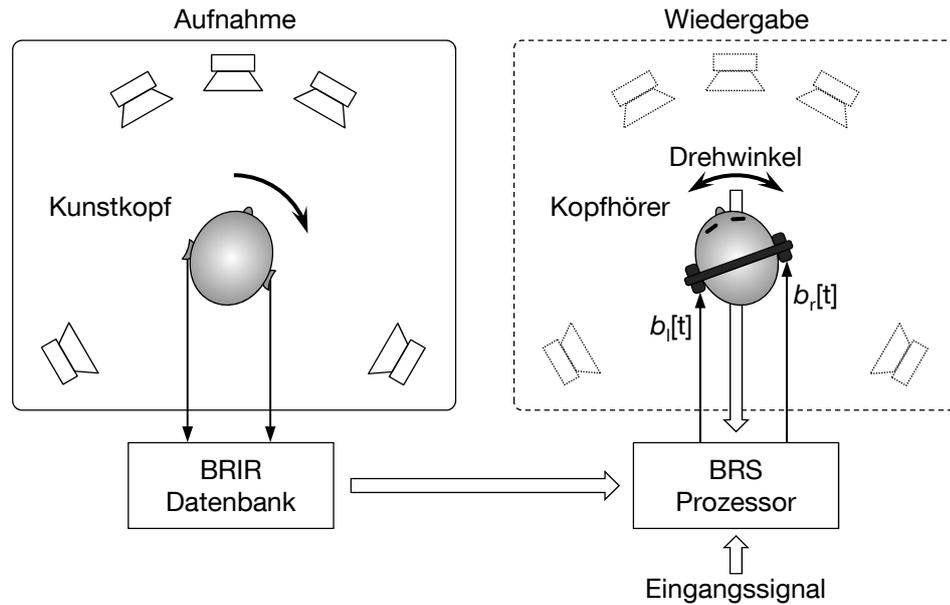
<sup>11</sup>intra-aurale Kopfhörer, *engl.*: in-ear headphones, earbuds, earphones

wird normalerweise von vorneherein darauf verzichtet, da bereits die Ohrsignale fehlerbehaftet sind. Neuere Untersuchungsergebnisse<sup>[107,116]</sup> unterstützen ein solches Vorgehen und empfehlen dann entweder eine am Aufnahmekopf vorgenommene<sup>[107]</sup> oder auch eine gemittelte<sup>[107,116]</sup> statt einer individuellen Kalibrierung.

Aus der Kunstkopf-Stereofonie stammt die Idee der entkoppelten Kalibrierung: Die Definition einer geeigneten Schnittstelle zwischen Aufnahme und Wiedergabe soll eine möglichst korrekte aber höreronabhängige Systemkette gewährleisten und das beliebige Austauschen von Komponenten auf beiden Seiten erlauben. Dazu wird eine Referenzsituation festgelegt, in der das Aufnahmesystem so entzerrt wird, dass es eine frequenzunabhängige Übertragungsfunktion aufweist. Die Wiedergabeseite wird derart kalibriert, dass die Referenzsituation möglichst unverfälscht reproduziert und damit auch die Aufnahmeentzerrung ausgeglichen wird<sup>[98]</sup>. Die Wahl der Referenz ist im Prinzip völlig beliebig, solange man an die Schnittstelle keine zusätzlichen Kompatibilitätsanforderungen stellt<sup>[24,S.90f.]</sup><sup>[190]</sup>. Man bezieht sich jedoch gerne auf eine standardisierte Situation wie eine frontale Beschallung im Freifeld oder ein ideales Diffusfeld. Für diese beiden Referenzsituationen existieren daher detaillierte Vorgaben, wie eine über mehrere Personen gemittelte Kalibrierung von Kopfhörern mit Hilfe von Lautstärkevergleichen oder Schalldruckmessungen im Gehörgang durchzuführen ist<sup>[52,79]</sup>. Mit solchen Kopfhörern lassen sich entsprechend entzerrte binaurale Aufnahmen direkt abhören.

Die sogenannte Diffusfeldentzerrung von Kunstkopf und Kopfhörer bietet weitere Vorteile, die theoretisch durch ein Assoziationsmodell für das räumliche Hören<sup>[189]</sup> begründet sind: erstens, entstehen beim Abhören von diffusfeldentzerrten Kunstkopfaufnahmen über Lautsprecher weniger starke Klangverfärbungen<sup>[24,S.92f.]</sup><sup>[190]</sup>, und zweitens, kann man einen diffusfeldentzerrten Kopfhörer als klangfarbenneutral ansehen<sup>[190,191]</sup>. Er wird auch von Tonmeistern in Hörversuchen mit Musikbeispielen gegenüber der Freifeldentzerrung oder der vom Hersteller vorgesehenen Entzerrung bevorzugt<sup>[24,S.94]</sup><sup>[191]</sup>. Außerdem stellt das ideale Diffusfeld eine für alle Schalleinfallrichtungen gemittelte Referenz dar, was messtechnisch in geringeren individuellen Abweichungen, im Vergleich zur Freifeldentzerrung, resultiert<sup>[98,129]</sup>.

Zusammengefasst lässt sich festhalten: Als Idealfall gilt eine Wiedergabe individuell aufgenommener Ohrsignale über einen ebenfalls individuell kalibrierten Kopfhörer. Aufnahme und Kalibrierung sollten dabei nach Möglichkeit direkt mit derselben Messanordnung durchgeführt werden. Wird nicht mit individuellen Ohrsignalen gearbeitet – die Aufnahmen eines Kopfes werden also mehreren Hörern vorgespielt – empfiehlt es sich zumindest, den Kopfhörer am Aufnahmekopf zu kalibrieren. Eine größere Flexibilität erreicht man jedoch durch eine entkoppelte Kalibrierung, indem man Aufnahmekopf und Wiedergabekopfhörer auf ein ideales Diffusfeld bezieht.



**Abbildung 2.2** Funktionsprinzip der Binauralen Raumsynthese (BRS). In einer separaten Aufnahme-prozedur wird die Referenzsituation, links dargestellt, mit einem drehbaren Kunstkopf vermessen. Dabei werden die binauralen Raumimpulsantworten (BRIR) der Lautsprecher am Abhörpunkt für unterschiedliche Kopfausrichtungen ermittelt und in einer Datenbank abgelegt. Der BRS-Prozessor, auf der Wiedergabeseite rechts, liest diese Daten abhängig vom aktuellen Drehwinkel aus und liefert ein binaurales Signal  $\mathbf{b} = (b_l[t], b_r[t])^T$ , das dem Benutzer über Kopfhörer dargeboten wird. Mit dem BRS-Verfahren wird die Wiedergabe eines beliebigen Eingangssignals über die Lautsprecher der Referenzsituation realitätsnah simuliert, es liegt sozusagen eine „akustische Kopie“ dieser Referenzsituation vor.

## 2.3 Binaurale Raumsynthese

Mit der Binauralen Raumsynthese (kurz: BRS) stand für Versuche im Rahmen dieser Arbeit ein leistungsfähiges binauraltechnisches System zur Verfügung. Es ist in der Lage, mehrere Schallquellen zu simulieren und dabei Kopfdrehungen des Hörers während der Wiedergabe dynamisch zu kompensieren. Dadurch besteht eine, wenn auch beschränkte, Bewegungsmöglichkeit in der auditiven virtuellen Umgebung. Das Verfahren ist speziell für den Einsatz in der Rundfunkproduktion ausgelegt. Es wurde in psychoakustischen Experimenten umfassend verifiziert<sup>[99,114]</sup>, in Feldstudien mit Endanwendern getestet<sup>[114]</sup> und bis zur Serienreife<sup>[99]</sup> weiterentwickelt.

Die Referenzsituation, beispielsweise eine im Regieraum eines Tonstudios installierte Mehrkanal-Lautsprecheranlage, muss zuerst mit einem Kunstkopf vermessen werden. Dazu werden mit Hilfe eines Drehtellers die binauralen Raumimpulsantworten jedes Lautsprechers automatisiert für alle relevanten Kopfausrichtungen – bis hin zu einer

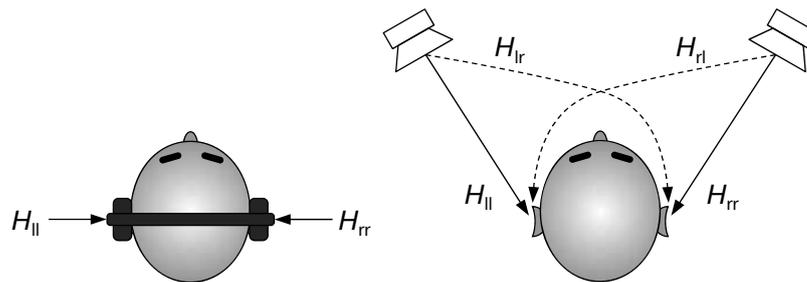
vollen Umdrehung – mit einem regelmäßigen Winkelabstand von  $6^\circ$  erfasst. Die Messergebnisse werden für die spätere Verwendung so interpoliert, dass für den berücksichtigten Winkelbereich Daten mit einer Rasterung von  $1^\circ$  vorliegen. Diese Auflösung entspricht in etwa der geringsten Lokalisationsunschärfe beim Richtungs-hören in der Horizontalebene<sup>[23,S.30 ff]</sup>. Damit hat man, anschaulich formuliert, eine akustische Kopie der Referenzsituation erstellt (vgl. Abb. 2.2 links).

Die Faltung von Eingangssignal und binauraler Raumimpulsantwort liefert ein binaurales Signal, das die Wiedergabe des Eingangssignals über den zugehörigen Lautsprecher der vorher vermessenen Referenzsituation realitätsnah simuliert. Da Linearität vorausgesetzt ist, können bei der Signalverarbeitung mehrere Kanäle additiv überlagert werden<sup>[145,S.334]</sup>. Abhängig vom aktuellen Drehwinkel des Kopfes wechselt der BRS-Prozessor während der Wiedergabe den Filtersatz mit binauralen Raumimpulsantworten (vgl. Abb. 2.2 rechts). Beim Umschalten der Filter treten wegen der feinen Winkelauflösung von  $1^\circ$  keine hörbaren Artefakte auf. Die Gesamtlatenz des benutzten Laborsystems liegt deutlich unterhalb der psychoakustisch ermittelten Wahrnehmbarkeitsschwelle von  $85\text{ ms}$ <sup>[55,114]</sup>.

Für eine verzerrungsfreie Übertragung zwischen Aufnahme und Wiedergabe sind beide Seiten entkoppelt mit Bezug auf ein ideales Diffusfeld zu kalibrieren. Solange dies gewährleistet ist, gilt die Systemkette als sehr robust gegen das Austauschen von Komponenten. Für die Aufnahme wird hauptsächlich das Kunstkopfmodell KU 100, hergestellt von der Georg Neumann GmbH, verwendet, weil es bereits ab Werk diffusfeldentzerrt ist. Beim Kopfhörer greift man bevorzugt auf hochwertige Modelle zurück, die auch in der Rundfunkproduktion häufig eingesetzt werden. Hier sollte jedoch die Kalibrierung immer überprüft werden<sup>[147,164]</sup>.

Das Verfahren zeichnet sich aus durch eine korrekte Richtungsabbildung in der Horizontalebene und eine weitgehende Klangfarbentreue für breitbandige Signale, beide Aspekte sind vergleichbar mit dem natürlichen Hören in der Referenzsituation selbst. Dies gilt sowohl für die Lautsprecher als auch für die mit ihnen erzeugten Phantomschallquellen. Die Abbildung der Elevation ist etwas ungenauer, simulierte Schallquellen werden im Mittel etwas zu hoch wahrgenommen. Die Berücksichtigung von Nickbewegungen erzielt aber keine entscheidende Verbesserung<sup>[113]</sup>.

Ein ähnliches Verfahren wird an der TU Berlin für Vergleichsstudien eingesetzt. Dort ist der Kunstkopf über ein steuerbares Gelenk mit einem Torso verbunden, wodurch sowohl Kopfdrehen als auch Kopfnicken kompensiert werden kann<sup>[106]</sup>. Mittlerweile wurde auch ein System mit individueller Anpassung an den Hörer vorgestellt. Mit Miniaturmikrofonen und einer vereinfachten Messprozedur mit drei Stützstellen werden personalisierte binaurale Raumimpulsantworten ermittelt. Wiedergabeseitig wird mit dem gleichen Messaufbau auch der Kopfhörer individuell entzerrt<sup>[159]</sup>.



**Abbildung 2.3** Systemtheoretischer Unterschied zwischen Kopfhörerwiedergabe (links) und Lautsprecherwiedergabe (rechts). Der von einem Lautsprecher abgegebene Schall erreicht jeweils das ipsilaterale, d. h. zugewandte, wie auch das kontralaterale, d. h. abgewandte Ohr. Die beiden Übertragungsstrecken zu den kontralateralen Ohren  $H_{lr}$ ,  $H_{rl}$  – hier gestrichelt dargestellt – werden als Übersprechen bezeichnet. Bei Kopfhörerwiedergabe hingegen tritt kein oder nur vernachlässigbar geringes Übersprechen auf, so dass jedes Ohr separat beschallt wird. Die zwei Übertragungsfunktionen eines Lautsprechers auf beide Ohren zusammen entsprechen im Zeitbereich der binauralen Raumimpulsantwort des Lautsprechers.

## 2.4 Wiedergabe über Lautsprecher

Die Herausforderung einer Wiedergabe binauraler Signale über Lautsprecher besteht darin, den Schalldruck in den beiden Ohren des Hörers korrekt zu kontrollieren. Der von einem Lautsprecher abgegebene Schall erreicht beide Ohren, die – sofern sich der Lautsprecher nicht in der Medianebene befindet – auch „ipsilaterales“, das heißt zugewandtes, und „kontralaterales“, das heißt abgewandtes, Ohr genannt werden. Spielt man die beiden Ohrsignale einer binauralen Aufnahme direkt über ein Lautsprecherpaar ohne zusätzliche Maßnahmen ab, so werden dadurch auf jedes Ohr unerwünschte Signalanteile, die eigentlich nur für das jeweils andere Ohr gedacht sind, übertragen. Diesen Effekt bezeichnet man als *Übersprechen*, das möglichst stark zu dämpfen ist, weil es das binaurale Signal beeinträchtigt. Zusätzlich müssen die beiden *Direktwege* zum jeweils ipsilateralen Ohr  $H_{ll}$  und  $H_{rr}$  – ähnlich der Kalibrierung bei Kopfhörerwiedergabe – entzerrt werden.

Man war – abgesehen von der exakten Lösung des Problems – einige Zeit lang bestrebt, eine akzeptable Wiedergabe von Kunstkopfsignalen direkt über Lautsprecher zu ermöglichen. Unbearbeitet können dabei deutliche Klangverfärbungen auftreten, die durch eine geeignete Verzerrung der binauralen Aufnahme minimiert werden sollen. Für eine korrekte Kopfhörerwiedergabe muss diese Verzerrung dann jedoch durch eine entsprechende Kopfhörerentzerrung wieder ausgeglichen werden. Die kontrovers diskutierten Vorschläge umfassen im Wesentlichen das Nivellieren des Kunstkopffrequenzgangs für eine oder zwei diskrete Schalleinfallrichtungen oder

für einen überlagerten Schalleinfall aus einem größeren Bereich, mit den Extremen Freifeldentzerrung einerseits und Diffusfeldentzerrung andererseits<sup>[24,S.90 ff.][63]</sup>.

Ein solcher Ansatz erfüllt allerdings nur das Minimalziel, ohne großen Zusatzaufwand eine weitgehend klangfarbentreue Lautsprecherwiedergabe von binauralen Aufnahmen zu realisieren. Eine korrekte, zur Kopfhörerwiedergabe vergleichbare Richtungsabbildung ist nur mit der exakten Wiedergabe der Ohrsignale erreichbar, für die sowohl eine Kompensation des Übersprechens als auch eine Entzerrung der Direktpfade erfolgen muss.

In Matrix-Vektor-Notation lässt sich das Problem sehr einfach formulieren. Für eine Lautsprecherwiedergabe wie in Abb. 2.3 rechts skizziert gilt im Frequenzbereich für die Spektren der Ohrsignale  $P_l$  und  $P_r$  folgende Gleichung:

$$\begin{bmatrix} H_{ll} & H_{rl} \\ H_{lr} & H_{rr} \end{bmatrix} \cdot \begin{pmatrix} B_l \\ B_r \end{pmatrix} = \begin{pmatrix} P_l \\ P_r \end{pmatrix} \quad (2.4)$$

oder kompakt notiert

$$\mathbf{H} \cdot \mathbf{b} = \mathbf{p} \quad (2.5)$$

$B_l$  und  $B_r$  stehen für den linken beziehungsweise rechten Kanal der binauralen Aufnahme. Sie werden zusammengefasst zum Vektor  $\mathbf{b} = (B_l, B_r)^T$ , die Ohrsignale entsprechend zum Vektor  $\mathbf{p} = (P_l, P_r)^T$ . Die Matrix  $\mathbf{H}$  enthält spaltenweise die in den Frequenzbereich transformierten binauralen Raumimpulsantworten der Lautsprecher. Ausmultipliziert für das linke Ohrsignal ergibt sich

$$\underbrace{H_{ll} \cdot B_l}_{\text{Direktpfad}} + \underbrace{H_{rl} \cdot B_r}_{\text{Übersprechen}} = P_l \quad (2.6)$$

und für rechts

$$\underbrace{H_{lr} \cdot B_l}_{\text{Übersprechen}} + \underbrace{H_{rr} \cdot B_r}_{\text{Direktpfad}} = P_r \quad (2.7)$$

Gesucht wird nun ein passendes, vor die Lautsprecher geschaltetes Netzwerk  $\mathbf{X}$ , das die beiden Direktpfade entzerrt und das Übersprechen unterdrückt:

$$\mathbf{H} \cdot \mathbf{X} \cdot \mathbf{b} = \mathbf{p} \quad (2.8)$$

Eine korrekte Wiedergabe der binauralen Aufnahme bedeutet, dass die resultierenden Ohrsignale  $\mathbf{p}$  identisch sein sollen mit den aufgenommenen Ohrsignalen  $\mathbf{b}$

$$\mathbf{b} \stackrel{!}{=} \mathbf{p} \quad (2.9)$$

woraus durch Kombinieren der Gln. (2.8) und (2.9) direkt folgt, dass  $\mathbf{H} \cdot \mathbf{X}$  gleich der Einheitsmatrix  $\mathbf{1}$  sein muss

$$\mathbf{H} \cdot \mathbf{X} = \mathbf{1} \quad (2.10)$$

und  $\mathbf{X}$  somit die Inverse von  $\mathbf{H}$  ist<sup>[33,S.266]</sup>

$$\mathbf{X} = \mathbf{H}^{-1} \quad (2.11)$$

Die Konstellation mit zwei Lautsprechern und einem Hörer mit zwei Ohren stellt den einfachsten Fall einer Wiedergabe von binauralen Signalen über Lautsprecher dar. Die Matrix  $\mathbf{H}$  ist dann quadratisch von der Ordnung 2, vgl. Gl. (2.4), und im Allgemeinen komplex. Sofern sie regulär ist, kann sie auch invertiert werden. Dazu muss ihre Determinante von Null verschieden sein. Das gesuchte Netzwerk  $\mathbf{X}$  ist beschrieben durch<sup>[33,S.266 & 269–271]</sup>

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} \frac{H_{rr}}{\det \mathbf{H}} & \frac{-H_{rl}}{\det \mathbf{H}} \\ \frac{-H_{lr}}{\det \mathbf{H}} & \frac{H_{ll}}{\det \mathbf{H}} \end{bmatrix} \quad (2.12)$$

mit<sup>[33,S.266]</sup>

$$\det \mathbf{H} \neq 0 \quad (2.13)$$

und<sup>[33,S.271]</sup>

$$\det \mathbf{H} = H_{ll} \cdot H_{rr} - H_{rl} \cdot H_{lr} \quad (2.14)$$

Die Signale, die über die Lautsprecher wiedergegeben werden müssen,  $\mathbf{w} = (W_l, W_r)^T$ , berechnen sich dann aus dem Netzwerk  $\mathbf{X}$  und der binauralen Aufnahme  $\mathbf{b}$

$$\mathbf{X} \cdot \mathbf{b} = \mathbf{w} \quad (2.15)$$

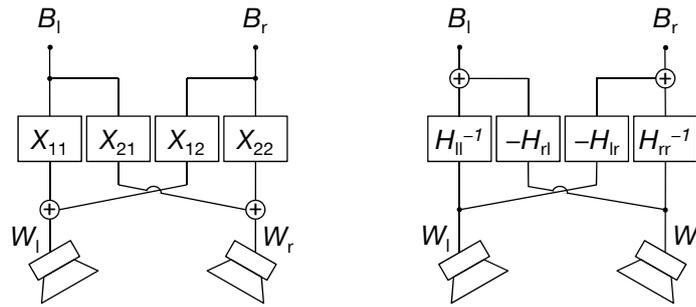
und als Gleichungssystem ausgeschrieben vgl. <sup>[101]</sup>

$$\frac{H_{rr}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_l + \frac{-H_{rl}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_r = W_l \quad (2.16)$$

$$\frac{-H_{lr}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_l + \frac{H_{ll}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_r = W_r \quad (2.17)$$

Die Gln. (2.16) und (2.17) lassen sich direkt in ein Blockdiagramm überführen, das in Abb. 2.4 im linken Teil dargestellt ist, wobei der allen vier Filtern gemeinsame Term  $\frac{1}{\det \mathbf{H}}$  auch in beiden Zweigen entweder vorgelagert oder nachgelagert eingefügt werden kann.

Im rechten Teil von Abb. 2.4 ist eine rekursive Filterstruktur skizziert, die den Ablauf des Korrekturprozesses auf intuitive Weise repräsentiert: Die Entzerrung der beiden Direktpfade erfolgt durch die Invertierung der ipsilateralen Übertragungsfunktionen, also durch  $H_{ll}^{-1}$  und  $H_{rr}^{-1}$ . Die unerwünschte kontralaterale Übertragung wird an den Ohren ausgelöscht, indem das Übersprechen über die Filter  $-H_{rl}$  und  $-H_{lr}$  auf dem jeweils anderen Lautsprecher negativiert wiedergegeben wird. Die Rekursion



**Abbildung 2.4** Filterstrukturen zur Vorverarbeitung eines binauralen Quellsignals  $\mathbf{b} = (B_l, B_r)^T$  für eine korrekte Reproduktion über Lautsprecher. Die Darstellung links folgt direkt aus der mathematischen Beschreibung und Lösung des Problems in Matrix-Vektor-Notation, rechts ist eine alternative rekursive Struktur aufgezeigt.

verdeutlicht, dass jedes Kompensationssignal seinerseits wieder auf das andere Ohr überspricht und in der Folge selbst kompensiert werden muss<sup>[57, S. 56] [102, S. 46 f.] [153]</sup>.

Beide Topologien sind mathematisch äquivalent und lassen sich wie folgt ineinander überführen. Die Systemgleichung für die rekursive Filterstruktur aus Abb. 2.4 rechts lautet ausmultipliziert

$$\frac{-H_{rl}}{H_{ll}} \cdot W_r + \frac{1}{H_{ll}} \cdot B_l = W_l \quad (2.18)$$

$$\frac{-H_{lr}}{H_{rr}} \cdot W_l + \frac{1}{H_{rr}} \cdot B_r = W_r \quad (2.19)$$

Setzt man nun Gl. (2.19) in Gl. (2.18) ein und isoliert  $W_l$  auf einer Seite, erhält man

$$W_l = \frac{H_{rr}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_l + \frac{-H_{rl}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_r \quad (2.20)$$

Gl. (2.20) in Gl. (2.19) liefert schließlich

$$W_r = \frac{-H_{lr}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_l + \frac{H_{ll}}{H_{ll}H_{rr} - H_{rl}H_{lr}} \cdot B_r \quad (2.21)$$

Da die Gln. (2.20) und (2.16) sowie (2.21) und (2.17) jeweils identisch sind, lösen beide Filterstrukturen das Problem, Ohrsignale über ein Lautsprecherpaar korrekt und vergleichbar mit einer Kopfhörerwiedergabe zu reproduzieren.

Für eine Stabilitätsbetrachtung kann man die geschlossene Lösung durch die in der Matrix  $\mathbf{X}$  zusammengefassten vier Filter in eine iterative Lösung umrechnen<sup>[101,102]</sup>.

Aus  $X_{11} = \frac{H_{rr}}{\det \mathbf{H}} = \frac{H_{rr}}{H_{ll}H_{rr} - H_{rl}H_{lr}}$  wird durch Erweitern mit  $\frac{1}{H_{ll}H_{rr}}$  [102, S. 49 f.]

$$X_{11} = \frac{\frac{H_{rr}}{H_{ll}H_{rr}}}{\frac{H_{ll}H_{rr}}{H_{ll}H_{rr}} - \frac{H_{rl}H_{lr}}{H_{ll}H_{rr}}} = \frac{1}{H_{ll}} \cdot \frac{1}{1 - \frac{H_{rl}H_{lr}}{H_{ll}H_{rr}}} \quad (2.22)$$

Gl. (2.22) stellt eine geometrische Reihe dar und lässt sich mit

$$x = \frac{H_{rl}H_{lr}}{H_{ll}H_{rr}} \quad (2.23)$$

umformen, sofern  $|x| < 1$  ist, zu [145, S. 183 & 188]

$$X_{11} = \frac{1}{H_{ll}} \cdot \sum_{k=0}^{\infty} x^k = \frac{1}{H_{ll}} (1 + x + x^2 + x^3 + \dots) \quad (2.24)$$

Ähnlich ergeben sich die anderen drei Filter zu

$$X_{12} = \frac{-H_{rl}}{H_{ll}H_{rr}} (1 + x + x^2 + x^3 + \dots) = -\frac{H_{rl}}{H_{rr}} X_{11} \quad (2.25)$$

$$X_{21} = \frac{-H_{lr}}{H_{ll}H_{rr}} (1 + x + x^2 + x^3 + \dots) = -\frac{H_{lr}}{H_{ll}} X_{22} \quad (2.26)$$

$$X_{22} = \frac{1}{H_{rr}} (1 + x + x^2 + x^3 + \dots) \quad (2.27)$$

Die Konvergenz der geometrischen Reihe und damit die Stabilität der vier Filter hängt von  $x$ , also im Wesentlichen vom Verhältnis des Übersprechens zu den Direktpfaden, ab [101] [102, S. 60]. Bei Bedarf lässt sich die geschlossene Lösung aus der iterativen durch einen Abbruch der Reihenentwicklung, beispielsweise nach der 2. Ordnung [102, S. 50], annähern.

Aus den Gln. (2.24) bis (2.27) leitet sich direkt auch eine weitere vorgeschlagene Filterstruktur ab [57, 127]. Dazu definiert man die interauralen Übertragungsfunktionen<sup>12</sup>  $ITF_l$  und  $ITF_r$  [57, S. 56]

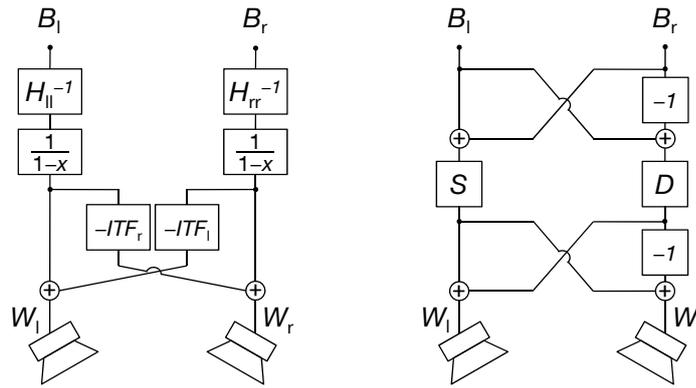
$$ITF_l = \frac{H_{lr}}{H_{ll}} \quad (2.28)$$

$$ITF_r = \frac{H_{rl}}{H_{rr}} \quad (2.29)$$

und erhält so

$$x = ITF_l \cdot ITF_r \quad (2.30)$$

<sup>12</sup>engl.: interaural transfer function (kurz: ITF)



**Abbildung 2.5** Weitere Filterstrukturen zur Vorverarbeitung eines binauralen Quellsignals  $\mathbf{b} = (B_l, B_r)^T$  für eine korrekte Reproduktion über Lautsprecher. Die Darstellung links ergibt sich aus der Umformung der geschlossenen in eine iterative Lösung. Die rechte Struktur beruht auf der vereinfachenden Annahme eines symmetrischen Problems, das heißt die beiden Übertragungstrecken zu den ipsilateralen Ohren sowie die beiden Übertragungstrecken zu den kontralateralen Ohren gelten als jeweils identisch. Sie wird als „Shuffler“-Struktur bezeichnet.

und die nun mit  $\mathbf{Y}$  bezeichneten Filter

$$Y_{11} = \frac{1}{H_{ll}} \cdot \frac{1}{1 - ITF_l \cdot ITF_r} \quad (2.31)$$

$$Y_{12} = -ITF_r \cdot Y_{11} \quad (2.32)$$

$$Y_{21} = -ITF_l \cdot Y_{22} \quad (2.33)$$

$$Y_{22} = \frac{1}{H_{rr}} \cdot \frac{1}{1 - ITF_l \cdot ITF_r} \quad (2.34)$$

oder in Matrix-Vektor-Notation<sup>[57, S. 56]</sup>

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{H_{ll}} & 0 \\ 0 & \frac{1}{H_{rr}} \end{bmatrix} \begin{bmatrix} 1 & -ITF_r \\ -ITF_l & 1 \end{bmatrix} \frac{1}{1 - ITF_l \cdot ITF_r} \quad (2.35)$$

Die zugehörige Filterstruktur ist in Abb. 2.5 auf der linken Seite dargestellt.

### Symmetrisches Problem

Die Wiedergabe von binauralen Signalen über Lautsprecher vereinfacht man schon von Anfang an oft zu einem symmetrischen Problem, indem man für die Lautsprecheranstellung das Stereoformat 2/0<sup>[80]</sup> zugrunde legt. Dann befinden sich die beiden

Lautsprecher an zur Medianebene gespiegelten kopfbezogenen Koordinaten und man kann – allerdings nur näherungsweise (vgl. Abschn. 2.1) – davon ausgehen, dass die Übertragungstrecken zu den ipsilateralen und kontralateralen Ohren jeweils paarweise identisch sind. Das heißt man nimmt an, dass

$$H_i = H_{ll} = H_{rr} \quad \text{Hinweis: Index } i \text{ steht für „ipsilateral“} \quad (2.36)$$

$$H_k = H_{lr} = H_{rl} \quad \text{Hinweis: Index } k \text{ steht für „kontralateral“} \quad (2.37)$$

und damit

$$ITF = ITF_l = ITF_r \quad (2.38)$$

gilt. Außerdem folgt daraus

$$\det \mathbf{H} = H_i^2 - H_k^2 \quad (2.39)$$

Damit können alle bisher aufgestellten Gleichungen entsprechend vereinfacht werden und die Blockdiagramme aus Abb. 2.4 sowie aus Abb. 2.5 links enthalten dann jeweils paarweise identische Filter.

Aus Gl. (2.12) wird so

$$\mathbf{X} = \frac{1}{\det \mathbf{H}} \begin{bmatrix} H_i & -H_k \\ -H_k & H_i \end{bmatrix} \quad (2.40)$$

Die symmetrische Matrix  $\begin{bmatrix} H_i & -H_k \\ -H_k & H_i \end{bmatrix}$  lässt sich mit ihren Eigenwerten

$$\lambda_1 = (H_i + H_k) \quad \text{und} \quad \lambda_2 = (H_i - H_k)$$

sowie den normierten Eigenvektoren

$$\mathbf{g}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad \text{und} \quad \mathbf{g}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

zerlegen<sup>[145,S.111–114]</sup>, so dass sich Gl. (2.40) insgesamt zu

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (2.41)$$

mit

$$S = \frac{1}{2} \cdot \frac{1}{H_i + H_k} \quad \text{Hinweis: } S \text{ steht für „Summe“} \quad (2.42)$$

$$D = \frac{1}{2} \cdot \frac{1}{H_i - H_k} \quad \text{Hinweis: } D \text{ steht für „Differenz“} \quad (2.43)$$

umformen lässt<sup>[11][57,S.63–64]</sup>. Aus den Gln. (2.41) bis (2.43) lässt sich eine neue Filterstruktur ableiten, die in Abb. 2.5 auf der rechten Seite dargestellt ist<sup>13</sup>. Dafür hat sich die Bezeichnung „Shuffler-Struktur“ eingebürgert<sup>[11,41,57,82,92,183]</sup>.

<sup>13</sup>leicht abweichende Strukturen resultieren aus alternativ festgelegten Eigenvektoren z. B. [41]

Eine mögliche – allerdings nur hypothetische und nicht überprüfte – Lösung des symmetrischen Problems wurde 1961 noch in Form eines analogen Schaltkreises publiziert<sup>[13]</sup>. Bereits zwei Jahre später taucht jedoch ein Ansatz auf, der auf digitaler Signalverarbeitung basiert<sup>[153]</sup> und als wesentlicher Eckpfeiler aller Überlegungen auf diesem Feld allgemein anerkannt ist. Der ihm zugeschriebene Begriff *Übersprechkompensation*<sup>14</sup> scheint jedoch erst um 1970 geprägt worden zu sein<sup>15</sup><sup>[48,154]</sup>. Damals wurde ein Verfahren zur akustischen Raumsimulation entwickelt, bei dem eines der Teilsysteme in der Lage ist, über ein bei kopfbezogenen Seitenwinkeln von  $\pm 22,5^\circ$  aufgetelltes Lautsprecherpaar verschiedene diskrete, aus der vorderen Halbebene einfallende Echos zu erzeugen. Die Implementierung entspricht dem Blockdiagramm aus Abb. 2.4 links mit symmetrischer Struktur kombiniert mit einer vorgelagerten statischen binauralen Synthese. Der Term  $\frac{1}{\det \mathbf{H}}$  ist dabei als separates Filter direkt am Eingang des Teilsystems realisiert<sup>[153,154]</sup>.

Eine praxisorientierte Lösung zur Wiedergabe von Kunstkopfaufnahmen wurde um 1970 beschrieben: Für das in Deutschland entwickelte TRADIS-Verfahren<sup>16</sup> werden aus den beiden Kanälen der Originalaufnahme vier neue Signale erzeugt, indem man sie einzeln und teilweise negativiert über ein Lautsprecherpaar wiedergibt und mit einem Kunstkopf an der späteren Hörerposition erneut aufnimmt. Eingeordnet in die bisher verwendete Systematik gilt für die Spektren dieser vier neu aufgenommenen Teilsignale

- $B_r \cdot H_{ll} \cdot A^{-1}$  („Original“-Rechts)
- $-B_r \cdot H_{rl} \cdot A^{-1}$  (Kompensations-Links)
- $B_l \cdot H_{rr} \cdot A^{-1}$  („Original“-Links)
- $-B_l \cdot H_{lr} \cdot A^{-1}$  (Kompensations-Rechts)

wobei  $A^{-1}$  ein beim Überspielen der Signale in den Übertragungsweg geschaltetes Filter darstellt, dessen Übertragungsfunktion in etwa  $\frac{1}{\det \mathbf{H}}$  entspricht. Sie werden in der Kombination „Original“-Links plus Kompensations-Links vom linken und „Original“-Rechts plus Kompensations-Rechts vom rechten Lautsprecher abgegeben, so dass auch dieses Vorgehen durch das linke Blockdiagramm aus Abb. 2.4 – genau genommen sogar ohne Beschränkung auf ein symmetrisches Problem – beschrieben ist<sup>[46]</sup>.

Will man den Aufwand für das Erstellen dieser Signale oder das Bestimmen der Übertragungsfunktionen vermeiden, kann man ein passendes Filter auch psychoakustisch ermitteln. Dazu lässt man beide Lautsprecher ein terzbreites Rauschen abstrahlen, wo-

<sup>14</sup>engl.: cross talk cancellation (kurz: CTC, XTC)

<sup>15</sup>Originalzitat: „Similarly, the ‚cross talk‘ from the left loudspeaker to the right ear has to be **cancelled** by appropriate pulses emitted from the right loudspeaker.“<sup>[154, S. 428]</sup>

<sup>16</sup>engl.: true reproduction of all directional information by stereophony (kurz: TRADIS)

bei ein Kanal invertiert ist. Für diesen Kanal muss eine Versuchsperson Verstärkung und Verzögerung so einregeln, dass sie das Rauschen möglichst lateral, und zwar auf der Seite des unbearbeiteten Kanals, hört. Diese Prozedur wird für alle Terzbänder des interessierenden Frequenzbereichs wiederholt. Aus den gewonnenen Daten lässt sich dann ein sogenanntes  $90^\circ$ -Filter ableiten, das vor die Lautsprecher geschaltet wird. Eine akzeptable Richtungsabbildung von binauralen Eingangssignalen ist für dieses Verfahren nachgewiesen<sup>[47,48]</sup>. Auch in einer deutlich jüngeren Konzeptstudie wird eine ähnliche Idee mit benutzerkalibrierten Filtern beschrieben<sup>[81]</sup>.

### Mehr als zwei Lautsprecher

Die Lösung für ein Stereolautsprecherpaar ist zwar theoretisch exakt, toleriert in der Praxis aber nur geringfügige Abweichungen vom optimalen Abhörpunkt. Der Spielraum beschränkt sich bei Kopfdrehungen auf etwa  $\pm 10^\circ$ <sup>[153]</sup> und bei seitlichen Bewegungen auf wenige Zentimeter<sup>[92]</sup>. Nur für symmetrische Anordnungen ist das Verfahren bei einer Translation des Hörers entlang der Mittelsenkrechten der Lautsprecherbasis, also entlang der Symmetrieachse, etwas robuster und erlaubt laut subjektiven Urteilen Abweichungen bis zu einem Meter nach vorne und hinten<sup>[92,126]</sup>. Daher wurde schon als Erweiterung für das TRADIS-Verfahren vorgeschlagen, die Kompensationssignale durch zwei zusätzliche hinter dem Hörer aufgestellte Lautsprecher abzuspielen, um so eine vergrößerte Abhörzone zu erhalten<sup>[96]</sup>.

Zudem haben die Hörer Schwierigkeiten, Schallereignisse wahrzunehmen, die sich hinter ihnen befinden würden<sup>[75]</sup>. Die Ursache ist vermutlich der ausschließlich von vorne aus den Lautsprechern einfallende Schall, wodurch Schallquellen hinter dem Kopf nicht plausibel sind. Eine Anordnung mit vier Lautsprechern, von denen zwei vor und zwei hinter dem Hörer aufgestellt sind, ist im direkten Vergleich mit einem Stereolautsprecherpaar wesentlich erfolgreicher, hinter dem Hörer liegende Quellen zu simulieren<sup>[74]</sup>. Es liegt also nahe, mehr als zwei Lautsprecher einzusetzen und rund um den Abhörplatz zu verteilen.

Wenn statt zwei nun  $n$  Lautsprecher berücksichtigt werden sollen, muss die Matrix  $\mathbf{H}$  pro Lautsprecher um eine neue Spalte vom Typ  $2 \times 2$  zum Typ  $2 \times n$  erweitert werden. Gl. (2.10), also  $\mathbf{H}\mathbf{X} = \mathbf{1}$ , gilt nach wie vor. Sie ist allerdings nicht mehr so leicht zu lösen, da  $\mathbf{H}$  jetzt nicht mehr regulär ist. Es ist aber möglich, das Problem aufzuteilen und  $\mathbf{X}$  spaltenweise zu berechnen<sup>[11]</sup>. Aus Gl. (2.10) wird dann

$$\mathbf{H} \cdot \mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (2.44)$$

und

$$\mathbf{H} \cdot \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (2.45)$$

wobei  $\mathbf{x}_1$  und  $\mathbf{x}_2$  Spaltenvektoren der Dimension  $n$  sind. Das bedeutet, man ordnet nun für die beiden binauralen Kanäle getrennt jedem der  $n$  Lautsprecher ein eigenes Filter zu. Durch den Filtersatz  $\mathbf{x}_1$  soll der linke binaurale Kanal unverfälscht auf das linke Ohr und überhaupt nicht auf das rechte Ohr übertragen werden, entsprechend umgekehrt gilt das ebenso für  $\mathbf{x}_2$  und den rechten binauralen Kanal. Für  $n > 2$  sind die beiden Gleichungssysteme unterbestimmt, das heißt sie enthalten mehr Unbekannte als Gleichungen, und sind daher nicht eindeutig lösbar. Mit Hilfe der Pseudoinversen  $\mathbf{H}^+$  erhält man die beiden eindeutig bestimmten Lösungsvektoren  $\mathbf{x}_1, \mathbf{x}_2$  mit minimaler euklidischer Norm<sup>[11]</sup>

$$\mathbf{x}_1 = \mathbf{H}^+ \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (2.46)$$

und

$$\mathbf{x}_2 = \mathbf{H}^+ \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (2.47)$$

In diesem Fall ist  $\mathbf{x}_1$  als erste und  $\mathbf{x}_2$  als zweite Spalte aus der Pseudoinversen  $\mathbf{H}^+$  direkt ablesbar. Man kann sie entweder mit Hilfe ihrer Adjungierten  $\mathbf{H}^H$  – das ist die konjugiert komplexe Transponierte  $(\mathbf{H}^*)^T$  [33,S.261] – gemäß der Formel<sup>[11,75]</sup>

$$\mathbf{H}^+ = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \quad (2.48)$$

berechnen, da hier  $\mathbf{H}\mathbf{H}^H$  regulär ist, oder allgemein durch Singulärwertzerlegung<sup>[11]</sup> ermitteln. Zusammengefasst gilt also für  $n$  statt zwei Lautsprecher  $\mathbf{X} = \mathbf{H}^+$  statt  $\mathbf{X} = \mathbf{H}^{-1}$ , wobei die Pseudoinverse  $\mathbf{H}^+$  vom Typ  $n \times 2$  ist. Für  $n = 2$  sind, sofern  $\mathbf{H}$  regulär ist,  $\mathbf{H}^+$  und  $\mathbf{H}^{-1}$  identisch<sup>[11]</sup>. Um die Verstärkung zu limitieren, erweitert man Gl. (2.48) in der Praxis oft mit einen frequenzabhängigen Regularisierungsparameter  $\beta \geq 0$  zu<sup>[115,119]</sup>

$$\mathbf{H}^+ = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H + \beta \mathbf{1})^{-1} \quad (2.49)$$

An jedem Lautsprecher liegt also die Summe der beiden jeweils individuell gefilterten binauralen Kanäle an. Bei einem symmetrischen Problem lassen sich die Filterstrukturen auch hier durch eine geeignete Zerlegung der Pseudoinversen noch vereinfachen<sup>[11,183,225]</sup>.

## Mehr als ein Hörer

Theoretisch lässt sich die Problemstellung auch auf mehrere Hörer erweitern. In der Praxis kann man sich diesen Zusatzaufwand oft ersparen, da das Übersprechen zwischen den Hörern bei ausreichend Abstand voneinander vernachlässigbar wird. Es existiert daher wenig einschlägige Literatur<sup>z. B. [11,86,105,117,120,121]</sup>. Sind die Distanzen zwischen den Hörern und auch zu den Lautsprechern aber verhältnismäßig kurz, kann eine Berücksichtigung aller Übertragungstrecken nützlich sein.

Für jeden zusätzlichen Hörer muss Gl. (2.8) um zwei Zeilen erweitert werden. Zudem muss festgelegt werden, wie die beiden binauralen Kanäle auf die jetzt  $h$  Hörer und die damit vorhandenen  $m = 2h$  Ohren übertragen werden sollen. Das erreicht man durch Einführung einer Übertragungsmatrix  $\mathbf{Z}$  vom Typ  $m \times 2$  in Gl. (2.9):

$$\mathbf{Z} \cdot \mathbf{b} \stackrel{!}{=} \mathbf{p} \quad (2.50)$$

Nimmt man an, dass alle ungeraden Zeilen für die linken Ohren und alle geraden Zeilen für die rechten Ohren gelten sollen, hat  $\mathbf{Z}$  die folgende Form

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (2.51)$$

Statt Gl. (2.10), also  $\mathbf{H}\mathbf{X} = \mathbf{1}$ , muss demnach

$$\mathbf{H}\mathbf{X} = \mathbf{Z} \quad (2.52)$$

zutreffen. Auch hier behilft man sich mit der Pseudoinversen, um das Problem zu lösen, und erhält für den allgemeinen Fall von  $n$  Lautsprechern und  $h$  Hörern, also  $m = 2h$  Ohren,

$$\mathbf{X} = \mathbf{H}^+ \mathbf{Z} \quad (2.53)$$

### Zusammenfassung

Gl. (2.53) ist die allgemeine Lösung für das Problem, zwei Eingangssignale so über  $n$  Lautsprecher wiederzugeben, dass die bei der Wiedergabe resultierenden Ohrsignale aller Hörer jeweils mit den zwei Eingangssignalen identisch sind. Wichtigster Anwendungsfall ist die korrekte – das heißt entzerrte und von Übersprechen freie – Wiedergabe binauraler Signale über Lautsprecher. Dabei ist die Matrix  $\mathbf{H}^+$  die Pseudoinverse der Matrix  $\mathbf{H}$ , die für jeden der  $h$  Hörer die in den Frequenzbereich transformierten binauralen Raumimpulsantworten aller  $n$  Lautsprecher enthält.  $\mathbf{H}$  ist vom Typ  $2h \times n$  und  $\mathbf{H}^+$  damit vom Typ  $n \times 2h$  und im Allgemeinen komplex.  $\mathbf{Z}$  enthält nur die Elemente 1 und 0 und ist vom Typ  $n \times 2$ , vgl. Gl. (2.51), so dass auch  $\mathbf{X}$  vom Typ  $n \times 2$  ist.  $\mathbf{X}$  spezifiziert für jeden der beiden Eingangskanäle pro Lautsprecher eine eigene Übertragungsfunktion. Für einen Hörer, also  $h = 1$ , wird  $\mathbf{Z}$  zur Einheitsmatrix und die Filter können direkt der Pseudoinversen  $\mathbf{H}^+$  entnommen werden. Für einen Hörer und zwei Lautsprecher, also  $h = 1$  und  $n = 2$ , gibt es eine exakte Lösung und die Pseudoinverse ist identisch mit der Inversen  $\mathbf{H}^{-1}$ . Gibt es über

doppelt so viele Lautsprecher als Hörer, ist also  $n > 2 \cdot h$ , bestimmt  $\mathbf{H}^+$  aus den möglichen Lösungen diejenige mit minimaler euklidischer Norm und damit die Lösung mit minimaler Leistungsaufnahme der Lautsprecher<sup>[11]</sup>. Andernfalls, also  $n < 2 \cdot h$ , ist das Problem überbestimmt und daher nicht lösbar. Die Pseudoinverse liefert dann die im quadratischen Mittel beste und ebenfalls normkleinste Näherungslösung<sup>17</sup>.

## Stereo-Dipol

Ausgehend vom symmetrischen Problem kann man zeigen, dass sich die Ohrsignale auch durch eine theoretische Wiedergabeordnung kontrollieren lassen, die sowohl aus einem Monopol als auch aus einem Dipol, überlagert an der selben Position in der Medianebene, besteht<sup>[11,184]</sup>. In der Praxis nähert man diese Situation an, indem man die Lösung des symmetrischen Problems auf zwei Lautsprecherboxen anwendet, die sehr nahe beieinander, das bedeutet bei kopfbezogenen Seitenwinkeln von etwa  $\pm 5^\circ$ , platziert sind. Das von einer solchen Anordnung erzeugte Schallfeld ändert sich wesentlich weniger stark über dem Ort als das von breiter aufgestellten Lautsprechern<sup>[88,89,184]</sup>. Man hält es daher für stabiler, falls Position oder Ausrichtung des Kopfes nicht genau dem vorgesehenen Abhörpunkt entsprechen<sup>[184]</sup>. Der Ansatz wurde zu einem Wiedergabesystem mit insgesamt drei solcher sogenannten „Stereo-Dipole“ weiterentwickelt, von denen sich je einer vor, einer über und einer hinter dem Hörer befindet<sup>[97]</sup>. Es gibt aber auch widersprechende Untersuchungen, die bei einem direkten Vergleich eine breitere Lautsprecheraufstellung von  $\pm 60^\circ$  bevorzugen<sup>z.B.[9]</sup>.

## Optimale Quellenverteilung

Im Zusammenhang mit dem Stereo-Dipol liegt die Frage nahe, ob sich ein optimaler Öffnungswinkel für das Lautsprecherpaar ermitteln lässt. Einige der durchgeführten Simulationsstudien liefern dafür ein frequenzabhängiges Ergebnis. Für den oberen Spektralbereich, etwa ab  $4\text{ kHz}$ , ist ein schmaler Öffnungswinkel von circa  $\pm 5^\circ$  – wie beim Stereo-Dipol – ideal, während sich für tiefere Frequenzen deutlich größere Winkelspannen empfehlen<sup>[207,208]</sup>. Aus dieser Tatsache leitet sich das Prinzip der „Optimalen Quellenverteilung“<sup>18</sup> ab. Basierend auf der theoretischen Vorstellung eines symmetrisch um den Hörer verteilten Schallwandlers, dessen Öffnungswinkel sich mit steigender Frequenz kontinuierlich von  $\pm 90^\circ$  bis  $0^\circ$  verkleinert, bietet es sich in der Praxis an, mit einem Mehrwegesystem zu arbeiten. Dazu wird je ein Teilband des Frequenzbereichs über verschieden weit aufgespannte Lautsprecherpaare wiedergegeben<sup>[12,185]</sup>. Durch einen zusätzlichen Lautsprecher in der Medianebene des Hörers können solche Anordnungen noch verbessert werden<sup>[186]</sup>.

---

<sup>17</sup>engl.: minimum-norm least-squares solution

<sup>18</sup>engl.: optimum source distribution (kurz: OSD, OPSODIS)

## Dynamische Systeme

Die in der Matrix  $\mathbf{H}$  zusammengefassten binauralen Raumimpulsantworten der Lautsprecher ändern sich bei jeder Bewegung des Hörers. Da sie aber die Ausgangsbasis der Filterberechnung bilden, funktionieren alle Verfahren mit vorab festgelegten Filtern nur, solange die vorgesehene Abhörposition eingehalten wird. Dynamische Systeme sind hingegen in der Lage, auf Bewegungen des Hörers zu reagieren.

Der naheliegende Ansatz besteht darin, die momentan gültigen binauralen Raumimpulsantworten zu schätzen und daraus die Filter neu zu berechnen. Umfassend beschrieben ist die Realisierung dieser Idee in einer am Massachusetts Institute of Technology 1997 eingereichten Dissertation<sup>[57]</sup>. Das dafür implementierte Labor-system mit zwei Lautsprechern greift auf eine Datenbank mit 710 kopfbezogenen Übertragungsfunktionen eines Kunstkopfs zurück. Zudem erfasst ein nach einem elektromagnetischen Prinzip arbeitendes Messgerät permanent Position und Ausrichtung des Kopfes. Basierend auf diesen beiden Informationen ist eine adäquate Kontrolle der Ohrsignale auch bei Bewegungen möglich, solange der Hörer sich nicht so positioniert, dass beide Lautsprecher auf der selben Seite seines Kopfes liegen<sup>[58]</sup>. Schon kurze Zeit später wurden weitere ähnliche Systeme vorgestellt. Sie arbeiten mit einer kamerabasierten<sup>[110]</sup> oder laserbasierten<sup>[60]</sup> Verfolgung der Kopfbewegungen, wurden aber nicht in Hörversuchen evaluiert. Auch in jüngerer Vergangenheit wurde die Idee wieder aufgegriffen<sup>[161,162]</sup>. Für solche Systeme wird oft speziell der Einsatz am Bildschirmarbeitsplatz in Betracht gezogen, da dann der zu berücksichtigende Bewegungsbereich des Hörers ausreichend limitiert ist.

Für etwas größere Dimensionen ist das dynamische System konzipiert, das in der Virtuellen Realität der Rheinisch-Westfälischen Technischen Hochschule Aachen eingesetzt wird<sup>[101,102]</sup>. Es benötigt ebenfalls eine Datenbank mit kopfbezogenen Übertragungsfunktionen. Vier Lautsprecher befinden sich dort in den oberen Ecken des von den Video-Projektionsflächen aufgespannten Quaders. Die Abhörzone ist in acht Sektoren eingeteilt, denen jeweils zwei der vier Lautsprecher zugeordnet sind. Wechselt der Hörer den Sektor, überblendet das System zwischen den beteiligten Lautsprecherpaaren. So kann dem Hörer ein angemessener Bewegungsspielraum in der Virtuellen Realität, insbesondere eine volle Drehung, eingeräumt werden. Eine Validierung, auch mit Kombination von akustischer und optischer Simulation, lieferte insgesamt gute Ergebnisse. Wegen der hoch angebrachten Lautsprecher zeigen sich aber Schwächen bei der Abbildung der Elevation virtueller Schallquellen<sup>[102]</sup>.

Andere Ansätze versuchen hingegen, die Neuberechnung der Filter zu vermeiden: Bei einem für Fluglotsen entwickelten Verfahren wird der Schall stark auf die Ohren gerichtet und bei Kopfbewegungen entsprechend nachgeführt. Es wurde für die Wiedergabe des Sprechfunkverkehrs am Arbeitsplatz entwickelt und ist daher für

den Frequenzbereich unterhalb von  $2,5\text{ kHz}$  ausgelegt<sup>[65,66]</sup>. Weitere Ideen beruhen auf konzeptionellen – also nicht realen – Schallquellen, die sich synchron zu einem oder mehreren Hörern bewegen, so dass ihre kopfbezogenen Koordinaten konstant bleiben<sup>[121,123,124,199]</sup>. Solche Ansätze fallen in den Rahmen dieser Arbeit und sind in Kap. 4 näher erläutert.

### **Adaptive Filter**

Liegen keinerlei kopfbezogene Daten vor, lässt sich prinzipiell auch der Einsatz sogenannter adaptiv invertierter Filter in Erwägung ziehen. Anstatt das System über die Berechnung der Pseudoinversen  $\mathbf{H}^+$  festzulegen, wird den Lautsprechern ein Filter  $\mathbf{H}_{inv}$  vorgeschaltet, das sich so justiert, dass es den Unterschied zwischen dem binauralen Signal und den tatsächlich resultierenden Ohrsignalen, abgesehen von einer unvermeidlichen Zeitverzögerung, minimiert. Theoretisch ist damit sogar eine Ermittlung der Filter in-situ, also direkt während der Wiedergabe, möglich<sup>[133]</sup>. Dazu müssen allerdings die Ohrsignale des Hörers bekannt sein, so dass das Verfahren in der Praxis nur eingeschränkt einsetzbar ist. Zumindest entbindet es den Anwender aber von der expliziten Ermittlung binauraler Raumimpulsantworten. Somit bieten adaptive Filter einen alternativen Ausgangspunkt, um die Wiedergabe binauraler Signale über Lautsprecher zu optimieren vgl. z. B. [10,37,84,85,105,146,151,209].



# 3

## Schallfeldrekonstruktion und Schallfokussierung

Die Idee zur Schallfeldrekonstruktion kann man auf eine simple Modellvorstellung zurückführen, die schon um 1930 formuliert wurde: Wäre man in der Lage, das Schallfeld in einem Konzertsaal entlang einer Ebene vor der Bühne mit unendlich vielen, infinitesimal kleinen Mikrofonen aufzunehmen, könnte man das bei der Aufführung erzeugte Schallfeld mit einer ebensolchen Lautsprecheranordnung an anderer Stelle wirklichkeitsgetreu nachbilden<sup>[56,160]</sup>. Dieses Prinzip ist auch unter dem Namen „Akustischer Vorhang“<sup>19</sup> bekannt. Dabei nimmt man grundsätzlich an, dass gleichartige Schallfelder auch gleichartig wahrgenommen werden. Wie bei der Binauraltechnik (vgl. Kap. 2) werden hierbei aber multimodale Effekte vernachlässigt.

Der Vorteil einer solchen Vorgehensweise ist, dass so das gesamte Schallfeld mit allen räumlichen Eigenschaften erhalten bleibt. Daher können sich darin mehrere Hörer frei bewegen, wobei für jeden eine individuelle Perspektive gilt und eine Relation zwischen allen Positionen im Schallfeld besteht. Nachteilig ist der hohe Materialaufwand, den die Kontrolle eines Schallfeldes in einem größeren Bereich erfordert.

Das allgemeine Ziel der Schallfeldrekonstruktion kann man wie folgt festhalten: Das natürliche Schallfeld einer Referenzszene soll mit solcher Genauigkeit nachgebildet werden, dass die Wahrnehmung dieses künstlichen Schallfeldes – trotz möglicher Rekonstruktionsfehler – identisch zu der des natürlichen Originals ist.

### 3.1 Wellengleichung für Schallfeldgrößen

Die Ausbreitung von Schall wird durch eine Wellentheorie beschrieben, die aus elementaren Gesetzen der klassischen Physik abgeleitet werden kann. Die folgenden drei Gleichungen sind in der angegebenen Form bereits durch eine lineare Näherung

<sup>19</sup>engl.: screen analogy; die teilweise verwendete Rückübersetzung „acoustic curtain“ bezeichnet hingegen einen Vorhang, der in der Raumakustik zur Schalldämpfung dient

vereinfacht. Dafür wird vorausgesetzt, dass die betrachteten Schwankungsgrößen Schalldruck  $p$  und Schalldichte  $\varrho$  klein gegen die mittlere Dichte  $\varrho_0$  und den statischen Druck  $p_0$  des Ausbreitungsmediums sind<sup>[140,S. 14 ff.]</sup>.

Gl. (3.1) folgt aus dem Massenerhaltungssatz,  $\vec{v}$  bezeichnet die Schallschnelle<sup>[50,S. 22]</sup>:

$$\frac{\partial \varrho}{\partial t} + \varrho_0 \nabla(\vec{v}) = 0 \quad (3.1)$$

Der Impulserhaltungssatz liefert mit Gl. (3.2) eine vereinfachte Form der Eulerschen Bewegungsgleichung<sup>[50,S. 22]</sup>:

$$\varrho_0 \frac{\partial \vec{v}}{\partial t} + \nabla p = 0 \quad (3.2)$$

Aus dem Energieerhaltungssatz ergibt sich – unter der Annahme adiabatischer Zustandsänderungen – aus Gesetzen der Thermodynamik mit Gleichung (3.3) eine Relation zwischen Druck  $p$  und Dichte  $\varrho$ , wobei  $c$  der Ausbreitungsgeschwindigkeit des Schalls<sup>20</sup> entspricht<sup>[50,S. 22]</sup>:

$$p = c^2 \varrho \quad (3.3)$$

Eliminiert man  $\varrho$  mit Hilfe von Gl. (3.3) aus Gl. (3.1) und differenziert das Ergebnis nach der Zeit, erhält man mit dem Einsetzen von Gl. (3.2) die homogene Wellengleichung<sup>[140,S. 17]</sup>:

$$\Delta p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 \quad (3.4)$$

Die homogene Wellengleichung (3.4) beschreibt in dieser Form das Verhalten des Schalldrucks über Raum und Zeit bei Schallausbreitung in einem quellfreien Fluid. Das Gebiet der Schallerzeugung wird also demnach nicht betrachtet, sondern nur das Gebiet der Schallausbreitung. Mathematisch wird Gl. (3.4) als hyperbolische partielle Differentialgleichung 2. Ordnung klassifiziert<sup>[145,S. 234]</sup>.

Durch die Fouriertransformation von Gl. (3.4) ergibt sich die sogenannte homogene Helmholtzgleichung (3.5)<sup>[220,S. 18 f.]</sup>:

$$\Delta P(\vec{x}, \omega) + k^2 P(\vec{x}, \omega) = 0 \quad (3.5)$$

mit der Wellenzahl

$$k = \frac{\omega}{c} = \frac{2\pi \cdot f}{c} \quad (3.6)$$

<sup>20</sup>unter Laborbedingungen, d. h. bei einer Raumtemperatur von 20°C und trockener Luft gilt  $c \approx 343 \frac{\text{m}}{\text{s}}$  <sup>[50,S. 18]</sup>

Im Gegensatz zu Gl. (3.4) ist Gl. (3.5) eine elliptische partielle Differentialgleichung 2. Ordnung<sup>[145,S.234]</sup>. Ausgehend von der Helmholtzgleichung (3.5) lassen sich unter Zuhilfenahme geeigneter Koordinatentransformationen elementare Lösungen der homogenen Wellengleichung ermitteln. So ergeben sich die Beschreibungen von ebenen Wellen, Kugelwellen und Zylinderwellen<sup>[50,S.25-27]</sup>.

## Green-Funktionen

Eine Green-Funktion  $G(\vec{x}|\vec{x}_0)$  ist allgemein die partikuläre Lösung einer inhomogenen linearen Differentialgleichung, wenn der Störterm eine Dirac-Deltafunktion  $\delta(\vec{x} - \vec{x}_0)$  ist<sup>[145,S.239]</sup>. Demgemäß gilt für die Helmholtzgleichung dann<sup>[220,S.265]</sup>

$$\Delta G(\vec{x}|\vec{x}_0, \omega) + k^2 G(\vec{x}|\vec{x}_0, \omega) = \delta(\vec{x} - \vec{x}_0) \quad (3.7)$$

Im Dreidimensionalen und unter Freifeldbedingungen lautet die zugehörige Green-Funktion<sup>[220,S.265]</sup>

$$G(\vec{x}|\vec{x}_0, \omega) = -\frac{e^{j\frac{\omega}{c}|\vec{x}-\vec{x}_0|}}{4\pi|\vec{x}-\vec{x}_0|} \quad (3.8)$$

Mit ihrer Hilfe kann nun die Lösung der inhomogenen Helmholtzgleichung innerhalb eines Volumens  $V$  für eine beliebige Störung  $\Sigma(\vec{x}, \omega)$  angegeben werden. Es gilt<sup>[145,S.239]</sup>

$$P(\vec{x}, \omega) = \iiint_V G(\vec{x}|\vec{x}_0, \omega) \Sigma(\vec{x}_0, \omega) d\vec{x}_0 \quad (3.9)$$

löst

$$\Delta P(\vec{x}, \omega) + k^2 P(\vec{x}, \omega) = \Sigma(\vec{x}, \omega) \quad (3.10)$$

für

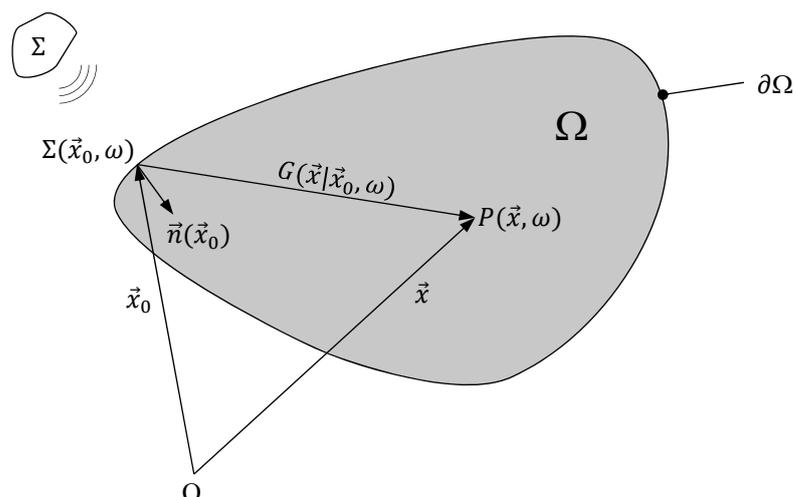
$$\vec{x}, \vec{x}_0 \in V \quad (3.11)$$

Die Gln. (3.7), (3.8)<sup>21</sup> beschreiben eine Punktschallquelle an der Stelle  $\vec{x}_0$ <sup>[220,S.198&265]</sup>. Die Green-Funktion  $G(\vec{x}|\vec{x}_0, \omega)$  kann in diesem Fall als Freifeldübertragungsfunktion, ermittelt am Ort  $\vec{x}$  für eine bei  $\vec{x}_0$  platzierte Punktschallquelle, interpretiert werden.

## 3.2 Kirchhoff-Helmholtz Integral

Eine allgemeine Lösung der homogenen Helmholtzgleichung für inhomogene Randbedingungen ist durch das Kirchhoff-Helmholtz-Integral beschrieben<sup>[5,S.26][165,S.33]</sup>.

<sup>21</sup>aus Gründen der Konsistenz sind in beiden Gleichungen die negativen Vorzeichen angepasst



**Abbildung 3.1** Geometrie des Kirchhoff-Helmholtz-Integrals. Damit kann der Schalldruck an jedem Punkt innerhalb eines quellenfreien Volumens  $P(\vec{x}), \forall \vec{x} \in \Omega$  bestimmt werden, wenn auf dem gesamten Rand  $\partial\Omega$  Schalldruck – hier mit  $\Sigma(\vec{x}_0)$  bezeichnet – sowie Schallschnelle (in Richtung der Normalen  $\vec{n}$ ) bekannt sind. Die Green-Funktion  $G(\vec{x}|\vec{x}_0, \omega)$  muss so bestimmt sein, dass sie die vorgegebenen Randbedingungen erfüllt vgl. [165, Figure 4.1].

Für den Schalldruck  $P$  an der Stelle  $\vec{x}$  gilt<sup>[5, S. 26]</sup>

$$a(\vec{x})P(\vec{x}, \omega) = - \iint_{\partial\Omega} \left[ G(\vec{x}|\vec{x}_0, \omega) \frac{\partial}{\partial \vec{n}(\vec{x}_0)} \Sigma(\vec{x}_0, \omega) - \Sigma(\vec{x}_0, \omega) \frac{\partial}{\partial \vec{n}(\vec{x}_0)} G(\vec{x}|\vec{x}_0, \omega) \right] dS(\vec{x}_0) \quad (3.12)$$

mit

$$a(\vec{x}) = \begin{cases} 1 & \forall \vec{x} \in \Omega \\ \frac{1}{2} & \forall \vec{x} \in \partial\Omega \\ 0 & \forall \vec{x} \notin \bar{\Omega} = \Omega \cup \partial\Omega \end{cases}$$

Darin bezeichnet  $\Omega \subset \mathbf{R}^3$  ein Volumen mit dem Rand  $\partial\Omega$  und  $dS(\vec{x}_0)$  ein infinitesimales Randelement mit nach innen zeigendem Normaleneinheitsvektor  $\vec{n}(\vec{x}_0)$ . Der Ausdruck  $\frac{\partial}{\partial \vec{n}(\vec{x}_0)}$  steht für die Normalableitung, das ist die Richtungsableitung entlang der Normalen  $\vec{n}(\vec{x}_0)$ <sup>[145, S. 254&221]</sup>. In Gl. (3.12) ist  $\vec{x}_0$  grundsätzlich ein Randpunkt, die Abhängigkeit des Integrals von den Werten am Rand des Volumens ist explizit verdeutlicht.

Eine übliche Interpretation lautet, dass der Schalldruck an der Stelle  $\vec{x}$  innerhalb eines quellenfreien Volumens  $P(\vec{x}), \forall \vec{x} \in \Omega$  eindeutig bestimmt werden kann, wenn

man am Rand des Volumens  $\partial\Omega$  den Schalldruck  $\Sigma(\vec{x}_0)$  und seine Normalableitung  $\frac{\partial}{\partial \vec{n}(\vec{x}_0)}\Sigma(\vec{x}_0)$  kennt. Mit der Euler-Gleichung (3.2), die im Frequenzbereich

$$j\omega\rho_0\vec{v} = \nabla P \quad (3.13)$$

lautet<sup>[220,S.18 f.]</sup>, kann  $\frac{\partial}{\partial \vec{n}(\vec{x}_0)}\Sigma(\vec{x}_0)$  als Schallschnelle in Normalrichtung aufgefasst werden. Die Green-Funktion  $G(\vec{x}|\vec{x}_0)$  gibt dabei an, wie sich der Schall – anschaulich beschrieben – vom Rand aus in das Volumen hinein ausbreitet. Bei ihrer Bestimmung müssen die vorgegebenen Randbedingungen berücksichtigt werden. Im Freifeld, also bei einer nur gedachten oder akustisch transparenten Berandung kann auf Gl. (3.8) zurückgegriffen werden<sup>[5,S.26]</sup>.

Das Volumen  $\Omega$  darf keine Schallquellen enthalten, da für die Herleitung von Gl. (3.12) die Gültigkeit der homogenen Helmholtzgleichung innerhalb des Volumens vorausgesetzt wird<sup>[165,S.31 ff.] [220,S.251 ff.]</sup>. Ein Schallfeld außerhalb des Volumens ist erlaubt, es kann jedoch nicht mit dem Kirchhoff-Helmholtz-Integral bestimmt werden, da sich gemäß der angegebenen Fallunterscheidung für alle Punkte außerhalb des Volumens der Wert Null ergibt<sup>vgl. [220,S.257]</sup>.

### 3.3 Wellenfeldsynthese

Die Wellenfeldsynthese<sup>22</sup> ist das bekannteste Verfahren zur Schallfeldrekonstruktion. Ihre Popularität ist auf das zwischen Januar 2001 und Juni 2003 von der Europäischen Kommission geförderte Forschungsprojekt CARROUSO<sup>23</sup> mit zehn Projektpartnern aus fünf verschiedenen Ländern – Deutschland, Frankreich, Schweiz, Niederlande und Griechenland – zurückzuführen<sup>[32,204]</sup>. Die ursprüngliche Idee wurde an der Technischen Universität Delft formuliert und entstammt der Seismik. In den 1990er Jahren wurde sie dort auch maßgeblich weiterentwickelt. Als Grundstein gelten zwei Aufsätze aus den Jahren 1988<sup>[20]</sup> und 1993<sup>[21]</sup>.

Der hohe Bekanntheitsgrad induzierte zahlreiche Forschungsarbeiten mit unterschiedlichsten Zielsetzungen. Das Resultat ist eine unüberschaubare Anzahl von Publikationen, die im Zusammenhang mit Wellenfeldsynthese stehen. Nachdem diverse Laborsysteme die praktische Anwendbarkeit untermauert hatten, wird das Verfahren mittlerweile weltweit kommerziell genutzt. Ein bekanntes Beispiel aus dem deutschsprachigen Raum ist die Beschallungsanlage der Seebühne in Bregenz<sup>[204]</sup>.

<sup>22</sup>engl.: wave field synthesis (kurz: WFS)

<sup>23</sup>Projektreferenz IST-1999-20993, das Akronym steht für engl.: creating, assessing and rendering in real time of high quality audio-visual environments in MPEG-4 context

Ausgangspunkt der Überlegungen zur Wellenfeldsynthese ist folgende Interpretation des Kirchhoff-Helmholtz-Integrals Gl. (3.12) aus Abschn. 3.2. Eine Green-Funktion  $G(\vec{x}|\vec{x}_0)$  nach Gl. (3.8) charakterisiert, wie erwähnt, das Abstrahlverhalten eines Monopols an der Stelle  $\vec{x}_0$ . Einen Dipol erhält man durch die Überlagerung zweier gegenphasiger Monopole mit infinitesimalem Abstand<sup>[50,S.32]</sup>. Formal ist dadurch eine Differentiation festgelegt<sup>[67,S.32]</sup>

$$\lim_{\vec{x}_1 \rightarrow \vec{x}_2} \frac{G(\vec{x}|\vec{x}_1) - G(\vec{x}|\vec{x}_2)}{|\vec{x}_2 - \vec{x}_1|} = -\frac{\vec{x}_2 - \vec{x}_1}{|\vec{x}_2 - \vec{x}_1|} \nabla G(\vec{x}|\vec{x}_2) \quad (3.14)$$

wobei der Einheitsvektor  $-\frac{\vec{x}_2 - \vec{x}_1}{|\vec{x}_2 - \vec{x}_1|}$  die Dipolachse, also die Hauptabstrahlrichtung, festlegt. Für die Normalableitung der Green-Funktion gilt allgemein<sup>[145,S.221]</sup>

$$\frac{\partial}{\partial \vec{n}} G(\vec{x}|\vec{x}_0) = \vec{n} \cdot \nabla G(\vec{x}|\vec{x}_0) \quad (3.15)$$

$$\begin{aligned} &= \vec{n} \cdot \frac{\partial G(\vec{x}|\vec{x}_0)}{\partial(\vec{x} - \vec{x}_0)} \\ &= \vec{n} \cdot \frac{\vec{x} - \vec{x}_0}{|\vec{x} - \vec{x}_0|} \left[ \left( j\frac{\omega}{c} - \frac{1}{|\vec{x} - \vec{x}_0|} \right) G(\vec{x}|\vec{x}_0) \right] \end{aligned} \quad (3.16)$$

Der direkte Vergleich von Gl. (3.15) mit Gl. (3.14) zeigt, dass die Normalableitung der Green-Funktion einen Dipol beschreibt, der entlang der Normalen ausgerichtet ist.

Aus dem Kirchhoff-Helmholtz-Integral Gl. (3.12) kann man damit auch schlussfolgern, dass sich ein Schallfeld innerhalb eines Volumens mit kontinuierlich auf dem Rand des Volumens verteilten Monopolen und Dipolen erzeugen lässt<sup>[30]</sup>. Der Ansatz der Wellenfeldsynthese ist, diese Aussage direkt in die Praxis umzusetzen. Dazu müssen einige vereinfachende Annahmen getroffen werden.

Um ausschließlich geschlossene Lautsprecherboxen verwenden zu können, rückt man von Freifeldbedingungen ab und definiert folgende Randbedingung, durch die eine Dipolverteilung auf dem Rand obsolet wird.

$$\frac{\partial}{\partial \vec{n}(\vec{x}_0)} G(\vec{x}|\vec{x}_0, \omega) = 0 \quad \forall \vec{x}_0 \in \partial\Omega \quad (3.17)$$

Gl. (3.17) fordert eine schallharte Begrenzung. In Anlehnung an die für die Laplace-Gleichung formulierten Randwertprobleme wird diese Vorgabe auch als Neumann-Randbedingung bezeichnet<sup>vgl. [145,S.235]</sup>. Der Subtrahend im Kirchhoff-Helmholtz-Integral verschwindet, allerdings muss die Green-Funktion an das veränderte Problem angepasst werden. Für ein Volumen in Form eines Halbraums mit unendlich großer Begrenzungsebene lässt sich zeigen, dass die Neumann Green-Funktion  $G_N(\vec{x}|\vec{x}_0)$  für die gleichnamige Randbedingung das Zweifache der Green-Funktion für Freifeldbedingungen ist<sup>[220,S.276]</sup>

$$G_N(\vec{x}|\vec{x}_0, \omega) = 2 G(\vec{x}|\vec{x}_0, \omega) \quad (3.18)$$

Das Kirchhoff-Helmholtz-Integral vereinfacht sich dann zum sogenannten Ersten Rayleigh-Integral<sup>[220,S. 34 ff. & 276]</sup>, das die mathematische Beschreibung des akustischen Vorhangs liefert und eine kontinuierliche Monopolverteilung voraussetzt. Damit folgt aus Gl. (3.12) nun

$$P(\vec{x}, \omega) = \oint_{\partial\Omega} -2 G(\vec{x}|\vec{x}_0, \omega) \frac{\partial}{\partial \vec{n}} \Sigma(\vec{x}_0, \omega) dS \quad (3.19)$$

Ist die Abmessung eines Volumens groß gegen die Wellenlänge, kann die Randfläche durch Tangentialebenen approximiert werden. Für konvexe Volumen kann man daher in guter Näherung ebenfalls die Gln. (3.18) und (3.19) anwenden<sup>[5,S. 29 ff.] [170]</sup>.

Die Realisierung eines akustischen Vorhangs für die dreidimensionale Schallfeldrekonstruktion ist mit hohem technischen Aufwand verbunden, weil wiedergabeseitig eine Art Lautsprecherwand benötigt wird<sup>vgl. z. B. [135,148]</sup>. Daher reduziert man in der Praxis das Problem um eine Dimension auf die Schallfeldrekonstruktion in einer Fläche auf Ohrhöhe des Publikums<sup>[21]</sup>. Da die Schallausbreitung in der Realität aber dreidimensional erfolgt, muss man zuerst annehmen, dass das Schallfeld entlang einer Richtung des Koordinatensystems konstant bleibt, und betrachtet anschließend eine Schnittebene. Genau genommen kann ein solches Schallfeld aber nur von einer kontinuierlichen Verteilung unendlich langer Linienquellen, die senkrecht zu dieser Schnittebene stehen und Zylinderwellen abstrahlen, erzeugt werden<sup>[220,S. 265 ff.]</sup>. Es sollen jedoch weiterhin Monopole, also Punktquellen, zum Einsatz kommen. Da sich die Schallintensität einer Kugelwelle aber umgekehrt proportional zum Quadrat des Abstands, und nicht wie bei einer Zylinderwelle nur umgekehrt proportional zum Abstand verändert<sup>[50,S. 26 f.]</sup>, ist es nicht möglich, ein Schallfeld innerhalb der gesamten durch eine Lautsprecherzeile begrenzten Halbebene oder einer von Lautsprechern umrandeten Hörfläche exakt zu reproduzieren. Der Fehler kann lediglich im ersten Fall für eine Referenzlinie oder im zweiten Fall für einen Referenzpunkt durch einen Korrekturfaktor kompensiert werden<sup>[21,170,175,202,203]</sup>.

In der Theorie wird bisher eine kontinuierliche Verteilung der sogenannten Sekundärquellen angenommen. Zusätzlich fordert das Erste Rayleigh-Integral Gl. (3.19) noch eine unendliche Randebene oder eine unendliche lange Randlinie. In der Praxis kann man allerdings nur eine endliche Anzahl von Lautsprechern an diskreten Punkten auf dem Rand positionieren, mathematisch geht damit das Integral in eine endliche Summe über. Ähnlich wie bei der Analyse von Zeitsignalen spricht man dabei von räumlicher Abtastung<sup>24</sup> und bei Lautsprecherzeilen zusätzlich von räumlicher Fensterung<sup>25 [21]</sup>.

---

<sup>24</sup>engl.: spatial sampling

<sup>25</sup>engl.: truncation

Das Zusammenwirken dieser beiden Effekte führt dazu, dass die räumliche Feinstruktur eines Schallfeldes nur in reduziertem Umfang rekonstruierbar ist, oder dass – umgekehrt ausgedrückt – eine diskrete und endliche Sekundärquellenverteilung nur ein Schallfeld mit begrenzter räumlicher Komplexität erzeugen kann<sup>[4]</sup>. Je niedriger die Abtastrate, also je größer der Abstand zwischen den Lautsprechern ist, desto deutlicher treten sogenannte Alias-Effekte auf. Für einen regelmäßigen Lautsprecherabstand  $\Delta x$  liefert die Berücksichtigung des Abtasttheorems im Ortsfrequenzbereich eine Abschätzung, für welches Frequenzband ein beliebiges Schallfeld verlustfrei darstellbar ist<sup>[21,172]</sup>:

$$f_{\text{alias}} \leq \frac{c}{2\Delta x} \quad (3.20)$$

Eine strikte mathematische Herleitung deckt zudem auf, dass durch die Abtastung auch evaneszente Wellen entstehen. Sie sind unvermeidbar, aber klingen gemäß ihrer Natur mit zunehmender Entfernung vom Rand sehr schnell ab<sup>[172]</sup>. Gl. (3.20) liefert bei realisierbaren Lautsprecherabständen von wenigen Dezimetern Grenzfrequenzen kleiner als  $2\text{ kHz}$ . Falls ein Testschall Spektralanteile oberhalb der Grenzfrequenz enthält, sind im direkten Vergleich zwischen Synthese und Original Unterschiede wahrnehmbar<sup>[109]</sup>. Zudem gibt es Anzeichen, dass die räumliche Unterabtastung Klangfarbenänderungen verursacht und die Lokalisation beeinträchtigt<sup>[4,172,193,222]</sup>. Es existieren zwar verschiedene Vorschläge für Gegenmaßnahmen vgl. z. B. [42,44,179], in der Praxis werden die Alias-Effekte jedoch meist toleriert und gelten nicht als gravierende Einschränkung<sup>[165,S. 165]</sup>.

Durch die räumliche Rechteck-Fensterung entstehen Beugungsartefakte, die in Form von Echos hörbar werden können. Sie lassen sich durch besser geeignete Fensterfunktionen<sup>26</sup> abmildern, mit denen die Lautsprecher zu den Enden der Anordnung hin stärker gedämpft werden<sup>[21,193]</sup>. Dafür muss jedoch eine Verkleinerung des Hörbereichs in Kauf genommen werden<sup>[193]</sup>.

## Datenbasierte und modellbasierte Wellenfeldsynthese

Es gibt zwei wesentliche Ansätze, mit denen den Randwerten eines Schallfeldes für die Wiedergabe ermittelt werden. Der erste, datenbasierte Wellenfeldsynthese, führt die Idee des akustischen Vorhangs konsequent fort. Mit einer geeigneten Mikrofonanordnung wird ein reales Schallfeld entlang eines Teilvolumens oder einer Teilfläche aufgenommen. Diese Signale können dann mit einer passenden Lautsprecheranordnung wiedergegeben werden und rekonstruieren weitgehend das Originalschallfeld innerhalb des von den Mikrofonen abgegrenzten Bereichs. Sind Randbedingungen und Geometrie auf der Aufnahmeseite und auf der Wiedergabe-

---

<sup>26</sup>engl.: tapering

seite nicht deckungsgleich, muss eine Umrechnung, die sogenannte Extrapolation, erfolgen<sup>[21,204]</sup>. Dieses Prinzip lässt sich noch weiter verallgemeinern, indem man das Schallfeld anhand der Informationen aus der Aufnahme in eine Überlagerung von ebenen Wellen oder Kugelflächenfunktionen zerlegt<sup>27</sup>. Damit erreicht man eine höhere Flexibilität bei der Weiterverarbeitung<sup>[122,204]</sup>.

Bei der modellbasierten Wellenfeldsynthese, dem zweiten Ansatz, ist der Aufnahmeschritt nicht notwendig. Stattdessen berechnet man den Schalldruckgradienten, den eine fiktive Primärschallquellenanordnung an den Lautsprecherpositionen erzeugen würde. Daraus ergibt sich für jeden Lautsprecher eine Ansteuerungsfunktion<sup>28</sup>. Relativ einfach lässt sich dies für elementare Schallfeldtypen wie ebene Wellen, Kugelwellen und Zylinderwellen durchführen<sup>[170,202]</sup>. Bei Bedarf lassen sich auch kompliziertere Schallfelder berücksichtigen, zum Beispiel von Punktschallquellen mit bestimmter Richtcharakteristik<sup>[1,43]</sup>, von schwingenden Objekten mit beliebiger konvexer Form<sup>[6,7]</sup> oder von sich bewegenden Schallquellen<sup>[3]</sup>.

Die Herleitung der Ansteuerungsfunktionen wird im Folgenden exemplarisch für eine primäre Quelle in Form eines kugelförmig abstrahlenden Monopols außerhalb des Hörbereichs kurz erläutert. Als Sekundärquellenverteilung wird eine geradlinige Lautsprecherzeile, bestehend aus  $n$  Lautsprechern mit regelmäßigem Abstand  $\Delta x$ , angenommen, die Primärschallquelle befindet sich bei  $\vec{x}_s$ . Gl. (3.19) wird zu

$$P(\vec{x}) \approx \sum_{i=1}^n \underbrace{-2 \frac{\partial}{\partial \vec{n}} \Sigma(\vec{x}_s, \vec{x}_i) \cdot C(\vec{x}_i, \vec{x}_{\text{ref}})}_{D_i} G(\vec{x}|\vec{x}_i) \quad (3.21)$$

$D_i$  bezeichnet die Ansteuerungsfunktion des  $i$ -ten Lautsprechers an der Position  $\vec{x}_i$ ,  $\Sigma(\vec{x}_s, \vec{x}_i)$  ist das Schalldruckspektrum, das die fiktive Primärquelle dort verursacht.  $C(\vec{x}_i, \vec{x}_{\text{ref}})$  ist der bereits erwähnte Korrekturfaktor, mit dem der Fehler, der bei der Vereinfachung auf zwei Dimensionen entsteht, zumindest für eine Referenzlinie parallel zur Lautsprecherzeile im Abstand  $\vec{x}_{\text{ref}}$  korrigiert werden kann. Die weiterhin bestehende Abhängigkeit von der Kreisfrequenz  $\omega$  wurde der Übersichtlichkeit halber in Gl. (3.21) nicht notiert. Für einen kugelförmig abstrahlenden Monopol gilt<sup>[50,S.25]</sup>

$$\Sigma(\vec{x}_s, \vec{x}, \omega) = \hat{S}(\omega) \cdot \frac{e^{-j\frac{\omega}{c}|\vec{x}-\vec{x}_s|}}{|\vec{x}-\vec{x}_s|} \quad (3.22)$$

wobei die frequenzabhängige Konstante  $\hat{S}(\omega)$  die Quellstärke bestimmt und auch als Monopolamplitude bezeichnet wird<sup>[140,S.159 ff.]</sup>. Für die Ansteuerungsfunktion ergibt sich also zusammen mit der Normalableitung von Gl. (3.22) schließlich vgl. [170]

$$D_i = -2 \cdot C(\vec{x}_i, \vec{x}_{\text{ref}}, \omega) \vec{n} \cdot \frac{\vec{x}_i - \vec{x}_s}{|\vec{x}_i - \vec{x}_s|} \left( \frac{1}{|\vec{x}_i - \vec{x}_s|} + j\frac{\omega}{c} \right) \hat{S}(\omega) \frac{e^{-j\frac{\omega}{c}|\vec{x}_i - \vec{x}_s|}}{|\vec{x}_i - \vec{x}_s|} \quad (3.23)$$

<sup>27</sup>engl.: plane wave/spherical harmonics decomposition

<sup>28</sup>engl.: driving function; früher auch: Syntheseoperator, engl.: synthesis operator

Die Ansteuerungsfunktionen lassen sich im Zeitbereich effizient implementieren. Der hintere Teil von Gl. (3.23) stellt eine Zeitverschiebung des Quellsignals  $s(t)$  um  $\frac{|\vec{x}_i - \vec{x}_s|}{c}$  dar, gewichtet mit  $\frac{1}{|\vec{x}_i - \vec{x}_s|}$ , was mit allen weiteren frequenzunabhängigen Termen einen Verstärkungsfaktor  $a_i$  ergibt. Die frequenzabhängigen Anteile fasst man im Filter  $h(t)$  zusammen. Nach der Transformation ergibt sich damit<sup>vgl. [170]</sup>

$$d_i = s(t) * h(t) * a_i \delta\left(t - \frac{|\vec{x}_i - \vec{x}_s|}{c}\right) \quad (3.24)$$

Nach Gl. (3.24) erhält man die Ansteuerungsfunktion des  $i$ -ten Lautsprechers  $d_i$  also durch eine Vorfilterung mit  $h(t)$  sowie eine Verstärkung und Verzögerung des Quellsignals  $s(t)$ . Verstärkungsfaktor und Verzögerungszeit sind im Wesentlichen durch den Abstand zwischen Primärquelle und Sekundärquelle  $|\vec{x}_i - \vec{x}_s|$  bestimmt. Das Vorfilter  $h(t)$  beruht unter anderem auf der Definition des Korrekturfaktors  $C(\vec{x}_i, \vec{x}_{\text{ref}}, \omega)$  und den bei seiner Herleitung getroffenen Annahmen. Im Idealfall ist es nach der Zusammenfassung und Vereinfachung aller frequenzabhängigen Terme nicht mehr abhängig von der Lautsprecherposition  $\vec{x}_i$  und somit in allen Ansteuerungsfunktionen identisch.

In den Korrekturfaktor  $C$  kann die räumliche Fensterfunktion zur Verringerung der Beugungsartefakte einfließen. Sollte statt einer Lautsprecherzeile eine Anordnung zum Einsatz kommen, bei der die Hörzone von Lautsprechern umschlossen ist, muss darin zusätzlich berücksichtigt werden, dass nur die in Ausbreitungsrichtung des Primärschallfeldes zeigenden Lautsprecher aktiviert werden<sup>[168]</sup>.

Wellenfeldsynthese erlaubt also die Rekonstruktion des Schallfeldes fiktiver Primärschallquellen mit einfachen Signalverarbeitungsschritten. Dazu werden aber während der Herleitung aus dem Kirchhoff-Helmholtz-Integral (3.12) diverse Vereinfachungen festgelegt, die zu teils erheblichen Abweichungen zwischen dem realen und dem rekonstruierten Schallfeld führen. Es ist daher empfehlenswert, die Gültigkeit der Annahmen am konkreten Anwendungsbeispiel zu überprüfen und die möglichen Auswirkungen der Fehler abzuschätzen. Besondere Aufmerksamkeit ist geboten bei Aufbauten mit wenigen Lautsprechern oder weiten Abständen, sowie Primärquellenpositionen nahe der Lautsprecheranordnung oder außerhalb des von den Sekundärquellen direkt kontrollierten Abschnitts<sup>vgl. [175]</sup>.

### 3.4 Schallfokussierung

Grundsätzlich setzt das Kirchhoff-Helmholtz-Integral voraus, dass das Volumen keine Schallquellen enthält (vgl. Abschn. 3.2). Es ist jedoch erlaubt, ein Schallfeld zu konstruieren, das von den Lautsprechern aus auf einen Punkt im Volumen zuläuft,

dort konvergiert und dahinter divergiert. Für Anordnungen, bei denen die Hörzone von Lautsprechern umschlossen ist, muss auch hier berücksichtigt werden, dass dafür nur ein Teil der Lautsprecher aktiviert werden darf<sup>[168]</sup>. Befindet sich der Hörer im divergierenden Teil, nimmt er eine Schallquelle am Konvergenzpunkt innerhalb des Volumens wahr<sup>[2]</sup>. Durch dieses Vorgehen ist es möglich, virtuelle Schallquellen im Gebiet zwischen den Lautsprechern und dem Hörer zu platzieren. Daraus ergeben sich erweiterte Möglichkeiten zur Tiefenstaffelung der Schallquellen. Zum Beispiel kann dadurch in Virtuellen Realitäten eine zur stereoskopischen Bildwiedergabe korrespondierende Tondarstellung realisiert werden.

Die Idee taucht im Rahmen der Wellenfeldsynthese Mitte der 1990er Jahre auf, wobei zur expliziten Unterscheidung von virtuellen Schallquellen vor und hinter den Lautsprechern auch die Bezeichnung „fokussierte Punktquelle“, oder verkürzt „Fokusquelle“, entstanden ist<sup>29</sup>.

Ähnliche Ansätze gibt es auch in anderen Teilgebieten der Akustik. Vor allem im Ultraschallbereich setzt man Methoden zur Schallfokussierung beispielsweise in der Medizintechnik, zur Unterwasserkommunikation, zur zerstörungsfreien Materialprüfung oder in der Seismik ein<sup>[168,182]</sup>. Anders als bei Wellenfeldsynthese besteht bei diesen Verfahren das Ziel hauptsächlich darin, Schallenergie an einem Punkt zu konzentrieren. Die Schallausbreitung vom Fokuspunkt weg wird nicht betrachtet<sup>[168]</sup>. Die Herausforderung liegt eher darin, komplizierte Ausbreitungsbedingungen wie ein inhomogenes Medium zu berücksichtigen<sup>[14]</sup>. Das dabei angewandte Prinzip wird als Zeitumkehr<sup>30</sup> bezeichnet<sup>[226]</sup>, dessen Übertragbarkeit auf den Hörschallbereich mittlerweile mehrfach bestätigt ist<sup>vgl. z. B. [14,163,182,226]</sup>.

Schallfokussierung beruht darauf, dass die Wellengleichung, wie sie in Abschn. 3.1 hergeleitet wurde, symmetrisch gegenüber einer Umkehrung der Zeit ist. Wenn also  $p(t)$  die Wellengleichung löst, gilt das theoretisch auch für  $p(-t)$ <sup>[14,226]</sup>. In der Praxis bedeutet das: Die Ansteuerungsfunktionen für die Schallfokussierung erhält man durch Zeitumkehr der an den Positionen aller Sekundärquellen gemessenen oder berechneten Impulsantworten  $h_i(t)$  einer am Fokuspunkt platzierten Schallquelle. Um Kausalität zu garantieren, müssen dazu alle Impulsantworten mindestens um die Dauer der längsten Impulsantwort  $T$  verschoben werden, demnach gilt  $d_i^{\text{fokus}} = h_i(T - t)$ . Ein so erzeugter Fokuspunkt kann – wenn sich wie oben beschrieben durch die Anordnung der Sekundärquellen tatsächlich ein divergierender Schallfeldanteil ergibt – als virtuelle Schallquelle interpretiert werden<sup>[206]</sup>.

---

<sup>29</sup>Originalzitat: „The potential of the method is further demonstrated by the **principle of focussing**, by which it is possible to position a virtual source *in front of* the reproduction array, instead of behind it.“<sup>[31,S.1]</sup>

<sup>30</sup>engl.: time reversal

Ein vergleichbares Ergebnis erhält man auch, wenn die Ansteuerungsfunktion im Frequenzbereich ermittelt wird. Dazu modelliert man entweder eine akustische Senke<sup>[171,173]</sup>, oder man arbeitet mit den Transformationspaaren<sup>[145,S.313]</sup>

$$\begin{aligned} f(-t) &\circ\text{---}\bullet F(-\omega) \\ f(t-T) &\circ\text{---}\bullet e^{-j\omega T}F(\omega) \end{aligned}$$

Durch die räumliche Abtastung und Fensterung, sowie durch die Vereinfachung auf zwei Dimensionen treten auch bei Fokusquellen Artefakte auf<sup>[171,173,215]</sup>. Am Fokuspunkt selbst ist das Schallfeld frei von Alias-Effekten, da sich dort die Sekundärwellen gleichphasig überlagern<sup>[171]</sup>. Der Radius der aliasfreien Zone  $r_{\text{alias}}$  lässt sich bei Lautsprecherzeilen mit der empirischen Formel<sup>[215]</sup>

$$r_{\text{alias}} = \frac{c}{f} \cdot \frac{d_{\text{fokus}}}{\Delta x}$$

abschätzen. Er hängt von der Entfernung der Fokusquelle zu den Lautsprechern  $d_{\text{fokus}}$ , dem Lautsprecherabstand  $\Delta x$  sowie der Wellenlänge  $\lambda = c/f$  ab und wird mit zunehmender Frequenz kleiner. Zusätzlich schränkt die räumliche Fensterung den Gültigkeitsbereich der Fokusquelle, in dem das divergierende Schallfeld wie gewünscht auftritt, näherungsweise auf ein Dreieck ein<sup>[173]</sup> und verursacht einen verbreiterten Fokuspunkt<sup>[214,215]</sup>. Ein spezielles Problem bei Fokusquellen ist, dass besonders bei großen Lautsprecheranordnungen wegen des Zeitumkehrprinzips hörbare Vorechos entstehen<sup>[171,173,214,215]</sup>. Im Rahmen der Wellenfeldsynthese sind die Artefakte von fokussierten Quellen und geeignete Gegenmaßnahmen aktuelle Forschungsthemen<sup>[173,214,215]</sup>.

Auch unter Schallfokussierung einzuordnen sind Methoden, mit denen man Teile eines Wiedergabebereichs selektiv beschallen kann. Sie verfolgen nicht das Ziel, Schall an einem Punkt zu konzentrieren, sondern die Schallenergie so zu verteilen, dass „akustisch helle“ und „akustisch dunkle“, das bedeutet stark und schwach beschallte, Zonen entstehen. So soll die Störung der Umgebung verringert oder die Privatsphäre erhöht werden<sup>vgl. z. B. [38,87,100]</sup>. Auch die Technik der Strahlenbündelung<sup>31</sup>, mit der sich ein Schallfeld mit einer räumlich stark begrenzten Hauptabstrahlrichtung erzeugen lässt, kann man hier einsortieren<sup>vgl. z. B. [65]</sup>.

---

<sup>31</sup>engl.: beam forming

# 4

## Ein virtueller Kopfhörer basierend auf Wellenfeldsynthese

Mit dem Ausdruck „virtueller Kopfhörer“ werden in dieser Arbeit Methoden zur Schallwiedergabe bezeichnet, welche die Systemeigenschaften einer Kopfhörerwiedergabe aufweisen, obwohl tatsächlich kein Kopfhörer zur Wiedergabe benutzt wird. Sie sind auch als „transaurale“ Wiedergabe<sup>32</sup> oder „Übersprechkompensation“ bekannt (vgl. Abschn. 2.4).

Ein wichtiges Einsatzgebiet für solche virtuellen Kopfhörer ist die Darbietung von auditiven virtuellen Umgebungen, die mittels Binauraltechnik simuliert werden. In Virtuellen Realitäten, wo ein realer Kopfhörer den Betrachter abschottet und taktil beeinträchtigt, kann eine transaurale Wiedergabe den Grad der Immersion steigern.

### 4.1 Funktionsprinzip

Wie in Kap. 2 bereits angedeutet, besteht bei virtuellen Kopfhörern eine große Herausforderung darin, das System dynamisch an Bewegungen des Hörers – genau genommen an eine sich verändernde Position der Ohren – anzupassen. Dafür muss generell vorausgesetzt werden, dass die Koordinaten des Hörers beziehungsweise seiner Ohren jederzeit durch eine geeignete Sensorik bestimmt werden können.

Sind die Quellenpositionen vorgegeben, wie durch fest installierte Lautsprecher, ändern sich schon mit jeder Kopfdrehung die Übertragungsfunktionen zwischen den Quellen und dem Hörer. Das Netzwerk für die transaurale Wiedergabe, das aus diesen Übertragungsfunktionen berechnet wird, muss dann also entsprechend aktualisiert werden. Problematisch ist dabei weniger die Anforderung einer ausreichend schnellen Reaktion des gesamten Systems, als vielmehr die Bestimmung der aktuell gültigen Übertragungsfunktionen, sei es durch das Abrufen und eventuelles Interpolieren aus einer Datenbank vorhergegangener Messungen oder durch eine entsprechend

<sup>32</sup>Originalzitat: „**Transaural** stereo (generic term) is a stereo-system plan that [...] takes the end point of the recording-reproducing chain to be the actual sounds at the ears.“<sup>[41,S.3]</sup>

genaue Modellierung. Dies gilt insbesondere dann, wenn dem Hörer ein großer Bewegungsradius erlaubt ist.

Der Bewegungsbereich des Hörers kann auch in mehrere Sektoren eingeteilt werden, wobei dann jeder Sektor von einer separaten Lautsprecheranordnung beschallt wird. Beim Wechsel in einen anderen Sektor werden die jeweils zugehörigen Lautsprecher aktiviert, während Positionsänderungen innerhalb eines Sektors durch eine adaptive Übersprechkompensation berücksichtigt werden. Vorteilhaft ist ein solches Vorgehen vor allem dann, wenn sich durch das Ausnutzen von Symmetrien der Aufwand entsprechend verringern lässt vgl. z. B. [101,102].

Ein adaptives System ist nur dann verzichtbar, wenn die Quellen an konstanten *kopfbezogenen* Koordinaten verharren. Dies trifft einerseits zu, falls sich der Hörer tatsächlich nicht bewegt. Das kann beispielsweise dann auftreten, wenn parallel zur Audiowiedergabe ein Bildschirm betrachtet wird. Für spezielle Anwendungen, wie elektronische Informations- und Werbeanzeigetafeln oder an Computerarbeitsplätzen könnte also sogar ein vollständig statisches System in Betracht gezogen werden.

Andererseits besteht eine zweite Möglichkeit zur Vermeidung einer adaptiven Übersprechkompensation darin, dass sich die Quellen synchron zum Hörer bewegen. Ein mechanisches Mitführen von Lautsprechern ist allerdings kaum mit vertretbarem Aufwand realisierbar. Ist der Hörer jedoch in den Freiheitsgraden seiner Bewegung ausreichend eingeschränkt, wäre vielleicht ein Durchschalten von mehreren passend positionierten Lautsprechern denkbar.

Löst man sich von der Vorstellung einer konkreten Realisierung der Quellen, wie in den vorhergehenden Fällen als Lautsprecher, kann der Ansatz des Mitbewegens auf eine andere Weise weiter verfolgt werden:

Methoden der Virtuellen Akustik erlauben die Erstellung von „virtuellen Quellen“. Das sind Schallquellen, die physikalisch nicht vorhanden sind, aber deren Ohrsignale entweder durch ein binauraltechnisches Verfahren oder eine Schallfeldrekonstruktion den Ohrsignalen einer real existierenden Quelle entsprechen und die damit auch so wahrgenommen werden. Die zur Erzeugung benötigten und tatsächlich vorhandenen Schallquellen werden in diesem Kontext als „sekundäre Quellen“ bezeichnet. Der Vorteil virtueller Quellen besteht darin, dass sie weitgehend frei platziert und bewegt werden können.

Solche virtuellen Quellen lassen sich auch für einen virtuellen Kopfhörer nutzen. Mittels Schallfeldrekonstruktion können sozusagen virtuelle Lautsprecher erzeugt und bei allen Bewegungen des Hörers so mitgeführt werden, dass sich ihre Position im kopfbezogenen Koordinatensystem nicht verändert. Über diese virtuellen Quellen

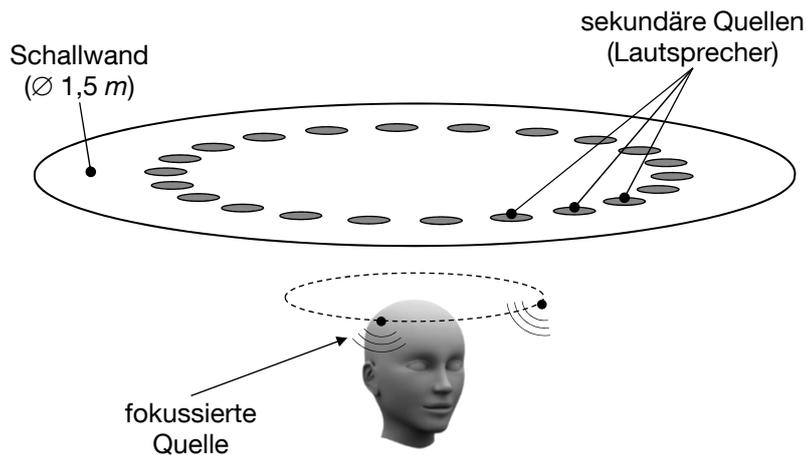
kann dann eine statische Übersprechkompensation betrieben werden. So erhält man einen virtuellen Kopfhörer, der dynamisch auf die Bewegungen des Hörers durch das Verschieben der Quellen reagiert, statt durch Adaption der Übertragungsfunktionen. Da die Schallfeldrekonstruktion mathematisch beschrieben werden kann (vgl. Kap. 3), lässt sich das Generieren und Bewegen von virtuellen Quellen verhältnismäßig einfach – und im Gegensatz zu binauraltechnischen Verfahren ohne die Ermittlung von kopfbezogenen oder raumbezogenen Übertragungsfunktionen, zum Beispiel durch aufwendige elektro-akustische Messreihen – implementieren.

Zwei Patente für virtuelle Kopfhörer, die auf einem derartigen Funktionsprinzip beruhen, wurden 2009 in Deutschland<sup>[221]</sup> und 2012 in den USA<sup>[95]</sup> erteilt. In beiden Fällen wird eine Wiedergabevorrichtung für binaurale Signale geschützt, bei der eine Übersprechkompensation über virtuelle Quellen erfolgt. In den Ausführungsbeispielen wird zur Erzeugung der virtuellen Quellen jeweils explizit auf die Methode der Wellenfeldsynthese zurück gegriffen. Im älteren Patent wird eine kreisförmige Lautsprecheranordnung oberhalb des Hörers beschrieben, die sich besonders für die Berücksichtigung von Kopfdrehungen eignet. Im Unterschied dazu wird im jüngeren Patent mit ebenen Wellen, also sehr weit entfernten virtuellen Quellen, gearbeitet, um Translationen des Hörers auszugleichen. Bei Kopfrotationen ist dann aber nach wie vor eine adaptive Übersprechkompensation notwendig. Genau genommen ist damit die Idee, die virtuellen Quellen bei allen Bewegungen des Hörers mitzuführen, um ausschließlich eine statische Übersprechkompensation einsetzen zu können, nur im älteren Patent konsequent umgesetzt. Dies wird im folgenden Abschnitt 4.2 genauer erläutert.

Es existiert auch ein Vorschlag, einen solchen virtuellen Kopfhörer mit Phantomschallquellen zu realisieren<sup>[199]</sup>. Da jedoch Ohrsignale von Phantomschallquellen und virtuellen Schallquellen, auch wenn sie zu identischen Hörereignissen führen, grundsätzlich verschieden sind<sup>[189,193]</sup>, ist die Anwendung der auf kopfbezogenen Übertragungsfunktionen beruhenden Übersprechkompensation in Kombination mit Phantomschallquellen theoretisch nicht begründbar. Es besteht zwar die Möglichkeit, dass sich der gewünschte Effekt perzeptiv trotzdem einstellt. Allerdings wird dies durch die Evaluierung des implementierten Laborsystems nicht bestätigt<sup>[199]</sup>.

## 4.2 Prototyp

Ein Prototyp eines virtuellen Kopfhörers, der mit synchron zum Hörer bewegten virtuellen Quellen arbeitet und auf dem das deutsche Patent<sup>[221]</sup> beruht, wurde am Institut für Rundfunktechnik (kurz: IRT) entwickelt und 2005 unter dem Namen

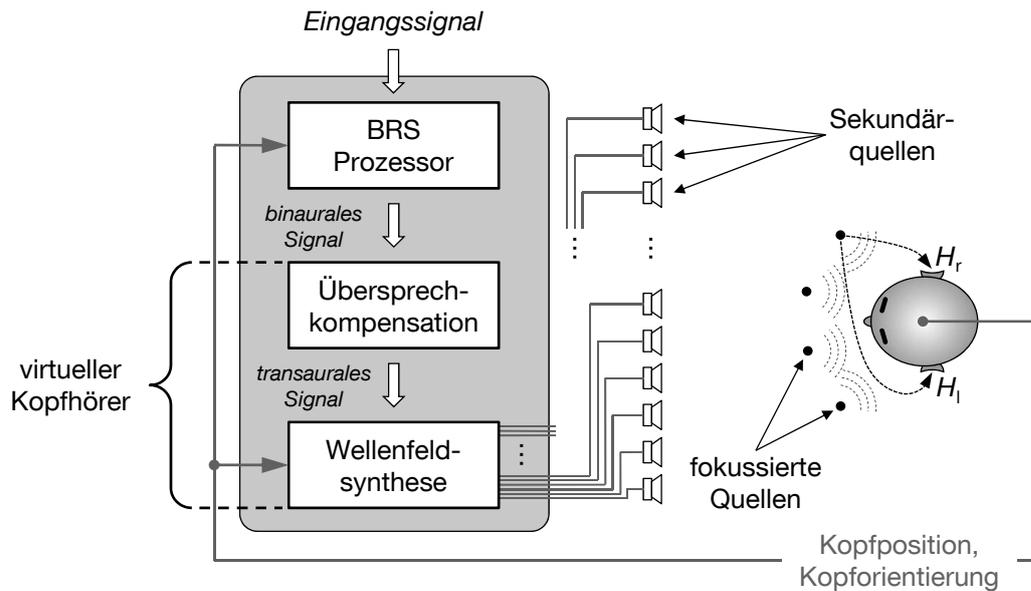


**Abbildung 4.1**  
Schematische  
Darstellung des  
Prototyps

„Binaural Sky“ erstmals vorgestellt<sup>[123]</sup>. Er ist der Ausgangspunkt für die im weiteren Verlauf beschriebenen Forschungsarbeiten.

Das System besteht aus einer kreisrunden Schallwand mit einem Durchmesser von 1,5 m, in die 22 Lautsprecherchassis, äquidistant auf einem konzentrischen Ring von 1 m Durchmesser verteilt, als Sekundärquellen eingelassen sind. Die Anordnung ist 0,4 m über dem Kopf des Hörers befestigt, so dass sich keine Lautsprecher in dessen direktem Blickfeld befinden. Die virtuellen Quellen zur Wiedergabe des transauralen Signals werden – angelehnt an die Prinzipien der Wellenfeldsynthese – als „fokussierte Quellen“ oberhalb der Ohren erzeugt. Die Anordnung ist schematisch in Abb. 4.1 dargestellt. Genau betrachtet eignet sich ein derart aufgebautes System nicht für die Wellenfeldsynthese, da es die geometrischen Grundannahmen nicht berücksichtigt. Die Lautsprecher müssten sich entweder in der Horizontalebene des kopfbezogenen Koordinatensystems – also auf Ohrhöhe – befinden, oder über die gesamte Fläche oberhalb des Hörers verteilt sein. Der Prototyp arbeitet daher in der hier beschriebenen Ausführung eher nach dem Prinzip der Zeitumkehr denn mit Wellenfeldsynthese (vgl. Abschn. 3.3 und 3.4).

Die Apparatur wurde für die Wiedergabe von binauralen Signalen des BRS-Systems (vgl. Abschn. 2.3) konzipiert. Sie ist daher speziell für die Berücksichtigung von Kopfdrehungen optimiert, was den rotationssymmetrischen Aufbau begründet. Ein Sensor ermittelt den aktuellen Drehwinkel des Kopfes und passt dementsprechend die Fokuspunkte der virtuellen Quellen an. Somit bewegen sich diese, wie in Abb. 4.1 angedeutet, synchron zum Hörer auf einer Kreisbahn oberhalb der Ohren. Das über diese virtuellen Quellen wiedergegebene transaurale Signal wird durch ein statisches Übersprechkompensationsfilter erzeugt, das vorab aus den dann jeweils konstanten kopfbezogenen Übertragungsfunktionen aller virtuellen Quellen berechnet wurde. Die Konstellation der virtuellen Quellen, das heißt ihre Anzahl und ihre genaue Lage



**Abbildung 4.2** Blockschaltbild des Prototyps. Der BRS-Prozessor (vgl. Abschn. 2.3) generiert ein binaurales Signal, das dem Hörer über einen virtuellen Kopfhörer dargeboten wird. Dieser basiert auf Wellenfeldsynthese (vgl. Abschn. 3.3), mit der, abhängig von Kopforientierung und Kopfposition, fokussierte Quellen (vgl. Abschn. 3.4) an konstanten kopfbezogenen Koordinaten erzeugt werden, die sich synchron zum Hörer bewegen. Dadurch kann ein statisches Filter für die Übersprechkompensation (vgl. Abschn. 2.4) eingesetzt werden, das aus den konstanten kopfbezogenen Übertragungsfunktionen der fokussierten Quellen – eine ist beispielhaft als  $H_l$  und  $H_r$  eingetragen – berechnet wird.

im kopfbezogenen Koordinatensystem, ist variabel und nicht explizit spezifiziert. Sie war daher Gegenstand einer später beschriebenen Untersuchung (vgl. Abschn. 6.5).

Eine Änderung der Kopfposition wird allerdings nicht kompensiert, weil sie vom BRS-System ebenfalls nicht berücksichtigt wird. Wenn der Hörer die vorgesehene Abhörposition unter dem Mittelpunkt des Lautsprecherrings verlässt, folgen ihm die virtuellen Quellen dabei nicht, da auch die binaurale Simulation nicht angepasst werden kann. Der Bewegungsradius ist also auf einen einzigen Punkt beschränkt. Durch das Mitbewegen der virtuellen Quellen bei einer Translation des Hörers kann der Arbeitsbereich des virtuellen Kopfhörers aber – unabhängig von der eingesetzten Binauraltechnik – erheblich erweitert werden, wie in Abschn. 4.3 noch gezeigt wird.

### 4.3 Konzeptionelle und technische Weiterentwicklung des Prototyps

Der in Abschn. 4.2 vorgestellte Prototyp konnte schrittweise verbessert werden. Da er für die schnelle Verifizierung des Konzepts zu Beginn nur auf vorberechnete und an die festgelegte Geometrie angepasste Filtersätze zurückgriff, wurde schließlich ein modulares Signalverarbeitungskonzept realisiert, das eine flexible Anpassung an unterschiedliche Vorgaben erlaubt.

#### Ausgangspunkt

Ähnlich wie bei der Binauralen Raumsynthese (vgl. Abschn. 2.3) wird durch die Beschränkung auf einen vorgegebenen Abhörpunkt eine einfache Implementierung ermöglicht. Dazu berechnet man die Ansteuerungsfunktionen der Wellenfeldsynthese für die erlaubten Kopforientierungen vorab, fasst sie mit der statischen Übersprechkompensation zusammen und speichert dieses Ergebnis jeweils als Filtersatz mit endlicher Impulsantwort<sup>33</sup> in einer Datenbank ab. Bei der Wiedergabe erfolgt der Zugriff auf die Datenbank abhängig von der Ausrichtung des Kopfes. Dadurch stehen mit nur einem Signalverarbeitungsschritt alle benötigten Ausgangssignale für die einzelnen Lautsprecher zur Verfügung.

Der offensichtliche Nachteil dieser Vorgehensweise liegt darin, dass so das System vollständig festgelegt ist. Jede Modifikation an einem der beiden Teilsysteme erfordert eine Neuberechnung aller Filter. Außerdem ist auf dieser Basis die Erweiterung des Bewegungsbereichs nicht sinnvoll, da man für jede mögliche Kombination aus Kopfposition und Kopforientierung einen separaten Filtersatz vorhalten müsste.

#### Kompensation von Laufzeitunterschieden

Eine in der Theorie unsaubere aber in der Praxis gut funktionierende Erweiterung des Prototyps wurde durch die adaptive Nachjustierung der Verzögerungszeiten bei einer Translation des Hörers erreicht. Bedingt durch die rotationssymmetrische Anordnung ist die Laufzeit des Schalls von jedem Lautsprecher zur vorgesehenen Abhörposition in der Mitte gleich. Wenn diese Abhörposition verschoben wird, entstehen Laufzeitdifferenzen, die sich durch einfache geometrische Überlegungen kompensieren lassen.

Messreihen und informelle Hörtests bestätigten, dass dieses pragmatische Vorgehen

---

<sup>33</sup>engl.: finite impulse response (kurz: FIR)

zum Erfolg führt. Die kopfbezogene Übertragungsfunktion einer Fokusquelle muss im Weltkoordinatensystem also weitgehend ortsunabhängig sein. Gleichzeitig ist damit belegt, dass das Einhalten der Rotationssymmetrie bei der Wiedergabe nicht erforderlich ist. Wenn sich der Hörer nur nach vorne oder hinten bewegt, bleibt die Anordnung aber bei gerader Kopfausrichtung zumindest symmetrisch zur Medianebene. Ähnlich wie bei lautsprecherbasierter Übersprechkompensation zeigte sich das System dann stabiler (vgl. Abschn. 2.4).

## **Dynamische Wellenfeldsynthese**

Das Ziel der Weiterentwicklung war jedoch die klare Trennung der beiden Funktionsblöcke Wellenfeldsynthese und Übersprechkompensation (vgl. Abb. 4.2). Nur so kann der Vorteil, dass sich die fokussierten Quellen praktisch frei platzieren und bewegen lassen, ausgenutzt und die ursprüngliche Idee vollständig umgesetzt werden.

Wie in Abschn. 3.3 erläutert, beruht Wellenfeldsynthese im Wesentlichen auf einer individuellen Verstärkung und Verzögerung der einzelnen Lautsprechersignale sowie einem vorgelagerten frequenzabhängigen Filter. Vereinfacht beschrieben benötigt man ein Teilsystem, das abhängig von der Position und Ausrichtung des Hörers alle Kanäle so verstärkt und verzögert, dass der Schall aus den Lautsprechern gleichzeitig und gleich stark an den gewünschten Fokuspunkten im kopfbezogenen Koordinatensystem eintrifft. Bei einer Bewegung des Hörers müssen beide Parameter artefaktfrei nachjustiert werden.

Dazu wurde ein Computerprogramm entworfen, implementiert und getestet, das die aktuell erforderlichen Verstärkungsfaktoren und Verzögerungszeiten für jeden Kanal anhand von Abstandsberechnungen ermittelt. Dann verzweigt und bearbeitet es das Eingangssignal entsprechend und stellt schließlich der Audioschnittstelle alle Ausgangssignale zur Wiedergabe bereit. Die Verstärkung ist durch eine einfache Multiplikation, die Verzögerung auf Basis eines modifizierten Ringpuffers umgesetzt. Das frequenzabhängige Vorfilter ist mit Hilfe der Schnellen Faltung<sup>34</sup> realisiert, wofür auf die am Massachusetts Institute of Technology entwickelte und frei verfügbare Unterprogramm-Bibliothek FFTW<sup>35</sup> zurückgegriffen wird. Bei einer Änderung der Verzögerungszeit werden die ursprünglich gültigen und die neu berechneten Ausgangssignale überblendet, um hörbare Störungen, die sonst durch eine Unstetigkeit im Signal entstehen können, zu unterdrücken.

---

<sup>34</sup>*engl.*: fast fourier transformation (kurz: FFT)

<sup>35</sup>*engl.*: fastest fourier transform in the west (kurz: FFTW)

## 4.4 Geräte und Systemkomponenten

Ein Großteil der Komponenten des Prototyps wurde in verschiedenen Experimenten wiederverwendet. Sie sind zusammen mit weiteren benutzten Gerätschaften im Folgenden aufgeführt und kurz beschrieben.

Der Lautsprecherring des Prototyps besteht aus Breitbandlautsprechern, die auch vom Hersteller zur Verwendung als Deckenlautsprecher vorgeschlagen werden (Visaton FRS 8). Sie weisen bei hohen Frequenzen allerdings eine deutliche Richtwirkung auf. Dies muss berücksichtigt werden, da sich der Hörer nicht im Bereich der Hauptachse der Lautsprecher bewegt. Zusätzlich ist, was in Abb. 4.1 nicht dargestellt ist, zur Klangverbesserung im Kreismittelpunkt ein zusätzlicher Tieftöner platziert, der die Reproduktion von Frequenzen unter 120 Hz übernimmt (Visaton W 200 S).

Alle Lautsprecher werden separat angesteuert. Die Signale werden über eine Audio-schnittstelle (RME Hammerfall DSP 9652, Digiface) digital ausgegeben, dann von zwei D/A-Wandlern (Creamware A16 Ultra) umgesetzt und von sechs baugleichen Endstufen mit je vier Kanälen (TerraTec-Producer SINE PA 460) verstärkt. Insgesamt stehen so maximal 24 diskrete Ausgangskanäle zur Verfügung, der Prototyp nutzt 23 davon. Eine A/D-Schnittstelle erlaubt die Zuspieldung von bis zu acht analogen Eingangskanälen (RME ADI 8-Pro). Alle digitalen Signale sind, sofern es nicht anders angegeben wird, im Zeitbereich mit 48 kHz abgetastet und mit 16 bit quantisiert. Im Verbund waren alle digitalen Audiogeräte durch einen externen Taktsignalgeber synchronisiert (STRAMP VCC; Rosendahl nanosyns).

Die Audiosignalverarbeitung erfolgt auf einem handelsüblichen Rechner mit einer aktuellen Linux-Distribution als Betriebssystem. Die Entscheidung für Linux beruht darauf, dass es zum Zeitpunkt der Entwicklung des Prototyps zusammen mit den verfügbaren Anwenderprogrammen eine sehr gute Basis für die Signalverarbeitung in Echtzeit bot. Im Verlauf der Forschungsarbeiten wurden Hard- und Software immer wieder aktualisiert: Der Prototyp arbeitete anfangs nur mit vorberechneten Filtern für die Kreisanordnung, die abhängig von der Kopfdrehung ausgetauscht wurden. Später war das System in der Lage, die Anregungssignale der Wellenfeldsynthese für unterschiedliche Sekundärquellenaufstellungen zur Laufzeit zu berechnen und zu modifizieren, so dass virtuelle Quellen tatsächlich flexibel platziert und bewegt werden konnten (vgl. Abschn. 4.3).

Zur Kopfverfolgung<sup>36</sup>, also zur Erfassung von Kopfposition und Kopforientierung in allen sechs Freiheitsgraden, wird ein elektromagnetisches System (Polhemus FASTRAK) eingesetzt. Es bietet eine sehr hohe Genauigkeit bei äußerst geringer

---

<sup>36</sup>engl.: head tracking

Latenz, ist jedoch empfindlich gegen Feldstörungen durch ferromagnetische Objekte, wie zum Beispiel Lautsprechermagnete. Außerdem ist die Befestigung eines Sensors am verfolgten Objekt erforderlich, wodurch der Hörer, wie im vorliegenden Fall durch einen Haarreif und das Anschlusskabel, gestört wird. Die ideale Lösung wäre ein berührungsloses System, das beispielsweise kamerabasiert oder laserbasiert arbeitet.

Abgesehen von der Schallwand mit den eingebauten Breitbandchassis wurden als Sekundärquellen auch geschlossene Lautsprecherboxen benutzt (Elac 301). Dieses 2-Wege System hatte sich bereits in anderen Experimenten zur Wellenfeldsynthese bewährt, genauso wie mehrere Lautsprecher mit würfelförmigem Metallgehäuse von etwa 10 cm Kantenlänge, mit denen ein Lokalisationsexperiment durchgeführt wurde (Studer, ohne Modellbezeichnung).

Lautsprecher mit möglichst geringen Abmessungen wurden für die Versuche mit sehr nahe am Kopf positionierten Schallquellen benötigt (vgl. Abschn. 6.5). Ein Fabrikant stellte dazu mehrere leistungsfähige Kleinlautsprecher mit einem Membrandurchmesser von etwa 3,5 cm zur Verfügung (Philips, ohne Modellbezeichnung). Sie erhielten jeweils ein zylindrisches Kunststoffgehäuse von 6 cm Durchmesser und 7 cm Länge mit abgerundeten Kanten an der Vorderseite. Solche selbstgebaute Lautsprecherboxen wurden sowohl bei einer Messreihe im reflexionsarmen Raum als auch in einem Hörversuchsaufbau benutzt. Die Ansteuerung erfolgte im ersten Fall über eine der bereits erwähnten 4-Kanal-Endstufen, im zweiten Fall über einen Verteilungsverstärker mit 12 Kanälen (Sonace Sonamp 1230).

Ohrsignale wurden generell mit einem Kunstkopf aufgenommen (Neumann KU 100). Für Schallsignale an diskreten Messpunkten kamen entweder Kondensatormikrofone (Schoeps CCM 2 S) oder Messmikrofonkapseln mit zugehörigem Impedanzwandler zum Einsatz (Brüel & Kjær 4165, 4190 mit 2269). Zusätzlich enthielten alle Messketten rauscharme Vorverstärker (Neumann PMV 70; img Stage Line MPA-102).

Bei Kopfhörerwiedergabe wurde auf hochwertige Modelle verschiedener Hersteller zurückgegriffen (STAX SR  $\lambda$  pro; Beyerdynamic DT 831; AKG 240 DF; Sennheiser HD 650), die als geeignet für die Binauraltechnik gelten<sup>vgl. [178]</sup>, und auch oft in der Hörfunkproduktion Verwendung finden.

Übertragungsfunktionen wurden größtenteils PC-gestützt<sup>37</sup> ermittelt. Als Anregungsfunktion diente ein logarithmisches Wobbelsignal, das über eine Audioschnittstelle (RME DIGI 96/8 PAD, Fireface 800) an das Wiedergabegerät ausgegeben wurde, je nach Anwendungsfall digital oder analog. Zusammen mit der gleichzeitig aufgenommenen Systemantwort lässt sich daraus die Impulsantwort beziehungsweise die

---

<sup>37</sup> engl.: personal computer (kurz: PC)

Übertragungsfunktion ableiten. Teilweise wurde ein für solche Zwecke spezialisiertes Messgerät benutzt, das nach einem ähnlichen Prinzip, allerdings auf Basis von Maximalfolgen, arbeitet (Audio Precision System Two Cascade 2522).

# 5

## Psychoakustische Untersuchungen am virtuellen Kopfhörer

Im Rahmen der Forschungsarbeiten am weiterentwickelten virtuellen Kopfhörer dienten verschiedene psychoakustische Untersuchungen dazu, das in Kap. 4 beschriebene Funktionsprinzip zu untermauern. Am Prototyp wurde in einem ersten anwendungsorientierten Schritt die Wahrnehmung binauraler Signale unter dem Aspekt Lokalisation getestet. Die weiteren Versuche sind als grundlagenorientiert einzuordnen. Sie vermeiden den für Lokalisationsversuche notwendigen binaural-technischen Verarbeitungsschritt und konzentrieren sich stattdessen auf die für Kopfhörerdarbietung charakteristischen Phänomene Lateralisation und Im-Kopf-Lokalisation sowie auf die Wahrnehmbarkeit des interauralen Kohärenzgrads. Die Ergebnisse sind in Kap. 7 zusammengestellt.

### 5.1 Lokalisation von binauralen Signalen

Der Prototyp des virtuellen Kopfhörers ist auf die Darbietung von binauralen Signalen des BRS-Systems (sh. Abschn. 2.3) ausgelegt. Es liegt also nahe, die Funktionsfähigkeit genau in dieser Kombination zu beurteilen. Darüber hinaus ist es ein allgemein übliches Vorgehen, Systeme zur Wiedergabe binauraler Signale über Lautsprecher mit Lokalisationsversuchen zu testen. Das Augenmerk liegt dabei – teilweise sogar ausschließlich – auf der Wahrnehmung seitlicher Schalleinfallrichtungen, also auf der korrekten Darstellung des Azimuts<sup>vgl. z. B. [9,46,48,58,75,91] [97,Manuscript E] [101,161,162,184,199]</sup>.

In vielen dieser Forschungsarbeiten wird dazu auf eine statische binaurale Synthese zurückgegriffen. Die Untersuchungen im Rahmen der Entwicklung des BRS-Systems zeigten jedoch, dass bei Kopfhörerdarbietung eine dynamische binaurale Synthese, die Kopfdrehungen des Hörers berücksichtigt, einer statischen binauralen Synthese, die lediglich mit einer nicht personalisierten kopfbezogenen Übertragungsfunktion arbeitet, besonders in Bezug auf die Richtungsdarstellung deutlich überlegen ist<sup>[99,111]</sup> (vgl. Abschn. 2.1). Versuche an der RWTH Aachen<sup>38</sup> bestätigten dieses Ergebnis später

<sup>38</sup>Rheinisch-Westfälische Technische Hochschule Aachen (kurz: RWTH Aachen)

auch explizit für die übersprechkompensierte Wiedergabe binauraler Signale<sup>[101]</sup>.

Bei dem im Folgenden beschriebenen Versuch war das System daher in der Lage, auf Kopfdrehungen des Hörers zu reagieren und sowohl die binaurale Synthese als auch die Übersprechkompensation entsprechend nachzujustieren. Zudem sollte neben der Wahrnehmung des Azimuts auch die Wahrnehmung der Elevation überprüft werden<sup>[123,124]</sup>.

### Versuchsaufbau

Dafür wurden sechs reale Schallquellen – Lautsprecher mit einem würfelförmigem Metallgehäuse von etwa 10 cm Kantenlänge (Studer, ohne Modellbezeichnung) – um den vorgesehenen Abhörpunkt des Prototyps, der etwa 40 cm mittig unterhalb der Schallwand mit den Lautsprechern liegt (sh. Abschn. 4.2), verteilt. In der Startposition befanden sie sich aus Sicht der Versuchsperson an den folgenden kopfbezogenen Koordinaten, die Nummerierung ergibt sich aus dem Uhrzeigersinn:

Nr.	$\alpha$	$\varepsilon$	$r$
1	0°	0°	1 m
2	-60°	+25°	1 m
3	-135°	-25°	1 m
4	180°	0°	1 m
5	110°	+10°	1 m
6	30°	-10°	1 m

Diese Referenzsituation mit allen sechs Quellen wurde dann in einer BRS-Messung vollständig erfasst (sh. Abschn. 2.3). Sie lieferte eine Datenbank mit binauralen Raumimpulsantworten, die eine – bezüglich Kopfdrehungen – dynamische binaurale Synthese erlaubte.

Um die Versuchsperson war ein zylinderförmiger Sichtschutz aus einem akustisch durchlässigen aber optisch intransparenten Material angebracht, der alle Lautsprecher, sowohl die Schallwand des virtuellen Kopfhörers als auch die zu lokalisierenden Kleinlautsprecher, verdeckte. Damit werden kreuzmodale Effekte durch den visuellen Kanal verhindert, es waren lediglich Anhaltspunkte zur Orientierung bei  $\alpha = 0^\circ$  und  $\pm 90^\circ$  sowie auf Höhe der vorgesehenen Horizontalebene, also bei  $\varepsilon = 0^\circ$  in regelmäßigen Abständen, auf der Innenseite des Vorhangs markiert.

## Stimuli

Der Grundstimulus besteht aus Rosa Rauschpulsen von 300 *ms* Dauer, die durch ebenfalls 300 *ms* lange Pausen unterbrochen sind. Jeder Puls wird mit einer über 20 *ms* linear ansteigenden Flanke eingeleitet und mit einer über 20 *ms* linear abfallenden Flanke beendet. Die Rauschpulse wurden bis zur endgültigen Bewertung durch den Versuchsteilnehmer über die jeweils aktive reale oder virtuelle Schallquelle in einer Endlosschleife dargeboten. Während des Versuchs lag der bewertete Schalldruckpegel aller Proben bei etwa 60 *dB(A)* am Abhörpunkt.

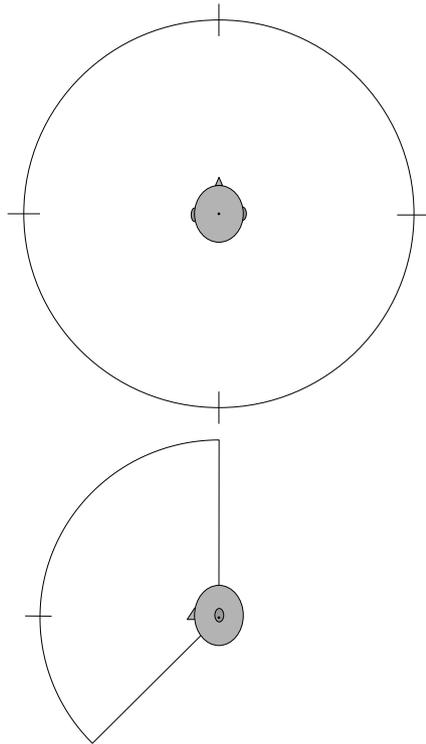
Die Wahl des Testsignals begründet sich wie folgt: Bei breitbandigen Signalen – wie Rauschen – zeigen sich insgesamt bessere Ergebnisse bei der Zuordnung zwischen wahrgenommener und tatsächlicher Schalleinfallrichtung<sup>[28,S.95]</sup>. Rosa Rauschen im Besonderen wird wegen der Ähnlichkeit seines Spektrums zu dem von Sprache und Musik eine höhere Praxisrelevanz zugeschrieben<sup>[51,S.44] [210,S.12f.]</sup>. Die Pulsierung soll dem Gehör schließlich zusätzliche zeitliche Informationen zur Auswertung liefern<sup>[157,S.97]</sup>. Die Flanke in der Hüllkurve verringert hörbare Artefakte, wobei eine gaußförmige Fensterung zu bevorzugen wäre<sup>[23,Abschn.1.3.2] [54,Abschn.1.1]</sup>. Ähnliche Signale, also gepulste Rauschen mit vergleichbaren zeitlichen Parametern, wurden auch zur Evaluierung verwandter Systeme zur Wiedergabe binauraler Signale über Lautsprecher mit Lokalisationsversuchen beispielsweise am Massachusetts Institute of Technology<sup>[58]</sup>, an der RWTH Aachen<sup>[101]</sup> oder an der Technischen Universität München<sup>[199]</sup> verwendet.

## Versuchsablauf

Zuerst setzte sich die Versuchsperson auf einen vorbereiteten Drehstuhl und nahm darauf mit Hilfe des Versuchsleiters die vorgesehene Abhörposition ein. Danach bekam sie den Zweck des Experiments und die Anzeigemethode erklärt. Dabei wurden Kopfdrehungen zur genaueren Lokalisation der Schalle ausdrücklich erlaubt, jedoch erfolgte auch die Anweisung, die Abhörposition einzuhalten und sich nicht um mehr als 90° zur Seite zu drehen. Zusätzlich schaltete der virtuelle Kopfhörer automatisch stumm, wenn sich ein Proband um mehr als 5 *cm* wegbewegte.

Der Teilnehmer markierte den gehörten Seitenwinkel und den gehörten Höhenwinkel in einer Schablone, wie sie in Abb. 5.1 dargestellt ist. Die Eignung einer solchen grafischen Anzeigemethode ist seit langem bestätigt<sup>[187]</sup>. Abb. 5.1 ist maßstabsgetreu, im Original war sie entsprechend vergrößert auf einem DIN A4-Blatt vorgedruckt. Für jeden Testschall wurde ein separater Bewertungsbogen ausgegeben.

Reale und BRS-simulierte Schallquellen wurden – unter anderem aus technischen



**Abbildung 5.1**

Vorgedruckte Skizze zur Bewertung des gehörten Seitenwinkels (Azimut) und des gehörten Höhenwinkels (Elevation) von BRS-simulierten Quellen bei Wiedergabe über einen virtuellen Kopfhörer (Prototyp).

Gründen – in zwei getrennten Durchgängen untersucht. Pro Durchgang wurde jede der sechs Quellen dreimal angesteuert. Die Reihenfolge war dabei zufällig, aber mit der Vorgabe, dass keine Quelle zweimal direkt nacheinander auftritt. Zu Beginn des Versuchs wurde noch ein verdecktes Training durchgeführt, indem die Versuchsperson alle Testschalle in zufälliger Reihenfolge bereits einmal bewertete, diese Ergebnisse später aber nicht berücksichtigt wurden. Weil die Teilnehmer die Anzeigemethode intuitiv verstanden, war kein längeres Training notwendig. So hatte eine Versuchsperson in jedem Durchgang 24 Testschalle zu beurteilen, was im Durchschnitt etwa 10 Minuten dauerte.

### **Versuchspersonen**

15 Personen im Alter von 23 bis 61 Jahren absolvierten die beiden Durchgänge des Lokalisationsversuchs. Alle Probanden wiesen ein altersgemäßes Hörvermögen auf und 13 von ihnen hatten Erfahrung mit Hörversuchen.

## 5.2 Lateralisation

Wird über Kopfhörer ein Signal dargeboten, das abgesehen von Laufzeitdifferenzen oder Pegelunterschieden zwischen beiden Kopfhörerkanälen<sup>39</sup> identisch ist, entsteht ein Hörereignis im Kopf, das sich auf der interauralen Achse zur lautereren oder früheren Seite hin verschiebt. Dieses lange bekannte und studierte Phänomen wird als „Lateralisation“ bezeichnet<sup>[23,S.111 ff.]</sup>. Im Gegensatz zur Lokalisation, die üblicherweise dreidimensional verstanden wird, erfasst man Lateralisation definitionsgemäß nur entlang einer Dimension, nämlich als seitliche Auslenkung des Hörereignisses projiziert auf die interaurale Achse<sup>[23,S.112][54,S.308]</sup>.

Bei diotischer Wiedergabe, also vollständig identischen Signalen in beiden Kopfhörerkanälen, liegt das Hörereignis in der Medianebene<sup>[23,S.115&126]</sup>. Bei Pegeldifferenzen oder Laufzeitdifferenzen wandert es gleichmäßig in Richtung des Kopfhörerkanals mit dem höheren Pegel oder dem zuerst auftretenden Signal. Ab Unterschieden von etwa 30 dB oder 1 ms ist das Hörereignis maximal ausgelenkt und wird vollständig auf der lautereren oder früheren Seite wahrgenommen. Im Detail stellen sich diese Effekte als frequenzabhängig dar und beeinflussen sich gegenseitig<sup>[23,S.111 ff.][54,S.308 f.]</sup>. Sie gelten allgemein als Indikator dafür, welche interauralen Signalmerkmale das Gehör für das Richtungshören auswertet<sup>[23,S.112]</sup>, wobei jedoch auch Einwände gegen eine solche Annahme bestehen<sup>[142,189]</sup>.

Lateralisation ist ein charakteristisches Phänomen der Kopfhörerwiedergabe. Daher wurde untersucht, ob sie bei dem Prototyp des virtuellen Kopfhörers (sh. Abschn. 4.2) ebenfalls auftritt und ob dabei Unterschiede im Vergleich zu einem normalen Kopfhörer zu beobachten sind. Im Gegensatz zur üblichen Definition wurde die Abfrage des Hörereignisortes in drei Dimensionen, *transversal* (links – rechts), *sagittal* (vorne – hinten) und *longitudinal* (oben – unten), vorgenommen (vgl. Abb. 5.2).

### Versuchsaufbau

Der Hörversuch lief rechnergestützt ab, wobei die Steuerung von der Versuchsperson selbst übernommen wurde. Dementsprechend stand ihr ein Computerarbeitsplatz zur Verfügung. Die Schallwand des virtuellen Kopfhörers war so über dem Sitzplatz abgehängt, dass der vorgesehene Abhörpunkt eingehalten werden konnte. Zusätzlich lag ein realer Kopfhörer (Beyerdynamic DT 831) bereit, der an einen Kopfhörerverstärker (Samson C-que 8) angeschlossen war. Da sich weder Lautsprecher noch andere Schallquellen im Blickfeld der Versuchsperson befanden, wurden keine Sicht-

---

<sup>39</sup>im Weiteren vereinfacht als „interaurale Pegel-/Zeitdifferenz“ bezeichnet; *engl.*: interaural level (auch: intensity)/time difference (kurz: ILD, IID/ITD)

schutzmaßnahmen vorgenommen. Beide Kopfhörer, der reale wie der virtuelle, waren diffusfeldentzerrt (sh. Abschn. 2.2) und wiesen damit identische Übertragungseigenschaften auf. Digitale Signale wurden im gesamten Versuchsaufbau in CD-Qualität<sup>40</sup>, also mit einer Abtastrate von 44,1 kHz verarbeitet.

## Stimuli

Der Versuch wurde mit vier verschiedenen Grundsignalen durchgeführt: ein Sinuston, zwei schmalbandige und ein breitbandiges Rauschen.

Nr.	Signalform	Frequenzbereich (Hz)
1	Sinuston	1000
2	Schmalbandrauschen	920–1080
3	Schmalbandrauschen	6400–7700
4	Rosa Rauschen	200–10000

Der Sinuston bei 1 kHz ist ein klassisches Referenzsignal der Psychoakustik und erlaubt den direkten Vergleich mit Ergebnissen in der Literatur<sup>vgl. [54, S. 308]</sup>. Die Schmalbandrauschen sind jeweils frequenzgruppenbreit mit einer Mittenfrequenz von 1 kHz beziehungsweise 7 kHz<sup>vgl. [54, Table 6.1]</sup>, um zwei verschiedene Frequenzbereiche selektiv zu testen. Als Breitbandsignal wurde wieder Rosa Rauschen gewählt, das aufgrund der Messergebnisse am Prototyp<sup>[123]</sup> zwischen 200 Hz und 10 kHz bandbegrenzt war. Es kann hier vorweg genommen werden, dass bei der Auswertung (vgl. Abb. 7.6) zwischen den einzelnen Grundsignalen jedoch keine signifikanten Unterschiede auftraten.

Diese vier Grundsignale sollten den Versuchspersonen mit Pegeldifferenzen zwischen  $\pm 30$  dB in 10 dB-Schritten dargeboten werden. Um gleichmäßig laute Testsignale zu erhalten, wurden die Pegel in den beiden Kanälen –  $L_1$  für links und  $L_2$  für rechts – in Anlehnung an eine 6 dB-Summenregel der Form<sup>[158]</sup>

$$L = 6 \cdot \log_2 \left( 2^{L_1/6} + 2^{L_2/6} \right)$$

gemäß der folgenden Übersicht eingestellt:

	digitaler Pegel (dB FS)							
$L_1$	86	85	84	80	74	65	56	
$L_2$	56	65	74	80	84	85	86	
$L_2 - L_1 = \Delta L$	-30	-20	-10	0	10	20	30	

<sup>40</sup>engl.: compact disc (kurz: CD)

Die Länge eines einzelnen Testsignals wurde nach Probeläufen auf zwei Sekunden festgelegt, die Signalflanken waren mit einer Anstiegszeit und einer Abfallzeit von jeweils 30 ms gaußmoduliert.

Die Verstärker wurden so eingeregelt, dass bei Beschallung mit dem virtuellen Kopfhörer am Abhörpunkt ohne Versuchsperson ein bewerteter Schalldruckpegel von etwa 70 dB(A) herrschte. Die Verstärkung des realen Kopfhörers wurde daran mit Hilfe eines Kunstkopfs angeglichen.

## Versuchsablauf

Alle Teilnehmer führten den Lateralisationsversuch zuerst mit dem realen Kopfhörer und in einem zweiten Durchgang mit dem virtuellen Kopfhörer durch. Die Versuchsperson nahm dazu vor dem bereitgestellten Computer Platz und erhielt eine schriftliche Anleitung. Sie setzte dann entweder den realen Kopfhörer auf oder der Versuchsleiter aktivierte den virtuellen Kopfhörer. Danach wurden alle Testsignale geordnet nach der Pegeldifferenz  $\Delta L$  von  $-30$  dB bis  $+30$  dB vorgespielt, um den Wahrnehmungshorizont abzugrenzen.

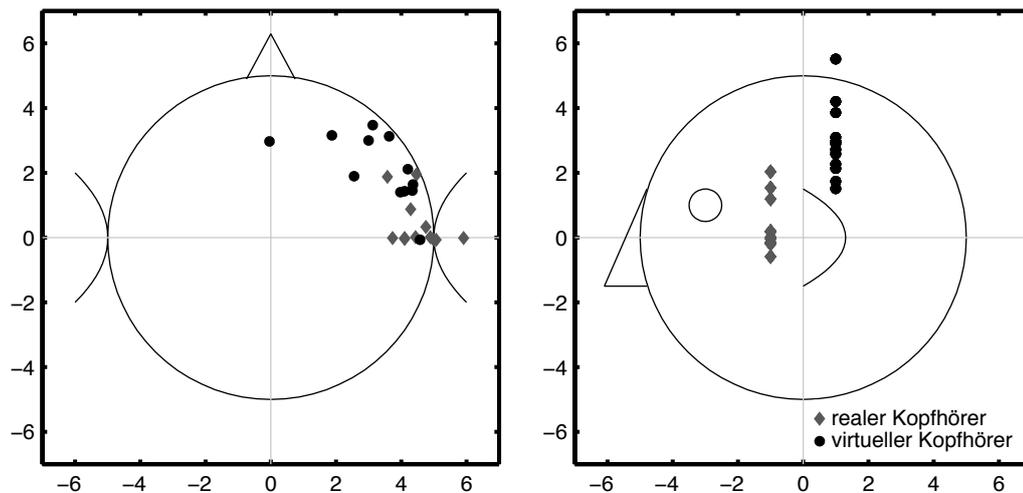
Während eines angeleiteten Trainings, bei dem fünf zufällig ausgewählte Testsignale zu bewerten waren, konnte sich die Versuchsperson anschließend mit der grafischen Benutzeroberfläche<sup>41</sup> vertraut machen und noch Rückfragen stellen. Bei Bedarf wurde dieses Training wiederholt.

Es bot sich ebenfalls an, eine grafische Bewertungsmethode anzuwenden, indem der Ort des Hörereignisses in einem stilisierten Kopf markiert werden sollte. Auch hier hat sich ein ähnliches Vorgehen schon in der Vergangenheit bewährt<sup>vgl. z. B. [108]</sup>. Das Testsignal wurde abgespielt, und dann war die erste Aufgabe der Versuchsperson, in der Aufsicht des Kopfes, Abb. 5.2 links, die Position des Hörereignisses in der Horizontalebene, also die laterale und sagittale Auslenkung, einzutragen. Nach einer endgültigen Bestätigung wurde das Testsignal wiederholt, um in der Seitenansicht des Kopfes, Abb. 5.2 rechts, schließlich noch die Höhe des Hörereignisses, also die longitudinale Auslenkung, festzulegen. Wurde auch diese Auswahl endgültig bestätigt, folgte das nächste Testsignal. Abb. 5.2 enthält exemplarisch die tatsächlichen Antworten einer Versuchsperson.

Die Bewertung erfolgte damit in den drei Dimensionen transversal, sagittal und longitudinal auf einer stufenlosen Skala von  $-7$  bis  $+7$ , wobei der vorgegebenen Skizze entsprechend Werte zwischen  $\pm 5$  und  $\pm 7$  auf einer Achse außerhalb des Kopfes lagen. Insgesamt waren die vier Grundsignale bei sieben Pegeldifferenzen

---

<sup>41</sup>engl.: graphical user interface (kurz: GUI)



**Abbildung 5.2** Grafische Benutzeroberfläche zur Beurteilung der Lateralisation in den drei Dimensionen *transversal* (links–rechts), *sagittal* (vorne–hinten) und *longitudinal* (oben–unten). Beispielhaft sind hier die Antworten von Versuchsperson 9 auf alle Signale mit einer interauralen Pegeldifferenz von +30 dB eingetragen, sowohl bei der Wiedergabe über einen realen (●) als auch über einen virtuellen (◆) Kopfhörer (Prototyp).

zu beurteilen. Alle Signale sollten dreimal getestet werden, so dass 84 Stimuli in zufälliger Reihenfolge abgespielt wurden. Die Gesamtdauer dafür betrug im Schnitt etwa 20 Minuten, wobei die Versuchsperson in der Lage war, das Tempo selbst zu bestimmen.

### Versuchspersonen

An beiden Durchgängen des Lateralisationsversuchs nahmen 17 Personen im Alter von 23 bis 64 Jahren teil. Alle Probanden zeigten ein altersgemäßes Hörvermögen und hatten Erfahrung mit verschiedenen Hörversuchen.

## 5.3 Wahrnehmbarkeit des interauralen Kohärenzgrads

Der interaurale Kohärenzgrad  $k$  ist definiert als absolutes Maximum der normierten Kreuzkorrelationsfunktion  $\Phi_{p_l p_r}(\tau)$  der Ohrsignale  $p_l(t)$  und  $p_r(t)$ , formal<sup>[23,S.162]</sup>

$$k = \max_{\tau} |\Phi_{p_l p_r}(\tau)| \quad (5.1)$$

Der Einfachheit halber wird der Begriff „interauraler Kohärenzgrad“ im Folgenden auch benutzt, wenn es sich eigentlich um den Kohärenzgrad der Eingangssignale eines Kopfhörers handelt. Laut Definition kann  $k$  nur einen Wert zwischen 0 und 1 annehmen, der die Ähnlichkeit der beiden Signale  $p_l(t)$  und  $p_r(t)$  charakterisiert. Bis auf frequenzunabhängige Pegeldifferenzen oder verzerrungsfreie Verschiebungen, die den Kohärenzgrad  $k$  nicht beeinflussen, weisen vollständig kohärente Signale, das heißt  $k = 1$ , keine Unterschiede auf. Sie liefern bei Kopfhörerwiedergabe, zum Beispiel auch in Lateralisationsexperimenten, ein einheitliches und scharf begrenztes Hörereignis. Sinkt der Kohärenzgrad – das bedeutet die Ohrsignale werden immer unähnlicher –, dehnt sich das Hörereignis aus und zerfällt für kleine Werte von  $k$  schließlich in zwei Teile<sup>[28,S.105f.]</sup>. Laut gängiger Interpretation wertet man diesen Effekt als Hinweis, dass das Gehör zusammengehörende Anteile der Ohrsignale segmentieren kann und daraus unterschiedliche Schallquellen erkennt<sup>vgl. z. B. [23,S.195]</sup>.

In Anlehnung an eine ältere Hypothese<sup>[149]</sup> sollte durch ein Experiment überprüft werden, ob Unterschiede in der Wahrnehmung des Kohärenzgrads zwischen realer und virtueller Kopfhörerwiedergabe bestehen und ob daraus ein Rückschluss auf die Wiedergabequalität des jeweiligen Systems gezogen werden kann. Als Maßzahl wurde die Wahrnehmbarkeitsschwelle des interauralen Kohärenzgrads ermittelt.

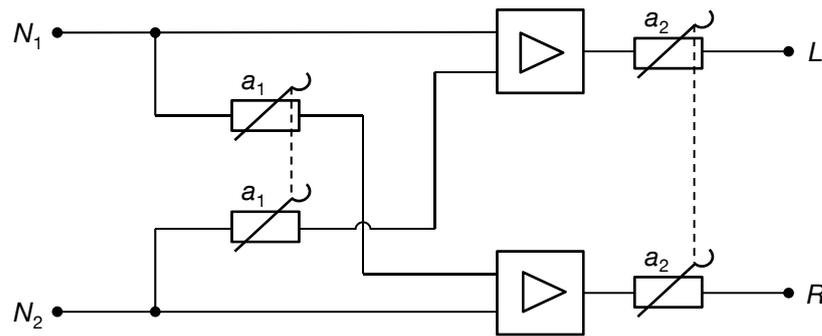
## Versuchsaufbau

Der Aufbau glich weitgehend dem des Lateralisationsversuchs (sh. Abschn. 5.2). Allerdings wurden, wie in der folgenden Tabelle aufgelistet, insgesamt drei Kopfhörermodelle unterschiedlicher Preisklassen mit dem virtuellen Kopfhörer verglichen.

Nr.	Kopfhörermodell
1	Beyerdynamic DT 831
2	Philips SBC HL 140
3	STAX SR $\lambda$ pro <b>mit</b> Diffusfeldentzerrung
4	STAX SR $\lambda$ pro <b>ohne</b> Diffusfeldentzerrung
5	virtueller Kopfhörer (Prototyp)

Kopfhörer 1 und 2 wurden wieder über den bereits erwähnten Kopfhörerverstärker (Samson C-que 8) betrieben, während der elektrostatische Kopfhörer STAX SR  $\lambda$  pro mit einem speziell abgestimmten Verstärker arbeitet, der über eine zuschaltbare Diffusfeldentzerrung verfügt.

Alle Probanden führten den Versuch mit jedem Kopfhörer einmal durch, allerdings in unterschiedlicher Reihenfolge, um einen systematischen Unterschied zwischen den



**Abbildung 5.3** Blockschaltbild zur Generierung eines Rauschsignalpaares  $[L, R]$  mit änderbarem Kohärenzgrad aus zwei Rauschquellen  $N_1$  und  $N_2$  durch Einstellung der Verstärkungsfaktoren  $a_1$  und  $a_2$  nach [149].

Kopfhörermodellen, beispielsweise verursacht durch einen Lerneffekt, auszuschließen. Vor Beginn eines Versuchsdurchgangs wurde daher für den Teilnehmer individuell einer der fünf Kopfhörer vorbereitet.

## Stimuli

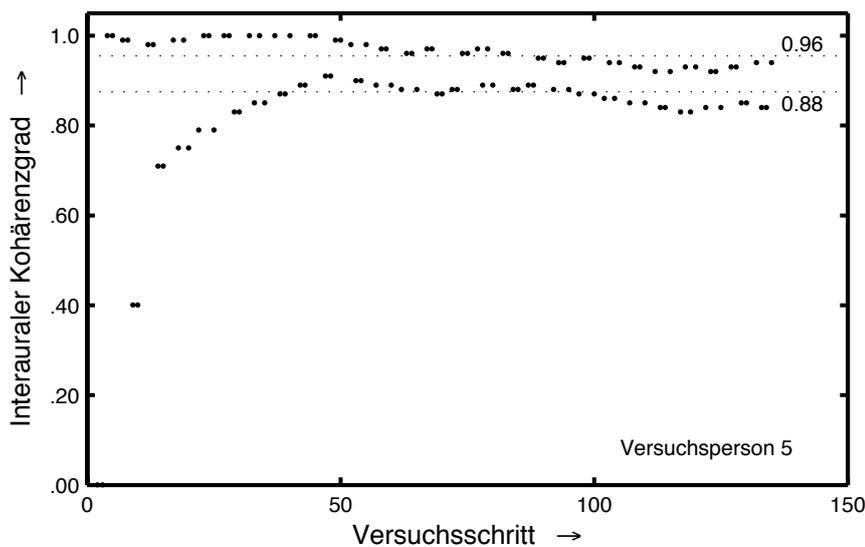
Die Stimuli waren einsekündige Rauschsignalpaare mit unterschiedlichem Kohärenzgrad. Sie wurden digital in CD-Qualität aus zwei unkorrelierten Rosa Rauschen mit Hilfe einer Überlagerungsmethode generiert, die durch das Blockschaltbild Abb 5.3 schematisch beschrieben ist. Die Signale waren im Zeitbereich gaußförmig mit einem Anstieg und Abfall von je  $20\text{ ms}$  begrenzt. Insgesamt wurden 38 Stimuli vorbereitet: für einen Kohärenzgrad 0 bis 0,7 in zehntel, darüber in hundertstel Schritten, da nach der Auswertung von Testläufen die Wahrnehmbarkeitsschwelle im oberen Drittel der Kohärenzgradskala zu suchen war.

## Versuchsablauf

Anhand einer schriftlichen Anleitung und mehrerer Beispielstimuli wurde jeder Proband vor seinem ersten Durchgang in das Experiment eingewiesen.

Der Hauptversuch war als sogenanntes transformiertes Treppenstufenverfahren<sup>42</sup> konzipiert, bei dem die bereits gegebenen Antworten den weiteren Verlauf des Versuchs beeinflussen [72, S. 227 ff.] [104, 197]. Der Versuchsperson wurde mit Wiederholung ein Stimulus immer im direkten Vergleich zum vollständig kohärenten Referenzstimulus präsentiert. Das heißt, es wurde ihr eine Sequenz  $A B - A B$  dargeboten,

<sup>42</sup>engl.: staircase method; auch: Auf-Ab Methode, engl: up-down method



**Abbildung 5.4** Typisches Versuchsprotokoll zur Ermittlung der Wahrnehmbarkeitsschwelle des interauralen Kohärenzgrads bei Kopfhörerwiedergabe mit einem adaptiven Verfahren. Die Abbildung zeigt exemplarisch das Ergebnis von Versuchsperson 5 für einen virtuellen Kopfhörer (Prototyp).

wobei die Referenz zufällig an erster oder zweiter Stelle, also bei A oder B, auftrat. Sie musste dann entscheiden, ob sie einen Unterschied zwischen den beiden Signalen wahrgenommen hatte. Allgemein gilt: Bewertete die Versuchsperson die beiden Signale einer Sequenz als gleich, verringerte sich der Kohärenzgrad des nächsten Testsignals, bewertete sie die Signale als verschieden, erhöhte er sich.

Die Annäherung an die Wahrnehmbarkeitsschwelle erfolgte zufällig abwechselnd von beiden Seiten und begann jeweils mit Testsignalen am Rand der Kohärenzgradskala, also  $k = 0$  und  $k = 1$ . Zusätzlich fand eine sogenannte 2-ab/1-auf Regel in leicht abgewandelter Form Anwendung. Das bedeutet, dass jede Sequenz zweimal direkt hintereinander zu bewerten war. Bei der Annäherung von unten erhöhte sich der Kohärenzgrad des nächsten Testsignals nur dann, wenn für beide Sequenzen das Urteil „verschieden“ abgegeben worden war, andernfalls wurde er wieder verringert. Für die Annäherung von oben galt diese Regel entsprechend umgekehrt. Sinnvolle Schrittweiten wurden vorab so ermittelt, dass sich der Versuch möglichst schnell im relevanten Bereich bewegte. Das Protokoll Abb. 5.4 zeigt einen typischen Verlauf des Experiments. Bei der Auswertung wurde für beide Kurven jeweils der Median aus den Werten aller Umkehrpunkte gebildet und als obere beziehungsweise untere Schwelle festgelegt. Dieses Ergebnis ist auch im Protokoll Abb. 5.4 mit gestrichelten Linien und Wertangaben markiert.

Pendelte der Kohärenzgrad des Testsignals bei der Annäherung von oben viermal und bei der Annäherung von unten dreimal um denselben Wert, wurde der Versuch automatisch beendet. Für den Fall, dass die Versuchsperson sehr unsicher war und diese Bedingung nicht eintrat, war ein Maximum von 150 Schritten vorgesehen. Diese Anzahl wurde in allen 80 Einzeldurchgängen allerdings nur ein einziges Mal erreicht und dabei konnte trotzdem ein verwertbares Endergebnis gewonnen werden. Abhängig vom Erreichen dieser Abbruchkriterien nahm ein Versuchsdurchgang zwischen 15 und 30 Minuten in Anspruch.

Die Ergebnisse des Experiments wurden zusätzlich noch durch einen Rückversuch nach dem Konstanzverfahren<sup>43</sup> mit Kopfhörer 1 auf Plausibilität kontrolliert.

### **Versuchspersonen**

Von den 16 Versuchspersonen haben 13 auch an dem im vorhergehenden Abschn. 5.2 beschriebenen Lateralisationsversuch mitgewirkt. Die Altersspanne lag daher genauso bei 23 bis 64 Jahren, wie sich alle Teilnehmer durch ein altersgemäßes Hörvermögen und entsprechende Hörversuchserfahrungen auszeichneten.

---

<sup>43</sup>*engl.*: method of constant stimuli

# 6

## Experimente mit geänderten geometrischen Parametern

Das vorhergehende Kap. 5 befasste sich mit Experimenten, die mit dem Prototyp des virtuellen Kopfhörers durchgeführt wurden. Das zugrunde liegende Funktionsprinzip ist aber nicht auf die für den Prototyp spezifizierte Anordnung (sh. Abschn. 4.2) beschränkt. Im Folgenden werden daher Versuche beschrieben, in denen geometrische Parameter verändert wurden. Dies betrifft einerseits die Position der Sekundärquellen, also die Lautsprecheranordnung des virtuellen Kopfhörers. Andererseits ist auch, wie schon in Abschn. 4.2 angedeutet, die Frage nach der optimalen Konstellation der virtuellen Quellen, also ihre Anzahl und Lage im kopfbezogenen Koordinatensystem, noch offen. Sie wird im letzten Abschnitt dieses Kapitels behandelt. Der Fokus der Untersuchungen liegt wie in Kap. 5 auf den Punkten Lokalisation, Lateralisation und Wahrnehmbarkeit des interauralen Koheränzgrads. Die Ergebnisse sind ebenfalls in Kap. 7 zu finden.

### 6.1 Lokalisation von binauralen Signalen

Allgemein zeichnen sich binauraltechnische Systeme durch eine akkurate Abbildung des Seitenwinkels der dargestellten Schallquellen aus. Es können jedoch, besonders bei statischer Binauraltechnik, Richtungsinversonen auftreten. Im Vergleich dazu ist die Abbildung des Höhenwinkels wesentlich ungenauer, nicht selten lassen sich Abweichungen von  $30^\circ$  und mehr beobachten vgl. z.B. [115]. Diese Probleme zeigten sich auch schon bei der Entwicklung des BRS-Systems und konnten selbst durch die zusätzliche Berücksichtigung von Nickbewegungen des Kopfes nicht zufriedenstellend verbessert werden [113][114, Fig. 4.17]. Auch mit einem Stereo-Lautsprecherpaar auf Ohrhöhe erzeugte Phantomschallquellen wurden in der BRS-Simulation durchschnittlich  $7^\circ$  höher als bei realer Wiedergabe wahrgenommen [147][114, Fig. 4.27] (vgl. Abschn. 4.2).

In Studien zur virtuellen Kopfhörerwiedergabe von binauralen Signalen wird die Elevationsabbildung oft überhaupt nicht betrachtet. Falls doch, sind Abweichungen in ähnlichen Größenordnungen wie bei realen Kopfhörern feststellbar vgl. z. B. [46,48,91]

[97,Manuscript E][115]. Auch der in Kap. 4 beschriebene Prototyp eines auf Wellenfeldsynthese basierenden virtuellen Kopfhörers zeigt entsprechende Schwächen in der Darstellung des Höhenwinkels (vgl. Abb. 7.1). Unklar ist jedoch, ob Abbildungsfehler der Schalleinfallrichtung bei virtueller Kopfhörerwiedergabe durch das Wiedergabesystem oder durch die binauralen Eingangssignale verursacht werden. Daher sollen in einem Experiment reale und virtuelle Kopfhörerwiedergabe einander gegenübergestellt werden, um die Richtungsabbildung und speziell die Abweichungen der Elevation zu untersuchen.

## Versuchsaufbau

Der Versuch wurde in einer schallgedämmten, abgedunkelten Hörversuchskabine mit einer etwa  $1\text{ m}^2$  großen, quadratischen Grundfläche durchgeführt. Für die reale Kopfhörerwiedergabe lag ein offener Kopfhörer (Sennheiser HD 650) bereit. Die virtuelle Kopfhörerwiedergabe erfolgte über ein Lautsprecherpaar (Bose Video RoomMate), das auf Ohrhöhe, also  $1.25\text{ m}$  über dem Kabinenboden, mit der maximal möglichen Basisbreite von  $0.85\text{ m}$  installiert war. Der Hörer befand sich aufgrund der räumlichen Gegebenheiten etwa  $0.6\text{ m}$  von der Basis entfernt, die Lautsprecher lagen damit bei kopfbezogenen Seitenwinkeln von  $\alpha \approx \pm 35^\circ$ . Die für die Berechnung der Übersprechkompensation nötigen binauralen Raumimpulsantworten waren direkt am Versuchsaufbau mit einem Kunstkopf (Neumann KU 100) gemessen worden. Da der reale Kopfhörer für das gleiche Kunstkopfmodell kalibriert war, wiesen virtuelle und reale Kopfhörerwiedergabe identische Übertragungseigenschaften auf und können zudem als diffusfeldentzerrt angesehen werden (sh. Abschn. 2.2, 2.4). In beiden Telexperimenten standen der Versuchsperson in der Kabine Bildschirm, Maus und Tastatur zur Verfügung, um eine grafische Benutzeroberfläche für die Eingabe der Bewertungen zu steuern.

## Stimuli

Grundstimulus war ein  $200\text{ ms}$  langer, zwischen  $180\text{ Hz}$  und  $14\text{ kHz}$  bandbegrenzter Rosa Rauschpuls mit gaußkurvenförmigen,  $5\text{ ms}$  langen Flanken. Der Rauschpuls wurde dreimal, getrennt durch  $300\text{ ms}$  dauernde Pausen, wiederholt. Um daraus binaurale Signale mit verschiedenen Schalleinfallrichtungen zu generieren, wurde eine von der Fachhochschule Köln veröffentlichte Datenbank mit kopfbezogenen Übertragungsfunktionen des Kunstkopfmodells KU 100 von Neumann<sup>[22,178]</sup> benutzt. Bei der realen Kopfhörerwiedergabe wurde auch auf die in die Datenbank integrierte Kalibrierung für verschiedene Kopfhörermodelle mehrerer Hersteller<sup>sh. [178, Table 3]</sup> zurückgegriffen. Für die virtuelle Kopfhörerwiedergabe wurde aus Stabilitätsgründen vor der Berechnung der Übersprechkompensation die Amplitude der binauralen Raumimpulsantworten mit einem gleitenden terzbreiten Rechteckfenster geglättet

und der Übertragungsbereich der resultierenden Filter durch einen Butterworth-Bandpass vierter Ordnung auf 120 Hz bis 14 kHz beschränkt. Die untere Grenzfrequenz der Stimuli musste dann aber sogar, wie oben beschrieben, auf 180 Hz erhöht werden, da in der Kabine sonst Resonanzeffekte auftraten.

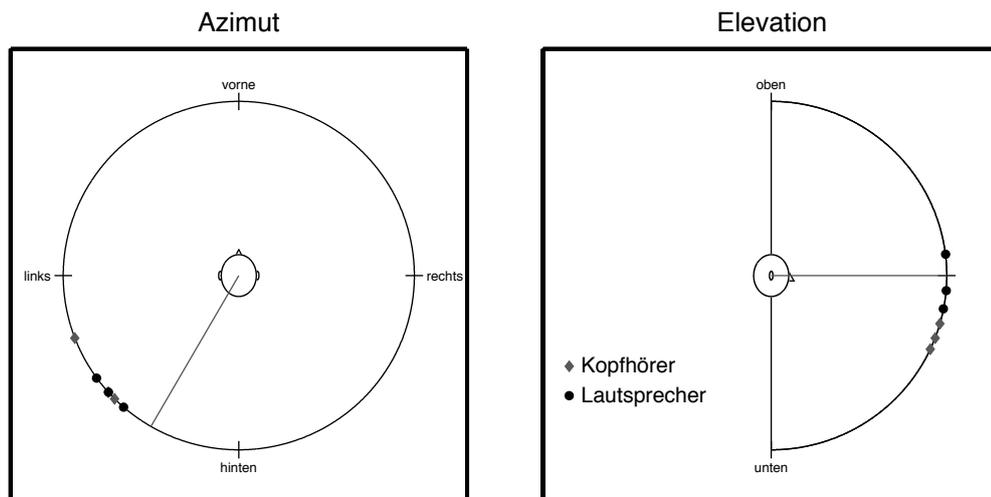
Getestet wurden acht unsymmetrisch angeordnete Seitenwinkel  $\alpha = 0^\circ, 30^\circ, 90^\circ, 150^\circ, 180^\circ, 240^\circ, 300^\circ, 345^\circ$  kombiniert mit den in  $15^\circ$ -Schritten gerasterten Höhenwinkeln  $\varepsilon = 0^\circ, \pm 15^\circ, \pm 30^\circ, \pm 45^\circ, \pm 60^\circ$ , so dass sich 72 Schalleinfallrichtungen ergeben. Zusätzlich wurden zwei Kontrollsignale erzeugt, das eine binaural mit einer Elevation von  $+90^\circ$ , also einem Schalleinfall direkt von oben, und das andere monofon, indem der Grundstimulus sowohl über Kopfhörer als auch über das übersprechkompensierte Lautsprecherpaar unbearbeitet auf dem rechten Kanal eingespielt wurde. Zusammen genommen erhält man damit 74 verschiedene Stimuli.

Die Verstärkung der Lautsprecher war so eingestellt, dass am Abhörpunkt in der geschlossenen Hörversuchskabine ein bewerteter Schalldruckpegel von etwa 70 dB(A) gemessen wurde, wenn beide Kanäle ein unbearbeitetes, zwischen 180 Hz und 14 kHz bandbegrenztes Rosa Rauschen abspielten. Die Kopfhörerverstärkung wurde mit Hilfe des auch in den anderen Messungen benutzten Kunstkopfs daran angepasst.

## Versuchsablauf

Der Versuch lief ähnlich ab wie das Lokalisationsexperiment mit dem Prototyp des virtuellen Kopfhörers, das in Abschn. 5.1 beschrieben ist. Die Testperson nahm in der Hörversuchskabine Platz, wo ihr die Aufgabe, die Anzeigemethode und die Bedienung der grafischen Benutzeroberfläche erklärt wurde. Dann bekam sie vom Versuchsleiter entweder den Kopfhörer überreicht oder wurde auf die vorgesehene Abhörposition der virtuellen Kopfhörerwiedergabe ausgerichtet. Da im Unterschied zu Abschn. 5.1 eine statische Binauraltechnik implementiert war, erfolgte die explizite Anweisung, Kopfposition und Kopfausrichtung während des Versuchs unverändert beizubehalten. Um die Versuchsperson dann auf das Experiment vorzubereiten, wurden ihr alle Stimuli in der Horizontalebene, also mit Elevation  $\varepsilon = 0^\circ$ , in geordneter Reihenfolge vorgespielt, so dass sie im Idealfall einen auf Augenhöhe gegen den Uhrzeigersinn um den Kopf laufenden Schalleinfall wahrnehmen konnte. Bei Bedarf wurde dieses Training noch einmal wiederholt. Nachdem eventuelle Unklarheiten beseitigt waren, startete das Experiment. Bei virtueller Kopfhörerwiedergabe wurde davor noch einmal die Sitzposition überprüft und gegebenenfalls korrigiert.

Wie bewährt gab die Versuchsperson die gehörte Schalleinfallrichtung am Bildschirm mit Hilfe der in Abb. 6.1 gezeigten Vorlage getrennt in den Koordinaten Azimut und Elevation an. Die Abbildung ist maßstabsgetreu und war auf einem 19-Zoll



**Abbildung 6.1** Grafische Benutzeroberfläche zur Angabe des gehörten Seitenwinkels (Azimut  $\alpha$ ) und des gehörten Höhenwinkels (Elevation  $\epsilon$ ) von binauralen Signalen. Beispielhaft sind hier alle Antworten von Versuchsperson 1 auf ein statisches binaurales Signal mit intendierter Schalleinfallrichtung aus  $\alpha = 150^\circ$ ,  $\epsilon = 0^\circ$  angegeben, sowohl bei der Wiedergabe über einen Kopfhörer ( $\blacktriangle$ ) als auch über ein übersprechkompensiertes Lautsprecherpaar ( $\bullet$ ).

Monitor dargestellt, so dass die Seitenlänge der beiden Quadrate etwa  $8\text{ cm}$  betrug. Die gewählten Winkel wurden mit Einheitszeigern markiert, eine Distanzbewertung war nicht vorgesehen. Den Ablauf des Versuchs kontrollierte der Proband selbst. Für jeden Stimulus konnte er sein Urteil bis zur endgültigen Bestätigung – durch Klick auf eine „Weiter“-Schaltfläche – beliebig oft nachjustieren und hatte auch die Möglichkeit das Signal bei Bedarf zu wiederholen. Die Versuchspersonen waren jedoch angewiesen, ein intuitives Urteil zu fällen und die Wiederholungsmöglichkeit nicht zu stark auszureizen.

Das gesamte Experiment bestand aus zwei Versuchsreihen, einmal mit realer und einmal mit virtueller Kopfhörerwiedergabe. In jeder Versuchsreihe wurden alle 74 Stimuli dreimal und in zufälliger Reihenfolge präsentiert, so dass insgesamt 222 Testschalle bewertet werden mussten. Deshalb wurden beide Versuchsreihen in je drei, insgesamt also sechs, Sitzungen aufgeteilt, die im Mittel etwa 13 Minuten dauerten. Alle Probanden führten zuerst die Versuchsreihe mit realer Kopfhörerwiedergabe durch.

## Versuchspersonen

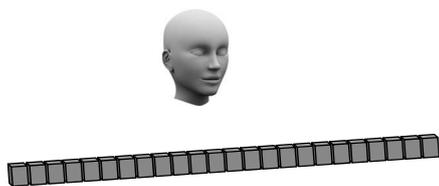
Die beiden Hörversuchsreihen wurden von 17 Probanden mit einer Altersspanne von 24 bis 49 Jahren durchgeführt. Das Hörvermögen aller Teilnehmer war altersgemäß und ein Großteil hatte Erfahrung mit psychoakustischen Experimenten. Ein unerfahrener Versuchsteilnehmer, der auch nach eigener Aussage Schwierigkeiten mit dem Bewertungsverfahren hatte, wurde bei der Auswertung (sh. Abb. 7.2, 7.3) ausgeschlossen, so dass insgesamt 16 Datensätze verarbeitet wurden.

## 6.2 Anordnung der Sekundärquellen

Das Funktionsprinzip des virtuellen Kopfhörers konnte mit den bisher beschriebenen Studien zwar grundsätzlich bestätigt werden, genau genommen allerdings nur für das präzise definierte Layout des Prototyps (sh. Abschn. 4.2). Gegenstand der folgenden Untersuchung war daher die Anordnung der Sekundärquellen, mit der die virtuellen Lautsprecher in Form von fokussierten Quellen erzeugt werden.

Dazu wurden drei Aufbauten entworfen, die man aufgrund ihrer Abmessungen für eine Anwendung in höhlenartigen virtuellen Realitäten mit Mehrseitenprojektion<sup>44</sup> oder bei Großbildleinwänden mit stereoskopischer Darstellung<sup>45</sup> in Betracht ziehen könnte. An ihnen sollte eine allgemeinere und zudem praxisnahe Anwendbarkeit des Funktionsprinzips gezeigt werden. Eine Anlehnung an die Wellenfeldsynthese wird mit der engen Aneinanderreihung der Lautsprecher beibehalten.

**Anordnung 1** erfüllt als einzige Wiedergabekonfiguration die Voraussetzungen der Wellenfeldsynthese und stellt einen klassischen Linienaufbau dar. Sie erreicht eine Gesamtbreite von 2,16 m und wurde so befestigt, dass sich die Mitte jeder Lautsprecherbox 1,2 m über dem Boden und damit auf Ohrhöhe einer sitzenden Versuchsperson befand. Im Folgenden wird sie kurz als „**Linie**“ bezeichnet.

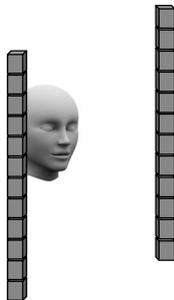


**Abbildung 6.2**  
Sekundärquellenanordnung 1  
(Linie)

<sup>44</sup>engl.: cave automatic virtual environment (kurz: CAVE)

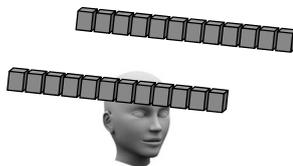
<sup>45</sup>engl.: 3D powerwall

**Anordnung 2** besteht aus zwei senkrechten Türmen mit je 12 Lautsprechern. Die Höhe der Türme beträgt  $1,44\text{ m}$ , der Abstand zwischen ihnen  $1,535\text{ m}$ . Die beiden untersten Lautsprecher waren  $0,48\text{ m}$  vom Boden entfernt, so dass die Mitte jedes Turms ebenfalls in  $1,2\text{ m}$  Höhe lag. Die Anordnung wird im weiteren Verlauf kurz „**Türme**“ genannt.



**Abbildung 6.3**  
Sekundärquellenanordnung 2  
(Türme)

**Anordnung 3** leitet sich aus Anordnung 2 ab. Der Unterschied besteht darin, dass die beiden Lautsprechertürme oberhalb des Hörers liegen und die Anordnung damit dem ursprünglichen Prototyp (vgl. Abb. 4.1) ähnelt. Die Abmessungen wurden beibehalten, beide Lautsprecherzeilen weisen eine Länge von  $1,44\text{ m}$  auf und liegen parallel in einem Abstand von  $1,535\text{ m}$  zueinander über dem Kopf des Hörers. Als Kurzbezeichnung wird „**Oberhalb**“ verwendet.



**Abbildung 6.4**  
Sekundärquellenanordnung 3  
(Oberhalb)

Alle drei Anordnungen bestehen aus 24 geschlossenen Lautsprecherboxen (Elac 301) mit den Abmessungen  $9\text{ cm} \times 12\text{ cm} \times 9,5\text{ cm}$ . Sie wurden ohne Zwischenräume aufgestellt, wodurch sich die beschriebenen Dimensionen erklären.

Für die Hörversuche wurden drei definierte Abhörscenarien festgelegt, mit denen jede Anordnung getestet wurde. Der Hörer befand sich dabei mittig in einem Abstand von  $1,2\text{ m}$  vor oder unter der jeweiligen Anordnung, womit die Sekundärquellen also wesentlich weiter von der Abhörposition entfernt sind als beim Prototyp. Während der große Abstand für die Anordnungen Linie und Türme einfach zu realisieren war, bedeutete dies für die Anordnung Oberhalb, dass die Lautsprecher in einer Höhe von etwa  $2,4\text{ m}$  an einer aufwendigen Rahmenkonstruktion befestigt werden mussten.

Jeder Aufbau erzeugte für einen virtuellen Kopfhörer vier Fokuspunkte an folgenden

kopfbezogenen Koordinaten: die Horizontalwinkel  $\alpha$  betragen  $\pm 22,5^\circ$  und  $\pm 67,5^\circ$ . Bei den Anordnungen Linie und Türme waren Radius  $r = 15 \text{ cm}$  und Elevation  $\varepsilon = 0^\circ$ , die Fokuspunkte waren also in der Horizontalebene in unmittelbarer Nähe zum Kopf platziert. Bei der Anordnung Oberhalb wurde diese Quellenkonstellation um  $20 \text{ cm}$  nach oben verschoben, womit sich  $r = 25 \text{ cm}$  und  $\varepsilon \approx 53^\circ$  ergaben. Alle Anordnungen wurden so mit einer gleichartigen Konstellation kopfnaher virtueller Quellen getestet, die den für die Schallfokussierung erlaubten Hörbereich berücksichtigte.

**Szenario 1** stellt die Ausgangsposition dar: der Hörer blickt so nach vorne, dass die jeweilige Sekundärquellenanordnung symmetrisch zur Medianebene ist. Damit sind die Lautsprecher gleichmäßig, das heißt in gleicher Anzahl und mit paarweise gleichen Abständen, auf der linken und rechten Seite des Kopfes verteilt. So wird vermieden, dass in der Ausgangsposition unerwünschte Effekte auftreten, die zum Beispiel durch Laufzeitunterschiede verursacht werden könnten. Die acht Übertragungsfunktionen zwischen den vier virtuellen Quellen und den beiden Ohren des Hörers wurden für jede Anordnung separat eingemessen und daraus ein spezifisches Filter für die Übersprechkompensation  $\mathbf{H}_0^+$  berechnet (sh. Abschn. 2.4).

**Szenario 2** prüft die Robustheit des virtuellen Kopfhörers bei Kopfdrotationen. Dazu dreht der Hörer seinen Kopf um  $30^\circ$  nach rechts. Die Lautsprecher sind dann aus seiner Sicht nicht mehr symmetrisch aufgestellt. Darüber hinaus führt dies bei den Anordnungen Linie und Türme dazu, dass das linke Ohr der Lautsprecheranordnung zugewandt ist, während das rechte Ohr durch den Kopf abgeschattet wird. Gemäß dem propagierten Funktionsprinzip werden die virtuellen Quellen der Kopfdrehung angepasst und so replaziert, dass sich ihre kopfbezogenen Koordinaten nicht ändern. Das Filter für die transaurale Wiedergabe entspricht jedoch weiterhin dem aus Szenario 1, also  $\mathbf{H}_0^+$ , so dass eine statische Übersprechkompensation vorliegt.

**Szenario 3** vergleicht die statische Übersprechkompensation aus Szenario 2 mit einer adaptiven Lösung. Der Kopf bleibt zusammen mit den virtuellen Quellen wie bei Szenario 2 um  $30^\circ$  nach rechts im Vergleich zur Ausgangsposition gedreht. Die Übertragungsfunktionen wurden für die veränderte Situation jedoch nochmals gemessen und ein neues Filter für die transaurale Wiedergabe  $\mathbf{H}_{30^\circ}^+$  berechnet.

Jede der drei beschriebenen Lautsprecheranordnungen Linie, Türme und Oberhalb wurde mit allen drei Abhörszenarien getestet, so dass sich insgesamt neun Teilexperimente ergaben.

Die Hörversuche fanden in einem reflexionsarmen Raum statt, um Fremdeinflüsse durch Raumakustik – wegen der im Vergleich zu den anderen Versuchen deutlich größeren Dimensionen der Wiedergabeeinrichtungen – zu vermeiden. Die gesamte Ansteuerungselektronik befand sich außerhalb des Versuchsraums. Ein Stuhl mit Kopfstütze diente dazu, in jedem Teilerperiment für alle Probanden eine identische Abhörposition gemäß der Vorgabe des jeweiligen Szenarios zu reproduzieren. Die Lautsprecheranordnungen Linie und Türme waren dabei sicherheitshalber von einem akustisch durchlässigen Vorhang verdeckt, um die Teilnehmer nicht durch visuelle Reize vom Versuch abzulenken. Bei der Anordnung Oberhalb befanden sich während des Versuchs keine Sekundärquellen im Blickfeld des Hörers, so dass hier auf einen Sichtschutz verzichtet wurde. Die Teilnehmer saßen auf dem vorbereiteten Stuhl und waren angewiesen, während des Versuchsablaufs die festgelegte und überprüfte Abhörposition einzuhalten und auch den Kopf nicht zu bewegen. Da sie so nicht in der Lage waren, ihre Bewertungen zu notieren oder eine grafische Benutzeroberfläche zu bedienen, teilten sie ihre Antworten dem Hörversuchsleiter mündlich per Gegensprechanlage mit.

Analog zu den Versuchen mit dem Prototyp bestand jedes Teilerperiment aus einem Hörversuch zur Lateralisation und einem Hörversuch zur Wahrnehmbarkeit des interauralen Kohärenzgrads.

### 6.3 Lateralisation

Im Unterschied zum Lateralisationsexperiment mit dem Prototyp (sh. Abschn. 5.2) sollten neben interauralen Pegeldifferenzen nun auch Zeitdifferenzen untersucht werden. Wegen der Beschränkung auf mündliche Bewertungen musste jedoch das Versuchskonzept vereinfacht werden. Deshalb wurde entschieden, nur noch die transversale Auslenkung des Hörereignisses auf einer elf-stufigen Kategorienskala beurteilen zu lassen und – anstatt der sagittalen und longitudinalen Auslenkung – direkt nach dem Auftreten einer Im-Kopf-Lokalisation zu fragen. Der zugehörige **Versuchsaufbau** ist in Abschn. 6.2 beschrieben.

#### Stimuli

Grundstimulus war ein Rosa Rauschpuls von 300 *ms* Dauer mit gaußkurvenförmigen, 20 *ms* langen Flanken, der im Abstand von 300 *ms* sechsmal wiederholt wurde, so dass daraus ein 3,3 *s* langes Gesamtsignal entstand. Gepulst war das Rauschen, um Zeitversätze zu verdeutlichen.

Zwischen beiden Kopfhörerkanälen wurden Pegeldifferenzen  $\Delta L$  bis zu  $\pm 30$  dB sowie Zeitdifferenzen  $\Delta t$  bis zu  $\pm 1000$   $\mu s$  in folgender Abstufung erzeugt:

Nr.	$\Delta L/dB$	Nr.	$\Delta t/\mu s$
1	-30	12	-1000
2	-24	13	-650
3	-18	14	-450
4	-12	15	-250
5	-6	16	-150
6	0	17	0
7	+6	18	+150
8	+12	19	+250
9	+18	20	+450
10	+24	21	+650
11	+30	22	+1000

Zusätzlich wurde Signal 6 in beiden Kanälen um 6 dB, die Signale 5 und 7 um 2 dB abgesenkt. Diese grobe Pegelanpassung bewirkte, dass subjektiv alle Stimuli ähnlich laut waren. Die Endverstärker der Lautsprecheranordnungen wurden so eingestellt, dass der bewertete Schalldruckpegel am Abhörpunkt etwa 70 dB(A) erreichte.

Aus den 22 Stimuli und den für jede Anordnung separat ermittelten Filtern zur Übersprechkompensation  $\mathbf{H}_{0^\circ}^+$  und  $\mathbf{H}_{30^\circ}^+$  wurden vorab vierkanalige transaurale Signale berechnet, um für diesen Schritt während des Versuchs eine Signalverarbeitung in Echtzeit zu vermeiden. Die so nachbearbeiteten Stimuli konnten stattdessen direkt an einen Wellenfeldsyntheseprozessor übergeben werden, der, wie beschrieben, vier fokussierte Quellen in unmittelbarer Nähe zum Kopf erzeugte (sh. Abschn. 6.2, vgl. auch Abb. 4.2). Drei zusätzliche Kontrollsignale entstanden dadurch, dass einzelne Lautsprecher aus der jeweiligen Anordnung den Grundstimulus auch unbearbeitet abspielten. Sie wurden von den Versuchspersonen in allen Teilerperimenten sehr zuverlässig erkannt. Wie schon in Abschn. 6.1 angemerkt, wurden auch hier die der Übersprechkompensation zugrunde liegenden Übertragungsfunktionen vor der Berechnung geglättet und die resultierenden Filter bandbegrenzt.

## Versuchsablauf

Die Versuchsperson wurde gebeten eine schriftliche Anleitung durchzulesen und konnte bei Unklarheiten Fragen stellen. Mit Hilfe des Versuchsleiters nahm sie dann die für Anordnung und Szenario vorgesehene Abhörposition auf dem Stuhl ein und wurde mit einem Training auf den Hauptversuch vorbereitet. Dazu setzte sie einen

Kopfhörer (STAX SR  $\lambda$  pro) auf, über den ihr mehrere Stimuli vorgespielt wurden. Der Versuchsleiter stand dabei für weitere Erklärungen bezüglich der Bewertung zur Verfügung, nahm abschließend den Kopfhörer entgegen und kontrollierte noch einmal die Abhörposition, bevor er den Raum verließ.

Im Hauptversuch waren mündlich für jeden Testschall zwei Urteile abzugeben: Die Versuchspersonen sollten, erstens, mit ja oder nein beantworten, ob sie den Schall im oder am Kopf wahrgenommen hatten und, zweitens, die transversale Auslenkung auf einer Kategorienskala bewerten. Wie beim Lateralisationsversuch mit dem Prototyp (sh. Abschn. 5.2) standen die Randwerte  $-5$  und  $+5$  für das linke und rechte Ohr,  $0$  lag in der Kopfmittle. Die Auslenkung sollte auf die interaurale Achse projiziert und ganzzahlig angegeben werden, so dass die Skala insgesamt elf Stufen umfasste <sup>vgl. [54, S. 308 f.]</sup>.

Die 22 Stimuli und die drei Kontrollsignale wurden jeweils dreimal präsentiert, und bis auf die Vorgabe, dass sich kein Testsignal direkt wiederholen soll, war die Abfolge zufällig. Insgesamt war eine Versuchsperson also 75-mal gefordert, einen Stimulus bezüglich Im-Kopf-Lokalisation und Lateralisation zu bewerten. Im Mittel benötigte sie dazu etwa 14 Minuten.

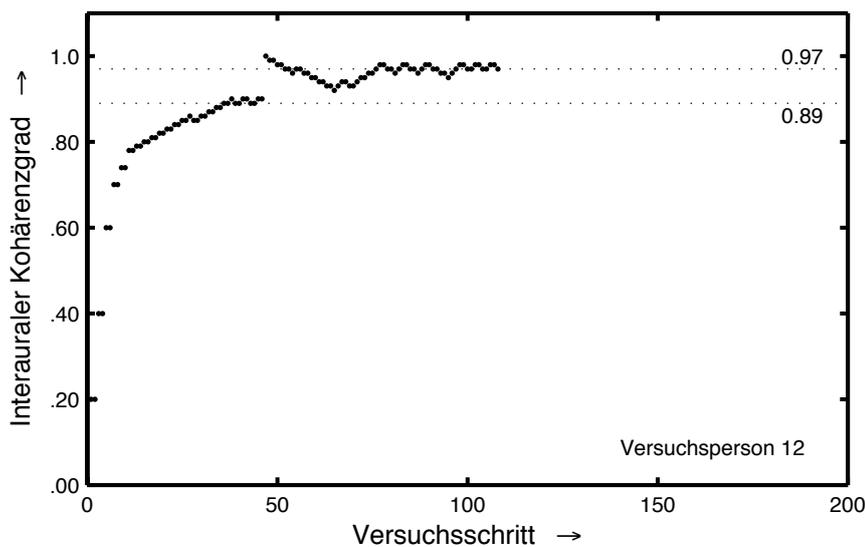
Zum Vergleich wurde der Versuch vor den Teilerperimenten mit den drei Lautsprecheranordnungen auch mit einem Kopfhörer (STAX SR  $\lambda$  pro) durchgeführt.

## 6.4 Wahrnehmbarkeit des interauralen Kohärenzgrads

Im Wesentlichen wurde hier das ursprüngliche Konzept des Experiments mit dem Prototyp (sh. Abschn. 5.3) übernommen. Einige leichte Änderungen optimierten den Versuchsablauf, vor allem um eine kürzere Gesamtdauer zu erreichen. Details zum **Versuchsaufbau** können Abschn. 6.2 entnommen werden.

### Stimuli

Die Signalpaare, bestehend aus Rosa Rauschen mit unterschiedlichem Kohärenzgrad, stammten direkt vom Versuch mit dem Prototyp (sh. Abschn. 5.3) und wurden ohne Änderungen wiederverwendet. Wie beim Lateralisationsversuch (sh. Abschn. 6.3) wurden daraus mit den Übersprechkompensationsfiltern  $H_{0^\circ}^+$  und  $H_{30^\circ}^+$  transaurale Wiedergabesignale erstellt.



**Abbildung 6.5** Typisches Versuchsprotokoll zur Ermittlung der Wahrnehmbarkeitsschwelle des interauralen Kohärenzgrades bei virtueller Kopfhörerwiedergabe mit unterschiedlichen Sekundärquellenanordnungen. Die Abbildung zeigt exemplarisch das Ergebnis von Versuchsperson 12 für den Versuchsaufbau „Türme (Szenario 3)“.

### Versuchsablauf

Beim Ablauf wurden jedoch Anpassungen vorgenommen, die auch in den Protokollen zu erkennen sind (vgl. Abb. 6.5 mit Abb. 5.4). Die Annäherungen an die obere und untere Schwelle liefen nicht mehr abwechselnd sondern hintereinander ab, und die adaptive 2-ab/1-auf Regel war ohne Abwandlung umgesetzt. Das heißt, bei der Annäherung von unten verringerte sich der Kohärenzgrad des nächsten Testsignals sofort, wenn die Versuchsperson angegeben hatte, einen Unterschied wahrgenommen zu haben. Erhöht wurde er nur dann, wenn die Sequenz zuvor zweimal direkt hintereinander als gleich bewertet worden war. Für die Annäherung an die obere Schwelle ist diese Regel entsprechend umgekehrt zu interpretieren. Damit konnte sich der Kohärenzgrad bereits nach einer – anstatt grundsätzlich nur nach zwei – Antworten ändern.

Die Sequenz aus Testsignal und Referenz wurde ohne Wiederholung vorgespielt und die maximale Anzahl an Annäherungsschritten war auf 100 beschränkt. Sie wurde in 20 von 218 und damit in etwa 9 % der Fälle ausgeschöpft. Zusätzlich war festgelegt, dass die Ermittlung der Schwelle dann abbricht, wenn die Versuchsperson dreimal um denselben Wert geschwankt hatte und dabei die Varianz der letzten zwölf erreichten Kohärenzgrade kleiner oder gleich 0.4 war. Erneut galt der Median aus den Werten aller Umkehrpunkte in einer Annäherung als Wahrnehmbarkeitsschwelle.

Mit allen Anpassungen konnte die Gesamtdauer von vormals 15 bis 30 Minuten auf nun 10 bis 15 Minuten verkürzt werden. Die ausführliche Einweisung vor dem Versuch mit schriftlicher Anleitung und Training wurde aber beibehalten.

### **Versuchspersonen**

An beiden Hörversuchsreihen mit den je neun Telexperimenten nahmen insgesamt 16 verschiedene Versuchspersonen im Alter von 22 bis 52 Jahren teil. Die Telexperimente mit der Konfiguration Linie (Szenario 1) wurden von 13, alle weiteren je von 12 Probanden absolviert. 10 Personen waren Stammteilnehmer, die sich für alle Versuche zur Verfügung stellten. Die Teilnehmer waren ihrem Alter entsprechend normalhörend und hatten größtenteils bereits Erfahrung mit psychoakustischen Experimenten.

## **6.5 Konstellation der virtuellen Quellen**

In Abschn. 2.4 wurde angedeutet, dass die übersprechkompensierte Wiedergabe von binauralen Signalen verbessert werden kann, wenn man mit mehr als zwei rund um den Hörer verteilten Lautsprechern arbeitet. Konkrete Experimente zum Einfluss der Lautsprecheraufstellung beziehen sich in den meisten Fällen allerdings auf Forschungsarbeiten zum Stereo-Dipol oder zur optimalen Quellenverteilung (sh. Abschn. 2.4). Zu den wenigen Ausnahmen zählt eine in Japan durchgeführte und 2001 veröffentlichte Studie<sup>[75]</sup>. Sie orientiert sich an den von der ITU empfohlenen Mehrkanaltonformaten<sup>[80]</sup> und vergleicht zwölf verschiedene Konstellationen mit zwei bis acht Lautsprechern in zwei Metern Entfernung zum Hörer mittels eines Hörversuchs zur Richtungswahrnehmung. Als Schlussfolgerung wird dort festgehalten, dass auf der transversalen Achse keine Lautsprecher platziert werden sollten.

Solche Studien zeigen, dass die Übersprechkompensation auch durch die Quellenkonstellation beeinflusst werden kann. Prinzipiell lassen sich beim Prototyp virtuelle Quellen weitgehend frei um den Kopf des Hörers platzieren, es existieren aber weder Angaben über ihre Anzahl noch über ihre Position im kopfbezogenen Koordinatensystem (vgl. Abschn. 4.2). Naheliegender praktischer Erwägungen, das sind vor allem die Abmessungen des Prototyps und mögliche Mehrbenutzer-Szenarien, ist jedoch das Anordnen von virtuellen Quellen mit geringem Abstand zum Kopf. Um diesen Punkt näher zu betrachten, sollten mehrere Konstellationen mit kopfnahen realen Quellen verglichen werden. Auch dabei sollte als Bewertungsmaßstab nicht die Richtungswahrnehmung bei binauralen Eingangssignalen dienen, sondern die für Kopfhörerwiedergabe typischen Effekte Lateralisation und Im-Kopf-Lokalisation.

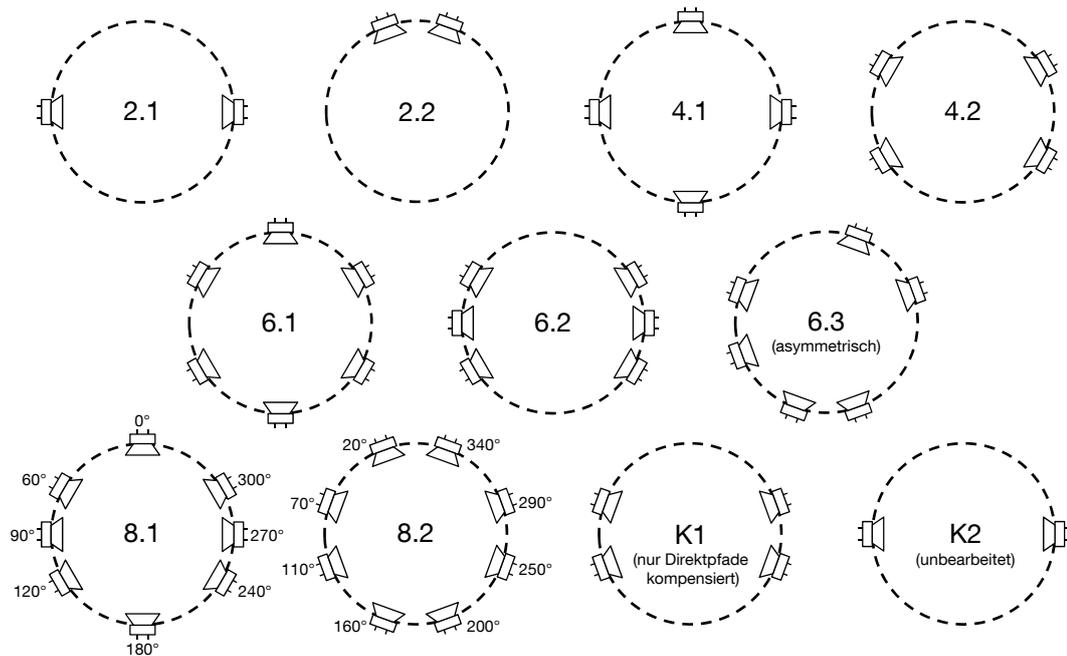
## Versuchsaufbau

Anders als bei den allgemein üblichen Messmethoden mit Kopfbezug<sup>vgl. z. B. [39,70,118]</sup> müssen bei kopfnahen Schallquellen zusätzliche Punkte berücksichtigt werden. Eine wichtige Rolle spielt die Größe des Lautsprechers. Berechnungen mit einem Kugelmodell des Kopfes ergaben zum Beispiel, dass bei Entfernungen  $r$  von 15 bis 20 *cm*, das ist ein Außenabstand von etwa 6 bis 11 *cm* zum Kopf, der Radius der Schallquelle maximal 3 bis 5 *cm* betragen sollte, um Streueffekte zu vermeiden<sup>[228]</sup>. Daher wurden sowohl für die Messreihen als auch beim Hörversuch Kleinlautsprecher mit einem Membrandurchmesser von etwa 3,5 *cm* (Philips, ohne Modellbezeichnung), die in ein zylindrisches Kunststoffgehäuse eingebaut waren, verwendet (sh. Abschn. 4.4). Zudem sind die kopfbezogenen Übertragungsfunktionen im Nahbereich des Hörers nicht nur winkelabhängig sondern auch entfernungsabhängig (vgl. Abschn. 2.1). Bei kurzen Distanzen vergrößert sich der Einfluss der akustischen Parallaxe<sup>[36]</sup>, außerdem können höhere interaurale Pegeldifferenzen und eine ausgeprägte Tiefpassfilterung auftreten, nur die interauralen Zeitdifferenzen verändern sich kaum<sup>[34,136]</sup>.

Gemessen wurde im reflexionsarmen Raum mit einem Kunstkopf (Neumann KU 100). Weil die Daten zur Erstellung von Übersprechkompensationsfiltern dienen sollten, wurde keine Bezugsmessung an der Stelle des kopfbezogenen Koordinatenursprungs durchgeführt. So bleiben die benötigten Schalllaufzeiten erhalten und müssen später nicht erneut berechnet werden. Genau genommen wurden also binaurale Raumimpulsantworten im Freifeld für zahlreiche kopfnaher Schallquellen in regelmäßigen Winkelabständen und verschiedenen Entfernungen ermittelt (vgl. Abschn. 2.1). Um daraus das Abstrahlverhalten des Lautsprechers zu eliminieren und gleichzeitig die akustische Parallaxe zu berücksichtigen, erfolgten anschließend Bezugsmessungen mit zwei Mikrofonen (Schoeps CCM 2S) an den beiden Stellen der Ohrkanäleingänge. Damit wurden die Amplitudenfrequenzgänge der zugehörigen Kunstkopfmessungen korrigiert.

Eigentlich verschlechtert sich die Übersprechkompensation durch diese Korrektur, da der Einfluss des Lautsprechers vernachlässigt wird, obwohl er bei der Wiedergabe vorhanden ist. Die Motivation, trotzdem wie beschrieben vorzugehen, bestand darin, kopfbezogene Übertragungsfunktionen für den Nahbereich zu erhalten, die möglichst nur von der Schallquellenposition und nicht von gerätespezifischen Merkmalen abhängig und damit in anderen Versuchen wiederverwendbar sind.

Für die Hörversuche wurden elf verschiedene kopfnaher Quellenkonstellationen zur Erzeugung einer virtuellen Kopfhörerwiedergabe festgelegt. Sie leiten sich aus zwei Grundkonstellationen mit je acht Lautsprechern ab, die im folgenden mit 8.1 und 8.2 bezeichnet sind. Bei Konstellation 8.1 befinden sich die Quellen an Azimuten von 0°, 60°, 90°, 120°, 180°, 240°, 270° und 300°. Alle Lautsprecher liegen in der Horizontalebene,



**Abbildung 6.6** Schematische Darstellung der 11 Lautsprecherkonstellationen mit den dafür verwendeten Bezeichnungen. Sie dienen zur Untersuchung möglicher Anordnungen von kopfnahen Quellen für die transaurale Wiedergabe.

das heißt ihre Elevation beträgt  $0^\circ$ , und damit sind je drei Quellen in der Nähe der Ohren und zusätzlich je eine direkt vor und direkt hinter dem Hörer platziert. Bei Konstellation 8.2 betragen die Azimute  $20^\circ$ ,  $70^\circ$ ,  $110^\circ$ ,  $160^\circ$ ,  $200^\circ$ ,  $250^\circ$ ,  $290^\circ$  und  $340^\circ$ , die Elevation ebenfalls  $0^\circ$ , so dass die Quellen gleichmäßiger um den Kopf verteilt sind und nicht auf der sagittalen und transversalen Achse liegen. Durch Deaktivieren von Lautsprechern entstehen aus 8.1 die Konstellationen 2.1, 4.1, 4.2, 6.1, 6.2 und K2, aus 8.2 ergeben sich die restlichen, nämlich 2.2, die asymmetrische 6.3 und K1. In dieser Aufteilung wurden sie daher auch in getrennten Versuchsdurchgängen, einmal mit sieben und einmal mit vier Konstellationen, getestet. Abb. 6.6 bietet eine grafische Übersicht, in der Benennung steht die erste Ziffer für die Anzahl der Lautsprecher, die zweite Ziffer ist eine laufende Nummerierung.

Als weiterer Parameter sollte außer der Anordnung auch die Entfernung der Quellen Berücksichtigung finden. Die elf Konstellationen wurden daher mit Radien von  $24\text{ cm}$  und  $39\text{ cm}$  realisiert, das entspricht seitlichen Abständen zwischen Quelle und Ohr von etwa  $15\text{ cm}$  und  $30\text{ cm}$ . Dafür wurde einen halben Meter über der Abhörposition eine passende kreisförmige schallgedämmte Holzplatte angebracht, von der acht der beschriebenen Kleinlautsprecher in einer der beiden Grundkonstellationen mit einem Gestänge abgehängt waren.

K1 und K2 sind Kontrollkonstellationen, bei denen die Übersprechkompensation nicht vollständig umgesetzt ist. K1 entzerrt nur die Direktpfade, das bedeutet, es werden nur die ipsilateralen Übertragungsfunktionen berücksichtigt. K2 gibt das Eingangssignal ganz ohne Bearbeitung wieder. Mit diesen Konstellationen sollte überprüft werden, ob die bei kopfnahen Quellen durch natürliche Abschattung auftretende Kanaltrennung, die hier laut Messung bis  $2\text{ kHz}$  um etwa  $10\text{ dB}$  und über  $2\text{ kHz}$  noch deutlich höher liegen kann, bereits ausreicht, eine virtuelle Kopfhörerwiedergabe zu induzieren. Für die restlichen neun Konstellationen wurde aus den Messdaten der kopfnahen Schallquellen jeweils ein Filter zur Übersprechkompensation berechnet (sh. Abschn. 2.4).

## Stimuli

Als Grundstimuli waren Rosa Rauschpulse und eine Sprachsequenz vorgesehen. Natürliche Stimuli wie Sprache werden in der Virtuellen Akustik verwendet, weil man sich, einerseits, an der Anwendung orientiert<sup>vgl. z. B. [18]</sup>, und andererseits, weil das Gehör bekannte Signale besser beurteilen kann und beispielsweise die tatsächliche Schalleinfallrichtung leichter erkennt<sup>[28, S. 95]</sup>.

Die Rosa Rauschpulse waren jeweils  $300\text{ ms}$  lang und hatten gaußkurvenförmige,  $20\text{ ms}$  lange Flanken. Mit zwölf Wiederholungen und  $600\text{ ms}$  Pause entstand daraus ein Gesamtsignal von etwa 10 Sekunden. Das Sprachsignal war ein 15 Sekunden dauernder Ausschnitt eines englischen Begrüßungstextes, der für Vorführungen des Prototyps in einem Rundfunkstudio professionell produziert worden war. Die Wiedergabelautstärke war so eingestellt, dass für beide Stimuli am Abhörpunkt ein bewerteter Schalldruckpegel von etwa  $65\text{ dB}(A)$  herrschte. Beim im Pegel stärker schwankenden Sprachstimulus wurde dazu der Wert des äquivalenten Dauerschallpegels über die gesamte Wiedergabedauer von 15 Sekunden gemessen.

Dazu lässt sich festhalten, dass es bei den Ergebnissen nur geringe Unterschiede zwischen beiden Stimuli gibt. Rosa Rauschen liefert in einigen Fällen extremere Werte, in der Aggregation der Daten führt das zu vergrößerten Interquartilbereichen, also zu einer höheren Streuung.

## Versuchsablauf

Zu Beginn des jeweiligen Telexperiments erhielt die Versuchsperson eine schriftliche Instruktion. Anschließend wurde ihr Rosa Rauschen mit einem Kopfhörer (AKG 240 DF) diotisch sowie monotisch links und monotisch rechts vorgespielt. An dieser Referenz sollte sie sich im weiteren Verlauf des Experiments orientieren. Dann setzte sie sich auf den bereit gestellten Stuhl und wurde unter Zuhilfenahme einer Messlatte

auf die vorgesehene Abhörposition im Mittelpunkt der jeweiligen Grundkonstellation ausgerichtet. Alle Probanden wurden angewiesen, die Position einzuhalten und Kopfbewegungen zu vermeiden. Eine Kopfstütze oder eine anderweitige Fixierung des Kopfes kam bewusst nicht zum Einsatz. Instabile Konstellationen, die empfindlich auf Abweichungen reagieren, sind so leichter zu identifizieren. Nach der Ausrichtung begann eine Trainingsrunde, die alle zu bewertenden Konstellationen beinhaltet. Bei Unklarheiten konnten noch Fragen gestellt werden, danach startete der Hauptversuch.

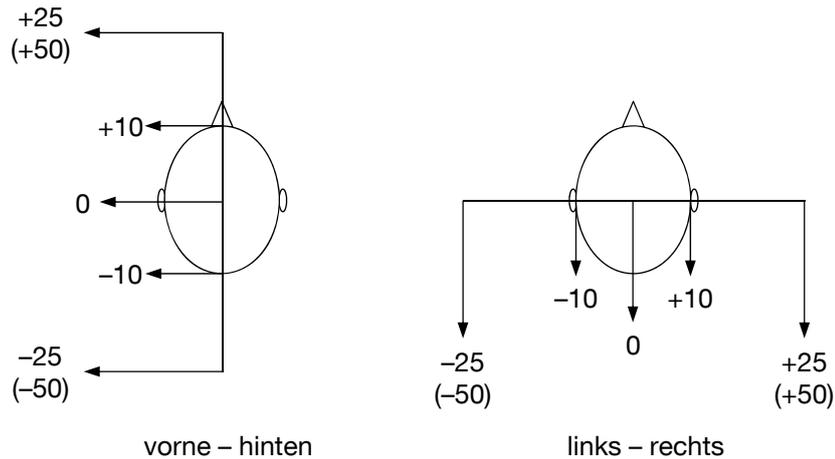
Gemäß Instruktion hatte die Versuchsperson für jeden präsentierten Stimulus vier Bewertungen abzugeben. Auch hier teilte sie – bedingt durch die spezifizierte Sitzposition und Kopfausrichtung – ihre Urteile wieder mündlich mit.

Zuerst sollte sie mit ja oder nein beantworten, ob das Hörereignis im Kopf stattgefunden hatte. Es wurde also explizit nach einer Im-Kopf-Lokalisation gefragt.

Als zweites war die Quellenausdehnung auf einer Skala von 0 bis 50 zu bewerten. Die Fragestellung lautete, ob sich das Signal gegenüber der Kopfhörerwiedergabe bei der Einführung verbreitert hätte. Die Antwort 0 stand für eine Punktquelle, also für sehr kompakt, und 50 für sehr breit. Diese Formulierung lässt sich berechtigt kritisieren, da Frage und Antwort nicht zusammen passen. Allerdings kann man davon ausgehen, dass die überwiegend erfahrenen Probanden die intendierte Skala richtig verstanden hatten. Das Ergebnis wird hier vorweg genommen, weil sich die Bewertungen zwischen 10 und 20 konzentrierten und kaum signifikante Unterschiede aufweisen. Bei monotonischer Wiedergabe über die Kontrollanordnungen K1 und K2 wurde die Quellenausdehnung am kompaktesten beurteilt, weil die Versuchspersonen vermutlich einzelne Lautsprecher erkannten. Mit Abstand am breitesten wurde – ebenfalls bei monotonischer Wiedergabe – Anordnung 2.2 mit einer Quellenentfernung  $r = 24\text{ cm}$  bewertet, die auch Auffälligkeiten bei den Lateralisationsergebnissen zeigt.

Die letzten beiden Urteile bezogen sich auf die sagittale und transversale Auslenkung des Hörereignisses. Entsprechend der vorgegebenen Skizze, die in Abb. 6.7 originalgetreu zu sehen ist, hatte die Skala zwei Bezugspunkte. Die Werte  $\pm 10$  markierten den Rand des Kopfes,  $\pm 25$  die Entfernung der Lautsprecher bei Konstellationen mit dem kleinen Radius  $r = 24\text{ cm}$ ,  $\pm 50$  entsprechend für den großen Radius  $r = 39\text{ cm}$ . Die Versuchsperson sollte auf die Konsistenz der Antworten achten. Hatte sie die Frage zur Im-Kopf-Lokalisation bejaht, war die sagittale und transversale Auslenkung folglich mit Werten zwischen  $\pm 10$  anzugeben. Die Idee hinter dieser Skalierung war, dass die Ergebnisse für die beiden unterschiedlichen Quellenentfernungen direkt vergleichbar sein sollten.

Da sich die 11 Lautsprecherkonstellationen aus den zwei Grundkonstellationen 8.1 und 8.2 ableiten, entstanden zwei unterschiedlich lange Durchgänge mit 7 und 4



**Abbildung 6.7** Skala zur Beurteilung der sagittalen (vorne – hinten) und transversalen (links – rechts) Auslenkung des Hörereignisses bei monotonischer und diotischer Wiedergabe über die 11 untersuchten Lautsprecherkonstellationen.

Konstellationen. Zusammen mit den beiden festgelegten Entfernungen,  $r = 24 \text{ cm}$  und  $r = 39 \text{ cm}$ , gab es also vier verschiedene Teilexperimente, für die ein Umbau der zugrunde liegenden Lautsprecheranordnung erforderlich war. In jedem Teilexperiment wurde die virtuelle Kopfhörerwiedergabe mit zwei Wiedergabemodi getestet: einmal diotisch und einmal monotonisch links. Pro Wiedergabemodus wurde in zufällig permutierter Reihenfolge jede Quellenkonstellation zweimal mit gepulstem Rauschen und zweimal mit dem Sprachsignal angesteuert. Damit waren je Modus 28 oder 16 und demnach im gesamten Teilexperiment 56 oder 32 Stimuli zu bewerten, was im Mittel ungefähr 32 oder 18 Minuten dauerte.

### Versuchspersonen

Insgesamt waren 11 Probanden an den Versuchen zur Quellenkonstellation beteiligt. Sie waren zwischen 24 und 52 Jahre alt, hatten ein ihrem Alter entsprechendes Hörvermögen und größtenteils Versuchserfahrung. Jedes einzelne Teilexperiment wurde mit 10 Personen durchgeführt, davon waren 9 identisch.



# 7

## Ein neuer Ansatz zur Qualitätsbeurteilung von virtuellen Kopfhörern

Die in den Kap. 5 und 6 beschriebenen Experimente stützen sich auf die These, dass das Gehör in der Lage ist, aus Erfahrung Lokalisationsinformationen, die in den Ohrsignalen enthalten sind, zu identifizieren und separat zu verarbeiten. Begründet wird sie so: Bei Kopfbewegungen können sich die Ohrsignalspektren einer Schallquelle wegen der Richtcharakteristik des Kopfes sehr stark verändern. Trotzdem nimmt der Hörer normalerweise keine Klangverfärbungen wahr. Man vermutet daher, dass das Gehör in einer ersten Stufe kopfbezogene Übertragungsfunktionen anhand erlernter Muster erkennt und aus den Ohrsignalen filtert<sup>vgl. [61,64,189]</sup>. Dadurch wird auch die wahrgenommene Richtung beziehungsweise der wahrgenommene Ort der Schallquelle festgelegt. Das nun einer bestimmten Schallquelle zugeordnete und von der Lokalisationsinformation befreite Signal wird erst in nachgelagerten Stufen weiter verarbeitet. Alle spektralen Veränderungen der Ohrsignale, die nur aufgrund der Richtcharakteristik des Kopfes auftreten, werden dabei aber nicht mehr beachtet.

In einer 1980 an der Technischen Universität Berlin eingereichten Dissertation<sup>[188]</sup> ist diese Vermutung detailliert erläutert. Sie wird dort als „Assoziationsmodell“ vorgestellt, das Elemente aus der Gestalttheorie enthält und diverse Phänomene des räumlichen Hörens kongruent erklären kann. Es liegt der Diskussion der Versuchsergebnisse als Hypothese zugrunde.

Ein Augenmerk richtet sich auf die *Im-Kopf-Lokalisation*. Sie ist typisch für Kopfhörerbeschallung mit nicht-binauralen Signalen, lässt sich aber auch mit Lautsprechern in ungewohnten Abhörsituationen oder durch unbekannte Schalle provozieren<sup>[141,142]</sup>. Der Effekt wird so erklärt, dass das Gehör einer Schallquelle nur dann einen Ort außerhalb des Kopfes zuordnet, wenn es eine bekannte oder plausible Lokalisationsinformation detektiert. Andernfalls entsteht ein Hörereignis im Kopf<sup>[188,189]</sup>.

Die Gesetzmäßigkeiten der *Lateralisation* (sh. Abschn. 5.2) bringt das Modell mit den bei Stereophonie auftretenden Phantomschallquellen in Zusammenhang. In der ersten Stufe registriert das Gehör zwar zwei Schallquellen, bei Stereophonie die beiden Lautsprecher und bei Lateralisation die Kopfhörermuscheln. Sind deren Signale

jedoch hinreichend kohärent, werden sie in einer nachgelagerten Verarbeitungsstufe verschmolzen, einschließlich eventuell vorhandener Lokalisationsinformationen. Es entsteht ein einziges Hörereignis, eine sogenannte Phantomschallquelle, bei Stereophonie außerhalb des Kopfes zwischen den beiden Lautsprechern, bei Lateralisation innerhalb des Kopfes zwischen den beiden Kopfhörermuscheln. Treten zwischen den beiden Schallquellen Laufzeitdifferenzen oder Pegelunterschiede auf, wird die Wahrnehmung durch die frühere beziehungsweise stärkere Schallquelle dominiert, so dass sich das Hörereignis in diese Richtung verschiebt<sup>[188,189]</sup>.

Aus der angenommenen Modellvorstellung wird folgende Überlegung abgeleitet: Ein virtueller Kopfhörer darf keine eigene Lokalisationsinformation produzieren, da sonst die Lokalisationsinformation binauraler Signale bei der Wiedergabe verfälscht werden könnte. Wird ein nicht-binaurales Signal dargeboten, sollten im Idealfall also keine für das Gehör erkennbaren Lokalisationsinformationen vorhanden sein. Daher müsste – genau wie bei einem realen Kopfhörer – eine Im-Kopf-Lokalisation auftreten. Bei Laufzeitdifferenzen oder Pegelunterschieden zwischen den beiden Kopfhörerkanälen ist dann analog ein im Kopf lateralisiertes Hörereignis zu erwarten. Stellen sich die Effekte Im-Kopf-Lokalisation und Lateralisation jedoch nicht zuverlässig ein, muss man davon ausgehen, dass unerwünschte Lokalisationsinformationen auftreten – beispielsweise weil das Gehör die beteiligten Sekundärquellen erkennt – und damit auch die Wiedergabe von binauralen Signalen beeinträchtigt sein wird.

Als Konsequenz wird somit vorgeschlagen, virtuelle Kopfhörer mit Lateralisationsexperimenten zu testen und nicht – wie es sonst oft praktiziert wird – mit Versuchen zum Richtungshören (sh. Abschn. 5.1, 6.1). Der wichtigste Vorteil besteht darin, dass für die Erzeugung passender Stimuli kein binauraltechnischer Verarbeitungsschritt nötig ist, der vor allem im statischen Fall, das bedeutet Kopfbewegungen werden nicht berücksichtigt, sowie bei unsauberer Umsetzung sehr fehleranfällig sein kann (sh. Abschn. 2.1). Dieser Vorschlag wird in Abschn. 7.1 anhand der in den Hörversuchen ermittelten Daten diskutiert.

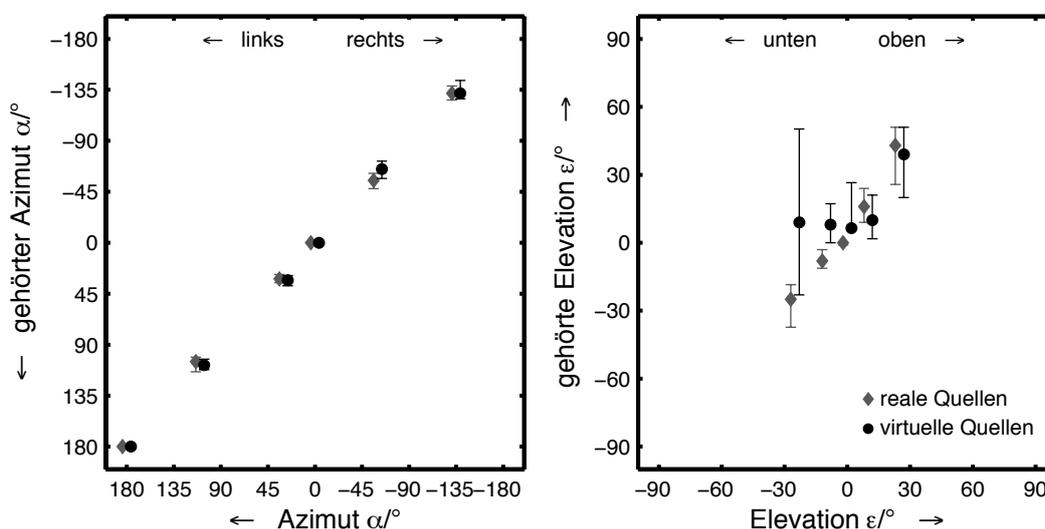
Grundsätzlich stützt sich dieser Ansatz auf das erwähnte Assoziationsmodell. Er stellt aber keinen Widerspruch zu anderen verbreiteten Modellen für das räumliche Hören dar, die sich mehr auf die Auswertung der interauralen Kreuzkorrelation und damit auf die Auswertung von interauralen Zeitdifferenzen und interauralen Pegeldifferenzen konzentrieren<sup>vgl. z. B. [24, Abschn. 3.4] [181]</sup>. Demnach muss ein virtueller Kopfhörer in der Lage sein, diese Ohrsignalunterschiede wahrnehmungsgerecht zu reproduzieren, was mit Lateralisationsversuchen leicht überprüft werden kann. Das Auftreten von Im-Kopf-Lokalisation erklären oder fordern solche Modelle jedoch nicht.

## 7.1 Diskussion der Versuchsergebnisse

Im folgenden Abschnitt sind alle Ergebnisse der in Kap. 5 und 6 beschriebenen Hörversuche zusammengefasst.

### Lokalisation

Aus Vergleichsgründen wurde beim Prototyp des virtuellen Kopfhörers (sh. Abschn. 4.2) auch die Richtungswahrnehmung bezüglich Azimut und Elevation bei der Wiedergabe binauraler Signale systematisch untersucht (sh. Abschn. 5.1). Sowohl der Prototyp als auch das umfassend erprobte BRS-System zur Erzeugung der binauralen Signale (sh. Abschn. 2.3) waren dabei in der Lage, Kopfbewegungen der Probanden dynamisch zu kompensieren.



**Abbildung 7.1** Gesamtergebnis des Hörversuchs zur Lokalisation. Die Abbildung zeigt Mediane und Interquartilbereiche der Antworten aller Versuchspersonen für sechs reale Quellen ( $\blacklozenge$ ) und sechs korrespondierende BRS-simulierte Quellen ( $\bullet$ ), vgl. Abschn. 5.1 ab S. 55. Links ist über dem tatsächlichen Azimut der Quelle der gehörte Azimut aufgetragen – die *umgekehrte Richtung der Koordinatenachsen* ergibt sich aus der Definition des kopfbezogenen Koordinatensystems. Rechts ist entsprechend die Elevation dargestellt, das Ergebnis der beiden Quellen auf Ohrhöhe (d. h.  $\varepsilon = 0^\circ$ ) ist dabei zusammengefasst<sup>adaptiert von [123,124]</sup>.

Abb. 7.1 links zeigt hervorragende Ergebnisse für die Richtungswahrnehmung in der Horizontalebene. Aufgetragen ist der gehörte Azimut über dem tatsächlichen für die sechs realen Quellen und ihre BRS-simulierten Pendanten, die korrekt und praktisch

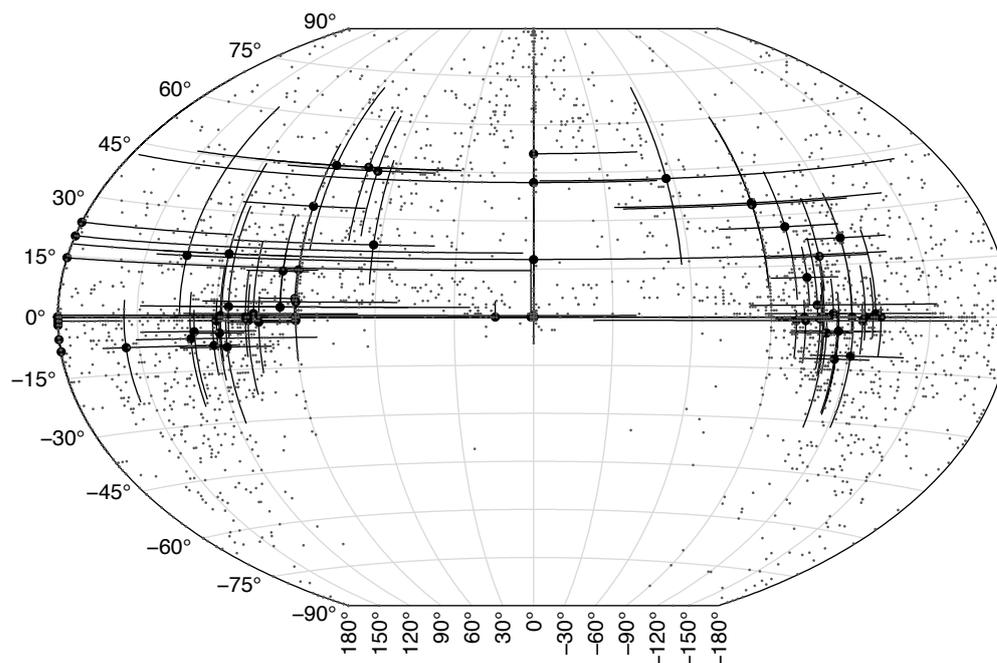
identisch wahrgenommen werden. Die Umkehrung der Koordinatenachsen wurde wegen der Definition des kopfbezogenen Koordinatensystems gewählt. Die Streuung der Daten, in den Grafiken als Interquartilabstand mit Fehlerbalken visualisiert, ist in beiden Fällen minimal. Somit funktioniert die Horizontalrichtungsabbildung auch bei der Simulation sehr präzise und es treten weder Richtungsinversionen noch größere Richtungsabweichungen auf.

Offensichtlich gibt es jedoch Probleme bei der Darstellung der Elevation, wie Abb. 7.1 rechts zu entnehmen ist. Reale Quellen werden im Mittel etwa um ihre tatsächlichen Erhebungswinkel wahrgenommen, aber selbst hier ist, vor allem bei den um  $\varepsilon = \pm 25^\circ$  nach oben und unten verschobenen Quellen, eine deutliche Unsicherheit feststellbar. Die Abbildung BRS-simulierter Quellen unterhalb der Horizontalebene, das bedeutet  $\varepsilon < 0^\circ$ , gelingt nur in Ausnahmefällen und die Streuung der Antworten ist wesentlich größer. Allein Quelle 2 bei  $(-60^\circ, +25^\circ, 1\text{ m})$  wird real und virtuell vergleichbar, aber in beiden Fällen im Mittel um etwa  $16^\circ$  zu hoch, bewertet.

Natürlich lässt sich hier vermuten, dass die Ursache hauptsächlich in den erhöht angebrachten Lautsprechern zu suchen ist vgl. [102, S. 112 f.]. Da aber das BRS-System bei realer Kopfhörerwiedergabe teilweise ähnliche Schwächen aufweist<sup>[113,147]</sup>, kann man die fehlerhafte Elevationsabbildung nicht ausschließlich dem virtuellen Kopfhörer zuschreiben. Daran lässt sich erkennen welche Schwierigkeiten auftreten, wenn virtuelle Kopfhörer mit Versuchen zum Richtungshören getestet werden. Es ist nicht eindeutig abzugrenzen, ob die Abbildungsfehler durch den virtuellen Kopfhörer oder das vorgeschaltete binauraltechnische System verursacht werden. Das alleinige Überprüfen der korrekten Azimutabbildung, wie es in moderneren Studien oft durchgeführt wird vgl. z. B. [9,58,75,101,161,162,184], mag zwar für die jeweils beabsichtigte Anwendung ausreichen. Problematisch wird meistens jedoch die Abbildung der Elevation sein, da sie wesentlich stärker auf eine korrekte Reproduktion spektraler Merkmale angewiesen ist<sup>[157, S. 14 f.]</sup>.

Aus diesem Grund wurde ein weiteres Experiment konzipiert und durchgeführt. Dabei wurde die Wahrnehmung von 72 verschiedenen, mit statischer Binauraltechnik erzeugten Schalleinfallrichtungen sowohl bei realer als auch bei virtueller Kopfhörerwiedergabe mit einem übersprechkompensierten Stereo-Lautsprecherpaar überprüft (sh. Abschn. 6.1).

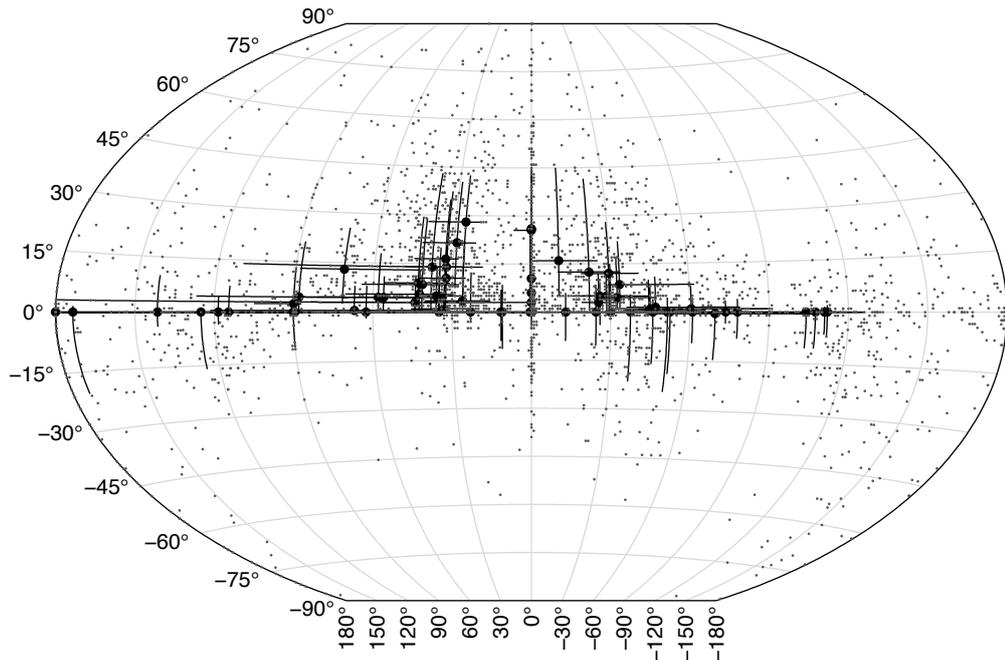
Angelehnt an die bei der Darstellung von Weltkarten benutzten Techniken sind die kopfbezogenen Koordinaten Azimut und Elevation in Abb. 7.2 und 7.2 mit einer Winkel-Tripel-Projektion veranschaulicht, wobei die Längengrade dem Azimut und die Breitengrade der Elevation entsprechen. In den Resultaten offenbaren sich eindeutige Schwächen der statischen Binauraltechnik:



**Abbildung 7.2** Ergebnis des Hörversuchs zur Lokalisation von mit statischer Binauraltechnik simulierten Quellen bei Kopfhörerwiedergabe, vgl. Abschn. 6.1 ab S. 67. Die Grafik zeigt sowohl die Einzelurteile ( $\cdot$ ), als auch Mediane und Interquartilbereiche der gültigen Antworten aller Versuchspersonen für jede der 72 simulierten Schalleinfallrichtungen ( $\bullet$ ), aufgetragen in einer Winkel-Tripel-Projektion des kopfbezogenen Koordinatensystems.

Abb. 7.2 zeigt, dass bei Kopfhörerwiedergabe für die untere Hälfte der vorderen Hemisphäre, also für  $|\alpha| < 90^\circ$  und  $\varepsilon < 0^\circ$ , kaum ein Einzelurteil abgegeben wurde, obwohl dieser Bereich eigentlich 16 von 72 simulierten Einfallrichtungen einschließt. Vier Versuchspersonen haben die gesamte vordere Halbkugel sogar fast vollständig ausgespart und damit nur von hinten einfallenden Schall gehört. Für Schalleinfall von vorne, also  $\alpha = 0^\circ$ , trat dementsprechend in etwa der Hälfte aller Fälle eine Richtungsinversion auf: über alle Elevationen lagen von 432 möglichen Urteilen 210 im Bereich  $|\alpha| \geq 160^\circ$  und nur 130 im Bereich  $|\alpha| \leq 20^\circ$ , die als korrekt in Bezug auf den Azimut angesehen werden können. Die restlichen 92 Urteile befinden außerhalb dieser beiden Bereiche und weisen damit eine deutliche seitliche Abweichung von mehr als  $20^\circ$  auf. Schalleinfall von hinten, also  $\alpha = 180^\circ$ , schneidet besser ab: hier waren 299 Urteile korrekt, 86 können als Richtungsinversion und 47 als Richtungsabweichung eingeordnet werden. Im Median wurde der Azimut für Schalleinfall von vorne und hinten aber immer richtig angegeben.

Weiterhin fällt auf, dass – abgesehen von den bereits erwähnten Werten für frontalen Schalleinfall – kaum ein Median in der vorderen Hemisphäre liegt. Ein Großteil der

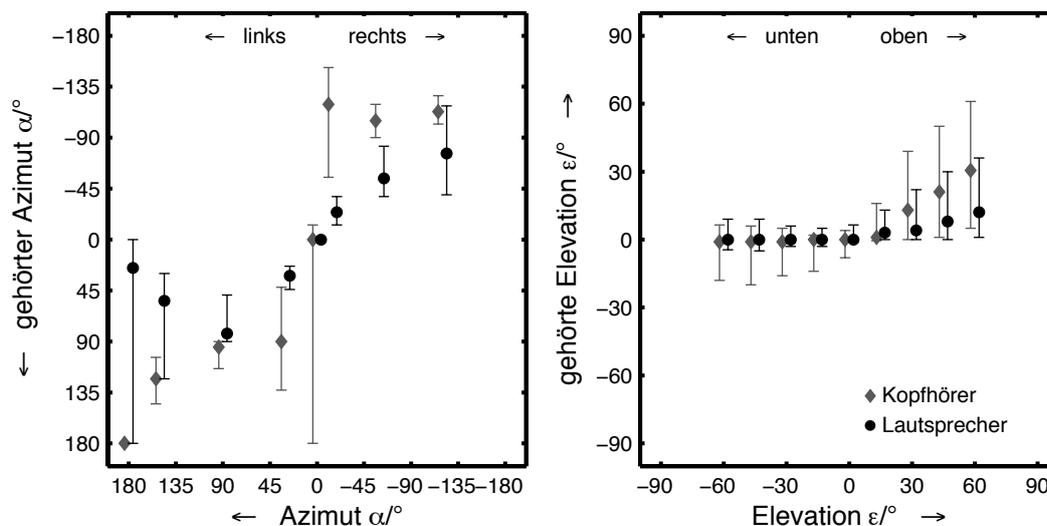


**Abbildung 7.3** Ergebnis des Hörversuchs zur Lokalisation von mit statischer Binauraltechnik simulierten Quellen bei Lautsprecherwiedergabe, vgl. Abschn. 6.1 ab S. 67. Die Grafik zeigt sowohl die Einzelurteile (·), als auch Mediane und Interquartilbereiche der gültigen Antworten aller Versuchspersonen für jede der 72 simulierten Schalleinfallrichtungen (●), aufgetragen in einer Winkel-Tripel-Projektion des kopfbezogenen Koordinatensystems.

Resultate konzentriert sich bei  $\alpha \approx \pm 120^\circ$ , die vorderen Einfallrichtungen aus den Seitenwinkeln  $+30^\circ$ ,  $-15^\circ$  und  $-60^\circ$  werden im Mittel alle hinten wahrgenommen.

Die Bandbreite der Elevation mit intendierten Werten von  $|\epsilon| \leq 60^\circ$  wird ebenfalls nicht vollständig abgebildet. Zwar erstrecken sich die Einzelurteile zumindest in der hinteren Hemisphäre über den gesamten Bereich, im Mittel wurde aber kein Schalleinfall mit einer Elevation  $\epsilon < -12^\circ$  bewertet.

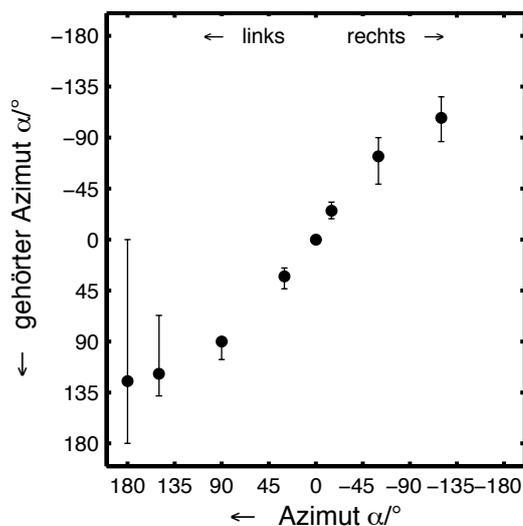
Im Unterschied dazu orientieren sich die mittleren Urteile bei virtueller Kopfhörerwiedergabe sehr stark an der Position der beiden Sekundärquellen bei  $\alpha \approx \pm 35^\circ$ , wie man in Abb. 7.3 erkennt. Und obwohl die Einzelurteile keine derart deutliche Lücke wie in Abb. 7.2 aufweisen, ist die Elevationsabbildung noch stärker limitiert. Alle Mediane bewegen sich im Bereich  $0^\circ \leq \epsilon \leq +28^\circ$ . Es wurden aber korrekte mittlere Richtungsurteile mit seitlichen und rückwärtigen Azimuten  $|\alpha| \geq 90^\circ$  abgegeben. Das virtuelle Kopfhörersystem war also in der Lage, Schalleinfall auf Ohrhöhe aus Richtungen weit außerhalb seiner Stereobasis darzustellen (vgl. Abb. 7.5).



**Abbildung 7.4** Gesamtergebnis des Hörversuchs zur Lokalisation von mit statischer Binauraltechnik simulierten Quellen, vgl. Abschn. 6.1 ab S. 67. Angegeben sind Mediane und Interquartilbereiche der gültigen Antworten aller Versuchspersonen für reale (◆) und virtuelle (●) Kopfhörerwiedergabe. Links ist über dem tatsächlichen Azimut der Quelle der gehörte Azimut aufgetragen – die *umgekehrte Richtung der Koordinatenachsen* ergibt sich aus der Definition des kopfbezogenen Koordinatensystems. Rechts ist entsprechend die Elevation dargestellt.

Schalleinfall direkt von hinten, also  $\alpha = 180^\circ$ , wurde jedoch in mehr als der Hälfte aller Fälle von vorne wahrgenommen: von ebenfalls 432 möglichen Urteilen über alle Elevationen waren hier bezüglich des Azimuts lediglich 139 korrekt, 251 lassen sich als Inversion und 42 als Abweichung charakterisieren. Bei Schalleinfall von vorne, also  $\alpha = 0^\circ$ , traten hingegen nur 56 Inversionen und 32 Abweichungen auf.

Fasst man das Ergebnis des gesamten Hörversuchs nach dem Vorbild von Abb. 7.1 kompakt zusammen, erhält man Abb. 7.4, in der reale und virtuelle Kopfhörerwiedergabe direkt verglichen werden können. Dafür wurde im linken Teil der mit Median und Interquartilbereich angegebene gehörte Azimut über alle Elevationen, und im rechten Teil analog die gehörte Elevation über alle Azimute, gemittelt. Deutlich zu sehen ist in Abb. 7.4 links, dass sich die wahrgenommene Schalleinfallrichtung bei realer Kopfhörerwiedergabe fast ausschließlich auf der hinteren Hemisphäre  $|\alpha| > 90^\circ$ , und bei virtueller Kopfhörerwiedergabe vollständig auf der vorderen Hemisphäre  $|\alpha| < 90^\circ$  verteilt. Abb. 7.4 rechts stellt noch einmal die Schwächen der Elevationsabbildung heraus. Ähnlich wie beim Versuch mit dem Prototyp (vgl. Abb. 7.1) gibt es hier im Mittel weder bei realer noch bei virtueller Kopfhörerwiedergabe Einfallrichtungen unterhalb der Ohrhöhe, also  $\varepsilon < 0^\circ$ .



**Abbildung 7.5**

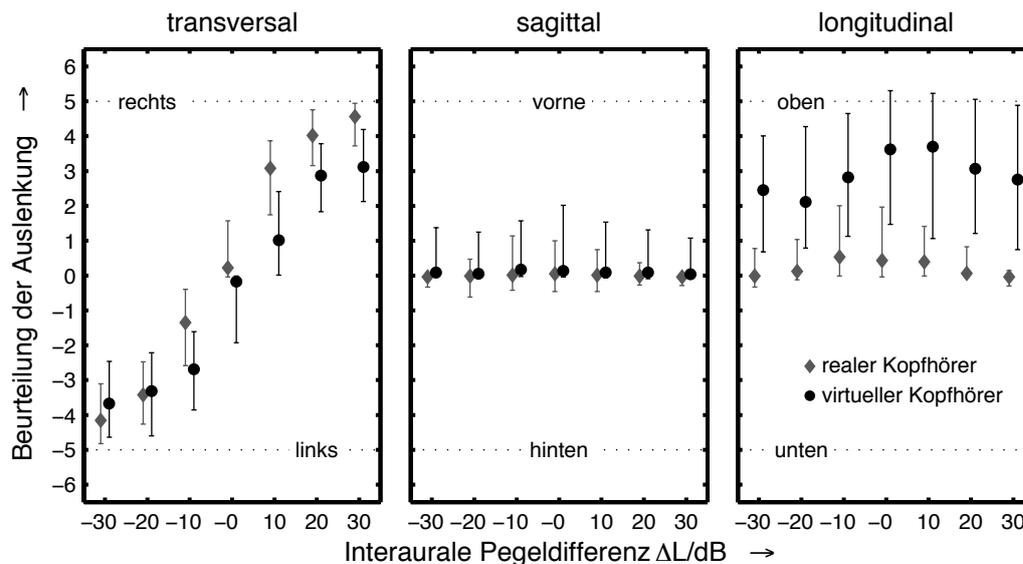
Selektiertes Teilergebnis des Hörversuchs zur Lokalisation von mit statischer Binauraltechnik simulierten Quellen, vgl. Abschn. 6.1 ab S. 67. Dargestellt sind Mediane und Interquartilbereiche der gehörten Azimute über dem tatsächlichen Azimut für Schalleinfall in der Horizontalebene, d. h.  $\varepsilon = 0^\circ$ , bei Lautsprecherwiedergabe – wiederum mit *umgekehrten Achsen*.

Abb. 7.5 beweist abschließend, dass die virtuelle Kopfhörerwiedergabe aber ein sehr gutes Ergebnis in Bezug auf die Abbildung des Seitenwinkels liefert, wenn man ausschließlich die Horizontalebene betrachtet. Lediglich die Wahrnehmung von direkt rückwärtigem Schalleinfall aus  $\alpha = 180^\circ$  zeigt sich in diesem Fall noch ziemlich instabil und könnte möglicherweise durch zusätzliche Sekundärquellen hinter dem Hörer oder ein dynamisches System verbessert werden (sh. Abschn. 2.4).

Das zweite Lokalisationsexperiment unterstreicht die bekannten Probleme beim Einsatz statischer Binauraltechnik: Obwohl das Experiment gemäß dem aktuellen Stand der Technik implementiert war, führten Richtungsinversonen und Richtungsabweichungen dazu, dass Schalleinfallrichtungen für bestimmte Teilbereiche des kopfbezogenen Koordinatensystems nur schwer oder überhaupt nicht zu simulieren waren. Anhand der vorliegenden Versuchsdaten kann nicht beurteilt werden, ob sich die Ergebnisse beispielsweise durch den Austausch des Kunstkopffmodells oder individuell ausgewählte Ohrsignale verbessern lassen. Die folgenden Versuche sollten deswegen ohne einen binauraltechnischen Verarbeitungsschritt auskommen.

### Lateralisation

Dem am Anfang dieses Kapitels propagierten Vorschlag entsprechend wurde der Prototyp im nächsten Schritt einem Lateralisationsexperiment unterzogen, dessen Ergebnisse in Abb. 7.6 zusammengefasst dargestellt sind. Die Versuchspersonen bewerteten die Auslenkung von insgesamt vier synthetischen Grundstimuli, die ihnen mit verschiedenen interauralen Pegeldifferenzen vorgespielt wurden, in den



**Abbildung 7.6** Gesamtergebnis des Hörversuchs zur Beurteilung der Im-Kopf-Lokalisation und Auslenkung eines Hörereignisses in den drei Dimensionen transversal, sagittal und longitudinal bei interauralen Pegeldifferenzen von bis zu 30 dB, vgl. Abschn. 5.2 ab S. 59. Gezeigt sind Mediane und Interquartilbereiche der Antworten aller Versuchspersonen auf alle Testsignale im direkten Vergleich zwischen realer (♦) und virtueller (●) Kopfhörerwiedergabe.

drei Dimensionen transversal, sagittal und longitudinal (sh. Abschn. 5.2). Da keine signifikanten Unterschiede zwischen den Grundstimuli festgestellt werden konnten, beruht die Grafik auf den kumulierten Versuchsdaten.

Der reale Kopfhörer zeigt ein erwartetes Verhalten, nämlich ein Hörereignis, das sich bei interauralen Pegeldifferenzen gleichmäßig und auf der interauralen Achse mit minimalen Abweichungen nach vorne und oben verschiebt<sup>vgl. z. B. [54, S. 308 f.]</sup>. Die Ergebnisse des virtuellen Kopfhörers weisen zwei wesentliche Unterschiede auf:

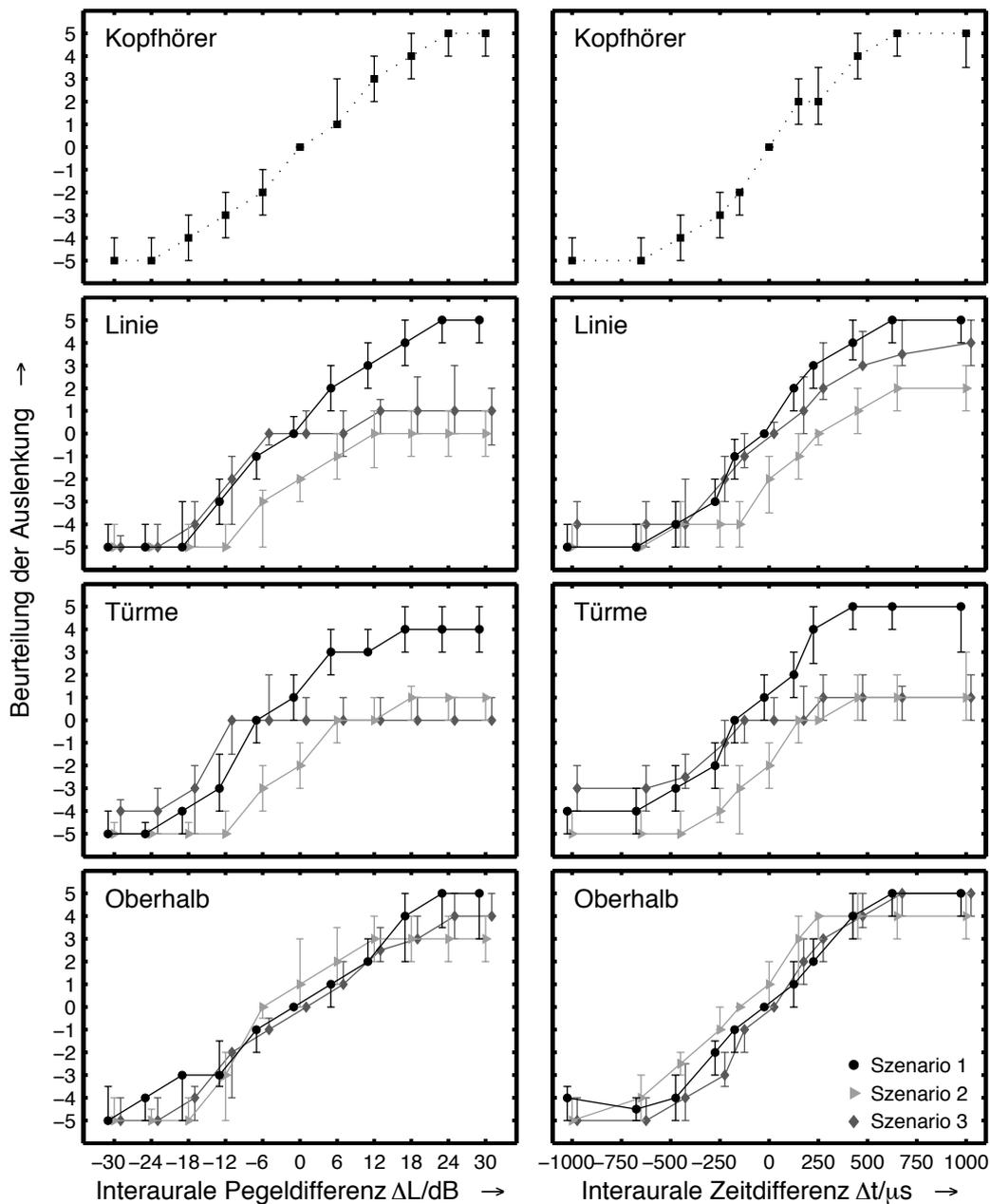
Bei der transversalen Auslenkung, Abb. 7.6 links, die üblicherweise als Lateralisation bezeichnet wird, ist der virtuelle Kopfhörer nicht in der Lage, die Hörereignisse bis an die Ohren, die sich auf der Skala bei  $\pm 5$  befinden, hier mit gestrichelten Linien gekennzeichnet, zu verschieben. Eine naheliegende Begründung ist die angegebene Kanaltrennung des Prototyps. Das System dämpft laut veröffentlichten Ergebnissen das unerwünschte Übersprechen unterhalb von 7 kHz um etwa 10 bis 20 dB<sup>[123,124]</sup>. Es ist damit also nicht fähig, Ohrsignale mit einer interauralen Pegeldifferenz von 30 dB korrekt wiederzugeben. Trotzdem scheint die Horizontalrichtungsabbildung dadurch nicht wesentlich beeinträchtigt zu sein (vgl. Abb. 7.1 links).

Während bei der sagittalen Auslenkung, Abb. 7.6 mitte, kaum Unterschiede auftreten, findet sich bei der longitudinalen Auslenkung, Abb. 7.6 rechts, möglicherweise die Erklärung der fehlerhaften Elevationsabbildung. Der virtuelle Kopfhörer verschiebt das Hörereignis im Kopf deutlich nach oben, in einzelnen Fällen wird es sogar etwas oberhalb des Kopfes wahrgenommen. Mit Hilfe des Assoziationsmodells kann man dieses Phänomen so interpretieren, dass beim Prototyp mit den erhöht angebrachten Lautsprechern noch unerwünschte Lokalisationsinformationen der Sekundärquellen übertragen werden, die für die Auslenkung der Phantomschallquelle im Kopf nach oben verantwortlich sind und damit auch die Elevationsabbildung verfälschen (vgl. Abb. 7.1 rechts).

Die Resultate zeigen, dass ein simpel konzipiertes Experiment zur Auslenkung von Hörereignissen im Kopf, das sich als Erweiterung von Lateralisationsversuchen auf die drei Dimensionen transversal, sagittal und longitudinal einordnen lässt, die Schwachstellen eines virtuellen Kopfhörers aufdecken kann. Ein potentiell fehlerbehafteter binauraltechnischer Verarbeitungsschritt lässt sich damit vermeiden.

Abb. 7.7 stellt mit acht Segmenten das Gesamtergebnis des Lateralisationsversuchs mit den drei unterschiedlichen Sekundärquellenanordnungen „Linie“, „Türme“ und „Oberhalb“ dar. Neben Pegeldifferenzen, Abb. 7.7 linker Teil, wurden hier auch Zeitdifferenzen, Abb. 7.7 rechter Teil, untersucht (sh. Abschn. 6.2 und 6.3). Als Vergleichsbasis sind in Abb. 7.7 oben die Ergebnisse für einen realen Kopfhörer angegeben. Die Abfrage der Auslenkung in drei Dimensionen konnte hier zwar nicht realisiert werden, aber das Ergebnis der transversalen Auslenkung ermöglicht bereits eine eindeutige Bewertung.

Für die Ausgangsposition, in den Teilgrafiken mit schwarzen Punkten (●) markiert und als Szenario 1 benannt, liefern alle Anordnungen dem realen Kopfhörer ähnliche Lateralisationskurven. Die Anordnung Türme weicht am stärksten ab, hier werden Stimuli ohne Pegeldifferenzen oder ohne Zeitdifferenzen im Mittel mit „1“, also schon mit einer leichten Auslenkung nach rechts bewertet. Szenario 2, dargestellt mit hellen Dreiecken (▸), offenbart eindeutig die Schwächen des in Abschn. 4.1 propagierten Funktionsprinzips. Bei diesem Telexperiment war der Kopf der Versuchsperson im Vergleich zur Ausgangsposition zwar um 30° nach rechts gedreht, die fokussierten Quellen waren aber der Kopfdrehung entsprechend repositioniert. Trotzdem lassen sich dann bei den Anordnungen Linie und Türme Hörereignisse kaum nach rechts auslenken. Die Ursache ist sicherlich in den geometrischen Randbedingungen zu suchen. Bei den beiden Anordnungen ist das rechte Ohr durch den Kopf abgeschattet von den Sekundärquellen. Man kann davon ausgehen, dass die virtuellen Quellen deshalb nicht mehr zuverlässig reproduziert werden und das ursprüngliche Übersprechkompensationsfilter  $H_0^+$  für diese Situation nicht mehr vollständig gültig ist. Das rechte Ohr kann in dem Fall offensichtlich nicht mehr selektiv beschallt werden.



**Abbildung 7.7** Gesamtergebnis des Hörversuchs zur Beurteilung der transversalen Auslenkung eines Hörereignisses (d. h. Lateralisation) bei interauralen Pegeldifferenzen bis zu 30 dB (links) oder Zeitdifferenzen bis zu 1000  $\mu\text{s}$  (rechts), vgl. Abschn. 6.3 ab S. 71. Dargestellt sind Mediane und Interquartilbereiche der Antworten aller Versuchspersonen für das Kontrollexperiment mit Kopfhörer ( $\blacksquare$ ) und die drei Sekundärquellenanordnungen Linie, Türme und Oberhalb. Vom symmetrischen Basisszenario 1 ( $\bullet$ ) aus wurde bei Szenario 2 ( $\blacktriangleright$ ) der Kopf um  $30^\circ$  gedreht und bei Szenario 3 ( $\blacklozenge$ ) die Übersprechkompensation nachjustiert.

Auch ein neu berechnetes Filter  $H_{30^\circ}^+$ , das in Szenario 3 eingesetzt wurde, kann das Ergebnis nicht verbessern, wie an den zugehörigen Lateralisationskurven mit den dunkleren Rauten (♦) erkennbar ist. Nur die Anordnung Linie nähert sich bei Zeitdifferenzen wieder dem erwarteten Verhalten an, erreicht aber im Mittel keine vollständige Auslenkung der Hörereignisse bis zu den Ohren.

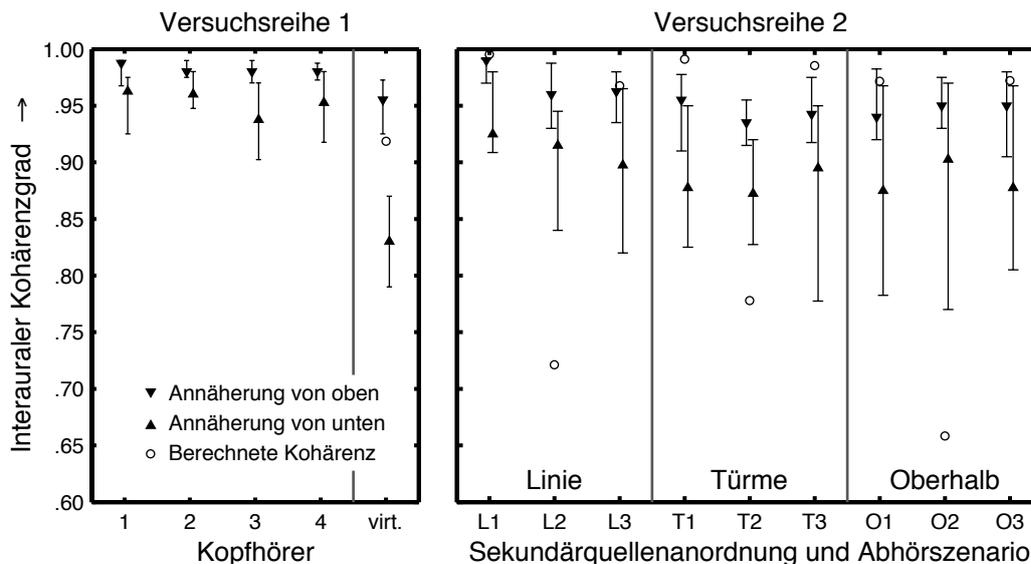
Wesentlich bessere Ergebnisse liefert die Anordnung Oberhalb: Sie zeigt zwar ebenfalls Schwächen bei Szenario 2, allerdings in geringerem Ausmaß. Auch hier ist mit Pegeldifferenzen keine vollständige Auslenkung zum rechten Ohr möglich, aber insgesamt entsprechen die Kurvenverläufe, auch bei Szenario 3, weitgehend denen des Kontroll-experiments mit dem realem Kopfhörer.

Neben der transversalen Auslenkung wurde in diesem Versuch direkt abgefragt, ob eine Im-Kopf-Lokalisation aufgetreten war. Dies bestätigten bei Pegeldifferenzstimuli im Mittel 62 %, bei Zeitdifferenzstimuli 47 % der Probanden. Die Kontrollsignale hingegen wurden mit einer Im-Kopf-Lokalisations-Quote von durchschnittlich 14 % zuverlässig erkannt. Interessant ist der Vergleich mit den Resultaten des Prototyps. Dort befinden sich sogar etwa 85 % der von den Versuchsteilnehmern in der grafischen Benutzeroberfläche festgelegten Hörereignisorte innerhalb des Kopfes. Interpretiert man diese Ergebnisse im Sinne des Assoziationsmodells, scheint der ursprüngliche Prototyp den anderen Anordnungen überlegen zu sein. Die Testschalle werden bei ihm merklich häufiger im Kopf lokalisiert, das bedeutet, er produziert selbst kaum verfälschende Lokalisationsinformationen. Vermutete Ursache dafür ist der kleinere Abstand der Sekundärquellen.

Bei der Anordnung von Sekundärquellen ist also offensichtlich darauf zu achten, dass bei Kopfdrehungen die Ohren nicht abgeschattet werden. Bewährt haben sich dafür über dem Hörer angebrachte Lautsprecher wie beim Prototyp (sh. Abb. 4.1) oder beim Versuchsaufbau Oberhalb (sh. Abb. 6.4). Damit besteht allerdings die Gefahr, dass die Elevationsabbildung beeinträchtigt wird. Zudem sollte ein möglichst geringer Sekundärquellenabstand angestrebt werden.

### **Wahrnehmbarkeit des interauralen Kohärenzgrads**

Schwieriger zu interpretieren ist das Gesamtergebnis der beiden Hörversuche zur Wahrnehmbarkeit des interauralen Kohärenzgrads (sh. Abschn. 5.3 und 6.4). Die ursprüngliche Hypothese, dass eine Korrelation zwischen der Wahrnehmungsschwelle der interauralen Kohärenz und der Wiedergabequalität des Kopfhörers besteht, muss jedoch zurückgewiesen werden. Nur die vergrößerte Darstellung lässt in Abb. 7.8 links Unterschiede zwischen den realen Kopfhörern erkennen, die in Wirklichkeit marginal und auch nicht signifikant sind. Allerdings zeigt der Prototyp



**Abbildung 7.8** Gesamtergebnis der beiden Versuche zur Wahrnehmbarkeit des interauralen Kohärenzgrads mit *angepasster* Ordinate. Die Abbildung zeigt Mediane und Interquartilbereiche der Antworten aller Versuchspersonen getrennt für die Annäherung an die Schwelle von oben ( $\blacktriangledown$ ) und von unten ( $\blacktriangle$ ). Im linken Teil sind die Ergebnisse für die Versuchsreihe 1 mit vier realen Kopfhörern (1–4) und einem virtuellen Kopfhörer (virt. = Prototyp) dargestellt, vgl. Abschn. 5.3 ab S. 62; im rechten für die Versuchsreihe 2 mit drei verschiedenen Sekundärquellenanordnungen (L, T, O) und je drei Abhörszenarien (1–3), vgl. Abschn. 6.4 ab S. 76. Zusätzlich ist für alle virtuellen Kopfhörersysteme der maximal erreichte Kohärenzgrad der resultierenden Ohrsignale, berechnet aus den gemessenen Gesamtübertragungsfunktionen, markiert ( $\circ$ ).

abweichende Resultate. Die untere Schwelle, also der Wert für die Annäherung von unten, liegt weitaus tiefer und der Wertebereich zwischen den beiden Schwellen, in dem den Probanden ein eindeutiges Urteil schwerfällt, ist größer.

Die meisten der übrigen virtuellen Kopfhörer mit den Sekundärquellenanordnungen Linie, Türme und Oberhalb, in Abb. 7.8 rechts abgekürzt mit L, T, und O bezeichnet, liefern für die drei Abhörszenarien ähnliche Resultate wie der Prototyp, wobei letzterer die im Mittel über beide Versuchsreihen niedrigste untere Schwelle aufweist. Ausnahme ist das mit „L1“ beschriftete Telexperiment, das sogar im Bereich der realen Kopfhörer angesiedelt ist. Wegen des leicht unterschiedlichen Ablaufs der Treppenstufenverfahren in den beiden Versuchsreihen sind die Ergebnisse jedoch nicht ohne Vorbehalt vergleichbar.

Eine klare Übereinstimmung mit der Auswertung der Lateralisationsversuche kann leider nicht festgestellt werden, hier liefert zum Beispiel die Anordnung Linie und

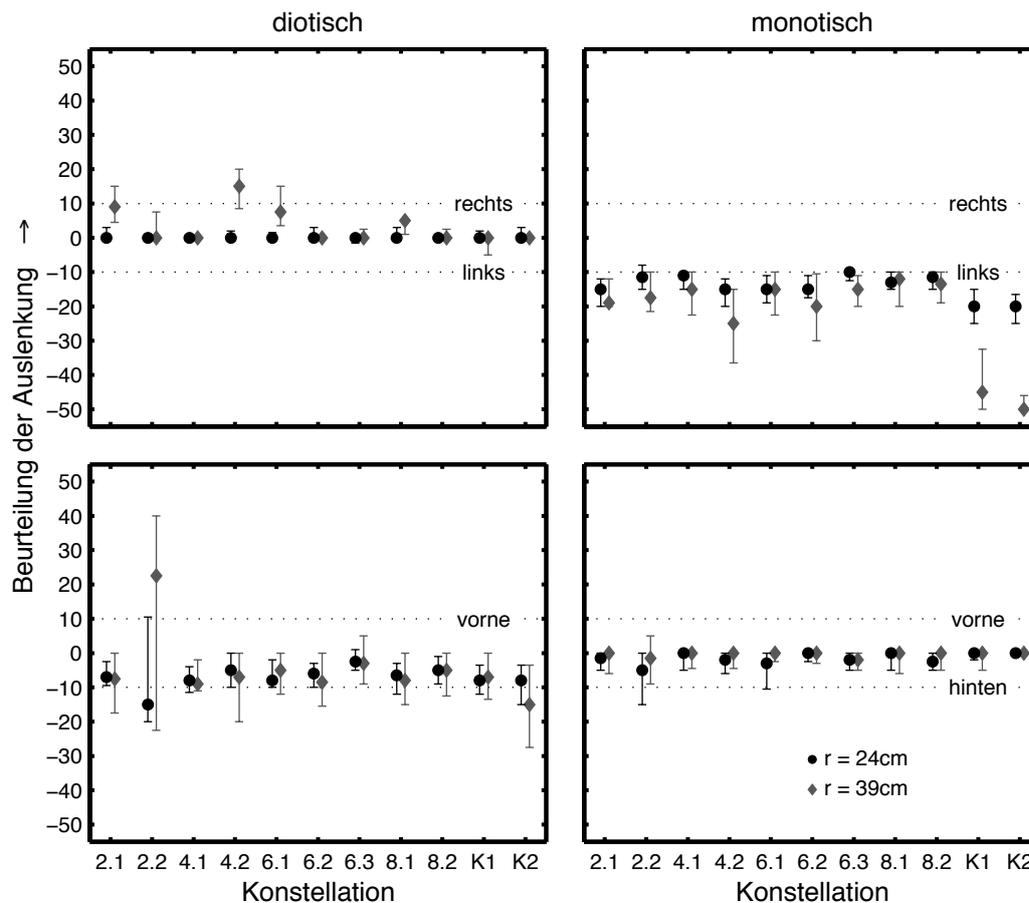
nicht die Anordnung Oberhalb die besseren Ergebnisse. Zudem sind speziell die Daten der unteren Schwellen von Versuchsreihe 2 mit einer erheblichen Streuung behaftet. Grundsätzlich kann man festhalten, dass alle getesteten virtuellen Kopfhörer in der Lage sind, mindestens einen interauralen Kohärenzgrad von etwa 0,85 abzubilden. Da bei solchen Werten bereits eine Verbreiterung des Hörereignisses auftreten kann<sup>[28,S.105 f.]</sup>, sollte im Idealfall ein Resultat wie in Telexperiment L1, das mit realen Kopfhörern vergleichbar ist, erreicht werden.

Anders als erwartet korreliert das Ergebnis nicht mit den zugehörigen Messwerten. Wenn man aus den gemessenen Gesamtübertragungsfunktionen der virtuellen Kopfhörersysteme den maximal erreichbaren Kohärenzgrad der resultierenden Ohrsignale berechnet, ergibt sich das in Abb. 7.8 mit kleinen Kreisen (○) markierte Ergebnis. Auffällig ist die Diskrepanz bei Abhörscenario 2, hier liegt die berechnete interaurale Kohärenz deutlich niedriger als die von den Probanden angegebenen Schwellenwerte. Somit wertet das Gehör statt dem Signalparameter Kohärenz vermutlich wesentlich komplexere Informationen aus.

### **Konstellation der virtuellen Quellen**

Schließlich ist noch die Frage nach der Konstellation der virtuellen Quellen im kopfbezogenen Koordinatensystem zu behandeln. Abb. 7.9 fasst die Beurteilung der transversalen und sagittalen Auslenkung für die elf getesteten Quellenkonstellationen zusammen. Wie im Versuchskonzept angedacht, werden darin auch die beiden Radien, mit denen die Konstellationen realisiert worden waren, direkt miteinander verglichen (sh. Abschn. 6.5).

Im linken Teil stellt Abb. 7.9 die Ergebnisse bei diotischer Wiedergabe, also bei Wiedergabe des gleichen Signals auf beiden Kopfhörerkanälen ohne Laufzeitdifferenzen oder Pegelunterschiede dar. Erwartet wird ein in der Kopfmittte wahrgenommenes Hörereignis, also eine mit „0“ beurteilte Auslenkung, sowohl transversal, Abb. 7.9 links oben, als auch sagittal, Abb. 7.9 links unten. Besonders auffällige Abweichungen nach rechts zeigen die Konstellationen 2.1, 4.2 und 6.1, aber nur für den größeren Radius  $r = 39\text{ cm}$ , sowie Konstellation 2.2, bei der das Hörereignis zwar mittig, aber fast ausschließlich mit häufigen Richtungsinversionen vor und hinter dem Kopf auftrat. Während beim Prototyp eine minimale Auslenkung nach vorne festgestellt wurde (vgl. Abb. 7.6 mitte), haben hier die meisten Konstellationen die Tendenz, das Hörereignis bei diotischer Wiedergabe bis an den hinteren Rand des Kopfes zu verschieben, die einzige positive Ausnahme ist die asymmetrische Konstellation 6.3. Die beiden Kontrollkonstellationen K1 und K2 reihen sich weitgehend unauffällig in das Ergebnis mit ein – abgesehen von der etwas größeren Auslenkung nach hinten für K2 beim größeren Radius.



**Abbildung 7.9** Gesamtergebnis des Hörversuchs zur Beurteilung der transversalen (oben) und sagittalen (unten) Auslenkung eines Hörereignisses bei diotischer (links) und monotonischer (rechts) virtueller Kopfhörerwiedergabe mit 11 verschiedenen Quellenkonstellationen, vgl. Abschn. 6.5. Die Abbildung zeigt für einen Abstand der Lautsprecher zum Kopfmittelpunkt von 24 cm (●) und 39 cm (◆) Mediane und Interquartilbereiche der Antworten aller Versuchspersonen auf beide Stimuli.

Dementsprechend liefert bei diotischer Wiedergabe Konstellation 2.2 mit weniger als 20 % die niedrigste Im-Kopf-Lokalisations-Quote, während für den kleineren Radius  $r = 24 \text{ cm}$  alle anderen Lautsprecheranordnungen teilweise deutlich über 50 % liegen, die Kontrollkonstellationen K1 und K2 eingeschlossen. Am besten schneiden 2.1, 6.3 und 8.2 mit etwa 80 % ab. Für den größeren Radius  $r = 39 \text{ cm}$  sind fast alle Prozentsätze kleiner: 2.2 liegt dann bei etwa 10 %, die besten Ergebnisse erreichen wieder 6.3 und 8.2, sowie 4.1, die alle drei knapp über 50 % liegen. Damit ist 4.1 die einzige Konstellation, bei der für beide Radien der Wert nahezu gleich ist.

Im rechten Teil von Abb. 7.9 sind die Beurteilungen bei monotischer Wiedergabe ausgewertet. Das heißt, die Signale wurden nur in einen – hier in den linken – Kopfhörerkanal eingespeist. Bei realer Kopfhörerwiedergabe werden Hörereignisorte am Ohr des aktiven, also linken, Kanals erwartet. Bei virtueller Kopfhörerwiedergabe wurden bisher Stimuli mit endlicher Pegeldifferenz bis zu  $\pm 30\text{ dB}$  getestet. Sie führen zu einer Verschiebung des Hörereignisses innerhalb des Kopfes, wobei nicht immer eine seitliche Auslenkung bis zum Ohr beobachtet werden konnte (vgl. Abb. 7.6 links und 7.7). Im Unterschied dazu ist die Pegeldifferenz zwischen den virtuellen Kopfhörerkanälen bei monotischer Wiedergabe gewissermaßen unendlich groß. In diesem Fall bewegt sich das Hörereignis, laut Abb. 7.9 rechts oben, vom Kopf weg nach außen. Bei den Kontrollkonstellationen K1 und K2, die ja keine korrekte Übersprechkompensation aufweisen, liegt es sogar jeweils fast am Ende der Skala bei  $-25$  beziehungsweise  $-50$ , also weit links neben dem Kopf. Hier identifizierten die Probanden die sekundären Quellen als Ursprung des Hörereignisses. Sagittal ist die Abweichung von der transversalen Achse hingegen geringer als bei diotischer Wiedergabe, wie man beim Vergleich von Abb. 7.9 rechts unten mit Abb. 7.9 links unten erkennt.

Bei monotischer Wiedergabe liegt die Im-Kopf-Lokalisations-Quote für beide Radien maximal bei nur knapp über 30%. Diese niedrigen Werte lassen sich damit erklären, dass für eine Versuchsperson „am Ohr“ nicht zwangsläufig auch „im Kopf“ bedeuten muss. Die Kontrollkonstellationen unterscheiden sich signifikant mit Quoten um 0%, was die extreme Beurteilung ihrer transversalen Auslenkung bestätigt. Insgesamt erreichen die Konstellationen 4.1, 6.3 und 8.2 beim kleineren Radius  $r = 24\text{ cm}$  gute Ergebnisse, Konstellation 8.1 schneidet beim größeren Radius  $r = 39\text{ cm}$  noch etwas besser ab.

Zusammengefasst lassen sich folgende Punkte festhalten:

- Ceteris paribus ist – wie auch schon in anderen Studien festgestellt wurde (sh. Abschn. 2.4 und 6.5) – ein Einfluss der Quellenkonstellation zu erkennen.
- Bei monotischer Wiedergabe lassen sich bei den Kontrollkonstellationen K1 und K2 die Lautsprecher identifizieren. Eine hohe Kanaltrennung, wie sie hier bei kopfnahen Quellen durch Abschattung auftritt, reicht alleine also nicht aus, um eine virtuelle Kopfhörerwiedergabe zu erzeugen. Und obwohl bei K1 sogar die Direktpfade entzerrt sind, erhält das Gehör in beiden Fällen ausreichende Lokalisationsinformationen, um die ursprünglichen Schallquellen zu erkennen. Der Vergleich von 2.1 und K2 zeigt, dass sich das Resultat deutlich ändert, wenn die Übersprechkompensation korrekt umgesetzt ist.
- Bei diotischer Wiedergabe verlagern sich die Hörereignisse bei Konstellation

2.2 auch in Richtung der Lautsprecher, zudem treten Richtungsinversonen auf. Gemäß Assoziationsmodell ist diese zwischen Stereo-Dipol und Stereo-Standardaufstellung einzuordnende Konstellation für eine virtuelle Kopfhörerwiedergabe ebenfalls nicht geeignet, da sie sogar doppeldeutige Lokalisationsinformationen entstehen lässt.

- Daneben fallen vor allem beim größeren Radius  $r = 39\text{ cm}$  einige der übrigen Konstellationen mit Abweichungen auf, insbesondere 4.2, deren Lautsprecher bei Azimuten von  $\pm 60^\circ$  und  $\pm 120^\circ$  platziert sind. Allgemein ist der kleinere Radius zu bevorzugen.
- Die insgesamt besten Ergebnisse, die in allen Fällen weitgehend den erwarteten Hörereignissen entsprechen, liefern die Konstellationen 8.1 und 8.2 sowie die asymmetrische Konstellation 6.3.

Es werden zwar auch von einigen Konstellationen mit zwei oder vier Lautsprechern zufriedenstellende Resultate erreicht, die virtuelle Kopfhörerwiedergabe verbessert sich aber durch eine höhere Anzahl von Quellen mit geringerer Entfernung zum Kopf. Bei der Positionierung muss weder die transversale Achse vermieden, wie in der in Abschn. 6.5 erwähnten Studie<sup>[75]</sup> gefordert, noch auf Symmetrie geachtet werden.

Nimmt man an, dass sich diese Ergebnisse auf den Prototyp übertragen lassen, wird eine Konstellation mit acht fokussierten Quellen empfohlen, die bei einem kopfbezogenen Radius von  $24\text{ cm}$  gleichmäßig um den Kopf verteilt sind. Es ist bei Bedarf vertretbar, die Anzahl zu reduzieren, um zum Beispiel die Komplexität der Signalverarbeitung zu verringern. Das Filter zur Übersprechkompensation ist aber auf jeden Fall vollständig zu implementieren. Abschattungseffekte, die bei kopfnahen Quellen eine durchaus deutliche Pegelreduzierung des übersprechenden Signals hervorrufen können, reichen für eine virtuelle Kopfhörerwiedergabe nicht aus.



# 8

## Ausblick und Zusammenfassung

In Kap. 4 wurde ein spezielles Funktionsprinzip für virtuelle Kopfhörer vorgestellt, das in der Theorie eine flexible Anpassung an unterschiedliche Anwendungsszenarien erlaubt. Die Kernidee beruht dabei auf der Verwendung virtueller Schallquellen, die in weiten Grenzen beliebig generierbar und frei platzierbar sind. Sie lassen sich synchron mit einem Hörer bewegen, so dass sie an konstanten kopfbezogenen Koordinaten verharren und daher mit einer statischen Übersprechkompensation kombiniert werden können.

### 8.1 Anwendung und Erweiterung des virtuellen Kopfhörers

Der Prototyp war ursprünglich eine Konzeptstudie für den Ersatz realer Kopfhörer beim BRS-System. Als vorrangiges Anwendungsszenario wurde daher ein Einsatz an Produktionsplätzen für Mehrkanalton im Rundfunk berücksichtigt. Hier wird neben der korrekten Richtungsabbildung vor allem eine absolut klangfarbentreue und artefaktfreie Wiedergabe erwartet. Aus dieser Anforderung heraus ergab sich die vorgestellte rotationssymmetrische Anordnung mit über dem Kopf des Hörers angebrachten Sekundärquellen. Sie erlaubt die bestmögliche Kompensation von Kopfdrehungen und liefert damit die Grundlage für eine Verwendung mit dem BRS-System. Zudem bietet sie den Vorteil, dass das Blickfeld des Hörers nicht durch Lautsprecher beeinträchtigt ist. Das Verlassen des festgelegten Abhörplatzes, das heißt eine Translation des Kopfes, ist aber nicht vorgesehen.

Es bietet sich daher an, das System speziell bei sitzenden Hörern anzuwenden, da deren Bewegungsbereich situationsbedingt ausreichend eingeschränkt ist. Neben Bildschirmarbeitsplätzen lassen sich hier vor allem Anpasskabinen in der Hörgeräteakustik oder Laborexperimente aus der Audiologie in Betracht ziehen.

Im Bereich der Virtuellen Realitäten wäre der Prototyp für Flugsimulatoren oder Fahrsimulatoren geeignet, die ebenfalls sitzend bedient werden. Der Einbau in reale Fahrzeuge oder Pilotenkanzeln ist damit auch denkbar. Im Gegensatz zum Simulator

dient das System dann aber nicht mehr der Erzeugung einer Virtuellen Akustik, sondern nur noch der Information, Kommunikation und Unterhaltung. In solchen Anwendungsfällen ist es ratsam, den Aufwand für eine vielkanalige Sekundärquellenanordnung zur Erzeugung virtueller Quellen gegen den Aufwand für wenige reale Lautsprecher mit einer dynamischen oder sogar statischen Übersprechkompensation abzuwägen und in Bezug zur gewünschten Klangqualität zu setzen.

Der vorgestellte virtuelle Kopfhörer ist nicht darauf ausgelegt, die Sicherheit der Privatsphäre zu gewährleisten. Soll er zur Kommunikation dienen, beispielsweise für binaurale Telefonkonferenzen am Arbeitsplatz, muss berücksichtigt werden, dass Gespräche oder Anweisungen – anders als bei realen Kopfhörern – ohne zusätzliche Maßnahmen problemlos mitgehört werden können. Um dies effektiv zu verhindern, sind zusätzliche Methoden zur Schallfokussierung notwendig, um eine selektive Beschallung des Zielbereichs zu erreichen (vgl. Kap. 3.4).

Die interessanteste Erweiterung des Prototyps besteht darin, die gesamte Fläche oberhalb der Abhörzone mit Lautsprechern zu besetzen. Gemäß der Herleitung der Wellenfeldsynthese ist es so möglich, fast im gesamten von dieser Fläche überspannten Bereich fokussierte Quellen mit ortsunabhängiger Abstrahlcharakteristik zu erzeugen und zu bewegen. Eine derartige Sekundärquellenanordnung erlaubt es, Kopfdrehungen und Kopfbewegungen sogar für mehrere Hörer zu berücksichtigen. Wird zum Beispiel die Deckenprojektionsfläche einer Virtuellen Realität vollständig mit Lautsprechern bestückt, ist eine individuelle Beschallung aller Betrachter ohne reale Kopfhörer möglich. In Kombination mit einem entsprechend leistungsfähigen binauraltechnischen System kann damit eine auf jeden Hörer separat zugeschnittene Virtuelle Akustik erzeugt und dargeboten werden. Eine solche Lautsprecherfläche kann aber abhängig von der Dimensionierung mit einem sehr hohen technischen Aufwand verbunden sein <sup>vgl. z. B. [148]</sup>.

Basierend auf den hier vorgestellten Ideen konnte im Rahmen des EU-Förderprojektes<sup>46</sup> 3D-VIVANT<sup>47</sup> am Institut für Rundfunktechnik mittlerweile ein Wiedergabesystem mit virtuellen Kopfhörern für zwei Benutzer realisiert werden <sup>sh. auch [120,121]</sup>. Dabei sind zwanzig Sekundärquellen unterhalb eines auto-stereoskopischen Monitors entlang einer geraden Linie aufgestellt. Nach den hier beschriebenen Versuchsergebnissen (sh. Abb. 7.7) lassen sich zwar mit einem solchen Aufbau größere Kopfdrehungen wegen Abschattungseffekten nicht sauber kompensieren. Das ist jedoch wegen der Kombination mit der visuellen Darstellung auch nicht vorgesehen, das Augenmerk bei der Entwicklung des Systems lag vorrangig auf einer Mehrbenutzerstudie.

---

<sup>46</sup>Europäische Union (kurz: EU)

<sup>47</sup>Projektreferenz 248420

## 8.2 Zusammenfassung

Der Ausdruck „Virtuelle Akustik“ bezieht sich auf das Teilsystem einer Virtuellen Realität, mit dem das Gehör angesprochen wird. Im Wesentlichen lässt sich dieses Fachgebiet in zwei Teilbereiche aufspalten: erstens, kopfbezogene Verfahren, mit denen die Ohrsignale der Betrachter individuell kontrolliert werden. Sie werden üblicherweise unter dem Oberbegriff „Binauraltechnik“ zusammengefasst. Und zweitens, raumbezogene Verfahren, mit denen man stattdessen ein für alle Hörer identisches Schallfeld rekonstruiert. In beiden Fällen soll erreicht werden, dass die zur Virtuellen Realität passende auditive Wahrnehmung ausgelöst wird.

In der Binauraltechnik stellt das sogenannte „Hören mit fremden Ohren“ eine große Problematik dar. Erfolgt bei der Aufnahme und Wiedergabe von Ohrsignalen, wie zum Beispiel in der Kunstkopf-Stereofonie, keine spezifische Anpassung an den jeweiligen Hörer, können erhebliche Abbildungsfehler, wie Richtungsabweichungen, Richtungsinversionen oder Im-Kopf-Lokalisation, auftreten. Hier können dynamische binauraltechnische Systeme Abhilfe schaffen. Sie erlauben dem Hörer, sich frei in der Virtuellen Akustik zu orientieren und damit unklare Situationen aufzulösen, indem sie bei Kopfbewegungen die Ohrsignale entsprechend nachjustieren. Als Anwendungsbeispiel wurde das praxiserprobte BRS-System vorgestellt.

Anders als bei Kopfhörern erfordert die korrekte Wiedergabe binauraler Signale über Lautsprecher eine Übersprechkompensation. Die Problemstellung lässt sich in Matrix-Vektor-Notation sehr einfach beschreiben und für einen Hörer und zwei Lautsprecher exakt lösen, ansonsten kann eine Näherung ermittelt werden. Die zugrunde liegende Theorie und relevante Weiterentwicklungen, wie Stereo-Dipol, Optimale Quellenverteilung, dynamische Systeme oder die Verwendung adaptiv invertierter Filter, wurden recherchiert, gegliedert und im Überblick dargestellt.

Von den raumbezogenen Verfahren wurde die besonders in Europa populäre Wellenfeldsynthese näher beleuchtet, die sich durch praxisgerechte Vereinfachungen aus dem Kirchhoff-Helmholtz-Integral herleitet. Sie erlaubt die Erzeugung fokussierter Quellen, die beim virtuellen Kopfhörer als konzeptionelle Schallquellen dienen.

Aufbauend auf den zusammengefassten Grundlagen wurde das Funktionsprinzip des auf Wellenfeldsynthese basierenden virtuellen Kopfhörers, eine Kombination von raumbezogenen und kopfbezogenen Verfahren, detailliert erläutert und der realisierte Prototyp mit einer kreisförmigen, über dem Hörer angebrachten Sekundärquellenanordnung vorgestellt. Auch die aufwendige technische Weiterentwicklung des Prototyps hin zu einem flexibel anpassbaren System wurde kurz skizziert.

Einen besonderen Beitrag zur Forschung auf dem Gebiet der Virtuellen Akustik liefert die Untersuchung der korrespondierenden Effekte Im-Kopf-Lokalisation und Lateralisation bei virtuellen Kopfhörern. Mit Lateralisationsversuchen, auf die drei Dimensionen transversal, sagittal und longitudinal erweitert, ist eine vereinfachte Funktionsprüfung möglich, die einen eventuell fehlerbehafteten binauraltechnischen Signalverarbeitungsschritt obsolet macht. Theoretisch basiert dieses Vorgehen auf Annahmen des Assoziationsmodells: Im-Kopf-Lokalisation und Lateralisation treten auf, wenn das Gehör keine bekannten oder plausiblen Lokalisationsinformationen erkennt. Daher sollen sich die Effekte so wie bei realen auch bei virtuellen Kopfhörern zeigen. Gewünschte Lokalisationsinformationen, wie sie in binauralen Signalen enthalten sind, werden dann bei der Wiedergabe nicht verfälscht.

Am Prototyp wurde sowohl ein binauraltechnisches Lokalisationsexperiment als auch ein auf drei Dimensionen erweiterter Lateralisationsversuch durchgeführt. Es zeigte sich, dass anhand der Lateralisationskurven sowohl die Größenordnung der erreichten Übersprechkompensation als auch die bei der Elevationsabbildung auftretenden Abweichungen identifiziert werden können.

Bei drei weiteren Sekundärquellenanordnungen deckte alleine die Bewertung der transversalen Auslenkung der Hörereignisse unerwünschte Abschattungseffekte auf. Als vorteilhaft haben sich daher über dem Hörer angebrachte Sekundärquellen erwiesen. Dabei muss aber eine mögliche Richtungsabweichung von Hörereignissen nach oben in Kauf genommen werden.

In einem stark vereinfachten Lateralisationsexperiment wurden auch kopfnaher Schallquellen in verschiedenen Konstellationen, das bedeutet Anzahl und Anordnung im kopfbezogenen Koordinatensystem, miteinander verglichen. Für die Übersprechkompensation besser geeignet sind demnach mehrere, in diesem Fall sechs bis acht, um den Kopf verteilte Quellen, die auch asymmetrisch angeordnet sein dürfen. Auffallend schlecht schnitt eine Konstellation mit zwei Quellen ab, die entsprechend der Seitenwinkel zwischen Stereo-Standardaufstellung und Stereo-Dipol eingeordnet werden kann. Sie ist für eine Übersprechkompensation mit kopfnahen Quellen nicht zu empfehlen.

Die Hypothese, dass auch die Wahrnehmbarkeit des interauralen Kohärenzgrads ein Indikator für die Wiedergabequalität realer und virtueller Kopfhörer sein könnte, musste in Anbetracht der Versuchsergebnisse allerdings verworfen werden.

Anders als die sonst üblichen Vorgehensweisen bietet die Versuchsmethodik, die in dieser Dissertation beschrieben wurde, neue Möglichkeiten zur Beurteilung der Qualität virtueller Kopfhörer. Sie beruht nicht primär auf physikalischen und messtechnischen Prinzipien, sondern berücksichtigt die Funktionsweise des Gehörs. Eine

Verfeinerung dieser Methodik bietet das Potential einfach zu implementierende Tests für virtuelle Kopfhörer auf der Basis bekannter Lateralisationsexperimente zu entwerfen, die keine zusätzliche Binauraltechnik benötigen.

Mögliche Anwendungen und Erweiterungen des auf Wellenfeldsynthese basierenden virtuellen Kopfhörers sind im vorhergehenden Abschn. 8.1 bereits zusammengefasst. Teile der Arbeit – insbesondere die Weiterentwicklung des Prototyps – sind in Deutschland patentrechtlich geschützt<sup>[194]</sup>.

*Man sollte immer versuchen, alle Sachen, auch die gewöhnlichsten, die ganz selbstverständlich dazusein scheinen, mit neuen, erstaunten Augen, wie zum ersten Mal, zu sehen. Dadurch gewinnen sie ihre Erstaunlichkeit zurück, die im Selbstverständlichen eingeschlafen war, und die Welt bleibt frisch; sonst aber schläft alles ein, Leben, Freude und Staunen.*

aus

BEKENNTNISSE DES HOCHSTAPLERS FELIX KRULL

Thomas Mann, 1875–1955  
deutscher Schriftsteller



## Literatur

- [1] AHRENS, J. ; SPORS, S.: Implementation of Directional Sources in Wave Field Synthesis. In: *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, NY, USA, Oktober 2007, S. 66–69
- [2] AHRENS, J. ; SPORS, S.: Notes on Rendering Focused Directional Virtual Sound Sources in Wave Field Synthesis. In: *Fortschritte der Akustik – DAGA 2008*. Dresden, März 2008, S. 229–230
- [3] AHRENS, J. ; SPORS, S.: Reproduction of Moving Virtual Sound Sources with Special Attention to the Doppler Effect. In: *124<sup>th</sup> AES Convention*. Amsterdam, Niederlande, Mai 2008. – Convention Paper 7363
- [4] AHRENS, J. ; SPORS, S.: Sound Field Reproduction using Planar and Linear Arrays of Loudspeakers. In: *IEEE Trans. Audio, Speech, Language Process.* Vol. 18 (2010), No. 8, S. 2038–2050
- [5] AHRENS, J.: *The Single-layer Potential Approach Applied to Sound Field Synthesis Including Cases of Non-enclosing Distributions of Secondary Sources*, Technische Universität Berlin, Dissertation, 2010
- [6] BAALMAN, M. A. J.: Discretisation of Complex Sound Sources for Reproduction with Wave Field Synthesis. In: *Fortschritte der Akustik – DAGA 2005*. München, März 2005, S. 209–210
- [7] BAALMAN, M. A. J.: Reproduction of Arbitrarily Shaped Sound Sources with Wave Field Synthesis – Physical and Perceptual Effects. In: *122<sup>nd</sup> AES Convention*. Wien, Österreich, Mai 2007. – Convention Paper 7017
- [8] BAALMAN, M. A. J.: *On Wave Field Synthesis and Electro-acoustic Music, with a Particular Focus on the Reproduction of Arbitrarily Shaped Sound Sources*, Technische Universität Berlin, Dissertation, 2007
- [9] BAI, M. R. ; LEE, C.-C.: Objective and Subjective Analysis of Effects of Listening Angle on Crosstalk Cancellation in Spatial Sound Reproduction. In: *J. Acoust. Soc. Am.* Vol. 120 (2006), No. 4, S. 1976–1989
- [10] BAI, M. R. ; LEE, C.-C.: Subband Approach to Bandlimited Crosstalk Cancellation System in Spatial Sound Reproduction. In: *EURASIP J. Adv. Signal Process.* Vol. 2007 (2007). – Article ID 71948

- [11] BAUCK, J. L. ; COOPER, D. H.: Generalized Transaural Stereo and Applications. In: *J. Audio Eng. Soc.* Vol. 44 (1996), No. 9, S. 683–705
- [12] BAUCK, J. L.: A Simple Loudspeaker Array and Associated Crosstalk Canceler for Improved 3D Audio. In: *J. Audio Eng. Soc.* Vol. 49 (2001), No. 1/2, S. 3–13
- [13] BAUER, B. B.: Stereophonic Earphones and Binaural Loudspeakers. In: *J. Audio Eng. Soc.* Vol. 9 (1961), No. 2, S. 148–151
- [14] BAVU, E. ; BESNAINOU, C. ; GIBIAT, V. ; ROSNY, J. de ; FINK, M.: Subwavelength Sound Focusing using a Time-reversal Acoustic Sink. In: *Acta Acust. united Ac.* Vol. 93 (2007), No. 5, S. 706–715
- [15] BEGAULT, D. R.: Perceptual Effects of Synthetic Reverberation on Three-dimensional Audio Systems. In: *J. Audio Eng. Soc.* Vol. 40 (1992), No. 11, S. 895–904
- [16] BEGAULT, D. R.: *3-D Sound for Virtual Reality and Multimedia*. Cambridge, MA, USA : AP Professional, 1994. – ISBN 0–12–084735–3
- [17] BEGAULT, D. R.: Auditory and Non-auditory Factors that Potentially Influence Virtual Acoustic Imagery. In: *AES 16<sup>th</sup> International Conference: Spatial Sound Reproduction*. Rovaniemi, Finland, April 1999, S. 13–26
- [18] BEGAULT, D. R. ; WENZEL, E. M. ; ANDERSON, M. R.: Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-related Transfer Functions on the Spatial Perception of a Virtual Speech Source. In: *J. Audio Eng. Soc.* Vol. 49 (2001), No. 10, S. 904–916
- [19] BEGAULT, D. R. (Hrsg.): *An Anthology of Articles on Spatial Sound Techniques. Part 1 : Virtual and Binaural Audio Technologies*. New York, NY, USA : Audio Eng. Soc., 2004. – ISBN 0–937803–53–7
- [20] BERKHOUT, A. J.: A Holographic Approach to Acoustic Control. In: *J. Audio Eng. Soc.* Vol. 36 (1988), No. 12, S. 977–995
- [21] BERKHOUT, A. J. ; VRIES, D. de ; VOGEL, P.: Acoustic Control by Wave Field Synthesis. In: *J. Acoust. Soc. Am.* Vol. 93 (1993), No. 5, S. 2764–2778
- [22] BERNSCHÜTZ, B.: A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100. In: *Conference on Acoustics – AIA-DAGA 2013*. Meran, Italien, März 2013, S. 592–595

- [23] BLAUERT, J.: *Räumliches Hören*. Stuttgart : S. Hirzel, 1974. – ISBN 3-7776-0250-7
- [24] BLAUERT, J.: *Räumliches Hören : Nachschrift, Neue Ergebnisse und Trends seit 1972*. Stuttgart : S. Hirzel, 1985. – ISBN 3-7776-0410-0
- [25] BLAUERT, J.: *Räumliches Hören : 2. Nachschrift, Neue Ergebnisse und Trends seit 1982*. Stuttgart [u. a.] : S. Hirzel, 1997. – ISBN 3-7776-0738-X
- [26] BLAUERT, J.: *Spatial Hearing*. Rev. Ed. Cambridge, MA, USA [u.a.] : MIT Press, 2001. – ISBN 0-262-02413-6
- [27] BLAUERT, J.: Analysis and Synthesis of Auditory Scenes. In: BLAUERT, J. (Hrsg.): *Communication Acoustics*. Berlin [u. a.] : Springer, 2005. – ISBN 978-3-540-22162-3, S. 1-20
- [28] BLAUERT, J. ; BRAASCH, J.: Räumliches Hören. In: WEINZIERL, S. (Hrsg.): *Handbuch der Audiotechnik*. Berlin [u. a.] : Springer, 2008. – ISBN 978-3-540-34300-4, S. 87-121
- [29] BOER, K. de ; VERMEULEN, R.: Eine Anlage für einen Schwerhörigen. In: *Philips' Technische Rundschau* 4. Jg. (1939), Heft 11, S. 329-332
- [30] BOONE, M. M. ; VERHEIJEN, E. N. G. ; TOL, P. F.: Spatial Sound-field Reproduction by Wave-field Synthesis. In: *J. Audio Eng. Soc.* Vol. 43 (1995), No. 12, S. 1003-1012
- [31] BOONE, M. M. ; VERHEIJEN, E. N. G. ; JANSEN, G.: Virtual Reality by Sound Reproduction based on Wave Field Synthesis. In: *100<sup>th</sup> AES Convention*. Kopenhagen, Dänemark, Mai 1996. – Convention Paper 4145
- [32] BRIX, S. ; SPORER, T. ; PLOGSTIES, J.: CARROUSO – An European Approach to 3D-Audio. In: *110<sup>th</sup> AES Convention*. Amsterdam, Niederlande, Mai 2001. – Convention Paper 5314
- [33] BRONSTEIN, I. N. ; SEMENDJAJEW, K. A. ; MUSIOL, G. ; MÜHLIG, H.: *Taschenbuch der Mathematik*. 5. überarb. und erw. Aufl. Frankfurt a. M. : Harri Deutsch, 2001. – ISBN 3-8171-2005-2
- [34] BRUNGART, D. S. ; RABINOWITZ, W. M.: Auditory Localization of Nearby Sources. Head-related Transfer Functions. In: *J. Acoust. Soc. Am.* Vol. 106 (1999), No. 3, S. 1465-1479
- [35] BRUNGART, D. S. ; DURLACH, N. I. ; RABINOWITZ, W. M.: Auditory Localization

- of Nearby Sources. II. Localization of a Broadband Source. In: *J. Acoust. Soc. Am.* Vol. 106 (1999), No. 4, S. 1956–1968
- [36] BRUNGART, D. S.: Auditory Parallax Effects in the HRTF for Nearby Sources. In: *1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, NY, USA, Oktober 1999, S. 171–174
- [37] CECCHHI, S. ; PALESTINI, L. ; PERETTI, P. ; PIAZZA, F. ; BETTARELLI, F.: Sub-band Adaptive Crosstalk Cancellation: A Novel Approach for Immersive Audio. In: *124<sup>th</sup> AES Convention*. Amsterdam, Niederlande, Mai 2008. – Convention Paper 7373
- [38] CHANG, J.-H. ; LEE, C.-H. ; PARK, J.-Y. ; KIM, Y.-H.: A Realization of Sound Focused Personal Audio System using Acoustic Contrast Control. In: *J. Acoust. Soc. Am.* Vol. 125 (2009), No. 4, S. 2091–2097
- [39] CHENG, C. I. ; WAKEFIELD, G. H.: Introduction to Head-related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. In: *J. Audio Eng. Soc.* Vol. 49 (2001), No. 4, S. 283–294
- [40] ĆIRIĆ, D. G. ; HAMMERSHØI, D.: Coupling of Earphones to Human Ears and to Standard Coupler. In: *J. Acoust. Soc. Am.* Vol. 120 (2006), No. 4, S. 2096–2107
- [41] COOPER, D. H. ; BAUCK, J. L.: Prospects for Transaural Recording. In: *J. Audio Eng. Soc.* Vol. 37 (1989), No. 1/2, S. 3–19
- [42] CORTEEL, E.: On the Use of Irregularly Spaced Loudspeaker Arrays for Wave Field Synthesis, Potential Impact on Spatial Aliasing Frequency. In: *9<sup>th</sup> International Conference on Digital Audio Effects (DAFx-06)*. Montreal, Kanada, September 2006
- [43] CORTEEL, E.: Synthesis of Directional Sources using Wave Field Synthesis, Possibilities, and Limitations. In: *EURASIP J. Adv. Signal Process.* Vol. 2007 (2007). – Article ID 90509
- [44] CORTEEL, E. ; PELLEGRINI, R. S. ; KUHN-RAHLOFF, C.: Wave Field Synthesis with Increased Aliasing Frequency. In: *124<sup>th</sup> AES Convention*. Amsterdam, Niederlande, Mai 2008. – Convention Paper 7362
- [45] DAMASKE, P. ; WAGENER, B.: Richtungshörversuche über einen nachgebildeten Kopf. In: *Acustica* Vol. 21 (1969), Heft 1, S. 30–35

- [46] DAMASKE, P. ; MELLERT, V.: Ein Verfahren zur richtungsgetreuen Schallabbildung des oberen Halbraumes über zwei Lautsprecher. In: *Acustica* Vol. 22 (1969/1970), Heft 4, S. 153–162
- [47] DAMASKE, P.: Richtungstreue Schallabbildung über zwei Lautsprecher. In: *Fortschritte der Akustik – DAGA 1970*. Berlin, September 1970, S. 403–406
- [48] DAMASKE, P.: Head-related Two-channel Stereophony with Loudspeaker Reproduction. In: *J. Acoust. Soc. Am.* Vol. 50 (1971), No. 4 (Pt. 2), S. 1109–1115
- [49] DANIEL, P. ; FASTL, H. ; FEDTKE, T. ; GENUIT, K. ; GRABSCH, H.-P. ; NIEDERDRÄNK, T. ; SCHMITZ, A. ; VORLÄNDER, M. ; ZOLLNER, M.: Kunstkopftechnik – Eine Bestandsaufnahme. In: *Nuntius Acusticus* (2007), Nr. 6
- [50] DEUTSCHE GESELLSCHAFT FÜR AKUSTIK E.V. (Hrsg.): *Akustische Wellen und Felder*. Berlin: DEGA, 2006. (DEGA Empfehlung 101)
- [51] DICKREITER, M. ; DITTEL, V. ; HOEG, W. ; WÖHR, M. ; ARD.ZDF MEDIENAKADEMIE (Hrsg.): *Handbuch der Tonstudioteknik*. Bd. 1. 7., völlig neu bearb. und erw. Aufl. München : KG Saur, 2008. – ISBN 978–3–598–11765–7
- [52] Norm DIN EN 60268-7 August 2011. *Elektroakustische Geräte – Teil 7: Kopfhörer und Ohrhörer*
- [53] EUROPÄISCHE RUNDFUNKUNION (Hrsg.): *Listening Conditions for the Assessment of Sound Programme Material: Monophonic and Two-channel Stereophonic*. 2<sup>nd</sup> Ed. Genf, Schweiz: EBU, 1998. (EBU Tech. 3276)
- [54] FASTL, H. ; ZWICKER, E.: *Psychoacoustics*. 3<sup>rd</sup> Ed. Berlin [u. a.] : Springer, 2007. – ISBN 978–3–540–23159–2
- [55] FELDERHOFF, U. ; REICHENAUER, K. ; THEILE, G.: Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale. In: *20. Tonmeistertagung*. Karlsruhe, November 1998, S. 229–237
- [56] FLETCHER, H.: Auditory Perspective—Basic Requirements. In: *Trans. Am. Inst. Elect. Eng.* 53 (1934), No. 1, S. 9–11
- [57] GARDNER, W. G.: *3-D Audio using Loudspeakers*, Massachusetts Institute of Technology, USA, Dissertation, 1997
- [58] GARDNER, W. G.: Head Tracked 3-D Audio using Loudspeakers. In: *1997 IEEE*

*Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).*  
New Paltz, NY, USA, Oktober 1997

- [59] GENUIT, K.: Ein Beitrag zur Optimierung eines Kunstkopfaufnahmesystems. In: *12. Tonmeistertagung*. München, November 1981, S. 218–243
- [60] GEORGIU, P. G. ; MOUCHTARIS, A. ; ROUMELIOTIS, S. I. ; KYRIAKAKIS, C.: Immersive Sound Rendering using Laser-based Tracking. In: *109<sup>th</sup> AES Convention*. Los Angeles, CA, USA, September 2000. – Convention Paper 5227
- [61] GERNEMANN-PAULSEN, A. ; NEUBARTH, K. ; SCHMIDT, L. ; SEIFERT, U.: Zu den Stufen im „Assoziationsmodell“. In: *24. Tonmeistertagung*. Leipzig, November 2006, S. 378–407
- [62] GOOSSENS, S. ; STUMPNER, R. ; LAMPARTER, H.: Hochwertige Auralisation der Schallfelder in Produktionsräumen. In: *22. Tonmeistertagung*. Hannover, November 2002
- [63] GRIESINGER, D.: Equalization and Spatial Equalization of Dummy-head Recordings for Loudspeaker Reproduction. In: *J. Audio Eng. Soc.* Vol. 37 (1989), No. 1/2, S. 20–29
- [64] GRIESINGER, D.: Frequency Response Adaptation in Binaural Hearing. In: *126<sup>th</sup> AES Convention*. München, Mai 2009. – Convention Paper 7768
- [65] GULDENSCHUH, M. ; SONTACCHI, A.: Transaural Stereo in a Beamforming Approach. In: *12<sup>th</sup> International Conference on Digital Audio Effects (DAFx-09)*. Como, Italien, September 2009
- [66] GULDENSCHUH, M. ; SHAW, C. ; SONTACCHI, A.: Evaluation of a Transaural Beamformer. In: *27<sup>th</sup> Congress of the International Council of the Aeronautical Sciences (ICAS 2010)*. Nizza, Frankreich, September 2010. – Paper Number: ICAS 2010-10.9.3
- [67] GUMEROV, N. A. ; DURAISWAMI, R.: *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Amsterdam, Niederlande [u. a.] : Elsevier, 2005 (Elsevier Series in Electromagnetism). – ISBN 978-08-044371-13
- [68] HAMMER, K. ; SNOW, W. B.: Binaural Transmission System at Academy of Music in Philadelphia. In: BEGAULT, D. R. (Hrsg.): *An Anthology of Articles on Spatial Sound Techniques*. Part 1 : *Virtual and Binaural Audio Technologies*. New York, NY, USA : Audio Eng. Soc., 2004. – ISBN 0-937803-53-7, S. 97–126

- [69] HAMMERSHØI, D. ; MØLLER, H.: Methods for Binaural Recording and Reproduction. In: *Acta Acust. united Ac.* Vol. 88 (2002), No. 3, S. 303–311
- [70] HAMMERSHØI, D. ; MØLLER, H.: Binaural Technique – Basic Methods for Recording, Synthesis, and Reproduction. In: BLAUERT, J. (Hrsg.): *Communication Acoustics*. Berlin [u. a.] : Springer, 2005. – ISBN 978–3–540–22162–3, S. 223–248
- [71] HAMMERSHØI, D. ; HOFFMANN, P. F.: Control of Earphone Produced Binaural Signals. In: *Forum Acusticum 2011*. Aalborg, Dänemark, Juni/Juli 2011, S. 2235–2239
- [72] HELLBRÜCK, J. ; ELLERMEIER, W.: *Hören*. 2. aktual. und erw. Aufl. Göttingen [u. a.] : Hogrefe, 2004. – ISBN 3–8017–1475–6
- [73] HELLBRÜCK, J. ; ELLERMEIER, W. ; KOHLRAUSCH, A. ; ZEITLER, A. ; DEUTSCHE GESELLSCHAFT FÜR AKUSTIK E.V. (Hrsg.): *Kompendium zur Durchführung von Hörversuchen in Wissenschaft und industrieller Praxis*. Berlin: DEGA, 2008
- [74] HILL, P. A. ; NELSON, P. A. ; KIRKEBY, O. ; HAMADA, H.: Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems. In: *J. Acoust. Soc. Am.* Vol. 108 (2000), No. 6, S. 2901–2910
- [75] HOKARI, H. ; FURUMI, Y. ; SHIMADA, S.: A Study on Loudspeaker Arrangement in Multi-channel Transaural System for Sound Image Localization. In: *AES 19<sup>th</sup> International Conference: Surround Sound—Techniques, Technology, and Perception*. Schloss Elmau, Juli 2001. – Paper Number 1922
- [76] HORBACH, U. ; PELLEGRINI, R. S. ; FELDERHOFF, U. ; THEILE, G.: Ein virtueller Surround Sound Abhörraum im Ü-Wagen. In: *20. Tonmeistertagung*. Karlsruhe, November 1998
- [77] HORBACH, U. ; KARAMUSTAFAOGLU, A. ; PELLEGRINI, R. S. ; MACKENSEN, P. ; THEILE, G.: Design and Applications of a Data-based Auralization System for Surround Sound. In: *106<sup>th</sup> AES Convention*. München, Mai 1999. – Convention Paper 4976
- [78] HUDDE, H. ; SCHRÖTER, J.: Verbesserungen am Neumann-Kunstkopfsystem. In: *Rundfunktechn. Mitt.* Jg. 25 (1981), Heft 1, S. 1–6
- [79] INTERNATIONALE FERNMELDEUNION (Hrsg.): *Determination of the Electro-acoustical Properties of Studio Monitor Headphones*. Genf, Schweiz: ITU, 1990. (Rec. ITU-R BS.708-0)

- [80] INTERNATIONALE FERNMELDEUNION (Hrsg.): *Multichannel Stereophonic Sound System with and without Accompanying Picture*. Genf, Schweiz: ITU, 2012. (Rec. ITU-R BS.775-3)
- [81] JOST, A. ; JOT, J.-M.: Transaural 3-D Audio with User-controlled Calibration. In: *COST-G6 Conference on Digital Audio Effects (DAFx-00)*. Verona, Italien, Dezember 2000
- [82] JOT, J.-M. ; LARCHER, V. ; WARUSFEL, O.: Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony. In: *98<sup>th</sup> AES Convention*. Paris, Frankreich, Februar 1995. – Convention Paper 3980
- [83] KAERNBACH, C.: Simple Adaptive Testing with the Weighted Up–Down Method. In: *Perception & Psychophysics* Vol. 49 (1991), No. 3, S. 227–229
- [84] KIM, J. ; KIM, S. ; YOO, C. D.: A Novel Adaptive Crosstalk Cancellation using Psychoacoustic Model for 3D Audio. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* Vol. 1. Honolulu, HI, USA, April 2007, S. I-185–188
- [85] KIM, S.-M. ; WANG, S.: A Wiener Filter Approach to the Binaural Reproduction of Stereo Sound. In: *J. Acoust. Soc. Am.* Vol. 114 (2003), No. 6, S. 3179–3188
- [86] KIM, Y. ; DEILLE, O. ; NELSON, P. A.: Crosstalk Cancellation in Virtual Acoustic Imaging Systems for Multiple Listeners. In: *J. Sound Vibration* Vol. 297 (2006), No. 1–2, S. 251–266
- [87] KIM, Y. ; KO, S. ; CHOI, J.-W. ; KIM, J.: Optimal Filtering for Focused Sound Field Reproductions using a Loudspeaker Array. In: *126<sup>th</sup> AES Convention*. München, Mai 2009. – Convention Paper 7688
- [88] KIRKEBY, O. ; NELSON, P. A. ; HAMADA, H.: The „Stereo Dipole“—A Virtual Source Imaging System using Two Closely Spaced Loudspeakers. In: *J. Audio Eng. Soc.* Vol. 46 (1998), No. 5, S. 387–395
- [89] KIRKEBY, O. ; NELSON, P. A. ; HAMADA, H.: Local Sound Field Reproduction using Two Closely Spaced Loudspeakers. In: *J. Acoust. Soc. Am.* Vol. 104 (1998), No. 4, S. 1973–1981
- [90] KOHLRAUSCH, A. ; VAN DE PAR, S.: Audio-visual Interaction in the Context of Multi-media Applications. In: BLAUERT, J. (Hrsg.): *Communication Acoustics*. Berlin [u. a.] : Springer, 2005. – ISBN 978–3–540–22162–3, S. 109–138

- [91] KÖRING, J. ; SCHMITZ, A.: Simplifying Cancellation of Cross-talk for Playback of Head-related Recordings in a Two-speaker System. In: *Acustica* Vol. 79 (1993), No. 3, S. 221–232
- [92] KOTORYNSKI, K.: Digital Binaural/Stereo Conversion and Crosstalk Cancelling. In: *89<sup>th</sup> AES Convention*. Los Angeles, CA, USA, September 1990. – Convention Paper 2949
- [93] KOYAMA, S. ; FURUYA, K. ; HIWASAKI, Y. ; HANEDA, Y.: Reproducing Virtual Sound Sources in Front of a Loudspeaker Array using Inverse Wave Propagator. In: *IEEE Trans. Audio, Speech, Language Process.* Vol. 20 (2012), No. 6, S. 1746–1758
- [94] KREBBER, W. ; GIERLICH, H.-W. ; GENUIT, K.: Auditory Virtual Environments: Basics and Applications for Interactive Simulations. In: *Signal Processing* 80 (2000), S. 2307–2322
- [95] Schutzrecht US 8,270,642 (18. September 2012). KUHN, C. ; PELLEGRINI, R. ; ROSENTHAL, M. ; CORTEEL, E. (Erfinder); Sonic Emotion AG, Oberglatt (Schweiz) (Anmelder).
- [96] KÜRER, R. ; PLENGE, G. ; WILKENS, H.: Wiedergabe von Kunstkopfsignalen über Lautsprecher. In: *radio mentor electronic* (1973), Heft 11, S. 512–513
- [97] LACOUTURE-PARODI, Y.: *A Systematic Study of Binaural Reproduction Systems through Loudspeakers: A Multiple Stereo-dipole Approach*, Aalborg Universitet, Dänemark, Dissertation, 2010
- [98] LARCHER, V. ; JOT, J.-M. ; VANDERNOOT, G.: Equalization Methods in Binaural Technology. In: *105<sup>th</sup> AES Convention*. San Francisco, CA, USA, September 1998. – Convention Paper 4858
- [99] LAUMANN, K.: Binaurale Raumsynthese: eine Zusammenfassung. In: *25. Tonmeistertagung*. Leipzig, November 2008, S. 712–722
- [100] LEE, J.-M. ; LEE, T.-W. ; PARK, J.-Y. ; KIM, Y.-H.: Generation of a Private Listening Zone; Acoustic Parasol. In: *20<sup>th</sup> International Congress on Acoustics (ICA 2010)*. Sydney, Australien, August 2010, S. 2975–2981
- [101] LENTZ, T.: Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments. In: *J. Audio Eng. Soc.* Vol. 54 (2006), No. 4, S. 283–294

- [102] LENTZ, T.: *Binaural Technology for Virtual Reality*, Rheinisch-Westfälische Technische Hochschule Aachen, Dissertation, 2007
- [103] LENTZ, T.: Near-field HRTFs. In: *Fortschritte der Akustik – DAGA 2007*. Stuttgart, März 2007, S. 677–678
- [104] LEVITT, H.: Transformed Up–Down Methods in Psychoacoustics. In: *J. Acoust. Soc. Am.* Vol. 49 (1971), No. 2 (Pt. 2), S. 467–477
- [105] LIM, J.-S. ; KYRIAKAKIS, C.: Virtual Loudspeaker Rendering for Multiple Listeners. In: *109<sup>th</sup> AES Convention*. Los Angeles, CA, USA, September 2000. – Convention Paper 5183
- [106] LINDAU, A. ; HOHN, T. ; WEINZIERL, S.: Binaural Resynthesis for Comparative Studies of Acoustical Environments. In: *122<sup>nd</sup> AES Convention*. Wien, Österreich, Mai 2007. – Convention Paper 7032
- [107] LINDAU, A. ; BRINKMANN, F.: Perceptual Evaluation of Individual Headphone Compensation in Binaural Synthesis based on Non-individual Recordings. In: *3<sup>rd</sup> ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems*. Dresden, September 2010
- [108] LINDEMANN, W. ; BLAUERT, J. ; GRUBER, K. ; BREUER, U.: Die räumliche Verteilung der Hörereignisse als Funktion des interauralen Korrelationsgrades. In: *Fortschritte der Akustik – DAGA 1985*. Stuttgart, März 1985, S. 499–502
- [109] LINDNER, F. ; VÖLK, F. ; FASTL, H.: Simulation und psychoakustische Bewertung von Übertragungsfehlern bei der Wellenfeldsynthese. In: *Fortschritte der Akustik – DAGA 2011*. Düsseldorf, März 2011, S. 663–664
- [110] LÓPEZ, J. J. ; GONZÁLEZ, A.: 3-D Audio with Dynamic Tracking for Multimedia Environments. In: *2<sup>nd</sup> COST-G6 Workshop on Digital Audio Effects (DAFx-99)*. Trondheim, Norwegen, Dezember 1999
- [111] MACKENSEN, P. ; REICHENAUER, K. ; THEILE, G.: Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören. In: *20. Tonmeistertagung*. Karlsruhe, November 1998, S. 218–228
- [112] MACKENSEN, P. ; FELDERHOFF, U. ; THEILE, G. ; HORBACH, U. ; PELLEGRINI, R. S.: Binaural Room Scanning – A New Tool for Acoustic and Psychoacoustic Research. In: *Joint Meeting ASA/EAA/DEGA – DAGA 1999*. Berlin, März 1999. – Paper 5APPA\_\_7

- [113] MACKENSEN, P. ; FRUHMANN, M. ; THANNER, M. ; THEILE, G. ; HORBACH, U. ; KARAMUSTAFAOGLU, A.: Head-tracker based Auralization Systems: Additional Consideration of Vertical Head Movements. In: *108<sup>th</sup> AES Convention*. Paris, Frankreich, Februar 2000. – Convention Paper 5135
- [114] MACKENSEN, P.: *Auditive Localization. Head Movements, an Additional Cue in Localization*, Technische Universität Berlin, Dissertation, 2004
- [115] MAJDAK, P. ; MASIERO, B. ; FELS, J.: Sound Localization in Individualized and Non-individualized Crosstalk Cancellation Systems. In: *J. Acoust. Soc. Am.* Vol- 133 (2013), No. 4, S. 2055–2068
- [116] MARTENS, W. L.: Individualized and Generalized Earphone Correction Filters for Spatial Sound Reproduction. In: *International Conference on Auditory Display (ICAD 2003)*. Boston, MA, USA, Juli 2003
- [117] MASIERO, B. ; QUI, X.: Two Listeners Crosstalk Cancellation System Modelled by Four Point Sources and Two Rigid Spheres. In: *Acta Acust. united Ac.* Vol. 95 (2009), No. 2, S. 379–385
- [118] MASIERO, B. ; DIETRICH, P. ; POLLOW, M. ; FELS, J. ; VORLÄNDER, M.: Design of a Fast Individual HRTF Measurement System. In: *Fortschritte der Akustik – DAGA 2012*. Darmstadt, März 2012, S. 329–330
- [119] MASIERO, B. ; VORLÄNDER, M.: A Framework for the Calculation of Dynamic Crosstalk Cancellation Filters. In: *IEEE/ACM Trans. Audio, Speech, Language Process.* Vol. 22 (2014), No. 9, S. 1345–1354
- [120] MEIER, M. ; LAWRENJUK, S.: Untersuchungen zur Crosstalk-Cancellation für mehrere Sweet-Spots. In: *Fortschritte der Akustik – DAGA 2012*. Darmstadt, März 2012, S. 329–330
- [121] MEIER, M.: Development of a Transaural 3D Audio Playback System for Multiple Listeners. In: *27. Tonmeistertagung*. Köln, November 2012, S. 500–509
- [122] MELCHIOR, F.: *Investigations on Spatial Sound Design based on Measured Room Impulse Responses*, Technische Universiteit Delft, Niederlande, Dissertation, 2011
- [123] MENZEL, D. ; WITTEK, H. ; THEILE, G. ; FASTL, H.: The Binaural Sky: A Virtual Headphone for Binaural Room Synthesis. In: *International Tonmeister Symposium*. Schloss Hohenkammer, Oktober/November 2005

- [124] MENZEL, D. ; WITTEK, H. ; THEILE, G. ; FASTL, H.: Binaurale Raumsynthese mittels Wellenfeldsynthese – Realisierung und Evaluierung. In: *Fortschritte der Akustik – DAGA 2006*. Braunschweig, März 2006, S. 255–256
- [125] MINNAAR, P. ; OLESEN, S. K. ; CHRISTENSEN, F. ; MØLLER, H.: Localization with Binaural Recordings from Artificial and Human Heads. In: *J. Audio Eng. Soc.* Vol. 49 (2001), No. 5, S. 323–336
- [126] MØLLER, H.: Reproduction of Artificial-head Recordings through Loudspeakers. In: *J. Audio Eng. Soc.* Vol. 37 (1989), No. 1/2, S. 30–33
- [127] MØLLER, H.: Fundamentals of Binaural Technology. In: *Applied Acoustics* Vol. 36 (1992), No. 3–4, S. 171–218
- [128] MØLLER, H. ; HAMMERSHØI, D. ; JENSEN, C. B. ; SØRENSEN, M. F.: Transfer Characteristics of Headphones Measured on Human Ears. In: *J. Audio Eng. Soc.* Vol. 43 (1995), No. 4, S. 203–217
- [129] MØLLER, H. ; JENSEN, C. B. ; HAMMERSHØI, D. ; SØRENSEN, M. F.: Design Criteria for Headphones. In: *J. Audio Eng. Soc.* Vol. 43 (1995), No. 4, S. 218–232
- [130] MØLLER, H. ; SØRENSEN, M. F. ; HAMMERSHØI, D. ; JENSEN, C. B.: Head Related Transfer Functions of Human Subjects. In: *J. Audio Eng. Soc.* Vol. 43 (1995), No. 5, S. 300–321
- [131] MØLLER, H. ; SØRENSEN, M. F. ; JENSEN, C. B. ; HAMMERSHØI, D.: Binaural Technique: Do We Need Individual Recordings? In: *J. Audio Eng. Soc.* Vol. 44 (1996), No. 6, S. 451–469
- [132] MÜLLER, S. ; MASSARANI, P.: Transfer-function Measurement with Sweeps. In: *J. Audio Eng. Soc.* Vol. 49 (2001), No. 6, S. 443–471
- [133] NELSON, P. A. ; HAMADA, H. ; ELLIOTT, S. J.: Adaptive Inverse Filters for Stereophonic Sound Reproduction. In: *IEEE Trans. Signal Process.* Vol. 40 (1992), No. 7, S. 1621–1632
- [134] NOVO, P.: Auditory Virtual Environments. In: BLAUERT, J. (Hrsg.): *Communication Acoustics*. Berlin [u. a.] : Springer, 2005. – ISBN 978-3-540-22162-3, S. 277–297
- [135] ONO, K. ; KOMIYAMA, S.: A Method of Reproducing Concert Hall Sounds

- by „Loudspeaker-Walls“. In: *102<sup>nd</sup> AES Convention*. München, März 1997. – Convention Paper 4490
- [136] OTANI, M. ; HIRAHARA, T. ; ISE, S.: Numerical Study on Source–Distance Dependency of Head-related Transfer Functions. In: *J. Acoust. Soc. Am.* Vol. 125 (2009), No. 5, S. 3253–3261. – ISSN 1520–8524
- [137] PAUL, S.: Binaural Recording Technology: A Historical Review and Possible Future Developments. In: *Acta Acust. united Ac.* Vol. 95 (2009), No. 5, S. 767–788
- [138] PELLEGRINI, R. S.: Quality Assessment of Auditory Virtual Environments. In: *International Conference on Auditory Display (ICAD 2001)*. Espoo, Finnland, Juli/August 2001, S. 161–168
- [139] PELLEGRINI, R. S. ; KUHN, C. ; GEBHARDT, M.: Headphones Technology for Surround Sound Monitoring – A Virtual 5.1 Listening Room. In: *122<sup>nd</sup> AES Convention*. Wien, Österreich, Mai 2007. – Convention Paper 7068
- [140] PIERCE, A. D.: *Acoustics*. Melville, NY, USA : Acoust. Soc. Am., 1989. – ISBN 0–88318–612–8
- [141] PLENGE, G.: Über das Problem der Im-Kopf-Lokalisation. In: *Acustica* Vol. 26 (1972), Heft 5, S. 241–252
- [142] PLENGE, G.: On the Differences between Localization and Lateralization. In: *J. Acoust. Soc. Am.* Vol. 56 (1974), No. 3, S. 944–951
- [143] PLENGE, G.: On the Behavior of Listeners to Stereophonic Sound Reproduction and the Consequences for the Theory of Sound Perception in a Stereophonic Sound Field. In: *83<sup>rd</sup> AES Convention*. New York, NY, USA, Oktober 1987. – Convention Paper 2532
- [144] RAAKE, A. ; SCHLEGEL, C. ; HOELDTKE, K. ; GEIER, M. ; AHRENS, J.: Listening and Conversational Quality of Spatial Audio Conferencing. In: *AES 40<sup>th</sup> International Conference: Spatial Audio—Sense the Sound of Space*. Tokio, Japan, Oktober 2010. – Paper Number 4-7
- [145] RÅDE, L. ; WESTERGREN, B. ; VACHENAUER, P.: *Springers Mathematische Formeln*. 3., durchges. Aufl. Berlin [u. a.] : Springer, 2000. – ISBN 3–540–67505–1
- [146] RAO, H. I. K. ; MATHEWS, V. J. ; PARK, Y.-C.: A Minimax Approach for the Joint

Design of Acoustic Crosstalk Cancellation Filters. In: *IEEE Trans. Audio, Speech, Language Process.* Vol. 15 (2007), No. 8, S. 2287–2298

- [147] RATHBONE, B. ; FRUHMANN, M. ; SPIKOFSKI, G. ; MACKENSEN, P. ; THEILE, G.: Untersuchungen zur Optimierung des BRS-Verfahrens (Binaural Room Scanning). In: *21. Tonmeistertagung*. Hannover, November 2000, S. 92–106
- [148] REUSSNER, T. ; SLADCEK, C. ; RATH, M. ; BRIX, S. ; PREIDL, K. ; SCHECK, H.: Audio Network-Based Massive Loudspeaker System for Flexible Use in Spatial Audio Research. In: *J. Audio Eng. Soc.* Vol. 61 (2013), No. 4, S. 235–245
- [149] RIPKA, A. von ; THEILE, G.: Die Beurteilung verschiedener stereofoner Wiedergabeeinrichtungen bezüglich der Abbildungsschärfe. In: *Fortschritte der Akustik – DAGA 1987*. Aachen, März 1987, S. 585–588
- [150] RUMSEY, F. (Hrsg.): *An Anthology of Articles on Spatial Sound Techniques. Part 2 : Multichannel Audio Technologies*. New York, NY, USA : Audio Eng. Soc., 2006. – ISBN 0–937803–56–1
- [151] SAMEJIMA, T. ; SASAKI, Y. ; TANIGUCHI, I. ; KITAJIMA, H.: Robust Transaural Sound Reproduction System based on Feedback Control. In: *Acoust. Sci. & Tech.* Vol. 31 (2010), No. 4, S. 251–259
- [152] SCHMITZ, A.: Vereinfachung eines Übersprechkompensators zur Wiedergabe kopfbezogener Aufnahmen über zwei Lautsprecher. In: *Fortschritte der Akustik – DAGA 1993*. Frankfurt a. M., März/April 1993, S. 502–505
- [153] SCHROEDER, M. R. ; ATAL, B. S.: Computer Simulation of Sound Transmission in Rooms. In: *1963 IEEE International Convention Record – Part 7*. New York, NY, USA, März 1963, S. 150–155
- [154] SCHROEDER, M. R.: Digital Simulation of Sound Transmission in Reverberant Spaces. In: *J. Acoust. Soc. Am.* Vol. 47 (1970), No. 2 (Pt. 1), S. 424–431
- [155] SCHULTZ, F. ; LINDAU, A. ; MAKARSKI, M. ; WEINZIERL, S.: An Extraaural Headphone for Optimized Binaural Reproduction. In: *26. Tonmeistertagung*. Köln, November 2010, S. 702–714
- [156] SEEBER, B. U. ; FASTL, H.: Subjective Selection of Non-individualized Head-related Transfer Functions. In: *International Conference on Auditory Display (ICAD 2003)*. Boston, MA, USA, Juli 2003

- [157] SEEBER, B. U.: *Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode*, Technische Universität München, Dissertation, 2003
- [158] SIVONEN, V. P. ; ELLERMEIER, W.: Directional Loudness in an Anechoic Sound Field, Head-related Transfer Functions, and Binaural Summation. In: *J. Acoust. Soc. Am.* Vol. 119 (2006), No. 5, S. 2965–2980
- [159] SMYTH, S. ; SMYTH, M. ; CHEUNG, S.: Smyth SVS Headphone Surround Monitoring for Studios. In: *AES 23<sup>rd</sup> UK Conference: Music Everywhere*. Cambridge, UK, April 2008
- [160] SNOW, W. B.: Basic Principles of Stereophonic Sound. In: *IRE Trans. Audio* Vol. 3 (1955), No. 2, S. 42–53
- [161] SONG, M.-S. ; ZHANG, C. ; FLORENCIO, D. ; KANG, H.-G.: Personal 3D Audio System with Loudspeakers. In: *2010 IEEE International Conference on Multimedia and Expo (ICME)*. Singapur, Juli 2010, S. 1600–1605
- [162] SONG, M.-S. ; ZHANG, C. ; FLORENCIO, D. ; KANG, H.-G.: Enhancing Loudspeaker-based 3D Audio with Room Modeling. In: *2010 IEEE International Workshop on Multimedia Signal Processing*. Saint-Malo, Frankreich, Oktober 2010, S. 34–39
- [163] SONTACCHI, A. ; HÖLDRICH, R.: „Getting Mixed up with WFS, VBAP, HOA, TRM ...“ From Acronymic Cacophony to a Generalized Rendering Toolbox. In: *Workshop Wave Field Synthesis – 1<sup>st</sup> DEGA Symposium*. Ilmenau, Juni 2007
- [164] SPIKOFSKI, G. ; FRUHMANN, M.: Optimisation of Binaural Room Scanning (BRS): Considering Inter-individual HRTF-characteristics. In: *AES 19<sup>th</sup> International Conference: Surround Sound—Techniques, Technology, and Perception*. Schloss Elmau, Juni 2001. – Paper Number 1907
- [165] SPORS, S.: *Active Listening Room Compensation for Spatial Sound Reproduction Systems*, Friedrich-Alexander-Universität Erlangen-Nürnberg, Dissertation, 2006
- [166] SPORS, S.: Spatial Aliasing Artifacts Produced by Linear Loudspeaker Arrays used for Wave Field Synthesis. In: *2<sup>nd</sup> International Symposium on Communications, Control, and Signal Processing (ISCCSP)*. Marrakesch, Marokko, März 2006
- [167] SPORS, S. ; RABENSTEIN, R.: Spatial Aliasing Artifacts Produced by Linear and Circular Loudspeaker Arrays used for Wave Field Synthesis. In: *120<sup>th</sup> AES Convention*. Paris, Frankreich, Mai 2006. – Convention Paper 6711

- [168] SPORS, S.: Extension of an Analytic Secondary Source Selection Criterion for Wave Field Synthesis. In: *123<sup>rd</sup> AES Convention*. New York, NY, USA, Oktober 2007. – Convention Paper 7299
- [169] SPORS, S.: Investigation of Spatial Aliasing Artifacts of Wave Field Synthesis in the Temporal Domain. In: *Fortschritte der Akustik – DAGA 2008*. Dresden, März 2008, S. 223–224
- [170] SPORS, S. ; RABENSTEIN, R. ; AHRENS, J.: The Theory of Wave Field Synthesis Revisited. In: *124<sup>th</sup> AES Convention*. Amsterdam, Niederlande, Mai 2008. – Convention Paper 7358
- [171] SPORS, S. ; AHRENS, J.: Spatial Sampling Artifacts of Focused Sources in Wave Field Synthesis. In: *International Conference on Acoustics – NAG/DAGA 2009*. Rotterdam, Niederlande, März 2009, S. 1556–1559
- [172] SPORS, S. ; AHRENS, J.: Spatial Sampling Artifacts of Wave Field Synthesis for the Reproduction of Virtual Point Sources. In: *126<sup>th</sup> AES Convention*. München, Mai 2009. – Convention Paper 7744
- [173] SPORS, S. ; WIERSTORF, H. ; GEIER, M. ; AHRENS, J.: Physical and Perceptual Properties of Focused Sources in Wave Field Synthesis. In: *127<sup>th</sup> AES Convention*. New York, NY, USA, Oktober 2009. – Convention Paper 7914
- [174] SPORS, S. ; AHRENS, J.: Reproduction of Focused Sources by the Spectral Division Method. In: *4<sup>th</sup> International Symposium on Communications, Control and Signal Processing (ISCCSP)*. Limassol, Zypern, März 2010
- [175] SPORS, S. ; AHRENS, J.: Analysis and Improvement of Pre-equalization in 2.5-Dimensional Wave Field Synthesis. In: *128<sup>th</sup> AES Convention*. London, UK, Mai 2010. – Convention Paper 8121
- [176] SPORS, S.: Schallfeldsynthese: Stand der Technik und Perspektiven. In: *Fortschritte der Akustik – DAGA 2011*. Düsseldorf, März 2011, S. 7–16
- [177] SPORS, S. ; WIERSTORF, H. ; RAAKE, A. ; MELCHIOR, F. ; FRANK, M. ; ZOTTER, F.: Spatial Sound with Loudspeakers and its Perception: A Review of the Current State. In: *Proc. IEEE* Vol. 101 (2013), No. 9, S. 1920–1938
- [178] STADE, P. ; BERNSCHÜTZ, B. ; RÜHL, M.: A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios. In: *27. Tonmeistertagung*. Köln, November 2012, S. 551–567

- [179] START, E. W. ; VALSTAR, V. G. ; VRIES, D. de: Application of Spatial Bandwidth Reduction in Wave Field Synthesis. In: 98<sup>th</sup> AES Convention. Paris, Frankreich, Februar 1995. – Convention Paper 3972
- [180] STEINBERG, J. C. ; SNOW, W. B.: Auditory Perspective—Physical Factors. In: *Trans. Am. Inst. Elect. Eng.* 53 (1934), No. 1, S. 12–17
- [181] STERN, R. M. ; TRAHOTIS, C.: Models of Binaural Interaction. In: MOORE, B. C. J. (Hrsg.): *Handbook of Perception and Cognition*. Vol. 6 : *Hearing*. New York, NY, USA : Academic Press, 1995. – ISBN 978-0-12-505626-7, S. 347–386
- [182] STRAUSS, M. ; NEIDHARDT, A.: Gestaltung der räumlichen Eigenschaften von Schallfeldern – Fokussierung von Schallereignissen. In: *Fortschritte der Akustik – DAGA 2008*. Dresden, März 2008, S. 27–28
- [183] SU, D.-J. ; HSIEH, S.-F.: Robust Crosstalk Cancellation for 3D Sound using Multiple Loudspeakers. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) Vol. 1. Honolulu, HI, USA, April 2007, S. 181–184
- [184] TAKEUCHI, T. ; NELSON, P. A. ; HAMADA, H.: Robustness to Head Misalignment of Virtual Sound Imaging Systems. In: *J. Acoust. Soc. Am.* Vol. 109 (2001), No. 3, S. 958–971
- [185] TAKEUCHI, T. ; TESCHL, M. ; NELSON, P. A.: Subjective Evaluation of the Optimal Source Distribution System for Virtual Acoustic Imaging. In: *AES 19<sup>th</sup> International Conference: Surround Sound—Techniques, Technology, and Perception*. Schloss Elmau, Juni 2001. – Paper Number 1924
- [186] TAKEUCHI, T. ; NELSON, P. A.: Extension of the Optimal Source Distribution for Binaural Sound Reproduction. In: *Acta Acust. united Ac.* Vol. 94 (2008), No. 6, S. 981–987
- [187] THEILE, G. ; PLENKE, G.: Localization of Lateral Phantom Sources. In: *J. Audio Eng. Soc.* Vol. 25 (1977), No. 4, S. 196–200
- [188] THEILE, G.: *Über die Lokalisation im überlagerten Schallfeld*, Technische Universität Berlin, Dissertation, 1980
- [189] THEILE, G.: Zur Theorie der optimalen Wiedergabe von stereofonen Signalen über Lautsprecher und Kopfhörer. In: *Rundfunktechn. Mitt.* Jg. 25 (1981), Heft 4, S. 155–170

- [190] THEILE, G.: Die Bedeutung der Diffusfeldverzerrung für die stereofone Aufnahme und Wiedergabe. In: *13. Tonmeistertagung*. München, November 1984, S. 112–124
- [191] THEILE, G.: On the Standardization of the Frequency Response of High-quality Studio Headphones. In: *J. Audio Eng. Soc.* Vol. 34 (1986), No. 12, S. 956–969
- [192] THEILE, G.: Neue Anwendung der Wellenfeldsynthese: Binaural Sky. In: *Fernseh- und Kinotechnik (FKT)* (2007), Nr. 10, S. 539–542
- [193] THEILE, G.: Wellenfeldsynthese. In: *25. Tonmeistertagung*. Leipzig, November 2008, S. 750–766
- [194] Schutzrecht DE 10 2007 032 272 (31. Juli 2014). THEILE, G. ; WITTEK, H. ; LAUMANN, K. ; FRANKE, R. ; MENZEL, D. (Erfinder); Institut für Rundfunktechnik GmbH, 80939 München (Anmelder).
- [195] TOOLE, F. E.: In-head Localization of Acoustic Images. In: *J. Acoust. Soc. Am.* Vol. 48 (1970), No. 4 (Pt. 2), S. 943–949
- [196] TOOLE, F. E.: The Acoustics and Psychoacoustics of Headphones. In: *AES 2<sup>nd</sup> International Conference: The Art and Technology of Recording*. Anaheim, CA, USA, Mai 1984. – Paper Number C1006
- [197] TREUTWEIN, B.: Adaptive Psychophysical Procedures. In: *Vision Res.* Vol. 35 (1995), No. 17, S. 2503–2522
- [198] VÖLK, F. ; KERBER, S. ; FASTL, H. ; REIFINGER, S.: Design und Realisierung von virtueller Akustik für ein Augmented-Reality-Labor. In: *Fortschritte der Akustik – DAGA 2007*. Stuttgart, März 2007, S. 673–674
- [199] VÖLK, F. ; MUSIALIK, T. ; FASTL, H.: Crosstalk Cancellation between Phantom Sources. In: *126<sup>th</sup> AES Convention*. München, Mai 2009. – Convention Paper 7722
- [200] VÖLK, F. ; STRAUBINGER, M. ; ROALTER, L. ; FASTL, H.: Measurement of Head Related Impulse Responses for Psychoacoustic Research. In: *International Conference on Acoustics – NAG/DAGA 2009*. Rotterdam, Niederlande, März 2009, S. 164–167
- [201] VÖLK, F.: Externalization in Data-based Binaural Synthesis: Effects of Impulse

- Response Length. In: *International Conference on Acoustics – NAG/DAGA 2009*. Rotterdam, Niederlande, März 2009, S. 1075–1078
- [202] VÖLK, F. ; LINDNER, F. ; FASTL, H.: Primary Source Correction (PSC) in Wave Field Synthesis. In: *International Conference on Spatial Audio (ICSA)*. Detmold, November 2011 (Kob, M. (Hrsg.): *Proceedings of ICSA 2011*. Detmold : Verband Deutscher Tonmeister/Erich Thienhaus Institut, Hochschule für Musik Detmold, 2011). – ISBN 978-3-9812830-2-0
- [203] VRIES, D. de: Sound Reinforcement by Wave Field Synthesis: Adaptation of the Synthesis Operator to the Loudspeaker Directivity Characteristics. In: *J. Audio Eng. Soc.* Vol. 44 (1996), No. 12, S. 1120–1131
- [204] VRIES, D. de: *Wave Field Synthesis*. New York, NY, USA : Audio Eng. Soc., 2009 (AES Monographs). – ISBN 978-0-937803-68-4
- [205] VROOMEN, J. ; BERTELSON, P. ; GELDER, B. de: Auditory-visual Spatial Interactions: Automatic versus Intentional Components. In: GELDER, B. de (Hrsg.) ; HAAN, E. de (Hrsg.) ; HEYWOOD, C. A. (Hrsg.): *Out of Mind : Varieties of Unconscious Processes*. Oxford, UK : Oxford University Press, 2001. – ISBN 0-19-850630-0, S. 140–150
- [206] WALKER, S. C. ; ROUX, P. ; KUPERMAN, W. A.: Synchronized Time-reversal Focusing with Application to Remote Imaging from a Distant Virtual Source Array. In: *J. Acoust. Soc. Am.* Vol. 125 (2009), No. 6, S. 3828–3834
- [207] WARD, D. B. ; ELKO, G. W.: Optimum Loudspeaker Spacing for Robust Crosstalk Cancellation. In: *1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* Vol. 6. Seattle, WA, USA, Mai 1998, S. 3541–3544
- [208] WARD, D. B. ; ELKO, G. W.: Effect of Loudspeaker Position on the Robustness of Acoustic Crosstalk Cancellation. In: *IEEE Signal Process. Lett.* Vol. 6 (1999), No. 5, S. 106–108
- [209] WARD, D. B.: Joint Least Squares Optimization for Robust Acoustic Crosstalk Cancellation. In: *IEEE Trans. Speech Audio Process.* Vol. 8 (2000), No. 2, S. 211–215
- [210] WEINZIERL, S.: Grundlagen. In: WEINZIERL, S. (Hrsg.): *Handbuch der Audiotechnik*. Berlin [u. a.] : Springer, 2008. – ISBN 978-3-540-34300-4, S. 1–39
- [211] WEINZIERL, S.: Aufnahmeverfahren. In: WEINZIERL, S. (Hrsg.): *Handbuch der Audiotechnik*. Berlin [u. a.] : Springer, 2008. – ISBN 978-3-540-34300-4, S. 551–607

- [212] WENZEL, E. M. ; ARRUDA, M. ; KISTLER, D. J. ; WIGHTMAN, F. L.: Localization using Nonindividualized Head-related Transfer Functions. In: *J. Acoust. Soc. Am.* Vol. 94 (1993), No. 1, S. 111–123
- [213] WENZEL, E. M.: What Perception Implies about Implementation of Interactive Virtual Acoustic Environments. In: *101<sup>st</sup> AES Convention*. Los Angeles, CA, USA, November 1996. – Convention Paper 4353
- [214] WIERSTORF, H. ; GEIER, M. ; SPORS, S.: Reducing Artifacts of Focused Sources in Wave Field Synthesis. In: *129<sup>th</sup> AES Convention*. San Francisco, CA, USA, November 2010. – Convention Paper 8245
- [215] WIERSTORF, H. ; RAAKE, A. ; GEIER, M. ; SPORS, S.: Perception of Focused Sources in Wave Field Synthesis. In: *J. Audio Eng. Soc.* Vol. 61 (2013), No. 1/2, S. 5–16
- [216] WIGHTMAN, F. L. ; KISTLER, D. J.: Headphone Simulation of Free-field Listening. I: Stimulus Synthesis. In: *J. Acoust. Soc. Am.* Vol. 85 (1989), No. 2, S. 858–867
- [217] WIGHTMAN, F. L. ; KISTLER, D. J.: Headphone Simulation of Free-field Listening. II: Psychophysical Validation. In: *J. Acoust. Soc. Am.* Vol. 85 (1989), No. 2, S. 868–878
- [218] WIGHTMAN, F. L. ; KISTLER, D. J.: Resolution of Front–Back Ambiguity in Spatial Hearing by Listener and Source Movement. In: *J. Acoust. Soc. Am.* Vol. 105 (1999), No. 5, S. 2841–2853
- [219] WIGHTMAN, F. L. ; KISTLER, D. J.: Measurement and Validation of Human HRTFs for Use in Hearing Research. In: *Acta Acust. united Ac.* Vol. 91 (2005), No. 3, S. 429–439
- [220] WILLIAMS, E. G.: *Fourier Acoustics*. San Diego, CA, USA [u. a.] : Academic Press, 1999. – ISBN 0–12–753960–3
- [221] Schutzrecht DE 10 2005 003 431 (15. Januar 2009). WITTEK, H. (Erfinder); Institut für Rundfunktechnik GmbH, 80939 München (Anmelder).
- [222] WITTEK, H. ; AUGUSTIN, T.: Räumliche Wahrnehmung von Wellenfeldsynthese: Der Einfluss von Alias-Effekten auf die Klangfarbe. In: *Fortschritte der Akustik – DAGA 2005*. München, März 2005, S. 199–200
- [223] XIE, B. ; ZHONG, X. ; YU, G. ; GUAN, S. ; RAO, D. ; LIANG, Z. ; ZHANG, C.: Report

on Research Projects on Head-related Transfer Functions and Virtual Auditory Displays in China. In: *J. Audio Eng. Soc.* Vol. 61 (2013), No. 5, S. 314–326

- [224] XU, S. ; LI, Z. ; SALVENDY, G.: Individualization of Head-Related Transfer Function for Three-Dimensional Virtual Auditory Display: A Review. In: *12<sup>th</sup> International Conference on Human-Computer Interaction (HCI International 2007)*. Peking, China, Juli 2007 (Shumaker, R. (Hrsg.): *Virtual Reality*. Lecture Notes in Computer Science Vol. 4563. Berlin [u. a.] : Springer, 2007). – ISBN 978–3–540–73334–8, S. 397–407
- [225] YANG, J. ; GAN, W.-S. ; TANG, S.-E.: Development of Virtual Sound Imaging System using Triple Elevated Speakers. In: *IEEE Trans. Consumer Electron.* Vol. 50 (2004), No. 3, S. 916–922
- [226] YON, S. ; TANTER, M. ; FINK, M.: Sound Focusing in Rooms: The Time-reversal Approach. In: *J. Acoust. Soc. Am.* Vol. 113 (2003), No. 3, S. 1533–1543
- [227] YON, S. ; TANTER, M. ; FINK, M.: Sound Focusing in Rooms. II. The Spatio-temporal Inverse Filter. In: *J. Acoust. Soc. Am.* Vol. 114 (2003), No. 6, S. 3044–3052
- [228] YU, G.-Z. ; XIE, B.-S. ; RAO, D.: Effect of Sound Source Scattering on Measurement of Near-field Head-related Transfer Functions. In: *Chinese Physics Letters* Vol. 25 (2008), No. 8, S. 2926–2929