

Dreidimensionale Lokalisierbarkeit und Differenzierbarkeit von Hörereignissen

Master-Thesis

zur Erlangung des akademischen Grades M.A.

Tobias Falke

1860589



Hochschule für Angewandte Wissenschaften Hamburg
Fakultät Design, Medien und Information
Department Medientechnik

Erstprüfer: Prof. Thomas Görne

Zweitprüfer: Prof. Dr.-Ing. Johann-Markus Batke

Hamburg, 30.09. 2016

Wir sehen die Welt nicht wie sie ist,
sondern wie wir sind. Talmud

Für meine Ehefrau.

Im Gedenken an Ingrid Holtermann.

Inhaltsverzeichnis

1	Das Gehör	8
1.1	Außen und Mittelohr	8
1.2	Innenohr	9
1.2.1	Signalwandlung an der Basilarmembran	9
1.2.2	Rückkopplungsprozesse im Innenohr	11
1.3	Der Hörnerv	12
1.3.1	Obere Olive	13
1.3.2	Colliculus inferior	13
1.3.3	Corpus geniculatum mediale	13
2	Das Richtungshören	14
2.1	Grundlegende Begriffe	14
2.2	Lokalisationsfähigkeit des Gehörs	15
2.2.1	Binaurales Richtungshören durch interaurale Pegeldifferenz . .	18
2.2.2	Binaurales Richtungshören durch interaurale Zeitdifferenzen .	20
2.2.3	Monaurales Richtungshören	20
2.3	Die Höhe als 3. Dimension	22
2.4	Entfernungshören	24
2.5	Lokalisation einer Quelle	26
2.5.1	Fehler bei interauraler Differenzlokalisierung	26
2.5.2	Auflösung der interauralen Fehllokalisierung	27
2.5.3	Intermodale Wechselwirkungen	28
2.6	Diskriminierung mehrerer Schallquellen	28
2.6.1	Hall als Wahrnehmung mehrerer Schallquellen	29
2.6.2	Grundsätzliche Voraussetzungen	30
2.6.3	Interaurale Kohärenz von Schallquellen	30
2.6.4	Binaurale Maskierungslevel-Differenz	32
2.6.5	Folgerungen für eine Differenzierbarkeit	34
2.7	Lokalisationsgenauigkeit bei mehreren Schallquellen	35
2.8	Summenlokalisierung	35
2.9	Die Lokalisierbarkeit vermindernde Störfaktoren	39
3	Messungen der interauralen Kohärenz	41
3.1	Vorüberlegungen	43
3.1.1	Anforderungen für eine automatisierte Datenauswertung . . .	43
3.1.2	Auswahl des Mess-Stimuli	44

Inhaltsverzeichnis

3.1.3	Länge des Messfensters	46
3.2	Durchführung	48
3.2.1	Kalibration des Messsystems	48
3.2.2	Messvorgang	49
3.2.3	Bewertung des subjektiven Höreindrucks	50
3.3	Datenauswertung	50
3.4	Schlussfolgerungen	50
3.4.1	Auswirkungen durch die Raumakustik	51
3.4.2	Vergleich der Wahrnehmung mit den Messwerten	51
3.4.3	Kohärenzgrade einzelner Quellen	53
3.4.4	Kohärenzgrade mehrerer Quellen	55
3.4.5	Aussagen zur Lokalisierbarkeit und Differenzierbarkeit	57
4	Modell zur Wahrnehmung eines Hörereignisses	59
4.1	Auditorische Verarbeitung im Kortex	59
4.1.1	Komplexe Reizverarbeitung	60
4.1.2	Was- und Wo-Ströme	60
4.1.3	Multimodale Wechselwirkung	62
4.1.4	Neuroanatomische Aufmerksamkeit	62
4.2	Theorie des globalen Arbeitsraums	63
4.2.1	Das Arbeitsgedächtnis	63
4.2.2	Die Aufmerksamkeit – das Spotlight	63
4.2.3	Konkurrierende Reize – die Schauspieler	65
4.2.4	Verarbeitungszentren – die Zuschauer	65
4.2.5	Das Selbst – der Direktor	66
4.2.6	Widersprüche zwischen Theorie und der Wahrnehmungspraxis	66
4.2.7	Wahrnehmung mehrerer Gedächtnisinhalte	67
4.2.8	Modifikation der Theorie des globalen Arbeitsraums	68
4.2.9	Auswirkungen auf die Differenzierbarkeit	69
4.2.10	Zusammenfassung der modifizierten Theorie	69
4.3	Auditorische Szenenanalyse	70
4.3.1	Auditorische Ströme	70
4.3.2	Stromseparation	71
4.3.3	Primitive Separation	71
4.3.4	Heuristiken der primitiven Stromseparation	72
	Herkunft	72
	Ähnlichkeit in Klangfarbe und Tonhöhe	73
	Zeitliche Nähe	73
	Guter Verlauf	74
4.3.5	Komplexe Separation	74
4.3.6	Abbildung von Hörereignissen über auditive Ströme	74
4.3.7	Auswirkungen auf die Lokalisierbarkeit und Differenzierbarkeit	75

Inhaltsverzeichnis

4.4	Wahrnehmungsmodell	76
4.4.1	Vorkortikale Verarbeitung	76
4.4.2	Kortikale Analyse	77
4.4.3	Aufmerksamkeit	77
4.4.4	Rückgekoppelte Analyse in der kortikalen Verarbeitung	77
4.4.5	Erkenntnisse aus dem Modell	78
5	Zusammenfassung der Ergebnisse	80
5.1	Begrenzungen durch kortikale Vorgänge	80
5.1.1	Wahrnehmung mehrerer Quellen	80
5.1.2	Die komplexe Lokalisationsgrenze	80
5.2	Lokalisierbarkeit	81
5.2.1	Wechselwirkungen	81
5.2.2	Allgemeine Lokalisierbarkeit	82
5.2.3	Einfluss der Raumakustik	82
	Bewertung der Akustik	82
	Die maximalen Punkt-Lokalisationsschärfe	83
5.2.4	Wechselwirkungen zwischen mehreren Hörereignissen	84
5.2.5	Lokalisierbarkeit unter speziellen Bedingungen	85
5.3	Differenzierbarkeit	88
5.3.1	Signalform	88
5.3.2	BMLD und BILD	88
5.3.3	auditive Separation	89
5.3.4	Einfluss der Akustik	89
5.4	Die dritte Dimension	90
5.5	Reproduktion von Schallfeld-Informationen	90
5.5.1	Abtastung in verschiedenen Frequenzbändern	90
5.5.2	Wahrnehmungsprozesse	91
5.5.3	Wiedergabesysteme	92
5.5.4	Möglichkeit zur verlustbehafteten Komprimierung	92
5.6	Sounddesigns und Mischung von 3D-Audio-Inhalten	92
5.6.1	Tempo und schnelle Schnitte	93
5.6.2	Entfernung und Höhe	93
5.7	Fazit	94
A	Material	96
A.1	Daten zur Messung des interauralen Kohärenzgrades	96
A.2	Quellcode des Matlabskript zur Analyse der interauralen Kohärenz	102
A.3	Protokoll zur Hörwahrnehmung bei den Messungen der interauralen Kohärenz	104
A.4	DVD	104
B	Abkürzungsverzeichnis	105

Inhaltsverzeichnis

C Glossar	106
D Danksagung	109
Abbildungsverzeichnis	110
Tabellenverzeichnis	111
Literaturverzeichnis	112

Abstract

How do humans understand soundwaves as separate auditory events in three dimensions?

To answer this question, existing research and theories of psychoacoustics have been reviewed and compared. Based on this, an experiment has been performed using a dummy head microphone to measure the interaural coherence of different sound source positions and numbers. The resulting data has been analysed and initial conclusions have been made about localizability and differentiability.

In the second approach, a model of the perception of auditory events has been created based on global workspace theory, the theory of auditory scene analysis, and evidence from psychoacoustics. The entire process that leads to auditory perception has been presented and analysed.

The results show the limitations that result from the hearing process and outside influences. Based on this, a perspective has been provided on the efficient reproduction and design of three-dimensional audio content.

Zusammenfassung

Wie nimmt der Menschen Schallquellen als Hörereignisse dreidimensional lokalisiert und voneinander differenziert wahr?

Zur Beantwortung dieser Frage werden als erster Ansatz Untersuchungen und Theorien der Psychoakustik verglichen. Darauf basierend wird eine Untersuchung durchgeführt, die die interaurale Kohärenz für die Lage und Anzahl verschiedener Schallquellen mittels Kunstkopf-Mikrofon bestimmt. Aus der Daten-Analyse kann dann eine erste Aussagen zur Lokalisationsschärfe und Differenzierbarkeit getroffen werden.

Im zweiten Ansatz wird auf Basis der Theorie des globalen Arbeitsraums, der Theorie zur auditorischen Szenenanalyse und den Aussagen der Psychoakustik ein Modell zur Hörereignis-Wahrnehmung erarbeitet. Darüber wird der gesamte Prozess abgebildet und analysiert, der zur auditiven Wahrnehmung führt.

Als Ergebnis können die Begrenzungen, die durch den Hörvorgang und durch äußere Faktoren entstehen, beschrieben werden. In einem kurzen Ausblick werden auf dieser Grundlage Ansätze zur effizienten Reproduktion und zur Gestaltung von 3-D-Audio-Inhalten entwickelt.

1 Das Gehör

Bevor eine Klang wahrgenommen werden kann, muss er in Nervenimpulse gewandelt werden. In dieser Form kann das Gehirn die Informationen verarbeiten. Wie diese sensorische Wandlung funktioniert, wird in diesem Kapitel behandelt, soweit es für die Thematik dieser Arbeit wichtig ist. Dabei liegt der Fokus auf natürlichen Hörbedingungen, wie sie im Alltag jedes normal hörenden gesunden Menschen vorliegen.

1.1 Außen und Mittelohr

Das Gehör des Menschen besteht äußerlich aus den zwei Ohren, die anatomisch unterteilt werden in äußeres Ohr, Mittelohr und Innenohr, wie in Abbildung 1.1 auf Seite 8 zu erkennen ist. Das äußere Ohr besteht aus der Ohrmuschel und dem äußeren Gehörgang (Ellermeier et al. 2008).

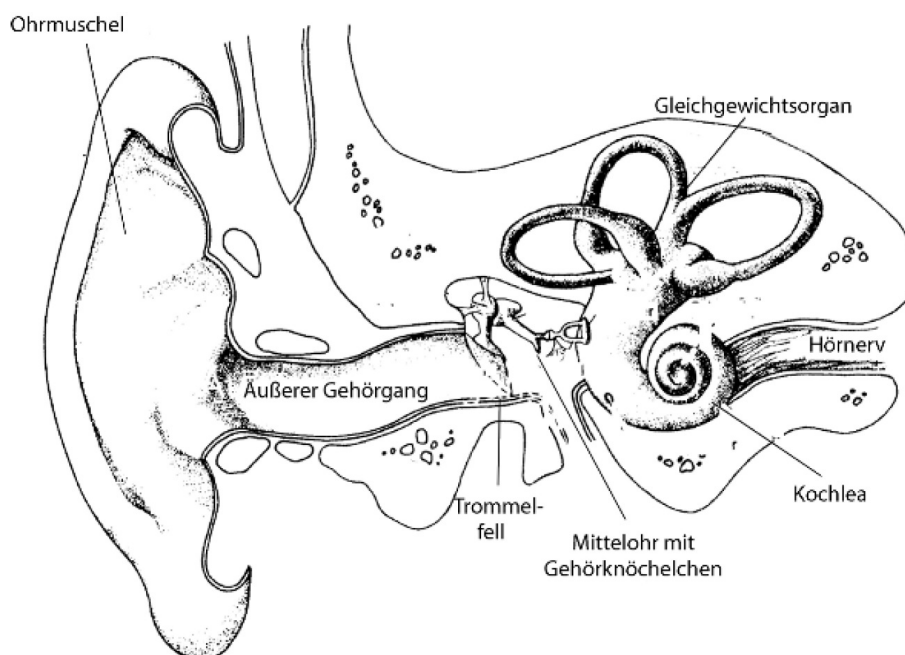


Abbildung 1.1: Anatomische Darstellung des Ohres (Ellermeier et al. 2008: S. 42)

Der äußere Gehörgang hat einen Durchmesser von etwa 8 mm. Daraus ergibt sich eine obere Grenzfrequenz von etwa 23 kHz. Darüber breiten sich die Schallwellen

im Kanal nicht mehr eben aus, was zu Verzerrung und Auslöschung führen würde (Blauert 1974). Die Grenze zwischen Außenohr und Mittelohr bildet das Trommelfell. Dieses dient der Impedanzanpassung zwischen der Luft im Außenohr und dem flüssigkeitsgefüllten Innenohr. Die Impedanzanpassung erfolgt über die Gehörknöchelchen, die die Schwingungen des Trommelfells auf das ovale Fenster übertragen (siehe Abbildung 1.2 auf S. 10). Dieses Übersetzungsverhältnis kann durch den Gehörmuskel geändert werden. Bei Pegeln von mehr als 80 dB über der Hörschwelle, die allgemein mit 20 μ Pa angenommen wird, kontrahiert der Gehörmuskel und erhöht so die Impedanz, um das Innenohr vor zu hohen Schalldrücken zu schützen. Diese Impedanz-Änderung führt jedoch zu einer nicht linearen Dämpfung. Hauptsächlich werden Frequenzen unter etwa 2 kHz gedämpft. Über dieser Grenze geht die Dämpfung schnell gegen Null (Ellermeier et al. 2008). Zudem bildet das Trommelfell einen akustischen Filter, der den empfangenen Schall auf Frequenzen von etwa 16 Hz bis 20 kHz begrenzt, wobei die obere Grenzfrequenz altersbedingt abnimmt (Ehret 2014).

1.2 Innenohr

Im Innenohr findet die eigentliche Umwandlung von Schall in Nervenimpulse statt. Das Innenohr besteht aus dem Gleichgewichtsorgan, welches Raumlage und Drehwinkel des Kopfes erfasst, sowie der Cochlea. In dieser werden die Druckwellen, die an die in ihr befindliche Flüssigkeit abgegeben wurden, in Nervenimpulse gewandelt. Die Cochlea ist ein schneckenförmiges Rohr, das zu seinem Ende leicht kegelförmig verjüngt verläuft (siehe Abbildung 1.2 auf S. 10). Am Röhrenbeginn befindet sich das ovale Fenster, abgeschlossen wird sie von der Schneckenspitze.

In der Cochlea befindet sich die Basilarmembran, auf der sich das Corti'sche Organ befindet (siehe Abbildung 1.3 auf Seite 10). Darin befinden sich etwa 3500 innere Haarzellen und etwa 12.000 äußere Haarzellen. Diese Haarzellen (Rezeptoren) sind die Sinneszellen, die den Schall in Nervenimpulse wandeln. Über die Stereozilien (auch Härchen genannt) sind die äußeren Haarzellen mit der Tektorialmembran, die das Corti'sche Organ bedeckt, verbunden. Die inneren Haarzellen haben keine oder nur lose Verbindung zur Tektorialmembran (Ellermeier et al. 2008).

1.2.1 Signalwandlung an der Basilarmembran

Durch den Druck, der über die Gehörknöchelchen auf das ovale Fenster und damit auf die Flüssigkeit in der Cochlea ausgeübt wird, beginnt die Basilarmembran zu schwingen. Durch hohe Frequenzen wird diese nahe des ovalen Fensters ausgelenkt. Tiefe Frequenzen hingegen führen zu einer Auslenkung nahe der Schneckenspitze (Ellermeier et al. 2008).

Diese Orts-Frequenz-Abhängigkeit entsteht dadurch, dass die Basilarmembran an der Basis, nahe dem ovalen Fenster, drei bis vier Mal schmaler, aber etwa hundert mal steifer ist als nahe der Schneckenspitze. Durch die Druckschwankungen entstehen auf

1 Das Gehör

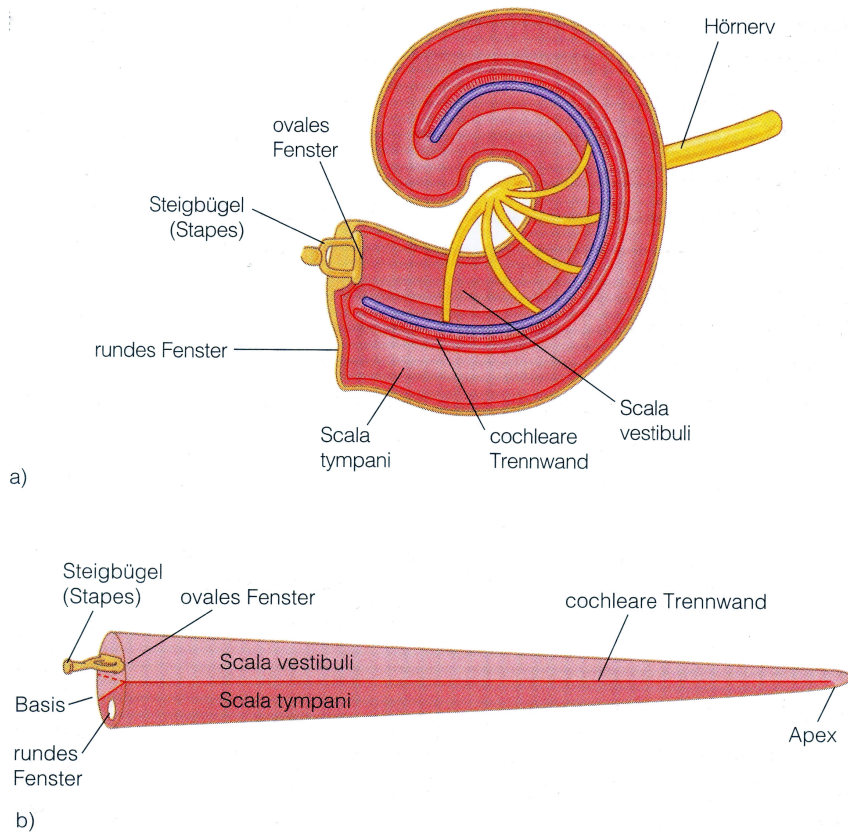


Abbildung 1.2: (a) Eine teilweise entrollte Cochlea. (b) Eine vollständig entrollte Cochlea. (Goldstein 2008: S. 269)

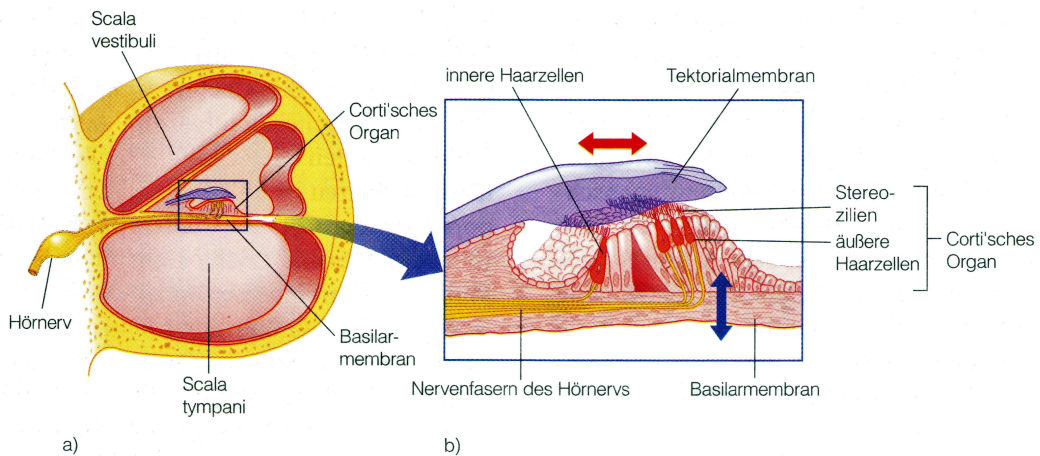


Abbildung 1.3: (a) Querschnitt durch die Cochlea. (b) Ausschnittsvergrößerung des Corti'schen Organs. Die Pfeile stellen die Bewegung der Basilar-membran und der Tektorialmembran dar, die durch die Schwingung der cochlearen Trennwand verursacht werden. (Goldstein 2008: S. 270)

der Basilarmembran Wanderwellen, die auf ihr entlanglaufen. Je nach Frequenz steilen sich diese Wellen an bestimmten Stellen durch Resonanz auf. Dieser Vorgang ist frequenz- und ortsspezifisch durch Dicke und Steifheit der Basilarmembran definiert (Goldstein 2008).

Die gesamte Energie wird dabei am jeweiligen Maximum abgegeben. Durch die so entstehende Ausbuchtung des Corti'schen Organs entstehen Scherbewegungen an den Stereozilien der Haarzellen, welche sodann ein Aktionspotential an den ihnen zugehörigen afferenten (zum zentralen Nervensystem hinführenden) Nervenfasern auslösen. Dabei haben die Haarzellen eine nur sehr kurze Regenerationszeit von etwa 1 ms. Somit löst das Gehör zeitlich sehr hoch auf (Ellermeier et al. 2008, Blauert 1974).

1.2.2 Rückkopplungsprozesse im Innenohr

Durch aktive Prozesse in den äußeren Haarzellen kann die Auslenkung der Basilarmembran sehr spitz und in einem engen Bereich stattfinden. Dadurch ist eine sehr genaue Frequenzzerlegung auf der Basilarmembran möglich (Ellermeier et al. 2008).

Die äußeren Haarzellen werden von efferenten (aus dem zentralen Nervensystem kommenden) Nerven aus dem oberen Olivenkomplex angesteuert. Durch Schwingen mit bis zu 30 kHz können sie dabei die Auslenkung der Basilarmembran verstärken. Einige efferente Nerven sind auch über Synapsen mit den Neuronen afferenter Nerven verbunden und dienen vermutlich der Dämpfung bestimmter Frequenzen. Dadurch wird zum einen die Empfindlichkeit und zum anderen die Frequenzselektivität auf der Basilarmembran signifikant erhöht. Durch die efferenten Nerven kann der obere Olivenkomplex Frequenzen selektiv verstärken, dämpfen oder die Filtergüte erhöhen, um das Gehör z.B. auf ein bestimmtes Hörereignis zu fokussieren (Ellermeier et al. 2008).

Das Gehör ist kein passives System, sondern basiert auf aktiver Rückkopplung. Zum einen kann das Gehör somit nicht wie ein Mikrofon betrachtet werden, zum anderen ist es dadurch ein nicht lineares System, das schwer zu untersuchen ist. Hinzu kommt, dass es Ermüdungserscheinungen unterliegt. Bei hohen, anhaltenden Schalldrücken kann die Cochlea nicht mehr genügend mit Glukose und Sauerstoff versorgt werden. Es bilden sich vermehrt Stoffwechselabfallprodukte, die das Gehör schädigen können. Meist kann sich das Gehör regenerieren und es tritt nur eine temporäre Schwerhörigkeit auf (Ellermeier et al. 2008). Währenddessen kommt es jedoch zu erheblichen Verschlechterungen der Lokalisationsleistung des Gehörs (Blauert 1974).

1.3 Der Hörnerv

Der Hörnerv stellt, wie in Abbildung 1.4 auf Seite 12 zu sehen ist, nicht nur eine einfache Verbindung zum auditorischen Kortex dar, er bewerkstelligt bereits einiges an Signalvorverarbeitung. Diese Verarbeitung und die dabei entstehenden Informationen sind subkortikal und stehen der Wahrnehmung erst nach der Verarbeitung im auditorischen Kortex zur Verfügung (Goldstein 2008).

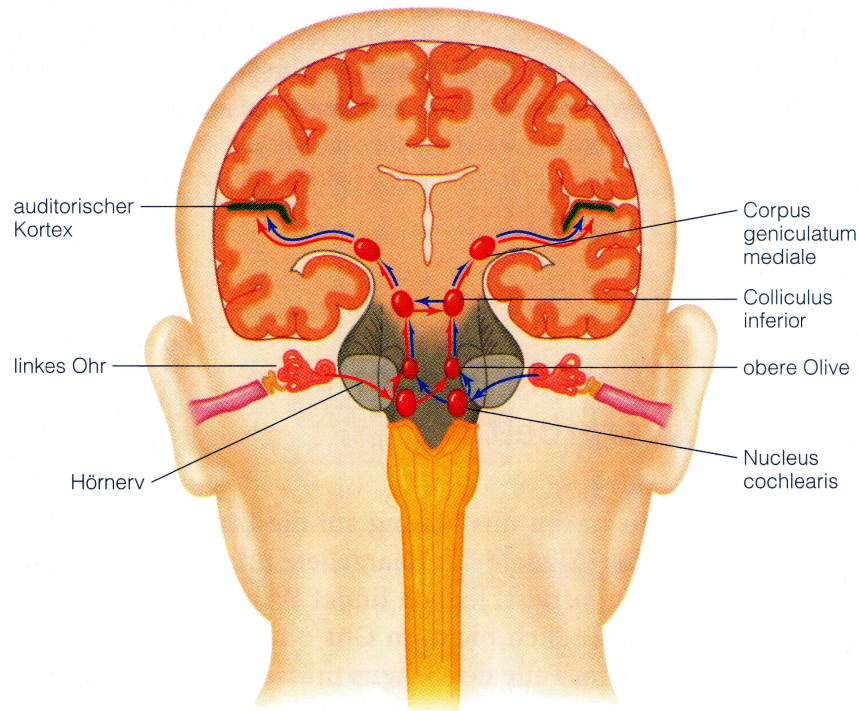


Abbildung 1.4: Vereinfachte schematische Darstellung der Bahnen des auditorischen Systems. (Goldstein 2008: S. 280)

Jede Haarzelle ist mit bis zu 20 Nerven verbunden, je nach Erregungsstärke feuern dabei proportional viele oder wenige Nerven. Die etwa 27.000 Nerven der Cochlea werden gebündelt und erreichen je Kopfseite den Nucleus cochlearis, der die Signale splittet und an die obere Olive beider Hirnseiten weiterleitet (Ellermeier et al. 2008). Es gilt, dass jede Struktur je Hirnseite einmal vorkommt (siehe dazu auch Abb.1.4, S.12). Dabei besteht ein fortwährender Austausch von Signalen zwischen den einzelnen Strukturen der beiden Hirnseiten (Ellermeier et al. 2008).

Aus Übersichtsgründen ist hier weiterführend nur die Weiterleitung der neuronalen Signale für eine Kopfseite beschrieben. Die Nerven sind rein tonotop organisiert. Das bedeutet, dass die Frequenz über den Nerv codiert ist und der Pegel über sein Feuerungsrate. Fricke (2007) ergänzt hier, dass genau genommen die Frequenz auch über die Feuerungsrate, die synchron und phasenstarr zur Frequenz des Schallereignisses ist, codiert wird. Sonst wäre die Frequenzauflösung des menschlichen Gehörs

wesentlich schlechter. In der oberen Olive, die einen Komplex mit mehreren Verarbeitungszentren darstellt, wird die Laufzeitdifferenz beider Ohren über Phasenkopplung in eine Impulsdichtemodulation der Neuronen umgewandelt. Auch die Pegelunterschiede werden entsprechend weiterverarbeitet (Seidl 2003).

1.3.1 Obere Olive

Ebenfalls werden durch die obere Olive motorische Nervenreize z.B. die des Gleichgewichtsorgans verarbeitet und Steuerreize ausgesendet, um z.B. den Kopf für eine besser Lokalisation eines Geräusches zu bewegen. Von hier aus steuern auch die efferenten Nerven die äußeren Haarzellen auf der Basilarmembran, um bestimmte Frequenzen zu bedämpfen oder zu verstärken (Ellermeier et al. 2008).

Zu beachten ist allerdings, dass der obere Olivenkomplex das Auslenken des Kopfes und das Tuning der Basilarmembran nicht selbst veranlassen kann. Denn dazu würde die obere Olive Richtungsinformationen und spektrale Vergleichsparameter benötigen, die erst in höheren Verarbeitungsstrukturen gebildet werden. Die obere Olive reicht lediglich die motorischen Befehle übergeordneter Strukturen weiter und setzt sie z.B. in Signale für spezifische Muskelgruppen um (Goldstein 2008).

1.3.2 Colliculus inferior

Im Colliculus inferior wird die eindimensionale Tonotopie der Cochlea in eine zweidimensionale überführt. Die Pegel- und Laufzeitdifferenzen werden verarbeitet. Es entsteht eine zweidimensionale, horizontal abbildende Karte, die für jede Frequenz Informationen zu Pegel und seitlicher Auslenkung¹ enthält. Durch die Prozesse im Colliculus inferior werden das Frequenzauflösungsvermögen des Gehörs bestimmt und die Fähigkeit, gleichzeitig vorhandene Frequenzkomponenten zu trennen² und einzeln zu identifizieren (Ehret 2014).

1.3.3 Corpus geniculatum mediale

Über den Corpus geniculatum mediale gelangen die Informationen an den jeweiligen auditorischen Kortex der beiden Hirnhälften (Goldstein 2008). Der Corpus geniculatum mediale bildet dabei eine Schnittstelle zu anderen Hirnregionen und fügt den auditiven Informationen nicht auditive hinzu. So gelangen ebenfalls Informationen der Amygdala, die Teil des emotionalen Zentrums ist, und des visuellen Kortex in den auditorischen Kortex (Celesia et al. 2015). Zur weiteren kortikalen Verarbeitung und dem Bewusstwerden eines Schallereignisses sei auf Kapitel 4 Seite 59 hingewiesen.

¹Eine Richtung im eigentlichen Sinne wird nicht abgebildet, lediglich ein Auslenkung. Vorne oder Hinten kann so z.B. nicht unterschieden werden. Siehe dazu Kapitel 2.5, Seite 26.

²z.B. Fomanden.

2 Das Richtungshören

Um eine Schallquelle lokalisieren zu können, bedient sich das Gehör verschiedener Mechanismen. Das Ohr arbeitet, akustisch betrachtet, als Druckempfänger, kann also keine Informationen über die Richtung erfassen. Durch die Form von Kopf und Ohrmuscheln kommt es aber zu richtungsabhängigen linearen Frequenzverzerrungen, die eine grobe Lokalisation ermöglichen. Außerdem kommt es zwischen den Ohren je nach Einfallrichtung des Schalls zu interauralen Pegelunterschieden, interaural intensity difference (IID), sowie zu Laufzeitdifferenzen, interaural time difference (ITD).

2.1 Grundlegende Begriffe

Zum weiteren Verständnis gilt es, einige Begriffe zu klären. Zuerst ist die Schallquelle³ zu nennen. Diese hat im Raum eine physikalische Position und strahlt ein akustisches Signal ab.

Als zweites ist das Schallereignis zu nennen, welches seine Ursache in einer oder mehrerer im Raum positionierten Schallquellen hat. Es beschreibt die Summe der Schallwellen, die jeweils von einem der beiden Ohren aufgenommen wird. Bei jedem gesunden Menschen gibt es unter normalen Umständen immer zwei Schallereignisse⁴, jeweils eines pro Ohr unabhängig von Richtung und Anzahl der Schallquellen.

Als letztes ist das Hörereignis zu nennen, das die Wahrnehmung einer Person beschreibt. Dabei gibt es keinen zwingenden Zusammenhang zwischen einer oder mehreren Quellen im Raum und einem wahrgenommenen Hörereignis (Blauert 1974). So werden z.B. mehrere Hörereignisse verschiedener Elongation zwischen zwei Lautsprechern bei Stereowiedergabe wahrgenommen, dennoch sind nur zwei Schallquellen vorhanden. Umgekehrt kann es auch bei vielen Quellen zu nur einem Hörereignis kommen. Das Hörereignis „Autobahn in der Ferne“ wird z.B. als eines wahrgenommen, obwohl es aus vielen unterschiedlichen und wechselnden Schallquellen besteht. Zuletzt kann ein Hörereignis auch ohne eine Schallquelle und ein Schallereignis vorkommen. Beispiele hierfür sind der Tinnitus oder das Hören einer virtuellen Tonhöhe⁵, was sogar noch beim Fehlen mehrerer Obertöne möglich ist (Goldstein 2008).

Weiter muss zwischen räumlichem Hören und Lokalisation unterschieden werden.

³In dieser Arbeit auch einfach nur Quelle genannt.

⁴Eine Ausnahme sind die einseitige Tonwiedergabe über Kopfhörer und die zu sehr hohen Frequenzen hin auftretende Abschottung zwischen den beiden Ohren durch den Kopf. Diese tritt ab ca. 8 kHz auf. siehe dazu auch Kapitel 2.2.1, Seite 18

⁵Es wird der Grundton wahrgenommen, obwohl dieser im Spektrum fehlt.

Gesunde Menschen hören immer räumlich. Ob sie aber ein Hörereignis lokalisieren können, hängt von verschiedenen Faktoren ab wie z.B. dem Nachhall eines Raumes (Blauert 1974).

2.2 Lokalisationsfähigkeit des Gehörs

Das Gehör kann sehr genau lokalisieren. In der Literatur werden Werte um 1° als bestes Auflösungsvermögen angegeben (Blauert 1974, Blauert et al. 2008). Allerdings ist dies kein allgemeingültiger Wert. Vielmehr ist es ein Spezialfall, denn die Lokalisationsschärfe⁶ ist abhängig von der Frequenz, der Quellenrichtung, der Signalform des Quellensignals und von der Quellenentfernung. Mit steigender Nachhallzeit eines Raumes wird die Lokalisation ebenfalls schlechter. Insbesondere gilt der genannte Wert nur für das Lokalisieren einer einzigen Schallquelle mit geringem Seiten- und Erhebungswinkel nahe dem Nullpunkt des Koordinatensystems, wie es in Abbildung 2.1 auf S. 15 zu finden ist. Als Stimulus führen nur Klicks und kurze Rauschpulse zu diesem guten Ergebnis. Je nach Art des Messstimulus lag die Lokalisationsschärfe zwischen $0,75^\circ$ (Klicks) und $11,8^\circ$ für bestimmte Tonimpulse (Blauert 1974).

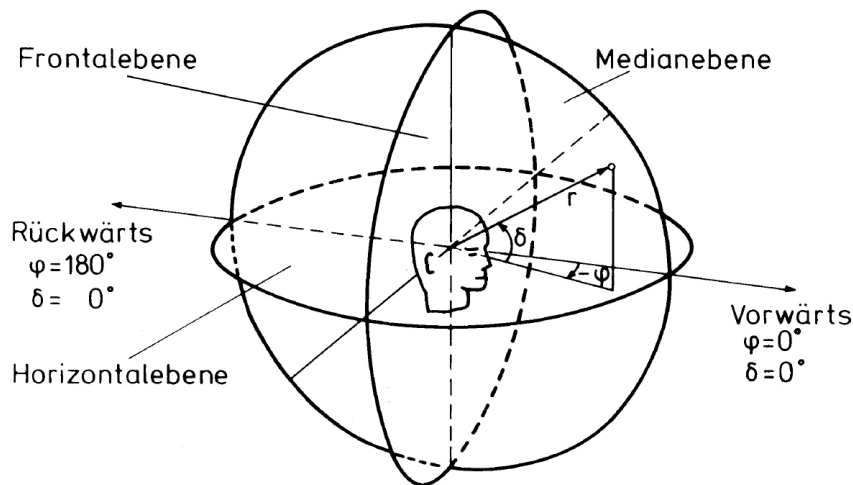


Abbildung 2.1: Kopfbezogenes Koordinatensystem; r Entfernung, φ Seitenwinkel, δ Erhebungswinkel (Blauert 1974: S. 11)

Je länger ein Signal unverändert erklingt, z.B. ein Dauerton, desto schlechter ist die Lokalisationsschärfe. Das gilt auch für die Bandbreite von Signalen: Je schmaler diese ist, um so geringer ist ebenfalls die Schärfe der Lokalisation.

⁶Sie gibt die durchschnittlich mögliche positive und negative Abweichung an. Eine Lokalisationsschärfe von einem Grad bedeutet also z.B. eine Abweichung der wahrgenommenen Richtung von einem Grad nach Links oder Rechts.

2 Das Richtungshören

Für die Frequenz gibt es keinen linearen Zusammenhang. Unterhalb von 500 Hz nimmt die Lokalisationsschärfe mit sinkender Frequenz ab. Darüber konnte nur gezeigt werden, dass für bestimmte Frequenzen eine bessere Lokalisation möglich ist als für andere (Blauert 1974). Die Abhängigkeit der Lokalisationsschärfe von der Frequenz oberhalb von 500 Hz kann der Abbildung 2.2 auf Seite 16 entnommen werden.

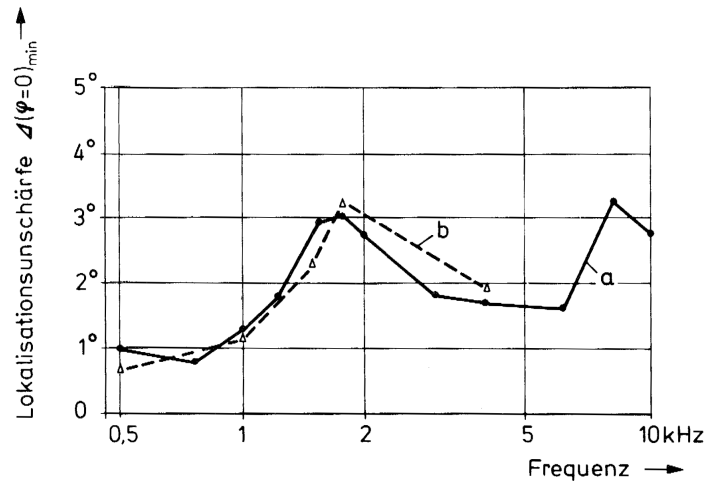


Abbildung 2.2: Lokalisationsschärfe in Abhängigkeit von der Frequenz für $\Delta(\varphi = 0)_{min}$, und $\delta = 0$; Kurve a: Dauertöne, Kurve b: Gauß-Töne (Blauert 1974: S. 32)

Wird davon ausgegangen, dass die meisten Signale in einer natürlichen Hörumgebung eher Schmalbandrauschen ähneln, so kann nach den Messungen die Blauert (1974) verglichen von einer Lokalisationsschärfe von $1,4^\circ$ bis $2,8^\circ$ ausgegangen werden. Ein praktischer Wert entspräche daher eher 2° für die frontale Lokalisation.

Bei anderen Raumwinkeln nimmt das Auflösungsvermögen weiter ab. Bei 30° Auslenkung auf der Horizontalen beträgt die Genauigkeit nur noch etwa $\pm 10^\circ$. Dieser Wert bleibt zu den Seiten hin konstant. Hinter dem Hörer nimmt die Lokalisationsschärfe wieder langsam zu. Für eine rückwärtige Quellenlage von 180° kann nach dem Vergleich von verschiedenen Studien, die Blauert (1974) anführte, geschlossen werden, dass die Lokalisationsschärfe etwa halb so gut ist wie für eine Quelle in 0° Lage. Ähnlich dem Sehen lässt auch das Hören in seiner Lokalisationsfähigkeit zu den Kopfseiten hin (lateral) stark nach (Goldstein 2008). In Extremfällen sind Lokalisationsschärfen von mehr als 60° möglich, bevor eine Lokalisation nicht mehr stattfindet (Jambrosic et al. 2005).

Wichtig ist, dass bei allen Untersuchungen die Köpfe der Versuchspersonen fixiert waren. Damit kann die maximale Lokalisationsschärfe auch für jene horizontalen Raumwinkel gelten, in die der Kopf frei gedreht werden kann. Im allgemeinen Idealfall wäre also eine Lokalisation von 1° durchaus in alle Horizontalrichtungen möglich, sofern sich der Hörende frei in jene Richtung drehen kann, aus der der Schall kommt.

Durch die vielen Abhängigkeiten der Lokalisationsschärfe und -richtung kann es

2 Das Richtungshören

zu Höranomalien kommen. So werden verschieden hohe Töne aus unterschiedlichen Richtungen kommend wahrgenommen. Zum Beispiel scheint ein Vogel, der seine Position nicht verändert, aus unterschiedlichen Richtungen zu kommen, je nachdem in welcher Tonhöhe er gerade zwitschert. Auch können Klangereignisse breiter wirken. So werden z.B. die Partialtöne eines Tones, der auf einem Instrument gespielt wird, leicht aus unterschiedlichen Richtungen wahrgenommen. Dies führte zur Wahrnehmung eines einzigen breiten Instrumententons. Ein zweiter Effekt ist die Falschlokalisierung bei schmalbandigen und verzerrten Schallereignissen. Dabei wird das Signal an der Ohrachse gespiegelt wahrgenommen. Zum Beispiel vorne links statt hinten links. Beide Effekte treten aber nur auf, wenn der Kopf fixiert ist oder das Signal unter 300 ms dauert. Da normalerweise automatisch eine Peilbewegung ausgeführt wird, die zur Auflösung der falschen Richtungsinformationen führt (Blauert 1974). Näheres dazu kann Kapitel 2.5.2, Seite 27 entnommen werden.

Für das Veranlassen einer Peilbewegung braucht das zentrale Nervensystem (ZNS) mindestens 300 ms. Die Kopfbewegung ist dabei minimal und beträgt in 95% aller Fälle unter einem Grad. Ist der Stimulus kürzer, kann keine genaue Lokalisation stattfinden. Dementsprechend ist es nicht möglich, kurze Schallereignisse genau zu lokalisieren (Blauert 1974).

Blauert (1974) konnte zeigen, dass bis etwa 1,6 kHz ein linearer Zusammenhang zwischen Pegel- und Laufzeitunterschied zwischen den Schallereignissen besteht, nicht aber darüber hinaus. Dabei geht er davon aus, dass Zeit und Pegeldifferenzen zur Lateration voneinander unabhängig vom Nervensystem ausgewertet werden. Das konnte später belegt werden (Seidl 2003). Aus diesem Zusammenhang kann eventuell auch die erhöhte Lokalisationsunschärfe um 1,6 kHz in Abbildung 2.2 auf Seite 16 erklärt werden.

Blauert (1974) sah für das Lokalisieren die Pegeldifferenz als wichtiger an als die Zeitdifferenz. Dies begründet er damit, dass die Pegeldifferenz für das Lokalisieren seitlicher Schallereignisse tragend sei. Insbesondere nimmt der Bereich der maximalen Schallquellenauslenkung durch ITD mit der Frequenzhöhe ab. Dadurch ist eine volle Seitwärtsauslenkung und damit eine Seitenlokalisierung in den für den Menschen wichtigen Frequenzbereichen nicht mehr möglich. Neuere Forschungen sprechen aber beiden Mechanismen ähnliche Bedeutung zu (Seidl 2003). Ausführlich wird dies im Kapitel 2.2.2, Seite 20 behandelt.

Schalldruckdifferenzen werden über das gesamte hörbare Spektrum ausgewertet. Interaurale zeitliche Trägerverschiebungen werden bis 1,6 kHz und Hüllkurvenverschiebungen im Zeitbereich von etwa 150 Hz bis zum Ende des wahrnehmbaren Spektrums ausgewertet. Diese Fähigkeit des Gehörs ist aber von der Trägerfrequenz und der Hüllkurve abhängig. Es ist also nicht so, dass eine Richtungslokalisierung unter 1,6 kHz nur über die Laufzeitdifferenz und darüber nur über die Pegeldifferenz stattfindet. Auch unter der genannten Frequenz kann über die Hüllkurven der Schallereignisse eine Laufzeitlokalisierung stattfinden.

Beide Mechanismen finden bei der Lokalisation durch das ZNS Beachtung. Es ist aber zu beobachten, dass – sobald signifikant Frequenzen über 1,6 kHz im Signal vorhanden sind – in der Regel bevorzugt eine Lokalisation über Pegelunterschiede erfolgt (Blauert 1974).

2.2.1 Binaurales Richtungshören durch interaurale Pegeldifferenz

Die Lokalisation durch Pegeldifferenz zwischen den Ohren ist bestimmten Grenzen und Abhängigkeiten unterlegen. Ein entscheidender Faktor ist dabei die Frequenz. So ist bei 2 kHz die Empfindlichkeit am höchsten. Hier führen im Vergleich zu 500 Hz bereits etwa 10 dB geringere Pegelunterschiede zu einer Vollausslenkung zu den Seiten. Bei 5 kHz werden 4 dB mehr benötigt (Blauert 1974). Wie in Abbildung 2.3, S. 18 erkennbar ist, sind die Kurven für unterschiedliche Auslenkungen des Zeigerstimulus nicht ganz gleich. Die Abweichungen können zum einen durch die geringe Anzahl der Testpersonen und zum andern durch die Begrenzung der Auslenkung des Zeigerstimulus, der durch ITD verschoben wurde, oder andere Abhängigkeiten entstanden sein. Dennoch kann eine Abhängigkeit der IID von der Frequenz zweifelsfrei nachgewiesen werden.

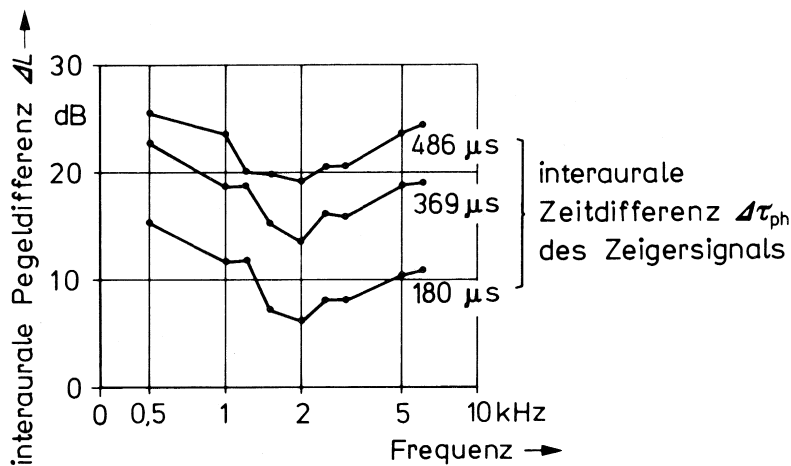


Abbildung 2.3: interaurale Pegeldifferenz eines Tones, dessen zugeordnetes Hörereignis seitlich gleich weit ausgelenkt ist wie ein akustischer Zeiger (zwischen 100 bis 300 0Hz Band begrenztes Rauschen), 6 Versuchspersonen, etwa 60 Phon (Blauert 1974: S. 128)

Bei Frequenzen unter 1,6 kHz kommt hinzu, dass das Hörereignis mit fallender Frequenz immer unschärfer lokalisiert wird und sogar zwei Hörereignisse auftreten können. Eines wird im Kopf lokalisiert das andere unscharf in einer der Pegeldifferenz entsprechenden Richtung. Für ungeübte Hörer ist das Hörereignis oft nicht mehr genau zu lokalisieren und wird sehr breit wahrgenommen.

2 Das Richtungshören

Die Möglichkeit bei tiefen Frequenzen noch über Pegelunterschiede lokalisieren zu können, gilt jedoch nur, sofern es kein Kanalübersprechen zwischen den Ohrsignalen gibt, das trifft jedoch nur bei Kopfhörerwiedergabe zu. Bei einer Schallquelle⁷ kommt es durch die große Wellenlänge mit sinkender Frequenz dazu, dass der Schall um den Kopf immer stärker gebeugt wird. Das führt zu starkem übersprechen. So liegt mit tiefer werdenden Frequenzen praktisch kein Pegelunterschied mehr zwischen den Ohren vor (Blauert 1974, Blauert et al. 2008).

Für hohe Frequenzen gilt bei Kopfhörerwiedergabe nur die Hörgrenze zur Lokalisation durch IID (Blauert 1974, Blauert et al. 2008). Bei Schallquellen im Raum kann ein Lokalisation nur bis etwa 8 kHz sicher erfolgen, da es oberhalb dieser Frequenz zu einer völligen Abschattung⁸ zwischen beiden Ohren kommt (Blauert et al. 2008). Ein Hörereignis würde dann lateral springen und könnte nicht mehr stabil lokalisiert werden. Damit ist ein Intensitäts-Signal-Panning zwischen Lautsprecher nur für Frequenzen zwischen 1,6 kHz und 8 kHz sinnvoll möglich.

Durch die langsam einsetzende Abschattung am Kopf ist davon auszugehen, dass bereits ab 2 kHz die Schallereignisse bei gleichen Differenzen immer weniger lateral auslenkbar sind (Blauert 1974). Ähnliches gilt für die Auslenkung für Pegel unter 2 kHz mit fallender Frequenz, da hier das Übersprechen problematisch ist.

Durch die Frequenzabhängigkeit gilt, dass, je breitbandiger ein Signal ist, desto unschärfer ist es lokalisierbar bzw. um so breiter wird es. Und Signale mit verschiedenen Frequenzlagen werden bei gleicher Pegeldifferenz verschieden weit ausgelenkt (Blauert 1974).

Weiter belegt Blauert (1974), dass beim lokalisieren über IID die Schärfe eines Schallereignisses vom Signalpegel und von der Pegeldifferenz abhängt. Bei Pegeln über 90 dB ermüdet das Ohr und das Hörereignis wandert zur Medianebene. Unter 30 dB steigt die Unschärfe ebenfalls stark an. Auch nimmt die Unschärfe mit steigenden Pegeldifferenzen zu. Damit wird die Lokalisation zu den Seiten hin unschärfer. Wie groß die Unschärfe zu den Seiten für IID wird, nennt Blauert (1974) nicht. Auch in aktueller Literatur wie Blauert et al. (2008) ist nichts näheres zu finden.

Bei welchem Pegelunterschied ein Hörereignis ganz zu einem Ohr hin ausgelenkt wird, ist nicht genau zu sagen. Ab einem Pegelunterschied von etwa 8 dB, führt dieser zu einer Verbreiterung des Hörereignisses und so zu einer damit einhergehenden breiteren Lokalisation. Für Töne wird bei einer Vollauslenkung ein Wert zwischen 15 und 20 dB angegeben und für Rauschen etwa 10 dB (Blauert 1974).

⁷Oder bei mehreren Quellen, die frei im Raum nach dem Intensitätsprinzip zur Bildung einer Phantomschallquelle abstrahlen.

⁸Allgemein lässt sich die Abschattungsgrenze mit $f = 5c/d$ annähern. Bei einer Schallgeschwindigkeit von $c_{20 \text{ °C}} = 343 \text{ m/s}$ und einem Kopfdurchmesser $d = 21 \text{ cm}$ ergibt sich so eine Grenze von 8166,6 Hz.

2.2.2 Binaurales Richtungshören durch interaurale Zeitdifferenzen

Ein zweiter Mechanismus zur Lokalisation in Horizontalebene ist die ITD. Diese funktioniert bis etwa 1,6 kHz und ergänzt die IID damit im Spektrum nach unten. Die Lokalisation basiert dabei nicht etwa auf der Phasenlage ankommender Signale zwischen den Ohren, sondern auf Laufzeitunterschieden. Ab einer Differenz von 630 μ s bis 1 ms wird ein Hörereignis komplett zu einer Seite ausgelenkt. Das bedeutet, dass die halbe Periodendauer einer Schwingung größer sein muss als 630 μ s, um eine laterale⁹ Lokalisation zu ermöglichen. Damit wird die maximale Lateration ab 800 Hz immer geringer und ist ab etwa 1,6 kHz nicht mehr möglich. Hinzu kommt, dass die auf der Basilarmembran erregte Haarzelle nach dem Auslösen eine Regenerationszeit von ein bis zwei Millisekunden benötigen. Über 500 Hz lösen so mit steigender Frequenz proportional weniger Haarzellen pro Schwingung ein Nervensignal aus. Eine sichere Triggerung ist nicht mehr möglich (Blauert 1974).

Dennoch kann über 1,6 kHz durch ITD lokalisiert werden. Dabei kommt es zu einer Auswertung der Hüllkurven der Schallereignisse an den Ohren. Auf welchem Effekt dies beruht, wurde noch nicht geklärt (Blauert 1974, Blauert et al. 2008).

Auch bei der ITD zeigt sich, dass die Lokalisationsschärfe zu den Seiten hin abnimmt. Hinzu kommt, dass die Auslenkung bis 630 μ s etwa proportional ist und dann bis 1 ms in etwa nur noch logarithmisch ansteigt, genaueres ist dazu der Abbildung 2.4, S. 21 zu entnehmen. Bei Differenzen größer einer Millisekunde, kommt es zu keiner weiteren Auslenkung des Hörereignisses. Bei größeren interauralen Verzögerungen würde die Signale lediglich in zwei Hörereignisse zerfallen, von dem das spätere als erste Reflexion oder im Extremfall als Echo wahrgenommen würde.

2.2.3 Monaurales Richtungshören

Der dritte Mechanismus zur Lokalisation basiert auf dem Erkennen von spektralen Veränderungen des Hörereignisses (Blauert 1974, 1997, Blauert et al. 2008). Dieses kann erst auf den höheren kortikalen Verarbeitungsstufen stattfinden, da hier das Signal, mit im Gedächtnis vorhandenen Mustern dieses Signals, verglichen werden muss. Untersuchungen zeigen, dass im Stamm und Mittelhirn keine Strukturen vorhanden sind, die diese Auswertung möglich machen würden (Seidl 2003, Ehret 2014).

Besonders diese Art des Richtungshörens setzt einen längeren Lernprozess vor allem in der Kindheit voraus, da diese Art von Richtungslokalisierung nur durch aufwendige kortikale Mustervergleiche funktionieren (Goldstein 2008).

Welche Frequenzverzerrung für welchen Raumwinkel stattfindet, lässt sich mit der Kopf-bezogenen Transferfunktion, head related transfer function (HRTF), beschreiben. Diese ist für jeden Menschen individuell. Dennoch gibt es bei einer breiten Menge von Personen Ähnlichkeiten (Blauert 1997, Blauert et al. 2008). Ebenfalls ist es möglich,

⁹90° seitlich des Hörers.

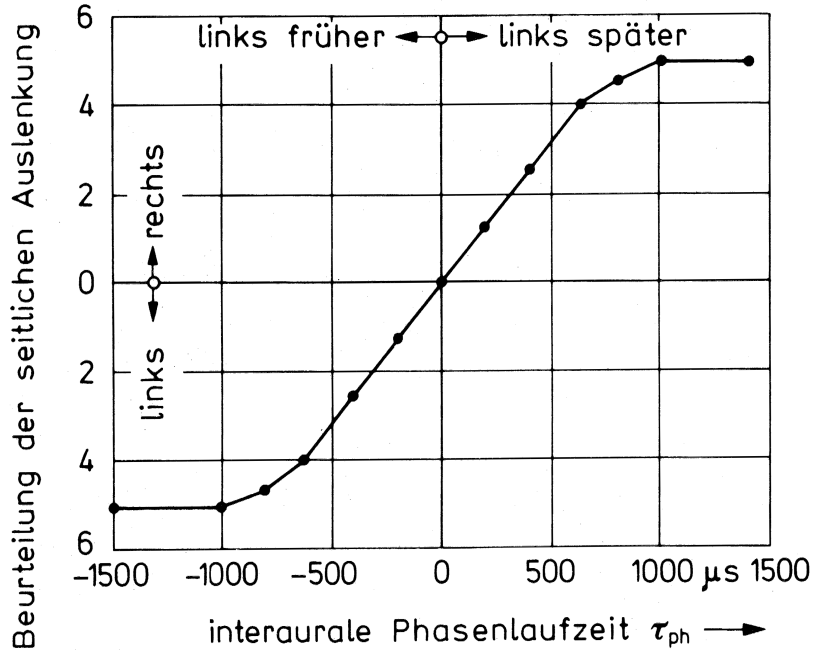


Abbildung 2.4: Seitliche Auslenkung des Hörereignisses als Funktion einer verzerrungsfreien interauralen Signalverschiebung. Gültig für Impulse und Impulshaltige Signale (Blauert 1974: S. 116).

eine neue HRTF durch massives Training zu erlernen, dies setzt jedoch einen längeren Lernprozess voraus (Goldstein 2008, Rozenn 2013).

Einen besonderen Einfluss bei dieser Art der Lokalisation hat dabei die Ohrmuschel und teilweise auch die Kopf- und Rumpfform. Dadurch werden, je nach Einfallswinkel eines Schallsignals, bestimmte Frequenzen verzerrt. Es findet also eine Umcodierung statt von räumlichen Informationen in zeitliche (Blauert 1974).

In der Literatur wird die Genauigkeit der Lokalisation in der Medianebene mit 4 bis 22° angegeben (Blauert 1974). Zum einen zeigt der Vergleich der bei Blauert (1974) genannten Arbeiten, dass die Genauigkeit mit steigendem Erhebungswinkel nachlässt, zum anderen zeigt er, dass die Lokalisierbarkeit stark von der Signalkennntnis des Probanden und der Signalform abhängen. Deswegen divergieren die Aussagen über die Lokalisationsschärfe in der Literatur auch stark, je nach dem in der Untersuchung genutzten Stimulus.

Aus Blauerts (Blauert 1974) Untersuchungen geht hervor, dass es bei jedem Menschen bestimmte individuelle Bänder gibt, die eine Lokalisation eines Hörereignisses Vorne oder Hinten sowie Oben erzeugen. Bei Messungen mit 10 Personen stellte er bei allen Personen einen Pegelanstieg um 8 kHz in Relation zu von vorne gehörten Tönen fest. Zumindest für das Hörereignis Oben scheint dieses Band für die meisten Personen zu gelten. Allerdings stellte Blauert (1974) bei weiteren Untersuchungen eine gewisse Streuung um dieses Band fest.

Weiter fand er Bänder, die jeweils nur für Vorne und Hinten eine Veränderung zeigten. Dabei betont Blauert (1974, 1997), Blauert et al. (2008) immer wieder das alle diese Bänder keine Allgemeingültigkeit haben, da zum Einen die Gruppenlaufzeiten und zum Anderen die Frequenzverzerrung bei jedem Hörer individuell sind. So können sich die nötigen Energiemaxima über die Frequenz verschieben oder verschmieren. Für die Höhenwahrnehmung konnte er aber zeigen, dass Signale im Spektrum um 8 kHz und darüber Energie benötigen, damit eine Höhenlokalisierung überhaupt möglich ist.

Blauert (1974) zeigt auch, dass im Schmalbandrauschen vom Gehör keine Information über die Einfallsrichtung des Schalls gefunden werden kann. Denn eine Lokalisation ist nur über die Veränderung der relativen Pegel der Frequenzanteile untereinander möglich, diese Veränderungen sind aber nur bei breitbandigen Signalen erkennbar. Für Signale unter 1-2 Terzen ist in der Medianebene keine korrekte Lokalisation möglich. Hier hängt der lokalisierte Punkt von der Terzmittenfrequenz ab. Je nach Frequenz wird dann ein Signal Vorne, Hinten oder Oben wahrgenommen¹⁰.

Blauert (1974) beschreibt, dass die Frequenzverzerrung für die Vorne-, Hinten- und Obenlokalisierung vor allem auf Kammfiltereffekten beruhen. Diese entstehen durch das akustischen Zusammenspiel von Kopf, Rumpf und Ohren. So ist es auch möglich, unbekannte Signale, sofern ihr Spektrum breit genug ist, richtig in diesen Raumlagen zu lokalisieren. Blauert (1974) zeigt aber auch beim Vergleich mehrerer Arbeiten, dass mit Kenntnis über das Signal eine genauere Lokalisation über HRTF möglich ist, als nur über die Polstellen der Kammfilter.

Ob eine Untenlokalisierung durch Frequenzverzerrung möglich ist, ließ sich in der Literatur nicht finden. Blauert (1974) schreibt hierzu nur, dass sie kaum möglich ist, meist führe dies zur in Kopf Lokalisation (IKL). Er stellte auch fest, dass ein Lokalisieren über Frequenzverzerrung erst ab Signallänge von mindestens 0,5 ms möglich ist. Für das Lokalisieren in der Medianebene konnte er zeigen, dass diese nur Frequenzabhängig, nicht aber von Veränderungen in Pegel, IID oder Zeit, ITD abhängig ist, sofern der Kopf der Probanden nicht bewegt werden konnte.

2.3 Die Höhe als 3. Dimension

Das Hören in der Höhe basiert, wie im vorherigen Kapitel 2.2.3, 20 gezeigt, auf linearen Frequenzverzerrungen und Kopfbewegungen (Blauert 1974). Sofern eine Kopfbewegung stattfindet, ist in Grenzen auch eine Höhenlokalisierung über IID und ITD möglich (siehe Kapitel 2.5.2, S.27). Für eine genauere Lokalisation in der Höhe oder besser Medianebene ist dabei die Frequenzverzerrung ein wichtiger Faktor, da über die Kopfbewegung effektiv nur eine Sektorlage bestimmt werden kann. Über IID und ITD kann bei frei beweglichem Kopf nur festgestellt werden, dass das Hörereignis

¹⁰Deswegen sind z.B. Zikaden schwer in ihrer Höhenlage zu lokalisieren. Obwohl sie in den Bäumen sitzen, werden sie nur diffus in Bodennähe wahrgenommen. Sie strahlen ein gepulstes Rauschen ab, das in Intervallen kurz unterbrochen ist und sehr stark Band-begrenzt ist. Die Pulse sind ziemlich gleichförmig (eigene Messungen).

2 Das Richtungshören

irgendwo in der oberen Hemisphäre liegt. Die genaue Lage muss durch monaurale Informationen gewonnen werden (Blauert 1974).

Sofern der Kopf fixiert war, konnte dabei für eine Erhebung von Null in der Medianebenen mit weißem Rauschen die beste mögliche Lokalisationsschärfe von $\pm 4^\circ$ ermittelt werden. Für einen bekannten Sprecher lag sie hingegen nur bei $\pm 9^\circ$ und viel für andere Stimuli weiter auf $\pm 17^\circ$ (Blauert 1974). Mit steigendem Erhebungswinkel fällt, wie auf Abbildung Seite zu sehen ist, die Lokalisationsschärfe deutlich.

Im allgemeinen Idealfall wäre also eine maximale Lokalisationsschärfe von 4° in der Medianebene möglich – sofern der Hörende den Kopf frei nach oben und unten kippen kann. In der Praxis ist es jedoch sehr unbequem, den Kopf stark nach unten oder oben zu kippen und nur unter Einsatz des ganzen Körpers möglich. Auch muss davon ausgegangen werden, dass weißes Rauschen kein übliches Signal für natürliche Schallquellen ist. So wird, praktisch betrachtet, die Lokalisationsschärfe eher um die 10° liegen.

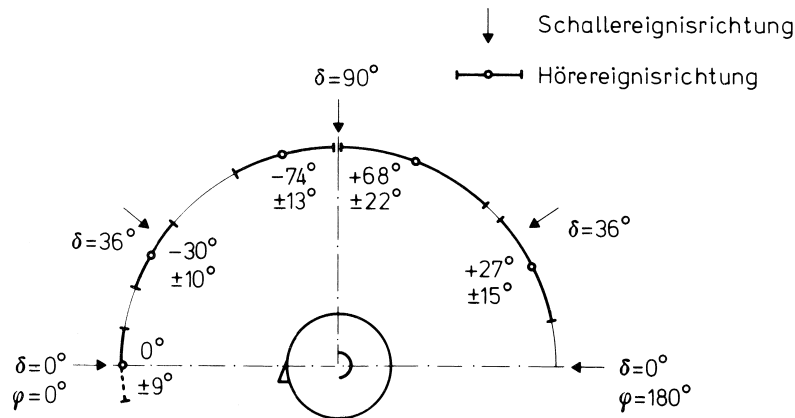


Abbildung 2.5: Minimale Lokalisationsunschärfe und Lokalisation in der Medianebene bei fortlaufender Sprache eines bekannten Sprechers. Kopf fixiert (Blauert 1974: S. 36).

Da über Laufzeit- oder Pegelunterschiede nur die Sektorlage bestimmt werden kann, ist über diese interauralen Unterscheidungsmerkmale ein Panning zwischen zwei vertikalen Lautsprechern nicht möglich: Das Hörereignis springt zwischen oberem und unterem Lautsprecher.

Das Erkennen einer Höhenlage zwischen den Lautsprechern ist insbesondere nicht möglich, da die monaurale Frequenzverzerrung für die jeweilige Winkelrichtung nicht korrekt ist. Bei einem halbwegs stabilen Horizontalpanning müsste also die individuelle HRTF des Rezipienten beachtet werden.

Es wäre möglich, eine HRTF für mehrere konkrete Punkte im Raum zu messen und zwischen den Punkten durch Überblenden zu interpolieren, ohne dass der natürliche räumliche Höreindruck verloren ginge. Diese Technik wird bereits seit mehreren Jahren zur Mehrkanalwiedergabe-Simulation mittels HRTF und Headtracking auf

Kopfhörern genutzt (Rozenn 2013, Theile et al. 1998). Dabei wird vorwiegend auf der Horizontalebene mit Überblenden der einzelnen HRTF gearbeitet. Die zu berücksichtigende vertikale Auslenkung ist, bedingt dadurch, dass der Kopf nicht sehr weit vertikal gekippt werden kann, sehr gering. Dennoch wäre der Aufwand für 3 Dimensionen wesentlich größer. Problematisch ist auch, dass diese Verfahren nur über Kopfhörer und durch die individuelle HRTF nur für eine bestimmte Person zufriedenstellend funktioniert.

Mit diesen Erkenntnissen ist klar, dass eine vertikale Positionierung einer virtuellen Punktschallquelle über Lautsprecher im Raum in der Höhe nicht durch Phantom-schallquellen möglich ist. Zur Positionierung braucht es daher mehrere Lautsprecher in der Vertikalen, die über Schallfeld-Syntheseverfahren, wie Higher Order Ambisonic (HOA) oder Wellenfeldsynthese WFS angesteuert werden. Dabei muss an den Ohr des Zuhörers eine Schallfeld erzeugt werden, das der einer realen Schallquelle ähnlich ist.

Entsprechendes konnte bei Arbeiten mit einem experimentellen HOA-3D-Mischsystem (Falke et al. 2014), das dem entsprechend arbeitet, beobachtet werden, dass ein Höhenpanning der Hörereignisse mit etwa $\pm 20^\circ$ Lokalisationsunschärfe möglich ist. Allerdings springt hier das Hörereignis oftmals merklich zwischen den Lautsprechern. Mit einer größeren Anzahl Lautsprechern in der Vertikalen ließe sich die Lokalisationsschärfe sicherlich verbessern und dadurch ein Springen vermeiden.

2.4 Entfernungshören

Beim Entfernungshören, wie auch bei der Höhenwahrnehmung, wirkt sich die Bekanntheit des Signals auf die Lokalisation aus. Generell ist das Entfernungshören sehr Fehleranfällig, da es vor allem auf Lautstärke sowie Hallanteile und Frequenzverzerrung basiert. Die Entfernungswahrnehmung basiert nicht auf einem objektiven Auswertungsverfahren, sondern auf einer subjektiven Schätzung. Es ist davon auszugehen, dass das beim natürlichen Hören oft die falsche Entfernung angenommen wird. Wie genau die Entfernungswahrnehmung im Gehirn funktioniert, ist dabei noch nicht klar (Blauert 1974, 1997, Blauert et al. 2008). Da sie, wie die Höhenwahrnehmung, auf einer komplexen Signalanalyse basiert, wird sie ebenfalls im oder oberhalb des auditorischen Kortex stattfindet.

Das Entfernungshören lässt sich, soweit durch Experimente in Messungen belegt, funktionsbedingt in drei Schallquellen-Entfernungsbereiche einteilen: unter 25 cm, 25 cm bis 15 m und über 15 m (Blauert 1974, 1997, Blauert et al. 2008).

Sehr nah am Kopf, unter 25 cm, kommt es durch Nahbesprechungseffekte und die starke Schallfeldkrümmung zu großen spektralen Verfärbungen (Blauert 1974). Dadurch ändern sich die monaurale und interaurale Übertragungsfunktion stark. Diese Änderung ist für jeden Menschen spezifisch, aber individuell in Grenzen, von seiner Anatomie bestimmt. Sobald sich die Schallquelle einem Ohr nähert, wird sie vom anderen Ohr je nach Frequenz teilweise bis vollständig abgeschirmt. Die interauralen Pe-

2 Das Richtungshören

gelunterschiede und der Signalpegel steigt dabei stark an. Diese Effekte sind von Kopfform und -größe abhängig. Durch die starke Änderung der Außenohr-Übertragungsfunktionen (HRTF) kann so im Bereich bis etwa 25 cm um den Kopf eine Entfernung bestimmt werden (Blauert et al. 2008).

Der interessanteste Bereich für die Tonwiedergabe ist das Entfernungshören zwischen 25 cm und 15 m. Es beruht vor allem auf Signalpegelveränderungen und Hallanteilen. Ausgehend von einer Punktschallquelle nimmt das Gehör dabei bei 6 dB Dämpfung eine Entfernungsverdopplung an. Um eine Entfernung so bestimmen zu können, muss dem Gehör dabei die Quellenlautstärke bekannt sein. Dadurch ist diese Abschätzung sehr unscharf. Zusätzlich wertet das Gehör den zu einem Quellsignal gehörenden Hallanteil aus. Dabei hängt diese Entfernungsbeurteilung vom Schalleinfallswinkel und von der Signalbeschaffenheit ab (Blauert 1974, Blauert et al. 2008).

Über 15 m sind Pegeldämpfungen über die Entfernung kaum noch wahrzunehmen. Auch der Anteil von Reflexionen ist hier oft so hoch, dass diese Parameter zur Entfernungsbestimmung wegfallen. Bei Entfernungen über 15 m kommt es aber zusätzlich zur frequenzabhängigen Dämpfung durch die Luft. Dadurch werden insbesondere die Höhen stark gedämpft. Eine Entfernungsbestimmung im eigentlichen Sinne ist dabei nicht mehr möglich. Es wird nur noch eine große Entfernung wahrgenommen. Der akustische Horizont beim Menschen liegt somit bei etwa 15 m. Entfernungen darüber hinaus werden nur auf Grund von Erfahrungswerten wahrgenommen (Blauert 1974, Blauert et al. 2008).

Neben den drei beschriebenen Arten des Entfernungshörens gibt es noch einen vierten Spezialfall, die IKL. Sie tritt z.B. bei Kopfhörerwiedergabe auf. Dies liegt nicht etwa daran, dass es zu einem Druckkammereffekt an den Ohren kommt, das ist auch lediglich für tiefe Frequenzen richtig. Stattdessen liegt ein Spezialfall des Entfernungshörens im Nahbereich bis 25 cm vor (Blauert 1974).

Sobald zwei Schallquellen nahe dem Kopf mit gleichen Abständen zu jeweils einem Ohr, ähnliche Signale abstrahlen, kommt es in der Regel zu IKL. Als ähnlich sind dabei nach Blauert (1974) Signale mit einer interauralen Kohärenz viel größer 0,3 anzunehmen, wie sie z.B. bei HiFi-Wiedergaben üblich sind. Kompensiert man die Effekte des Nahbereichshörens durch elektrische Endzerrung findet keine IKL mehr statt.

Ebenfalls kann es zu IKL kommen, wenn Schallquellen, direkt vor und hinter einem Zuhörer bzw. links und rechts von ihm positioniert, gegenphasige Signale abstrahlen. Auch bei einigen Signalen in der Medianebene wie z.B. Schmalbandrauschen kann es zu IKL kommen (Blauert 1974, Blauert et al. 2008). Die IKL ist dabei sehr von der Zuhörerposition abhängig. Sobald dieser diesen schmalen Hörbereich verlässt, hat er eine normale aber diffuse Lokalisation (Blauert 1974). Entsprechen sind solche Signale bei der Mischung zu vermeiden, da sie bei einzelnen Zuhörern zu IKL führen können.

2.5 Lokalisation einer Quelle

Wie bereits im Kapitel 2.2, S. 15 kurz beschrieben, ist es für den Menschen möglich, mit einer Genauigkeit von bis zu einem Grad zu lokalisieren. Allerdings gilt das nur für eine Schallquelle, die horizontal vorne und nahe dem Schnittpunkt zur Medianebene positioniert ist. Für mehrere Schallquellen und einzelne außerhalb dieses Bereiches ist die Lokalisation wesentlich schlechter. Insbesondere gilt dies bei Hörereignissen, die durch Summenlokalisierung wahrgenommen werden (Blauert 1974). Diese beschreibt einen Effekt, der z.B. bei der Stereowiedergabe genutzt wird. Die Summenlokalisierung entsteht durch Laufzeit und/oder Pegelunterschiede eines Signals, das über zwei oder mehr Quellen wiedergegeben wird. Dabei entsteht ein Hörereignis, das zwischen den Quellen entsprechend der Schallereignis-Differenz lokalisiert wird (Blauert et al. 2008). Allgemein wird diese Hörereignis als Phantomschallquelle bezeichnete.

Die Hörereignisposition ist dabei nur bedingt proportional zu Pegel- und Zeitunterschieden, da die im Kapitel 2.2.1, S. 18 und Kapitel 2.2.2, S. 20 genannten Einschränkungen gelten. Bei kleinen Auslenkungen von etwa $\pm 20^\circ$ ist die Auslenkung noch in etwa proportional zur interauralen Pegel- und Zeitdifferenz. Darüber hinaus besteht keine direkte Proportionalität bzw. ein in etwa logarithmischer Zusammenhang. Ein Hörereignis kann dabei maximal um $\pm 30^\circ$ ausgelenkt werden. Deswegen ist z.B. auch das Stereodreieck zur optimalen Wiedergabe mit 60° Lautsprecherabstand angegeben.

2.5.1 Fehler bei interauraler Differenzlokalisierung

Bei der Lokalisation in der Horizontalebene über IID und ITD besteht eine große Fehleranfälligkeit (Blauert 1974). Diese beruht auf einem einfachen mathematischen Problem: Ist zu einem Punkt im dreidimensionalen Raum die Entfernung (oder auch Laufzeit bzw. Pegel) von 2 Punkten bestimmt (die Ohren), deren Position bekannt ist, kann die Position des gesuchten Punktes nur auf einer Kreisbahn, die parallel zur Medianebene liegt, bestimmt werden. Wo der Punkt tatsächlich auf der Kreisbahn liegt, kann nicht festgestellt werden. Beim Hören kommt es zusätzlich zu dem Problem, dass nur die Differenz von Laufzeit und Pegel zwischen den Ohren bestimmt werden kann. Dadurch ergibt sich zusätzlich eine Hyperbelfunktion für Entfernungen, bei denen die Differenz gleich ist. So wird aus der Kreisbahn ein Hyperbelkegel, auf dessen Fläche das Hörereignis liegt (Abb. 2.6, S. 27). Blauert (1974) spricht hier vom sogenannten „*cone of confusion*“ (ebd. S. 143 fff).

Selbst auf die Horizontale reduziert ist es für eine Zuhörer nicht möglich, rein über interaurale Differenzen zu bestimmen, ob eine Quelle vor oder hinter ihm liegt. Er könnte sie lediglich auf einer Hyperbel verorten, auf der Laufzeit und Pegel in beiden Ohren gleich sind. So sind für die IID- und ITD-Lokalisation zusätzliche Informationen über die Raumlage nötig. Diese kann über monaurale Richtungsinformationen im Spektrum erfasst werden. Wie bereits beschrieben, ist jedoch diese Möglichkeit begrenzt und Fehleranfällig.

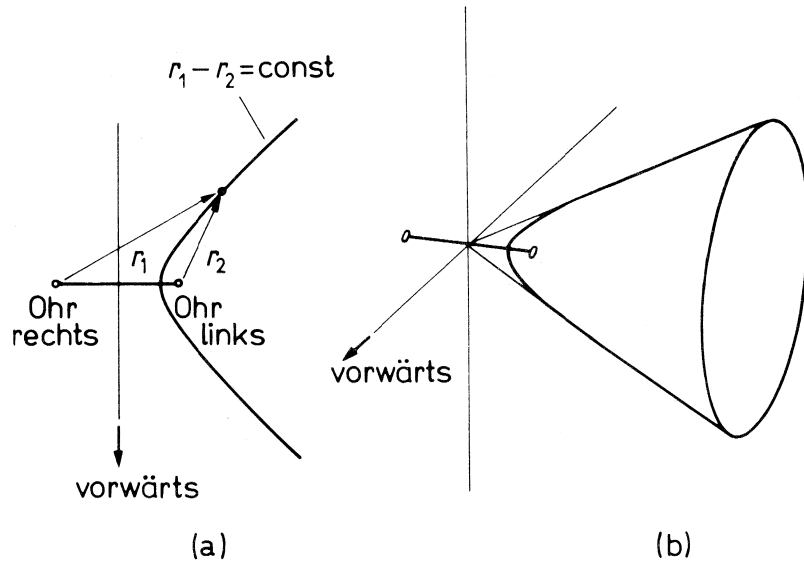


Abbildung 2.6: (a) Bei diesem Model ist der geometrische Ort aller Punkte gleicher Abstandsdifferenz vom linken und rechten Ohr in der Horizontalebene eine Hyperbel, (b) im Raum ein Hyperbelkegel, der sich, je nach dem in welcher Medianhemisphäre sich der Punkt befindet, nach links oder rechts ausdehnt (Blauert 1974: S. 144).

Im alltäglichen Leben ist ein Lokalisieren in Horizontalrichtung dennoch sehr genau und in der Regel nicht fehlerbehaftet, insbesondere vordere und hintere Lage werden normalerweise richtig lokalisiert. Daher muss ein weiterer Mechanismus vorhanden sein, der diese Differenzierung ermöglicht.

2.5.2 Auflösung der interauralen Fehllokalisation

Blauert (1974) stellt dazu seine Drehtheorie auf, die mittlerweile breiten Konsens gefunden hat. Blauert (1974) zeigt, dass, sobald eine grobe Lokalisation auf dem „*cone of confusion*“ vorliegt, ein Mechanismus ausgelöst wird, der durch kleine Kopfbewegungen die Unbestimmtheiten auflöst. Mathematisch gesehen werden hier durch die Bewegung neue „*cone of confusion*“ mit anderer Raumlage erzeugt, über deren gemeinsamen Schnittpunkt dann die Unbestimmtheit aufgelöst werden kann.

Es wäre zwar auch möglich, die Lokalisation durch monaurale Richtungsinformationen zu präzisieren, Blauert (1974) zeigt aber anhand verschiedener Untersuchungen, dass dies nicht der Fall ist und immer eine Kopfbewegung stattfindet, sofern sie nicht von äußeren Faktoren unterbunden wird.

Die Drehung geschieht dabei immer in Richtung Hörereignis in der Regel aber nicht so weit, dass es genau frontal läge. Meist führten die Probanden in den von Blauert (1974) genannten Untersuchungen nur leichte Nick- und Drehbewegungen aus. Die Untersuchungen zeigten, dass das ZNS Informationen über die Raumwinkeländerung

erhält und auswertet. Diese Informationen werden über den Gleichgewichtssinn, der ebenfalls im Ohr liegt, sowie motorische Reize der Nackenmuskeln und des Sehsinns erlangt. Dabei gilt: je mehr Sinnesinformationen bei der Bewegungserkennung beteiligt sind, desto genauer kann lokalisiert werden (Blauert 1974).

Durch die leichten Drehbewegungen, die meist nur wenige Grad betragen, ist es möglich, eine sehr genaue Karte zu erzeugen, die für jede Frequenzen die Horizontaltrichtung zeigt.

Zusätzlich ermöglicht die Roll- und Nickbewegungen des Kopfes eine Bestimmung der Sektorlage einer Schallquelle in die Sektoren vorne oben, vorne unten, hinten oben, hinten unten. Ebenfalls können so nach Blauert (1974) grobe Informationen über den Erhebungswinkel des Hörereignisses erlangt werden. Blauert (1974) konnte damit zeigen, dass Höheninformationen vom Gehör nicht nur über monaurale Informationen gewonnen werden.

Da im Colliculus inferior die Richtungen für einzelne sehr schmale Spektren nur lateral abgebildet werden (die Lokalisation ist nur auf dem „*cone of confusion*“ möglich), kann ein Auswertungsmechanismus, der aus diesen Informationen den Erhebungswinkel bestimmt, nur in höheren Hirnarealen liegen. Die durch Kopfbewegung gewonnenen Lageinformationen einer Schallquelle dominieren dabei über die durch monaurale Auswertung gewonnenen Lageinformationen. Diese können aber ergänzend wirken, um insbesondere bei Erhebung in der Medianebene eine Lokalisation zu präzisieren (Blauert 1974).

2.5.3 Intermodale Wechselwirkungen

Blauert (1974) beschreibt einige Versuche zur Untersuchung intermodaler Wechselwirkung zwischen Seh- und Hörsinn im Bezug auf die Lokalisation. Bei einigen Versuchen konnte dabei fest gestellt werden, dass Seh- und Hörereignis sich von ihrer tatsächlichen Position zu einer Position verschieben, die als zwischen den Reizquellen liegend wahrgenommen wird. Bei anderen Versuchen war dieser Effekt nicht zu beobachten. Bei näherer Untersuchung der Versuchsaufbauten ist zu erkennen, dass dort, wo ein Zusammenhang¹¹ zwischen auditiven und visuellen Reiz besteht, eine Art Phantomquelle zwischen Ton- und Bildquelle entsteht. In anderen Fällen wurde vermutlich kein Zusammenhang zwischen Bild und Ton gebildet, wodurch keine „räumliche Komplikation“ entstand, wie es Blauert (1974) nannte.

2.6 Diskriminierung mehrerer Schallquellen

Bei mehreren Schallquellen kommt es in beiden Ohren zu Signalüberlagerungen. Ob und in wie weit diese Signale von einander wieder differenziert werden können, hängt von verschiedenen Faktoren ab.

¹¹Die Annahme, dass es sich bei dem akustischen und dem visuellen Reiz um denselben Gegenstandes handelt.

Als Sonderfall des Hörens mehrerer Schallquellen nennt Blauert (1974) das Hören einer Schallquelle in natürlichen Räumen. Dabei kommt es zu Reflexionen der ausstrahlten Wellen an den akustischen Begrenzungsflächen des Raumes (z.B. Wände). Diese können als Spiegelschallquellen angesehen werden. Natürliches räumliches Hören, unabhängig von der Anzahl der Primärschallquellen ist demnach immer ein Hören vieler, im Extremfall unendlich vieler Schallquellen.

Blauert (1974) geht davon aus, dass das Gehör jedes Signal spektral zerlegt und selektiv für jedes Frequenzband eine Richtungsbestimmung vornimmt. Dies wird durch aktuelle Forschungen der Neurologie belegt, die zeigen, dass durch Neuronen im Colliculus inferior für jede Frequenz die horizontale Richtung abgebildet wird (Seidl 2003, Ehret 2014: siehe dazu auch Kapitel 1.3.2, S. 13).

2.6.1 Hall als Wahrnehmung mehrerer Schallquellen

Für die Betrachtung der Reflexionen als Spiegelschallquellen ist zu beachten, dass unterhalb der Echschwelle nicht lokalisiert werden kann. Darüber hinaus ist die Anzahl der einzeln lokalisierbaren Quellen begrenzt. Das Gesetz der ersten Wellenfront besagt, dass das Gehör Verzögerungen von mehr als $630 \mu\text{s}$ ¹² als Reflexionen wahrnimmt, sie aber bis zum Erreichen der Echschwelle nicht auf Richtungsinformationen hin auswertet. Diese frühen Reflexionen sorgen aber für eine vollere und lautere Wahrnehmung (Görne 2008). Sie sind somit insbesondere für die Sprachverständlichkeit wichtig.

Die Echschwelle ist individuell von verschiedenen Parametern abhängig. Bedingend sind die Frequenz, die Richtung der Reflexion zur Primärquelle, sowie Pegel und Pegeldifferenzen in Relation zur Primärquelle. In einigen Fällen kommt es bereits bei einer Echschwelle von 2 ms zu einer Lokalisation der Reflexionen. Unter anderen Umständen erst bei wesentlich längeren Verzögerungen von bis zu 215 ms. Für Geräusche und Musik liegt die Echschwelle oft über 50 ms, für Sprache bei etwa bei 20 ms. Reflexionen innerhalb dieser Zeit werden nicht lokalisiert, erzeugen aber eine Klangfärbung und eine Räumlichkeit, die den Raum in Größe und Material der reflektierenden Flächen wahrnehmen lassen (Blauert 1974).

Eine Diskriminierung und damit die Ortung erster Reflexionen unterhalb der Echschwelle ist nicht möglich. Erst spätere Reflexionen von z.B. Wänden werden lokalisiert, sofern sie diskriminierbar sind. Dafür gelten dieselben Gegebenheiten wie für die Diskriminierung mehrerer Primärschallquellen (siehe folgendes Kapitel 2.6.2).

Die Reflexionen des Nachhalls könne wegen ihrer großen Zahl und dem gleichzeitigen Auftreten nicht lokalisiert werden. Es entstehen vielen Überlagerung in beiden Ohren, wodurch unkorrelierte Schallereignisse entstehen, die eine Lokalisation nicht mehr zulassen (Blauert 1974).

¹²Blauert (1974) gibt die Schwelle mit $630 \mu\text{s}$ bis 1 ms an. Die Grenze ist unscharf und von verschiedenen Faktoren abhängig. Allgemein wird sie meist mit 0,6 ms angegeben vgl. Görne (2008).

2.6.2 Grundsätzliche Voraussetzungen

Für die Lokalisationsschärfe von Schallquellensignalen die ein Hörereignis hervorrufen, ist zu beachten, dass die Ortungsschärfe abhängig ist von Signalhüllkurve, Frequenz und Frequenzbreite des Signals (Blauert 1974). Die Differenzierbarkeit ist ebenfalls abhängig von der Form des Quellensignals. Sind die Signale einzelner Quellen sehr ähnlich, verschmelzen sie zu einem diffusen oder breit lokalisierbarem Hörereignis.

Kurze breitbandige Klicks sind besser ortbar als lange Sinustöne. Klicks könne daher auch noch bei sehr naher Lage der Quellen zueinander im Gegensatz zu Sinustönen voneinander differenziert werden.

Ein weitere Faktor ist der Pegelunterschied einzelner Quellen. Bei Pegelunterschieden kommt es, abhängig von den Signalbeschaffenheiten, ab 15 dB, spätestens jedoch bei 40 dB dazu, dass eine Quelle von einer lauteren betreffend der Lokalisierbarkeit maskiert wird. Die leisere Quelle kann klanglich noch wahrgenommen werden, nicht aber in ihrer Position bestimmt werden.

Das bedeutet für die 3D-Mischung, dass, wenn viele Quellen lokalisierbar sein sollen, die Pegel ähnlich laut sein müssen. Andererseits bedeutet dies für leise Atmos im Hintergrund, das es nicht sinnvoll ist, dies aus vielen virtuellen Punktschallquellen¹³ zu mischen. da sie durch vordergründige Quellen in ihrer Lokalisierbarkeit maskiert würden. Ein diffuser Teppich aus wenigen¹⁴ teilkohärente virtuellen Punktschallquellen, z.B. aus den Signalen einer quadrophonische Aufnahme, genügen.

2.6.3 Interaurale Kohärenz von Schallquellen

Eine wichtige Beschaffenheit zur Differenzierbarkeit mehrerer Signale in ihre jeweiligen Hörereignisse, ist die interaurale Kohärenz¹⁵. Die Signale mehrerer Quellen und die Schallereignisse der Ohren weisen dabei unter natürlichen Bedingungen und üblichen Hörsituationen nicht dieselbe Kohärenz auf. Ein Hauptgrund sind die Übersprecheffekte der Schallwellen zwischen beiden Ohren. So ist der mögliche Schwankungsbereich des Kohärenzgrades der Ohrsignale in der Regel kleiner als zwischen den Schallquellen (Blauert 1974).

Der Kohärenzgrad der Schallereignisse zwischen beiden Ohren bedingt die Ausdehnung eines resultierenden Hörereignisse. Je kleiner dieser ist, desto breiter wird ein Hörereignis. Die Lokalisationsschärfe lässt nach. Bei einem Kohärenzgrad von Eins ist diese maximal. Mit sinkendem Kohärenzgrad wird die Lokalisation immer breiter

¹³Im Bereich der 3-D Tonmischung wird oft von Objekten gesprochen, dies Bezeichnung ist jedoch missverständlich.

¹⁴Z.B. zwei bilaterale Quellen auf Ohrhöhe oder frontal je links und rechts und hinten je links und rechts auf Ohrhöhe.

¹⁵In dieser Arbeit werden die Bezeichnungen Kohärenzgrad und Kohärenz synonym für den interauralen Kohärenzgrad genutzt. Dieser lässt sich nach Blauert (1974) ersatzweise mit dem Maxima des Betrages der normierten Kreuzkorrelation (Kor_{inau}) berechnen. Sollten andere Kohärenzen gemeint sein, werden diese explizit benannt.

2 Das Richtungshören

und unschärfer. Bei einem Grad unter 0,2 beginnt das Hörereignis in zwei separate Hörereignisse zu zerfallen. Bereits ab einem Kohärenzgrad von 0,4 dehnt sich das Hörereignis bei Binauraler Kopfhörerwiedergabe über die gesamte vordere Hörhemisphäre aus. Teilkohärente Ohrsignale führen also zu ausgedehnteren Hörereignissen als kohärente. Mit sinkendem Kohärenzgrad wird die Lokalisation unschärfer (Blauert 1974).

In Untersuchungen mit Kopfhörerwiedergabe zeigte Blauert (1974), dass durch die Inkohärenz der Schallereignisse diese separiert und als eigenständiges Hörereignis wahrgenommen werden können. Bei sehr geringen oder inkohärenten Signalen wird jedem Ohrsignal ein getrenntes Hörereignis zugeordnet. Für Teilkohärenzen gilt dabei, dass nur ein Hörereignis stattfindet, das mit sinkendem Kohärenzgrad immer diffuser lokalisiert wird, bis es bei der Grenze von etwa 0,2 in zwei Hörereignisse zerfällt.

Diese Gegebenheiten sind nach Blauert (1974) auf das Hören von Schallquellen im freien Schallfeld übertragbar, unter der Berücksichtigung der Übersprecheffekte zwischen beiden Ohren, sowie raumakustischer Effekte. Dadurch bilden sich bei inkohärenten Quellen, die sehr nah beieinander liegen, kohärente Ohrsignale. Zwei räumlich nahe gelegene inkohärente Quellen führen zur Wahrnehmung eines einzigen Hörereignisses mit einer bestimmten Ausdehnung. Liegen sie sehr nah nebeneinander, ist die interaurale Kohärenz nahe eins, es wird nur ein Hörereignis mit minimaler Ausdehnung wahrgenommen. Umgekehrt gilt, dass bei zwei kohärenten Quellen, die weiter auseinander liegen, teilkohärente Schallereignisse entstehen. Dies führen ebenfalls zu einem Hörereignis mit einer bestimmten Ausdehnung (Blauert 1974).

Es ist allerdings zu beachten, dass die Untersuchungen nur mit Rauschsignalen gemacht wurden. Daher kann nicht auf einen Allgemeinfall geschlossen werden (siehe Kapitel 3.4.5, Seite 57).

Im freiem Raum sind beide Ohrsignale nur bei Beschallung mit einer Punktschallquelle kohärent, wenn diese auf der Medianebene liegt. Bei seitlicher Auslenkung kann der Grad unter 0,5 absinken. Hierin ist mit ein Grund dafür zu sehen, dass die Lokalisationsschärfe zu den Seiten hin stark abnimmt (Blauert 1974).

Schallreflexionen haben, obwohl sie zumeist einer linearen Frequenzverzerrung unterliegen, noch einen hohen Kohärenzgrad. Damit werden Reflexion und Quelle nicht als getrennte Hörereignisse wahrgenommen, sondern lediglich ein breiteres Hörereignis, das weniger scharf lokalisierbar ist (Blauert 1974).

Eine Differenzierung von Hörereignissen allein auf Basis der interauralen Kohärenz ist nicht möglich. Allerdings gibt sie einen Anhaltspunkt, ob eine Diskriminierung in einzelne Hörereignisse überhaupt möglich ist. Näheres dazu im Kapitel 3, S. 41. Ein Großteil der Hörereignisdifferenzierung findet dabei im auditorischen Kortex statt (Goldstein 2008).

Ein Hinweis für eine rudimentäre Hörereignis-Differenzierung unterhalb des auditorischen Kortex gibt Blauert (1974) mit der Vermutung, dass Hörereignisse wohl auf Grund von Frequenzbandseparation voneinander differenziert werden. Auf der Basilarmembran findet eine Zerlegung in mehrere Frequenzbänder statt, in der oberen Olive werden die Frequenzbänder in ihrer Zahl und Güte durch weitere Signalanalysen er-

höht (Ellermeier et al. 2008). Danach wird für jedes Band unabhängig die laterale Richtung bestimmt. Hier könnte für jedes Band eine interaurale Kohärenz gebildet werden, um fest zu stellen, ob es, im Falle einer Kohärenz von nahe Eins, zu vermutlich einem genau lokalisierbaren Hörereignis, im Falle von einem Kohärenzgrad kleiner 1 zu einem Hörereignis mit einer bestimmten Ausdehnung kommt und in Falle einer Kohärenz unter 0,2 das Signal des Bandes zu mehreren Hörereignissen gehört.

Qualitativ lässt sich damit zumindest feststellen, dass für die Differenzierbarkeit von Quellen-Signalen die Spektren differenzierbar sein müssen.

Weiterführende Untersuchungen zur Quellendifferenzierbarkeit in Abhängigkeit vom Kohärenzgrad speziell bei dreidimensionaler Quellenlage im Raum waren nicht zu finden. Eine weiterführende Arbeit, bei der mit mehreren horizontal im vorderen Hörraum positionierten Quellen gearbeitet wurde, ist bei Jambrosic et al. (2005) zu finden. In dieser Arbeit wird nicht direkt untersucht, wieviele Quellen voneinander differenziert werden können. Es wurde aber der Lokalisationsfehler in Abhängigkeit von der Störschallquellenanzahl untersucht. Näheres dazu im Kapitel 2.7, Seite 35. Über die Differenzierbarkeit, abhängig vom Kohärenzgrad, fanden Jambrosic et al. (2005) heraus, dass zwei Quellen im Raum unabhängig von ihrer Winkellage nicht differenzierbar sind, wenn die Kohärenz der Signale 0,814 oder höher beträgt. Ab welchem Kohärenzgrad eine Differenzierung möglich ist, ist damit nicht geklärt. Insbesondere da Rauschen verwendet wurde (siehe Kapitel 3.4.5, S. 57). Betreffend des Korrelationsgrades der Quellensignale wäre zumindest schließbar, dass umso mehr Quellen voneinander differenziert werden können, je kleiner dieser ist. Fortführend sei dazu auf Kapitel 3, S. 41 verwiesen.

2.6.4 Binaurale Maskierungslevel-Differenz

Weiterführend ist zur Differenzierbarkeit mehrerer Signale, in die jeweiligen Hörereignisse, der Cocktailparty-Effekt zu betrachten. Dabei gilt, dass ein Nutzsignal aus einer bestimmten Richtung, das von einem Störsignal aus einer anderen Richtung überlagert wird, bei binauralem Hören nicht so stark verdeckt wird wie bei monauralem Hören (Blauert 1974).

In Untersuchungen wurde dabei festgestellt, unter welchen Umständen und bis zu welcher binauralen Maskierungsleveldifferenz, binaural masking level difference (BMLD), ein Signal separiert werden konnte. In anderen Untersuchungen wurde speziell die binauralen Maskierungsleveldifferenzen zur Sprachverständlichkeit, binaural intelligibility level difference (BILD), untersucht (Blauert 1974).

Dort lässt sich ein Anhaltspunkt finden, wie viele Quellen maximal differenziert wahrnehmbar sind, bevor die Summe der Schallsignale aller Quellen bei binauraler Rezeption das Signale jeder einzelnen Quelle maskiert. In diesem Fall ist nur noch die Summe aller Signale als Hörereignis wahrnehmbar. Eine diffuse Lokalisation des resultierenden Hörereignisse kann zwar noch möglich sein (Blauert 1974). Jedoch dehnt es sich mindestens über die Lage der einzelnen Quellpositionen aus. Unter Umständen ist die Ausdehnung sogar größer. Wenn die Quellen in verschiedenen Raumsektoren

2 Das Richtungshören

verteilt sind, ist keine Lokalisation mehr möglich. In diesem Fall wird lediglich ein diffuses Schallfeld wahrgenommen, dass zu einer starken Umhüllung führt.

Für die BILD ist zu berücksichtigen, dass sie lediglich den Punkt angibt, ab dem Sprache, wie z.B. Dialoge in der Kinomischung, nicht mehr verstanden werden kann. Für Sprache wurde eine Maximale BILD von 9 dB gemessen. Sobald ein Sprecher um 9 dB leiser ist als die Summe der Signale anderer Sprecher, kann er nicht mehr verstanden werden. Geht man davon aus, dass die Signale der einzelnen Sprecher kaum korrelieren, kann mit der Verdopplung der Sprecher eine Verschlechterung des BILD von 3 dB angenommen werden.

Daraus resultiert, dass im Optimalfall bei acht Sprechern die Grenze erreicht ist, bei der jeweils der Dialog eines Sprechers verstanden werden kann. In der Praxis wird der Wert sicherlich geringer sein. Im Umkehrschluss ergibt sich, dass, sofern mindestens 9 Sprecher gleich laut sprechen, es unmöglich ist, einen einzelnen zu verstehen. Beim Unterschreiten der BILD ist eine Differenzierung der einzelnen Sprecher und eine Lokalisation durchaus noch möglich. Daher ist diese nicht für die Differenzierbarkeit zu berücksichtigen. Erst mit Unterschreiten des BMLD ist eine Differenzierung und damit eine Lokalisation nicht mehr möglich. Wichtig ist die BILD aber insoweit, dass damit definiert ist, unter welchen Umständen ein Dialog gerade noch verstanden werden könnte.

Die BMLD basiert vor allem auf interaurale Phasenbeziehungen von Nutzsignal und Störsignal und ihrer Phasenlage zueinander. Weniger wichtig, dennoch bestimmende Parameter, sind Bandweite des Störsignals und der Frequenzabstand der Bänder von Nutz- und Störsignal, sowie deren interaurale Kohärenz. Da die Phasenkopplung der Neuronen im Hörnerv mit steigender Frequenz abnimmt und bereits bei 1 kHz kaum noch vorhanden ist (Goldstein 2008), wirkt sich das stark auf die BMLD aus. Für Frequenzen über 2 kHz ist sogar gar keine BMLD mehr messbar (Blauert 1997).

Unter folgenden Optimalbedingungen ist eine BMLD mit einem Maximum bei 250 Hz von 15 dB ermittelt worden: das Störsignal ist ein Rauschen und hat eine interaurale Phasenlage von Null und das Nutzsignal ist ein Dauerton mit einer interaurale Phasenlage von 180° an den Ohren. Zwischen 100 Hz und 1 kHz liegt der Wert immerhin über 6 dB. Bei anderen Phasenlagen war der BMLD bei 250 Hz wesentlich geringer und läuft sehr schnell gegen 0 dB (siehe dazu Abbildung 2.7, Seite 34). Arbeiten darüber, wie groß die BMLD mit natürlichen Signalen ist, wie z.B. Umweltgeräusche, waren nicht zu finden.

Wird davon ausgegangen, dass sich mit einer Verdoppelung der Quellenanzahl die BMLD, mit der Annahmen von unkorrelierten Signalen, um 3 dB verschlechtert, lässt sich zumindest eine erste Eingrenzung vornehmen. Bei 15 dB BMLD könnten so bis zu 32 Quellen geortet werden. Allerdings müsste jeder Quellen jeweils einen Dauerton abstrahlen, der zu einem Schallereignis mit einer interauralen Phasenlage von 180° führen würde. Die Summe aller Signale dürfte aber keine interaurale Phasendifferenz bilden und müsste ein Rauschen erzeugen. Diese Bedingung ist praktisch nicht erfüllbar. Ginge man von einem möglichen BMLD von 6 dB bis 12 dB aus, wären so 4 bis 16 Schallquellen differenzierbar.

2 Das Richtungshören

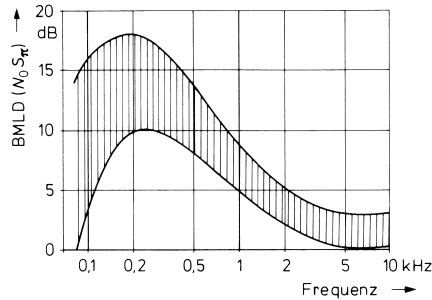


Abbildung 2.7: Abhängigkeit der BMLD eines Dauertons bei $N_0 S_\pi$ von der Frequenz. Schraffiert der Streubereich unterschiedlicher Messungen (Blauert 1974: S. 207).

2.6.5 Folgerungen für eine Differenzierbarkeit

Für die Höhenwahrnehmung gilt nach Blauert (Blauert 1974: siehe dazu auch Kapitel 2.2.3, S. 20) dass diese nur möglich ist, wenn Frequenzen im Band um 8 kHz vorhanden sind. Da bereits ab 2 kHz keine BMLD mehr vorhanden ist, kann das ein Anzeichen dafür sein, dass mehrere Quellen in der Höhe nur schwer voneinander differenzierbar sind. Auf jeden Fall müssen Quellen, die in der Höhe lokalisierbar sein sollen, vor allem im Frequenzbereich um 8 kHz und darüber hinaus mehr Energie abstrahlen, als die Summe aller anderen Quellen.

Abgeleitet aus der Frequenzgrenze des BMLD lässt sich schlussfolgern, dass mehrere Quellen in der Horizontalen besser lokalisierbar und differenzierbar sind, wenn sie ihre Energiemaxima unterhalb von 2 kHz besitzen¹⁶.

Bei Forschungen zur Pegelabhängigkeit des BMLD wurde fest gestellt, dass ein maximaler BMLD erst ab einen Signalpegel von 60 dB SPL erreicht wird. Bei 40 dB ist er bereits um 7 dB geringer, und bei 20 dB um mehr als 10 dB verringert. Zu erklären ist dies mit dem inneren Rauschen der Ohren, das das Signal überlagert und so für eine zusätzliche Maskierung sorgt (Blauert 1974).

Durch die BMLD wird auch die Dynamik mehrerer Quellen eingeschränkt, sofern eine gute Lokalisation möglich sein soll. Quellen, die 40 dB leiser sind als der Gesamtpegel, können nicht lokalisiert werden. Bei spektraler Verdeckung können Quellen bereits nicht mehr lokalisiert werden, die 15 dB leiser sind (siehe dazu Kapitel 2.6.2, S. 30). Virtuelle Punktschallquellen, die entsprechend leise sind, könnten nicht

¹⁶Eine interessante Beobachtung hierzu bieten Grillen: Ihr Zirpen hat ein Energiemaxima deutlich über 2 kHz. Nach Blauerts Untersuchungen müssten sie gut lokalisierbar sein, da ihr Zirpen eine ausgeprägte Hüllkurve besitzt und das Signal selbst aus einem Rauschen mit einer Bandbreite von einigen Terzen im Bereich von etwa 2 - 10 kHz besteht (eigene Messungen und Erfahrungen). Für eine einzelne Grille gilt dabei, dass sie in der Tat sehr gut lokalisierbar ist. Nur durch Hören ist ihre Position sehr gut auszumachen, lange bevor sie visuell lokalisierbar ist. Sobald mehrere Grillen zirpen, ist es jedoch unmöglich, ihre Position zu bestimmen, sofern man nicht einer einzelnen so nah ist, dass sie deutlich lauter ist als die anderen Grillen.

mehr wahrnehmbar im Raum positioniert werden. Vermutlich würden sie mit einer lauterer Quelle zu einem etwas breiterem Hörereignis verschmelzen und dessen Klang verändern.

Für die Mischung gilt, dass es sinnvoll ist, statt vieler virtueller Punktschallquellen, die verschiedenen Geräusche wiedergeben, eine mehrkanalige Atmomischung, deren Einzelkanäle eine geringe Kohärenz aufweisen, über virtuelle Punktschallquellen in den Raumecken wiederzugeben. Denn in beiden Fällen kommt es zu einer diffusen, nicht lokalisierbaren Wahrnehmung. Es ist nicht sinnvoll, mehrere virtuelle Punktschallquellen bei Pegeln unter 40 dB SPL im Raum zu positionieren, da sie bei diesem Pegel nicht mehr differenzierbar oder lokalisierbar sind (siehe dazu ebenfalls Kapitel 2.6.2, S. 30).

2.7 Lokalisationsgenauigkeit bei mehreren Schallquellen

Für eine Schallquelle ist die Lokalisationsgenauigkeit, insbesondere im frontalen Bereich, sehr hoch. Bei mehreren Schallquellen kommt es jedoch zu Wechselwirkungen, die zu einer Verschlechterung der Lokalisationsleistung führen. In der bereits genannten Arbeit von Jambrosic et al. (2005) finden sich einige interessante Ergebnisse. Dort wurde mit 17 Probanden getestet, wie genau ein Bezugssignal in Abhängigkeit von mehreren Störsignalen lokalisiert werden kann und ob eine Unterscheidung des Bezugssignals von den Störern möglich ist. Die Störsignale bestanden aus Rauschen, das Bezugssignal war ebenfalls ein Rauschen, das amplitudenmoduliert war. Bei dieser Kohärenz von 0,814 zwischen Stör- und Bezugssignal lag der Lokalisationsfehler bei nur einem Störsignal bereits bei $\pm 60^\circ$. Bei höheren Kohärenzen verschmolzen beide Quellen zu einem breiten Hörereignis. Die Lokalisationsgenauigkeit läge demnach allgemein etwa zwischen 1° und ca. 60° . Bei korrelierten Quellen gilt, dass sie nicht einzeln als Hörereignisse lokalisierbar sind sofern sie nicht in einem größeren horizontalen Abstand zueinander liegen. Eine gute Lokalisation mehrerer Quellen ist nur möglich, wenn die Lautstärkeunterschiede und die Signalkohärenz einzelner Spektren gering ist (siehe dazu Kapitel 2.6, S. 28)

2.8 Summenlokalisierung

Eine Summenlokalisierung entsteht, sobald mindestens zwei Quellen ein korreliertes oder teilkorreliertes Signal abstrahlen. Dabei entsteht die Wahrnehmung eines Hörereignisses zwischen den Quellen, die Phantomschallquelle. Dieses kann durch Verzögerung oder Pegeländerung zwischen den Quellen ausgelenkt werden. Begrenzt können auch mehrere Hörereignisse unterschiedlicher Auslenkung dargestellt werden. Generell ist die Lokalisationsschärfe und Differenzierbarkeit von Hörereignissen die auf Phantomschallquellen beruhen wesentlich geringer, als die von jenen, deren Ursache ei-

2 Das Richtungshören

ne einzelne Schallquelle ist. Die Schärfe nimmt ebenfalls mit zunehmender lateraler Auslenkung ab. Blauert (1974) gibt für eine Auslenkung von $\pm 30^\circ$, für Pegeldifferenz etwa $\pm 15^\circ$ Lokalisationsunschärfe, für die Laufzeitdifferenz je nach Signalform $\pm 6^\circ$ bis $\pm 11^\circ$ Lokalisationsunschärfe an. Dabei ist zu beachten, dass die maximale Auslenkbarkeit über die Laufzeitdifferenz ab etwa 500 Hz abnimmt und über 2 kHz nicht mehr möglich ist.

In den vorangegangenen Kapiteln wurde die Auslenkbarkeit eines Hörereignisses für reale Schallquellen oder für Kopfhörer-Darbietung beschrieben. Dabei wurde festgestellt, dass unter optimalen Bedingungen ¹⁷ eine Vollauslenkung zu den Seiten hin möglich ist.

Für eine Hörereignis-Auslenkung im Raum über IID und/oder ITD zwischen zwei Quellen, wie z.B. Lautsprecher, gelten jedoch einige Einschränkungen. So ist es nach verschiedenen Messungen, wie sie Blauert (1974), Blauert et al. (2008) dokumentieren, nicht möglich, ein Hörereignis weiter als $\pm 30^\circ$ horizontal auszulenken (siehe dazu auch Abbildung 2.8, S. 36). Deswegen werden Stereolautsprecher idealerweise auch 60° auseinander gestellt, da so die volle Hörereignisauslenkung zwischen den Lautsprechern abhängig von der Pegel- und Zeitdifferenz möglich ist.

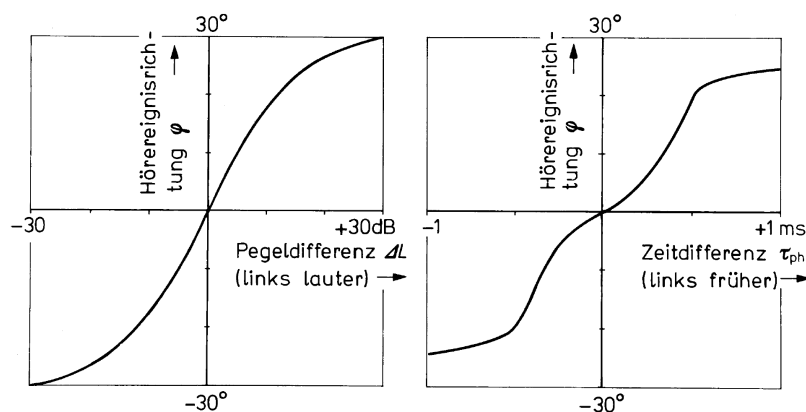


Abbildung 2.8: Summenlokalisationskurven für Pegel- bzw. Laufzeitdifferenzen der beiden Lautsprechersignale. Breitbandsignale, Zuhörerkopf fixiert (Blauert et al. 2008: S. 101).

Werden die Lautsprecher weiter auseinander gestellt, besteht kein direkter Zusammenhang mehr zwischen Pegel- und Laufzeitänderung und dem Hörereignis. Das Hörereignis kann nicht mehr voll ausgelenkt werden. Nur bis zu einem Winkel von etwa $\pm 30^\circ$, darüber hinaus springt es zwischen den Lautsprechern. Bei sehr großen Lautsprecherabständen kommt es zusätzlich dazu, dass jeder Lautsprecher als eigenständiges Hörereignis wahrgenommen wird. Die Phantomschallquelle kann gar nicht

¹⁷Es bestehen diverse Abhängigkeiten, die die volle Auslenkbarkeit beschränken; siehe dazu Kapitel 2.2, S. 15.

2 Das Richtungshören

mehr ausgelenkt werden, sie springt, abhängig von IID und ITD, nach links oder rechts, eine Lokalisation zwischen den Quellen ist nicht mehr möglich. Das liegt darin begründet, dass der interaurale Kohärenzgrad für eine Quelle zu den Kopfseiten in abnimmt. Eine zu breite Positionierung der Lautsprecher führt zu einer Dekorrelation der Schallereignisse in den Ohren und damit dazu, dass keine Summenlokalisierung mehr möglich ist, für die eine hohe Kohärenz nötig ist.

Das bedeutet nicht, dass ein Hörereignis in Form von Phantomschaquellen nicht weiter als 30° zu den Seiten hin ausgelenkt werden kann. Es bedeutet aber, dass die Abstände zwischen den wiedergebenden Lautsprechern nicht größer als 60° sein dürfen. Mit mehreren Lautsprechern sind auch größere Winkel möglich. Allerdings wird das Hörereignis bei seitlicher Summenlokalisierung sehr unscharf, wie in Abbildung 2.9, S. 38 zu sehen ist. Unter realen Gegebenheiten betrachtet, springt es einfach zu dem Lautsprecher, der den höheren Pegel abstrahlt. Eine lineare Abhängigkeit zur interauralen Differenz ist nicht mehr gegeben. Die zu den Seiten hin stark abnehmende interaurale Kohärenz verschlechtert die Wahrnehmung massiv.

Die technische Wiedergabe durch Summenlokalisierung, bei der eine seitliche Lokalisation möglich sein soll, ist nicht sinnvoll. Insbesondere bei einer präzisen 3D-Wiedergabe sollten Hörereignisse möglichst nicht durch Summenlokalisierung dargestellt werden. Vor allem an den Seiten und hinter dem Hörer. Dreidimensionale Audiowiedergabe-Verfahren, die Summenlokalisierung nutzen, wie z.B. AURO 3D¹⁸ zeigen die genannte Schwäche deutlich. So verhält sich z.B. ein Auto, das während der Aufnahme seitlich an einem AURO-Mikrofonaufbau¹⁹ vorbei fuhr, als resultierendes Hörereignis so, dass es von den hinteren zu den vorderen Lautsprechern sprang, ohne dass eine Bewegung wahrnehmbar war (Erfahrungen des Autors) und belegt damit die Ergebnisse der Messungen, die in Abbildung 2.9, S. 38, zu sehen sind.

¹⁸Nach Gesprächen des Autors mit Wilfried Van Baelen, der das System federführend entwickelte, nutzt AURO3D bis AURO 13.1 nur Pegeldifferenzen und teilweise Laufzeitdifferenzen zur Darstellung von Phantomschallquellen zwischen den Lautsprechern. Eine anderer Form der Wiedergabe sei eventuell bei zukünftigen Systemen, die wesentlich mehr Lautsprecher diskret ansteuern, möglich. Van Baelen äußerte sich hier jedoch zurückhaltend.

¹⁹Im Prinzip ein sehr weiter 5-Kanal-INA-Aufbau, über dem in einem Abstand von etwa 70 cm ein IRT-Kreuz für die vier Höhenlautsprecher aufgebaut war.

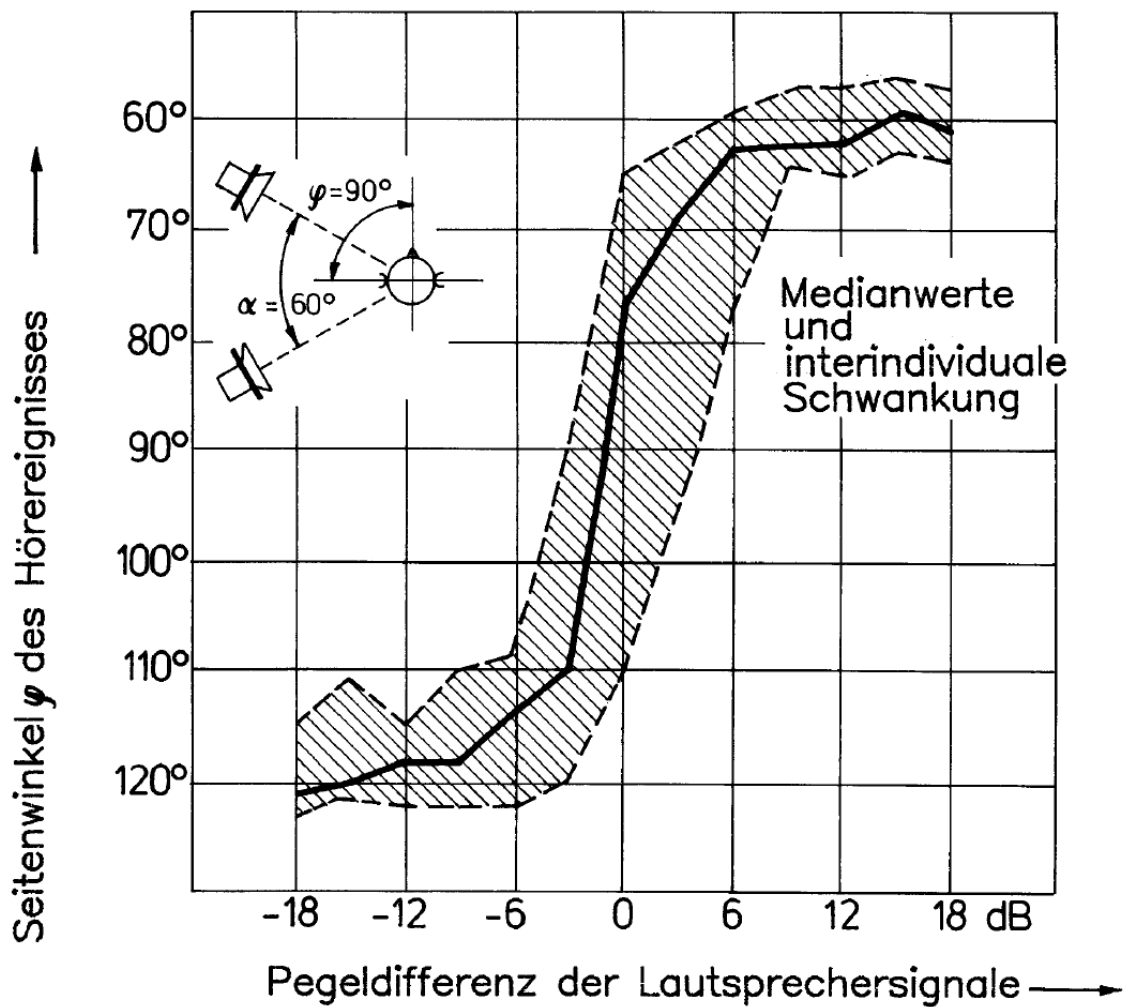


Abbildung 2.9: Summenlokalisierung bei seitlicher Lautsprecheranordnung (Blauert et al. 2008: S. 102).

2.9 Die Lokalisierbarkeit vermindernde Störfaktoren

Untersuchungen, auf die Blauert (1974) hinweist und z.T. selbst durchgeführt hat, zeigen, dass die Lokalisation sehr anfällig ist auf zu hohe oder zu niedrige Pegel im Bezug auf die Hörschwelle. Unter 30 dB SPL verschlechtert sich die Lokalisation (siehe Abb. 2.10, S. 39). Mit ein Grund dafür ist das Eigenrauschen des Gehörs. Bei 30 dB SPL ist die Lokalisationsschärfe am größten und nimmt hin zu 40 dB SPL etwas ab, bleibt dann aber lange Zeit konstant. Für die Schallfeldreproduktion zu Unterhaltungszwecken ist diese Schwelle aber weniger relevant, da in der Regel Pegel deutlich über 40 dB SPL genutzt werden, da die wiedergegebenen Töne sonst im Raumgrundrauschen²⁰, das besonders im Kino durch die anwesenden Personen höher ist, unterzugehen drohen.

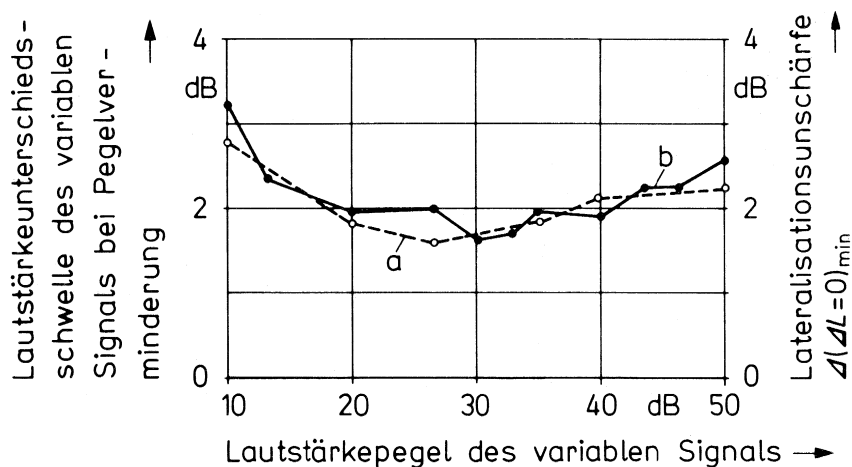


Abbildung 2.10: (a) Lautstärke-Unterscheidungsschwelle, (b) Laterationsunschärfe (Blauert 1974: S. 130).

Ein weitaus größeres Problem sind hohe Schalldrücke, die im Kino, vor allem bei zeitgenössischen Filmen, sehr viel und oft vorkommen. So sind im Kino z.B. Spitzenschalldrücke über 110 dB SPL²¹ möglich (Dolby Laboratories Inc. 2000). Ab einem Schalldruck von 90 dB SPL der Übertragungsweg von der Schallquelle zum Trommelfell aber bereits nicht mehr als linear anzusehen (Blauert 1974). Ab diesem Be-

²⁰Nach Erfahrung des Autors liegt das Grundrauschen im Kino meist über 40 dB SPL.

²¹Bei -20 dBFs RMS Signal-Pegel ist das System mit rosa Rauschen auf 85 dB(A), bei langsamer Integrationszeit einzumessen. Nach dem Messvorgang wird der Subwoofer um 10 dB im Pegel angehoben. Beachtet man den Crestfaktor des Rauschens, bleibt ein Headroom von ca. 12 dB, was 97 dB(A) Spitze entspricht. Weiter ist zu beachten, dass es sich um Signale handelt, die in der Messung integriert sind. Der Spitzen-Pegel ist somit wesentlich höher. Wird ebenfalls der A-Bewertungsfaktor und die Subwoofer-Anhebung berücksichtigt, kommt es zu Signalspitzen deutlich über 110 dB SPL.

2 Das Richtungshören

reich kann es zu starken Lokalisationsfehlern aufgrund von nichtlinearen Verzerrungen kommen.

Weiter ist zu beachten, dass das Signal für einzelne Lautsprecher, bei hohen Pegeln unter Umständen mechanisch durch die Lautsprechermembran oder elektronisch durch Schutzschaltungen limitiert wird. Das führt zusätzlich zu Verzerrungen bei Pegel und Frequenzgang und zu Unschärfe im Zeitbereich. Bei Pegeln über 90 dB SPL beginnt das Gehör nach wenigen Sekunden zu adaptieren. Die Adaption ist nach drei bis fünf Minuten vollständig abgeschlossen. Die Readaption beträgt ein bis zwei Minuten. Bei hohen Schalldruckexpositionen, länger als fünf Minuten, beginnt das Gehör zu ermüden. Dieser Effekt hält lange Zeit bis hin zu Tagen an und lässt nur langsam nach. Bei konstanten Signalen wie Rauschen oder Sinustöne kommt es, unabhängig von der Lautstärke, ebenfalls zur Adaption und Ermüdung (Blauert 1974, Ellermeier et al. 2008). Entsprechend ist eine längere Rezeption solcher Töne zu vermeiden.

Ermüdung und Adaption führen dazu, dass sich das Hörereignis immer mehr zur Mitte verschiebt und unschärfer lokalisiert wird (Blauert 1974). Gerade bei der 3D-Mischung muss für laute Szenen beachtet werden, dass die Lokalisationsleistung des Gehörs nachlässt. Laute Szenen sollten daher nicht zu lange andauern, um eine maximale Adaption und damit eine minimierte Lokalisierbarkeit zu vermeiden. Ebenfalls sollten zwischen lauten Szenen möglichst Pausen mit moderaten Pegeln (deutlich unter 90 dB SPL) liegen, damit es nicht zur Ermüdung des Gehörs kommt und es wieder readaptieren kann²².

²²In der aktuellen Vorführungspraxis in 3D-Audio wird nach Erfahrungen des Autors darauf meist viel zu wenig Rücksicht genommen.

3 Messungen der interauralen Kohärenz

Um weitere Erkenntnisse über die Differenzierbarkeit von Hörereignissen zu erlangen, wurde die interaurale Kohärenz in natürlicher Hörumgebung wie es z.B. im Kino vorkommt, untersucht. Dabei wurden unterschiedliche Quellen-Positionen und -Anzahlen berücksichtigt.

Das Tonlabor der HAW besitzt einen experimentellen Mischarbeitsplatz zur 3D-Audiomischung in HOA, der dafür ein gutes Testumfeld bietet. Es sind bereits 16 Lautsprecher (Schallquellen) vorhanden, die etwa sphärisch gleichmäßig verteilt sind. Die Akustik ist, wie aus Abbildung 3.1, S. 42 und Abbildung 3.2, S. 42 gefolgert werden kann, dem eines Kinos zumindest ähnlich. Durch akustische Absorber und Vorhänge an den Wänden sind die Reflexionen im Raum einigermaßen gut bedämpft.

Zur Messung des interauralen Kohärenzgrades wurde ein Kunstkopfmikrofon (Neumann KU 100) in der Mitte zwischen den Lautsprechern positioniert. Die mittlere Lautsprecherebene lag auf Ohrhöhe, der Abstand zu allen Lautsprechern war in etwa gleich. Eine detaillierte Aufstellung der Positionen der einzelnen Lautsprecher und der Konfiguration des Systems kann der Materialsammlung Tabelle A.3, Seite 97 und Tabelle A.4, Seite 98 im Anhang der Arbeit entnommen werden. Als Koordinatenursprung wurde der Standpunkt des Kunstkopfes auf dem Fußboden gewählt. Dieser war so aufgestellt, dass er zu allen Lautsprechern möglichst denselben Abstand hatte. Eine Dokumentation der verwendeten Geräte ist auf S. 96 zu finden. Alle im Rahmen der Arbeit erzeugten Daten, Audioaufnahmen und Tabellen sind auf der beiliegenden DVD zu finden. Die Ergebnisse der Datenauswertungen auf die Bezug genommen wird, sind soweit nötig ebenfalls in Tabellenform im Anhang der Arbeit zu finden. außerdem ist eine ausführliche Datensammlung in Tabellenform auf der beiliegenden DVD zu finden. Dort sind auch sämtliche Daten zu finden, die für dies Arbeit erstellt und erfasst wurden.

3 Messungen der interauralen Kohärenz



Abbildung 3.1: Vorderansicht des HOA-Mischplatzes.



Abbildung 3.2: Rückansicht des HOA-Mischplatzes.

3.1 Vorüberlegungen

Nach Blauert (1974) kann der interaurale Kohärenzgrad (Kor_{inau}) mit dem Maxima des Betrages der normierte Kreuzkorrelationsfunktion berechnet werden. Bei der genannten Funktion handelt es sich dabei lediglich um ein mathematisches Modell, das die Fähigkeit des Gehörs, Hörereignisse separieren zu können, nur grob nachbildet. Über die tatsächlichen Vorgänge im Gehirn ist noch nicht genug bekannt, um auf eine konkrete Funktion schließen zu können (Blauert 1974, Goldstein 2008). Nach den Untersuchungen von Blauert (1974) bildet die beschriebene Funktion jedoch eine gute Annäherung, die für die folgenden Untersuchungen genügt.

$$\Phi_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{\frac{1}{2T} \int_{-T}^{+T} x(t) y(t + \tau) dt}{\frac{1}{2T} \sqrt{\int_{-T}^{+T} x^2(t) dt \int_{-T}^{+T} y^2(t) dt}}$$

Abbildung 3.3: Gleichung zur Berechnung der normierten Kreuzkorrelation nach (Blauert 1974: S. 162)).

Die normierte Kreuzkorrelationsfunktion lässt sich, wie die Gleichung in Abbildung 3.3, Seite 43 erkennen lässt, berechnen. Aus dieser Funktion muss nun, wie die Gleichung 3.1, S. 43 zeigt, das Maximum des Betrages nach τ berechnet werden. Entsprechend dieser Funktion wurde ein Matlab Skript programmiert, um die interaurale Kohärenz mathematisch nachbilden zu können.

$$Kor_{\text{inau}} = \max_{\tau} |\Phi_{xy}(\tau)| \quad (3.1)$$

3.1.1 Anforderungen für eine automatisierte Datenauswertung

Für eine nachfolgende Auswertung soll das Skript automatisch Grafiken und Datensätze aus einer Wave-Datei erzeugen. Die interaurale Kohärenz soll in Schritten mit bestimmbarer Zeitlängen, also einzelnen Messfenstern, über die gesamte Dateilänge berechnet werden. Zum einen ist dadurch eine eher dem Gehör entsprechende Untersuchung möglich, da das Gehör nicht unbegrenzt Zeit zum Integrieren hat²³. Zum andern ergeben sich so zwei weitere Vorteile:

Erstens ist es möglich, durch starke Schwankungen Fehlmessungen zu identifizieren, bzw. durch die vielen Einzelmessungen, die über eine lange Messzeit gemacht werden, den Messfehler durch vereinzelte Störgeräusche klein zu halten. Die Messergebnisse hatte bei statischen Quellenpositionen eine Standardabweichung unter 0,06,

²³Die Länge dieser Integrationszeit ist bisweilen nicht bekannt. In der Literatur finden sich lediglich Vermutungen

lediglich bei Fehlmessungen kam es zu größeren Abweichungen. Diese Messungen wurden aussortiert. Die genannten Messfehler wurden vor allem durch Verkehrslärm und andere Lärmstörungen verursacht, die während der Messungen nicht immer wahrnehmbar waren.

Zweitens ist es möglich, die Veränderung der interauralen Kohärenz über die Zeit zu messen. So kann z.B. auch ein Ausschnitt aus einem in HOA-3D-Audio gemischten Film analysiert werden.

Das dem Überlegungen entsprechende Matlab-Skript ist im Anhang der Arbeit ab S. 102 und auf der DVD zu finden.

3.1.2 Auswahl des Mess-Stimuli

Damit die interaurale Kohärenz für mehrere Quellen bestimmt werden kann, muss das Signal aller Quellen zueinander inkohärent sein. Sonst würde die Kohärenz der Quellensignale die zu messende interaurale Kohärenz verändern. Wie sich die interaurale Kohärenz verändert, ist dabei von der Lage der Quellen zueinander abhängig (siehe dazu Kapitel 2.6.3, S. 30). Diese Abhängigkeit darf für die Messung jedoch nicht gegeben sein.

Prinzipiell wäre es möglich, Sinustöne zu benutzen, die im Bezug auf ihre Frequenz kein ganzzahlig teilbares Verhältnis und spektral genügend Abstand zueinander haben. Diese würden nur sehr gering korrelieren. Eine Messung mit zwei Tönen, 213 Hz und 749 Hz, erbrachte dabei einen guten geringen Kohärenzgrad von rund 0,0037 (Datei: Sinuston_213_749.wav). Allerdings bestünde die Gefahr, dass einige Töne stark mit Raumresonanzen interagieren. Messfehler wären nicht mehr auszuschließen. Ausschlaggebend ist aber, dass die HRTF auf Einzeltöne keine messbare Auswirkung hat. Die Untersuchung würde daher keine brauchbaren Daten liefern. Daher kann nur breitbandiges Rauschen als Stimuli genutzt werden.

Da rosa Rauschen die zu verwendenden Lautsprecher weniger beanspruchen würde²⁴, galt es diesen, Stimulus zu erst zu testen. Dazu wurden mit Audacity zwei Tonspuren die jeweils generiertes rosa Rauschen enthielten erstellt und als Stereo Wav-Datei exportiert. Anschließend wurde mit dem Matlab-Skript die Kohärenz gemessen. Dabei lag der Kohärenzgrad über einer Messlänge von 10 s bei rund 0,056 (Datensatzname: Rosa_unkorel). Weißes Rauschen hingegen lag mit rund 0,007 knapp eine Zehnerpotenz unter dem Kohärenzgrad von rosa Rauschen (Datensatzname: Weiss_unkorel).

Für weißes Rauschen gilt, dass es ein konstantes Leistungsdichte-Spektrum besitzt. Für ein unbegrenztes Spektrum würde daher die Autokorrelation Null ergeben. Dasselbe gilt für die Kreuzkorrelation von zwei Rauschsignalen. Für reale Signale gilt aber eine Begrenzung des Spektrums. Bei den Messungen ist dabei zum einen die Bandbegrenzung durch die Kopfmikrofone und die Lautsprecher ausschlaggebend. Zum

²⁴Durch die Dämpfung proportional zur Frequenz mit 3 dB pro Oktave, kann eine Überlastung der Hochtöner vermieden werden.

Ändern die Quantisierung und das Sampling bei der digitalen Aufnahme, was eine weitere systematische Begrenzung des Spektrums bedeutet. Da weißes Rauschen dadurch in seiner statistische Verteilung begrenzt ist, kann es immer wieder vorkommen, dass bei zwei Rauschsignalen zufällig Signalanteile ähnlich sind. Da für die interaurale Kohärenz jeweils das Maxima des Betrages der normierte Kreuzkorrelationsfunktion berechnet wird, führen diese zufälligen Ähnlichkeiten zu einem signifikanten Anstieg der Kohärenz. Dieser Effekt ist aber über das Spektrum des Aufnahmesystems von circa. 20 Hz bis 24 kHz sehr gering.

Rosa Rauschen wird mit einem Leistungsdichte-Spektrum definiert, das mit $1/f$ also 3 dB pro Oktave abnimmt. Rosa Rauschen entspricht demnach weißem Rauschen, das mit einem Tiefpass-Filter mit einer Steilheit von 3 dB pro Oktave gefiltert ist. Durch diese Filterung steigt die statistische Verteilung zu den tiefen Frequenzen hin. Die Kohärenz wird größer. Diese Zunahmen fällt insbesondere auf, wenn die Kohärenz über eine längere Zeit mit einer Messfenstergröße von 200 ms betrachtet wird. Wie in Abbildung 3.4, S. 46 zu sehen ist, schwankt die Kohärenz bei rosa Rauschen sehr stark, während sie bei weißem Rauschen ziemlich konstant ist. Beim Bilden eines Durchschnittswertes liegt weißes Rauschen mit 0,037 weit vor dem von rosa Rauschen mit 0,199.

Der erwartete interaurale Kohärenzgrad für zwei getrennte Hörereignisse soll nach (Blauert 1974) unter circa 0,2 liegen. Dieser Wert könnte für zwei getrennt lokalisierbare Quellen ebenfalls gegeben sein. Daher müssen die Mess-Stimuli deutlich unter diesem Wert liegen. Rosa Rauschen ist daher ungeeignet. Für die Messungen wurde aus diesem Grund, sofern nicht anders angegeben, weißes Rauschen benutzt.

Die Inkohärenz aller Rausch-Stimuli zueinander galt es ebenfalls festzustellen, um aussagekräftige Daten erheben zu können. Da es Fehleranfällig und Zeitaufwändig ist, jeweils ein Rauschsignal einzeln gegen alle anderen zu korrelieren, wurde ein anderer Ansatz gewählt²⁵. Da bekannt ist, dass die Kohärenz von zwei weißen Rauschsignalen theoretisch Null ist (praktisch ergab sich ein Wert von etwa 0,037) und der Überlagerungssatz gilt, kann eine Inkohärenz aller Rauschsignale zueinander einfacher festgestellt werden: Dazu wird jeweils ein Rauschsignal mit der Summe der übrigen verglichen. Dieses wird für alle 17 Rausch-Stimuli mit dem Matlabskript durchgeführt. Sofern sich mindesten zwei Rauschsignale ähneln, kommt es zur signifikanten Erhöhung des Korrelationsgrad. Er würde deutlich über 0,037 steigen.

Die 17 Messungen ergaben jedoch alle einen Durchschnittswert zwischen 0,0371 und 0,0373. Die Standardabweichung lag bei allen Messungen unter 0,003. Hier kann von einer üblichen Streuung, wie sie bei statistischen Signalen vorkommt, ausgegangen werden. Die Inkohärenz aller Stimuli ist so hinreichend erwiesen.

²⁵Bei 17 Rauschsignalen die jeweils miteinander korreliert werden, ergeben sich sonst mit $\binom{17}{2} = 136$ nötige Messungen.

3 Messungen der interauralen Kohärenz

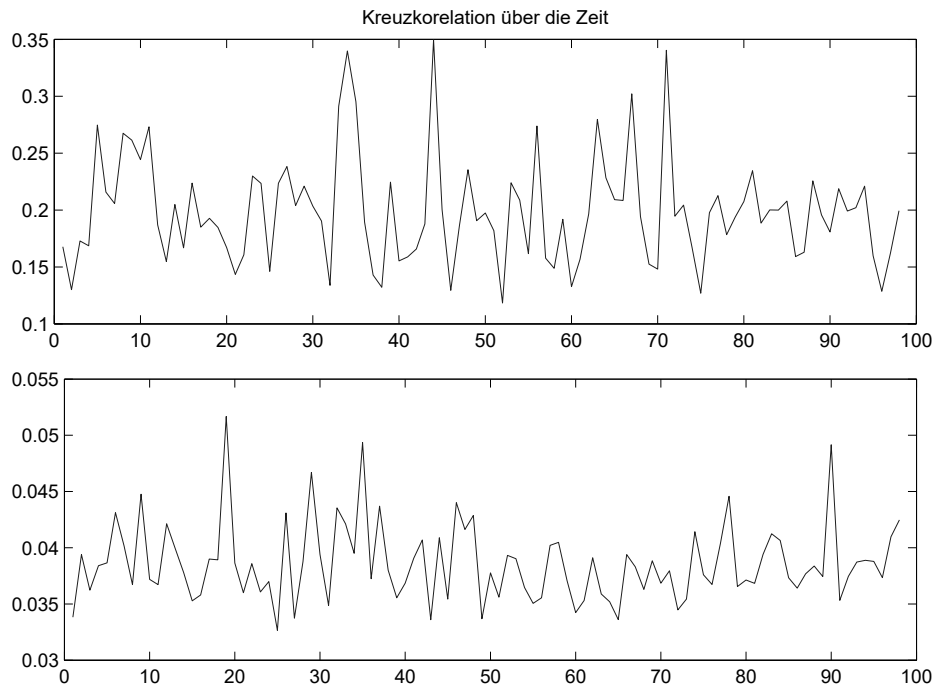


Abbildung 3.4: Vergleich des interauralen Kohärenzgrades von rosa Rauschen (oben) und weißem Rauschen (unten) über eine Länge von 10s.

3.1.3 Länge des Messfensters

Wie bereits beschrieben, wird zur Ermittlung der interauralen Kohärenz im Matlaskript mit einer einstellbaren Messfensterbreite gearbeitet. Die Länge des Fenster ist mit sec/n einstellbar. Dabei schreitet das Fenster nach und nach in Halblängenschritten die Probe ab. Die Fenster überdecken sich so jeweils zur Hälfte. Bei jedem Schritt wird eine Sampleprobe in Länge des Fensters genommen und das Maxima des Betrages der normierten Kreuzkorrelationsfunktion berechnet und für die jeweilige Fensterposition gespeichert. Am Ende wird aus allen Werten der Durchschnittswert, Minima, Maxima und die Standardabweichung berechnet. Zusätzlich wird eine Grafik erstellt, die die Änderungen des Kohärenzgrades über die Zeit anzeigt.

Um eine optimale zeitliche Auflösung bei geringstmöglichem Fehler zu erhalten, muss bestimmt werden, wie groß das Messfenster mindestens sein muss. Die Fensterung bedeutet, einfach betrachtet, eine Hochpassfilterung des Signals. Werden Rauschsignale, die durch das Messsystem bereits begrenzt sind, passgefiltert, steigt die Kohärenz zwischen ihnen. So steigt z.B., wie gezeigt, bei einem rosa Rauschsignal, das nichts anderes als ein Tiefpass-gefiltertes weißes Rauschsignal ist, ebenfalls die Kohärenz. Entsprechend führt eine Hochpassfilterung des weißen Rauschsignals durch die Fensterung genauso zu einer Steigerung der Kohärenz. In beiden Fällen wird die statistische Verteilung zu bestimmten Frequenzen hin begünstigt, was zu einer größeren Ähnlichkeit führt.

3 Messungen der interauralen Kohärenz

Um die optimale Fensterlänge zu ermitteln, wurden zwei unkorrelierte weiße Rauschsignale mit jeweils 10 s Länge mit Fensterlängen zwischen 12,5 bis 200 ms abgetastet und der durchschnittliche Kohärenzgrad ermittelt (Dateien: Weiss_unkorel 12ms5 bis Weiss_unkorel 200ms). Anschließend wurden die Werte in eine Diagramm eingezeichnet und eine Regressionskurve berechnet. Wie auf diesem Diagramm in Abbildung 3.5, S. 47 zu sehen ist, besteht ein negativ potentieller Zusammenhang zwischen der Fensterlänge und dem resultierenden Kohärenzgrad. Es ist gut zu erkennen, dass die Funktion bei 200 ms bereits stark abgeflacht ist. Theoretisch würde sogar bereits ein Fenster von 100 ms die Bildung der interauralen Kohärenz kaum negativ beeinflussen. Um sicher zu gehen, wurde eine Auflösung von 200 ms gewählt. Der erreichte Wert von 0,052 ist merklich größer als die Kohärenz über die gesamten 10 s mit 0,007. Die Kohärenz sinkt jedoch lediglich um den Faktor 7, während die zeitliche Auflösung um den Faktor 100^{26} steigt. Alle späteren Messergebnisse für die interaurale Kohärenz des Kunstkopfes lagen über 0,1. Die Kohärenz der Messstimuli ist damit noch genügend klein.

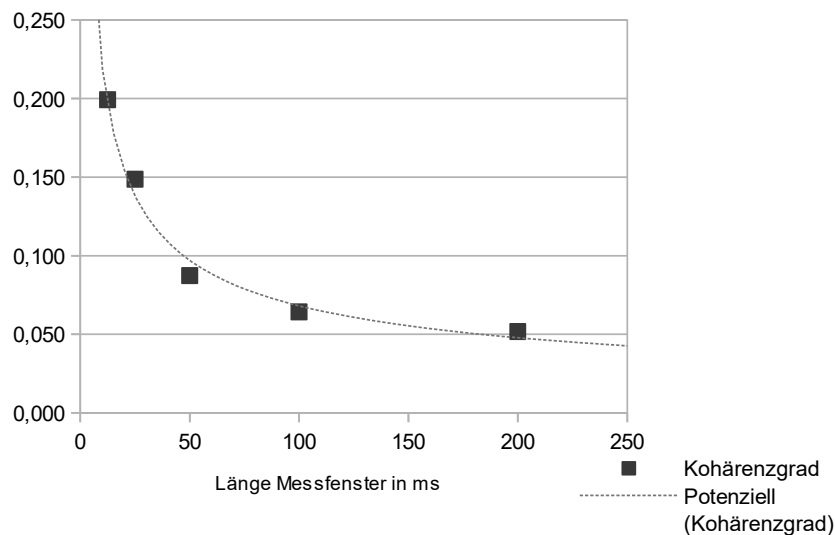


Abbildung 3.5: Werte für Kor_{inau} bei unterschiedlichen Fensterlängen.

Zur Validierung wurden zusätzlich Dateien mit der der Fensterung entsprechenden Länge erzeugt und vermessen (Dateien: 12ms5Nois bis 1000msNois). Wie Abbildung 3.6, S. 48 erkennen lässt, besteht ein ähnlicher Zusammenhang wie bei der gefensternten Messung über eine Datei von 10 s. Da hier für jeden Wert nur eine Messung ausgeführt wurde, sind die Daten für eine quantitative Aussage betreffend der Fensterlänge unbrauchbar. Für eine ähnliche Aussagekraft müssten mehr als 100 Proben mit jeweils extra generiertem Rauschen jeweils für eine der Fensterung entsprechenden Länge generiert werden. Das würde mehr als 500 Proben bedeuten. Dieser Aufwand

²⁶die Fenster überlappen um 50% daher verdoppelt sich der Faktor nochmals. Sonst würde $10^s/0,2s$ natürlich 50 ergeben

wäre nicht gerechtfertigt. Dennoch zeigt sich auch ohne Mehrfachmessung qualitativ, dass der negativ-potentielle Zusammenhang gegeben ist.

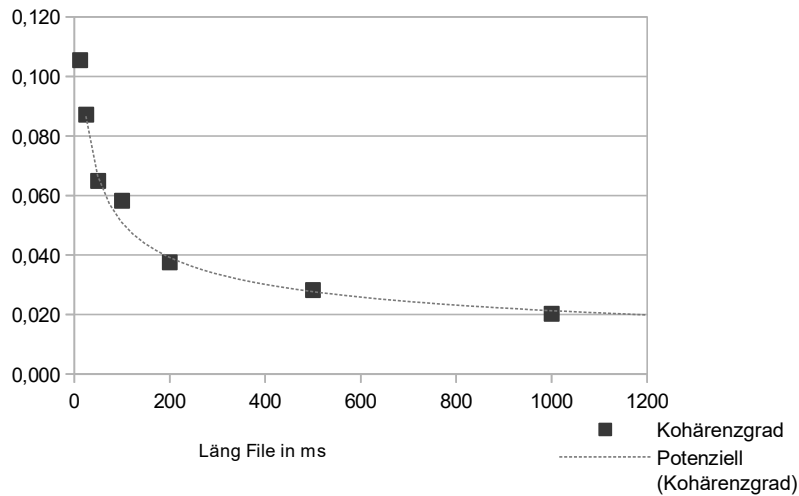


Abbildung 3.6: Werte für Kor_{inau} bei unterschiedlichen Filelängen.

3.2 Durchführung

Zur Durchführung der Messung wurde der Arbeitsplatz (siehe Abbildung 3.1, S. 42) in der Mitte der Lautsprechersphäre zur Seite geräumt. An dieser Stelle wurde später der Kunstkopf entsprechend der Dokumentation auf Seite 97, bezogen auf die Ohrachse in 1,20 m, auf Höhe der mittleren Lautsprecherebene montiert. Er wurde so ausgerichtet, dass der vordere und hintere Center-Lautsprecher der mittleren Lautsprecher-Ebene in dessen Medianebene lagen.

3.2.1 Kalibration des Messsystems

Zuerst wurden die Lautsprecher mit einem Schalldruckmessgerät (NTi XL2) auf einen möglichst gleichen Wert eingemessen. Der Schalldruckmesser wurde dazu genau an der Stelle des Mittelpunktes der Ohrachse, des sich später dort befinden Kunstkopfes, positioniert. Technisch bedingt gab es Abweichungen von etwa $\pm 0,7$ dB. Die Abweichungen lagen damit aber bereits im Bereich der Messungenauigkeit des Schalldruckmessers.

Nach dem Einmessen der Lautsprecher wurde der Kunstkopf aufgestellt. Für die Messung wurden die Mikrofonanschlüsse des Kunstkopfes an den Mikrofoneingängen eins und zwei des ersten Audiointerfaces angeschlossen. Beide Mikrofonverstärker wurden auf den gleichen Wert so eingestellt, dass es nicht zu Signalbegrenzungen

kommen konnte. Um etwaige Messunsymmetrien zwischen linkem und rechten Kunstkopfhör durch unterschiedliche Mikrofonkapsel-Empfindlichkeiten oder Abweichungen der Vorverstärker auszuschließen, wurde die Signale mit der genutzten Nuendo-DAW auf Pegelgleichheit überprüft.

Dazu wurde mit einem Testtongenerator (Sennheiser MZL 400), der über einen kalibrierten Lautsprecher einen definierten Pegel abgibt, ein Sinus-Testton direkt in die Gehörgänge des Kunstkopfs gespielt. Dabei wurde der Testtongenerator möglichst genau an der selben Stelle leicht an das linke und später rechte Ohren gedrückt. Dabei wurde jeweils der Pegel in Nuendo gemessen. Eine Abweichung war dabei nicht zu erkennen bzw. tendierte mit Differenzen unter einem Dezibel mal zum einen mal zum anderen Ohr. Diese kleine Abweichung ist unter den gegebenen Umständen mehr als ausreichend.

3.2.2 Messvorgang

Im Nuendo wurde jedem Lautsprecher fest je eines der 17 nicht-kohärenten Rauschsignale zugeordnet. Eine Ausnahme bilden einige Messungen, die bewusst mit kohärentem Rauschsignalen gemacht wurden. Dazu wurde ein einziges Rauschsignal gleichzeitig auf alle zu messenden Lautsprecher ausgespielt. In Nuendo wurde das Projekt so eingerichtet das die gewählten Lautsprecher bereits zwei Sekunden lang das Rauschsignal abspielten, bevor die Aufnahmen der beiden Kunstkopfsignale beginnt. So wird gewährleistet, dass der Raum zum Zeitpunkt der Messung bereits eingeschwungen ist. Die Aufnahme wird nach 10 s^{27} zusammen mit der Wiedergabe automatisch beendet. Von einer kurzen technischen Blende am Anfang und Ende wurde abgesehen. Bezogen auf die zu messende Kohärenz ist der Fehler bei hartem Abschneiden, insbesondere bei Rauschen, geringer als beim Blenden. Die Länge von 10s wurde gewählt, da so pro Messung 100 Messwerte bei der unter Kapitel 3.1.3, Seite 46 genannten Messfensterlänge zur Verfügung standen.

Durch die hohen Energiegehalte im oberen Spektrum des genutzten weißen Rauschens musste eine Limitierung oder gar Beschädigung der Hochtöner der genutzten Lautsprecher ausgeschlossen werden. Insbesondere eine Limitierung der Hochtonsignale durch die mögliche Überlastung hätte die Messung verfälscht. Daher wurde nicht sehr laut mit einem Ausgangspegel von -38 dBFS pro Lautsprecher gemessen. Die genutzten Lautsprecher hatten eine Eingangsempfindlichkeit von $+4\text{ dBu}$. Das Audiointerface gibt in der gewählten Einstellung bei 0 dBFS einen Pegel von $+13\text{ dBu}$ ab. Die einzelnen Lautsprecher wurden dementsprechend mit -29 dB unter Vollaussteuerung betrieben, was in etwa 70 dB SPL pro Lautsprecher entsprach.

²⁷Genau genommen waren es $10,0345\text{ s}$, wie sich später bei der Auswertung herausstellte. Nuendo scheint nicht immer so genau zu sein.

Durch die geringere Lautstärke kam es immer wieder zur Störung der Messungen durch Umgebungsgeräusche. Sofern dies bei der Messung auffiel, wurde sie wiederholt. In anderen Fällen wurde der Messsatz später entfernt, wenn sich die Fehler als signifikant herausstellten.

Ebenfalls wurde eine in HOA-3D-Audio gemischte Filmsequenz über den Kunstkopf aufgenommen. Ferner wurden ein Rauschsignal mit verschiedenen Panning-Techniken einmal komplett um den Kunstkopf herum geführt. Dabei wurde zum Einen Summenlokalisierung und zum Andern ein HOA-Panning 3. Ordnung benutzt.

3.2.3 Bewertung des subjektiven Höreindrucks

Nachdem alle Messungen durchgeführt worden waren, wurde der Kunstkopf abgebaut und beiseite geräumt. Anschließend nahm der Autor so zwischen den Lautsprechern platz so, dass seine Kopf in etwa die Position des Kunstkopfes einnahm. Darauffolgend wurden verschiedene Stimuli angehört und Auffälligkeiten notiert. Das Protokoll dazu ist im Anhang zu finden.

3.3 Datenauswertung

Die ursprüngliche Version des Matlab-Skripts wertete jeweils nur eine Datei aus, zeichnete zwei Figuren, um den Verlauf der Kohärenz über die Zeit darzustellen und gab die Messwerte im Kommandofenster aus. Die Daten hätten alle manuell gespeichert und übertragen werden müssen. Bei über 70 Messätzen die jeweils 2 Grafiken und mehrere Ergebniswerte gerierten, war diese manuelle Methode zu zeitaufwendig.

Daher wurde das Skript mit einigem Recherche- und Zeitaufwand modifiziert. Im Ergebnis muss nun lediglich ein Ordner angegeben werden, in dem alle Wav-Dateien automatisch nacheinander ausgewertet werden. Die generierten Grafiken werden automatisch in einem Unterverzeichnis des angelegten Ordners in verschiedenen Formaten gespeichert. Ebenfalls werden dort alle Auswertungsergebnisse des Datensatzes in verschiedenen Formatierungen gespeichert.

Diese Datensätze wurden anschließend in ein Tabellenverarbeitungsprogramm eingelesen, formatiert, weiter ausgewertet und zur bessern Verständlichkeit grafisch aufbereitet.

3.4 Schlussfolgerungen

Bei der ersten Untersuchung der Datenlage ergab sich für einzeln gemessene Schallquellen, dass sie in etwa den Kohärenzgraden entsprach, die Blauert (1974) angab. Allerdings lagen die Messergebnisse für die jeweilige Horizontalposition der Quellen unter den Messergebnissen von Blauert (1974). Dies lässt sich mit der Akustik des Raumes erklären.

3.4.1 Auswirkungen durch die Raumakustik

Die Messung der interauralen Kohärenz für den vorderen Centerlautsprecher der mittleren Ebene betrug lediglich 0,77, für den oberen mittleren Center über dem Kunstkopf ergab sich ein Wert von 0,84. Theoretisch hätten beide Werte bei 1 liegen müssen. Es zeigte sich auch, dass alle linken Lautsprecher ausnahmslos geringere Werte aufwiesen als die rechten (Die Messergebnisse aller Lautsprechersignale finden sich in der Tabelle der Materialsammlung auf Seite 99). Die Messungen für die jeweils gegenüberliegenden Lautsprecher hätten jedoch gleich sein müssen. Bei der genauen Betrachtung der Wiedergabesituation war festzustellen, dass alle linken Lautsprecher sehr nah an einer Wand oder in Raumecken standen. Dadurch wurden Reflexionen begünstigt. Bei dieser Betrachtung zeigte sich auch, warum bei der Nutzung des oberen mittleren Centers bessere Werte zu messen waren, als bei der des vorderen Centerlautsprechers. Der obere Lautsprecher hatte einen deutlich größeren Abstand zu allen akustischen Begrenzungsflächen. Sehr klar ist dieser Trend auch zu erkennen, wenn die Werte für die obere und untere Lautsprecherebene in Abbildung 3.8, Seite 53 verglichen wird. Theoretisch müssten auch hier die Werte in etwa gleich sein. Durch die Nähe zum Fussboden ergeben sich aber für die unteren Lautsprecher auf der linken wie auf der rechten Seite schlechtere Werte. Dabei gelten für beide Lautsprecherebenen, dass links durch die größere Nähe zur Wand schlechtere Ergebnisse entstehen.

Somit lässt sich zeigen, dass die Akustik einen entschiedenen Einfluss auf die Lokalisierbarkeit hat. Bereits in dem gemessenen Raum, der keine allzu schlechte Akustik hat, sinkt die maximale Punkt-Lokalisationsschärfe (näheres dazu im Kapitel 5.2.3 Seite 82).

3.4.2 Vergleich der Wahrnehmung mit den Messwerten

Schon ab zwei Rauschquellen war eine genaue Lokalisierung schwierig. Durch Kopfbewegungen war es aber noch gut möglich, beide Hörereignisse voneinander zu differenzieren. Bei drei und vier Rauschquellen war die Differenzierbarkeit stark von der Lage abhängig und nur durch leichte Kopfbewegung möglich. Der mittlere vordere Centerlautsprecher war gut von den Lautsprechern seitlich links und rechts auf der mittleren Ebene zu unterscheiden. Allerdings wirkten die Hörereignisse schon sehr diffus und breit. Leider war die dazugehörige Messung wegen der darin enthaltenen Störungen nicht auswertbar. Eine messtechnische Bewertung und damit ein Vergleich ist an dieser Stelle nicht möglich.

Bei mehr als vier Rauschquellen war, wenn sie gleichmäßig um den Zuhörer verteilt waren, nur noch etwas diffuses zu hören, das beschrieben wurde als nicht lokalisierbare und unzählbare Hörereignisse. Insbesondere bei der Nutzung der Lautsprecher der Höhen- und Mittelebene entstand ein Gefühl starker Umhülltheit. Durch die Bodenlautsprecher wurde dieses Gefühl kaum besser.

3 Messungen der interauralen Kohärenz

Bei einer Verteilung in nur einer Hemisphäre war nur wahrnehmbar, dass unzählbar viele Hörereignisse in dieser Hemisphäre lagen.

Bei korrelierten Rauschsignalen von mehr als zwei Quellen²⁸, die um den Zuhörer gleichmäßig verteilt waren, entstand der Eindruck, dass ein Hörereignis wahrnehmbar, aber nicht lokalisierbar war. Dabei scheinen einige Frequenzen überbetont zu sein. Es war eine deutliche Klangfärbung zu vernehmen.

Beim messtechnischen Vergleich von korrelierten und unkorrelierten Rauschsignalen, die über mehrere Quellen abgespielt wurden, fällt auf, dass bei korrelierten Rauschsignalen einen signifikant höherer Kohärenzwert messbar war (Einen Vergleich bieten Abbildung 3.7 und 3.10, Seite 52 und Seite 56).

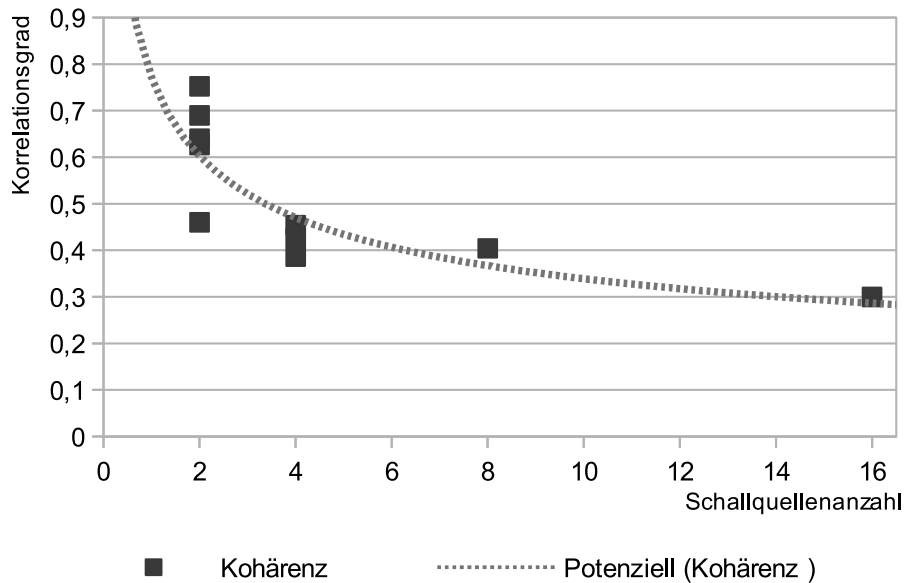


Abbildung 3.7: Werte für die interaurale Kohärenz in Abhängigkeit von der Lautsprecheranzahl bei kohärentem Rauschsignalen der Quellen.

Wird davon ausgegangen, dass, wie im Kapitel 1.3.2, Seite 13 beschrieben ist, die Schallereignisse spektral zerlegt werden, kann auch davon ausgegangen werden, dass für diese einzelne Spektren eine Kohärenz gebildet wird. Eine entsprechende Vermutung stellt Blauert (1974) auf.

Bei korrelierten Rauschsignalen aus mehreren Quellen ist eine deutliche Klangfärbungen zu hören. Diese kann nur durch Interferenzen entstehen, die durch die unterschiedlichen Signallaufzeiten der Lautsprecher zu den einzelnen Ohren bedingt sind. Dabei entsteht ein Kammfiltereffekt, der bestimmte Frequenzbänder verstärkt. Bei diesen Bändern scheint es zu hohen Kohärenzen zu kommen. Entsprechend werden sie zu einem Hörereignis gruppiert. Da aber die Richtungsinformationen für die einzelnen Frequenzbänder nicht übereinstimmen, bleibt die Wahrnehmung diffus.

²⁸bei zwei Quellen kam es in der Regel zur Bildung einer Phantomschallquelle oder zu zwei sehr diffusen Hörereignissen, gelegentlich kam es zu IKL, sofern der Kopf nicht bewegt wurde

Bei unkorrelierten Signalen kam es zu Hörereignissen, die weder zählbar noch in ihrer Richtung bestimmbar waren. Vermutlich konnte hier durch die komplett fehlende Kohärenz in den Frequenzbändern kein einzelnes Hörereignis ausgebildet werden. Stattdessen gilt für jedes Band, dass es zu mehreren Hörereignissen gehören muss. Denn nach Blauert (1974) sind unkorrelierte Bänder immer Teil mehrerer Hörereignisse. So entsteht die Wahrnehmung unzählbarer Hörereignisse.

3.4.3 Kohärenzgrade einzelner Quellen

Bei der Betrachtung der Kohärenzgrade, siehe Abbildung 3.8, Seite 53, zeigte sich ein Zusammenhang zwischen lateraler Auslenkung einer Quelle und dem Kohärenzgrad. Blauert (1974) wies zwar darauf hin, dass dieser Zusammenhang besteht, führte dazu aber keine weiteren Messungen durch, die eine quantitative Aussage zulassen.

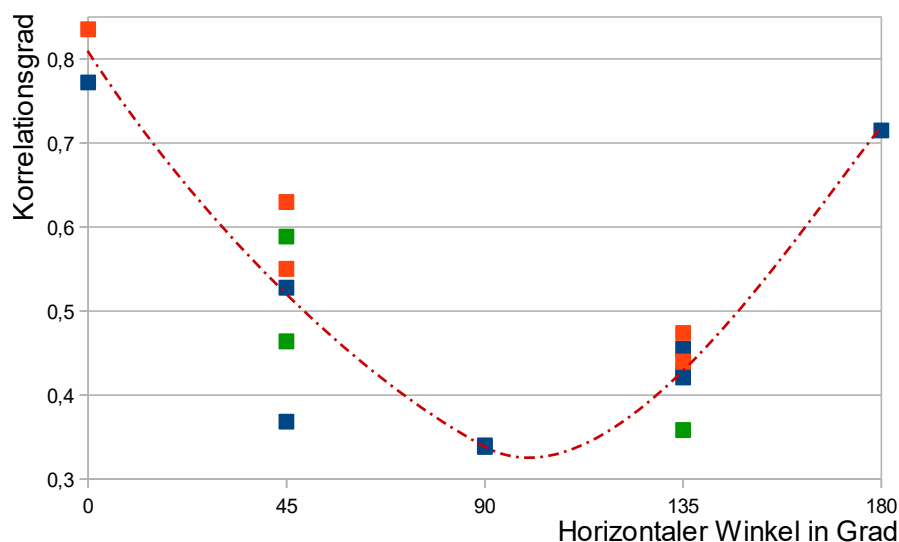


Abbildung 3.8: Werte für die interaurale Kohärenz in Abhängigkeit von der Quellenlage. Die oberen Lautsprecher sind rot, die mittleren blau und die unteren grün eingezeichnet. Die Regressionsfunktion ist dunkelrot gestrichelt eingezeichnet. Es gilt, dass die linken Quellen an der Medianebene gespiegelt dargestellt sind, es gilt $90^\circ = 270^\circ$

Aus Kapitel 2.2 Seite 15 ist bekannt, dass ab einer horizontalen Quellenauslenkung von 30° die maximale Lokalisationsfähigkeit seitlich gleichbleibend bei einer Unschärfe von etwa 10° liegt.

Für die seitliche 90° Lage ergab sich ein Kohärenzgrad von rund 0,34. Für $\pm 45^\circ$ ergaben sich jeweils unterschiedliche Werte, hier führte die Asymmetrie des Raumes zu Abweichungen. Der beste Wert ergab rund 0,54 der schlechtere 0,37. Der Durchschnitt entsprach 0,45. Der Wert für 30° wurde linear zwischen 0° und 45° interpoliert

und ergab 0,53. Nach den durchgeführten Messungen würde das bedeuten, dass ab einem interauralem Kohärenzgrad zwischen 0,53 bis 0,34 die Lokalisationsunschärfe 10° beträgt.

Durch die gegebene Raumakustik waren die Messwerte durchweg geringer als bei den Messungen, die Blauert (1974) unter idealen Bedingungen durchführte. Nach seinen Messungen lag für eine seitliche Auslenkung von 90° der minimale interaurale Kohärenzgrad bei etwa 0,4. Für eine frontale Quelle in der Medianebene bei eins. Für weitere Positionen (insbesondere 30° und 45°) führte er keine Messungen durch. Beim Vergleich der Werte für 0° und 90° Lage lag der Offset²⁹ zwischen Blauerts und den selber erstellten Messwerten bei rund 0,2 und 0,1.

Aus diesen korrigierten Werten lässt sich schließen, dass für Signale die eine Kohärenzen zwischen etwa 0,6 und 0,4 erzeugen eine Lokalisationsunschärfe von 10° in der Horizontalebene gilt.

Nach Blauerts Ergebnissen liegt der Schluss nahe, dass die Lokalisationsschärfe unter einem interauralen Kohärenzgrad von 0,4 weiter sinkt. Blauert (1974) zitiert einen Versuch, der klar einen Zusammenhang zwischen dem sinkenden Kohärenzgrad und einer kleiner werdenden Lokalisationsschärfe belegt. Bis zu einem Kohärenzgrad von etwa 0,3 bestand dabei ein linearer Zusammenhang (siehe Abbildung 3.9, Seite 55). Unter 0,2 steigt die Funktion stark an. Wobei ab diesem Wert zwei getrennte Hörereignisse wahrgenommen wurden, von dem keines durch Laufzeitdifferenz der Kopfhörersignale lateral auslenkbar war. Da durch diesen Effekt keine sinnvolle Lokalisation mehr möglich ist, kann dieser Wert als Grenzwert gelten.

Da diese Versuche mit Kopfhörern gemacht wurden, sind sie nur unter bestimmten Gesichtspunkten auf Schallquellen im Raum zu übertragbar. So besteht im Raum immer eine gewisse interaurale Kohärenz. Wie die Messungen gezeigt haben, kann diese aber durchaus signifikant unter 0,2 liegen. In der Messung wurden Werte unter 0,1 ermittelt. Für das Hören im Raum kommt es bei Kohärenzen unter 0,2 jedoch nicht zum Hören von zwei Hörereignissen³⁰, sondern – nach Erfahrungen des Autors während der Tests – zu Hörereignissen, die weder in ihrer Anzahl noch in ihrer Richtung bestimmt werden konnten. Jambrosic et al. (2005) konnten in ihren Untersuchungen Lokalisationsunschärfen von über 60° verzeichnen. Daher kann davon ausgegangen werden, dass, bevor unter 0,20 keine Lokalisation mehr möglich ist, eine Lokalisationsunschärfe von mindestens 60° erreicht wird.

Auch die nach hinten etwas schlechtere Lokalisationsschärfe konnte am schlechteren Kohärenzgrad (rund 0,1 schlechter) für den hinteren Center im Vergleich zum vorderen gezeigt werden.

²⁹Gemittelt jeweils über die Center- und die Seitenlautsprecher.

³⁰Eventuell habe die zwei Hörereignisse bei der Messung die Blauert (1974) angab, ihre Ursache in den zwei Kopfhörerlautsprechern.

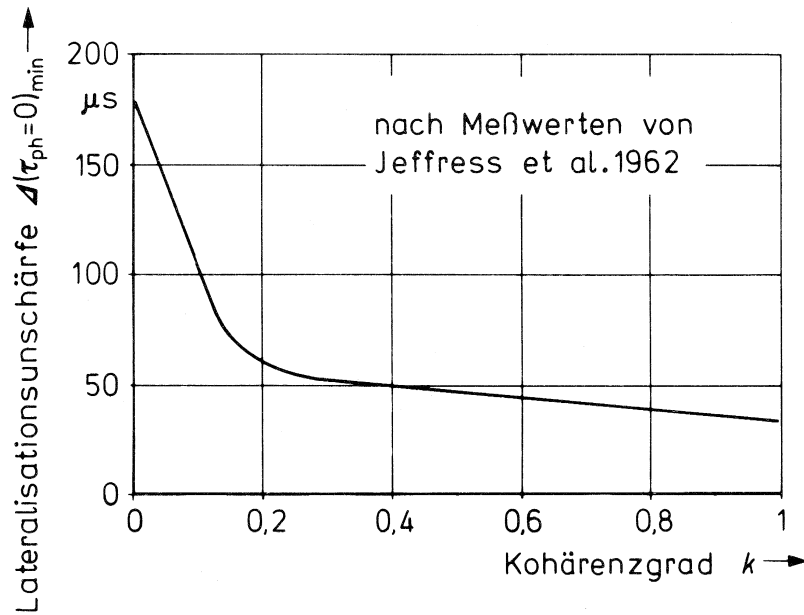


Abbildung 3.9: Laterationsunschärfe als Funktion des Kohärenzgrades der Ohrsignale. Sie zeigt, bei welcher Verzögerung eines Kopfhörer-Signals eine seitliche Auslenkung bemerkt wurde. Wobei bei $k=0$ auch bei $190 \mu s$ keine Auslenkung mehr feststellbar war (Blauert 1974: S. 36).

3.4.4 Kohärenzgrade mehrerer Quellen

Sofern nicht eine der beteiligten Quelle der vordere oder hintere Centerlautsprecher war, fiel die Kohärenz bereits bei zwei Quellen unter 0,6. Bei mehr als drei Quellen lag die Kohärenz, sofern nicht ein Centerlautsprecher beteiligt war, unter 0,4. Ab acht Quellen fiel der Kohärenzgrad unter 0,2.

Nach den Ausführungen in Kapitel 3.4.3 Seite 53 lässt sich daraus schließen, dass mit zunehmender Quellenanzahl die Lokalisationschärfe fällt, da die interaurale Kohärenz ebenfalls abnimmt. Für Werte unter 0,2 kann es nach diesen Ausführungen zu keiner Lokalisation mehr kommen.

Die Lokalisationsunschärfe liegt somit nach den Messungen zum Teil schon bei zwei Quellen über 10° . Wie in Abbildung 3.10, Seite 56 zu erkennen ist, können sogar im Extremfall bereits zwei Quellen nicht mehr lokalisiert werden. Das ist der Fall, wenn je eine Quelle links und rechts genau seitlich des Kopfes liegen. bei spätestens acht Quellen ist keine Lokalaistion mehr möglich, da der Grenzwert von 0,2 unterschritten ist.

Berücksichtigt man jedoch die Beeinflussungen des Raumes, kann es unter besseren akustischen Bedingungen durchaus möglich sein, dass bei mehr als 8 Quellen eine Kohärenzen über 0,2 messbar ist. Bei acht und mehr Quellen kann es aber durch den zu geringen Abstand zueinander zu Verschmelzungen der Hörereignisse kommen.

Dennoch lässt sich keine allgemeingültige Aussage ableiten, da die Messstimuli

3 Messungen der interauralen Kohärenz

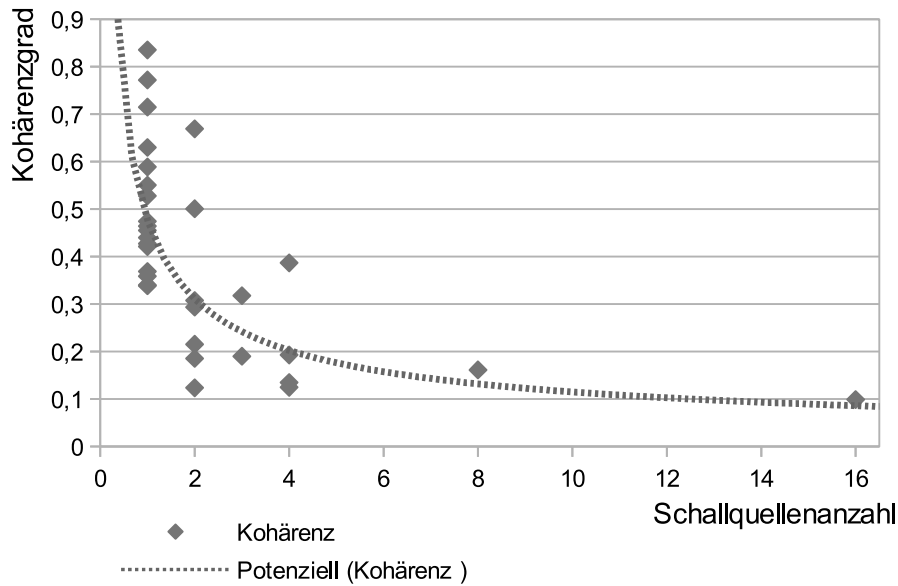


Abbildung 3.10: Werte für die interaurale Kohärenz in Abhängigkeit von der Lautsprecheranzahl bei inkohärentem Rauschsignalen der Quellen.

breitbandige Rauschsignale waren. Sofern sich die Spektren mehrerer Quellen nicht überlagern, müsste jede Quelle genau lokalisierbar sein.

Das gilt aber nur so lange, wie nicht einzelne Schallquellen zu Hörereignissen verschmelzen und damit eine diffuse Lokalisation nur eines Hörereignisses entsteht³¹. Quellen, die nahe beieinander liegen, bilden immer interaurale Teilkohärenzen. Dabei ist es nicht von Belang, ob die Signalspektren auseinander liegen. Denn in beiden Ohren entsteht ein Mischsignal, das eine spektrale Zergliederung in verschiedene Hörereignisse nicht ohne weiteres zulässt. So entsteht ein breites Hörereignis. Nahe beieinander bedeutet dabei, dass der horizontale Winkelabstand der Quellen kleiner oder gleich der für die Richtung geltenden Lokalisationsunschärfe ist.

Dass eine spektrale Unterscheidbarkeit von Quellensignalen deren Differenzierbarkeit in Hörereignisse begünstigt, gälte damit nur solange, wie ihr räumlicher Abstand zueinander groß genug ist. Die Messungen zeigen, dass in der Regel bei 4 Quellen bereits eine Verschmelzung einzelner Signale zu einem Hörereignis entsteht. Für eine korrekte Aussage jedoch Kapitel 4, Seite 59 berücksichtigt werden³².

³¹Es sei auf Kapitel 2.6.3, Seite 30 hingewiesen, wonach Quellen, die nahe nebeneinander liegen, auch bei Inkohärenz ein teilkohärentes Schallereignis erzeugen, was zu einem Hörereignis mit breiter Lokalisation führt

³²Es ist zu beachten dass für die Messung Rauschsignale benutzt wurden, die kein erkennen von Mustern zulassen. Für natürliche Signale gilt das nicht. So verschmilzt das Signal einer Trompete und eines Saxophons, die etwa gleich laut spielen und nah beieinander stehen, nicht unbedingt zu einem Hörereignis. Durch Mustererkennung kann der auditorische Kortex beide Signale wieder trennen und dem jeweiligen Instrument zuordnen. Die Lokalisation wird aber ungenauer sein als

Wie groß die Lokalisationsunschärfe ist, hängt von der Raumakustik ab, von der Quellenanzahl und, ob die Quellen seitlich oder nahe der Medianebene liegen. Da die Parameter teilweise abhängig sind, können deren einzelne Teil-Werte für die Lokalisationsschärfe nicht einfach summiert werden. Daher kann nur gesagt werden, dass die Lokalisationsunschärfe auf jeden Fall kleiner ist als die des signifikantesten Einzelparameters.

Würde z.B. die Raumakustik als Parameter mit dem größten Einfluss eine bestmögliche Kohärenz von 0,6 für den Raum ergeben, würde eine Quelle selbst auf der Medianebene nur noch mit einer Genauigkeit von $\pm 10^\circ$ wahrgenommen werden können. Kommt eine zweite seitliche Quelle hinzu, kann nach dem aktuellen Erkenntnisstand nur gesagt werden, dass die seitliche Quelle eine größere Lokalisationsunschärfe besitzt als die in Medianebene. Die in Medianebene kann durch die 2. Quelle wesentlich schlechter als mit $\pm 10^\circ$ Genauigkeit wahrgenommen werden. Mit dieser Argumentation ist zumindest qualitativ eine Verschlechterung der Lokalisierbarkeit der beteiligten Quelle feststellbar.

3.4.5 Aussagen zur Lokalisierbarkeit und Differenzierbarkeit

Vor der Auswertung der Messungen bestand die Annahme, dass, je kleiner der interaurale Korrelationsgrad ist, desto mehr Quellen können voneinander differenziert werden. Diese Annahmen bestätigt sich nur zum Teil. Sobald der interaurale Kohärenzgrad unter 0,2 fällt, ist keine sichere Differenzierung einzelner Quellen mehr möglich. Wie viele Quellen dabei maximal differenziert werden können, ist von ihrer Lage zueinander und von deren Signaleigenschaften abhängig.

Weiter bestand die Annahmen, dass die Lokalisierbarkeit mit fallendem Kohärenzgrad sinkt. Dies konnte bestätigt werden. Dabei ist davon auszugehen, dass eine Lokalisierbarkeit unter 0,2 nicht mehr gegeben ist.

Bekräftigt werden die Messergebnisse auch durch Messungen der 3-D-Audiowiedergabe eines Films, der experimentell in HOA 3. Ordnung gemischt worden ist. Das Material wurde über die Anlage wiedergegeben, in der auch gemessen wurde. Wie in Abbildung 3.11, Seite 58 zu erkennen ist, liegt der interaurale Kohärenzgrad zumeist deutlich über 0,6 und erreicht teilweise Kohärenzen nahe eins. Zumindest horizontal war eine sehr gute Lokalisation möglich. In der Höhe war die Lokalisationsschärfe deutlich geringer. Bei 3 bis 4 Quellen war eine deutliche Verschlechterung der Lokalisierbarkeit zu bemerken. Bereits bei der Mischung, die vom Autor dieser Arbeit durchgeführt wurde, war zu bemerken, dass maximal 4 Quellen einigermaßen lokalisierbar waren und nicht in der Gefahr standen, als Hörereignis zu verschmelzen. Daher wurde darauf geachtet, dass möglichst wenig virtuelle Quellen gleichzeitig lokalisierbar positioniert wurden. Es wurde sehr viel mit einem diffusen Grund, der nur schwach Richtungen erkennen ließ, gearbeitet und wenigen virtuellen Quellen im

bei nur einer Quelle. näheres dazu in Kapitel 4, Seite 59.

3 Messungen der interauralen Kohärenz

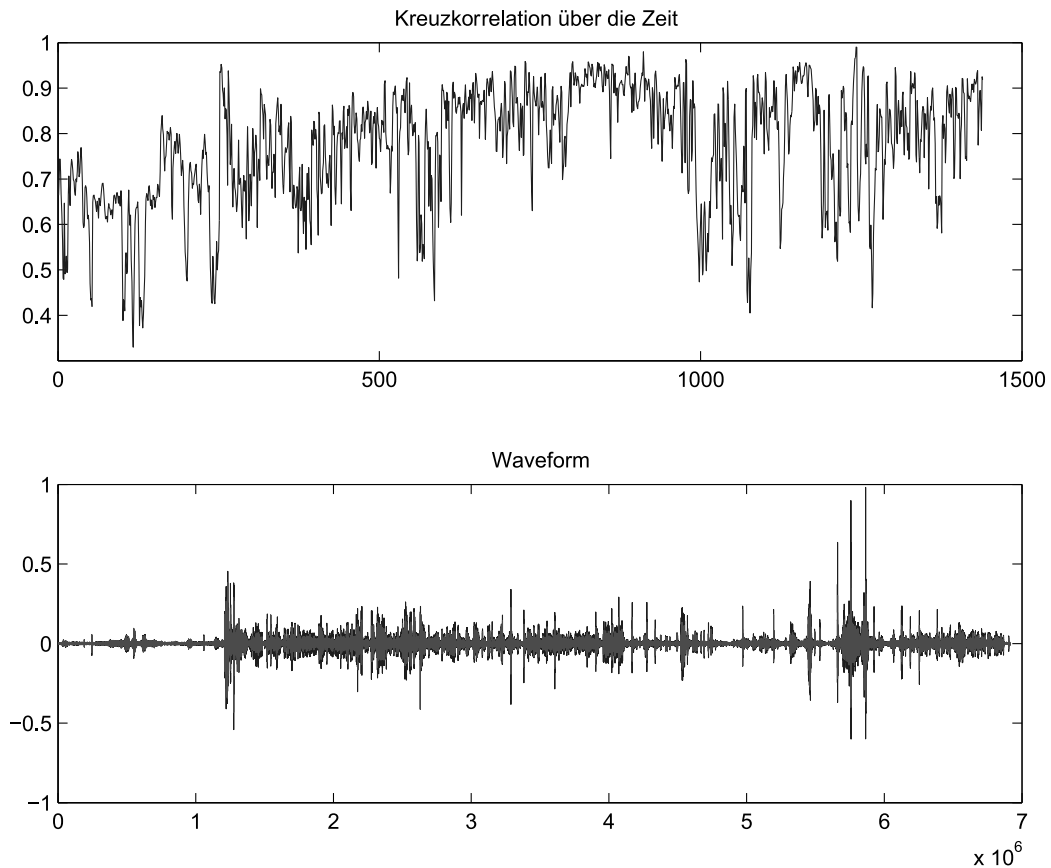


Abbildung 3.11: Werte für die interaurale Kohärenz über die Zeit (oben) und Waveform der Ohrsignale (unten) für die Aufnahme einer Filmsequenz in HOA-3-D-Audio mit einem Kunstkopf.

Vordergrund³³. So wurde eine maximal Lokalisierbarkeit begünstigt.

Nach diesen Erfahrungen und den Auswertungen der Daten liegt der Schluss nahe, dass zumindest aus psychoakustischer Sicht lediglich 2 bis 8 Quellen differenzierbar sind. Wobei die Lokalisationsschärfe mit steigender Anzahl stark abnimmt. Dabei gilt, dass die Quellen nur differenzierbar sind, sofern ihre Abstände zueinander groß genug sind. Für mehrere Quellen gilt, dass ihre Lokalisationsunschärfe größer ist als 10° und mit der Anzahl der Quellen weiter steigt.

Beider Entwicklung des Modelles in Kapitel 4, Seite 59 zeigte sich jedoch, dass das nutzen von mehreren Rauschquellen einen Spezialfall des Hörens bedeutet. Denn durch das Rauschen kann keine oder eine nur sehr eingeschränkte Separation von Hörereignissen stattfinden. Die Differenzierbarkeit nimmt stark ab, dadurch verschmelzen die Signale mehrere Quellen zu einem Hörereignis, dass nur noch eine Phantomschallquelle wahrnehmen lässt. Daher kann mit den Ergebnissen der Messung nur bedingt eine Aussage getroffen werden.

³³Durch deutlich höhere Pegel vom Hintergrund abgesetzt.

4 Modell zur Wahrnehmung eines Hörereignisses

Damit der Eindruck eines Hörereignisses entsteht, bedarf es der Wahrnehmung des entsprechenden akustischen Reizes. Dabei ist es nicht selbstverständlich, dass alle sensorischen Reize, die über das Gehör aufgenommen werden, zu einer Wahrnehmung führen. So werden manche Signale, die das Gehör aufnimmt, nicht wahrgenommen: z.B. der Pulsschlag des Herzens oder das Rauschen, das im Innenohr entsteht, weil dieses von Blut durchströmt wird. Es besteht kein zwingender Zusammenhang zwischen Eingangssignal (Schallereignis) und Ausgangssignal (der Wahrnehmung).

Ein Großteil des akustischen Wahrnehmungsprozesses findet im Kortex statt, der Hirnrinde. Dort liegen verschiedene Zentren, die gleichzeitig verschiedene Reize verarbeiten können. Zwar wurden in den letzten Jahren viele neue Erkenntnisse über die Funktionen der verschiedenen kortikalen Bereiche erlangt, dennoch ist insbesondere beim menschlichen Gehirn ein Großteil der Funktionen des Gehirns noch unerforscht (Goldstein 2008, Deike 2010). Seit 2013 gibt es durch das Human Brain Project das Bestreben, die Funktion des gesamte Gehirns zu entschlüsseln (Human Brain Project 2013). Zur Zeit gibt es aber noch keine verwertbaren Studien. In Zukunft könnte dieses Projekt zu einem weit besseren Verständnisses des Hörens führen, als es der aktuelle Stand der Forschung zulässt.

4.1 Auditorische Verarbeitung im Kortex

Im Kapitel 1 ab Seite 8 wurde beschrieben, wie der Schall, der durch die Ohren aufgenommen, in Nervensignale gewandelt und weiter verarbeitet wird, am Ende dieser Kette im auditorischen Kortex ankommt.

Bis vor kurzem wurde in der Forschung noch davon ausgegangen, dass die Hörreize ausschließlich zentral im auditiven Kortex verarbeitet werden. Neue Forschungen belegen nun, dass es mehrere dezentrale Verarbeitungszentren gibt. Einige dieser Zentren arbeiten crossmodal³⁴. Das trifft insbesondere auf visuelle Reizinformationen zu, die in die auditiven Zentren gelangen (Goldstein 2008).

³⁴Informationen anderer sensorischer Reize werden zur Verarbeitung des Reizes genutzt, im Gegensatz zur multimodalen Verarbeitung besteht aber kein unidirektionaler Informationsfluss. Es besteht kein Austausch zwischen den sensorischen Zentren. Mehrere sensorische Zentren projizieren Informationen in ein Zentrum.

Die Informationen des Hörnervs gelangen gleichzeitig an verschiedene Regionen des Gehirns. Verschiedene Aspekte der auditiven Informationen werden parallel verarbeitet und mit anderen sensorischen Reizen verknüpft. Erst durch diese parallele Bearbeitung kann in kürzester Zeit ein komplexer Höreindruck entstehen (Leifert-Fiebach 2008, Goldstein 2008).

4.1.1 Komplexe Reizverarbeitung

Es gibt mehrere auditive Verarbeitungsbereiche (siehe dazu Abbildung 4.1, Seite 61), die um den auditorischen Kortex herum verteilt liegen. Während der auditive Kortex auf einfache Stimuli wie Sinustöne reagiert, antworten die umliegenden Gebiete nur auf komplexe Reize wie z.B. Rauschen oder Akkorde und Melodien.

Im Gürtel werden Muster über einzelne Neuronen abgebildet und erkannt. Je weiter außen die Neuronen im Gürtel liegen, desto komplexer sind die erkannten Muster. Es besteht demnach eine hierarchische Verarbeitungsstruktur, ähnlich wie beim Sehsinn. In der unteren visuellen Hierarchie können Neuronen nur Linien erkennen, in den höheren auch Landmarken und Gesichter (Goldstein 2008).

Wie genau die Bereiche um den auditiven Kortex funktionieren, wie sie konkret einen Reiz abbilden und verarbeiten, ist noch nicht belegt. Untersuchungen zeigen jedoch, dass bereits im erweiterten Gürtel eine crossmodale Verarbeitung zusammen mit visuellen Reizen stattfindet. Bis jetzt konnte für mehrere Areale nur belegt werden, dass sie sowohl auf auditive wie auch auf visuelle Reize reagieren.

4.1.2 Was- und Wo-Ströme

Seit einigen Jahren ist bekannt, dass auditive wie visuelle Informationen, getrennt in Was- und Wo-Ströme, in den Frontallappen projizieren (siehe dazu Abbildung 4.2, Seite 61). Es besteht dabei die Ansicht, dass dort die unterschiedlichen Reize zur audiovisuellen Abbildung eines Objektes zusammengeführt werden und so eine audiovisuelle Karte mit der Raumlagen verschiedener Objekte erstellt wird (Goldstein 2008).

Leifert-Fiebach (2008) ergänzt, dass die Hörereignislokalisation und -identifikation nach aktueller Forschungslage unabhängig voneinander stattfindet. Es gibt Hinweise auf verschiedene Mechanismen, die für die Lokalisation, die Bewegungsdetektion und die räumliche Aufteilung von Hörereignissen in einzelne Geräuschquellen zuständig sind. Ob diese Informationen im Frontallappen oder erst in späteren Verarbeitungsstufen verbunden werden, ist noch nicht belegt.

4 Modell zur Wahrnehmung eines Hörereignisses

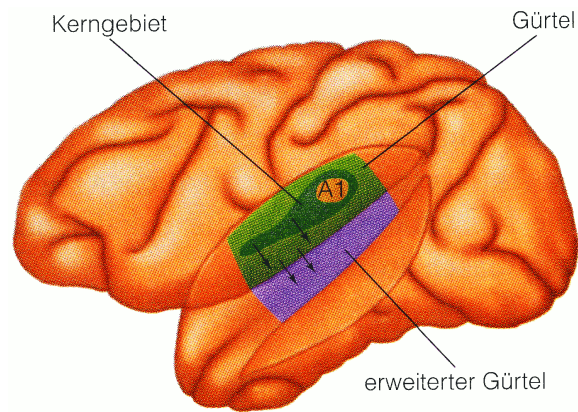


Abbildung 4.1: Die drei wichtigsten auditorischen Areale im Kortex. Der primär-auditorische Kortex (A1), der Gürtel und der erweiterte Gürtel. Die Signale breiten sich dabei vom Kerngebiet über den Gürtel zum äußeren Gürtel hin aus, wie die Pfeile zeigen (Goldstein 2008: S. 281).

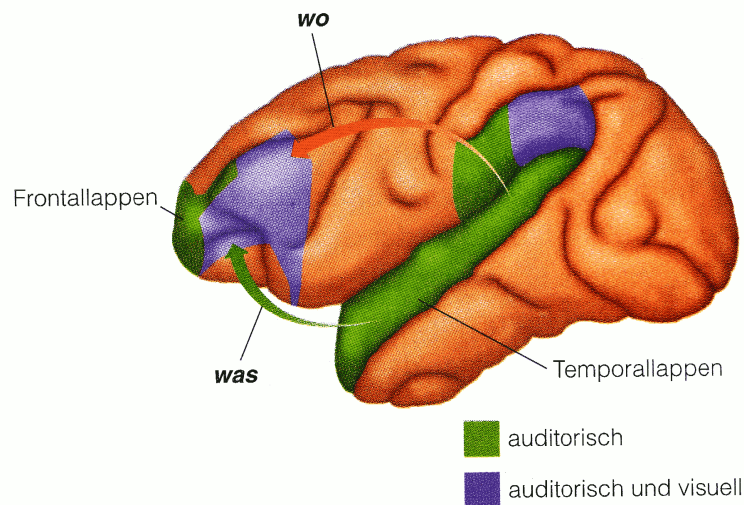


Abbildung 4.2: Untersuchungen am Kortex des Affen zeigen, dass es Areale gibt, die nur auf auditive Reize (grün) reagieren und solche die auf auditive und visuelle Reize (violett) reagieren. Ebenfalls bekannt ist, dass Reize in Form von Was- und Wo-Strömen vom auditiven Kortex zum Frontallappen gelangen (Goldstein 2008: S. 281).

4.1.3 Multimodale Wechselwirkung

Als belegt gilt, dass verschiedene sensorische Reize, die ihre Ursache in einer Reizquelle haben, durch die Aufmerksamkeit verbunden und als Einheit wahrgenommen werden (Leifert-Fiebach 2008). Es gibt also z.B. nicht die Wahrnehmung einer Amsel, die wir hören und einer (zweiten), die wir sehen, die sich (beide) an einer Position befinden. Sondern die Wahrnehmung einer Amsel, an einer Stelle sitzend, zu der verschiedene Attribute wie z.B. Klang der Stimme und Position gehören. Dabei kann es zu einer so starken Bindung der unterschiedlichen Sinnesinformationen kommen, dass nicht mehr zwischen den Sinnen unterschieden wird³⁵. Das führt z.B. im Film bei der Rezeption in 3D-Audio dazu, dass ein Auto kurzzeitig noch bildlich wahrgenommen wird, obwohl es die Leinwand optisch bereits verlassen hat und objektiv nur noch zu hören ist (Falke 2014). Einen ähnlichen Effekt beschreibt Goldstein (2008): Bei einem Experiment wurden in schneller Abfolge Einzelbilder mit zwei Bällen gezeigt. Die Bilder wurden schnell nacheinander wiedergegeben. Dabei wurde durch die Einzelbilder die Wahrnehmung erzeugt, dass sich die Bewegungsbahnen der Bälle zu kreuzen schienen. Die Versuchspersonen beschrieben anschließend, dass die Bälle sich aneinander vorbei bewegten. Sobald jedoch im Moment der Kreuzung ein Klick eingespielt wurde, nahmen die meisten Probanden wahr, dass die Bälle kollidierten und voneinander abprallten.

Die multimodale Wechselwirkung führt unter Umständen dazu, dass, wenn ein auditiver und ein visueller Reiz dem desselben Objekt der Wahrnehmung zugeordnet wird, sich zwischen gehörtem und gesehenem eine Art Phantomschallquelle bildet. Die auditive und die visuelle Quelle rückt also, von ihrer realen Positionen aus, in der Wahrnehmung näher zusammen. In anderen Fällen kann es auch dazu kommen, dass beides aus Richtung der visuellen Empfindung wahrgenommen wird. In der Regel ist der visuelle Reiz ausschlaggebend für die Lokalisation (Goldstein 2008).

4.1.4 Neuroanatomische Aufmerksamkeit

Die Aufmerksamkeit ist ein wichtiger mentaler Prozess, der dem Bewusstsein nur bestimmte Reize zugänglich macht. Diese Fokussierung ist nötig, da das Bewusstsein in seiner Kapazität, Reize parallel zu verarbeiten, begrenzt ist. Um möglichst viel gleichzeitig verarbeiten zu können, ist das Gehirn bestrebt, eine möglichst große Anzahl von Reizen zu einem Objekt zusammenzuführen. Im Mittelhirn konnten entsprechende neuronale Strukturen nachgewiesen werden, die visuelle und auditive Informationen integrieren, sofern eine gewisse räumliche und zeitliche Koinzidenz besteht. Verschiedene Reize, die einen gemeinsamen Ursprung haben, werden dort als Attribute eines Objektes zu einem intermodalen Strom zusammengefasst (Leifert-Fiebach 2008).

³⁵Natürlich kann sich jeder gesunde Mensch jederzeit bewusst machen, was gehört und was gesehen wird. Dies benötigt aber eine aktive Aufmerksamkeitslenkung. Siehe dazu auch Kapitel 4.2.7, Seite 67.

Noch ist nicht genau erforscht, welche Prozesse im Gehirn zur Aufmerksamkeitslenkung sowie zur Separation und Integration von Reizen führen (Leifert-Fiebach 2008, Goldstein 2008). Es bestehen aber verschiedene, in der Forschung allgemein anerkannte Theorien, die darstellen, wie das Gehirn einen eingehenden (akustischen) Reiz zu einer Wahrnehmung verarbeitet. Diese Theorien werden im Folgenden dargestellt.

4.2 Theorie des globalen Arbeitsraums

Die Theorie des globalen Arbeitsraums (Global-Workspace-Theorie) wurde von Baars (1997) aufgestellt. Er fasst dazu verschiedene Modelle und Untersuchungen unterschiedlicher wissenschaftlicher Disziplinen zum Thema Bewusstsein zusammen. Dabei stützt er sich vor allem auf Untersuchungen zum Arbeitsgedächtnis und auf Ergebnisse der Kognitionswissenschaft (Spieß 2002). Für Baars (1997) stellt das Bewusstsein ein zentrales neuronales Modul dar, das er globalen Arbeitsraum (global workspace) oder auch Arbeitsgedächtnis nennt und dessen Funktion er in seiner Theorie erörtert. Dabei nutzt er sehr plastisch das Modellbeispiel einer Theaterbühne, wie in Abbildung 4.3, Seite 64 zu sehen ist.

4.2.1 Das Arbeitsgedächtnis

Das Arbeitsgedächtnis, das Baars (1997) in seinem Theatermodell als Bühne bezeichnet, ist nach seinen Angaben kapazitiv auf etwa 7 Einheiten begrenzt. Darauf wiesen auch verschiedene Studien hin (Spieß 2002). Für diese Einheiten gilt, dass sie unabhängig sind Baars (1997). Das bedeutet, dass z.B. die visuelle und die auditive Projektion eines Autos, die aus einer Richtung kommen im Arbeitsgedächtnis eine Einheit bilden würden. Währenddessen würden unterschiedliche Wahrnehmungsrichtungen, sofern die auditive außerhalb des Sehfeldes liegt, zu zwei getrennten Einheiten führen, die je die Information „Auto“ enthielten.

Was die Anzahl der Einheiten betrifft, macht Baars (1997) noch weitere Einschränkungen: So seien die sieben Einheiten nicht gleichzeitig bewusst. Dies ließe sich z.B. an einer Zahlenfolge belegen:

16 7 9 74 2 16 3

Es ist nach (Baars 1997) Ansatz nicht möglich, alle Zahlen auf einmal zu erfassen oder nach dem Betrachten synchron zu erinnern. Es ist lediglich möglich, sie nacheinander in das Bewusstsein zu rufen. Das Arbeitsgedächtnis enthält also mehrere Einheiten, von denen jeweils nur eine bewusst ist.

4.2.2 Die Aufmerksamkeit – das Spotlight

Die Bewusstseinsbeschränkung erklärt Baars (1997) in seinem Theatermodell durch das Spotlight. Dieses lässt nur den Bereich der Bühne erkennen, der durch das Licht

4 Modell zur Wahrnehmung eines Hörereignisses

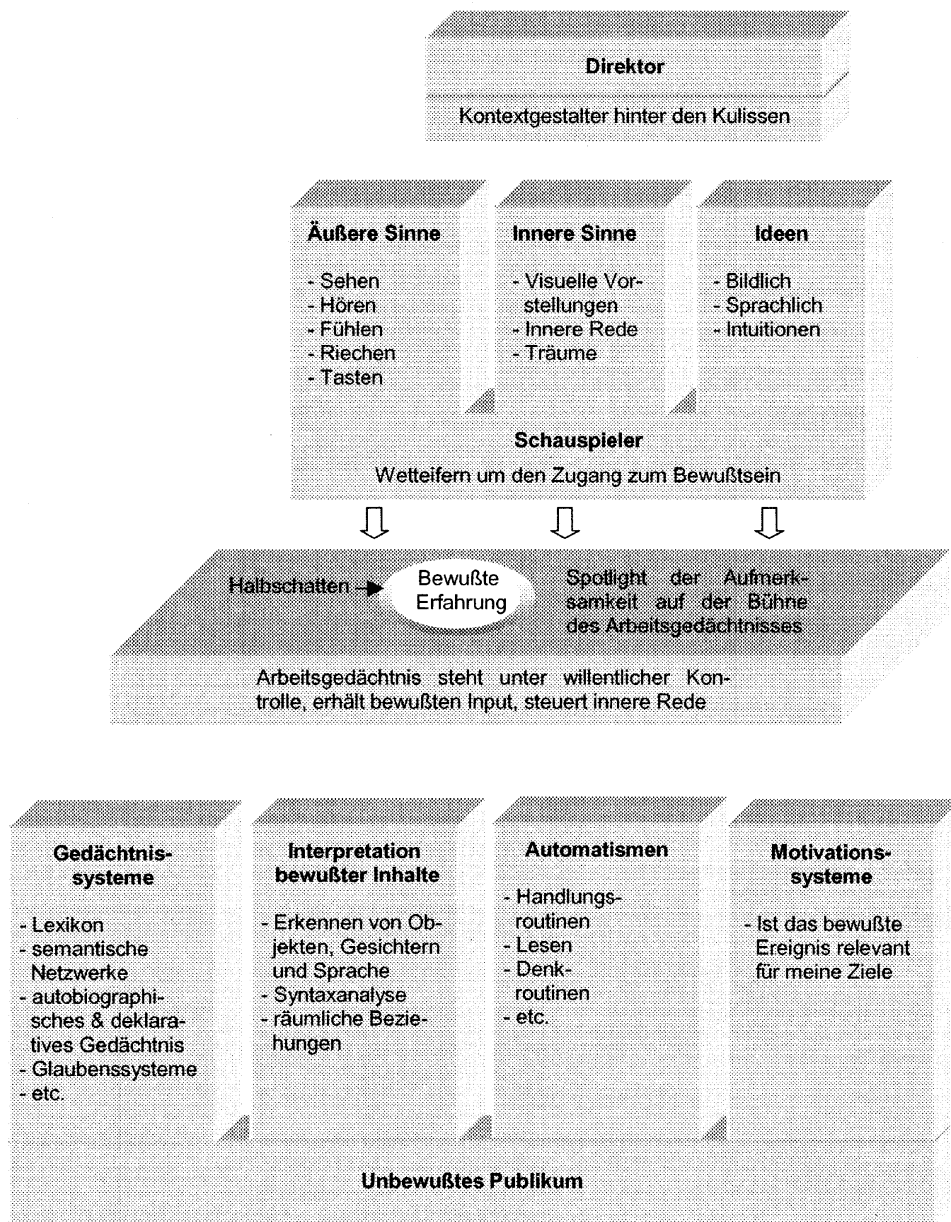


Abbildung 4.3: Baars Theatermodell zur Theorie des globalen Arbeitsraums (Spieß 2002: S. 71)

– die Aufmerksamkeit – erleuchtet ist. Nur durch diese Aufmerksamkeit kann das gesamte Publikum den jeweiligen Gedächtnisinhalt sehen. Wobei das Publikum in seiner Metapher für die verschiedenen Analyse- und Auswertungszentren im Gehirn steht, wie auf Abbildung 4.3, Seite 64 detaillierter zu erkennen ist. Dazu gehören unter anderem der visuelle und auditive Kortex.

Weiter redet Baars (1997) von einem Halbschatten. Dieser beschreibt Informationen, die nicht voll bewusst sind. Auf diese besteht dabei kein direkter Zugriff, dennoch sind diese Informationen wichtig für eine weitere Verarbeitung. Ein Beispiele hierfür sind Gefühle.

4.2.3 Konkurrierende Reize – die Schauspieler

Da das Arbeitsgedächtnis in seiner Verarbeitungskapazität begrenzt ist, konkurrieren innere und äußere Reize wie Schauspieler um einen Platz auf der Bühne. Äußere Reize sind z.B. die Sinnesreize, innere sind z.B. Vorstellungen oder Ideen (zur weiteren Aufzählung siehe Abbildung 4.3, Seite 64). Um möglichst viel Reize verarbeiten zu können, werden sie, soweit möglich, zu Einheiten gruppiert. So wird z.B. das Krähen eines Raben mit seinem visuellen Abbild zu einer Einheit verknüpft. Dabei unterscheidet die Aufmerksamkeit andauernd, welche Reize wichtig und welche weniger wichtig sind, um diese zugunsten neuer wichtiger Informationen verdrängen zu können (Baars 1997).

4.2.4 Verarbeitungszentren – die Zuschauer

Erst durch die Zuschauer – also die verschiedenen Verarbeitungszentren – ist es möglich, die komplexen Sinnesreize zu erfassen und weiter zu verarbeiten. Über die Aufmerksamkeit bekommen die einzelnen Zentren die Aufforderung, die Inhalte einer bestimmten Arbeitsgedächtnis-Einheit zu präzisieren und mit weiteren Informationen anzureichern (Baars 1997). Durch die Aufmerksamkeit werden alle kortikalen Verarbeitungszentren angewiesen, Daten, die dem Inhalt der fokussierten Einheit entsprechen, auszutauschen und zur Gewinnung neuer Informationen zu interagieren. So kann z.B. das auditive Zentrum die Lokalisation verbessern, indem es den motorischen Kortex eine Peilbewegung ausführen lässt und dadurch Information über die Bewegung zurückgemeldet bekommt, die dann eine Auflösung des „*cone of confusion*“ ermöglichen. So kommt es zur multimodalen Reizverarbeitung.

Neben dieser bewussten Verarbeitung finden weiterhin die unbewussten und automatisierten Vorgänge der Reizerfassung statt, die den größten Teil der Hirnaktivität ausmachen. Mit diesen Vorgängen können jedoch nur Reize erfasst werden, zu denen vorher durch bewusstes Handeln eine Erkennungsstrategie entwickelt wurde. Auch sind die Strategien größtenteils auf monomodale Analysen beschränkt (Goldstein 2008).

Unter diesen „*Zuschauern*“, wie Baars (1997) sie nennt, sind für die Wahrnehmung insbesondere die semantischen Netzwerke wichtig. Denn sie setzen unter anderem die

sensorische Darstellung eines realen Objekts (also alle Daten, die die menschlichen Sinne über dieses Objekt erfasst haben) in eine semantische, also zeichenhafte Darstellung um. Dadurch kann beispielsweise im Extremfall ein Objekt, das sensorisch mit zehntausenden von Neuronen dargestellt werden muss, semantisch mit nur einem Neuron dargestellt werden (Leifert-Fiebach 2008, Goldstein 2008).

So tritt z.B. an die Stelle einer großen Menge von auditiven und visuellen Daten, die ein Auto beschreiben, lediglich die Semantik „Auto“, die als Symbol aus nur vier Zeichen besteht (in dieser Analogie also vier Neuronen).

4.2.5 Das Selbst – der Direktor

Baars (1997) beschreibt das Selbst als mentale Darstellung, die alles umfasst, was die eigene Person ausmacht und von ihrer Umwelt abgrenzt. Es gibt den eigenen Erfahrungen und Handlungen Stabilität und Richtung. Dadurch entsteht ein Kontextrahmen, der in jeder Situation wissen lässt, was und wer das Individuum ist. Dieser Rahmen führt zu einer individuellen Steuerung des Bewusstseins. Es ist nach Baars (1997) Metapher der Direktor, der anordnet, auf welche Bereiche der Bühne der Spot der Aufmerksamkeit ausgerichtet werden soll. Das Selbst ist einer der ausschlaggebenden Faktoren dafür, welche Reize bevorzugt verarbeitet werden.

Durch die Aufmerksamkeit werden nur jene Reize bewusst gemacht, die den Zielen, Bedürfnissen und Trieben des Individuums entsprechen. Insbesondere werden solche Reize bevorzugt, die das menschliche Überleben sichern (Leifert-Fiebach 2008).

Es gibt jedoch Reflexe, die den Fokussierungswillen des Selbst außer Kraft setzen. Ein spezieller Reflex, der dies ermöglicht, ist in der Sounddesigntheorie als Orientierungsreflex bekannt. Ein plötzliches, lautes Geräusch oder das Abreißen eines Geräusches fokussiert sofort die Aufmerksamkeit auf diesen Reiz (Mikunda 2002). Der Drang des Selbst nach Aufmerksamkeitslenkung kann, zumindest kurzfristig, durch starke Reize zurückgedrängt werden.

4.2.6 Widersprüche zwischen Theorie und der Wahrnehmungspraxis

Bei der Überprüfung des Zahlenbeispiels von Baars (1997) (Abschnitt 4.2.1, Seite 63), mit dem er verdeutlicht, dass nur jeweils eine Einheit bewusst wird, besteht das Problem, nicht beschreiben zu können, wie mehrere Reize gleichzeitig wahrgenommen werden können. Denn Goldstein (2008) definiert die Wahrnehmung als eine bewusste sensorische Erfahrung.

Verschiedene Reize könnten jedoch durchaus gleichzeitig wahrgenommen werden. Zum Ändern gibt es Untersuchungen, die belegen, dass mehrere auditive Ströme gleichzeitig wahrgenommen werden können. So weist Leifert-Fiebach (2008) darauf hin, dass Schallereignisse³⁶ maximal zur Ausbildung von drei oder vier auditiven

³⁶Leifert-Fiebach grenzt diesen Begriff in ihrer Arbeit nicht deutlich ab, es bleibt unklar, ob die

Strömen führen, die als unabhängig voneinander wahrgenommen werden können. Damit wären allein durch den auditiven Sinn mehrere Einheiten einzeln und damit bewusst wahrnehmbar. Auch individuelle Erfahrungen zeigen, dass mehrere Dinge gleichzeitig wahrnehmbar sind.

Dieser Widerspruch ist nicht einfach zu lösen. Insbesondere, da der Begriff des Bewusstseins nicht eindeutig definiert ist und von verschiedenen Theorien unterschiedlich ausgelegt wird. Spieß (2002) sieht darin ein Problem unter anderem in der Kognitionsforschung, dass ohne eine genaue Definition des Basisbegriffes auch die sich darauf beziehenden Begriffe nicht wirklich definiert werden können. So ist auch der Begriff der Wahrnehmung nicht eindeutig definiert.

Goldstein (2008) stellt für die visuelle Reizverarbeitung fest, dass es ohne Aufmerksamkeit keine Wahrnehmung gibt. Er differenziert allerdings zwischen fokussierter Aufmerksamkeit, die der von Baars (1997) beschriebenen Aufmerksamkeitsfleck (dem Spotlight) entspricht und unfokussierter Aufmerksamkeit. Diese Unterscheidung entspricht dabei Untersuchungsergebnissen der Neuroanatomie, die auf eben jene unterschiedlichen Aufmerksamkeitsniveaus hinweisen. Dabei zeigen die Untersuchungen, dass diese Unterscheidung auch für auditive Reize gültig ist Leifert-Fiebach (2008).

Das anfängliche Problem die Wahrnehmung über die Theorie des globalen Arbeitsraums erklären zu können lässt sich auflösen, indem zwischen fokussierter und unfokussierter Aufmerksamkeit unterschieden wird. Die Wahrnehmung beinhaltet damit alle sensorischen Reize, die durch fokussierte oder unfokussierte Aufmerksamkeit abgebildet werden.

4.2.7 Wahrnehmung mehrerer Gedächtnisinhalte

Spieß (2002) Untersuchungen weisen darauf hin, dass die unfokussierte Aufmerksamkeit nur semantisch stattfindet. Zwar beschreibt er in seiner Arbeit Vorgänge der unbewussten Wahrnehmung, weist jedoch gleichzeitig auf die Unschärfe des Bewusstseinsbegriffes hin. Der Vergleich der von ihm gemachten auditiven Untersuchungen und der von ihm zitierten Arbeiten mit Untersuchungen zum Nachweis der unfokussierten Aufmerksamkeit im visuellen Bereich, wie Goldstein (2008) sie zitiert, zeigt eine signifikante Ähnlichkeit. In allen Arbeiten wird mit einem Reiz und einer Maskierung gearbeitet, die ein fokussiertes Wahrnehmen unmöglich machen. Daher liegt der Schluss nahe, dass unbewusstes Wahrnehmen nach Spieß (2002) und unfokussierte Aufmerksamkeit nach Goldstein (2008) identisch sind. Ähnliches formuliert auch Goldstein (2008). Somit kann nach Spieß (2002) geschlossen werden, dass die unfokussierte Aufmerksamkeit nur semantisch stattfindet. Das bedeutet, dass neuronal nicht der Reiz an sich abgebildet wird, sondern lediglich eine kompakte Musterbeschreibung in das Arbeitsgedächtnis projiziert wird.

Goldstein (2008) beschreibt, dass durch die unfokussierte Wahrnehmung mehre

Schallereignisse in den Ohren oder mehrere Schallquellen gemeint sind. Für die Argumentation in diesem Abschnitt ist dies jedoch unerheblich.

Reize direkt in die Wahrnehmung gelangen können. Allerdings sind diese Abbildungen (oder Projektionen) des Reizes sehr ungenau. So kann z.B. bei Gesichtern zwischen solchen von Frauen und solchen von Männern unterschieden werden, während eine genaue Zuordnung zu einer bestimmten Person nicht möglich ist. Diese Forschungen zeigen, dass nur bekannte und einfache Muster (also semantische Abbildungen) gleichzeitig in die Wahrnehmung gelangen können. Komplexe Reize werden auf einfach bekannte Reizmuster reduziert. Die Gesichter können so nur in der Form einfacher semantischer Symbole, also Frau oder Mann, in die Wahrnehmung dringen. Wie komplex diese Mustererkennung maximal sein kann, ist dabei nicht bekannt. Es liegt jedoch die Vermutung nahe, dass es keine feste Grenze gibt. Vielmehr ist diese Erkennung abhängig von bewussten Lernprozessen. Ein Gesicht, das jeden Tag immer wieder genau angesehen wird, könnte so durchaus auch unfokussiert wahrgenommen und der entsprechenden Person zugeordnet werden. Selbiges gilt für komplexe akustische Reize, wie z.B. den Klang einer Stimme. Auf der anderen Seite zeigt Goldstein (2008), dass es möglich ist, durch fokussierte Aufmerksamkeit auf bestimmte Gedächtnisinhalte die Wahrnehmung durch unfokussierten Aufmerksamkeit zu begrenzen. Dabei zeigt er, dass Personen, die sich stark auf bestimmte Reiz konzentrieren, einen weiteren Reiz, den sie unfokussiert hätten wahrnehmen müssen, nicht wahrnehmen. Ist das Arbeitsgedächtnis bereits mit bestimmten Reizen ausgelastet, können keine weiteren Reize in Form von Strömen aufgenommen werden, sie gelangen nicht in die Wahrnehmung. Diese Informations-Überflutung konnte auch neuronal nachgewiesen werden (Leifert-Fiebach 2008). Ebenfalls weist Goldstein (2008) drauf hin, dass eine fein strukturierte Wahrnehmung immer eine fokussierte Aufmerksamkeit benötigt ³⁷.

4.2.8 Modifikation der Theorie des globalen Arbeitsraums

Die beschriebenen Erkenntnisse zur unfokussierten Aufmerksamkeit lassen folgenden Schluss zu: Die Einheiten des Arbeitsgedächtnisses werden mit den Strömen verschiedener Reize gefüllt. Dabei versucht das Arbeitsgedächtnis, möglichst viele Ströme in eine Einheit zu integrieren. Die Ströme erhalten dabei nur unscharfe Beschreibungen (Semantiken) eines Reizes. Bis zu sieben Einheiten können so durch die unfokussierte Aufmerksamkeit wahrgenommen werden. Durch schnelles Steuern der fokussierten Aufmerksamkeit wird jeder Inhalt im Arbeitsgedächtnis nacheinander abgetastet. Durch die Fokussierung werden die verschiedenen kortikalen Zentren veranlasst, zu interagieren und durch genauere Erfassung Ströme zu generieren, die die jeweilige Gedächtniseinheit mit genaueren Informationen füllt (siehe dazu für den auditiven Bereich Kapitel 4.3, Seite 70). Dieser fein strukturierte Inhalt der Gedächtniseinheit

³⁷Ein einfaches Beispiel dazu ist der Text auf dieser Seite. Jeder Absatz kann mit einem Blick auf die Seite als solcher als aus Buchstaben bestehend erkannt werden. Für eine Erfassung des Textes in seiner Bedeutung ist jedoch eine Fokussierung nötig. Erst dadurch kann – wie es gerade geschieht – der Text erfasst und der Inhalt verstanden werden. Während die gesamte Seite durch unfokussierte Aufmerksamkeit semantisch nur als Ansammlung von Buchstaben, die in Worten und Absätzen gegliedert sind, wahrgenommen wird.

kann nun durch die unfokussierte Aufmerksamkeit wahrgenommen werden. Durch anhaltend wiederkehrende Fokussierung aller Gedächtniseinheiten ist so eine präzise Wahrnehmung vieler Reize möglich.

Wird die fokussierte Aufmerksamkeit stark auf eine Einheit gelenkt und von dieser exklusiv beansprucht, z.B. durch das Lesen eines Textes, kann diese Abtastung nicht mehr weiter erfolgen. Neue Reize können so nicht mehr genauer erfasst werden. Die unfokussierten Gedächtnisinhalte werden nur noch unscharf wahrgenommen. Wobei Alarm- und Orientierungsreize jederzeit den Fokus verschieben können³⁸ (Mikunda 2002).

Da diese Abtastung des Wahrnehmungsraumes und das Erfassen von Details zeitkritisch ist, wäre zu schließen, dass schnelle Änderungen vieler unterschiedlicher sensorischer Reize nicht genau verfolgt werden können, beziehungsweise zu Fehlwahrnehmungen führen³⁹.

4.2.9 Auswirkungen auf die Differenzierbarkeit

Da maximal sieben Einheiten wahrgenommen werden können, besteht darin eine Grenze für die Differenzierung von Schallquellen in Hörereignisse. Strahlen mehr als sieben Quellen Schall ab, werden sie zu maximal sieben Hörereignissen zusammengefasst.

4.2.10 Zusammenfassung der modifizierten Theorie

Mit der Unterscheidung in unfokussierte und fokussierte Aufmerksamkeit ist die Wahrnehmung über die Theorie des globalen Arbeitsraums abbildbar, ohne von Baars (1997) Paradigma abzuweichen, dass nur ein Arbeitsgedächtnisinhalt bewusst (fokussiert) sein kann. Im Modell werden alle Reize in Form von Mustern (oder auch Semantiken) als Ströme in das Arbeitsgedächtnis projiziert. Diese Ströme werden vom Arbeitsgedächtnis weitmöglichst als Einheiten zusammen gefasst.

Können durch die Limitierung des Arbeitsgedächtnisses nicht alle Reize in Form von Strömen aufgenommen werden, führt dies zu einer Bevorzugung wichtiger Reize. Alle anderen Reize gelangen nicht mehr in das Arbeitsgedächtnis und somit auch nicht in die Wahrnehmung.

³⁸Diese Alarmreize verkomplizieren die Untersuchung der auditiven Wahrnehmung. Beim Sehen ist es möglich, Bildbereiche, auf die nicht fokussiert wird, zu manipulieren, ohne dass eine Wahrnehmung einsetzt. Im auditiven Bereich wird normalerweise die gesamte Szenerie erfasst. Dadurch führt eine Manipulation meist zu einem Orientierungsreiz und zur Wahrnehmung der Manipulation.

³⁹Interessant ist in dieser Hinsicht auch das Verlangen der Wahrnehmung nach Konsistenz. So dauert es z.B. beim Betreten eines unbekanntes Raumes eine gewisse Zeit, bis alle seine Details wahrgenommen werden. Die Wahrnehmung produziert jedoch das Gefühl, dass der Raum im Moment des Betretens bereits im Detail wahrgenommen wurde. Die detaillierten Informationen werden im Nachhinein untergeschoben, als wären sie bereits vorhanden gewesen.

Alle Reize projizieren mehre Ströme in das Arbeitsgedächtnis, die nur grobe Informationen enthalten. Durch die Fokussierung der Aufmerksamkeit auf eine Einheit werden alle kortikalen Zentren angewiesen, zu interagieren und die enthaltenen Ströme zu präzisieren (siehe dazu für den auditiven Bereich Kapitel 4.3, Seite 70). Dadurch wird die Einheit mit detaillierten Informationen gefüllt.

Durch die unfokussierte Aufmerksamkeit sind alle Einheiten gleichzeitig wahrnehmbar. Wurden diese fokussiert, können sie im Detail wahrgenommen werden.

4.3 Auditorische Szenenanalyse

Mit der auditorischen Szenenanalyse (auditory scene analysis) beschreibt Bregman (1994) die Regeln, auf der die auditive Wahrnehmung beruht. Dabei stellte er fest, unter welchen Gegebenheiten einzelne auditive Reize zu einer sinnhaften Einheit gruppiert werden⁴⁰. Diese Theorie konnte dabei nicht nur beim Menschen erfolgreich überprüft werden, sondern auch bei diversen Tierarten. An diesen wird zur Zeit die neuronale Grundlage, auf der die Strom-Separation beruht, untersucht. (Deike 2010)

4.3.1 Auditorische Ströme

Die Analyse der auditiven Szenerie beruht vor allem auf Ideen der Gestalttheorie, die als grundlegende Gesetze zur Bildung einer Gestalt die Übersummativität, die Nähe und die Ähnlichkeit, nennt (Deike 2010). Bregman (1994) bezieht die Nähe und die Ähnlichkeit auf allgemeine Korrelationen im Frequenz- (z.B. Tonhöhe, Klangfarbe) sowie im Zeitbereich (z.B. Rhythmus, Gleichzeitigkeit). Auf Basis ähnlicher Merkmale werden Gruppen gebildet und voneinander separiert. Diesen Vorgang der Abgrenzung nennt er auditorische Strom-Separation (auditory stream segregation).

Bregman (1994) beschreibt Heuristiken, wie Gruppen gebildet werden und dadurch mehrerer Schallquellen der realen Welt im Bewusstsein voneinander abgegrenzt abgebildet werden. Diese Abbildungen werden auditorische Ströme (auditory streams) genannt. Sie projizieren in das Bewusstsein und bilden so den die Person umgebenden akustischen Raum ab (Bregman 1994, Deike 2010). Bregman (1994) definiert einen auditorischen Strom als die wahrgenommene Gruppierung aus Teilen des neuronalen Spektrogramms als Ganzes.

Ein (auditorischer) Strom kann verschiedene Muster zu einem übergeordneten Muster zusammenfassen, also z.B. mehrere Töne zu einem Akkord. Genauso können übergeordneten Muster zergliedert und als einzelne Ströme wahrgenommen werden. Ein

⁴⁰Sehr interessant ist auch die Seite des McGill Auditory Research Laboratory das Albert S. (Al) Bregman bis zur Schließung leitete. Hier finden sich große Teile des Standardwerkes von Bregman (1994). Zudem sind Arbeiten und Forschungen zur auditorischen Szenenanalyse zu finden. Die Seite war ein Zufallsfund, nachdem diese Arbeit bereits zum großen Teil verfasst war. Daher konnte diese Quelle nicht weiter berücksichtigt werden. Da es schade wäre, das Vorhandensein dieser gehaltvollen Quelle zu verschweigen, sei auf diesem Wege auf sie hingewiesen: <http://webpages.mcgill.ca/staff/Group2/abregm1/web/index.htm>

Akkord wird z.B. wieder in einzelne Töne zergliedert. Gesteuert wird dies durch die Aufmerksamkeit. Ähnliches gilt auch für verschiedene Quellen. So können mehrerer Sprecher zu einem Strom als Hintergrundgeräusche zusammengefasst werden, während das Muster der Sprache eines einzelnen Sprechers als Strom separiert wird (Bregman 1994).

4.3.2 Stromseparation

Die auf Muster-Erkennung basierende auditorische Strom-Separation muss erlernt werden. Dazu benötigt es Basisregeln, um damit komplexere Regeln zur Musterbildung zu erlernen. Als eine Basisregel führt Bregman (1994) dazu die Diskontinuität auf. So lernt z.B. ein Kind, was ein Wort ist, indem seine Eltern vor und nach einem Wort eine Pause (als Begrenzung des Wortes) machen. Durch diese Diskontinuität widmet das Kind seine Aufmerksamkeit dem Muster zwischen den Pausen und nimmt es als Einheit wahr. Mit der Zeit entsteht dabei ein genaues Muster dieses Wortes, das mit einer Bedeutung verknüpft wird. Später sind die Pausen nicht mehr nötig. Erwachsene erkennen Worte, ohne dass diese durch Pausen voneinander abgesetzt werden müssen.

Es gibt zwei Arten, wie die Repräsentation eines Stromes in der Wahrnehmung entstehen kann. Zum Einen ungelernt auf Basis primitiver Mustererkennung und zum Andern durch komplexe Mustererkennung. Die primitiven Prozesse (primitiven Separation) der auditiven Gruppierung basieren auf dem Erkennen einfacher Muster. Die komplexen Prozesse (komplexe Separation) entstehen hingegen erst durch Lernvorgänge und benötigen Vorwissen über einen wahrzunehmenden Klang (Bregman 1994).

4.3.3 Primitive Separation

Durch den Colliculus inferior (siehe 1.3.2, Seite 13) besteht bereits eine Karte, die Energiegehalt (Pegel) und binaurale Lageninformationen über die Frequenz abbildet. Durch fortwährendes Abrufen und Zwischenspeichern entsteht eine Karte, die zusätzlich über die Zeit abbildet. Bei der primitiven Separation werden daraus Ströme separiert, die bestimmte Bereiche im Spektrum über die Zeit beschreiben (Bregman 1994).

Das energetische Spektrum wird durch eine Vielzahl von Analysen in mehrere lokale Muster gewandelt, die sich über einen bestimmten Zeit- und Frequenzbereich erstrecken. Jedes dieser Muster kann dabei einen auditorischen Strom erzeugen. Wobei noch nicht belegt ist, inwieweit die Anzahl der Ströme begrenzt ist.

Leifert-Fiebach (2008) zitiert mehrere Studien, die die Anzahl gleichzeitiger auditiver Ströme auf maximal vier begrenzen. Es ist jedoch nicht klar, ob sich diese Studien auf die auditiven Ströme der komplexen oder der primitiven Separation beziehen. Es wäre denkbar, dass die primitiven Ströme in hoher Anzahl direkt über die unfokussierte Aufmerksamkeit wahrgenommen werden, während die komplexen Ströme nur in stark

begrenzter Anzahl wahrgenommen werden. Das wird vor allem durch die Aussagen von Goldstein (2008) über die unfokussierte Aufmerksamkeit gestützt. Zum Ändern löst die Unterscheidung in primitive und komplexe Ströme auch ein triviales Problem: Die komplexe Separation braucht mitunter sehr viel Zeit für die Mustererkennung. Dadurch käme es während dieser Zeit zu keiner Wahrnehmung, es gäbe Lücken. Das ist jedoch nicht der Fall⁴¹.

Je näher beieinander auditive Ströme im Frequenzbereich liegen, desto schwieriger ist es für den Hörenden, sie zu separieren. Bregman (1994) geht dabei von einem logarithmischen Zusammenhang für den Frequenzbereich aus. Mit zunehmender Nähe steigt die Chance exponentiell, dass zwei auditive Ströme nicht mehr separiert werden können. Falls die Aufmerksamkeit auf einen der Ströme gelenkt wird, kann dieser jedoch weiterhin problemlos separiert werden. Auch bei der Strom-Separation ist die fokussierte Aufmerksamkeit wichtig. Nur dadurch ist es möglich, Ströme zu separieren, die sich überlagern oder nahe beieinander liegen.

4.3.4 Heuristiken der primitiven Stromseparation

Strahlen mehrere Schallquellen ein Signal ab, ist es nicht ohne weiteres möglich, deren Signale wieder voneinander zu trennen. Denn im Gehör entstehen nur zwei Schallergebnisse, die jeweils eine Summe der Energie der abstrahlenden Quellen darstellen. Es ist nicht möglich, diese ohne Informationen über die einzelnen Quellensignale wieder zu separieren. Die nötigen Informationen werden durch Heuristiken erlangt, wie sie die Gestalttheorie beschreibt. Damit ist eine auditive Gruppierung möglich, die zu einer Separation führen kann (Bregman 1994).

Herkunft

Ein wichtiger Faktor für die Gruppierung ist die Herkunftsrichtung. Über die binauralen Richtungsinformationen, die im Colliculus inferior frequenzbezogen abgebildet sind, können Quellen unterschiedlicher Raumlagen voneinander separiert werden. Das ist jedoch nur solange möglich, wie die Quellen spektral unterschiedliche Signale abstrahlen und weit genug voneinander entfernt liegen. Andernfalls sind die Informationen nicht eindeutig. Spektren, die Informationen über verschiedenen Quellen erhalten, zeigen eine falsche bzw. eine nur unscharfe Richtung des resultierenden Hörereignisses an (Bregman 1994).

Temporäre Überschneidungen von zwei Mustern im Frequenzbereich können durch Korrelation im Zeitbereich aufgelöst werden. Allerdings kann die Energie den beiden

⁴¹Das sei am Klangbeispiel näher erklärt: erst ist ein Zischen und Stampfen zu hören, dann ein Pfeifen. Aus diesen Informationen entsteht ein komplexer auditiver Strom, der eine Dampfklok abbildet. Das Zischen, Stampfen und Pfeifen wird aber bereits vorher in Form von primitiven Strömen wahrgenommen. Wäre das nicht so, wäre die Wahrnehmung lückenhaft, bis es zum Erkennen der Dampfklok kommt.

Mustern nicht wieder im ursprünglichen Verhältnis zugeteilt werden, da ein Muster geschlossen sein und sich abgrenzen muss. Zur Lösung der Überdeckung gibt es zwei Heuristiken: Sind beide Muster unterschiedlich laut, greift das Prinzip der Verdeckung: Die Energie wird dem Muster zugeordnet, das näher (lauter) ist. Bei ähnlicher Lautstärke kommt es zur Doppelwahrnehmung: Die Energie wird beiden Mustern jeweils vollständig zugeteilt (Bregman 1994).

Da die binauralen Richtungsinformationen z.B. durch Reflexionen und andere akustische Effekte oft verfälscht werden, wird bei der Mustererkennung im spektralen Bereich weniger auf diese Informationen zurückgegriffen. Dies findet eher im Zeitbereich statt (Bregman 1994). Es braucht weitere Gruppierungsvorschriften, die es ermöglichen, nahe gelegene und ähnliche Quellen von einander zu unterscheiden.

Ähnlichkeit in Klangfarbe und Tonhöhe

Eine wichtige Vorschrift, die Bregman (1994) nennt, ist die Ähnlichkeit in Klangfarbe und Tonhöhe. Schallquellen strahlen in der Natur meist in einer ähnlichen Tonhöhe ab und ändern diese nur langsam. Auf diesem Prinzip beruhend werden Tonfolgen ähnlicher Tonhöhe als Einheit, als Melodie wahrgenommen. Diese Einheit besteht, solange die Pausen zwischen den Tönen nicht zu lang sind. Bei schneller Folgen von wechselnd hohen und tiefen Tönen entsteht so die Wahrnehmung zweier Melodien, obwohl nur eine Instrument monophon gespiegelt wird (siehe Abbildung 4.4, Seite 73). Damit diese Gruppierung in zwei Melodien zustande kommt, darf der zeitliche Abstand nicht zu groß sein, und der spektrale Abstand der hohen und tiefen Töne muss groß genug sein (siehe Abbildung 4.5, Seite 74).



Abbildung 4.4: Bei schnellem Spielen werden jeweils die oberen und unteren Noten in der Wahrnehmung als Melodie gruppiert (Goldstein 2008: S. 303).

Zeitliche Nähe

Ein weiteres Prinzip, das auch bei der Bildung von Melodien mitwirkt, ist die zeitliche Nähe. Denn Schall, der in rascher Folge oder über verschiedene Spektren gleichzeitig abgestrahlt wird, stammt im natürlichen Umfeld meist von derselben Quelle. Dementsprechend wird gruppiert, was zeitlich nah ist oder gleichzeitig im Spektrum liegt, sofern die spektrale Distanz nicht zu groß ist.

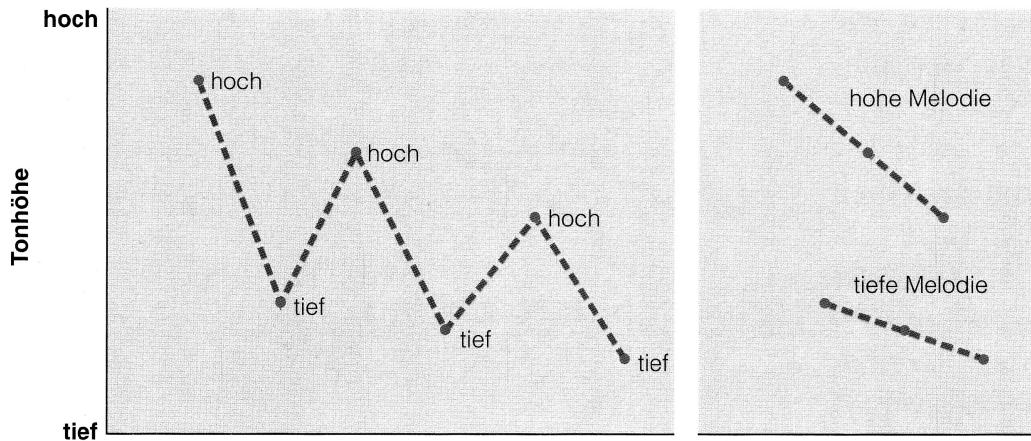


Abbildung 4.5: Links: Wechseln sich hohe und tiefe Töne langsam ab, kommt es nicht zur auditiven Gruppierung. Jeder Ton ist einzeln wahrnehmbar; Rechts: Bei schnellem Wechsel der Töne kommt es zur Gruppierung, es wird eine Melodie wahrgenommen (Goldstein 2008: S. 304).

Guter Verlauf

Die letzte Heuristik ist der gute Verlauf. Das bedeutet, dass Schallereignisse, sofern sie von nur einer Quelle stammen, sich nicht abrupt sondern nur allmählich ändern und dass diese Änderung einen festen Verlauf hat (z.B. das Ansteigen der Tonhöhe).

4.3.5 Komplexe Separation

Die Grundlage der komplexe Mustererkennung ist, wie Bregman (1994) beschreibt, das Erlernen eines Musters, das im Gedächtnis abgelegt wird. Unbekannte Muster können so nicht erkannt und deshalb nur als primitive Ströme abgebildet werden. Da ein Zurückgreifen auf Gedächtnisinhalte und ein Überprüfen zeitlich langer Muster auf Ähnlichkeit viel Zeit in Anspruch nimmt, unterliegen die komplexen Erkennungsprozess einer großen Latenz. Einige Forschungen weisen drauf hin, dass durch die große benötigte Leistung, insbesondere durch das Abrufen großer Gedächtnisinhalte, die Strom-Separation durch komplexe Separation auf etwa vier Ströme begrenzt ist (Leifert-Fiebach 2008, Goldstein 2008). Durch primitive Separation können weitaus mehr Ströme gebildet werden. Überlagern sich aber mehrere Quellen stark in Frequenz und Zeitbereich, ist eine komplexe Separation nötig (Bregman 1994).

4.3.6 Abbildung von Hörereignissen über auditive Ströme

(Blauert 1974) definiert den Begriff des Hörereignisses als die Wahrnehmung, die entsteht, wenn eine oder mehrere Quellen im Raum Schall abstrahlen und dieser über die Ohren aufgenommen wird. Das Hörereignis kann damit einem auditiven Strom

gleichgesetzt werden. Allerdings nur, wenn der auditive Strom auch über das Arbeitsgedächtnis in die Wahrnehmung projiziert. Da dieses, wie im Kapitel 4.2.9, Seite 69 beschrieben, begrenzt ist, kann unter Umständen kein Hörereignis entstehen, obwohl die auditiven Information vom Kortex verarbeitet wurden. Dieser Effekt ist bekannt als Informationsüberflutung. Wenn eine zu große Menge visueller, auditiver, emotionaler und anderer sensorischer Reize vorliegen, die nicht alle vom Arbeitsgedächtnis erfasst werden können (siehe Kapitel 4.2.3, Seite 65), muss die Aufmerksamkeit unwichtige Ströme verwerfen (Leifert-Fiebach 2008). Unwichtige auditive Reize gelangen demnach nicht in die Wahrnehmung und führen nicht zu einem Hörereignis.

Ein weiterer Aspekt besteht darin, dass bei nur einer abstrahlenden Quelle (z.B. einem Instrument) mehrere Ströme in Form von Melodien entstehen können. Es wird nur eine Quelle wahrgenommen die zu mehreren Hörereignissen führt. Das liegt daran, dass Ströme sowohl sequenziell als auch spektral organisiert sein können. Sequenziell (horizontal) bezeichnet den Zeitbereich oder, musikalisch betrachtet, die Melodie. Spektral (vertikal) bezeichnet die gleichzeitige Verteilung von Tönen zur gleichen Zeit, musikalisch der Akkord und die Harmonie.

Für die Differenzierung von Hörereignissen ist dies insoweit von Belang, da die Anzahl der wahrnehmbaren Ströme begrenzt ist. Strahlt eine Quelle eine mehrstimmige Melodie ab, entstehen mehrere auditive Ströme, die z.B. im Arbeitsgedächtnis durch Overflow Ströme verdrängen können, die andere Quellen darstellen. Diese Quellen werden nicht mehr wahrgenommen und lokalisiert.

4.3.7 Auswirkungen auf die Lokalisierbarkeit und Differenzierbarkeit

Ob eine Quelle einzeln als Hörereignis oder summiert mit anderen Quellen zu einer Wahrnehmung führt, hängt davon ab ob die Quellensignale jeweils in einen auditorischen Strom separiert werden können. Dabei gilt, dass mehrere, auch nahe Quellen, primitiv separiert werden können, wenn sie nur eine geringe zeitliche oder spektrale Überdeckung besitzen. So können durch die Einschränkungen des Arbeitsgedächtnisses bis zu sieben Quellen separiert werden.

Überdecken sich mehrere Quellen im Spektrum stark, können nur bis zu vier voneinander durch komplexe Separation in Ströme getrennt werden. Das gilt jedoch nur, sofern ihre Muster abgrenzbar sind, sich die Muster also nicht zu sehr ähneln⁴². Bei spektral stark überlappenden Quellensignalen ist es demnach im günstigsten Fall nicht möglich, mehr als vier Quellen zu separieren. Alle anderen Quellen werden lediglich über die primitive Separation in Ströme getrennt, bzw., da keine Muster-Grenzen erkannt werden, zusammengefasst und so zu einem diffus lokalisierten Hörereignis verschmolzen.

Aus Kapitel 2.2, Seite 15 ist bekannt, dass eine spektral breit abstrahlende Quelle

⁴²Dementsprechend können z.B. zwei baugleiche Instrumente, die unisono nebeneinander gespielt werden, schlecht voneinander separiert und in ihrer Raumlage differenziert lokalisiert werden.

schärfer lokalisiert werden kann als eine mit schmalbandiger Abstrahlung. Da insbesondere die primitive Separation in mehrere Ströme eine exklusive Zuordnung von Spektren bedeutet, entsteht so faktisch eine Bandbegrenzung. Dadurch nimmt die Lokalisationsschärfe mit steigender Anzahl von separierten Strömen ab. Speziell für die Höhe, deren Erkennung auf Kammfilter-Effekten und Veränderungen der HRTF bei Frequenzen um und über 8 kHz beruht (siehe Kapitel 2.2.3, Seite 20) – was ebenfalls eine Bandbegrenzung für die Lokalisation bedeutet – führen so mehrere Quellen, die als Ströme separiert werden, zu größerer Lokalisationsunschärfe in der Medianebene. Durch die komplexe Mustererkennung kann die Lokalisationsschärfe verbessert werden, da eine Analyse über einen längeren Zeitraum mit einer komplexen Mustererkennung möglich ist. Allerdings können dann nur maximal vier Ströme voneinander abgegrenzt werden.

Ein Strom besteht aus mehreren zeitlich und spektral abgegrenzten Bändern, die alle eine Richtungsinformation enthalten. Um eine Richtung bestimmen zu können, muss aus diesen Bändern der Mittelwert gebildet werden. Weisen alle Bänder in dieselbe Richtung, kann der Strom sehr genau lokalisiert werden und die Lokalisationsunschärfe ist sehr gering. Wenn die Bänder jedoch in unterschiedliche Richtungen weisen – da z.B. eine schlechte Raumakustik zu starken Reflexionen führt – können einzelne Bänder in verschiedene Richtungen weisen. Die Abweichung und damit auch die Lokalisation ist unscharf. Insbesondere, wenn wenige Bänder in einem Strom enthalten sind, führt dies zu einer schlechteren Lokalisationsschärfe, da die Fehler einzelner Bänder ein höheres Gewicht haben.

4.4 Wahrnehmungsmodell

Führt man die Theorie des globalen Arbeitsraums, die Theorien zur Analyse der auditiven Szenerie und Aussagen der Psychoakustik, insbesondere von Blauert (1974), zusammen, kann ein Modell erstellt werden, das abbildet, wie Schallereignisse zur Wahrnehmung von mehreren Hörereignissen führen. Daran können der Hörvorgang nachvollzogen und beschrieben sowie Einschränkungen für die Lokalisation und die Differenzierung definiert werden.

4.4.1 Vorkortikale Verarbeitung

Wie in Kapitel 1, ab Seite 8 beschrieben wurde, entsteht aus der Schallenergie, die an beiden Basilmembranen in Nervenreize gewandelt wird, im Colliculus inferior eine tonotopische Karte, in der die binauralen Richtungsinformationen jeweils getrennt in vielen schmalen Frequenzbändern abgebildet werden. Diese Informationen stehen dann dem auditiven Kortex und anderen kortikalen Verarbeitungszentren zur Verfügung.

Zu beachten sind dabei die Erkenntnisse aus Kapitel 2.5, Seite 26: Die binauralen Richtungsinformationen sind nicht eindeutig. Die Sektorlage einer Quelle kann nicht bestimmt, sondern ihre Lage nur ungenau auf dem „*cone of confusion*“ verortet wer-

den. Es handelt sich also lediglich um primitive Richtungsinformationen, die keine genaue Lokalisation zulassen. Dem entsprechend kann von primitiver Lokalisation gesprochen werden.

4.4.2 Kortikale Analyse

Im auditiven Kortex⁴³ wird die tonotope Karte, die die primitive Richtungen enthält, durch fortwährende Abtastung zusätzlich über die Zeit abgebildet (siehe Kapitel 4.3.3, Seite 71. Aus dieser Karte werden sowohl mit primitiver als auch mit komplexer Separation auditive Ströme gewonnen, die in das Arbeitsgedächtnis projizieren. Durch den Separationsprozess können die primitiven Richtungsinformationen, die in den verschiedenen tonotopen Bänder des Strom enthalten sind, ausgewertet werden. Jeder auditive Strom, der in das Arbeitsgedächtnis projiziert, enthält primitive Richtungsinformationen.

4.4.3 Aufmerksamkeit

Durch die unfokussierte Aufmerksamkeit gelangen die Inhalte des Arbeitsgedächtnisses in die Wahrnehmung. Bis zu sieben Hörereignisse können so primitiv lokalisiert werden.

Durch Reflexe oder auch das Selbst wird die fokussierte Aufmerksamkeit auf eine Arbeitsgedächtnis-Einheit gelenkt, in die die kongruenten Ströme⁴⁴ verschiedener sensorischer Kortexe projizieren. Dadurch werden sämtliche Kortexe angewiesen, Informationen auszutauschen und zu interagieren, um die enthaltenen Ströme der jeweiligen Einheit zu präzisieren. Für die auditive Lokalisation bedeutet das, dass z.B. auch visuelle Informationen zur genaueren Lokalisation abgerufen werden.

4.4.4 Rückgekoppelte Analyse in der kortikalen Verarbeitung

Durch die fokussierte Aufmerksamkeit wird ebenfalls ein Rückkopplungsprozess in Gang gesetzt. Der auditive Kortex gibt an die obere Olive Steuerbefehle, Spektren – die der zu präzisierende Strom enthält – zu verstärken und störende Spektren zu dämpfen. Über die efferenten Nervenfasern werden die äußeren Haarzellen angesteuert, um die entsprechenden Frequenzen zu dämpfen oder zu verstärken. Über den motorischen Kortex wird eine Peilbewegung ausgeführt und über den Corpus geniculatum mediale dem auditiven Kortex mitgeteilt. Durch diese Informationen kann der auditive Kortex durch komplexe Separation den „*cone of confusion*“ auflösen und eine Sektorlage bestimmen. Die genaue vertikale Lage kann durch eine HRTF-Analyse des

⁴³genau genommen werden die Schallreize nicht nur im auditiven Kortex verarbeitet. Siehe dazu Kapitel 4.1.1, Seite 60. Im Rahmen des Modells ist eine Unterscheidung jedoch nicht notwendig, da lediglich Heuristiken beschrieben werden. Welchen Bereichen diese zugesprochen werden, ist dabei nicht von Belang.

⁴⁴Ströme die das selbe Objekt beschreiben.

Stromes bestimmt werden. Es entsteht ein präziser auditiver Strom, der eine genaue dreidimensionale Lage beschreibt. So entsteht ein dreidimensional scharf lokalisiertes Hörereignis.

Da die Annahme besteht, dass nur bis zu vier komplexe Ströme ausgebildet werden können, ist eine präzise dreidimensionale Lokalisation nur für maximal vier Hörereignisse möglich. In diesem Sinne kann von einer komplexen Lokalisation gesprochen werden. Alle anderen Hörereignisse werden dagegen nur primitiv lokalisiert. Weil die fokussierte Aufmerksamkeit sehr schnell wechseln kann, entsteht durch den fortlaufenden Prozess jedoch das Erleben, alle Hörereignisse (gleichzeitig) scharf und dreidimensional lokalisieren zu können.

4.4.5 Erkenntnisse aus dem Modell

Aus dem Modell geht hervor, dass die Informationen der primitiven Separation in Form von (primitiven) auditorischen Strömen direkt in das Arbeitsgedächtnis projizieren und lediglich primitiv lokalisiert wahrgenommen werden. Die genaue dreidimensionale Lokalisation durch komplexe Separation in maximal vier Ströme ist nur durch die fokussierte Aufmerksamkeit möglich, die die verschiedenen Kortexe koordiniert und eine Peilbewegung ausführen lässt.

Wieviele Ströme wahrgenommen werden können, ist nicht abschließend definierbar. Da nicht bekannt ist, wie die Ströme in das Arbeitssgedächtnis gelangen und wie groß die Arbeitssgedächtnis-Kapazität genau ist. Leifert-Fiebach (2008) spricht von maximal vier auditiven Strömen, die gleichzeitig ausgebildet werden können. Es wäre möglich, dass es sich dabei – wie in dieser Arbeit angenommen – nur um eine Begrenzung der komplex separierter Ströme handelt, während primitiv separierte Ströme in größerer Zahl abgebildet werden können.

Wie viele Hörereignisse maximal differenziert werden können, kann durch die Theorie des globalen Arbeitsraums auf eine maximale Anzahl von sieben eingegrenzt werden. Teilweise wird in der Literatur aber auch eine geringere Anzahl von Einheiten angegeben.

Durch die aktive Steuerung beider Basilarnebrane können verschiedene Frequenzen gedämpft und andere verstärkt werden⁴⁵. Dadurch kann der interaurale Kohärenzgrad künstlich erhöht werden, um eine Schallquelle besser lokalisieren und differenzieren zu können. Ein Rückschluss auf die Lokalisierbarkeit und insbesondere die Differenzierbarkeit durch das Messen der interauralen Kohärenz ist dadurch nur bedingt möglich. Dementsprechend müssen die Untersuchungen und die Schlüsse aus Kapitel 3, Seite 41 zumindest relativiert werden, insbesondere in der Hinsicht, wie genau mehrere Quellen lokalisierbar sind und ab welcher Anzahl keine Differenzierung mehr möglich ist. Die Aussage, dass mit der Anzahl der Quellen die Lokalisations-Genauigkeit abnimmt, ist jedoch richtig.

⁴⁵Eine unterschiedliche Verstärkung und Bedämpfung der beiden Basilarnebrane wäre auch möglich.

Es ist also davon auszugehen, dass durch die kortikalen Prozesse im optimalen Fall genauer lokalisiert werden kann, als aus den Untersuchungen für mehrere Quellen in Kapitel 3, Seite 41, hervor geht. Es können auch mehr als acht Quellen in Hörereignisse differenziert werden, allerdings nicht gleichzeitig. Nach den beschriebenen Beschränkungen des Arbeitsgedächtnisses würden einige Quellen zu einem diffusen Hörereignis zusammengefasst, während bis zu vier Quellen gleichzeitig durch fokussierte Aufmerksamkeit und komplexe Separation scharf dreidimensional lokalisiert werden können. Durch eine Verschiebung der fokussierten Aufmerksamkeit ist es so im Optimalfall möglich, jede Quelle einzeln als Hörereignis lokalisieren und differenzieren zu können. Die Aussage, dass nahe gelegene Quellen, auch wenn sie in unterschiedlichen Spektren abstrahlen, nicht separiert werden können, ist – unter Berücksichtigung der kortikalen Prozesse – nicht richtig, da insbesondere durch die rückgekoppelten Prozesse eine Separation in Hörereignisse und eine selektive Lokalisation möglich ist.

5 Zusammenfassung der Ergebnisse

Die Fragestellung dieser Arbeit wurde aus verschiedenen wissenschaftlichen Perspektiven umfassend beleuchtet. Eine abschließende Antwort kann zum derzeitigen Stand der Forschung nicht gegeben werden. Dennoch können einige grundlegende Aussagen getroffen werden.

5.1 Begrenzungen durch kortikale Vorgänge

Aus den Ergebnissen des 4. Kapitels (ab Seite 59) geht hervor, dass unsere Wahrnehmung eine scharfe dreidimensionale Lokalisation nur beschränkt zulässt. Es können nur bis zu sieben Hörereignisse gleichzeitig voneinander differenziert werden, von denen maximal vier gleichzeitig scharf und dreidimensional – also komplex – lokalisierbar sind.

5.1.1 Wahrnehmung mehrerer Quellen

Dass bis zu vier Hörereignisse komplex lokalisierbar sind, gilt nur, solange die Aufmerksamkeit nicht auf einem Hörereignis fokussiert bleibt. In diesem Fall werden alle weiteren Hörereignisse nur primitiv und damit ungenau lokalisiert. Diese Bindung der Aufmerksamkeit kann unter anderem durch komplexe semantisch Quellensignale – wie z.B. Sprache – entstehen, die, um deren Inhalt aufnehmen zu können, eine stete Fokussierung der Aufmerksamkeit benötigen. Ebenfalls kann es durch Reizüberflutung des Arbeitsgedächtnisses dazu kommen, dass auditive Reize nicht in die Wahrnehmung gelangen können. Starke Reize, insbesondere visuelle können unter Umständen dazu führen, dass Schallquellen nicht mehr wahrgenommen werden.

Bei mehr als sieben Quellen verschmelzen einige zu einem Hörereignis, da nur sieben Hörereignisse wahrnehmbar sind. Durch Prozesse der Aufmerksamkeit ist es jedoch möglich, jede Quelle als Hörereignis zu separieren und dreidimensional scharf zu lokalisieren, sofern die Signalformen der Quellen, deren Lagen und die Raumakustik dies zulassen.

5.1.2 Die komplexe Lokalisationsgrenze

Für die Lokalisierbarkeit von Hörereignissen muss immer zwischen komplexer und primitiver Lokalisation unterschieden werden. Die Prozesse zur komplexen Lokalisation sind zeitkritisch. Bei schnellen Veränderungen der auditiven Szenerie kann daher

ein Hörereignis oft nur primitiv lokalisiert werden. Wieviel Zeit zur komplexen Lokalisation nötig ist, hängt von der Komplexität der Hörereignisse und von deren Anzahl ab. Eine verbindliche Angabe ist daher nicht möglich.

Eine untere Grenze lässt sich jedoch durch die Angabe von Blauert (1974) festlegen, dass bis zum Ausführen einer Peilbewegung etwa 300 ms vergehen. Quellen, die kürzer als 300 ms abstrahlen, sind in der auditiven Szenerie daher nur primitiv lokalisierbar. So kann für die komplexe Lokalisation aller Quellen im Raum folgende Faustformel gelten: Alle Quellen können sicher komplex lokalisiert werden, wenn sie gleichzeitig kontinuierlich ein Signal abstrahlen, das länger ist als deren Anzahl multipliziert mit 300 ms. Dieses Maß wird weiterführend komplexe Lokalisationsgrenze genannt. Darüber hinaus gelten die Einschränkungen der Kapitel 5.2 ab Seite 81 und 5.3 ab Seite 88.

5.2 Lokalisierbarkeit

Aus den Ausführungen in Kapitel 2 ab Seite 14 und den Auswertungen aus Kapitel 4 ab Seite 59 kann eine präzisere Aussage getroffen werden, wie genau lokalisiert werden kann. Dabei muss zwischen dem allgemeinen und dem speziellen Fall unterschieden werden. Es ist dabei zu beachten, dass die Akustik eines Raumes und eventuelle Wechselwirkungen zwischen Quellensignalen einen erheblichen Einfluss auf die Lokalisierbarkeit haben.

5.2.1 Wechselwirkungen

Die Wechselwirkungen entstehen, indem die Signale einer Schallquelle durch Separationsprozesse nicht mehr vollständig von Signalen anderer Quellen getrennt werden können. Durch die entstehenden Überdeckungen können die Richtungsinformationen nicht mehr vollständig unabhängig ausgewertet werden. So werden teilweise falsche Richtungsinformationen ausgewertet, es entsteht eine breitere Lokalisation. Im Extremfall ist keine Lokalisation möglich, wodurch die Signale zu einem einzigen Hörereignis zusammen gefasst werden.

5.2.2 Allgemeine Lokalisierbarkeit

Für ein Hörereignis, das seine Ursache in einer punktförmigen Schallquelle hat, gilt unabhängig von deren Lage, dass es mit einer Genauigkeit von etwa einem Grad horizontal und etwa vier Grad in Medianebene lokalisiert werden kann, wenn:

- Der Rezipient genug Zeit zum lokalisieren hat
- Der Rezipient den Kopf so bewegen kann, dass die Quelle frontal auf dem Schnittpunkt von Median- und Horizontalebene liegt
- Keine störenden Raumreflexionen entstehen
- Das Quellensignal sehr breitbandig ist
- Das Quellensignal signifikant Frequenzen um und über 8 kHz enthält
- Keine Signale von weiteren Quellen abgestrahlt werden
- Die Quelle eine Punktschallquelle ist

Der Allgemeinfall kann nur gelten wenn ein Hörereignis das Resultat eines Signals einer Quelle ist. Sofern eine virtuelle Punktschallquelle an den Ohren des Rezipienten dasselbe Schallfeld wie diese Quelle erzeugt, gilt für sie ebenfalls der Allgemeinfall.

5.2.3 Einfluss der Raumakustik

Die Raumakustik hat einen entscheidenden Einfluss auf die Lokalisierbarkeit einer Quelle. So zeigte die Auswertung der Messungen in Kapitel 3 ab Seite 41, dass der genutzte Raum nicht optimal war und von einer Verschlechterung der Lokalisierbarkeit ausgegangen werden muss. Die für diese Arbeit erstellten Messreihen zeigen, dass es sinnvoll ist, die Lautsprecher für eine 3-D-Wiedergabe in die Wände einzubauen, da sich so die störende Reflexionen der Lautsprecher an den Wänden verringern. Eine sehr trockene Akustik begünstigt die Lokalisierbarkeit bei der Wiedergabe von 3D-Audio-Inhalten⁴⁶.

Bewertung der Akustik

Aus den Untersuchungen im Kapitel 3 ab Seite 41 kann folgender Mess-Aufbau abgeleitet werden, um die akustische Eignung eines Raumes für die 3D-Wiedergabe zu beurteilen:

Ein omidirektionaler Lautsprecher wird mit maximal möglichem Abstand zu allen akustischen Begrenzungsflächen aufgestellt. Über diesen Lautsprecher wird ein weißes Rauschen so abgespielt, dass es nicht zu Signalverzerrungen oder zur Zerstörung

⁴⁶Der Einsatz von Hallgeräten, um einen filmischen Raum besser erzählen zu können, wird durch eine trockene Raumakustik ebenfalls begünstigt.

des Lautsprechers kommt. Dieses Signal wird mit einem Kunstkopf-Mikrofon aufgenommen, das so aufgestellt ist, dass der Lautsprecher auf dem vorderen Schnittpunkt von Horizontal- und Medianebene liegt. Der Abstand zwischen Kunstkopf-Mikrofon und Lautsprecher sollte dem späteren Abstand zwischen Lautsprechern und Rezipienten entsprechen, um eine Aussage treffen zu können, wie gut eine Wiedergabe unter den gegebenen Umständen ist. Um verschiedene Räume miteinander vergleichen zu können, wäre es sinnvoll, einen Abstand zu normieren, da die Kohärenz auch vom Abstand abhängig ist.

Aus den aufgenommenen Signalen wird die interaurale Kohärenz berechnet. Ist der Wert kleiner eins, verschlechtert sich dadurch auch die maximale -Lokalisations-schärfe. Unter 0,6 (siehe zu diesem Wert Kapitel 3.4.3, Seite 53) liegt die maximale -Lokalisationsschärfe nur noch bei 10° auf dem vorderen Schnittpunkt zwischen Median- und Horizontalebene. Unter 0,4 fällt die Lokalisierbarkeit weiter ab. Unter 0,3 ist eine sichere Lokalisation eventuell nicht mehr gegeben, nach Blauert (1974) entsteht eine zu diffuse Wahrnehmung. Die Angaben müssten jedoch durch weitere Untersuchungen noch besser belegt werden.

Ob für die Messungen ein Kunstkopf-Mikrofon zwingend erforderlich ist, könnten nur weitere Untersuchungen zeigen. Eventuell reichen bereits leicht gerichtete Mikrofone oder ein Trennkörper zwischen den Mikrofonen. Weiterführend wäre auch zu untersuchen, ob ein Zusammenhang mit andern akustischen Größen wie Klarheitsmaß, Deutlichkeitsmaß, STI etc. besteht. Eventuell könnte aus diesen Daten ebenfalls ein Rückschluss auf die maximale Lokalisationsschärfe gezogen werden.

Die maximalen Punkt-Lokalisationsschärfe

Die Ausführungen zur Raumakustik legen nahe, den Begriff der maximalen Punkt-Lokalisationsschärfe einzuführen. Sie beschreibt den unter den gegebenen Umständen maximal zu erreichenden Wert der Lokalisierbarkeit für eine Quelle, die auf dem vorderen Schnittpunkt von Median- und Horizontalebene liegt.

Da unter optimal Bedingungen für diesen Punkt die bestmögliche Lokalisationsschärfe mit etwa 1° horizontal und 4° in Medianebene angegeben ist, kann so ein Rückschluss gezogen werden, wie gut z.B. die maximale akustische Abbildungsfähigkeit eines Wiedergabesystems in einem Raum sein kann.

Für alle anderen Raumwinkel gilt, dass – abhängig von Median- und Horizontalwinkel – die Lokalisationsschärfe entsprechend der Ausführungen in Kapitel 2.2, Seite 15 geringer ist als die maximale Punkt-Lokalisationsschärfe.

5.2.4 Wechselwirkungen zwischen mehreren Hörereignissen

Nach den Schlussfolgerungen in Kapitel 3.4, Seite 50 und Kapitel 4.3.7, Seite 75 ist bei mehreren Signal abstrahlenden Quellen eine Verschlechterung der Lokalisation der resultierenden Hörereignisse festzustellen.

Die Untersuchungen, die für diese Arbeit durchgeführt worden sind, weisen darauf hin, dass bei zwei Quellen die Lokalisationsunschärfe auf 10° steigt und mit weiter wachsender Anzahl größer wird. Durch die Analyse aus Kapitel 4.3.7 konnte jedoch gezeigt werden, dass es sich nicht um einen Allgemeinfall handelt. So können auch mehrere Quellen über ihr resultierendes Hörereignis durchaus sehr genau lokalisiert werden. Es konnte aber auch belegt werden, dass mit wachsender Anzahl von Schallquellen eine genaue Lokalisation immer unwahrscheinlicher wird bzw. diese in der Regel nachlässt.

Dabei ergeben sich verschiedene Qualitäten der Lokalisierbarkeit:

- Diffuse Lokalisation: Umschließend und nicht ortbar
- Hemisphärische Lokalisation: In einer bestimmten Raumsektorlage⁴⁷ nicht näher lokalisierbar
- Breite Lokalisation: Mit einer bestimmten Ausdehnung (oder Unschärfe) in einer Richtung lokalisierbar
- Punkt-Lokalisation: Die Lokalisationsschärfe ist maximal

Die diffuse Lokalisation kann bei mehreren teilkoheränt⁴⁸ abstrahlenden Quellen, die um den Rezipienten verteilt sind, auftreten, oder wenn sehr viele inkohärent abstrahlende Quellen (mehr als 16) um den Zuhörer verteilt sind. Liegen die genannten Quellen vornehmlich in einem bestimmten Raumsektor, entsteht eine hemisphärische Lokalisation.

Kann eine Quelle als Hörereignis separiert werden, findet eine breite Lokalisation statt, wenn das Quellensignal geringe Kohärenzen zu anderen Quellensignalen aufweist. Besteht keine Kohärenz, kommt es zur Punktlokalisierung, eine Quelle kann mit größtmöglicher Genauigkeit lokalisiert werden.

Die Punkt-Lokalisation ist nur sehr eingeschränkt im vorderen Hörraum möglich, nahe dem vorderen Schnittpunkt von Horizontal und Medianebene. Da der Kopf frei beweglich ist, gilt diese Lokalisationsfähigkeit für jeden Bereich des Hörraums, auf den der vordere Schnittpunkt gerichtet werden kann. Nach Blauert (1974) muss der Kopf nicht voll in Richtung der Schallquelle ausgelenkt werden, um den „*cone of confusion*“

⁴⁷vordere, hintere, obere, untere, linke oder rechte Hemisphäre.

⁴⁸Die Kohärenzen sind so groß, dass die Quellen nicht in einzelne Hörereignisse separiert werden können. Eine explizite Aussage ist schwer möglich, da die Signalform mit berücksichtigt werden muss. Eigene Erfahrungen zeigen, dass die Signale von Mehrkanal-Aufnahmen, die mit konventionellen Mikrofonen außerhalb des Hallradius gemacht wurden die Bedingung in der Regel erfüllen.

aufzulösen und komplex lokalisieren zu können. Je nach Signalform kann aber eine Vollausrichtung des Kopfes nötig sein, um eine maximale Lokalisationsschärfe zu erreichen.

5.2.5 Lokalisierbarkeit unter speziellen Bedingungen

Da die allgemeinen Bedingungen für die Lokalisierbarkeit praktisch nicht relevant sind, gilt es, spezielle Bedingungen zu formulieren. Diese können über fünf Faktoren definiert werden:

- Die maximale vertikale und horizontale Beweglichkeit des Kopfes
- Die Signalform der abstrahlenden Quellen
- Die Lage der Quellen relativ zum Kopf
- Die Stärke der Wechselwirkung zwischen mehreren Signalen
- Die Raumakustik

Die Beweglichkeit des Kopfes ist maßgeblich davon abhängig, ob der Rezipient steht oder sitzt und ob er durchgehend einen Bildinhalt verfolgen soll.

Nach Günther et al. (2013) kann der Kopf maximal jeweils um 40 Grad nach vorne und hinten geneigt werden. Nach Volkmann (1872) ist es möglich, den Kopf jeweils maximal 91 Grad nach links und rechts zu drehen, sofern der gesamte Oberkörper dabei beteiligt ist.

Unter typischen Rezeptionsbedingungen – wie sie z.B. im Kino herrschen – sind diese Werte nicht zu erreichen. Wird von einer sitzenden Haltung und einer Kopfbewegung ausgegangen, die die Wirbel nicht belastet und ein uneingeschränktes Verfolgen des Bildes zulässt, kann von einer maximalen Drehbewegung von 32° und einer maximalen Neigung von 15° ausgegangen werden (Volkmann 1872, Günther et al. 2013, Schmidt 2005).

Die Signalform der Quelle kann ebenfalls zu einer Verschlechterung der Lokalisation beitragen. So gilt eine Lokalisationsunschärfe auf dem frontalen Schnittpunkt von Horizontal- und Medianebene, die je nach Signalart zwischen $0,75^\circ$ und $11,8^\circ$ liegt. Für reale Signale kann von etwa 2° horizontal (siehe Kapitel 2.2, Seite 15) und 10° vertikal (siehe Kapitel 2.3, Seite 22) ausgegangen werden.

Die relative Quellenlage beeinflusst die Lokalisierbarkeit ebenfalls negativ, wie Untersuchungen zur Auslenkung in Median- und Horizontalebene zeigen (vgl. Kapitel 2.3, Seite 22, Kapitel 2.6.3, Seite 30, Kapitel 3.4.3, Seite 53). Inwieweit eine Abhängigkeit zwischen Signalform und relativer Quellenlage besteht, ist zur Zeit noch nicht untersucht. Es ist nicht klar, wie sich beide Faktoren gemeinsam auf die Lokalisationsschärfe auswirken. Als gesichert kann aber gelten, dass die Lokalisationsunschärfe mindestens so hoch ist, wie der Wert des signifikantesten Faktors.

Über die Wechselwirkung der Faktoren kann mit dem Wissensstand dieser Arbeit keine quantitative Aussage getroffen werden. Es kann nur belegt werden, dass keine Wechselwirkung vorliegt, wenn sich verschiedene Signale spektral nur kurzzeitig überdecken. In diesem Fall kommt es zu keiner Verschlechterung der Lokalisations-schärfe. Weiter kann belegt werden, dass eine maximale Wechselwirkung vorliegt, wenn mehrere Quellen das gleiche Signal abstrahlen. In diesem Fall verschlechtert sich die Lokalisierbarkeit erheblich (siehe dazu Kapitel 4.3.7, Seite 75).

Zur Akustik sei auf Kapitel 5.2.3, Seite 82 hingewiesen. Auch für diesen Faktor gilt, dass noch nicht klar ist, wie dieser Abhängigkeiten mit anderen Faktoren bildet. So kann auch in diesem Fall nur gelten, dass die Lokalisationsunschärfe so hoch ist wie der Wert des signifikantesten Faktors.

Mit der Ausgangslage, dass die Signale nicht zu Wechselwirkungen führen und die Raumakustik keine signifikante Verschlechterung der Lokalisationsschärfe verursacht, können konkrete Aussagen zur maximalen Lokalisierbarkeit im Kino gemacht werden. Davon ausgehend, dass der Kopf des Rezipienten maximal um 32° gedreht und um 15° nach vorne und hinten geneigt wird, die Signalform der abstrahlenden Quellen eine maximale Lokalisationsschärfe von 2° horizontal und 10° vertikal zulassen, kann auf eine explizite richtungsabhängige Lokalisationsunschärfe geschlossen werden. Diese kann den Polardiagrammen in Abbildung 5.1, Seite 87 entnommen werden.

Da einige Abhängigkeiten – wie beschrieben – noch nicht geklärt sind, kann die Grafik nur unter Vorbehalt gelten. Es sei drauf hingewiesen, dass für die vertikale Lage nicht genügend Daten vorliegen, um mit Bestimmtheit sagen zu können, wie groß die Lokalisationsunschärfe für 90° tatsächlich ist. Eventuell kann sie kleiner als 22° ausfallen. Auch besteht der Vorbehalt, dass die analysierten Theorien und damit auch das in Kapitel 4, Seite 59) aufgestellte Modell gültig sind. Für die vertikale Auslenkung muss auch beachtet werden, dass die Signale mehrerer Quellen insbesondere für die Höhenlokalisierung in der Regel Wechselwirkungen zeigen. Dadurch fällt die erreichbare Lokalisationsschärfe meist geringer aus, als die Grafik angibt (siehe dazu Kapitel 5.3, Seite 88).

An diesem Beispiel wird klar, dass eine genaue quantitative Aussage nur schwer möglich ist, da sehr viel Abhängigkeiten bestehen, die oft nur eine Tendenz als Aussage zulassen. Im Falle der Abbildung 5.1, Seite 87 kann zumindest belegt werden, dass im Kino seitliche Quellen und insbesondere solche in der Höhe deutlich schlechter lokalisiert werden können als jene in Frontallage.

5 Zusammenfassung der Ergebnisse

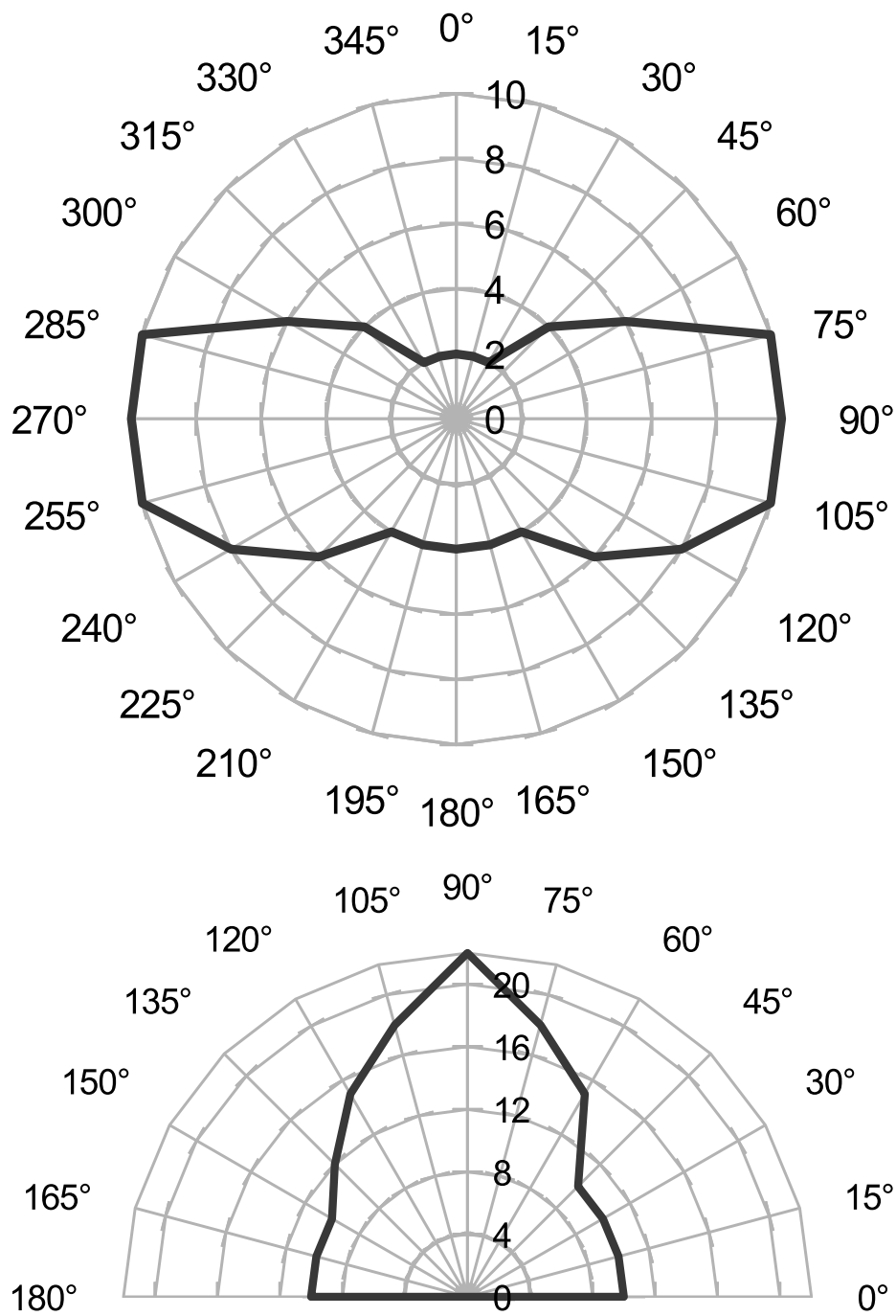


Abbildung 5.1: Lokalisierbarkeit in Kino. Oben: Horizontale richtungsabhängige Lokalisationsunschärfe in Grad; Unten: Vertikale richtungsabhängige Lokalisationsunschärfe in Grad.

5.3 Differenzierbarkeit

Ob Schallquellen in einzelne Hörereignisse differenziert werden können, hängt von verschiedenen Faktoren ab. Die Analyse der kortikalen Prozesse zeigt, dass das Signale einer Quelle auch in mehrere Hörereignisse in Form von Melodien separiert werden kann⁴⁹ (siehe Kapitel 4.3.6, Seite 74). Ein ähnlicher Effekt entsteht auch bei der Phantomschallquellenbildung (siehe Kapitel 2.8, Seite 35). Dabei erzeugen lediglich zwei Schallquellen mehrere Hörereignisse unterschiedlicher seitlicher Auslenkung.

5.3.1 Signalform

Die Differenzierbarkeit ist insbesondere von der Signalform abhängig. Kann keine Separation von Mustern und damit keine Separation in Hörereignisse im auditiven Kortex stattfinden, können mehrere Quellensignale nicht in einzelne Hörereignisse separiert werden. Insbesondere Rauschsignale lassen nur sehr begrenzt eine Separation, auf Basis der Quellen-Richtungsinformationen zu. Daher ist es unmöglich durch Test an Probanden, die maximal mögliche Lokalisationschärfe, im Falle mehrerer Quellen die ein Signal ausstrahlen, zu bestimmen. Den, damit die Lokalisationschärfe maximal sein kann muss ein Rauschen verwendet werden, dieses lässt aber bei mehreren Quellen keine Separation mehr zu. Ohne Separation ist jedoch keine Lokalisation möglich. Des wegen kann eine Aussage betreffen der Lokalisation und Differenzierbarkeit nur jeweils für einen Spezialfall gemacht werden. Eine allgemeine Aussage ist nicht möglich.

5.3.2 BMLD und BILD

Die Untersuchungen zur BMLD – wie sie in Kapitel 2.6.4, Seite 32 geschildert wurden – zeigen, dass es eine absolute Grenze für die Anzahl von gleichzeitig abstrahlenden Schallquellen gibt, die keine Differenzierung in die den Quellensignalen entsprechenden Hörereignisse mehr zulässt. Eine Messung der BMLD ist jedoch nur mit Kopfhörern möglich, da – wie in Kapitel 2, Seite 14 und Kapitel 4 Seite 59 geschildert wurde – für Quellen im Schallfeld eines Raumes zu viele Abhängigkeiten bestehen, die eine allgemeine Bestimmung der BMLD nicht zulassen. Dementsprechend ist keine allgemeine Aussage zur Differenzierbarkeit im Raum möglich. Es konnten jedoch einige Messbedingungen, die bei Kopfhörer-Wiedergabe möglich sind, für Quellen im Raum ausgeschlossen werden. Dadurch ist zumindest die Aussage möglich, dass eine Differenzierbarkeit von mehr als 16 Quellen in einem Raum unwahrscheinlich ist. Jenseits dieser Grenze kommt es mit hoher Wahrscheinlichkeit zur Verschmelzung einzelner Quellensignale zu einem Hörereignis.

⁴⁹Forschungen von Blauert (1974) zeigen, dass diese Melodien getrennt lokalisiert werden. In der Regel werden sie zwar in Quellenrichtung verortet, durch Wechselwirkungen im Separationsprozess kann es jedoch zu leicht unterschiedlichen Richtungsempfindungen kommen.

Aus Untersuchungen der BILD geht hervor, dass es nicht möglich ist, mehr als 8 Dialoge so von einander zu separieren, dass der Dialog, der im Fokus der Aufmerksamkeit liegt, verstanden werden kann. Dieser Wert ist für die Produktion von 3-D-Audio-Inhalten interessant, wenn z.B. mehre Dialoge gleichzeitig stattfinden und der Rezipient selbst entscheiden soll, welchem Dialog er folgen will⁵⁰.

Speziell für die Differenzierbarkeit von Hörereignissen in der Höhe ergibt sich eine wichtige Aussagen aus der BMLD (sieh dazu Kapitel 2.6.5, Seite 34). Da die BMLD zu hohen Frequenzen hin abfällt und über 2 kHz bereits unter 3 dB beträgt, kommt es bei Hörereignissen in Höhenlage zu starken Wechselwirkungen, da die Höhenlage nur monaural durch Frequenzverzerrungen oberhalb von 8 kHz bestimmt werden kann. Bereits bei zwei Quellen entsteht, betreffend ihrer Medianlage, eine große Lokalisationsunschärfe. Durch die komplexe Separation ist es möglich, die einzelnen Quellen in Hörereignis aufzugliedern, dies ist jedoch nur zu Lasten einer steigenden Lokalisationsunschärfe in der Medianebene möglich (siehe Kapitel 4.3.7, Seite 75)

5.3.3 auditive Separation

Die Untersuchungen zur auditiven Separation zeigen, dass gleichzeitig maximal sieben Hörereignisse ausgebildet und damit differenziert werden können. Ist die Quellenanzahl größer, verschmelzen die Signale einzelner Quellen jeweils zu einem Hörereignis. Durch die fokussierte Aufmerksamkeit kann jedoch jede Quelle jeweils in ein Hörereignis separiert werden, sofern die einzelnen Signale eine Separation zulassen (siehe Kapitel 4, Seite 59).

5.3.4 Einfluss der Akustik

In einem Raum entsteht Hall, indem das Signal einer Quelle von den Wänden reflektiert wird. Diese Reflexionen und die folgenden Reflexionen können jeweils als eigene Schallquelle, die Spiegel-Schallquelle, betrachtet werden. Diese Spiegel-Schallquelle strahlt naturgemäß ein Signal ab, dass dem der primären Quelle ähnlich ist. Durch diese Ähnlichkeit hat der Hall eine negative Auswirkung auf die Differenzierbarkeit. Es ist dem auditiven Kortex möglich, den Hall stark zu unterdrücken, dennoch werden die primitiven und komplexen Separationsprozesse behindert. In sehr halliger Umgebungen ist die Separation von Hörereignissen deutlich eingeschränkt (siehe Kapitel 2.6.1, Seite 29).

⁵⁰Insbesondere für Computer-Rollenspiele wäre dieses Wissen interessant, da man dem Rezipienten bis zu 8 Dialoge gleichzeitig in einer Handlung anbieten kann und er selber entschieden kann, welchem Dialog er folgen will, um Informationen zu erlangen, die für eine Spiel-entscheidende Wahl von Bedeutung sind.

5.4 Die dritte Dimension

Von einer dreidimensionalen Wahrnehmung kann gesprochen werden, wenn auch eine Tiefenwahrnehmung stattfindet, also Entfernungen erkannt werden können. Physikalisch ist dies jedoch durch das Gehör nicht möglich (siehe Kapitel 2.4, Seite 24). Allerdings hat das Gehirn Heuristiken entwickelt, die eine Abschätzung der Entfernung ermöglichen. Diese sind jedoch sehr subjektiv. Durch diese subjektive Entfernungsbestimmung und die Möglichkeit, horizontal und vertikal zu lokalisieren, kann von einer dreidimensionalen Hörfähigkeit gesprochen werden.

Die Lokalisation in der Vertikalen ist dabei nur durch Rückkopplungsprozesse der fokussierten Aufmerksamkeit möglich (siehe Kapitel 4.4, Seite 76). Diese Vorgänge unterliegt mitunter hohen Latenzen und sind fehleranfällig. Da binaurale Informationen nur sehr beschränkt für die Höhen-Lokalisation genutzt werden können, ist die Lokalisationsschärfe in der Medianebene wesentlich geringer als in der Horizontalebene (siehe Kapitel 2.3, Seite 22). Mit steigendem Erhebungswinkel nimmt die Lokalisationsschärfe stark ab. Bei mehreren in der Höhe lokalisierten Hörereignissen funktionieren die komplexen Separationsprozesse nicht mehr einwandfrei (siehe Kapitel 4.3.7, Seite 75). So entstehen zusätzliche Lokalisationsunschärfen. Im Extremfall ist es möglich, dass keine Hörereignis-Separation mehr stattfindet. Mehrere Quellen bilden dann zusammen eine Phantomschallquelle, die nur noch sehr breit lokalisiert werden kann⁵¹.

5.5 Reproduktion von Schallfeld-Informationen

Um eine dreidimensionale Lokalisierbarkeit zu gewährleisten, ist es notwendig, Verfahren zu benutzen, die an den Ohren des Zuhörers ein Schallfeld erzeugen, das dem einer realen Schallquelle ähnlich ist. Das ist z.B. durch WFS oder HOA möglich.

5.5.1 Abtastung in verschiedenen Frequenzbändern

Ab 8 kHz ist eine binaurale Lokalisation in Raum, z.B. bei der Wiedergabe über Lautsprecher, nicht mehr möglich, da es zwischen beiden Ohren zur Abschattung kommt (siehe Kapitel 2.2.1, Seite 18). Es kann lediglich unterschieden werden, ob eine (virtuelle) Quelle links oder rechts vom Kopf liegt⁵².

Der Schluss, dass deswegen das Schallfeld oberhalb dieser Frequenz nicht mehr reproduziert werden muss, da in diesem Bereich keine Lokalisation mehr möglich wäre, ist jedoch falsch und wäre fatal für die Lokalisierbarkeit in der Höhe. Denn die

⁵¹Unter Umständen kann es zu einem Orientierungsreflex kommen, der den Kopf des Hörenden einer Quelle zuwenden lässt, um sie separieren zu können.

⁵²Der Rezipient könnte eine Quelle über 8 kHz binaural durch eine akustische Regula falsi approximieren, indem er durch Auslenkung des Kopfes den Punkt findet, über dem die Wahrnehmung der Quelle lateral springt. Praktisch hat das jedoch keine Relevanz. Bei Kopfhörer-Wiedergabegilt die Grenze von 8 kHz nicht.

monauralen Informationen können bis zum Ende des hörbaren Spektrums ausgewertet werden. Insbesondere für die Lokalisation in der Höhe sind dabei Frequenzen um 8 kHz relevant (siehe Kapitel 2.3, Seite 22). Gerade in diesem Bereich muss also das Schallfeld sehr genau rekonstruiert werden, damit eine präzise Lokalisation in der Höhe überhaupt möglich ist.

Im Rahmen dieser Arbeit ist es nicht möglich, belegen zu können, ob eine fehlerhafte Reproduktion des Schallfeldes einen größeren Fehler bei der monauralen Lokalisation verursacht als bei der binauralen Lokalisation. Sicher kann aber gesagt werden, dass eine fehlerhafte Reproduktion des Schallfeldes bei hohen Frequenzen durch die kürzere Wellenlänge wesentlich größere Fehler erzeugt, als bei tiefen Frequenzen.

Zusätzlich nimmt die Lokalisationsschärfe zu den tiefen Frequenzen hin ab, sodass bei sehr tiefen Tönen nur noch die Raumsektorlagen bestimmt werden können. Eine belegte Grenze, ab welcher unteren Frequenz dies eintritt, war in der Literatur allerdings nicht zu finden. In der Praxis werden oft Werte um 100 Hz angegeben, die jedoch fachlich nicht belastbar sind (Görne 2008).

Aus diesen Erkenntnissen kann geschlossen werden, dass je tiefer die Frequenz ist, desto ungenauer kann das Schallfeld reproduziert werden, ohne dass die Lokalisationsfähigkeit eines Zuhörers beeinträchtigt wird. Bei hohen Frequenzen, insbesondere um 8 kHz, muss das Schallfeld sehr hoch aufgelöst werden, um eine (gute) Höhenlokalisierung gewährleisten zu können.

Darin ist auch ein Grund zu finden, dass aktuelle Wiedergabesysteme insbesondere bei der Höhenlokalisierung keine zufriedenstellenden Resultate liefern. Oft werden einfache Mehrwegelautsprecher einzeln angesteuert, um im Verbund ein Schallfeld zu erzeugen. Bei tiefen und mittleren Frequenzen ist damit eine sehr genaue Rekonstruktion des Schallfeldes möglich (unter Umständen genauer als nötig). Bei hohen Frequenzen ist dann jedoch die Anzahl der (einzeln ansteuerbaren) Hochtöner zur gering, um ein Schallfeld genügend genau reproduzieren zu können.

Es wäre sinnvoll, ein Schallfeld, in mehrere Frequenzbänder aufzuteilen und jeweils nur so genau abzutasten und wiederzugeben, wie nötig. Für eine gute Höhenlokalisierung könnte es eventuell reichen, das Schallfeld in einem relativ schmalen Band um 8 kHz sehr genau abzutasten und wiederzugeben.

5.5.2 Wahrnehmungsprozesse

Durch den Wahrnehmungsprozess kommt es unter bestimmten Umständen zeitweise zu größeren Lokalisationsunschärfen, die eine genauere Reproduktion des Schallfeldes unnötig machen. Insbesondere, wenn sehr viele Schallquellen gleichzeitig ein Signal abstrahlen, kann es zu einer verminderten Lokalisationsfähigkeit beim Rezipienten kommen. Diese Effekte könnten genutzt werden, um zu diesen Zeitpunkten die Auflösung des Schallfeldes entsprechend zu verringern.

5.5.3 Wiedergabesysteme

Für Wiedergabesysteme muss die Abhängigkeit zwischen Frequenz und nötiger Schallfeldauflösung beachtet werden (siehe dazu Kapitel 5.5.1, Seite 90). Daher ist es sinnvoll, unabhängig voneinander Tief-, Mittel- und Hochtöner, entsprechend der nötigen Schallfeldauflösung, im Raum zu verteilen und getrennt anzusteuern. Zusätzlich kann berücksichtigt werden, dass die mögliche Lokalisationsschärfe an den Seiten und in der Höhe geringer ist. Dementsprechend kann die Anzahl von Tief- Mittel- und Hochtönern reduziert werden. Insbesondere in Horizontallage (auf Ohrhöhe) kann die Anzahl der Hochtöner stark verringert werden, da hier eine genaue Lokalisation vor allem über binaurale Informationen stattfindet. Für diese sind bei mehreren Quellen vorrangig Signalanteile unter 2 kHz ausschlaggebend, darüber hinaus sinkt die Lokalisationsschärfe bei mehreren Quellen durch die zu geringe BMLD ab (siehe dazu Kapitel 2, Seite 14).

5.5.4 Möglichkeit zur verlustbehafteten Komprimierung

Das Erfassen der Informationen eines Schallfeldes benötigt sehr viel Speicherplatz. Daher ist es nötig, die Daten zu reduzieren, ohne dass es zu einer Verringerung der Lokalisationsfähigkeit des Rezipienten kommt. Dafür würde sich einerseits anbieten, das Schallfeld in verschiedenen Frequenzbändern unterschiedlich genau abzutasten (siehe dazu Kapitel 5.5.1, Seite 90) und anschließend jeweils verlustbehaftet zu komprimieren.

Eine weitere Möglichkeit bestünde darin, ein heuristisches Modell zu entwickeln, das analysiert, ob Bedingungen vorliegen, die beim Rezipienten zu einer Verschlechterung der maximalen Lokalisationsschärfe führen (siehe dazu Kapitel 5.2, Seite 81). In diesem Fall könnte das gesamte Schallfeld ebenfalls unschärfer abgetastet werden, wodurch sich die Datenmenge verringern würde. Theoretisch wäre es sogar möglich, durch Aufmerksamkeitsvorhersage verlustbehaftet komprimieren zu können. In diesem Fall würde nur diejenige Richtung eines Schallfeldes genauer abgetastet, die im Fokus der Aufmerksamkeit liegt, weil nur diese Richtung mit einer hohen Lokalisationsschärfe erfasst wird. Möglich wäre dies z.B., indem im Postproduktionsprozess virtuelle Schallquellen, die den Fokus an sich binden, wie z.B. Dialoge, extra für die Kompression gekennzeichnet würden.

5.6 Sounddesigns und Mischung von 3D-Audio-Inhalten

Durch falsche Vorgehensweisen in der Mischung kann es zu einer eingeschränkten Lokalisierbarkeit kommen und zum Verschmelzen einzelner Quellensignale zu einem Hörereignis. Insbesondere die Ausführungen zu Faktoren im Kapitel 2.9, die die Lokalisierbarkeit vermindern (Seite 39) sind daher zu berücksichtigen.

Mehreren Quellen können sich gegenseitig in der Lokalisierbarkeit beeinflussen. Dem kann mit der „Vier-aus-Sieben“-Regel entgegengewirkt werden: Maximal sieben Quellen können gleichzeitig als Hörereignis erfasst werden, von denen nur vier komplex lokalisierbar sind. Diese vier Quellen sollten in den Vordergrund gemischt werden, um eine Lokalisation zu begünstigen.

Es können mehr als sieben Quellen lokalisiert werden, allerdings ist bei steigender Quellenzahl mit einer Verringerung der Lokalisierbarkeit zu rechnen. Die Untersuchungen der Arbeit weisen darauf hin, dass es nicht möglich ist, mehr als 16 (virtuelle) Quellen gleichzeitig im Hörfeld zu positionieren, ohne dass einzelne Quellen unseparierbar zu einem Hörereignis verschmelzen. Um ein natürliches Klangbild zu erreichen, sollte eher mit einem diffusen Hintergrund gearbeitet werden, der in Form von teilkohärenten Signalen über wenige virtuelle Quellen verteilt um den Zuschauer herum ein diffuses Hörereignis hervorruft. Aus diesem Grund brauchen dann nur wenige vordergründige Quellen platziert werden, die eine gute Punkt-Lokalisation erlauben.

Für Hörereignisse, die bei einem Zuhörer entstehen, ist zu überlegen, in welcher Qualität diese lokalisiert werden sollen. Dabei kann die Auflistung in Kapitel 5.2.4, Seite 84 hilfreich sein, um dementsprechend eine oder mehrere virtuelle Schallquellen zu setzen.

5.6.1 Tempo und schnelle Schnitte

Schnelle Schnitte und Bewegungen führen immer zu einer Verschlechterung der Lokalisierbarkeit. Im Extremfall können virtuelle Schallquellen nicht mehr wahrgenommen werden. Bei schnellen Schnittfolgen können Szenen, die viele virtuelle Quellen enthalten, gar nicht komplex erfasst werden. Der Aufwand, der im Sounddesign betrieben worden ist, kann nicht wahrgenommen werden. Ob die virtuellen Quellen einer schnell geschnittenen Szene überhaupt sicher lokalisiert werden können, kann durch Berechnung der komplexen Lokalisationsgrenze überprüft werden (siehe Kapitel 5.1.2, Seite 5.1.2)

5.6.2 Entfernung und Höhe

In der Höhe sollten nur wenige virtuelle Schallquellen gleichzeitig ein Signal abstrahlen, da die resultierenden Hörereignisse stark wechselwirken, wodurch die Lokalisierungsschärfe abnimmt. Im Extremfall bildet sich nur eine Phantomschallquelle (siehe dazu Kapitel 2.3, Seite 22 und Kapitel 5.4, Seite 90).

Das Entfernungshören ist ein rein subjektiver Prozess der Wahrnehmung, der vor allem auf Erfahrungswerten des Zuhörers beruht. Daher kann die Entfernung nicht durch das Wiedergabesystem erzeugt werden. Die Entfernung entsteht erst durch die Gestaltung des Mischtonmeisters (siehe dazu Kapitel 2.4, Seite 24).

5.7 Fazit

In der Psychoakustik gibt es einige Untersuchungen, die sich mit der Lokalisation beschäftigen. Sie beleuchten jedoch nur Teilaspekte. Daher wurden die Untersuchungen analysiert und die Aspekte zusammengefasst, um erste Aussagen treffen zu können. Durch eine ergänzende eigene Untersuchung konnten neben qualitativen Aussagen auch erste quantitative Aussagen getroffen werden.

Eine spätere Analyse zeigte jedoch, dass diese Untersuchung keinen allgemeinen Rückschluss zulässt, da die Prozesse, die im Gehirn zur Bildung von Hörereignissen führen, nicht berücksichtigt wurden. Die interaurale Kohärenz hat nicht die signifikante Auswirkung auf die Lokalisierbarkeit und Differenzierbarkeit, wie zum Beginn der Untersuchungen erwartet wurde. Durch die Messung der interauralen Kohärenz können aber dennoch Tendenzen gezeigt werden. Für eine objektive Bewertung der Beeinflussung durch die Raumakustik ist die Messung der interauralen Kohärenz sogar sehr aussagekräftig.

Wo und wie die akustische Lokalisation und Separation in Hörereignisse im Gehirn stattfindet, ist noch nicht belegt. Neue Erkenntnisse zeigen lediglich, dass die Verarbeitung für die Lokalisation nicht nur im auditiven Kortex stattfindet und dass es bei der Lagebestimmung eines Objektes eine starke Wechselwirkung zwischen den Sinnen gibt. Eine explizite Aussage, wie die auditive Lokalisation und die Separation in Ströme durch neuronale Prozesse im Gehirn funktioniert, kann aufgrund des derzeitigen Forschungsstandes nicht getroffen werden.

Daher wurden gängige Theorien gesucht, die die Prozesse im Gehirn modellhaft beschreiben und als belegt gelten. Auf Basis dieser Theorien konnte ein Modell zur Hörereigniswahrnehmung entwickelt werden. Durch die Analyse des aufgestellten Modells und durch Verknüpfung mit Untersuchungsergebnissen der Psychoakustik ist es nun möglich, verschiedene Grenzen der Lokalisierbarkeit und Differenzierbarkeit zu definieren.

Es zeigte sich, dass eine allgemeine Aussage, wie gut ein oder mehrere Hörereignisse voneinander differenziert und lokalisiert werden können, nicht sinnvoll ist. Es muss daher immer ein Spezialfall betrachtet werden, um eine Aussage treffen zu können. Erst dadurch ist es möglich, explizite Grenzen zu definieren.

So konnten neue Erkenntnisse gewonnen werden, unter welchen Bedingungen ein Hörereignis entsteht und unter welchen Bedingungen es wie genau lokalisiert werden kann.

Das dreidimensionale Hören ist nur durch einen komplexen rückgekoppelten Wahrnehmungsprozess möglich, der auf Heuristiken beruht. Diese Heuristiken können jedoch zu falschen Wahrnehmungen führen. Insbesondere bei der Lokalisationsrichtung von Hörereignissen. In diesen Fehlern liegen die Grenzen der Lokalisierbarkeit und Differenzierbarkeit.

5 Zusammenfassung der Ergebnisse

Abschließend sei gesagt, dass die begrenzte Lokalisierbarkeit und Differenzierbarkeit von Hörereignissen auch positive Seiten hat: Ohne diese Begrenztheit würden keine Phantomschallquellen in der Wahrnehmung entstehen. Es würden lediglich die Signale der beteiligten Quellen wahrgenommen werden. Ein räumlicher Musikgenuss wäre dann nur denjenigen vergönnt, die sich eine Eintrittskarte für ein philharmonisches Konzert leisten können. Dagegen wäre es um die räumliche Abbildung von Populärmusik schlecht bestellt, da es das Stereoverfahren⁵³ nicht geben würde.

⁵³Und andere auf Phantomschallquellen-Bildung basierende Verfahren.

A Material

A.1 Daten zur Messung des interauralen Kohärenzgrades

Tabelle A.1: Materialliste

Anzahl	Gegenstand
1	Neumann KU 100 (Kunstkopfstereomikrofon)
1	MacBook 17“
3	XLR-Kabel 10 m
1	Adapter Kabel Klinke auf XLRm
1	Case
2	Fireface 800
1	D/A-Wandler Behringer
17	JBL Boxen mit Netzkabel
16	XLR auf Klinke Kabel 10 m
4	Lautsprecherstative 2 m mit Neiger
4	Lautsprecher-Bodenstative mit Neiger
7	verstellbare Lautsprecherstative
1	Lautsprecherstativ 1,2 m
1	Sennheiser Rauschgenerator (vermutlich MZL 400)
1	Schalldruckmessgerät (NTi XL2)

Tabelle A.2: Benennung der Lautsprecher: Dazu wurden Dreiertupel gewählt. Der erste Tupelwert beschreibt die Ebene oder auch den Layer, in dem der Lautsprecher liegt. Der zweite die Längsposition auf der Ebene und der Dritte die Querposition auf der Ebene. Sofern nur Einertupel mit dem Vermerk „Layer“ genannt sind, bezeichnet dies alle Lautsprecher der genannten Ebene; TMC entspricht dem Lautsprecher über dem Kunstkopf

1. Tupelstelle		2. Tupelstelle		3. Tupelstelle	
Wert	Bedeutung	Wert	Bedeutung	Wert	Bedeutung
M	Mittlere Ebene	F	Front	L	Links
T	Obere Ebene	S	Seitlich	C	Center
B Bottom	Untere Ebene	B	Hinten	R	Rechts
		M	Mitte		

Tabelle A.3: Lautsprecher-Positionen im Bezug auf dem Koordinatenursprung. Abkürzungen: Nr. = Nummer des Lautsprechers, X = Abstand auf der X-Achse (Links-Rechts in Hörrichtung), Y = Abstand auf der Y-Achse (Vorne – Hinten), Z = Abstand auf der Z-Achse (Höhe)

Lautsprecher	Nr.	X	Y	Z
Top Front Left	2	-1,6 m	1,4 m	1,9 m
Top Front Right	14	1,6 m	1,4 m	1,9 m
Top Back Left	6	-1,6 m	-1,4 m	1,9 m
Top Back Right	10	-1,6 m	-1,4 m	1,9 m
Top Mid Center	-	0	0	2,2 m
Mid Front Center	1	0	2 m	1 m
Mid Front Left	3	-1,6 m	1,4 m	1,2 m
Mid Front Right	15	1,6 m	1,4 m	1,2 m
Mid Side Left	5	-1,8 m	0	1,2 m
Mid Side Right	13	1,8 m	0	1,2 m
Mid Back Center	9	0	-1,7 m	1,2 m
Mid Back Left	7	-1,6 m	-1,4 m	1,2 m
Mid Back Right	11	1,6 m	-1,4 m	1,2 m
Bot Front Left	4	1,1 m	1,3 m	0,55 m
Bot Front Right	16	1,1 m	1,3 m	0,55 m
Bot Back Left	8	1,1 m	-1,3 m	0,55 m
Bot Back Right	12	1,1 m	-1,3 m	0,55 m

Tabelle A.4: Lautsprecher Anschlusskonfiguration. Abkürzungen: Nr. = Nummer des Lautsprechers, FF = Fireface, Fireface 1 das obere Fireface 2 das untere

Lautsprecher	Nr.	Interface Anschluss
Top Front Left	2	FF 2 AN2
Top Front Right	14	FF 1 AN6
Top Back Left	6	FF 2 AN6
Top Back Right	10	FF 1 AN2
Mid Front Center	1	FF 2AN1
Mid Front Left	3	FF 2 AN3
Mid Front Right	15	FF 1 AN7
Mid Side Left	5	FF 2 AN5
Mid Side Right	13	FF 1 AN5
Mid Back Center	9	FF 1 AN1
Mid Back Left	7	FF 2 AN7
Mid Back Right	11	FF 1 AN3
Bot Front Left	4	FF 2 AN 4
Bot Front Right	16	FF1 AN8
Bot Back Left	8	FF2 AN8
Bot Back Right	12	FF1 AN4
Top Mid Center	-	D/A Behringer Out 1 über FF1 Adat 1/1

Tabelle A.5: Messungen des interauralen Kohärenzgrades mit korrelierten Signalen. Abkürzungen: DrsCh = Durchschnittswert, g-Kor = Kohärenz über die gesamte File-Länge, LS = Anzahl verwendeter Lautsprecher, l(s) = File-Länge in Sekunden.

Dateiname/Stimulus	Werte gefensterte Kohärenz			g-Kor	LS	l(s)
	Maxima	Minima	Drsch			
M Layer Kor	0,46806	0,37210	0,40420	0,40480	8	10,0345
M T B Layer Kor	0,33201	0,26498	0,29966	0,29995	16	10,0345
MFC MBC Kor	0,78116	0,73687	0,75230	0,75282	2	10,0345
MFC MSL MSR MBC Kor	0,47384	0,39384	0,42766	0,42813	4	10,0345
MFL MBR Kor	0,93901	0,51200	0,62608	0,69069	2	10,0345
MFL MFR Kor	0,74624	0,66156	0,68979	0,69128	2	10,0345
MFL MFR MBL MBR Kor	0,61026	0,33993	0,38597	0,39018	4	10,0345
MSL MSR Kor	0,67672	0,61353	0,64002	0,64264	2	10,0362
T Layer Kor	0,50128	0,43181	0,45392	0,45526	4	10,0345
TFL TFR Kor	0,50096	0,43472	0,46025	0,46156	2	10,0345

Tabelle A.6: Messungen des interauralen Kohärenzgrades mit unkorrelierten Signalen. Abkürzungen: Drsch = Durchschnittswert, g-Kor = Kohärenz über die gesamte File-Länge, LS = Anzahl verwendeter Lautsprecher, l(s) = File-Länge in Sekunden.

Dateiname/Stimulus	Werte gefensterter Kohärenz			g-Kor	LS	l(s)
	Maxima	Minima	Drsch			
B Layer	0,21370	0,09212	0,13460	0,13369	4	10,0345
BBL	0,45671	0,30912	0,35853	0,36178	1	10,0345
BBR	0,53653	0,38189	0,42702	0,42968	1	10,0345
BFL	0,53183	0,42171	0,46407	0,46682	1	10,0345
BFR	0,63199	0,56165	0,58873	0,58969	1	10,0345
M Layer	0,18896	0,13101	0,16112	0,16121	8	10,0345
M T B Layer	0,13270	0,07467	0,09874	0,09830	16	10,0345
MBC	0,82432	0,64711	0,71491	0,72026	1	10,0345
MBL MBR	0,31071	0,14858	0,21522	0,21122	2	10,0345
MBL	0,53871	0,36373	0,42110	0,42397	1	10,0345
MBR	0,54204	0,39581	0,45517	0,45787	1	10,0345
MFC MBC	0,68365	0,65502	0,66927	0,66958	2	10,0345
MFC MSL MSR MBC	0,42991	0,35783	0,38670	0,38709	4	10,0345
MFC MSL	0,52508	0,47825	0,50064	0,50082	2	10,0345
MFC	0,79804	0,75597	0,77205	0,77247	2	10,0345
MFL MFR MBL MBR	0,16017	0,09465	0,12446	0,11782	4	10,0345
MFL MFR	0,23345	0,15192	0,18528	0,18599	2	10,0345
MFL	0,47890	0,31334	0,36844	0,37265	1	10,0345
MFR MBR	0,36224	0,26984	0,30748	0,30418	2	10,0345
MFR MSR MBR	0,28863	0,14547	0,19007	0,18674	3	10,0345
MFR	0,55827	0,49940	0,52795	0,52943	1	10,0345
MSL MSR	0,19977	0,09046	0,12370	0,11763	2	10,0345
MSL	0,45238	0,27952	0,33854	0,34116	1	10,0345
MSR	0,43193	0,28587	0,34002	0,34373	1	10,0345
T Layer	0,25610	0,16028	0,19276	0,19291	4	10,0345
TBL	0,59888	0,33791	0,43982	0,44756	1	10,0345
TBR	0,64395	0,37674	0,47399	0,48063	1	10,0345
TFL TFR MFC	0,36988	0,27475	0,31760	0,31792	3	10,0345
TFL TFR	0,35421	0,25145	0,29343	0,29389	2	10,0345
TFL	0,62458	0,49413	0,55023	0,55302	1	10,0345
TFR	0,65628	0,61588	0,62970	0,62984	1	10,0345
TMC	0,88328	0,81283	0,83531	0,83671	1	10,0345

Tabelle A.7: Verschiedene andere Messungen des Kohärenzgrades Teil 1. Abkürzungen:
 DrsCh = Durchschnittswert, g-Kor = Kohärenz über die gesamte File-
 Länge, l(s) = File-Länge in Sekunden.

Dateiname/Stimulus	Werte gefensterter Kohärenz			g-Kor	l(s)
	Maxima	Minima	Drsch		
Rosa_unkorel	0,34967	0,11836	0,19883	0,05554	10,00
Weiss_unkorel 12ms5	0,19928	0,09610	0,12514	0,00682	10,000
Weiss_unkorel 25ms	0,14869	0,07280	0,09397	0,00682	10,000
Weiss_unkorel 50ms	0,08737	0,05546	0,06975	0,00682	10,000
Weiss_unkorel 100ms	0,06420	0,04367	0,05167	0,00682	10,000
Weiss_unkorel 200ms	0,05168	0,03266	0,03863	0,00682	10,000
Weiss_unkorel	0,05168	0,03266	0,03863	0,00682	10,000
12ms5Nois	NaN	NaN	NaN	0,10543	0,013
25msNois	NaN	NaN	NaN	0,08717	0,025
50msNois	NaN	NaN	NaN	0,06488	0,050
100msNois	NaN	NaN	NaN	0,05823	0,100
200msNois	NaN	NaN	NaN	0,03756	0,200
500msNois	0,03970	0,03893	0,03925	0,02825	0,500
1000msNois	0,03979	0,03273	0,03716	0,02026	1,000
Hulot_kunstkopf_Probe_Leise	0,99902	0,20835	0,77841	0,51221	120,905
Hulot_kunstkopf_Probe_Laut	0,99012	0,32968	0,77506	0,81027	144,007
Kor_Test_Musik	0,83134	0,47028	0,63584	0,56181	1,000
M Layer HOA-Pan 360Grdref	0,67479	0,17477	0,36786	0,15445	20,061
M Layer Intensety-Pan 360grd	0,75974	0,20786	0,39550	0,13914	20,048
M Layer Intesety-Pan MFL to MFR	0,75987	0,23566	0,46009	0,26978	20,048

Tabelle A.8: Verschiedene andere Messungen des Kohärenzgrades Teil 2. Abkürzungen:
 Drschr = Durchschnittswert, g-Kor = Kohärenz über die gesamte File-
 Länge, l(s) = File-Länge in Sekunden.

Dateiname/Stimulus	Werte gefensterter Kohärenz			g-Kor	l(s)
	Maxima	Minima	Drschr		
WRbbg_200-400_1Okt	0,51833	0,24408	0,31421	0,06286	10,000
WRbbg_200-300_1Quinte	0,64732	0,28752	0,40626	0,08076	10,000
WRbbg_300-400_1Quarte	0,57532	0,25429	0,39558	0,09221	10,000
WRbbg_500-600_1klTerz	0,52280	0,24330	0,38533	0,08195	10,000
WRbbg_800-900_1Ton	0,54660	0,27602	0,38557	0,07832	10,000
WRbbg_4000-4240_1Halbton	0,47907	0,20940	0,29236	0,05975	10,000
Sinuston_213_749	0,00329	0,00366	0,00416	0,00006	10,000
weis_No_kor_Test	0,03077	0,03717	0,04778	0,00408	30,200
kor_test_weis_1	0,05467	0,03064	0,03718	0,00933	120,000
kor_test_weis_2	0,05023	0,03036	0,03724	0,00522	120,000
kor_test_weis_3	0,05003	0,03073	0,03715	0,01943	120,000
kor_test_weis_4	0,05133	0,03085	0,03734	0,00467	120,000
kor_test_weis_5	0,05207	0,03011	0,03725	0,00494	120,000
kor_test_weis_6	0,05192	0,03072	0,03715	0,00467	120,000
kor_test_weis_7	0,05236	0,03054	0,03709	0,00531	120,000
kor_test_weis_8	0,04888	0,03028	0,03713	0,00461	120,000
kor_test_weis_9	0,05185	0,02953	0,03720	0,00391	120,000
kor_test_weis_10	0,05374	0,03116	0,03713	0,00431	120,000
kor_test_weis_11	0,05011	0,03064	0,03723	0,00335	120,000
kor_test_weis_12	0,05197	0,03072	0,03707	0,00486	120,000
kor_test_weis_13	0,05276	0,03142	0,03729	0,00515	120,000
kor_test_weis_14	0,04977	0,03158	0,03708	0,00493	120,000
kor_test_weis_15	0,05087	0,03088	0,03714	0,00512	120,000
kor_test_weis_16	0,04943	0,03088	0,03716	0,00961	120,000
kor_test_weis_17	0,05667	0,03064	0,03733	0,01689	120,000
Real_test_dif	0,09736	0,00944	0,03737	0,01280	25,550

A.2 Quellcode des Matlabskript zur Analyse der interauralen Kohärenz

18.08.16 20:26 D:\Dateien\Dokumente\MA...\Auswertung Interaurale Korrelation.m 1 of 2

```

Pfad = 'D:\Dateien\Desktop\Masterthesies\Messungen Korelationagrad 3D\' ;%Pfad des
Ordners mit den auszuwertenden Dateien muss als einziges Angepasst werden
%
Save_Pfad = [Pfad 'Auswertung\']; % Ort an dem die Auswertung gespeichert wird
mkdir(Save_Pfad); %erstellen des Ordners für die Auswertung
f = dir([Pfad '*.wav']); % Suche der Wav-Dateien die Ausgewertet werden sollen im
vorgegebenen Ordner

for i = 1:length(f) % Schleife um alle Dateien nacheinander auszuwerten

    filename = f(i,1).name; % Definition des Dateinamen der Auszuwertenden Datei
    disp(['Auswertung der Datei ' filename]) % Ausgabe welche Datei Ausgewertet wird
    File = regexprep(filename, '.wav', ''); %Name der zu ladenden WAV-Datei zur Berechnung der
X-Korelation
    Name_file =[Pfad File '.wav'];%Pfad der Probe

    [Probe,Fs]=wavread(Name_file);% Einlesen der Probe
    h=wavread(Name_file,'size');% Länge der Probe in Samplen bestimmen
    time_Probe=h(1)/Fs;% Anzeige wie lang die Probe insgesamt in sec ist
    res=5;%Auflösung der Messfenster zur Abtastung in sec/res; 5 für 200ms
    k_step=floor(Fs/2/res);%Schrittweite der Fenster mit 2 so eingestellt,
    %dass sich die Fenster 50% überdecken
    k_max=(floor(h(1)-Fs/res));% Abbruchbedingung für Schleife
    k_diff=floor(Fs/res);% Breite des Messfensters in Sample
    kor=zeros(ceil((2*res*h(1)/Fs)-2),1);%Vorbelegung der Matrix mit Nullen
    n=1;%Zähloperator um die Matrix für die Auswertung Feldweise mit der Lösung
    %zu beschreiben
    for k=1:k_step:k_max; %Die Schleife mit der die Korrelation Schrittweise mit dem
vorgegebenem Zeitfenster berechnet wird
        kor_win = xcorr(Probe(k:(k+k_diff),1),Probe(k:(k+k_diff),2),'coeff');
        kor(n,1)=(max(abs(kor_win)));
        n=n+1;
    end;
    kor_all = xcorr(Probe(:,1),Probe(:,2),'coeff');% Korrelation über die gesamte Länge
als Vergleich

%Auswertung

Maxima=max(kor);
Minima=min(kor);
Durchschn=mean(kor);
St_Abweich=std(kor);
Max_Kor_Alles=max(abs(kor_all));
Durchschn_Alles=mean(kor_all);
Abweich_Alles=std(kor_all);

Save_Data_to = [Save_Pfad File '.txt'];% Speicherpfad der Datei in der die Variablen
gespeichert werden sollen
% Datei erstellen
fid = fopen(Save_Data_to, 'w');
% Schreiben des Auswertungstextes in Die Datei:
fprintf(fid, '%s\r\n\r\n', File);
fprintf(fid, '%s%g sec \r\n\r\n', 'Länge der Probe = ',time_Probe);
fprintf(fid, '%s\r\n', 'Werte gefensterte Korelation');
fprintf(fid, '%s%g\r\n', 'Maximale Korelation = ',Maxima);

```

A Material

18.08.16 20:26 D:\Dateien\Dokumente\MA...\Auswertung Interaurale Korrelation.m 2 of 2

```
fprintf(fid, '%s%g\r\n', 'Minimale Korelation =', Minima);
fprintf(fid, '%s%g\r\n', 'Durchschnittliche Korelation =', Durchschn);
fprintf(fid, '%s%g\r\n', 'Standardabweichung Korelation =', St_Abweich);
fprintf(fid, '\r\n%s\r\n', 'Werte Korelation über Alles');
fprintf(fid, '%s%g\r\n', 'Maximale Korelation über alles =', Max_Kor_Alles);
fprintf(fid, '%s%g\r\n', 'Durchschnittliche Korelation über alles =', Durchschn_Alles);
fprintf(fid, '\r\n%s\r\n', 'Standardabweichung Korelation über alles =', Abweich_Alles);
fprintf(fid, '\r\n');
fclose(fid);% Datei schliessen

Save Data to = [Save_Pfad File '_bl.txt'];
fid2 = fopen(Save_Data_to, 'w');
% Schreiben des Auswertungstextes ohne Text:
fprintf(fid2, '%g %g %g %g %g %g %g %g', time_Probe, Maxima, Minima, Durchschn, St_Abweich,
Max_Kor_Alles, Durchschn_Alles, Abweich_Alles);
% Datei schliessen
fclose(fid2);

% Grafik erstellen und Speichern
figure(1)
subplot(2,1,1)
plot(kor)
title('Kreuzkorelation über die Zeit')
subplot(2,1,2)
plot(Probe)
title('Waveform')
% exportfig ist eine Erweiterung die installiert werden muss war unter Win
% 10 nicht Lauffähig
%export_fig ([Save_Pfad File '_bl'], '-jpg')
saveas(1, [Save_Pfad File '_bl.pdf'], 'pdf')
saveas(1, [Save_Pfad File '_bl.jpg'], 'jpg')

figure(2)
plot(abs(kor_all))
title('Kreuzkorelation über alles ')
%export_fig ([Save_Pfad File '_alles_bl'], '-jpg')
saveas(2, [Save_Pfad File '_alles.pdf'], 'pdf')
saveas(2, [Save_Pfad File '_alles.jpg'], 'jpg')

end
```

A.3 Protokoll zur Hörwahrnehmung bei den Messungen der interauralen Kohärenz

Nach dem die Messungen mit dem Kunstkopf durchgeführt worden sind, wurde dieser zur Seite geräumt. Die Stimuli wurden erneut abgespielt und diesmal von mir angehört. Dabei stellte ich folgendes fest: Schon ab 2 Rauschquellen war eine genaue Lokalisierung schwierig. Durch Kopf-Bewegungen war es aber noch gut möglich, beide Hörereignisse voneinander zu differenzieren. Bei drei und vier Schallquellen war die Differenzierbarkeit stark von der Lage abhängig und nur durch leichte Kopfbewegung möglich. Mitte Front Center war gut von Mitte Side Left und Mitte Side Right zu unterscheiden, allerdings wirkten die Hörereignisse schon sehr diffus und breit. Bei mehr als vier Schallquellen war, wenn sie gleichmäßig um mich verteilt waren, nur noch etwas Diffuses zu hören – als Hörereigniss weder zählbar noch lokalisierbar. Insbesondere bei Nutzung der Höhen- und MittenLautsprecher entstand bereits ein Gefühl starker Umhülltheit, dies wurde durch die Bodenlautsprecher gefühlt kaum besser. Bei Verteilung hauptsächlich in einer Hemisphäre war nur wahrnehmbar, dass viele nicht zählbare Quellen in dieser Hemisphäre lagen. Bei korrelierten Rauschquellen entstand der Eindruck, dass ein Hörereignis zu hören war, das aber nicht lokalisierbar war. Einige Frequenzen scheinen überbetont, eine deutliche Klangfärbung ist wahrzunehmen. Bei unkorrelierten Rausch-Signalen kann es zu der Wahrnehmung kommen, dass sehr viele nicht zählbare und in ihrer Richtung nicht bestimmbare Hörereignisse wahrnehmbar waren.

A.4 DVD

Auf der beiliegenden DVD befinden sich alle ausgewerteten Wav-Dateien, das Nuendo-Projekt der Messungen, das Matlab-Skript und alle damit erzeugten Datensätze und Grafiken. Weiter ist dort eine OpenOffice-Calc-Tabelle mit der kompletten Datensammlung sowie eine Kopie als Excel-Tabelle, das Protokoll zur Wahrnehmung des Autor beim anhören der Messstimuli und eine digitale Kopie dieser Arbeit zu finden.

B Abkürzungsverzeichnis

- BILD** binaural intellegibility level difference dt. binauralen Maskierungsleveldifferenzen zur Sprachverständlichkeit
- BMLD** binaural masking level difference dt. binaurale Maskierungsleveldifferenz
- HOA** Higer Order Ambisonic dt. Ambisonic höherer Ordnung
- HRTF** head relatet tranfer function, dt. Kopf bezogenen Transferfunktion
- IID** interaural intensety diference, dt. Interaurale Intensitätendifferenz
- IKL** in Kopf Lokalisation
- ITD** interaural time difference, dt. interaurale Laufzeitdifferenz
- WFS** Wellenfeldsynthese
- ZNS** zentrale Nervensystem, das

C Glossar

afferent	von lat. affere: hintragen, zuführen, siehe auch Afferenz.
Afferenz	Fortsätze von Nervenzellen, über die einem bestimmten Bereich in Gehirn Signale zufließen.
auditorischer Kortex	Hörrinde (Hörkortex) des Großhirns.
binaural	in beiden Ohren verursacht oder beide Ohren betreffend.
Cochlea	Gehörschnecke des Innenohres.
Corti-Organ	Haarsinneszellen mit Stützzellen auf der Basilarmembran in der Cochlea.
crossmodal	Bezeichnet die Information eines oder mehrerer kortextfremder Reize die unidirektional in den jeweiligen Kortex gelangen. Die Reize werden dort verarbeitet ohne, dass eine Interaktion zwischen den Kortexen unterschiedlicher Sinnen besteht.
efferent	von lat. effere: hinaustragen, hinausführen, siehe auch Efferenz.
Efferenz	Nervenfasern, über die Signale von Nervenzellen fort und an andere Zellen weitergeleitet werden.
Elongation	Auslenkung auf einer Ebene.
Hörereignis	Beschreibt die auditive Wahrnehmung einer Person. Dabei gibt es keinen zwingenden Zusammenhang zwischen einer Quelle im Raum oder den Schallereignissen und einem Hörereignis. Siehe dazu auch Kapitel 2.1, S. 14.
interaural	In Relation der beiden Ohren zueinander.
komplexe Lokalisation	Räumlich exakte Lokalisation durch Rückkopplungsprozesse und multimodale Verarbeitung in den Kortexen. Siehe dazu auch Kapitel 4.4.4, S. 77.

lateral	Seitwärts zur Hauptrichtung des Körpers; in Richtung Seitenrand des Gehirns.
Lateration	Die Bestimmung der Auslenkung eines möglichen Hörereignisses zu den Seiten des Kopfes hin auf Basis der Auswertung der interauralen Differenz. Da der „ <i>cone of confusion</i> “ bei dieser Art der Lokalisation noch nicht aufgelöst ist, kann keine Raumlage bestimmt werden, lediglich eine seitliche Auslenkung. Diese Art der Lokalisation kann auch primitive Lokalisation genannt werden.
Medianebene	orthogonal zur Achse zwischen den Ohren verlaufende Ebene. Siehe dazu Abbildung 2.1, S. 15.
multimodal	Bezeichnet den multidirektionalen Informationsfluss zwischen den Kortexen verschiedener Sinne. Dabei besteht eine Interaktion zwischen den Kortexen.
Phantomschallquelle	Die Wahrnehmung eines Hörereignisses zwischen mindestens zwei Schallquellen, dass entsteht, wenn alle Quellen dasselbe oder ein sehr ähnliches Signal abgeben. Durch Pegel- und/oder Zeitdifferenzen kann das Hörereignis zwischen den Quellen ausgelenkt werden. Siehe dazu auch Kapitel 35, S. 35.
primitive Lokalisation	Art der Lokalisation bei der die Sektorlage eines Hörereignisses nicht bestimmt werden kann. Ihre Lage kann nur auf dem „ <i>cone of confusion</i> “ verortet werden. Siehe dazu auch Kapitel 4.4.1, S. 76.
Schallereignis	Die Summe der Schallenergie einer oder mehrerer, im Raum abstrahlender Schallquellen in jeweils einem der beiden Ohren. Siehe dazu auch Kapitel 2.1, S. 14.
Schallquelle	Ein Schall-abstrahlender Körper im Raum.
Semantik	Ein Zeichen das für eine Bestimmte Bedeutung oder auch ein Muster, das einen komplexen Zusammenhang beschreiben kann, steht. Der Datenumfang des Zeichen ist in der Regel wesentlich kleiner als der Zusammenhang auf den es verweist.
tonotop	siehe Tonotopie.

Tonotopie

Die Abbildung von Frequenzen eines Schallereignisses an bestimmten Stellen auf der Basilarmembran. Im weiteren Sinn bezeichnet sie die Abbildung der Frequenzwahrnehmung über Pegel und Lateration.

D Danksagung

An dieser Stelle möchte ich Thomas Görne und Johann-Markus Batke danken, dass sie sich bereit erklärt haben diese umfangreiche Arbeit zu prüfen und mich bei der Erstellung unterstützten.

Weiter danke ich herzlich: Eberhard Falke, der sehr dazu beigetragen hat, dass die Auswirkungen meiner Legasthenie nicht allzu deutlich zu bemerken sind. Kira Raguse, die einige schwache und falsche Formulierungen und argumentative Fehler in der Arbeit aufdeckte. Ines Prokoph die mir bei der orthopädischen Literaturrecherche hilfreich war. Adrew Lightsey, der mir bei der Übersetzung der Zusammenfassung ins Englische behilflich war. Und meinen Freunden, die mir die, für diese Arbeit, essenzielle Espressomaschine schenkten.

Zu letzte möchte ich Jesus meinem gutem Freund und Herrn danken, der mich geduldig begleitet und mir auch für diese Arbeit immer wieder gute Einfälle und die nötige Ausdauer schenkte.

Abbildungsverzeichnis

1.1	Anatomische Darstellung des Ohres	8
1.2	Die Cochlea	10
1.3	Details der Cochlea	10
1.4	Das auditorische System	12
2.1	Kopfbezogenes Koordiantensystem	15
2.2	Lokalisationsschärfe und Frequenz	16
2.3	Frequenzabhängigkeit der Lokalisationsschärfe	18
2.4	Auslenkung des Hörereignisses	21
2.5	Lokalisationsunschärfe in der Medianebene	23
2.6	Cone of confusion	27
2.7	Abhängigkeit der BMLD	34
2.8	Summenlokalisationskurven für IID und ITD	36
2.9	Seitliche Summenlokalisierung	38
2.10	Laterationsunschärfe und Lautstärkeunterscheidungsschwelle	39
3.1	HOA-Mischplatz vorne	42
3.2	HOA-Mischplatz hinten	42
3.3	Normierten Kreuzkorelation	43
3.4	Interaurale Kohärenzgrade von Rauschen	46
3.5	Kor_{inau} bei unterschiedlichen Fenstertlängen	47
3.6	Kor_{inau} bei unterschiedlichen Filelängen	48
3.7	Interaurale Kohärenz in Abhängigkeit von der Lautsprecheranzahl	52
3.8	Interaurale Kohärenz in Abhängigkeit von der Quellenlage	53
3.9	Laterationsunschärfe als Funktion des Kohärenzgrades	55
3.10	Interaurale Kohärenz zur Lautsprecheranzahl bei inkoh.Rauschs.	56
3.11	Werte für die interaurale Kohärenz über die Zeit	58
4.1	Die drei wichtigsten auditorischen Areale	61
4.2	Was- und Wo-Ströme	61
4.3	Theatermodell zur Theorie des globalen Arbeitsraums	64
4.4	Melodische Gruppierung	73
4.5	Grenzen der melodische Gruppierung	74
5.1	Richtungsabhängige Lokalisationsunschärfe im Kino	87

Tabellenverzeichnis

A.1	Materialliste	96
A.2	Benennung der Lautsprecher	97
A.3	Lautsprecher-Positionen	97
A.4	Lautsprecher Anschlusskonfiguration	98
A.5	Messungen des interauralen Kohärenzgrades mit korrelierten Signalen	98
A.6	Messungen des interauralen Kohärenzgrades mit unkorrelierten Signalen	99
A.7	Verschiedene andere Messungen des Kohärenzgrades 1.	100
A.8	Verschiedene andere Messungen des Kohärenzgrades 2.	101

Literaturverzeichnis

- Baars, Bernard J.: „In the Theater of Consciousness: Global Workspace Theory, A Rigorous Scientific Theory of Consciousness.“, in: *Journal of Consciousness Studies* 4, Nr.4, S. 292-309, 1997
- Blauert, Jens: *Räumliches Hören*, 1. Auflage, S.Hirzel Verlag Stuttgart, 1974
- Blauert, Jens: *Spatial Hearing: The Psychophysics of Human Sound Localization*, 1. Auflage S. 288-422, MIT Press, 1997
- Blauert, Jens & Braasch, Jonas: „Räumliches Hören“, in: Weinzierl, S. (Hrsg.): *Handbuch der Audiotechnik*, 9. Auflage, S. 87-122, Springer, 2008
- Bregman, Albert S. : *Auditory scene analysis: the perceptual organization of sound*, 1. Auflage, S. 1-45, S.641-704, MIT Press, 1994
- Celesia, G.G. & Hickok, G.): *The Human Auditory System: Fundamental Organization and Clinical Disorders*, 1. Auflage, S. 1-116, Elsevier Science, 2015
- Deike, Susann: *FMRT-Studien am auditorischen Kortex des Menschen bei der auditiven Wahrnehmung in Cocktail-Party-Situationen*, S. 1-21, 2010
- Dolby Laboratories: „Dolby 5.1 Channel Production Guidelines“, 2000
- Ellermeier, W. & Hellbrück, J.: „Hören – Psychoakustik – Audiologie.“, in: Weinzierl, S. (Hrsg.): *Handbuch der Audiotechnik*, 9. Auflage, S. 42-85, Springer, 2008
- Ehret, Günter: *Zentrales Hörsystem (Übersicht) Wie hört das Gehirn? – Neurobiologie des Hörsystems*, <https://www.uni-ulm.de/nawi/nawi-neurobio/arbeitsgruppen/prof-dr-guenter-ehret/zentrales-hoersystem.html?print=1>, 2007, letzter Zugriff: 14. 06. 2016
- Falke, Tobias & Batke, Johann-Markus & Görne, Thomas: „Remixing a Historic Film in Higher Order Ambisonics 3D Audio – Workflow and Technical Solutions “, *136th Audio Engineering Society Convention e-Brief*, 2014
- Falke, Tobias: „Wie viel 3D braucht unser Gehirn – über die Schaffung realitätsnaher Höreindrücke “, *28th Tonmeistertagung – VDT international Convention* , November 2014

Literaturverzeichnis

- Fricke, Jobst P.: *Eine Konsonanztheorie auf der Grundlage von Autokorrelation unter Berücksichtigung der Unschärfe*, <http://www.konsonanztheorie.de/>, 2007, letzter Zugriff: 12. 08. 2016
- Günther, R. & Jantsch, H.: *Physikalische Medizin*, 2. Auflage, S. 25, Springer-Verlag, 2013
- Goldstein, E. Bruce: „Sensation and Perception, 7th edition,“, in der Übersetzung: Plata, Goido (Übersetzung aus dem Amerikanischen) & Irtel, Hans (Hrsg.): *Wahrnehmungspsychologie Der Grundkurs*, 7. Auflage, S. 21-136, S. 267-332, Springer-Verlag Berlin Heidelberg, 2008
- Görne, Thomas: *Tontechnik*, 2. Auflage, S. 118-121, Hanser, 2008
- Human Brain Project: *Human Brain Project*, <https://www.humanbrainproject.eu/>, 2013, letzter Zugriff: 14. 09. 2016
- Leifert-Fiebach, Gundhild: *Auditorischer Neglekt und auditorische Extinktion bei Patienten mit visuell räumlichen Neglekt*, S. 1-21, 2008
- Mikunda, Christian: *Kino spüren Strategien der emotionalen Filmgestaltung*, 1. Auflage, S. 193-198, WUV Universitätsverlag, 2002
- Rozenn, Nicol: „3D audio binaural: what do you mean by “individual“?“, 2013
- Schmidt, Ulrich : *Professionelle Videotechnik*, 4. Ausgabe, S. 18-21, Springer-Verlag Berlin Heidelberg, 2005
- Seidl, Armin H.: *Entwicklung und erfahrungsabhängige Plastizität neuronaler Mechanismen für Schalllokalisation bei Säugern*, S. 1-22, 2003
- Spieß, Ralf : „Unbewusste Informationsverarbeitung Forschungsansätze, Ergebnisse und methodische Probleme unter besonderer Berücksichtigung des akustischen Bereichs“, in: *Studienreihe Psychologische Forschungsergebnisse Band 88*, S. 1-12, S. 70-104, Verlag Dr. Kovač, 2002
- Theile, Günther & Felderhoff, Uwe & Mackensen, Philip: „Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale“, *TMT*, 1998
- Jambrosic, Kristian & Ivancevic, Bojan & Petosic, Antonio: „Referent sound Source localisation in multisource environment for acoustically different spaces“, *ForumAcusticum*, S. 1619-1623, 2005
- Volkman, Alfred Wilhelm : *Die Drehbewegungen des Körpers: (Abdruck aus Virchows Archiv für Pathologische Anatomie Bd LVI.)*, Ausgabe 4, Buchdruckerei d. Waisenhauses, 1872

Ich versichere, die vorliegende Arbeit selbstständig ohne fremde Hilfe verfasst und keine anderen Quellen und Hilfsmittel als die angegebenen benutzt zu haben. Die aus anderen Werken wörtlich entnommenen Stellen oder dem Sinn nach entlehnten Passagen sind durch Quellenangaben eindeutig kenntlich gemacht.

Ort, Datum

Tobias Falke