

# **Zur Wahrnehmbarkeit von Diskretisierungen in der dynamischen Binauralsynthese**

## *(On the perceptibility of discretisation in dynamic binaural synthesis)*

*Alexander Lindau\*, Stefan Weinzierl\*\**

\* TU Berlin, Fachgebiet Audiokommunikation, alexander.lindau@tu-berlin.de

\*\* TU Berlin, Fachgebiet Audiokommunikation, stefan.weinzierl@tu-berlin.de

### **Kurzfassung**

Eine auf der Auralisation binauraler Raumimpulsantworten (BRIRs) beruhende dynamische Binauralsynthese kann eine Nachführung des Schallfelds auf Kopfbewegungen des Hörers in allen Freiheitsgraden leisten. Um die Hörbarkeit der bei der Messung der Impulsantworten stets vorgenommenen, räumlichen Diskretisierung zu untersuchen, wurden in zwei Experimenten die Schwellwerte der eben wahrnehmbaren Auflösung des Gridrasters in den drei rotatorischen Freiheitsgraden der Kopfbewegung mittels eines adaptiven 3AFC-Verfahrens ermittelt. Die Versuche wurden mit verschiedenen Audioinhalten und mit binauralen Datensätzen aus verschiedenen akustischen Umgebungen durchgeführt. Die Ergebnisse der Untersuchung liefern Anhaltspunkte für die, für eine plausible binaurale Simulationen von akustischen Umgebungen erforderliche, räumliche Auflösung der Orientierungen des Empfängers.

### **1. Motivation**

Die Simulationen akustischer Umgebungen durch dynamische Binauralsynthese auf der Basis gemessener binauraler Raumimpulsantworten (Binaural Room Impulse Responses, BRIRs) können ein hohes Maß an perzeptiver Plausibilität erreichen [1]. Dabei stellt die realistische Interaktion zwischen den Kopfbewegungen des Hörers und dem synthetisierten Schallfeld einen wesentlichen Qualitätsparameter dar. So ist es wesentlich, ob die dynamische Auralisation den drei rotatorischen Freiheitsgraden der Kopfbewegung (horizontal, vertikal, lateral, s. Abb. 1) folgt und wie fein die räumliche Auflösung der BRIR-Datensätze bzgl. dieser Freiheitsgrade ist. Je feiner die Datensätze aufgelöst sind, desto höher ist allerdings der Aufwand bei der Auralisation ebenso wie die Messdauer der zu resynthetisierenden akustischen Szene. Aus diesem Grund wurden auditive Schwellwertuntersuchungen im Hinblick auf die eben wahrnehmbare Diskretisierung von BRIR-Datensätzen durchgeführt, auf deren Grundlage sich der Aufwand für die Messung und Simulation von akustischen Szenen bei gleichzeitiger Wahrung der perzeptiven Plausibilität optimieren lässt.

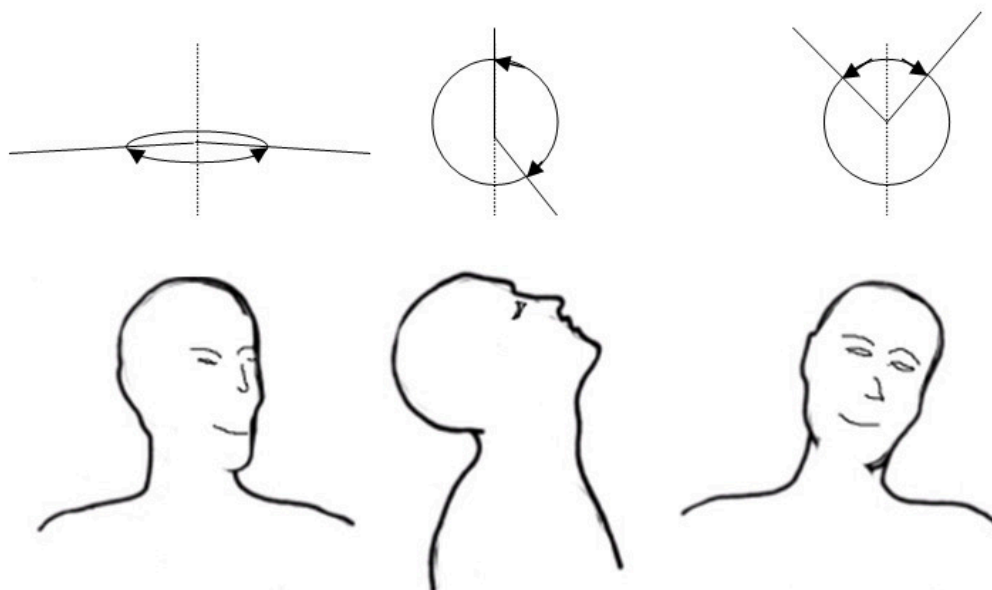


Abb. 1: Rotatorische Freiheitsgrade der Kopfbewegung, v.l.n.r.: horizontale, vertikale und laterale Rotation, sowie hörtrypische Bewegungsbereiche (schematisch)

## 2. Stand der Forschung

Das räumliche Auflösungsvermögen des auditorischen Systems wird in der Literatur durch Werte wie den kleinsten hörbaren Winkelunterschied (Minimum Audible Angle, MAA) oder, für bewegte Hörer oder Quellen, durch den kleinsten hörbaren Bewegungswinkel (Minimum Audible Movement Angle, MAMA) beschrieben. In nachhallfreier Umgebung wurden gerade hörbare Winkelunterschiede von bis zu  $1^\circ$  für frontalen Schalleinfall gefunden [2], [3]. Der im Hinblick auf Kopfbewegungen relevante Schwellwert des kleinsten hörbaren Winkelunterschieds bei bewegter Quelle (MAMA) liegt zwischen  $5^\circ$  und  $20^\circ$ , abhängig von der Geschwindigkeit der Quelle und der Bandbreite des Signals [4]. Bei einer systematischen Untersuchung zur Hörbarkeit der Auflösung von HRTF-Datensätzen wurde, ausgehend von einem Ursprungsdatensatz mit  $2^\circ$ -Auflösung (horizontal und vertikal), die Hörbarkeit stufenweise auflösungsreduzierter und dann wieder auf das  $2^\circ$ -Raster linear interpolierter Datensätze für statische und bewegte Quellen in der Frontal-, Horizontal- und Medianebene untersucht. Im statischen Falle waren richtungsabhängig Auflösungsreduktionen von  $2^\circ$ - $36^\circ$  hörbar, im bewegten Falle  $8^\circ$ - $90^\circ$ . Insgesamt erschien somit eine erhebliche Reduzierung des ursprünglich mit  $2^\circ$ -Auflösung vermessenen Datensatzes möglich [5]. Da bei realen Räumen jedoch ständig Reflexionen aus allen Einfallrichtungen vorliegen, können diese Daten nicht ohne weiteres auf den Fall einer Simulation anhand von BRIRs übertragen werden.

Die binaurale Simulation von akustischen Umgebungen findet heute in verschiedenen Anwendungen Eingang. So bieten etwa Programmpakete zur numerischen Raumakustiksimulation (EASE [6], ODEON [7]) Auralisationsmodule zur auditiven Überprüfung raumakustischer bzw. elektroakustischer Entwürfe an, wobei die Berechnung binauraler Raumimpulsantworten offline, zumeist für einige wenige Hörerpositionen und jeweils nur für eine gewählte Blickrichtung geschieht. Die Auralisation erfolgt dann durch Faltung nachhallfreier Audiodaten mit den berechneten Impulsantworten ohne Berücksichtigung von

Kopfbewegungen. Interaktive Auralisationssysteme, die häufig unter den Bezeichnungen 3D- oder Spatial-Audio System, Virtual Auditory Display (VAD) oder Virtual Acoustic Environment (VAE) firmieren, basieren in der Regel auf - häufig öffentlich zugänglichen - Datenbanken von Außenohrübertragungsfunktionen (Head Related Transfer Functions, HRTFs) mit unterschiedlichen Winkelauflösungen. Neueste Systeme leisten sogar eine echtzeitfähige physikalische Simulation der Raumakustik und berechnen, oft unter Ausnutzung auditorischer Wahrnehmungsgrenzen, BRIRs 'on the fly'.

System	Auralisation	Auflösung horizontal / vertikal	Quelle
EASE 4	physik. modelliert, offline, statisch	5-30° / 10°	[8]
ODEON 9	physik. modelliert, offline, statisch	5° / 5,6°	[9]
Raven	physik. modelliert, Echtzeit, dynamisch	1-2° / 1-2°	[10]
IKA-SIM	physik. modelliert, Echtzeit, dynamisch	15° / 10° (5° interp.)	[11]
DIVA	physik. modelliert, Echtzeit, dynamisch	10° / 15° (1° interp.)	[12]
SLAB	reale Messung, Echtzeit, dynamisch	10° / 10°	[13]
SSR	reale Messung, Echtzeit, dynamisch	5° (1° interp.)	[14]
BRS	reale Messung, Echtzeit, dynamisch	6° (1° interp.)	[15]

Tab. 1: Auflösungen der Datensätze von HRTFs bzw. BRIRs in aktuellen Auralisationssystemen

Derzeit eignet sich nur das System der TU Berlin [1] für eine automatisierte Vermessung von binauralen Raumimpulsantworten in allen Freiheitsgraden der Kopfbewegung. Marktgängige Kunstkopfsysteme leisten nur eine Kopfbewegung in der Horizontalebene [15], [16] oder verwenden HRTFs, die mit einem als Ganzes bewegten Kunstkopf ohne Torso im reflexionsarmen Raum vermessen wurden. Die Wirkung lateraler Kopfbewegungen über einem feststehenden Torso wurde bisher mit keinem System untersucht.

Die horizontalen und vertikalen Auflösungen der HRTF- bzw. BRIR-Datensätze aktueller Auralisationssysteme sind sehr unterschiedlich. Häufig werden Datensätze zunächst mit niedrigerer Auflösung vermessen, um sie dann auf eine feinere Auflösung zu interpolieren [11], [14], [15], [17]. Bei einer systematischen Untersuchung zur Auflösung von HRTF-Datensätzen [5] erwies sich die richtungsabhängige Reduzierung eines ursprünglich mit 2°-Auflösung vermessenen und anschließend durch Interpolation auf die gleiche Auflösung rekonstruierten Datensatzes auf ca. ein Zehntel möglich. Inwieweit diese Ergebnisse auf den Fall realer Räume übertragen werden können, wo zusätzliche Raumreflexionen die Wahrnehmbarkeit von durch Kopfdrehungen bedingten Änderungen des Schallfelds beeinflussen, soll durch die vorliegende Untersuchung ermittelt werden. Hierbei wurden Datensätze aus nachhallfreier Umgebung sowie aus drei Räumen mit unterschiedlichem Volumen und unterschiedlicher Nachhallzeit (siehe Tabelle 2) verwendet. Darüber hinaus wurde der Einfluss verschiedener Stimuli (rosa Rauschen, akustische Gitarre) untersucht.

### 3. Untersuchungsmethodik

#### 3.1. BRIR-Datensätze

Zur Bestimmung der für eine artefaktfreie binaurale Simulation erforderlichen Auflösung von binauralen Impulsantworten wurden Datensätze im reflexionsarmen Raum (RAR), in einem Tonstudio und in zwei Hörsälen der TU Berlin für jeweils eine frontale Schallquelle vermessen (s. Tabelle 2). Für die Messungen wurde der am Fachgebiet Audiokommunikation entwickelte Kunstkopf- und Torsosimulator FABIAN benutzt [1], dessen Konzeption es erlaubt, binaurale Raumimpulsantworten auch für laterale Kopfrotationen über dem Torso zu vermessen.



Abb. 2: FABIAN im Hörsaal H 104, Aufbau zur Messung mit für laterale Kopfrotationen. Links: Detailaufnahme des Halsgelenks, rechts: Zustand während der Messung

Im RAR wurde FABIAN im Fernfeld der Quelle positioniert, in allen anderen Räumen wurde er an Sitzplätzen positioniert, die etwa dem 2- bis 3-fachen Hallradius entsprachen, so dass von einem überwiegend diffusen Schallfeld ausgegangen werden kann. In allen Fällen wurde derselbe aktive 2-Weg-Lautsprecher vom Typ Meyersound UPL-1 als Anregungsquelle benutzt.

Raum	Volumen	Nachhallzeit	Hallradius	Abstand	Datensätze hor / ver / lat
Reflexionsarmer Raum	1800 m <sup>3</sup>	$f_u = 60\text{Hz}$	-	3 m	$\pm 80^\circ / \pm 35^\circ / \pm 60^\circ$
Tonstudio	235 m <sup>3</sup>	0,36 s	1,4 m	2,8 m	$\pm 80^\circ / \pm 35^\circ / \pm 60^\circ$
Hörsaal H 105	8600 m <sup>3</sup>	2,1 s	3,6 m	7,5 m	$\pm 80^\circ / \pm 35^\circ / 0^\circ$
Hörsaal H 104	3000 m <sup>3</sup>	0,95 s	3,2 m	10 m	$0^\circ / 0^\circ / \pm 60^\circ$

Tab. 2: Raumakustische und messtechnische Details der erstellten binauralen Datensätze

Die Datensätze wurden mit einer als unhörbar fein angenommenen Auflösung von  $1^\circ$  [5] und für physiologisch motivierte, maximale Drehwinkel des Kopfes von horizontal  $\pm 80^\circ$ , vertikal  $\pm 35^\circ$  und lateral  $\pm 60^\circ$  vermessen. Für die Untersuchung horizontaler und vertikaler Kopfbewegungen (Experiment I) wurden je Raum 11.431 BRIRs, für die Untersuchung lateraler Bewegungen (Experiment II) entsprechend 121 BRIRs vermessen. Bei der Untersuchung der lateralen Kopfbewegungen wurde trotz des größeren vermessenen Bereichs nur ein Datenbereich von  $\pm 30^\circ$  auralisiert wurde, s. Abschnitt 3.5 und 5. Die

gewählten Winkelbereiche basieren auf typischen Rotationsbereichen beim natürlichen Hören und auf den physiologisch motivierten Komfort- und Maximalbereichen (s. Tabelle 3). Für horizontale und laterale Rotationen sind sowohl die hörtypischen als auch die Komfortbereiche berücksichtigt, für vertikale Rotationen nur die hörtypischen Werte.

	Horizontale Rotation		Vertikale Rotation				Laterale Rotation	
	MW	$\sigma$	„Retroflexion“		„Anteflexion“		MW	$\sigma$
Komfortbereiche [18]	35°	-	50°	-	-45°	-	25°	-
hörtypisch [19]	42°	20,4°	15,2°	12,9°	nur Gesamtangabe		11,6°	8,3°
Maximalbereiche [20]	79°	14°	61°	27°	-60°	12°	41°	7°

Tab. 3: Übersicht über Kopfbewegungsbereiche

### 3.2. Stimuli

In den Experimenten wurden zwei Stimuli verwendet: Rosa Rauschen von 5 s Dauer, das mit 20 ms Fade-In/Out versehen war, sowie ein ebenfalls etwa 5-sekündiger Ausschnitt aus einem Stück für akustische Gitarre (Bourrée von J. S. Bach). Letzteres hatte sich bereits in früheren Hörversuchen als besonders geeignet erwiesen, Artefakte in einer Simulation aufzudecken [1]. Zudem sollte es als typisches, natürliches Musiksignal dienen, da es sowohl transiente als auch quasi-stationäre, harmonische Anteile enthält. Das auch in [9]-[12] verwendete rosa Rauschen wurde dagegen als besonders geeignet erachtet, spektrale Änderungen hörbar zu machen, die als Hauptartefakt bei der adaptiven Veränderung der BRIR-Gridauflösung erwartet wurden. Aufgrund des Bandpasses, der als Zielfunktion der Kopfhörerkompensation benutzt wurde, war die Bandbreite aller Signale auf 50-21 kHz begrenzt.

### 3.3. Auralisation

Zur Auralisation der Datensätze kam eine am Fachgebiet entwickelte Software zur partitionierten Faltung zum Einsatz. Bei einer Abtastrate von 44.1 kHz wurden jeweils nur die ersten  $2^{14}$  Samples jeder Impulsantwort dynamisch aktualisiert, da sich Veränderungen des späteren, diffusen Nachhalls aufgrund von Kopfbewegungen oder einer veränderten Quellenpositionen als nicht hörbar erwiesen haben [1]. Die Blockgröße der schnellen Faltung betrug für den frühen Teil der Impulsantworten 256 Samples, für den späten Nachhall 8192 Samples. Die minimale Latenz von einem Audioblock wird dabei durch die zugrunde liegende Jack-Audioserver-Architektur [www.jackaudio.org] verursacht. Die ersten Resultate eines durch Kopfbewegungen getriggerten Crossfades sind im ungünstigsten Fall nach Ablauf von 2 Audioblöcken hörbar. Die Ergebnisse der Zeitbereichsüberblendung werden blockweise ausgespielt und führen so für die wichtigen frühen Direktschallanteile zu einer minimalen Reaktionszeit der Simulation auf Kopfbewegungen. Die Überblendungsergebnisse späterer Anteile der Impulsantwort entsprechend später im Ausgangssignal hörbar. Die Dauer einer vollständigen Überblendung entspricht der Länge des frühen Teils der Impulsantwort [21].

Um sicherzustellen, dass der Überblendprozess an sich bei der Wiedergabe nicht hörbar war, wurden in einem Vorversuch Sinustöne und rosa Rauschen mit andauernder Überblendung zwischen identischen HRTFs auralisiert. Da dabei keinerlei Diskontinuitäten oder

Schaltvorgänge hörbar waren, kann angenommen werden, dass alle in den Hörversuchen entstandenen Artefakte auf durch die Auflösung der BRIR Datensätze bedingte Unterschiede zurückgehen.

Zur Verfolgung der Kopfbewegungen in allen drei rotatorischen Freiheitsgraden wurde ein Head Tracker mit einer nominalen Updaterate von 120 Hz eingesetzt (Polhemus Fastrack), entsprechend einem Detektionsintervall von 8.6 ms. Untersuchungen zur Hörbarkeit der aus Tracker-Updaterate, Dauer der seriellen Übertragung der Trackingdaten, dem Auslesen des seriellen Ports und der Übermittlung der Daten an die Audioverarbeitung resultierenden Systemlatenz sind derzeit in Planung.

Der Frequenzgang des benutzten Kopfhörer vom Typ STAX Lambda SR202 wurde durch ein linearphasiges 1024-Punkte-Filter kompensiert. Das Kompensationsfilter wurde aus dem mittleren Betragsfrequenzgang von 10 Messungen des jeweils neu auf dem Kunstkopf aufgesetzten Kopfhörers bestimmt und auf der Grundlage eines frequenzabhängigen LMS-Kriteriums optimiert [22]. Zu einer vergleichenden Untersuchung verschiedener Filterentwurfsverfahren in der Binauraltechnik siehe [23]. Da das Hörversuchsverfahren auf dem instantanen Vergleich von voller und reduzierter Datensatzauflösung beruht, musste der gesamte Datensatz (ca. 22 GB) im Arbeitsspeicher verfügbar gehalten werden. Dazu kam ein 8-CPU-Rechner mit Servermainboard, 32 GByte Hauptspeicher und einem Linux-Betriebssystem zum Einsatz.

Um Schaltartefakte beim Wechsel zu einer benachbarten BRIR zu vermeiden, wurde ein linearer Crossfade eingesetzt, dessen Überblenddauer mit 5.6 ms (entsprechend einer Audioblocklänge) so kurz gehalten wurde, dass sie nicht als lineare Interpolation hörbar werden sollte. Damit unterscheidet sich die vorliegende Untersuchung von ähnlichen Studien zur Hörbarkeit abrupter Veränderungen von A) der interauralen Laufzeitunterschiede [24], bzw. B) des minimalphasigen Spektrums [25], [26] beim Umschalten zwischen mehr oder weniger dicht benachbarten HRTFs. Diese Studien ermöglichen Aufschlüsse bezüglich der Wahrnehmbarkeit einer diskreten Stufung der fundamentalen Lokalisationscues; die für eine separate Untersuchung von interauralen Laufzeit- und Pegeldifferenzen notwendige Künstlichkeit der benutzten Stimuli wurde in der hier beschriebenen Studie jedoch bewusst vermieden.

### 3.4. Versuchspersonen

Die Studie wurde anhand zweier zeitlich aufeinander folgender Experimente durchgeführt. Am Experiment I (horizontale und vertikale Kopfbewegung) nahmen 21 Versuchspersonen im Alter von 24-35 Jahren teil (19 männliche, 2 weibliche), am Experiment II (laterale Kopfbewegung) 23 Versuchspersonen im Alter von 23-65 Jahren (20 männliche, 3 weibliche). Alle Versuchspersonen hatten Erfahrung als Probanden in Hörversuchen, der überwiegende Teil verfügte über eine musikalische Ausbildung. Beide Experimente wurden als vollfaktorielle Versuche mit Messwiederholungen durchgeführt, d.h. die Schwellwerte jeder Person wurden unter jeder Bedingungsvariation gemessen, resultierend in  $12 \times 21 = 252$  Schwellwerten in Experiment I und  $6 \times 23 = 138$  Schwellwerten in Experiment II.

### 3.5. Hörversuchsverfahren

Zur Messung der Schwellwerte wurde ein adaptives 3-AFC Verfahren eingesetzt [27]. Dabei wurden jeweils drei Stimuli in ihrer Reihenfolge randomisiert und aufeinander folgend in einer Triade präsentiert. Jede Triade enthielt zweimal einen Referenzstimulus (hier: die binaurale Simulation in 1°-Auflösung) und einmal ein verändertes Beispiel (hier: binaurale Simulation mit vergrößerter Auflösung).

In Experiment I waren sowohl horizontale als auch vertikale Kopfbewegungen erlaubt. Da aber der Schwellwert für die eben wahrnehmbare Auflösung unabhängig für vertikale und horizontale Kopfbewegungen getestet werden sollte, wurde die Auflösung jeweils nur für eine der beiden Kopfbewegungsrichtungen variiert, während die andere weiter mit höchster Auflösung simuliert wurde. Jeder Versuchsdurchlauf begann nach einer Trainingssession mit dem mittleren eines als sinnvoll betrachteten Bereichs von Gridrastrerauflösungen. Ein Maximalwert für die zu testende Auflösungsreduktion wurde bei beiden Experimenten in Vorversuchen bestimmt. Bei Experiment I betrug die größte Schrittweite raum- und richtungsabhängig 20° bzw. 30° (HRTFs, horizontal bzw. vertikal), und 30° bzw. 35° (Studio & Hörsaal H 105, horizontal bzw. vertikal). Die Datensätze in Experiment II enthielten nur BRIRs für laterale Kopfbewegungen bei konstant frontaler Orientierung. Hier betrug die maximale Schrittweite in allen Fällen 30°. Die Kopfbewegungsgeschwindigkeiten wurden in keinem der Versuche beschränkt oder vorgegeben. Die VP wurden stattdessen während des Trainings angewiesen, individuelle Kopfbewegungen zu finden, um ihre Detektionsleistung zu maximieren.

Die Gridrastrerauflösung wurde während eines Versuchs abhängig von den Antworten der VP nach einer Maximum-Likelihood-Adaptionsregel angepasst (Best PEST, [28]). Die kleinstmögliche Auflösungs-Schrittweite betrug 1°. Die Durchläufe wurden nach 10 Trials abgebrochen, woraus sich Versuchsdauern von zwischen 35 Minuten (Exp. II) und 1:30 h (Exp. I) ergaben. Aufgrund der 1°-Auflösung der BRIR-Datensätze und der Tatsache, dass im AFC Verfahren die 1°-Referenz sinnvoll nur gegen höhere Auflösungen verglichen werden kann, betrug die niedrigste messbare Auflösung damit 2°; ein Wert, der in beiden Experimenten nur zwei Mal erreicht wurde.

Die zulässigen Kopfbewegungsbereiche entsprachen, bis auf den Fall lateraler Kopfbewegungen, den in Tabelle 2 dargestellten Winkelbereichen. In Vorversuchen zeigte sich, dass bei einer lateralen Kopfneigung von mehr als 30° deutliche wahrnehmbare Kammfiltereffekte auftraten. Diese vermutlich auf Schulterreflexionen zurückzuführenden Veränderungen der Übertragungsfunktion wiesen dabei eine so starke Positionsabhängigkeit auf, dass auch für die feinste Gridrastrerstufung von 1° eine sichere Detektion möglich war. Aus diesem Grund wurden laterale Kopfbewegungen von >30° im vorliegenden Hörversuch ausgeschlossen. Eine weitergehende Diskussion findet sich in Abschnitt 5.

### 3.6. Versuchsdesign

In den Experimenten wurden die Schwellwerte der wahrnehmbaren Gridrastrerauflösung für horizontale, vertikale und laterale Kopfbewegungen unter den zwei Bedingungsvariationen Stimulus (2-stufig) und Raumtyp (3-stufig) erhoben. Die sich somit ergebenden Faktorstufenkombinationen beider Experimente sind in den Tabellen 4 und 5 dargestellt.

Raum	Stimulus	Auflösungsreduktion
Reflexionsarmer Raum	rosa Rauschen	horizontal
		vertikal
	akustische Gitarre	horizontal
		vertikal
Tonstudio	rosa Rauschen	horizontal
		vertikal
	akustische Gitarre	horizontal
		vertikal
Hörsaal H 105	rosa Rauschen	horizontal
		vertikal
	akustische Gitarre	horizontal
		vertikal

Tab. 4: Die 12 in Experiment I getesteten Bedingungsvariationen

Raum	Stimulus	Auflösungsreduktion
Reflexionsarmer Raum	rosa Rauschen	lateral
	akustische Gitarre	lateral
Tonstudio	rosa Rauschen	lateral
	akustische Gitarre	lateral
Hörsaal H 104	rosa Rauschen	lateral
	akustische Gitarre	lateral

Tab. 5: Die 6 in Experiment II getesteten Bedingungsvariationen

#### 4. Ergebnisse

Abbildung 3 zeigt die unter allen Bedingungsvariationen (Kopfbewegungsrichtungen, Räume, Stimuli) ermittelten Versuchsergebnisse als Verteilungen der gefundenen Schwellwerte über jeweils alle getesteten Versuchspersonen. Da die ermittelten Schwellwertverteilung überwiegend nicht normalverteilt waren (Kolmogoroff-Smirnoff Test,  $\alpha = 0.2$ ), werden zentrale Tendenzen und Streuungen im Folgenden durch Mediane und Perzentile beschrieben. Die Mediane der über alle Versuchspersonen je Bedingungsvariation ermittelten eben hörbaren Gridauflösungen bewegen sich in einem Bereich von  $4^\circ$  bis  $18^\circ$ . Bis auf zwei Ausnahmen (HRTFs, Rauschen, vertikale Kopfbewegung) konnte keine Versuchsperson Gridauflösungen von  $<3^\circ$  detektieren.

Während für den Rauschstimulus in horizontaler und vertikaler Bewegungsrichtung alle Versuchspersonen eine Auflösung von  $> 11^\circ$  zuverlässig detektieren konnten, waren für den Gitarrenstimulus in vertikaler Richtung reduzierte Auflösungen für einige Versuchspersonen selbst im direkten Vergleich zwischen feinsten und größter Auflösung ( $35^\circ$ -Raster) nicht detektierbar. In lateraler Bewegungsrichtung gab es unter *jeder* Bedingungsvariation (Raum, Stimulus) Personen, für die die größte Auflösung ( $30^\circ$ -Raster) nicht detektierbar war. Auch die in Abb. 3 für laterale Rotationen sichtbaren großen Streuungen verdeutlichen das Problem. Diese sind nicht auf eine zu kleine Stichprobe zurückzuführen, ebenso war der



angebotene Bereich der Auflösungen mit  $1^{\circ}$ - $30^{\circ}$  genügend hoch. Eine weitere Vergrößerung des lateralen Kopfbewegungsbereichs wurde aufgrund der in Abschnitt 3.5 beschriebenen Problematik nicht zugelassen. Die gefundenen Messwerte sind aufgrund des gewählten Messverfahrens reliabel, die großen Streuungen sind daher als typisch für die Wahrnehmung der Gridrasterreduktion bei lateralen Kopfbewegungen anzusehen.

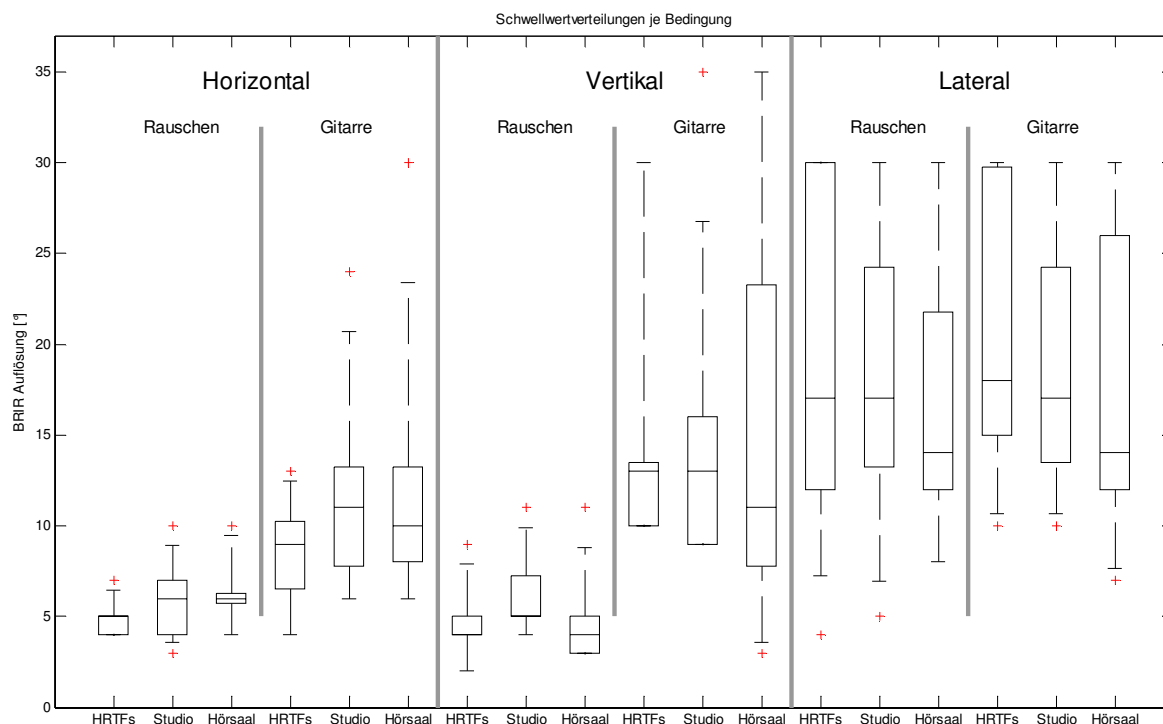


Abb. 3: Eben wahrnehmbare Diskretisierung der BRIR-Datensätze für alle Bedingungen (Mediane, Interquartile, Interdezilbereich, Extremwerte)

Die inferenzstatistische Überprüfung der in Abbildung 3 sichtbaren Trends erfolgte für Experiment 1 mittels einer univariaten  $3 \times 2 \times 2$  bzw. in Experiment II mittels einer univariaten  $3 \times 2$  ANOVA für Messwiederholungen. Die dafür notwendigen Voraussetzungen homogener Varianzen und homogener Korrelationen unter allen Faktorstufen wurden mit einem Sphärizitätstest [29] nach Mauchly überprüft. Nötigenfalls wurden entsprechende Freiheitsgradkorrekturen vorgenommen. Im Experiment I (horizontale und vertikale Bewegungen) zeigen die gefundenen Schwellwerte innerhalb der gesamten Stichprobe eine sehr hohe Reliabilität (Kronbachs Alpha = 0.934). Im Experiment II (laterale Bewegungen) lag sie deutlich niedriger (Kronbachs Alpha = 0.165), ein Indiz für starke individuelle Unterschiede bei der Wahrnehmung der Auflösung für laterale Kopfbewegungen.

Laut den Ergebnissen der ANOVA erschien die Empfindlichkeit für Diskretisierungen bei horizontalen Kopfbewegungen zunächst signifikant höher als bei vertikalen Rotationen. Bei näherer Betrachtung wurde dieser Effekt allerdings hauptsächlich durch die überdurchschnittlich schlechtere Erkennungsleistung bei vertikalen Kopfbewegungen während der Präsentation des Gitarrenstimulus' verursacht. Diese Abhängigkeit der in einer bestimmten Bewegungsrichtung eben wahrnehmbaren Diskretisierung vom Stimulustyp wurde als signifikante Interaktion bestätigt: Demnach wird bei Rauschsignalen eine Auflösungsreduktion in vertikaler Richtung früher wahrgenommen, während beim Gitarrenstimulus eine

horizontale Diskretisierung früher hörbar wird. Dementsprechend und entgegen althergebrachten Erwartungshaltungen wurden die geringsten Schwellwerte für vertikale Kopffrotationen bei Anregung durch das Rauschsignal gefunden.

Eine daraufhin unter den Bedingungen Rauschen für horizontale und vertikale Schwellwerte und für alle 3 Räume durchgeführte 2x3 ANOVA mit Messwiederholungen bestätigte die vertikalen Schwellwerte nun tatsächlich als signifikant niedriger (im Mittel  $5^\circ$  statt  $5.6^\circ$ ). Aufgrund der für frontalen Schalleinfall und vertikale Kopffrotationen geringen Änderungen der interauralen Laufzeitdifferenz (ITD) dürften spektrale Cues bei der Wahrnehmung der Granularität bestimmend sein. Die Erkennungsleistung beim Gitarrenstimulus fällt daher aufgrund dessen geringerer Bandbreite und des weniger stationären Verhaltens ab. In Kombination mit den bei horizontalen Kopfbewegungen induzierten ITD-Unterschieden sind es aber wohl genau diese Eigenschaften die für eine nun wiederum bessere Erkennbarkeit der Granularität beim Gitarrenstimulus sorgen.

Die offensichtlich deutlich geringere Empfindlichkeit bei der Wahrnehmung lateraler Gridrastrereduktionen gegenüber der bei Reduktion in der horizontalen bzw. vertikalen Richtung (Abb. 3, rechts), bestätigten zwei Mann-Whitney-Tests für unabhängige Stichproben als signifikant.

Im Falle des Gitarrenstimulus lagen Schwellwerte und Streuungen zumeist deutlich höher. Die für rosa Rauschen gerade wahrnehmbare Gridrastrerauflösung wird erwartungsgemäß als hochsignifikant geringer bestätigt (im Mittel  $5^\circ$  vs.  $12,3^\circ$ ). In Übereinstimmung mit den bereits in Abschnitt 2 zitierten Untersuchungen zum Minimum Audible Movement Angle (MAMA, [4]) führte die höhere Bandbreite des Rauschsignals zu besseren Detektionsleistungen, der beobachtete Abfall beim natürlichen Signal ist ausgesprochen groß. Bei lateralen Rotationen wird diese höhere Empfindlichkeit für Reduktionen des Gridrasters bei Präsentation von Rauschen nur noch als schwacher Trend sichtbar. Die unter dieser Bedingung vorherrschende Unsicherheit der Versuchspersonen kann durch das breitere Rauschspektrum kaum verringert werden.

Der Einfluss der verschiedenen Räume führte zu keinem signifikanten Unterschied bei der Wahrnehmung horizontaler, vertikaler oder lateraler Gridauflösung. Für horizontale und vertikale Kopfbewegungen war jedoch ein Trend zu höherer Empfindlichkeit bei HRTFs sichtbar war (im Mittel um ca.  $1^\circ$ ). Bei lateralen Rotationen trat der umgekehrte Trend auf. Mit geringer werdendem Raumanteil stiegen offenbar die Schwierigkeiten eine gröbere Rasterung der Datensätze für eine frontale Schallquelle zu erkennen, bei der keine interauralen Pegel- und Laufzeitdifferenzen als Cue zur Verfügung stehen (s. a. Abschnitt 5).

## 5. Laterale Kopfbewegungen

Den Gründen für die in Vorversuchen (s. Abschnitt 3.5) festgestellte, rapide Zunahme der Empfindlichkeit für Diskretisierungen der lateralen Bewegungsrichtung bei Kopfneigungen von mehr als  $\pm 30^\circ$  wurde nachträglich durch eine Analyse der binauralen Impulsantworten nachgegangen. Da hier, zumindest für den Direktschallanteil einer frontalen Quelle, keine interauralen Pegel- und Laufzeitdifferenzen zur Verfügung stehen, kommen für die Lokalisation nur spektrale Differenzen der BRIRs als Cue in Frage. Diese sind in Abb. 4 für

die drei untersuchten HRTF- bzw. BRIR-Betragspektren über dem lateralen Auslenkungswinkel für die gemessenen Neigungswinkel von  $\pm 60^\circ$  dargestellt. Für den im Hörversuch dargebotenen Winkelbereich von  $\pm 30^\circ$  zeigen sich nur geringe spektrale Veränderungen. Im nachhallfreien Fall (Abb. 4 links) zeigt lediglich der Einbruch zwischen 5 und 10 kHz, eine Ohrkanalantiresonanz, mit zunehmender Neigung des Ohrs zur Schulter Veränderungen in Folge eines beginnenden Kammfilters. Insgesamt ist die geringe Empfindlichkeit für laterale Diskretisierungen in diesem Winkelbereich - zahlreiche Hörer konnten auch eine nur alle  $30^\circ$  nachgeführte Impulsantwort nicht von der maximalen Auflösung von  $1^\circ$  unterscheiden - nicht überraschend. Jenseits eines Neigungswinkels von  $30^\circ$  in Richtung der Schulter (für das linke Ohr also für Neigungen  $< -30^\circ$  in Abb. 3) zeigt sich jedoch ein deutlicher Kammfiltereffekt, der vermutlich durch Reflexionen an Schulter und Torso bedingt ist. Mit zunehmender Neigung des Ohrs verschieben sich die Peaks des Kammfilterfrequenzgangs in Richtung höherer Frequenzen.

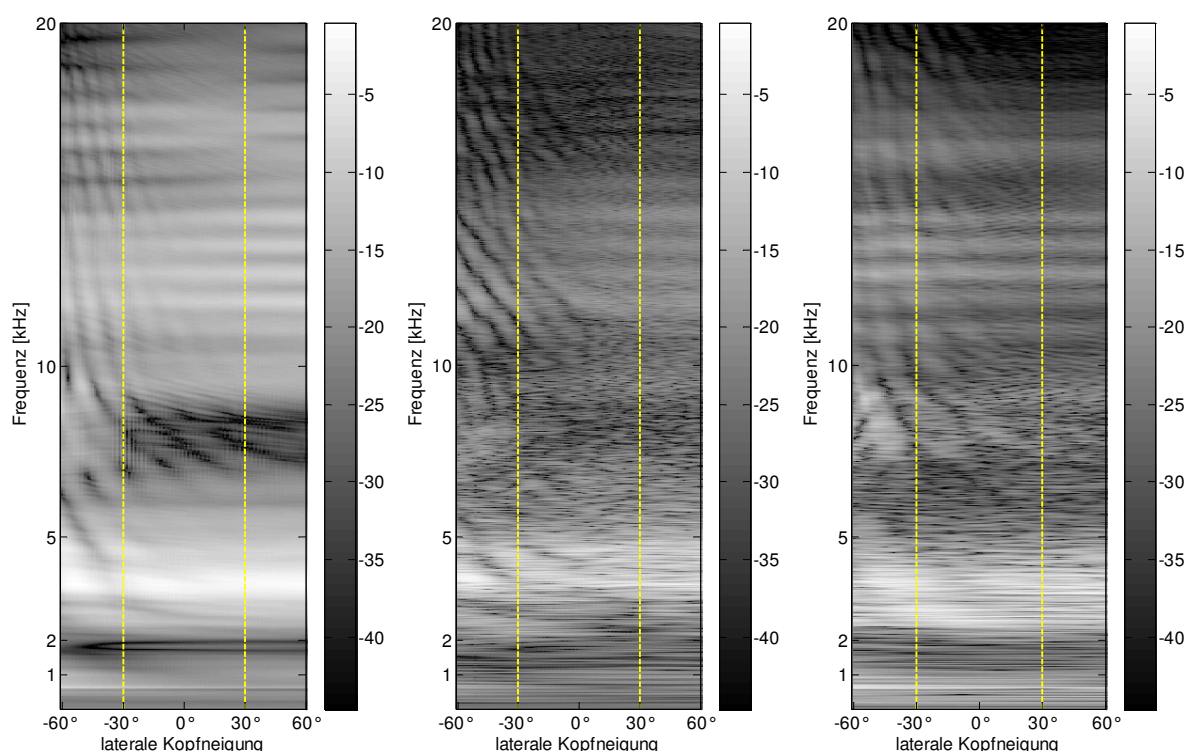


Abb. 4: Spektren der binauralen Impulsantworten (linkes Ohr) über dem lateralen Kopfneigungswinkel, gestrichelte Bereiche kennzeichnen den im Hörversuch benutzten Winkelbereich von  $\pm 30^\circ$ , v.l.n.r.: HRTFs, Studio, Hörsaal H 104

Bei der Auralisation des gesamten Winkelbereichs von  $\pm 60^\circ$  konnten zahlreiche Versuchspersonen bereits eine Diskretisierung in  $2^\circ$ -Schritten zuverlässig von der feinsten  $1^\circ$ -Auflösung unterscheiden. Somit gelten die untersuchten Schwellen für laterale Bewegungen tatsächlich auch nur für einen eingeschränkten Winkelbereich von  $\pm 30^\circ$ .

Die Erkennung von lateralen Diskretisierungen kann in nicht nachhallfreier Umgebung durch Reflexionen aus nicht-horizontalen Einfallrichtungen verbessert werden, etwa durch Deckenreflexionen, die bei lateraler Kopfneigung eine interaurale Laufzeit- und Pegeldifferenz induzieren. Dies könnte eine Erklärung sein, warum Versuchspersonen im Hörsaal H

104, der tatsächlich über eine stark reflektierende Decke verfügt, im Mittel - wenn auch nur geringfügig - empfindlicher für eine Reduktion der lateralen Auflösung waren als in nachhallfreier Umgebung ( $14^\circ$  versus  $17^\circ/18^\circ$ , s. Abb. 3). Die zusätzlichen Raumreflexionen in Studio und Hörsaal (Abb. 4, mittig und rechts) führen offensichtlich zu einer größeren Varianz des Spektrums über dem lateralen Neigungswinkel. Eine genauere Untersuchung dieses Effekts ist Gegenstand aktueller Versuche.

## 6. Zusammenfassung

Durch die vorliegende Studie wurde eine eben hörbare Diskretisierung für binaurale Raumimpulsantworten in drei Freiheitsgraden der Kopffrotation ermittelt. Schwellwerte wurden für verschiedene akustische Umgebungen und mit zwei Audioinhalten in adaptiven 3-AFC-Hörversuchen für die drei rotatorischen Kopfbewegungen bestimmt. Mit dem musikalischen Inhalt war es, zumindest für horizontale und vertikale Auflösung, deutlich schwieriger, Auflösungsreduktionen zu hören, als für rosa Rauschen. Im Falle der lateralen Bewegungen war dieser Unterschied weniger deutlich. Überraschenderweise lagen die Werte der eben wahrnehmbaren vertikalen Gridrausterauflösung bei breitbandiger Anregung wie rosa Rauschen niedriger als in der horizontalen Richtung. Die lateralen Schwellwerte dagegen lagen signifikant höher als die vertikalen und horizontalen.

Auflösung war hörbar für ...	Rauschen hor/ver/lat	Gitarre hor/ver/lat
50%	$6^\circ \times 5^\circ \times 16^\circ$	$9^\circ \times 12^\circ \times 16^\circ$
25%	$4^\circ \times 4^\circ \times 12^\circ$	$7^\circ \times 9^\circ \times 12^\circ$
5%	$4^\circ \times 3^\circ \times 8^\circ$	$5^\circ \times 4^\circ \times 8^\circ$
0%	$2^\circ \times 1^\circ \times 3^\circ$	$3^\circ \times 2^\circ \times 3^\circ$

Tab. 6: Für einen bestimmten Anteil der Stichprobe gerade wahrnehmbare horizontale, vertikale und laterale Gridrausterauflösung (getrennt dargestellt für die signifikant unterschiedliche Stimulusbedingung)

Angesichts dieser Schwellwerte sollte eine BRIR-Gridauflösung von  $2^\circ$  horizontal,  $1^\circ$  vertikal und  $3^\circ$  lateral eine für virtuelle akustische Umgebungen ausreichende Auflösung auch für sensibles Audiomaterial und kritische Hörer gewährleisten. Bei zukünftigen Untersuchungen zur Detektierbarkeit interpolierter binauraler Datensätze sollte dieses Raster als Zielrastrer verwendet werden. Für musikalische Inhalte in nicht nachhallfreien akustischen Umgebungen kann dagegen auch eine BRIR-Gridauflösung von  $5^\circ$  horizontal,  $4^\circ$  vertikal und  $8^\circ$  lateral für 95% der Hörer eine plausible binaurale Simulationen leisten.

## 7. Quellenverzeichnis

- [1] Lindau, A.; Hohn, T.; Weinzierl, S.: "Binaural resynthesis for comparative studies of acoustical environments." 122<sup>nd</sup> AES Conv., Vienna: 2007, preprint no. 7032
- [2] Mills, A.W.: "On the Minimum Audible Angle." In: *J. Acoust. Soc. Am.* (1958), Vol. 30, pp. 237-246

- [3] Perrott, D. R.; Musicant, A. D.: "Minimum audible movement angle: Binaural localization of moving sound sources." In: *J. Acoust. Soc. Am.*, Vol. 62 (1977), No. 6, pp. 1463-1466
- [4] Chandler, D. W.; Grantham, D.W.: "Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity." In: *J. Acoust. Soc. Am.*, Vol. 91, No. 3, pp. 1624-1636
- [5] Minnaar, P.; Plogsties, J.; Christensen, F.: "Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis." In: *J. Audio Eng. Soc.*, Vol. 53 (2005), No. 10, pp. 919-929
- [6] <http://www.ada-acousticdesign.de/set/setsoft.html>
- [7] <http://www.odeon.dk/>
- [8] Gardner, W. G.; Martin, K.: "HRTF measurements of a KEMAR." In: *J. Acoust. Soc. Am.*, Vol. 97 (1995), No. 6, pp. 3907-3908
- [9] Algazi, V. R. et al.: "The CIPIC HRTF Database." IEEE-WASPAA Workshop, New Paltz, New York, 2001
- [10] Lentz, T. et al.: "Virtual Reality System with Integrated Sound Field Simulation and Reproduction." In: *EURASIP J. of Advances in Signal Processing* (Article ID 70540), 2007
- [11] Silzle, A.; Novo, P.; Strauss, H.: "IKA-SIM: A System to Generate Auditory Virtual Environments". 116<sup>th</sup> AES Convention, Berlin: 2004, preprint no. 6016
- [12] Savioja, L. et al.: "Creating interactive virtual acoustic environments." In: *J. Audio Eng. Soc.*, Vol. 47 (1999), No. 9, pp. 675-705
- [13] Anderson, M. et al.: "Design and Verification of HeadZap, a Semi-automated HRIR Measurement System." 120<sup>th</sup> AES Conv., Paris: 2006, preprint no. 6655
- [14] Geier, M.; Ahrens, J.; Spors, S.: "The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods". 124<sup>th</sup> AES Conv., Amsterdam. 2008, preprint no. 7330
- [15] Mackensen, P.: *Auditive Localization. Head Movements, an additional cue in Localization*. Dissertation. Berlin: Technische Universität, 2004
- [16] Christensen, F. et al.: "A Listening Test System for Automotive Audio - Part 1: System Description". 118<sup>th</sup> AES Convention, Barcelona: 2005, preprint no. 6358
- [17] Hartung, K.; Braasch, J.; Sterbing, S.J.: "Comparison of different methods for the interpolation of head-related transfer functions." In: *Proc. of the AES 16<sup>th</sup> International Conference*. Rovaniemi, 1999
- [18] DIN 33408-1: *Körperumrißschablonen für Sitzplätze*, Berlin: Beuth, 1987
- [19] Thurlow, W.R.; Mangels, J.W.; Runge, P.S.: "Head movements during sound localization." In: *J. Acoust. Soc. Am.*, Vol. 42 (1967), pp. 489-493
- [20] Morgan, C. T. et al.: *Human Engineering Guide to Equipment Design*, New York: McGraw-Hill, 1963
- [21] Müller-Tomfelde, C.: "Time varying Filters in non-uniform Block Convolution." In: *Proc. of the COST G-6 Conference on Digital Audio Effects*. Limerick, 2001, Vol. 2001
- [22] Kirkeby, O.; Nelson, P.A.: "Digital Filter Design for Inversion Problems in Sound Reproduction." In: *J. Audio Eng. Soc.*, Vol. 47 (1999), No. 7/8, pp. 583-595
- [23] Schäfer, Z.; Lindau, A.: "Evaluation von Verfahren zur Entzerrung von binauralen Signalen". Beitrag zur 25. Tonmeistertagung, Leipzig: 2008
- [24] Hoffmann, P. F.; Møller, H.: "Audibility of Time Switching in Dynamic Binaural Synthesis". 118<sup>th</sup> AES Conv., Barcelona: 2005, preprint no. 6326

- [25] Hoffmann, P. F.; Møller, H.: "Audibility of Spectral Switching in Head-Related Transfer Functions". 119<sup>th</sup> AES Conv., New York: 2005, preprint no. 6537
- [26] Hoffmann, P. F.; Møller, H.: "Audibility of Spectral Differences in Head-Related Transfer Functions". 120<sup>th</sup> AES Conv., Paris: 2006, preprint no. 6652
- [27] Leek, M. R.: "Adaptive procedures in psychophysical research." In: *Perception & Psychophysics*, Vol. 63 (2001), Vol. 8, pp. 1279-1292
- [28] Pentland, A.: "Maximum likelihood estimation: The best PEST." In: *Perception & Psychophysics*, Vol. 28 (1980), No. 4, pp. 377-379
- [29] Bortz, J.: *Statistik für Sozial- und Humanwissenschaftler*. 6. vollst. überarb. und akt. Aufl.: Springer, 2005