Institut für
Musikpädagogische Forschung

FORSCHUNGSBERICHTE

12

PSYCHOAKUSTISCHE UND PSYCHOLOGISCHE ASPEKTE DER AUDIODATENREDUKTION

Eine Bestandsaufnahme aus Forschung und Praxis

MICHAEL CAMMANN

Hochschule für
Musik und Theater Hannover

Institut für
Musikpädagogische Forschung

FORSCHUNGSBERICHTE

12

PSYCHOAKUSTISCHE UND PSYCHOLOGISCHE ASPEKTE DER AUDIODATENREDUKTION

Eine Bestandsaufnahme aus Forschung und Praxis

MICHAEL CAMMANN

Hochschule für
Musik und Theater Hannover

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Cammann, Michael:
Psychoakustische und psychologische Aspekte der Audiodatenreduktion. Eine Bestandsaufnahme aus Forschung und Praxis
Michael Cammann. Institut für Musikpädagogische Forschung,
Hochschule für Musik und Theater Hannover. – Hannover :
Inst. für Musikpädag. Forschung, 2000

(Forschungsberichte / Institut für Musikpädagogische Forschung,
Hochschule für Musik und Theater ; 12)
ISBN 3-931852-11-3
NE: Institut für Musikpädagogische Forschung <Hannover>:
Forschungsberichte

Forschungsberichte des
Instituts für Musikpädagogische Forschung
an der
Hochschule für Musik und Theater Hannover
(Emmichplatz 1, 30175 Hannover)

Herausgeber dieser Ausgabe: Klaus-Ernst Behne

Redaktion: Ingo Laufs, Karl-Jürgen Kemmelmeyer

Nachdruck nur mit Genehmigung des Autors

Hannover 2000

ISSN 1430-8088

ISBN 3-931852-11-3

Institut für Musikpädagogische Forschung

Vorstand: Prof. Dr. K.-J. Kemmelmeyer (Direktor), Prof. Dr. R. Kopiez, Prof. Dr. F. Riemer
Prof. Dr. E. Altenmüller, Prof. Dr. F. Amrhein, Prof. Dr. H. Bäßler, Prof. P. Becker, Prof. K.-E. Behne,
Prof. Dr. A. Edler, Prof. Dr. J. Herwig, Prof. Dr. E. Hickmann, OStR A. Lehmann-Wermser, Bärbel Siefert
Sekretariat: Anette Zeuch

Externe Mitglieder: Prof. Dr. P. Brünger, Prof. Dr. H,-J. Kaiser, Martin Weber

Institutsgebäude: Schiffgraben 48

Vorwort des Herausgebers

Datenreduktion gehört zu den wichtigsten und zugleich meistumstrittenen Verfahren der heutigen Medien- und Kommunikationswelt, die entscheidend mit dazu beitragen, dass beispielsweise das Internet für Musik- und Sprachübermittlung genutzt werden kann.

Die Kosten für physischen Speicherplatz sinken, der Bedarf an Übertragungskapazitäten in Datennetzen wie dem Internet nimmt ständig zu. Um die Effizienz der Speicherung und der Übertragung von Audiodaten zu steigern, wurden unter Berücksichtigung der Eigenschaften des Gehörs Datenkompressions- und Datenreduktionsverfahren entwickelt, die das Datenvolumen verringern, ohne dabei zwingend die Klangqualität hörbar zu mindern.

Mit der Digitalisierung moderner Informations- und Kommunikationstechnologie wurde eine hohe Konvergenz aller Datenträger und Übertragungseinrichtungen erreicht – lediglich die Datenformate bzw. die Codierung der digitalen Daten entscheiden über Einsatzzweck und Qualität. Neue Verbreitungsformen führen zu einem veränderten Käuferverhalten und haben eine Umstrukturierung des Musikmarktes zwangsläufig im Gefolge. Mit der Verbreitung leistungsfähiger Computer in den Haushalten und den digitalen Möglichkeiten des Internet ist einerseits Musik fast jeder Art im Internet für die private Nutzung verfügbar, andererseits entstehen dadurch neue, bisher noch nicht gelöste Urheberrechtsprobleme.

Die vorliegende Forschungsarbeit von Michael Cammann entstand in Verbindung mit einer im Staatsexamen gestellten Aufgabe. Da diese Arbeit synoptisch Verfahren der Audiodatenreduktion einschließlich einer kritischen

Betrachtung gegenüberstellt, wird sie als schnelle Informationsquelle für weitere Studien dienen können.

Hannover, im November 2000

Klaus-Ernst Behne

Inhaltsverzeichnis

Einleitung **1**

Teil 1 - Das Gehör

1. Aufbau und Funktionsweise des menschlichen Gehörs **4**

1.1. Außen- und Mittelohr 5
1.2. Innenohr 6
1.3. Hörnerv 12
1.4. Zentrales auditorisches Nervensystem 13

2. Leistungsvermögen des Gehörs **14**

2.1. Tonhöhe 14
2.2. Lautstärke 16
2.3. Kritische Frequenzbandbreite 18
2.4. Zeitliches Auflösungsvermögen 21
2.5. Lautheit 21
2.6. Räumliches Hören 23
2.7. Maskierung: Verdeckungseigenschaften des Gehörs 24
2.7.1. Maskierung eines Tons durch stationäres Weißes Rauschen 25
2.7.2. Maskierung eines Tons durch stationäres Schmalbandrauschen 26
2.7.3. Maskierung eines Tons durch einen anderen, stationären Ton ... 28
2.7.4. Maskierung eines Tons durch komplexe, stationäre Töne 30
2.7.5. Maskierung durch instationäre Schallereignisse 31
2.7.6. „Addition“ mehrerer Maskierer 33

2.7.7. Erklärungsmodell für Maskierungseffekte 34

Teil II - Digitale Tonaufzeichnung

3. Digitale Signalverarbeitung 36

3.1. Duales Zahlensystem 36
3.2. Signalabtastung 37
3.3. Analog-Digital-Wandlung 39
3.4. Digital-Analog-Wandlung 41

4. Digitale Tonaufzeichnung ohne Datenreduktion 43

4.1. Compact Disc 43
4.2. Digital Audio Tape 46

Teil III - Datenreduktion unter psychoakustischen und psychischen Aspekten

5. Digitale Tonaufzeichnung unter Datenreduktion 47

5.1. Datenreduktionsverfahren der Digital Compact Cassette: PASC 50
5.2. Datenreduktionsverfahren der MiniDisc: ATRAC 52
5.3. MPEG Audio - Layer 3 58
5.3.1. Signalzerlegung - Polyphasenfilterbank 60
5.3.2. Hybrid-Filterbank 61
5.3.3. Psychoakustisches Modell 62
5.3.4. Joint Stereo Coding 63
5.3.4.1. M/S-Stereokodierung 64
5.3.4.2. Intensitäts-Stereokodierung 65
5.3.5. Huffman Coding 67
5.4. Kritische Betrachtung der gehörangepassten Datenreduktion 68
5.4.1. Qualitätsbeurteilung als Grundproblem der Datenreduktion 69

5.4.1.1. Subjektive Qualitätsbewertung nach der Triple stimulus-hidden reference-double blind - Methode 70
5.4.1.2. „CD-Qualität“ - ein realistisches Gütesiegel? 75
5.4.2. Schutz der Urheberrechte 79
5.4.2.1. Music-on-Demand 81

6. Schlußbetrachtung und Ausblick 83

7. Literatur 90

Anhang 98

Email von Niels Rump, Business Relations Manager des Fraunhofer-Instituts für Integrierte Schaltungen Erlangen vom 07.09.1998

Email der Music-On-Demand-Kundenbetreuung vom 07.12.1998

Stichwortverzeichnis 100

Abbildungsverzeichnis

Abb. 1: Schnitt durch das Gehör....4
http://www.dasp.uni-wuppertal.de/akustikshow/geamtdars.html (Stand: 14.12.98)

Abb. 2: Ansicht des Mittelohres....6
http://www.dasp.uni-wuppertal.de/akustikshow/mittelohr.html (Stand: 14.12.98)

Abb. 3: Ansicht des Innenohres mit Schnitt durch die Schnecke.....7
http://www.dasp.uni-wuppertal.de/akustikshow/innenohr.html (Stand: 14.12.98)

Abb. 4: Schematische Darstellung des Ohres.....8
Hall, Donald E.: *Musikalische Akustik - Ein Handbuch*, Mainz 1997, S. 107

Abb. 5: Aktiv-mechanische Verstärkung der Wanderwelle in schematischer Darstellung als Auslenkung der Basilarmembran.....11
Zenner, Hans-Peter: *Hören: Physiologie, Biochemie, Zell- und Neurobiologie*, Stuttgart 1994, S. 21

Abb. 6: Aktive Auslenkung der Wanderwelle der Basilarmembran. 11
Zenner, Hans-Peter, a.a.O., S. 18

Abb. 7: Eben merkbare Unterschiede in der Frequenz in Abhängigkeit von der Anfangsfrequenz für Sinuswellen.....15
Hall, Donald E., a.a.O., S. 113

Abb. 8: Schallempfindlichkeit in Relation zur Frequenz 16
Borucki, Hans: *Einführung in die Akustik*, Zürich 1980, S. 201, nach Zwicker, Eberhard/ Feldtkeller, Richard: Das Ohr als Nachrichtenempfänger, Stuttgart 1967, S. 3

Abb. 9: Eben merkbarer Unterschied im Schallpegel L_I in Abhängigkeit des anfänglichen Schallpegels für Sinuswellen mit verschiedenen Frequenzen.....17
Hall, Donald E., a.a.O., S. 112

Abb. 10: Änderung der wahrgenommenen Lautstärke von schmalbandigem Rauschen bei Erhöhung der Bandbreite und konstanter Gesamtintensität.....19
Hall, Donald E., a.a.O., S. 392

Abb. 11: Auswirkungen der kritischen Bandbreite auf die Tonhöhenwahrnehmung.....19
Hall, Donald E., a.a.O., S. 393

Abb. 12: Mithörschwelle eines Tons, eben verdeckt durch Weißes Rauschen 25
Zwicker, Eberhard/Fastl, Hugo: *Psychoacoustics - Facts and Models*, Berlin/Heidelberg/New York 1990, S. 57

Abb. 13: Mithörschwellen eines Testtons, eben verdeckt durch Schmalbandrauschen mit der Mittenfrequenz f_C. 27
Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 59

Abb. 14: Mithörschwellen eines Testtons, eben verdeckt durch Schmalbandrauschen mit f_c=1 kHz und verschiedenen Störpegeln L_{CB}.....27
Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 60

Abb. 15: Mithörschwelle eines Tons, eben verdeckt durch einen Störton mit einer Frequenz von 1 kHz bei einem Pegel von 80 dB.....29
Eigene Grafik, nach: Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 62

Abb. 16: Mithörschwelle eines Tons, eben verdeckt durch neun Obertöne mit gleichem Schallpegel über der Grundtonfrequenz 200 Hz.. 30
Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 65

Abb. 17: Schematische Darstellung des zeitlichen Verlaufs bei Maskierung durch instationäre Stimuli 32
Herre, Jürgen: *Fehlerverschleierung bei spektral kodierten Audiosignalen*, Erlangen 1995, S. 14

Abb. 18: Addition von Verdeckungseffekten 33
Zwicker, Eberhard/H. Fastl, a.a.O., S. 97

Abb. 19: A/D-Wandlung mit 4-Bit-Auflösung 40
Dickreiter, Michael: *Handbuch der Tonstudiotechnik - Band 2*, 6. verbesserte Auflage, München 1997, S. 281

Abb. 20: Quantisierungsrauschabstand eines digitalen Systems 41
Webers, Johannes: *Tonstudiotechnik - Schallaufnahme und -wiedergabe bei Rundfunk, Fernsehen, Film und Schallplatte*, München 1989, S. 156

Abb. 21: Blockdiagramm eines Enkodierers mit psychakustischem Modell..... 49
Eigene Grafik

Abb. 22: ATRAC-Enkoder in Blockdarstellung..... 54
http://www.audiomap.de/infos/minidisc.htm (Stand: 01.06.2000)

Abb. 23: Bitzuweisung durch ATRAC 55
http://www.audiomap.de/infos/minidisc.htm

Abb. 24: ATRAC - MDCT-Zeitfenstergrößen. 56
Eigene Grafik

Abb. 25: Übergang Vor- zu Simultanverdeckung 57
Nach Herre, Jürgen: a.a.O., S. 14

Abb. 26: MPEG-Audio Layer 3: Subbandunterteilung (oben) im Vergleich zur Anordnung der Frequenzgruppen (unten). 61
Pan, Davis: *A Tutorial on MPEG/Audio Compression.* Reprinted IEEE Multimedia Journal, Sommer 1995, nach: www.eskimo.com/~miyaguch/mp3info.html (Stand: 01.06.2000)

Abb. 27: MPEG-Audio Layer 3: Bit-Einsparungspotential in Abhängigkeit von der Stereoredundanz..... 65
Eigene Grafik

Abb. 28: Signalweg während der Kodierung und Dekodierung bei der Intensitäts-Stereokodierung MPEG-Audio Layer 3 66
Nach: Herre, Jürgen, a.a.O., S. 32

Abb. 29: Bewertungsskalen für die subjektive Qualitätsbewertung. 71
Hoeg, Wolfgang/Steffen, Edith: *Methoden zur subjektiven Qualitätsbewertung von Audiosignalen*, in: Highlights aus der Forschung – Quellenkodierung und Qualitäts-beurteilung als Impulsgeber für Innovation bei Sprache, Audio und Video (Deutsche Telekom AG Hg.), Darmstadt 1996, S. 30

Abb. 30: MPEG-Audio Layer 3: Qualitätsminderung gegenüber der Referenz. 74
Nach: Hoeg, Wolfgang/Steffen, Edith, a.a.O., S. 38

Abb. 31: Preisvergleich zwischen Music-on-Demand und CD-Versand anhand zweier willkürlich ausgewählter Alben. 85
Eigene Grafik

Tabellenverzeichnis

Tab. 1: Bark-Skala 20
Nach: Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 142

Tab. 2: Lautheitswerte eines 1 kHz-Tons. 22

Tab. 3: Gängige Kompressionsraten der MPEG-1 Audio-Layer.....59

Tab. 4: MPEG-Audio Layer 3: Qualitätsminderung gegenüber der Referenz -Übersicht der verwendeten Programmausschnitte. 74
Nach: Hoeg, Wolfgang/Steffen, Edith, a.a.O., S. 38

Abdruckgenehmigungen zu den Abbildungen und Tabellen

Abb. 4/7/9/10/11: Hall, Donald E.: *Musikalische Akustik – Ein Handbuch*,
Mainz 1997, S. 107/113/112/392/393;
© INTERNATIONAL THOMSON PUBL., Andover, GB

Abb. 5/6: Zenner, Hans-Peter: *Hören: Physiologie, Biochemie, Zell- und Neurobiologie*,
Stuttgart, 1994, S. 21/18;
© GEORG THIEME VERLAG, Stuttgart

Abb.8: Boruki, Hans: *Einführung in die Akustik*,
Zürich, 1980, S. 201;
© SPEKTRUM AKADEMISCHER VERLAG, Heidelberg

Abb. 12/13/14/16/18: Zwicker, Eberhard/Fastl, Hugo*: Psychoakustics – Facts and Models*,
Berlin/Heidelberg/New York, 1990, S. 57/59/60/65/97;
© SPRINGER VERLAG, Heidelberg

Abb. 19: Dickreiter, Michael: *Handbuch der Tonstudiotechnik – Band 2*,
6. verbesserte Auflage, München, 1997, S. 281;
© K.-G. SAUR-VERLAG, München

Abb. 20: Webers, Johannes: *Tonstudiotechnik – Schallaufnahme und -wiedergabe bei Rundfunk, Fernsehen, Film und Schallplatte*,
München 1989, S. 156;
© FRANZIS VERLAG GmbH, München

Tab. 1: Zwicker, Eberhard/Fastl, Hugo: *Psychoackoustics – Facts and Models*,
Berlin/Heidelberg/New York, 1990; S. 142;
© SPINGER VERLAG, Heidelberg

Abkürzungsverzeichnis

A/D	Analog / Digital
ADSL	Asymmetric Digital Subscriber Line
AES	Audio Engineering Society
ATRAC	Adaptive Transform Acoustic Coding
Bit	Binary digit
CAV	Constant Angular Velocity
CD	Compact Disc
CD-DA	Compact Disc Digital Audio
CD-ROM	Compact Disc - Read Only Memory
CLV	Constant Linear Velocity
codec	coder and decoder
D/A	Digital / Analog
DAB	Digital Audio Broadcasting
DAT	Digital Audio Tape
DCC	Digital Compact Cassette
DVD	Digital Versatile Disc
Fraunhofer IIS	Fraunhofer-Institut für Integrierte Schaltungen, Erlangen
GEMA	Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte
GfK	Gesellschaft für Konsumforschung
IEC	International Electrotechnical Commission
ISDN	Integrated Services Digital Network
ISO	International Organization for Standardization
ITU-R	International Telecommunication Union - Radiocommunication Sector
JND	just noticeable difference
MC	MusiCassette
MD	MiniDisc
MDTC	Modifizierte diskrete Cosinus-Transformation
MO	Magneto Optical
MoD	Music-on-Demand
MP3	MPEG Audio-Layer 3
MPEG	Moving Pictures Experts Group
PASC	Precision Adaptive Subband Coding
PCM	Pulse-Code-Modulation
RIAA	Recording Industry Association of America
SNR	signal-to-noise ratio
www	world wide web

Einleitung

In der vorliegenden Arbeit wird das Spannungsfeld zwischen optimaler Klangqualität und gleichzeitig optimaler Datenreduktion erörtert, das auch in der Forschung nach neuen bzw. verbesserten Verfahren zur Reduktion von Audiodaten im Mittelpunkt steht.
Ferner wird diskutiert, inwiefern die Frage der Qualitätsbeurteilung in rein wissenschaftlicher und in wirtschaftlicher Hinsicht ihren Stellenwert hat.

Auf eine empirische Untersuchung musste verzichtet werden, da die technischen Anforderungen an einen repräsentativen Hörversuch letztlich zu hoch und somit nicht zu realisieren sind. So wird die Beurteilung von kodiertem Tonmaterial in der vorliegenden Arbeit in Berufung auf vergangene Versuche untersucht.

Im ersten Teil wird der Aufbau und das Leistungsvermögen des menschlichen Gehörs beschrieben, da hierin das Potential derjenigen Datenreduktionskonzepte zu finden ist, die auf den Verdeckungseigenschaften des Gehörs basieren.

Da als Qualitätsmaßstab für datenreduziertes Audiomaterial häufig die Compact Disc (CD) angeführt wird, soll im zweiten Teil ein grundlegender Einblick in die Digitaltechnik und die nicht-datenreduzierende Tonaufzeichnung gegeben werden. Diese Zusammenhänge sind Voraussetzung für das Verständnis der limitierenden Faktoren, die eine Datenreduktion notwendig machen.
Im dritten Teil, der zugleich den Schwerpunkt dieser Arbeit darstellt, wird eine Auswahl an Audiodatenreduktionsverfahren beschrieben. Der Schwerpunkt liegt auf den gegenwärtig am häufigsten angewandten Systemen *ATRAC*, das in der *MiniDisc (MD)*-Technologie verwendet wird, und *MPEG Audio-Layer 3 (MP3)*, das im Bereich der Datennetze zunehmend an Bedeutung gewinnt.
Das dritte führende System, *RealAudio*, ist in Hinblick auf seine Reduktionsprinzipien mit den beiden erstgenannten vergleichbar. Dennoch wird

in dieser Arbeit nicht weiter auf RealAudio eingegangen, da es das *Streaming*-Verfahren zur Übertragung verwendet: Zunächst wird nur der Anfang einer Audio- oder Videodatei übertragen; während dieser abgespielt wird, erfolgt im Hintergrund der Transfer der restlichen Daten. Weil keine Möglichkeit vorgesehen ist, die Daten beim Empfänger dauerhaft zu speichern, muss der Reduktionsfaktor ständig an die momentane Netzauslastung angepasst werden, um einen kontinuierlichen Datenfluß zu gewährleisten.[1] Folglich existieren keine konstanten, vergleichbaren Werte hinsichtlich der Reduktionsrate und der Klangqualität, wodurch eine inhaltliche Auseinandersetzung nahezu unmöglich gemacht wird.

Datenreduktion basiert nicht ausschließlich auf psychoakustischen Erkenntnissen über die Eigenschaften des Gehörs. Daher werden auch allgemeine, für die Bitreduktion ebenso wichtige Grundlagen der digitalen Signalverarbeitung und der statistischen Datenkompressionsverfahren berücksichtigt.

Generell muss festgestellt werden, dass das Material, das im world wide web angeboten wird, einer sachlichen Hinterfragung oftmals nicht standhalten kann und in Hinblick auf Quellenangaben häufig der wissenschaftlichen Grundlage entbehrt.

Der letzte Abschnitt der vorliegenden Arbeit behandelt abschließend den Komplex der Qualitätsbeurteilung. Dabei wird untersucht, nach welchen Methoden, Kriterien und Absichten die Qualität datenreduzierten Materials

[1] MP3-Dateien hingegen werden erst abgespielt, wenn die Übertragung vollständig abgeschlossen ist, so dass in diesem Verfahren keine Qualitätsveränderungen durch Schwankungen in der Übertragung verursacht werden;
nach: Music-on-Demand-Kundenbetreuung, persönliche Stellungnahme per Email (07.12.1998), s. Anhang.

beurteilt wird. Besonders das häufig verwendete Gütesiegel „CD-Qualität“[2] steht im Mittelpunkt der Betrachtung.

[2] Der Suchbefehl nach den Begriffen „CD-Qualität“ und „MPEG“ oder „MP3“ ergab über 14715 bzw. 9370 Einträge (Suchmaschine: http://www.altavista.com – Stand: 08.12.1998).

Teil I: Das Gehör

1. Aufbau und Funktionsweise des menschlichen Gehörs

Das Ohr dient dem menschlichen Hörapparat als Schallwahrnehmungsorgan, das die ankommenden Luftdruckschwankungen in elektrische Nervenimpulse transformiert, die zur Verarbeitung an das Gehirn weitergeleitet werden. Es ist anatomisch untergliedert in die Abschnitte *des Außenohres*, des *Mittelohres* und des *Innenohres* (s. Abb.1).

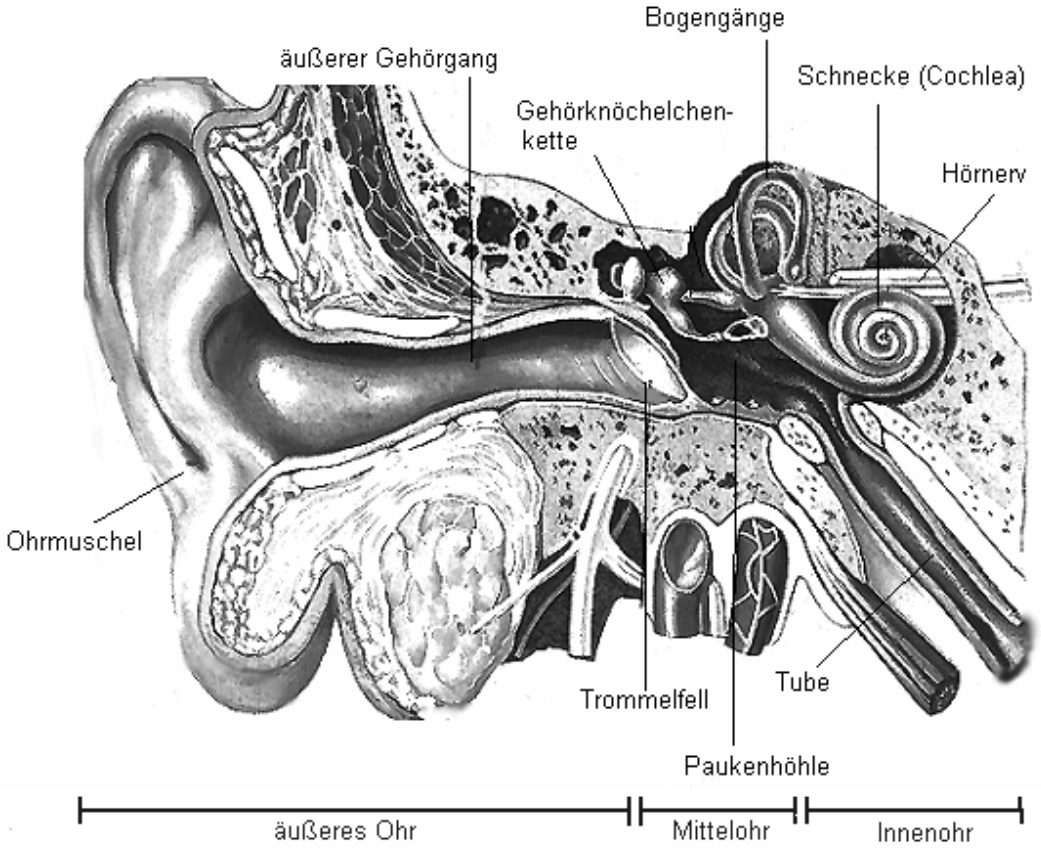


Abb. 1: Schnitt durch das Gehör

1.1. Außen- und Mittelohr

Das Außenohr leitet den an der *Ohrmuschel* ankommenden Schall durch den *Gehörgang* zum *Trommelfell*, einem dünnen Häutchen aus faserigem Gewebe. Dabei dient der innere Teil der Ohrmuschel als Schalltrichter, der kurzwellige Schallwellen in Richtung des Trommelfells bündelt.

Da sich der äußere Gehörgang mit einem offenen und einem geschlossenen Ende als Resonanzraum auswirkt, ist bereits ein Schalldruckpegelzuwachs von bis zu 20 dB am Trommelfell messbar. Die Frequenz, bei der sich die Resonanz am stärksten auswirkt, liegt bei ca. 2500 Hz.[3]

Entsprechend der schallbedingten Luftdruckschwankungen im Gehörgang resoniert das Trommelfell und überträgt die Schwingungen an das Mittelohr, eine geschlossene Kammer, die unmittelbar an das Trommelfell angrenzt. Das Mittelohr ist über die *Eustachische Röhre (Tube)* mit der Mundhöhle verbunden, so dass auf beiden Seiten des Trommelfells gleicher Luftdruck herrscht; Luftdruckunterschiede würden sonst das Trommelfell zu einer Seite hin auslenken und das Hörvermögen beeinträchtigen. Diese Verbindung des Ohres zum Nasen-Rachen-Raum wird über die aneinanderhaftenden Wände der Eustachischen Röhre verschlossen gehalten und nur beim Schlucken und Gähnen kurzzeitig geöffnet.

Im Mittelohr (s. Abb. 2) wird die Schallenergie durch mechanische Schwingungen vom Trommelfell über die drei Gehörknöchelchen *Hammer, Amboß* und *Steigbügel* zum Innenohr weitergeleitet. Hammer und Steigbügel sind mit je einem Muskel in der Paukenhöhle befestigt, der sich zum Schutz des Ohres anspannt, sobald zu hohe Lautstärken auf den Hörapparat einwirken. Allerdings tritt diese Muskelkontraktion erst verzögert ein, so dass dieser Schutz bei plötzlich auftretenden Schallereignissen mit zu großer Energie (z.B. bei

[3] Zenner, Hans-Peter: *Hören: Physiologie, Biochemie, Zell- und Neurobiologie*, Stuttgart 1994, S. 4.

lautem Hupen) nicht arbeitet. Der Schall erreicht das Innenohr und kann dort irreparable Schäden (z.B. ein Knalltrauma) an den Sinneszellen verursachen.

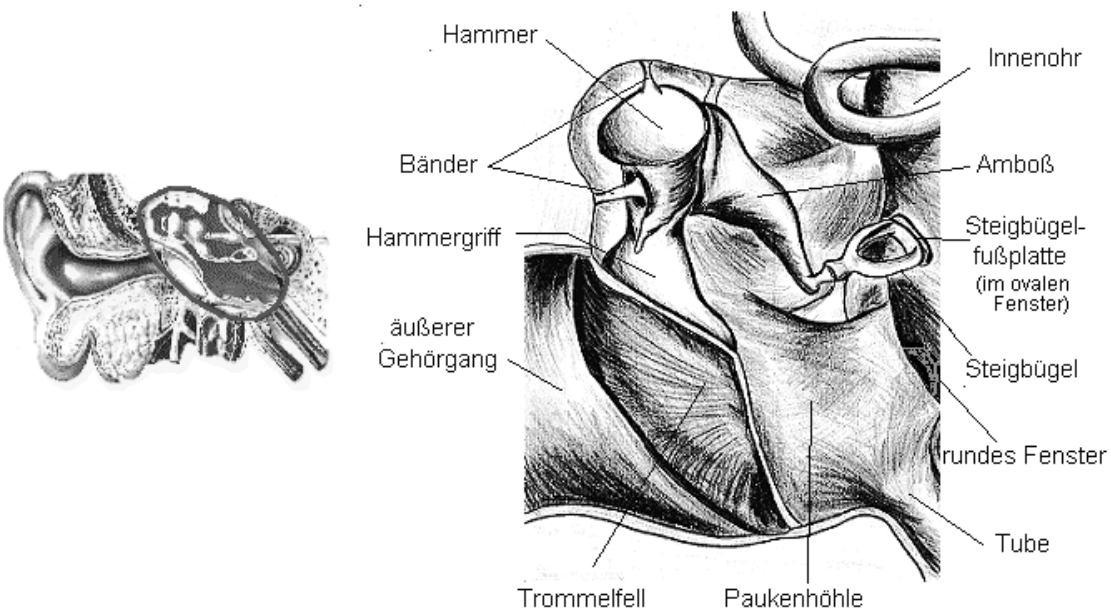


Abb. 2: Ansicht des Mittelohrs

1.2. Innenohr

Der Übergang von Mittel- zu Innenohr besteht aus zwei Membranen, dem *Ovalen Fenster* und dem *Runden Fenster*, hinter denen sich die *Perilymphe*, die Flüssigkeit des Innenohres, befindet. Die Gehörknöchelchen wirken als Hebelsystem, so dass der Steigbügel am *Ovalen Fenster* eine Kraft ausübt, die um den Faktor v_H=1,3 höher ist als die, mit der das Trommelfell den Hammer auslenkt. Um zwischen Trommelfell und Steigbügelfußplatte das Kräftegleichgewicht aufrecht zu erhalten, muss der an der Steigbügelfußplatte ausgeübte Druck an die auf das Trommelfell einwirkende Kraft angepasst

werden. Dieser Verstärkungsfaktor v_F=17 ergibt sich nach GRIEFANH/JANSEN[4] direkt aus dem Flächenunterschied zwischen Trommelfell (A_T=55 mm^2) und Steigbügelfußplatte (A_S=3,2 mm^2).[5] Zusammen mit der Kraftverstärkung durch das Hebelsystem der Gehörknöchelchen resultiert daraus der Gesamtverstärkungsfaktor v_{ges}=22.[6] Der Druck am Ovalen Fenster weist also den 22-fachen Wert des Druckes am Trommelfell auf.

Neben den Bogengängen, die dem Gleichgewichtssinn dienen, befindet sich im Innenohr die *Schnecke (Cochlea)* (s. Abb. 3).

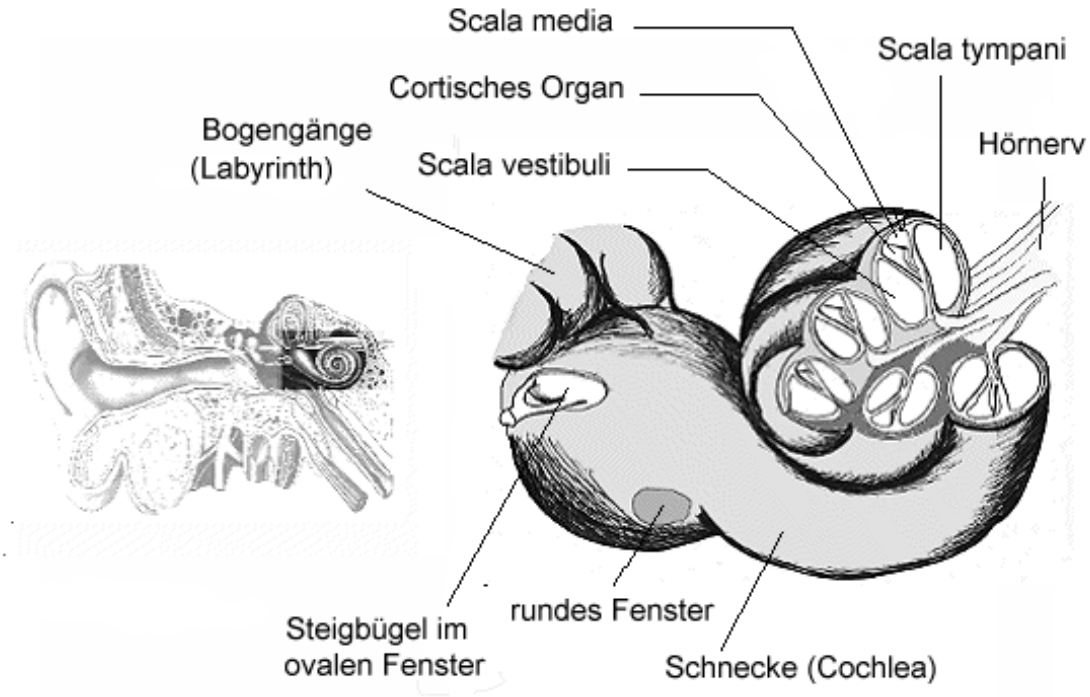


Abb. 3: Ansicht des Innenohres mit Schnitt durch die Schnecke

[4] Vgl. Griefahn, Barbara/Jansen, Gerd: *Schallwirkung beim Menschen und Fragen des Gehörschutzes*, in: Taschenbuch der Technischen Akustik (Heckl, Manfred/Müller, Helmut A. Hg.), Berlin/Heidelberg/New York ^{2}1994 [^{1}1975], S. 71.

[5] $v_F = A_T / A_S = 55\ mm^2 / 3{,}2\ mm^2 \approx 17$.

[6] $v_{ges} = v_H \cdot v_F = 1{,}3 \cdot 17 \approx 22$.

Die Cochlea wird durch ein komplexes System aus Membranen[7] und Lymphflüssigkeit geteilt in den *Paukengang (scala timpani)* und den *Vorhofsgang (scala vestibuli),* die lediglich an der Schneckenspitze durch eine kleine Öffnung, das *Schneckenloch (Helicotrema),* verbunden sind (s. Abb. 4).

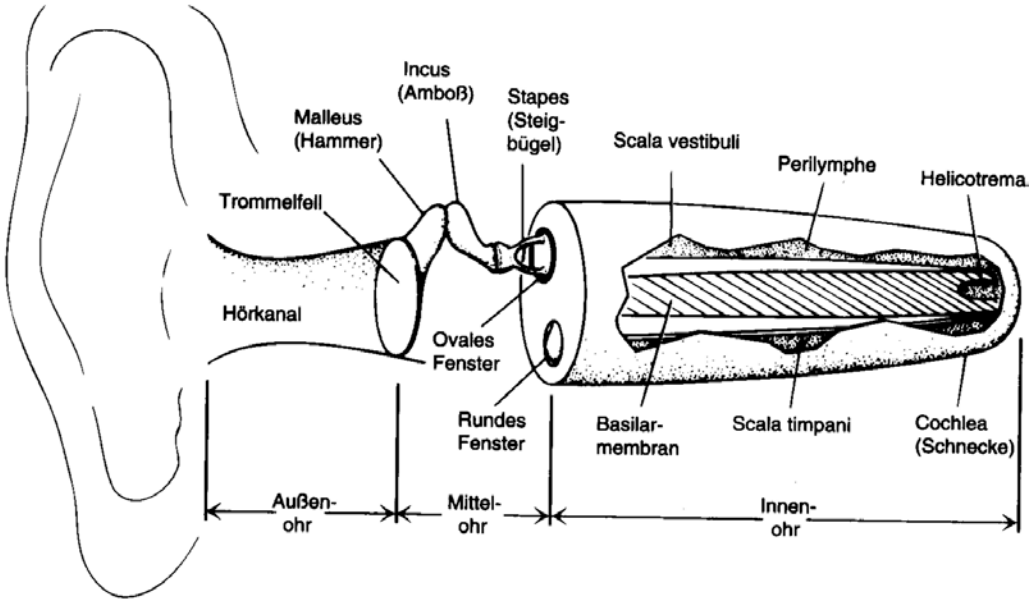


Abb. 4: Schematische Darstellung des Ohres.
Wichtige Teile sind stark vergrößert und die Schnecke in ausgerolltem Zustand dargestellt.

Wird nun das Ovale Fenster einwärts in das Innenohr ausgelenkt, überträgt sich diese Auslenkung als Wanderwelle auf die im Paukengang befindliche Flüssigkeit, die ihrerseits zu einer Auslenkung der *Basilarmembran* führt und sich in der Flüssigkeit im Vorhofsgang fortsetzt.

Die gesamte Bewegung innerhalb des Innenohres als geschlossenes System wird mechanisch erst durch das Runde Fenster ermöglicht, das sich dem Ovalen Fenster entgegengerichtet auslenkt: drückt die Vibration am Steigbügel das Ovale Fenster ein, beginnt die Flüssigkeit zu schwingen und dehnt das Runde Fenster als letztes Glied dieser Schwingungskette aus und umgekehrt.

[7] Zur Vereinfachung wird der Apparat der cochleären Trennwand namentlich reduziert auf seinen Hauptbestandteil, die Basilarmembran.

Die Basilarmembran bildet das Bindeglied zwischen Schall und Schallwahrnehmung: auf ihr sitzt das *Cortische Organ*, das aus 3500 *inneren* und 20000 *äußeren Haarzellen* besteht.[8] Die Bewegung unterschiedlicher Bereiche der Basilarmembran regt unterschiedliche Haarzellengruppen an, die dann Signale an die sie umspinnenden Nervenfasern aussenden.

Eine schallbedingte Wanderwelle breitet sich innerhalb der Cochlea nicht gleichmäßig von der Schneckenbasis bis zur Spitze aus, da die Basilarmembran die Wellen verstärkt bzw. abdämpft. Die Amplitude einer Welle steigt auf ihrem Weg vom Ovalen Fenster zur Schneckenspitze bis zu einem Maximum an und fällt darauf sofort wieder stark ab. Diese Erscheinung beruht auf dem Aufbau der Basilarmembran: während sie an der Schneckenbasis schmal (< 0,1 mm) und steif ist, verbreitert sie sich zum Helicotrema (ca. 0,5 mm) und wird schlaffer und massiver. Die am Ovalen Fenster angeregte Wanderwelle breitet sich also entlang der Basilarmembran aus und versetzt diese in Bewegung. Mit zunehmender Entfernung der Auslenkungswelle von der Cochleabasis steigt die Schwingungsamplitude, da sie durch die schlaffer werdende Basilarmembran weniger gedämpft wird. Parallel zur Steifigkeitsabnahme wächst jedoch die Massenträgheit, die ihrerseits die Schwingung dämpft. An einer von der Schwingungsfrequenz abhängigen Position auf der Basilarmembran löschen sich die Auswirkungen der Steifigkeit und der Massenträgheit aus, wodurch die Trägheit des Gewebes minimiert und dieses maximal ausgelenkt wird. Danach dominiert die Trägheit und die Amplitude nimmt rapide ab.

Die Frequenzselektivität der Basilarmembran beruht auf der Eigenschaft, dass der Bereich mit maximaler Auslenkungsamplitude von der Tonfrequenz abhängt: Hohe Frequenzen bewirken ein Maximum in der Nähe der Schneckenbasis; mit sinkender Frequenz verlagert sich diese Position zur Schneckenspitze. So wird ein aus mehreren Tonhöhen bestehendes Schallsignal

[8] Griefahn, Barbara/Jansen, Gerd, a.a.O., S. 69.

entlang der Basilarmembran aufgespreizt (*Frequenzdispersion*). Dieser Vorgang lässt sich als eine grobe Frequenzauflösung interpretieren und wird daher als *Ortsprinzip* bezeichnet.

Bereits bei HELMHOLTZ ist zu lesen, *„dass es verschiedene Teile des Ohres sein müssen, welche durch verschieden hohe Töne in Schwingung versetzt werden und diese Töne empfinden.*“[9] Diese Vorstellung deckt sich in ihren Grundzügen mit dem heutigen Stand der Forschung. Allerdings weiß man heute, dass Schwingungen im Innenohr einer ständigen Dämpfung unterliegen[10], so dass sich die stärksten Trennwandauslenkungen an ihrem frequenzspezifischen Ort nur breit und flach ausprägen. Durch diese passiven Wanderwellen kann die hohe Frequenzselektivität der Cochlea nicht ausreichend erklärt werden.[11]

ZENNER bezeichnet die aktive Schallvorverarbeitung als *„die bedeutendste Stufe in der Frequenzanalyse des Schalls durch das auditorische System“*.[12] Dabei formt die Cochlea die flache und breite passive Wanderwelle aktiv in eine solche mit hohen und scharfen Spitzen um, die sich wiederum auf einem sehr schmalen Bereich der cochleären Trennwand abbildet. Nach heutigen Hypothesen geht man davon aus, dass es entlang der Cochlea separat wirkende Vibrationsverstärker gibt, die der Wanderwelle Energie zuführen, sobald sie ihre frequenzspezifische Region durchläuft.[13] Dadurch wird die Auslenkung noch steiler bis hin zu einer ausgeprägten Spitze, die dann umso stärker abfällt: in dem Moment, in dem die Wanderwelle ihren frequenzspezifischen Bereich verlässt, setzt auch die aktive Verstärkung durch die Cochlea aus (s. Abb. 5).

---

[9] Helmholtz, Hermann von: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Braunschweig [6]1913 [[1]1862] , S. 235.

[10] Zenner, Hans-Peter, a.a.O., S. 18.

[11] ebd.

[12] ebd.

[13] Zenner, Hans-Peter, a.a.O., S. 21.

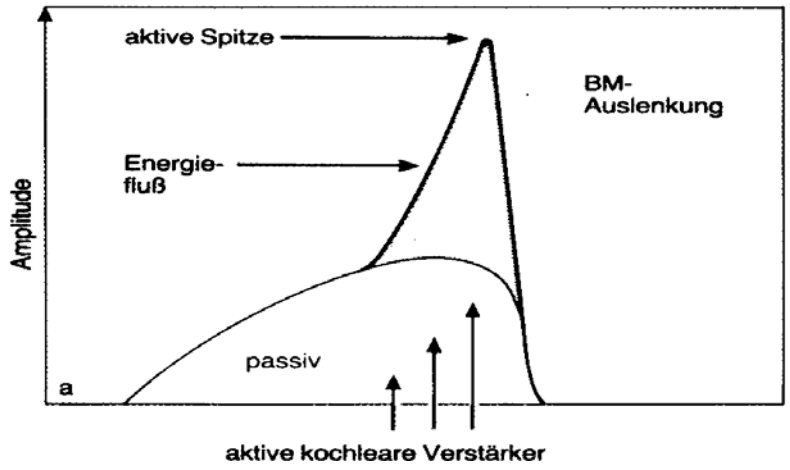


Abb. 5: Aktiv-mechanische Verstärkung der Wanderwelle in schematischer Darstellung als Auslenkung der Basilarmembran (BM).

Beobachtet wurde dieses Phänomen anhand der Basilarmembranauslenkung lebender und toter Organe (s. Abb. 6), so dass man der Cochlea eindeutig eine Fähigkeit zur aktiven Beeinflussung der Wanderwelle zusprechen kann.

Die die Schnecke durchlaufende Wanderwelle besteht demnach aus zwei Komponenten: einem passiven Teil, der durch den äußeren Schall ausgelöst wird, und einem aktiven Teil, den die Cochlea selbst mit eigenen Energiequellen erzeugt, feinabgestimmt auf die passive Wanderwelle. Die Herkunft dieser *„Energie biologischen Ursprungs“* [14] ist noch nicht wissenschaftlich erwiesen.

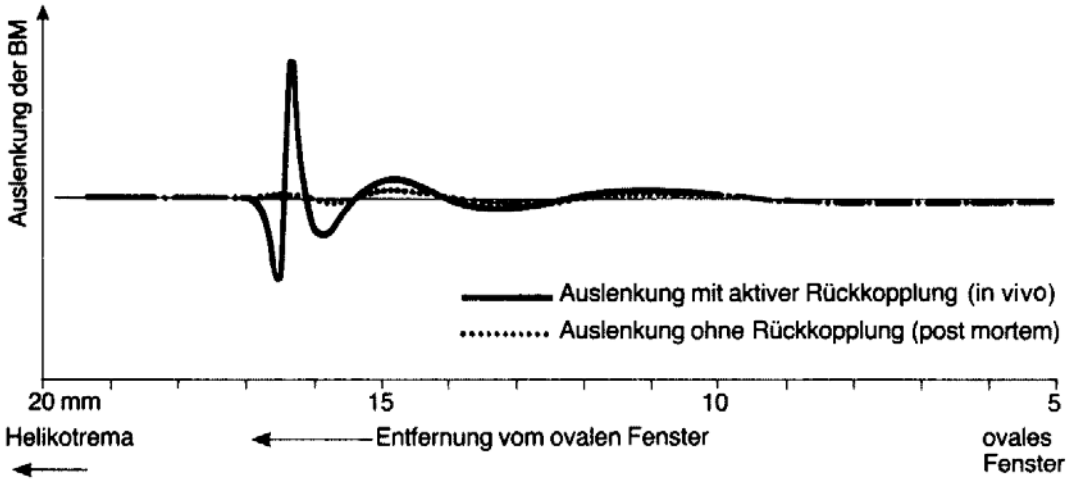


Abb. 6: Aktive Auslenkung der Wanderwelle der Basilarmembran

[14] Zenner, Hans-Peter, a.a.O., S. 31.

1.3. Hörnerv

Der Hörnerv verbindet die Cochlea mit dem Hirnstamm und besteht aus Nervenfasern, die jede für sich von einer bestimmten Haarzelle aktiviert werden.[15]

Die Dauer eines Stimulus wird durch die Zeitdauer der Nervenfaseraktivierung übermittelt; die Höhe des Schalldruckpegels wird kodiert als die *Entladungsrate*. Diese kann zwischen der Hörschwelle (s. Kap. 2.2., Abb. 8) und 80 dB bis 90 dB nichtlinear zunehmen.[16] Ab einer bestimmten Lautstärke kann eine Nervenfaser ihre Entladungsrate nicht weiter steigern, so dass bei noch höheren Lautstärken die entsprechende Lautstärkeinformation durch die zunehmende Aktivität benachbarter Fasern übermittelt wird. Die *„Wahrnehmung der Lautstärke im Gehirn* [ist] *also eine Interpretation des Gehirns für die durchschnittlich ankommende Zahl von Impulsen auf dem auditorischen Nervenkanal“*.[17]

Neben der Frequenzselektivität nach dem Ortsprinzip findet eine Periodizitätsanalyse statt, ein Zählen der Schwingungsperioden der Trennwand: Ein Schwingungshoch auf der Basilarmembran veranlasst die entsprechenden Haarzellen, die an ihr anliegende Nervenfaser zu aktivieren; bei einem Schwingungstief tritt das Gegenteil ein. Auf diese Weise ist die Phase der ankommenden Schwingung an die Entladungsrate der Nervenfaser gekoppelt. Oberhalb von etwa 1000 Hz verliert sich diese Übereinstimmung, so dass dieses Erklärungsmodell nur im Frequenzbereich unter 1000 Hz treffend ist.[18] In vereinfachter Darstellung wird so die Phasenhäufigkeit gezählt, die der Frequenz des Schallstimulus entspricht.

---

[15] Zenner, Hans-Peter, a.a.O., S. 92.

[16] ebd.

[17] Hall, Donald E./Goebel, Johannes (Hg.), *Musikalische Akustik – Ein Handbuch*, Mainz 1997, S. 116.

[18] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 385.

1.4. Zentrales auditorisches Nervensystem

Im Hirn wird das über den Hörnerv ankommende Signal mittels eines hierarchischen Systems von synaptischen Neuronenverknüpfungen weiterverarbeitet. Auf den unteren Ebenen dieses Systems durchläuft es eine einfache Sinus-Signal-Kodierung, die vorgibt, dass – nach dem Ortsprinzip – bestimmte Schallfrequenzen an bestimmten Orten der Großhirnrinde angesiedelt sind.

> *„Man nimmt daher an, dass diese bis zur Hörrinde zunehmende Spezialisierung der Neurone auf bestimmte Eigenschaften des Schallreizes es erlaubt, Merkmalsmuster innerhalb des Schallreizes herauszuarbeiten und für die kortikale Beurteilung vorzubereiten. Nur der Informationsgehalt, nicht das ganze Signal, wird dadurch zum Kortex weitergeleitet (...).“*[19]

Die meisten höheren Neuronen haben also die Eigenschaft, nicht auf reine Sinustöne, sondern auf spezielle Merkmale eines Schallmusters zu reagieren. So gibt es Neuronen, die auf Frequenzmodulation reagieren, auf Amplitudenänderung oder auf zeitliche Kriterien eines Schallsignals.[20]

[19] Zenner, Hans-Peter, a.a.O., S. 118.

[20] ebd.

2. Leistungsvermögen des Gehörs

Das Leistungsvermögen des Gehörs bildet die Grundlage für die gehörangepasste Datenreduktion, die, ausgehend von den Erkenntnissen der Psychoakustik, für den Höreindruck irrelevante Informationen herausfiltert und unterdrückt.
Die Psychoakustik beschreibt nach ZWICKER[21] die Zusammenhänge zwischen Schallereignissen und den durch sie beim Menschen hervorgerufenen Hörempfindungen.
Für die Reduktion von irrelevanten Signalkomponenten sind das Auflösungsvermögen des Gehörs und dessen Verdeckungseigenschaften am wichtigsten. Das Auflösungsvermögen bezieht sich auf die Aspekte Tonhöhe, Lautstärke und Lautheit, zeitliche Auflösung und räumliches Hören. Die Verdeckungseigenschaften sind, abhängig von der Gestalt des maskierenden Schallereignisses, unterschiedlich ausgeprägt.

In Studien zum Leistungsvermögen des Gehörs wird durch Hörversuche die *Abstandsempfindung* bestimmt, also der kleinste gerade noch wahrnehmbare Unterschied zwischen zwei Klängen hinsichtlich Tonhöhe und Lautstärke. In der Literatur findet sich häufig anstelle des Begriffs Abstandsempfindung die Bezeichnung *JND*[22].

2.1. Tonhöhe

Das menschliche Gehör ist in der Lage, Schallsignale mit Frequenzen zwischen 20 Hz und 20 kHz wahrzunehmen. Personen mit einer Hörfähigkeit bis zu dieser oberen Grenze sind Ausnahmeerscheinungen; ein gesunder, junger Mensch

[21] Zwicker, Eberhard: *Psychoakustik*, Berlin/Heidelberg/New York 1982, S. 1.

[22] JND = just noticeable difference.

registriert im Durchschnitt Frequenzen bis etwa 17 kHz. Dieser Wert fällt auf durchschnittlich 12 kHz bei Frauen und 5 kHz bei Männern in der Altersgruppe der Sechzigjährigen.[23] Die Wahrnehmung tiefer Töne unterliegt nicht diesem Schwund; allerdings sind Töne unter 30 Hz nur bei starker Intensität und ohne störende Nebengeräusche hörbar.

Untersuchungen des frequenzbezogenen JND ergaben, dass er für Sinuswellen im Tonhöhenbereich unter 500 Hz ungefähr 1 Hz beträgt.[24] Oberhalb von 500 Hz steigt er an und ab 5 kHz nimmt die Steigung stärker zu, da die Tonhöhenunterscheidungsfähigkeit weiter abnimmt und ab 10 kHz nicht mehr vorhanden ist (s. Abb.7).

Betrachtet man den alltäglicheren Fall, nämlich den Umgang mit komplexen Wellen, so beträgt der JND im tiefen Frequenzbereich etwa 0,1 Hz. [25]

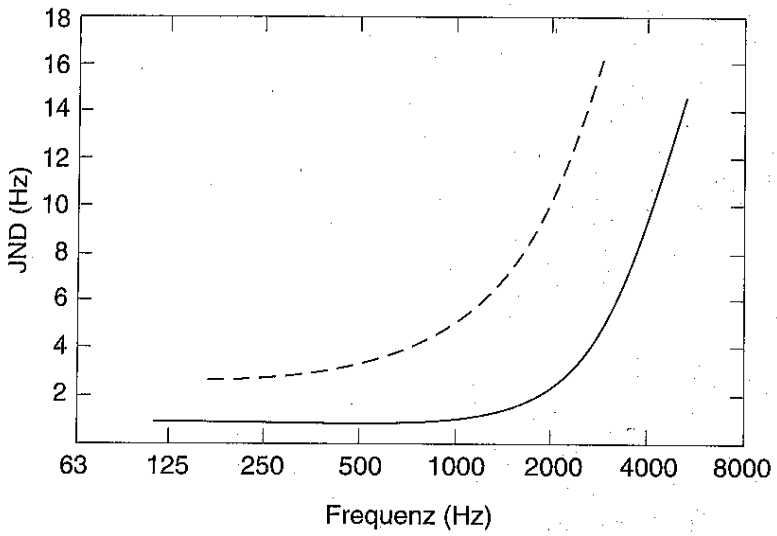


Abb. 7: Eben merkbare Unterschiede in der Frequenz in Abhängigkeit von der Anfangsfrequenz für Sinuswellen. Die gestrichelte Linie gilt für kaum noch hörbare Schallpegel, die durchgezogene für in der Musik übliche Pegel von 60 dB bis 90 dB.

[23] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 110.

[24] Zwicker, Eberhard/Fastl, Hugo: *Psychoacoustics – Facts and Models*, Berlin/ Heidelberg/New York 1990, S. 166.

[25] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 113.

2.2. Lautstärke

Tiefe und hohe Töne müssen für ihre Hörbarkeit mit wesentlich stärkerer Intensität erklingen als solche von mittlerer Tonhöhe: BORUCKI nennt einen Mindestschalldruck von 10^{-16} W/cm^2 für Frequenzen zwischen 1 kHz und 2 kHz; ein Ton mit 15 kHz benötigt dagegen eine Schallintensität von $5 \cdot 10^{-14}$ W/cm^2, um eine Schallempfindung auszulösen.[26] Trägt man die Schallempfindlichkeit gegen die Frequenz auf, stellt sich der höhere Intensitätsbedarf tiefer und hoher Töne hinsichtlich ihrer Wahrnehmbarkeit im Vergleich zu mittleren Tonhöhen um etwa 2000 Hz dar. Der Graph, der diesen frequenzspezifischen Sensibilitätsunterschied ausdrückt, zeigt die *Ruhehörschwelle* oder auch *Hörschwelle* (s. Abb. 8).[27] Der Bereich zwischen Ruhehörschwelle und *Schmerzgrenze* wird als *Hörfläche* bezeichnet.

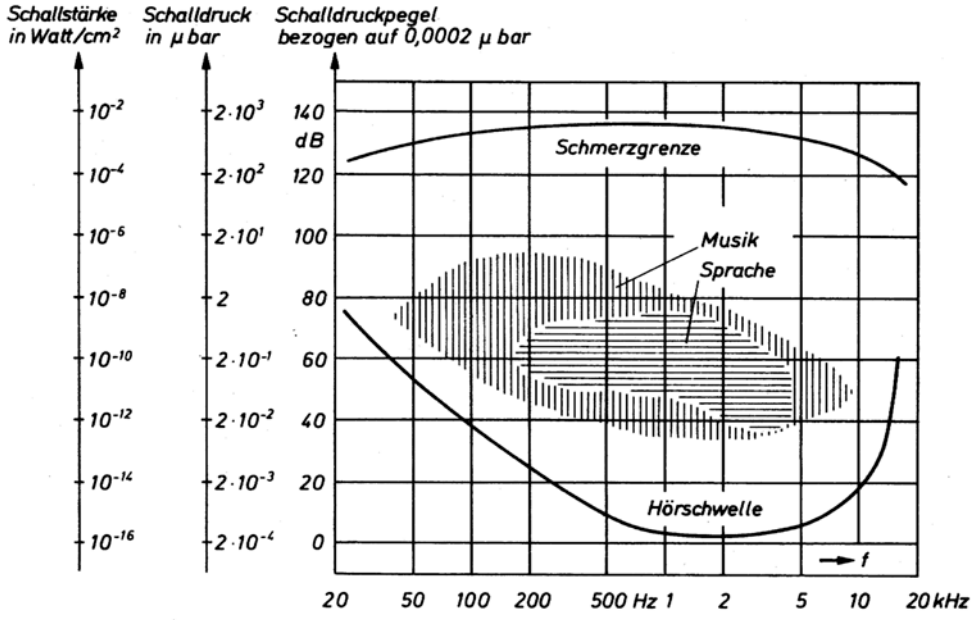


Abb. 8: Schallempfindlichkeit in Relation zur Frequenz

[26] Borucki, Hans: *Einführung in die Akustik*, Zürich 1980, S. 199.

[27] Die Bezeichnung „Hörschwelle“ ist insofern problematisch, als dass sie auch als Sammelbegriff für Ruhehör- und Mithörschwelle verwendet wird.

Da die Ergebnisse der Schallempfindlichkeitsuntersuchungen stark von individuellen Schwankungen und Beeinflussungen durch die Versuchsanordnung abhängen, stellt HALL verallgemeinernd fest, dass der Lautstärke-JND zwischen 0,5 dB und 1 dB liegt.[28]

Diese Verallgemeinerung vernachlässigt die starke Tonhöhenabhängigkeit der JND für Schallpegel, die BORUCKI anhand eines 1 kHz hohen Tons verdeutlicht: bei einem Ausgangspegel von 15 dB ist eine Zunahme um 1,5 dB gerade noch wahrnehmbar; ausgehend von 90 dB genügen dazu bereits 0,6 dB (s. Abb. 9).

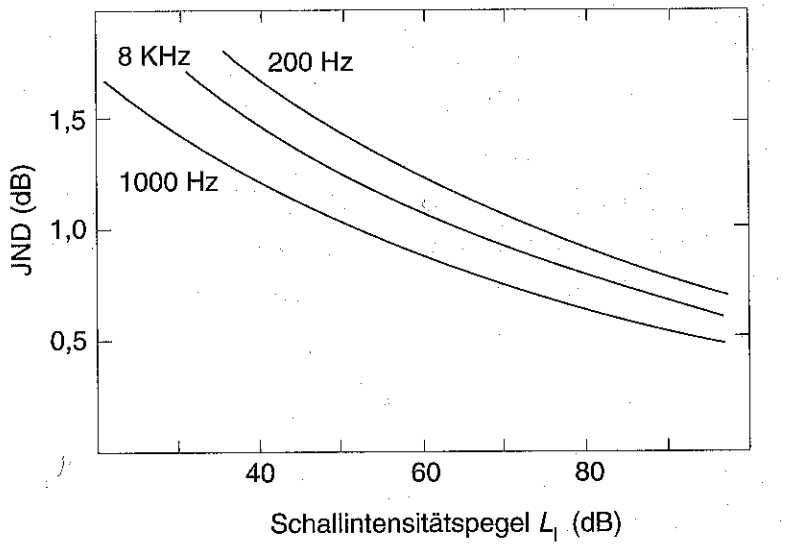


Abb. 9: Eben merkbarer Unterschied im Schallpegel L_I in Abhängigkeit des anfänglichen Schallpegels für Sinuswellen mit verschiedenen Frequenzen. Die 1000-Hz-Linie stellt nach HALL eine zulässige Annäherung an alle Frequenzen zwischen 500 Hz und 4000 Hz dar.

[28] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 113.

2.3. Kritische Frequenzbandbreite

Versuche haben gezeigt, dass zwei Töne mit nahe beieinanderliegender Frequenz ungefähr den gleichen Bereich der Basilarmembran, zwei Töne in größerem Abstand voneinander zwei getrennte Regionen anregen.[29] Der Frequenzbereich, in dem eine Überlappung der Reaktionskurven der Basilarmembran stattfindet, wird als *kritische Frequenzbandbreite* oder *Frequenzgruppe* bezeichnet: Wird die Bandbreite bei gleichbleibender Mittenfrequenz vergrößert, verändert sich ab einem bestimmten Moment die subjektive Wahrnehmung.

Die kritische Bandbreite beträgt nach ZWICKER etwa 100 Hz unterhalb von 500 Hz und nimmt oberhalb von 500 Hz proportional zur Frequenz um ein Fünftel der Mittenfrequenz zu.[30]

Auch HALL nennt als Annäherung für die kritische Bandbreite bei Mittenfrequenzen über 500 Hz eine Spanne von 15% bis 20% der Mittenfrequenz; dies entspricht etwa 2½ bis 3 Halbtönen.[31] Er skizziert als Beispiel (s. Abb.10):

> *„Nehmen wir an, Sie hören zunächst ein schmalbandiges Rauschen, z B. eine Mischung aller Frequenzen von 980 bis 1020 Hz, was wir mit den Angaben Mittenfrequenz=1000 Hz und Bandbreite=40 Hz beschreiben können. Nun wird die Bandbreite allmählich vergrößert, während Mittenfrequenz und Intensität gleich bleiben (...). Eine Zeitlang bleibt die wahrgenommene Lautstärke unverändert, aber sobald die Bandbreite ungefähr den Wert von 160 Hz überschreitet, nimmt die subjektive Lautstärke plötzlich zu.“*[32]

---

[29] Moore, Brian C. J.: *An Introduction to the Psychology of Hearing*, London/San Diego/New York ^{3}1992 [^{1}1977], S. 85.

[30] Zwicker, Eberhard/Feldtkeller, Richard: *Das Ohr als Nachrichtenempfänger*, 2. neubearbeitete Aufl., Stuttgart 1967, S. 134.

[31] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 392.

[32] ebd.

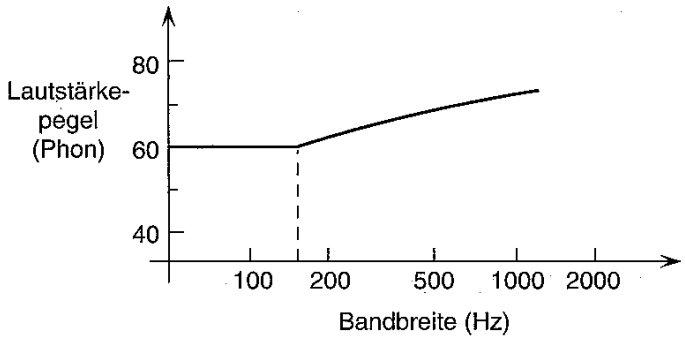


Abb. 10: Änderung der wahrgenommenen Lautstärke von schmalbandigem Rauschen bei Erhöhung der Bandbreite und konstanter Gesamtintensität

Die kritische Bandbreite wirkt sich auch hinsichtlich der Tonhöhenunterscheidung aus: Bei zwei gleichzeitig wahrnehmbaren Sinuswellen mit der konstanten Frequenz f_1 und der variablen Frequenz f_2, die sehr nahe beieinander liegen, hört man deutliche Schwebungen, die mit zunehmendem Abstand von f_2 zu f_1 schneller werden, um schließlich in Klangrauhigkeit überzugehen. Die Schwebungen werden wegen ihrer Häufigkeit nicht mehr als solche wahrgenommen. Dies gilt zu beiden Seiten der Frequenz f_1, also sowohl für $f_2>f_1$ als auch $f_2<f_1$. Innerhalb des Bereichs bis zum Erreichen der kritischen Bandbreite wird nur eine Tonhöhe wahrgenommen, die ungefähr dem arithmetischen Mittel von f_2 und f_1 entspricht, und erst außerhalb hört man zwei separate Tonhöhen (s. Abb.11).

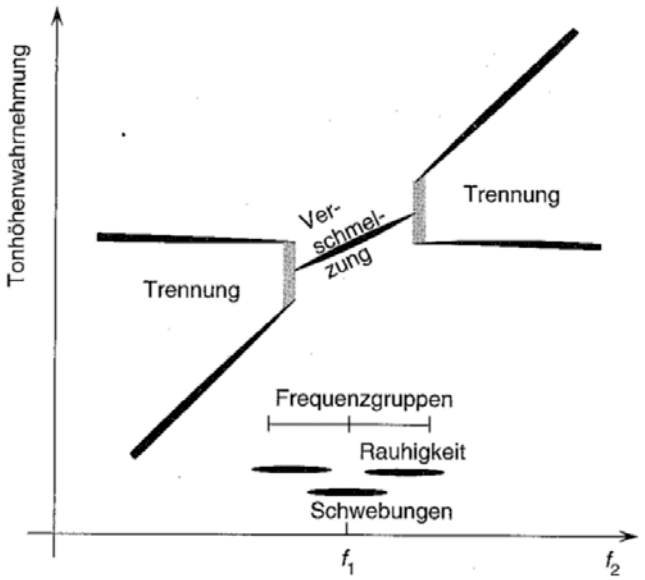


Abb. 11: Auswirkungen der kritischen Bandbreite auf die Tonhöhenwahrnehmung

Da die Einteilung des Hörspektrums in Frequenzgruppen für die Erklärung psychoakustischer Phänomene große Bedeutung hat, werden diese zur einheitlichen Handhabung anhand der Einheit *Bark* kategorisiert. Die Unterteilung beruht darauf, dass das menschliche Gehör das breite Frequenzspektrum in Abschnitte unterteilt, die mit den Frequenzgruppen korrespondieren. Fügt man die einzelnen Frequenzgruppen mit der Mittenfrequenz f_c so hintereinander, dass die obere Grenze f_u der einen gleichzeitig die untere Grenze f_l der nächsthöheren Frequenzgruppe ist, erhält man daraus die nach BARKHAUSEN benannte *Bark-Skala*. Die Spanne von einer Grenze zur nächsten gibt die kritische Bandbreite Δf_G an (s. Tab. 1).

z	f_l, f_u	f_c	z	Δf_G	z	f_l, f_u	f_c	z	Δf_G
Bark	Hz	Hz	Bark	Hz	Bark	Hz	Hz	Bark	Hz
0	0				12	1720			
		50	0.5	100			1850	12.5	280
1	100				13	2000			
		150	1.5	100			2150	13.5	320
2	200				14	2320			
		250	2.5	100			2500	14.5	380
3	300				15	2700			
		350	3.5	100			2900	15.5	450
4	400				16	3150			
		450	4.5	110			3400	16.5	550
5	510				17	3700			
		570	5.5	120			4000	17.5	700
6	630				18	4400			
		700	6.5	140			4800	18.5	900
7	770				19	5300			
		840	7.5	150			5800	19.5	1100
8	920				20	6400			
		1000	8.5	160			7000	20.5	1300
9	1080				21	7700			
		1170	9.5	190			8500	21.5	1800
10	1270				22	9500			
		1370	10.5	210			10500	22.5	2500
11	1480				23	12000			
		1600	11.5	240			13500	23.5	3500
12	1720				24	15500			
		1850	12.5	280					

Tab. 1: Bark-Skala nach ZWICKER

2.4. Zeitliches Auflösungsvermögen

Die kürzeste Dauer eines wahrnehmbaren Impulses beträgt 2 ms, doch werden Stimuli, die kürzer als 5 ms dauern, vom Gehör noch nicht in Frequenzgruppen zerlegt. Die unterscheidbaren Frequenzstufen werden bei ansteigender Dauer kleiner, bis ab einer Signaldauer von 500 ms das Auflösungsvermögen nicht weiter zunimmt.[33]

Ein Schallereignis muss so lange anhalten, bis seine einzelnen Schwingungskomponenten ihre Sinusförmigkeit angenommen haben.[34] Ist das Signal zu kurz – ist also die Mindest-Kennzeit des betreffenden Tons unterschritten –, prägen sich diese sinusförmigen Schalländerungen nicht aus; auf das Gehör wirkt ein komplexes Frequenzspektrum ein, das als Knacken registriert wird. Im Bereich von 10000 Hz bis etwa 400 Hz beträgt die Kennzeit ungefähr 4 ms bis 10 ms und steigt bei tiefen Tönen mit einer Frequenz von weniger als 100 Hz auf über 30 ms an.[35]

2.5. Lautheit

Zur besseren Umschreibung eines möglichen Versuchsergebnisses „Ton 1 ist doppelt so laut wie Ton 2“ wird die *Lautheit* anhand der *Sone*-Skala gemessen. Per Definition hat ein 1000 Hz-Ton mit einer Schallintensität von 40 dB einen Lautheitswert von 1 sone. Experimentell wird nun die Intensität des 1 kHz-Tons

---

[33] Feiten, Bernhard: *Beurteilung von Quellenkodierungsverfahren für Audiosignale bei Berücksichtigung der Verdeckungseigenschaften des Gehörs*, Dissertation an der Technischen Universität Berlin 1989, S. 23.

[34] Nach Fourier kann jede stetige, nichtsinusförmige, periodische Schwingung durch Sinus- und Cosinus-Schwingungen dargestellt werden. Vgl. Franz, Dieter: *Elektroakustik – Grundlagen der Schallverarbeitung praxisnah dargestellt*, München 1990, S. 34.

[35] Webers, Johannes: *Tonstudiotechnik – Schallaufnahme und -wiedergabe bei Rundfunk, Fernsehen, Film und Schallplatte*, München ^{5}1989 [^{1}1968], S. 98.

bestimmt, den ein Proband als halb, doppelt oder viermal so laut empfindet. Die Lautheitswerte für 1000 Hz-Töne lauten[36]:

LAUTHEIT	INTENSITÄT
Halb so laut: **0,5 sone**	33 dB
(*Def.*: Lautheit **1 sone**)	40 dB
Doppelt so laut: **2 sone**	48 dB
Viermal so laut: **4 sone**	55 dB

Tab. 2: Lautheitswerte eines 1 kHz-Tons

Eine weitgehende Übereinstimmung individueller Testergebnisse erlaubt diese Kategorisierung von subjektiven Zuordnungen:

> *„Es gibt durchaus eine Näherungs-Skala für die Wahrnehmung von Lautstärke, mit der viele Leute übereinstimmen können; die Wahrnehmung der Lautstärke ist somit nicht ganz und gar verschieden von Individuum zu Individuum.“*[37]

Derartige Untersuchungen beziehen sich auf die Lautheitswahrnehmung von einzelnen Sinuswellen. Der überwiegende Teil der Studien zu komplexen Wellenformen kommt zu gleichgearteten Ergebnissen, die HALL wie folgt interpretiert:

> *„Der von einer Frequenzgruppe (=kritische Frequenzbandbreite) gelieferte Eindruck von Lautheit hängt ausschließlich von der in diesem Bandbereich empfangenen Energie ab; zwei ausreichend gut getrennte Bandbreiten liefern jeweils einen unabhängigen Beitrag zur Gesamt-Lautheit.“*[38]

Also werden die Intensitäten der Komponenten innerhalb einer kritischen Bandbreite addiert; außerhalb dieser Grenzen addiert sich deren wahrgenommene Lautheit. Beispielsweise erreicht eine Sinuswelle mit f_1=1000

[36] Borucki, Hans, a.a.O., S. 212.

[37] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 117.

[38] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 397.

Hz und Lautstärkepegel L_1=70 phon, gekoppelt mit einer zweiten mit f_2=1050 Hz und Lautstärkepegel L_2=65 phon, eine Lautstärke von 71 phon, die einer Lautheit von etwa 15 sone entspricht. Verlässt jedoch f_2 die kritische Bandbreite mit f_2=1500 Hz, haben beide Sinuswellen eine Gesamtlautheit von 23 sone.[39]

2.6. Räumliches Hören

Das Gehör nutzt nach HALL mindestens vier simultan arbeitende Mechanismen zur Bestimmung der Position einer Schallquelle: Ausgewertet werden wahrgenommene *Laufzeitunterschiede*, *Phasenunterschiede* und *Intensitäts-unterschiede.* Diese ersten drei Mechanismen werden als *binaurale Hörmechanismen* bezeichnet, da sie auf dem Zusammenspiel beider Ohren beruhen.

Mithilfe dieser Mechanismen bestimmt das Gehör die laterale – nach rechts oder links versetzte – Position der Schallquelle bis auf 5° genau.[40]

Eine seitlich ankommende Schallwelle muss bis zum gegenüberliegenden Ohr einen längeren Weg zurücklegen, so dass sie zeitverzögert wahrgenommen wird. Dieser Laufzeitunterschied beträgt maximal 0,6 ms für den Fall, dass der Schall genau von der Seite auftrifft. Laufzeitunterschiede werden nur bei Klängen von weniger als 100 ms Dauer oder bei Einschwingphasen längerer Klänge erkannt.[41]

Befindet sich die Schallquelle also seitlich des Kopfes, so wird sie zusätzlich durch die Wahrnehmung von Phasenunterschieden lokalisiert, die auf der Distanz der beiden Ohren zur Schallquelle beruhen. Dieser Mechanismus deckt jedoch nur Frequenzen bis etwa 1500 Hz ab, da bei höheren Klängen die

[39] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 397f., vgl. ders.: a.a.O., Tab. 5.1, S. 93.

[40] Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 262.

[41] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 343.

Hörnervenzellen nicht schnell genug feuern können, um den Phasenunterschied noch wiederzugeben.[42]

Der dritte Mechanismus nimmt interaurale Intensitätsunterschiede wahr. Die Schallfrequenz muss hierzu allerdings oberhalb von 2 kHz bis 3 kHz liegen, da tiefere Klänge wegen ihrer größeren Wellenlänge den Kopf so ‚umwandern', dass die Intensität an beiden Ohren übereinstimmt, auch wenn der Schall seitlich auftrifft.

Der vierte Mechanismus nutzt die Form des Außenohres, durch die die Schallwellen – besonders solche mit einer Frequenz von über 5000 Hz – in Abhängigkeit von ihrer Auftreffrichtung gebündelt und verstärkt werden, so dass „*die relative Stärke von hochfrequenten Schallanteilen sich zu ändern* [scheint], *wenn diese von vorn nach hinten und zurück wandern*" [43].

2.7. Maskierung: Verdeckungseigenschaften des Gehörs

Ein weiteres Phänomen, das im Zusammenhang mit der Wahrnehmung von Lautheitsunterschieden steht und einen wichtigen Aspekt dieser Arbeit darstellt, ist das der *Maskierung* oder *Verdeckung*. Hiermit wird der Vorgang bezeichnet, in dem „*ein Stimulus die Antwort auf einen anderen zum Verschwinden bringt oder verkleinert*."[44]

Die Wahrnehmung vollkommener Stille ist nur unter bestimmten, aufwendigen Versuchsbedingungen herzustellen: in der Regel ist der Mensch von Störschall („*Rauschen, Klang oder ein Ton*" [45]) umgeben, der die Hörschwelle anhebt. Diese ist der Hörsituation angepasst und heißt *Mithörschwelle*: Sie gibt den Pegel des Schalldrucks an, den ein hinzutretender Stimulus haben muss, um

[42] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 343.

[43] ebd.

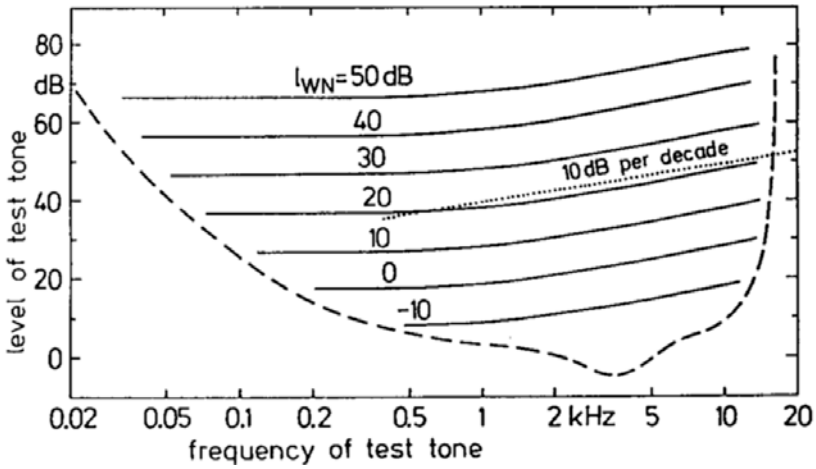
[44] Zenner, Hans-Peter, a.a.O., S. 29.

[45] Zwicker, Eberhard/Feldtkeller, Richard, a.a.O. , S. 55.

eben hörbar zu sein. Bleibt ein Testton unterhalb dieser durch das Störsignal vorgegebenen Mithörschwelle, so wird er durch den Störschall *verdeckt* (*maskiert*). Die im Folgenden beschriebenen Ergebnisse von ZWICKER und FASTL[46] spiegeln die Erkenntnisse über Verdeckungseigenschaften des Gehörs wider.

2.7.1. Maskierung eines Tons durch stationäres Weißes Rauschen[47]

Weißes Rauschen ist definiert als ein Rauschen, bei dem alle Frequenzen gleiche Amplitudenwerte haben. Das Frequenzspektrum umfasst, unter Berücksichtigung des menschlichen Hörspektrums, den Bereich von 20 Hz bis 20 kHz. Das Weiße Rauschen ließe sich als eine horizontale Gerade im Frequenzspektrum veranschaulichen. Man könnte nun vermuten, dass sich die



Hörschwelle entsprechend diesem Frequenzspektrum parallel verschoben erhöht.

Abb. 12: Mithörschwelle eines Tons, eben verdeckt durch Weißes Rauschen mit dem Pegel l_{WN}. Die gestrichelte Linie entspricht der Ruhehörschwelle

[46] Zwicker, Eberhard/Fastl, Hugo: *Psychoacoustics – Facts and Models*, Berlin, Heidelberg, 1990.

[47] Vgl. Zwicker, Eberhard/Fastl, Hugo, a.a.O. , S. 57–58. Im Folgenden beziehen sich die Quellenangaben hinter den Kapitelüberschriften 2.7.1. bis 2.7.6. auf das gesamte jeweilige Kapitel.

Die tatsächlichen Mithörschwellen verlaufen jedoch nur unterhalb von etwa 500 Hz in horizontaler Richtung. Oberhalb der 500 Hz steigt die Mithörschwelle aber um 10 dB pro Verzehnfachung der Frequenz (s. Abb. 12, gepunktete Linie). Die Mithörschwelle des Testtons liegt für tiefe Frequenzen um etwa 17 dB oberhalb der *Schallintensitätsdichte* (l_{WN}) des Weißen Rauschens. Eine Erhöhung des Pegels um 10 dB hat eine Anhebung der Mithörschwelle um ebenfalls 10 dB zur Folge. Bei ganz tiefen und ganz hohen Frequenzen sind Mithörschwelle und Ruhehörschwelle identisch.

2.7.2. Maskierung eines Tons durch stationäres Schmalbandrauschen[48]

In seinen Versuchen verwendet ZWICKER Rauschen, dessen Bandbreite kleiner als die kritische Bandbreite oder maximal mit dieser identisch ist (ca. 100 Hz unter bzw. 20% der Mittenfrequenz über 500 Hz; s. Kap. 2.3.). Seine Untersuchungen mit Schmalbandrauschen als Maskierer lassen sich in zwei Kategorien unterteilen, abhängig von den Parametern Störpegel und Mittenfrequenz:

a) Bei konstantem Störpegel L=60 dB werden die Mithörschwellen von Testtönen untersucht, die maskiert werden durch Schmalbandrauschen mit den Mittenfrequenzen 250 Hz, 1 kHz und 4 kHz und den jeweiligen kritischen Bandbreiten 100 Hz, 160 Hz und 700 Hz (vgl. Tab. 1, S. 19).
b) Als Maskierer dient Rauschen mit der Mittenfrequenz 1 kHz und der Bandbreite 160 Hz, dessen Pegel L_{CB} zwischen 20, 40, 60, 80 und 100 dB variiert.

Abb. 13 veranschaulicht die Beobachtungen zu Versuch a): Das Maximum der Mithörschwelle scheint für f_c=250 Hz am breitesten zu sein, während die mit f_C=1 kHz und f_C=4 kHz korrespondierenden Mithörschwellen stärkere Ähnlichkeit haben. Außerdem vergrößern sich mit steigender Mittenfrequenz f_c die

[48] Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 59–60.

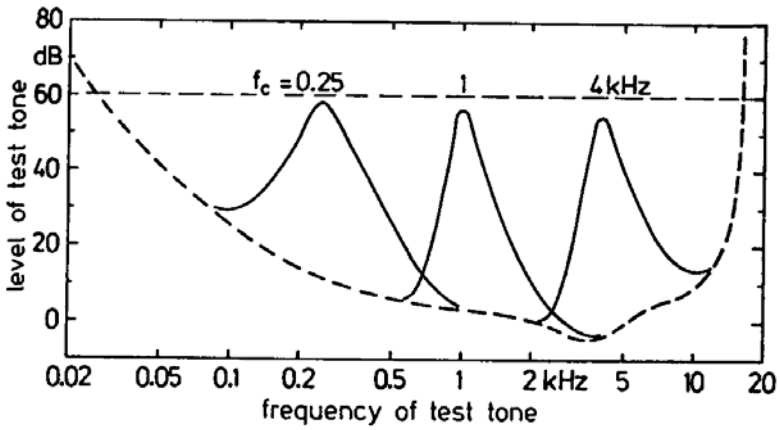


Abb. 13: Mithörschwellen eines Testtons, eben verdeckt durch Schmalbandrauschen mit der Mittenfrequenz f_C. Die gestrichelte Linie entspricht der Ruhehörschwelle

Abstände zwischen dem Störpegel L=60 dB und den Maxima der jeweiligen Mithörschwelle: die Differenzen liegen bei 2 dB für f_c=250 Hz, 3 dB für f_c=1 kHz und 5dB für f_c=4 kHz (s. Abb. 13). Schließlich fällt auf, dass alle drei Mithörschwellen zu ihrem jeweiligen Maximum sehr steil ansteigen und, sobald sie es überschritten haben, weniger steil abfallen.

Versuch a) und b) lassen sich über den Einzelfall f_c=1 kHz und L_{CB}=60 dB als tertium comparationis verknüpfen (s. Abb. 14). Der Anstieg hin zum Maximum, das jeweils bei L_{CB}-3dB erreicht wird, verläuft in allen Fällen fast parallel und folglich mit gleicher Steilheit und klingt in umgekehrt proportionalem Verhältnis zu den Pegeln L_{CB} ab: je kleiner diese werden, desto steiler das Gefälle, je höher

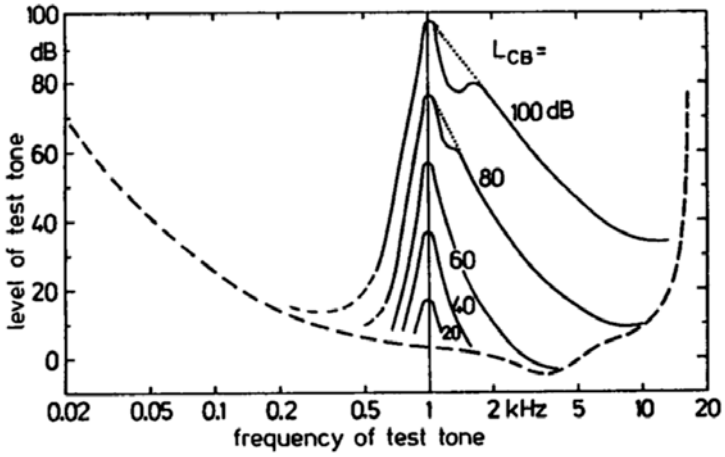


Abb. 14: Mithörschwellen eines Testtons, eben verdeckt durch Schmalbandrauschen mit f_c=1 kHz und verschiedenen Störpegeln L_{CB}

die Pegel, desto flacher läuft die Mithörschwelle aus.
Die Einbuchtungen der Mithörschwellen zu den Störpegeln von 80 dB und 100 dB (s. Abb. 14) führt ZWICKER auf non-lineare Effekte des Gehörs zurück, die zu einem Kombinationsrauschen führen, das aus Testton und Schmalbandrauschen resultiert: Die Testperson nimmt eine Veränderung des Rauschbildes und nicht des eigentlichen Testtons wahr. Dieser wird erst hörbar, wenn sein Lautstärkepegel die in Abb. 14 gepunktet dargestellten Werte erreicht.

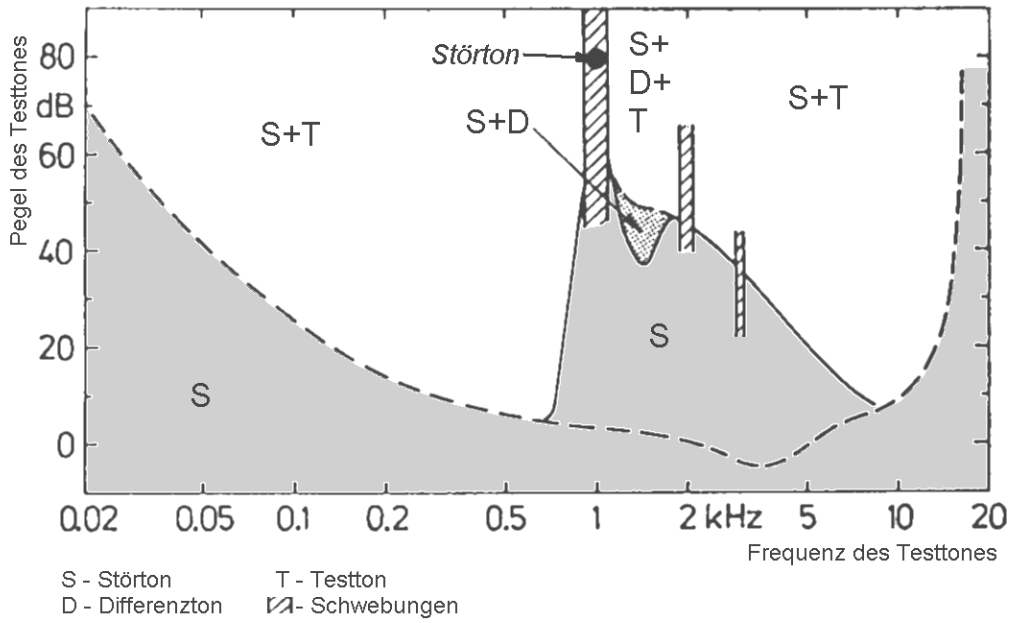
2.7.3. Maskierung eines Tons durch einen anderen, stationären Ton[49]

Maskierer ist ein Sinuston der Frequenz 1 kHz und dem Schallpegel L_M=80 dB. Bleibt die Frequenz des Testtons unterhalb von 500 Hz, so ist die Mithörschwelle identisch mit der Ruhehörschwelle (gestrichelte Linie in Abb. 15). Ab einer Frequenz von rund 700 Hz steigt die Mithörschwelle des Testtons steil an und liegt weit über der Hörschwelle. Auch bei höheren Testfrequenzen bis 9 kHz lässt sich der Schalldruck des Testtons weit über die Ruhehörschwelle steigern. Dennoch ist nur der Störton allein zu hören. Erhöht man den Pegel nun weiter, tritt im Bereich von 1 kHz bis 2 kHz neben dem Störton ein Differenzton auf, dessen Frequenz geringer ist als die des Testtons. Diese Situation entspricht der schwarz schraffierten Fläche in Abb. 15.
Im Frequenzbereich um 1,4 kHz (s. Abb. 15, Minimum der schwarz schraffierten Fläche) hörten die Testpersonen bei relativ geringem Testtonpegel von 40 dB einen *kubischen Differenzton*[50] mit der Frequenz 600 Hz. Bei weiterer Pegelanhebung kommt zum Störton und zum Differenzton der Testton selbst hinzu.

---

[49] Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 61–65.

[50] $f_{diff} = 2f_1-f_2$ mit $f_1<f_2$; vgl. Parncutt, Richard: Art. *Leistungen des Gehörs*, in: Musikpsychologie: ein Handbuch (Bruhn, Herbert/Oerter, Rolf/Rösig, Helmut Hg.), Reinbek bei Hamburg 1993, S. 674.



Hier zeigt sich, dass Sinustöne als Maskierer die Hörschwelle in komplizierterer Form beeinträchtigen als Rauschen. Auch die Beurteilung von Mithörschwellen

Abb. 15: Mithörschwelle eines Tons, eben verdeckt durch einen Störton mit einer Frequenz von 1 kHz bei einem Pegel von 80 dB

wird problematischer, da die Testperson auf jede Veränderung des Klangbildes reagiert, die über das des reinen Störtons hinausgeht.

Diese wahrgenommene Veränderung muss sich jedoch nicht zwingend auf den Testton selbst beziehen, sondern kann auch aus anderen physikalischen Gesetzmäßigkeiten resultieren. Hat der Testton z.B. die Frequenz 990 Hz bei einem Pegel von 60 dB, so wird ein deutlich hörbares Pulsieren der Frequenz 10 Hz registriert. Nähert sich der Testton der 1000 Hz-Marke des Maskierers weiter an, machen sich Schwebungen bemerkbar. Die Testperson reagiert auf das hinsichtlich des reinen Maskiertons veränderte Hörbild, obwohl das eigentliche Kriterium, das Hinzutreten des Testtons zum Störton, nicht erfüllt ist. Derselbe Effekt tritt ein, sobald der Testton den Bereich um 2 kHz und 3 kHz erreicht. Diese Bereiche sind in Abb. 15 durch Schraffierung gekennzeichnet.

2.7.4. Maskierung eines Tons durch komplexe, stationäre Töne[51]

Maskierungsphänomene treten im Alltag kaum in der Form auf, dass ein reiner Ton einen anderen maskiert. Diesem Umstand wird im Folgenden Rechnung getragen.

Als Maskierer dienen in diesem Versuch ein 200 Hz-Ton als Grundton und die folgenden neun Elemente der Partialtonreihe über diesem. Der Schallpegel der einzelnen Teiltöne ist gleich und liegt bei je 40 dB. In einer zweiten Versuchsreihe wird der Pegel auf 60 dB erhöht (s. Abb. 16).
Der Kurvenverlauf zeigt deutliche Analogien zu dem der Maskierungseffekte durch Schmalbandrauschen: Die Mithörschwelle zu denjenigen Frequenzen, welche die des höchsten Teiltons übersteigen, ist bei höheren Störpegeln flacher ausgeprägt als bei geringeren Pegeln, d.h. die Mithörschwelle geht erst bei höheren Testtonfrequenzen in die Ruhehörschwelle über.

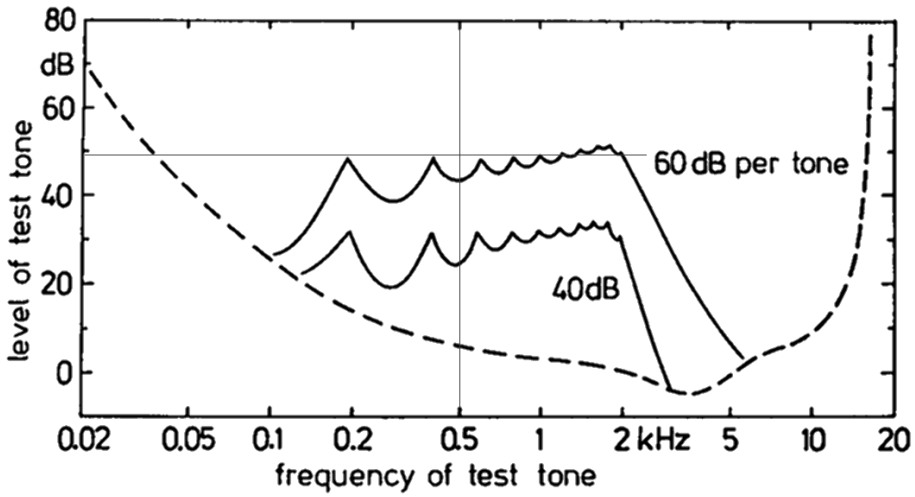


Abb. 16: Mithörschwelle eines Tons, eben verdeckt durch neun Obertöne mit gleichem Schallpegel über der Grundtonfrequenz 200 Hz. (Die Hilfsgeraden werden zu einem späteren Zeitpunkt erläutert.)

[51] Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 65–68.

Da die oberen Partialtöne auf der logarithmischen Skala mit zunehmender Frequenz immer dichter nebeneinander liegen, prägen sich die Minima dazwischen entsprechend weniger stark aus. Aufgrund der Tatsache, dass in der Musik viele komplexe Klänge aus einzelnen Partialtönen zusammengesetzt sind, äußern ZWICKER und FASTL die Vermutung, dass diese Klänge Mithörschwellen hervorrufen, die mit den in Abb. 16 dargestellten vergleichbar sind. Lediglich die ‚Spitze-Tal'-Ausprägung variiert, da mit steigender Teiltonanzahl die ‚Spitzen' zunehmen und folglich die ‚Täler' weniger deutlich auftreten.

2.7.5. Maskierung durch instationäre Schallereignisse[52]

Zusätzlich zu den bisher beschriebenen Maskierungsphänomenen in Abhängigkeit vom Frequenzspektrum hängt die Verdeckung bei instationären[53] Schall-ereignissen von der Zeit ab.

Abb. 17 zeigt anhand eines Tons, der zum Zeitpunkt t=0 ms eingeschaltet und nach 200 ms ausgeschaltet wird, den Verdeckungsverlauf mit den Phasen *Vorverdeckung*, *Simultanverdeckung* und *Nachverdeckung*.
Bereits vor dem Einschalten eines Maskierers erhöht sich die Hörschwelle, daher spricht man vom Phänomen der Vorverdeckung. Dieser Vorgang beginnt etwa 20 ms vor dem Einsatz des Störschalls. Jedoch ist die Abhängigkeit des Vorverdeckungsmaßes von Störschalldauer und -pegel nicht erklärt.

ZWICKER warnt davor, aus dem Phänomen der Vorverdeckung den Schluss zu ziehen, das Gehör könne ‚in die Zukunft' hören; vielmehr hängt die

[52] Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 72–77.

[53] In den bisher beschriebenen Versuch wurde mit Dauersignalen, also stationären Ereignissen gearbeitet.

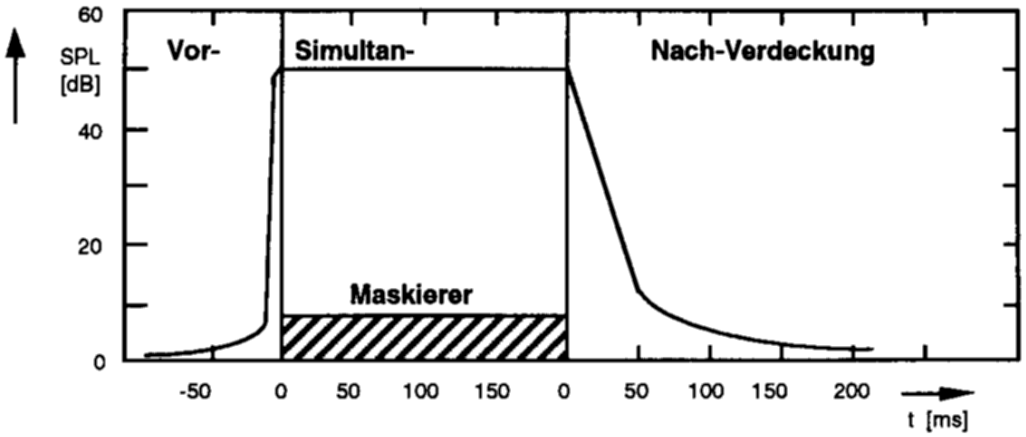


Abb. 17: Schematische Darstellung des zeitlichen Verlaufs bei Maskierung durch instationäre Stimuli

Vorverdeckung damit zusammen, „(...) *dass die Empfindungen nicht sofort auftreten, sondern Verarbeitungszeiten notwendig sind, um sie zu erzeugen.*“[54]

Der Vorgang der Simultanverdeckung beschränkt sich auf die Zeitspanne, in der das Störsignal tatsächlich eingeschaltet ist. Die Verdeckungseigenschaften von Störtönen mit einer Dauer von mehr als 200 ms entsprechen während dieser Phase denen stationärer Töne (s. Kap. 2.7.1. – 2.7.4.). Dauert der Stimulus weniger als 200 ms an, erhöht sich die Hörschwelle im umgekehrt proportionalen Verhältnis zur Tondauer: je kürzer die Tondauer, desto stärker steigt die Mithörschwelle an.

Der Nachverdeckungseffekt hängt von der Dauer (t) des Maskierers ab. Ein Störton mit t=200 ms erzeugt eine später abklingende Mithörschwelle als ein 5 ms langer Ton und geht später als dieser wieder in die Ruhehörschwelle über.

[54] Zwicker, Eberhard a.a.O., S. 94

2.7.6. „Addition“ mehrerer Maskierer[55]

Auch die Verdeckungseigenschaften eines einzigen Störstimulus entsprechen nicht den Anforderungen, denen das Gehör in der Realität ausgesetzt ist. Vielmehr ist die Hörwirklichkeit durch komplexe Gebilde bestimmt, deren Elemente, Einzelschwingungen, in einer bestimmten Hierarchie hinsichtlich ihrer maskierenden Wirkung bzw. ihrer Verdeckung simultan präsent sind.

ZWICKER und FASTL bestätigen ihre Vermutung, dass der Zuwachs der Gesamtmithörschwelle ΔL_T maximal ist, wenn die Mithörschwellen der einzelnen Störtöne identisch sind, also die Differenz $\Delta L_{T(1-2)}$ gleich Null ist.

Im einfachsten Fall sind die Störtöne selbst identisch: jeder dieser Töne, isoliert erklingend, bewirkt dieselbe Mithörschwellenänderung. Folglich ist die Gesamtmithörschwelle maximal.

Darüber hinaus existieren Sonderfälle von Störtonkombinationen, in denen die Mithörschwellen übereinstimmen, obwohl die Maskierer verschieden sind. Dies ist der Fall, wenn ein Ton mit etwa 800 Hz und Schmalbandrauschen als Maskierer verwendet werden.

Versuche haben gezeigt, dass bei sich verringerndem Abstand der Einzelmithörschwellen die Gesamthörschwelle um bis zu 12 dB nach oben verschoben sein kann (s. Abb.18).

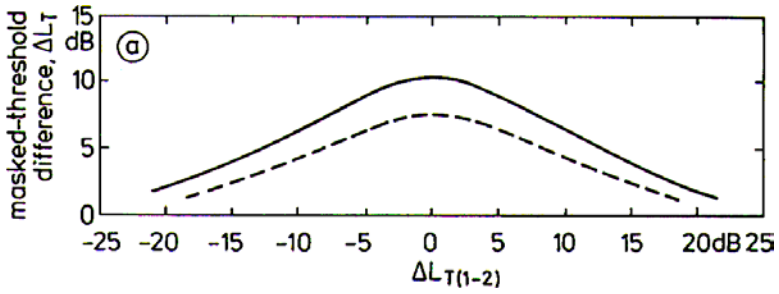


Abb. 18: Addition von Verdeckungseffekten
Die durchgezogene Linie resultiert aus Mithörschwellen mit mehr als 15 dB Abstand zur Ruhehörschwelle, die gestrichelte aus dem Abstand von 5 bis 15 dB.

[55] Zwicker, Eberhard/Fastl, Hugo, a.a.O., S. 95–99.

2.7.7. Erklärungsmodell für Maskierungseffekte

HALL differenziert den Grad der Maskierung nach dem Kriterium ‚innerhalb und außerhalb der kritischen Bandbreite': innerhalb dieser Grenzen darf ein Ton eine bestimmte Lautstärke gegenüber einem anderen nicht unterschreiten, um nicht maskiert zu werden. Verlässt der zweite, hinzutretende Ton jedoch die Frequenzgruppe des ersten, erfolgt die Informationsübermittlung an das Gehirn über verschiedene Nervenkanäle, so dass beide Töne wahrgenommen werden.[56] Eine Erklärung liegt im *„Leitung-besetzt-Effekt"*[57], so die Hypothese PICKLES': ein Stimulus, der an den entsprechenden Haarzellen eine maximale Entladungsrate der Nervenfasern hervorruft, blockiert einen hinzutretenden schwachen Stimulus, da die Entladungsrate nicht mehr gesteigert werden kann.
Auffällig ist, dass sowohl in dem Abb. 12 zugrundeliegenden Versuch als auch in dem zu Abb. 16 bis zur Frequenz von 500 Hz die Mithörschwellen horizontal verlaufen, oberhalb von 500 Hz jedoch ansteigen. ZWICKER und FELDTKELLER schließen daraus, „dass *das Gehör durch Geräusche unterhalb 500 Hz entsprechend der absoluten Schallintensitätsdichte und oberhalb 500 Hz entsprechend der relativen Schallintensitätsdichte gestört wird."*[58]

HALL beschreibt den Effekt der ‚Aufwärts-Maskierung', nach dem *„ein zweiter Ton stärker maskiert wird, wenn seine Frequenz höher als die des maskierenden ist, und umgekehrt."*[59] Der Grund hierfür liegt darin, dass ein maskierender tiefer Ton im Gegensatz zu einem hohen, an der Schneckenbasis lokalisierten Ton größere Teile der Basilarmembran in Schwingung versetzt:

---

[56] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 399.

[57] Pickles, James O.: *An Introduction to the Physiology of Hearing*, London 1998, S. 104.

[58] Zwicker, Eberhard/Feldtkeller, Richard, a.a.O., S. 70.

[59] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 399.

„Daher sind mehr ‚Ausgabe-Nervenkanäle' beschäftigt, wenn der maskierende Ton eine tiefere Frequenz hat, und weniger, wenn sie höher ist."[60]

Alle diese Erkenntnisse über die Ursachen und die Auswirkungen bestimmter Verdeckungseigenschaften des Gehörs bilden die Voraussetzung für die effektive Reduktion von Audiodaten auf psychoakustischer Grundlage. Da die Forschungsergebnisse auf der ceteris-paribus-Klausel basieren, also individuelle Unterschiede im Hörvermögen nicht berücksichtigen und daher allgemeingültigen Charakter haben, lassen sie sich durch entsprechende Algorithmen für die digitale Signalverarbeitung nutzbar machen.

[60] Hall, Donald E./Goebel, Johannes (Hg.), a.a.O., S. 399.

Teil II: Digitale Tonaufzeichnung

3. Digitale Signalverarbeitung

Die digitale Tontechnik ermöglicht, nicht das analoge Signal selbst, sondern seine Beschreibung zu übertragen. Das Originalsignal ist dadurch genau reproduzierbar, da Störungen in der Übertragungskette nicht den Sinngehalt modifizieren können.

3.1. Duales Zahlensystem

Die Übertragung und Speicherung digitaler Informationen erfolgt nach dem *dualen Zahlensystem*. Dessen Wurzeln liegen bei LEIBNIZ (1679), der das Dualsystem, einen Vertreter der binären Systeme[61], als Stellenwertsystem zur Basis 2 entwickelte. Die Zahl 27 wird beispielsweise im allgemein gebräuchlichen Dezimalsystem dargestellt als $27_{10}=2\cdot10^1+7\cdot10^0$, im Dualsystem als $11011_2=1\cdot2^4+1\cdot2^3+0\cdot2^2+1\cdot2^1+1\cdot2^0$.

Wenn auch Dualzahlen für den Betrachter sehr unübersichtlich erscheinen und die Informatik deshalb auf das *Hexadezimalsystem* zurückgreift[62], so birgt das Dualsystem den Vorteil, dass sich die abstrakten Zahlen „0“ und „1“ als die möglichen Zustände einer Ziffer gut durch hohe und niedrige Spannungswerte ausdrücken lassen. Eine Binärzahl besteht aus einzelnen zweiwertigen Zeichen, die den Zustand „0“ oder „1“ annehmen können. Jedes einzelne wird *Bit* (Abkürzung für „binary digit“) bezeichnet.

---

[61] Als binäres System gilt z.B. auch das Morsealphabet (Punkt, Strich).

[62] Vgl. Stahlknecht, Peter/Hasenkamp, Ulrich: *Einführung in die Wirtschaftsinformatik*, Berlin/Heidelberg ^{8}1997 [^{1}1983], S. 19.

3.2. Signalabtastung

Zur Umwandlung von analogen Audiosignalen in Folgen digitaler Werte wird die *Pulse-Code-Modulation* (PCM) verwendet: Dem Tonsignal werden durch einen Abtaster periodisch Momentanwerte in Form von Spannungszuständen entnommen, die ein A/D-Wandler[63] in Zahlen umwandelt. Dafür muss jeder Zahlenstelle die entsprechende Information durch einen eigenen Kontakt zugeführt werden. Bei der Übertragung über einen einzigen Datenkanal wird daher dieses breite Datenbündel durch einen Parallel/Serien-Wandler zeitlich aufeinanderfolgend angeordnet. Die Rückwandlung in ein hörbares Signal geschieht durch einen Serien/Parallel-Wandler und einen D/A-Wandler[64].

Analoge Tonsignale besitzen zwei Eigenschaften, in denen sie sich streng von Digitalsignalen unterscheiden: zum einen ist das Signal zeitkontinuierlich, d.h., es existiert zu jedem unendlich kleinen Zeitwert ein Signalwert, und zum anderen ist das Signal wertkontinuierlich, was besagt, dass jeder beliebige Wert innerhalb des vorgegebenen Wertebereichs angenommen werden kann. Für die Verarbeitung der Signale mittels Zahlenkombinationen hätten diese Eigenschaften zur Folge, dass diese unendlich viele Ziffern zur Beschreibung eines Momentanwertes benötigten.

Die Charakterisierung des analogen Signals durch eine Folge von Zahlenwerten ist deshalb beschränkt auf eine endliche Zahl von Augenblickswerten; das zeitkontinuierliche Signal wird also in ein zeitdiskretes überführt. Diese

[63] Analog/Digital-Wandler.

[64] Digital/Analog-Wandler.

Methode ist „*technisch notwendig, da die anschließende Umwandlung in einen Zahlenwert jeweils auch eine gewisse Zeit benötigt.*“[65] Hinzu kommt, dass sonst auch die zu verarbeitende Datenmenge in unendliche Dimensionen steigt.

Ein Impulsgeber steuert den Abtaster, der einen Wert liefert, sobald eine Impulsspannung anliegt. Dies geschieht kontinuierlich im Abstand einer gleichbleibenden Abtastperiode, deren Kehrwert die *Abtastfrequenz* oder *Samplerate* ergibt.[66]

Die Wahl der Samplerate ist nicht beliebig: Einerseits werden bei geringer Samplerate bedeutende Signalveränderungen innerhalb einer längeren Abtastphase nicht berücksichtigt, andererseits ist bei zu kürzer Abtastperiode die Transformation in ein digitales Signal im A/D-Wandler nicht möglich, da dieser nicht schnell genug die jeweiligen Signaleigenschaften identifizieren und der digitalen Entsprechung zuordnen kann. Der vom Abtaster gelieferte Momentwert wird in einem *sample&hold*-Schaltkreis für kurze Zeit gespeichert und so bis zum Beginn des nächsten verlängert. Dieser Abtastvorgang folgt dem von SHANNON formulierten *Abtasttheorem*, nach dem gelten muss[67]:

DICKREITER[68] wie auch KÄMMER[69] lassen den Spezialfall $f_A=2\ f_{max}$ zu, der von

$$f_A > 2\ f_{max}$$

f_A = Abtastfrequenz
f_{max} = höchste zu übertragende Frequenz

[65] Dickreiter, Michael: *Handbuch der Tonstudiotechnik – Band 2*, München [6]1997 [[1]1977], S. 273.

[66] Beispiel: Abtastfrequenz 44100 Hz, Abtastperiodendauer 1/44100 s (mit Hz:=1/s).

[67] Shannon, Claude E./Weaver, Warren: *The Mathematical Theory Of Communication*, Urbana/Chicago 1963, S. 86.

[68] Dickreiter, Michael, a.a.O., S. 279: *„Zusammenhang zwischen Abtastfrequenz und oberer Grenzfrequenz“: „$f_O \leq ½\ f_A$“*

[69] Kämmer, Heinz W.: *Die Compact Disc*, München [2]1988 [[1]1983], S. 29: *„(...) für 20000 Hz braucht man eine Abtastfrequenz von mindestens 40000 Hz (...)“*.

der Mehrzahl der Autoren jedoch ausgeschlossen wird. Hier sei verwiesen auf WINKLER und KRIEG. Beide schreiben:

> *„Für ein Audiosignal [mit der höchsten Frequenz] von 20 kHz braucht man also eine Abtastrate von über 40 kHz. Oder – anders ausgedrückt – bei einer Abtastrate von 40 kHz darf das zu digitalisierende [Analog-] Signal keine 20 kHz sein.“*[70]

Shannon selbst formuliert:

> *„If a function of time f(t) is limited to the band from 0 to W cycles per second it is completely determined by giving its ordinates at a series of discrete points spaced 1/(2W) seconds* (...) *.“*[71]

Daraus lässt sich ableiten, dass zur Beschreibung von n Perioden pro Sekunde Mindestens genau 2n+1 Abtastwerte benötigt werden: Setzt man n=1, ist also die Periodendauer T=1 sec, so werden 3 Abtastwerte benötigt, um die Schwingung eindeutig zu charakterisieren.

3.3. Analog-Digital-Wandlung

Das zeitdiskrete Signal wird im A/D-Wandler nach seiner Größe einem Wert zugeordnet. Als Zuordnungskriterium dient dabei nicht der absolute Spannungswert, sondern ein gewisses Spannungsintervall, da sich z.B. mit vierstelligen Dualzahlen, also in 4 Bit-Auflösung, nur 16 verschiedene Werte von 0000_2 bis 1111_2 erzeugen lassen, die gleichmäßig über den Bereich zwischen Maximum und Minimum der Kurve des analogen Signals verteilt werden (s. Abb. 19).
Treffen zwei Abtastwerte dasselbe Spannungsintervall, so werden sie durch dieselbe Zahl repräsentiert. Diese Signalverfälschung lässt sich interpretieren als

---

[70] Winkler, Gernot: *Tonaufzeichnung Digital – Moderne Audiotechnik mit Computerhilfe*, Aachen 1990, S. 15. Vgl. auch Krieg, Bernhard: *Digitale Audiotechnik ohne Ballast – Moderne Verfahren, Technik, Geräte, praktische Hinweise und Tips*, München 1992, S. 21; abweichende Formulierungen in *[ ]*.

[71] Shannon, Claude E./Weaver, Warren, a.a.O., S. 86.

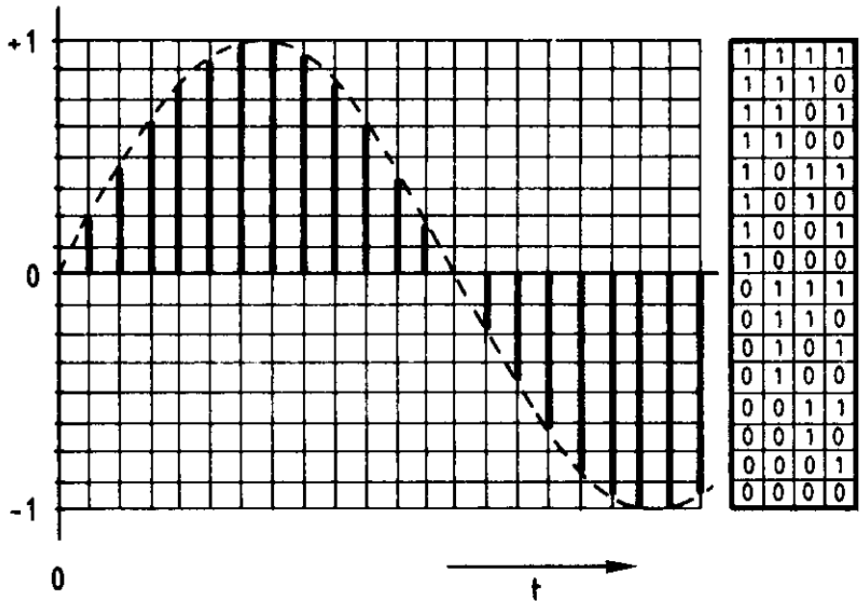


Abb. 19: A/D-Wandlung mit 4 Bit-Auflösung
In diesem Beispiel stellt das höchstwertige Bit das Vorzeichen dar (1 ⇒ positiv / 0 ⇒ negativ), die folgenden drei Bits den Betrag, der acht (2^3) verschieden Zustände annehmen kann.

Addition eines Fehler- oder Störsignals, dessen Werte zufällig verteilt sind im Bereich ±q/2 (q=Quantisierungsintervall). Da diese zufällige Verteilung ein Merkmal des gleichverteilten Rauschens ist, wird diese Störung als *Quantisierungsrauschen* bezeichnet. Bei der Rückwandlung vom digitalen in ein analoges Signal ist dieser Quantisierungsfehler nicht mehr zu beheben und wird als Rauschen wahrnehmbar.

Dieses Phänomen verringert sich, wenn die Anzahl der Intervalle zunimmt, da parallel die Intervallgröße abnimmt. Dazu wird die Stellenanzahl der Binärzahlen, die *Wortbreite*, erweitert, so dass sich durch jede zusätzliche Stelle die Anzahl der Spannungsintervalle verdoppelt ($2^{x+1}=2*2^x$). Jedes zusätzliche Bit verbessert den *Quantisierungsrauschabstand* um 6 dB (s. Abb. 20).[72] Dieser wird auch als *Signal-Rauschabstand*, *Störabstand* oder *SNR (signal-to-noise*

[72] Zölzer, Udo: *Digitale Audiosignalverarbeitung*, Stuttgart 1996, S. 62.

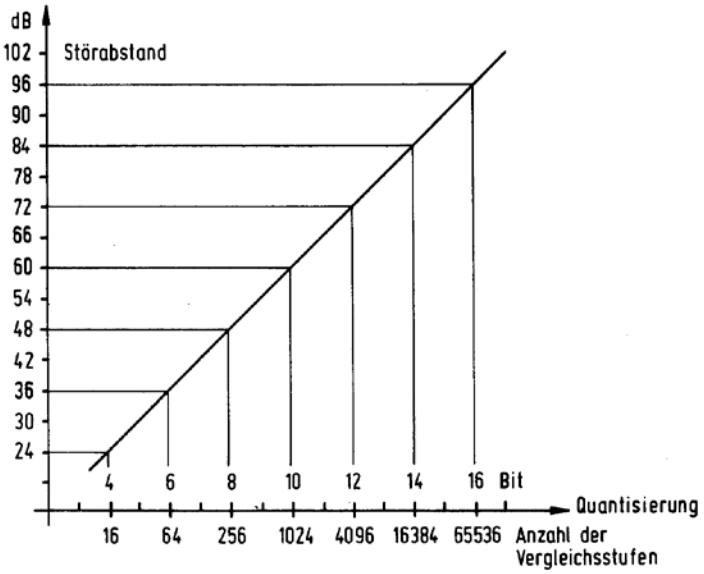


Abb. 20: Quantisierungsrauschabstand eines digitalen Systems

ratio) bezeichnet und dient als Maß für die durch den Analog-Digital-Wandler verursachten Signalverfälschungen.[73]

Künstliche Störschwingungen mit Frequenzen über 20 kHz wie z.B. Rauschen können das *Aliasing*-Phänomen verursachen: Aliasing bedeutet, dass die Signale, die gleich der Hälfte der Samplingfrequenz sind oder sie überschreiten, in den hörbaren Bereich *„umgeklappt“* werden. Beispielsweise würde sich eine Trägerfrequenz von 39 kHz aus der Stereo-UKW-Übertragung bei einer Samplerate von 44,1 kHz als Pfeifton der Frequenz 44,1 kHz-39 kHz=5,1 kHz auswirken. Daher entfernen steilflankige Analog-Filter die ‚zu hohen‘ Anteile, bevor diese den Digital-Analog Wandler erreichen.

3.4. Digital-Analog-Wandlung

Der D/A-Wandler hat die Funktion, aus einer an ihm anliegenden Zahl die entsprechende Spannung am Ausgang zu erzeugen. Er wandelt die Zahlenzustände in Ströme um, die über eine Summierschaltung addiert und als

[73] Der SNR ist definiert als das Verhältnis $SNR=P_S/P_Q$ (P_S=Leistung des analogen Signals, P_Q=Leistung des Quantisierungsrauschens), nach: Meyer, Martin: *Signalverarbeitung – Analoge und digitale Signale, Systeme und Filter*, Braunschweig/Wiesbaden 1998, S. 216.

entsprechende Spannung am analogen Ausgang des D/A-Wandlers abgegriffen werden.

Auf weitere technische Einzelheiten soll in dieser Arbeit verzichtet werden, da die Rückwandlung von der digitalen in die analoge Ebene erst nach der Datenreduktion stattfindet und somit keine unmittelbare Bedeutung für das Thema dieser Arbeit hat.

4. Digitale Tonaufzeichnung ohne Datenreduktion

Die Bezeichnung „Digitale Tonaufzeichnung ohne Datenreduktion“ birgt genaugenommen einen Widerspruch in sich, da der Vorgang der Digitalisierung weder zeit- noch wertkontinuierlich ist und somit bereits zu einem Informationsverlust führt. Regelrechte Datenreduktion bezieht sich jedoch nur auf die nachträgliche Bearbeitung eines bereits in digitaler Form vorliegenden Datenmaterials.

Eine der Hauptaufgaben eines Systems zur Reduktion von digitalen Audiodaten ist die Einhaltung der *High Fidelity Audio*-Maßstäbe; folglich sollten keine hörbaren Unterschiede zwischen dem reduzierten Signal und dem digitalen Original entstehen.[74] Die Träger für solche Originale sind vor allem Compact Discs und Digital Audio Tapes.

4.1. Compact Disc

Die *Compact-Disc (CD)* oder genauer *Compact Disc Digital Audio (CD-DA)* hat sich als Standard auf dem Tonträgermarkt etabliert, da sie unter anderem die Vorzüge der digitalen Tontechnik für den privaten Gebrauch zugänglich machte. Weitere Erscheinungsformen der CD, wie z.B. die *CD-ROM*[75] als Massenspeicher für Computer oder *CD-Video*, sollen an dieser Stelle nur erwähnt werden, da sie mit dem Kern dieser Arbeit nur indirekt zusammenhängen. So lässt sich der Audioteil einer CD-ROM im *Mixed-Mode*, also sowohl mit einer Daten- als auch einer Audiospur, auf den Aufbau einer CD-DA zurückführen.

---

[74] Brandenburg, Karlheinz 1998: *Perceptual Coding Of High Quality Digital Audio,* in: Applications Of Digital Signal Processing To Audio And Acoustics (Brandenburg, Karlheinz/Kahrs, Mark Hg.), Boston/Dordrecht/London 1998, S. 40.

[75] Compact Disc - Read Only Memory.

Um dem Abtasttheorem (s. Kap. 3.2.) zu genügen, liegt dem CD-Standard eine Abtastrate von 44,1 kHz zugrunde, die unter Gewährleistung gewisser Toleranzen ein Übertragungsspektrum von 20 kHz gewährleistet.
Die 16 Bit-Auflösung ermöglicht eine Unterteilung in 65.536 Quantisierungsstufen, mit denen ein Signal-Rauschabstand von etwa 96 dB erreicht wird (s. Abb. 20),[76] so dass das Quantisierungsrauschen nicht wahrnehmbar ist.[77]

Auch der Aufbau einer CD ist in Hinblick auf Datenreduktion interessant, da die Konstruktion den zur Verfügung stehenden Speicherplatz begrenzt. Wenn dieser nicht ausreicht, müßte z.B. der Speicherplatzbedarf gesenkt werden.
Auf der Oberfläche einer optischen Speicherplatte sind mikroskopisch kleine, kontinuierlich einer Spur von innen nach außen folgende Vertiefungen eingeprägt, die 600 nm breit und 120 nm bis 200 nm tief sind und als *Pits* bezeichnet werden. Die nicht vertieften Bereiche der Reflexionsschicht heißen *Lands*.
Die Unterseite der CD wird mit Laserlichtstrahlen mit der Wellenlänge λ=780 nm abgetastet. Trifft das Licht auf die ungeprägte Spiegeloberfläche, wird der Laserstrahl voll reflektiert. Trifft der Lichtstrahl jedoch auf ein Pit, das so hoch ist wie etwa ein Viertel der Wellenlänge λ, so verringert sich die bei den Empfänger-Dioden ankommende Energie aufgrund der Interferenz zwischen Hin- und Rückstrahl um ½λ. Wären Hin- und Rückstrahl genau gegenphasig, so würden sich beide gegenseitig vollkommen auslöschen. In der Praxis lässt sich jedoch nur eine deutliche Schwächung des reflektierten Lichtes erreichen. Die Summe aller Empfänger-Diodensignale ergibt, in hohe und niedrige Spannungen transformiert, die Information der CD.

[76] Webers, Johannes, a.a.O., S. 156.

[77] Franz, Dieter, a.a.O., S. 373.

Während des Auslesevorgangs wird die auf der CD gespeicherte Information als Lichtsignal der Qualität hell/dunkel verarbeitet. Daher zählt die CD zur Familie der *optischen Speichermedien*.

Ein Fokussiersystem aus drei Servo-Schaltungen sorgt dafür, dass beim Auslesevorgang der Laserstrahl korrekt in der Spur bleibt, das optische System den präzisen Abstand zur Plattenoberfläche einhält und sich die Spur mit gleichbleibender Geschwindigkeit unter dem Lesekopf bewegt. Letzteres bedeutet, dass die Umdrehungsgeschwindigkeit nach dem *CLV-Modus* (Constant Linear Velocity) mit größer werdendem Spurradius abnehmen muss, um einen gleichbleibenden Datenstrom sicherzustellen.[78] Im Gegensatz dazu steht der bei Festplattensystemen verwendete *CAV-Modus* (Constant Angular Velocity), bei dem die Motordrehzahl konstant gehalten wird.[79]

Auf einem Millimeter Plattenradius befinden sich mehr als 600 Spuren, die bei dem Durchmesser einer CD von 12 cm eine Spur der Gesamtlänge von rund 6 km bilden.[80] Aus dieser extrem hohen Speicherdichte ergibt sich eine Kapazität von bis zu 74 Minuten stereophoner Audiodaten.
Wird die CD mit der Standardgeschwindigkeit von 1,3 m/s ausgelesen, so werden pro Sekunde 176.400 Bytes übertragen, die *Audiodatenrate* beträgt also 1,4112 MBit/s.[81] Dieser Wert bezieht sich nur auf die reinen Audiodaten, denn der *Kanalbitstrom* beinhaltet außerdem die *Subcodes*, zu denen u.a. Daten zur Fehlerkorrektur und zur Synchronisationskontrolle gehören. Die Gesamtdatenrate liegt bei 4,32 MBit/s.[82] Die Frage, wie diese zusätzlichen Daten zustande kommen, soll an dieser Stelle unbeantwortet bleiben, da sie in

[78] Meissner, Hansgeorg: *Digitale Multimediasysteme*, Berlin 1994, S. 92.

[79] Meissner, Hansgeorg, a.a.O., S. 91.

[80] Meissner, Hansgeorg, a.a.O., S. 88.

[81] Steinmetz, Ralf: *Multimedia-Technologie – Einführung und Grundlagen*, Berlin/Heidelberg 1993, S. 161.

[82] Webers, Johannes, a.a.O., S. 610.

keinem Zusammenhang mit Audioqualität stehen, sondern in erster Linie der Einhaltung des CD-DA- oder auch *Red Book*-Standards[83] für Audio-CDs dienen.

4.2. Digital Audio Tape

Das *Digital Audio Tape (DAT)* ist ein Magnetband, das PCM-Daten aufzeichnet. Dieses Aufzeichnungsverfahren ist verwandt mit dem der Videoaufzeichnung, bei der ebenfalls rotierende Magnetköpfe entlang einer schräg zur Bandrichtung orientierten Spur speichern bzw. lesen. Durch diese Lesemethode lässt sich die Datendichte auf dem Band erhöhen.

Die Abtastung erfolgt mit 48 kHz im Standardmodus oder mit 32 kHz im Longplaymodus. Mit der Verminderung der Samplingfrequenz verlängert sich die Spielzeit der DAT-Kassette von standardmäßigen 120 Minuten auf 240 Minuten. Hochwertige DAT-Aufzeichnungsgeräte verfügen über eine dritte Abtastrate, die bei 44,1 kHz liegt. Solche Studio-DAT-Rekorder haben den Vorteil, dass sie als Mastermedium für eine CD-Produktion keine abschließende verlustbehaftete Konvertierung der Abtastrate von 48 kHz auf 44,1 kHz benötigen, also die digitale Audioinformation des DAT der zukünftigen CD entspricht. Alle drei Modi quantisieren die Daten mit 16 Bit-Auflösung.

Da Studio-DAT-Recorder mit einer Abtastfrequenz von 44,1 kHz bei 16 Bit-Quantisierung dasselbe Aufzeichnungsverfahren verwenden wie die Audio-CD, ergibt sich eine „Q*ualitäts-Äquivalenz*“[84] zwischen DAT und CD. Daher wird hier auf die DAT-Technologie nicht weiter eingegangen, und sie ist im Folgenden miteinbezogen, wenn von „CD-Qualität“ gesprochen wird.

[83] Vgl. Meissner, Hansgeorg, a.a.O., S. 96.

[84] Winkler, Gernot, a.a.O., S. 127.

Teil III: Datenreduktion unter psychoakustischen und psychologischen Aspekten

5. Digitale Tonaufzeichnung unter Datenreduktion

An dieser Stelle soll erläutert werden, weshalb zwischen den im Alltag häufig synonym verwendeten Begriffen Datenreduktion und Datenkompression differenziert werden muss. Der Unterschied zwischen beiden Termini ist fundamental: Der Vorgang der Reduktion bedeutet eine Verringerung der Daten, der der Kompression eine Zusammenpressung.

Algorithmen zur Verringerung der Audiodatenmenge basieren auf zwei Strategien, die sich in ihrem Resultat grundlegend unterscheiden: Werden für den Höreindruck vermeintlich irrelevante Signalanteile nicht mitkodiert, so lässt sich anhand der verbleibenden Daten die Urgestalt des Schallsignals nicht reproduzieren, die Semantik der Daten wird also verändert. Die Kodierung nach rein mathematischen oder statistischen Verfahren wie der Wiederholungskompression[85] jedoch ändert die Semantik nicht, sie sorgt nur für eine Verringerung der Datenmenge. Datenreduktion und Datenkompression implizieren deshalb verschiedene Aussagen bezüglich der Qualität des kodierten Signals.

[85] „*Folgt dasselbe Zeichen (z.B. die Zahl 0) mehrmals aufeinander, wird es nur einmal zusammen mit der Anzahl dargestellt*.“ Nach: Stahlknecht, Peter/Hasenkamp, Ulrich, a.a.O., S. 23.

In der Literatur sind sowohl die synonyme Verwendung[86] als auch die strikte Unterscheidung nach obengenannten Kriterien[87] anzutreffen. RUMSEY umgeht Verwechslungen, indem er differenziert nach „*lossless and lossy coding*“.[88]

Der begrifflichen Unterscheidung zwischen Datenkompression und Datenreduktion entsprechend werden die Algorithmen der gehörangepassten Audiokodierung differenziert nach *Source Coding* und *Perceptual Coding*[89].

Das *Source Coding* nutzt statistische Eigenschaften eines Datenvorrats, um die redundanten Datenanteile zwar einzusparen, wohl aber nach einer Dekodierung das Original vollständig wiederherstellen zu können. Die wohl bekanntesten nach diesem Prinzip arbeitenden Kodierer stammen aus dem Bereich Datensicherung (z.B. Packer-Software WINZIP, ARJ, u.a.).

Das grundlegende Prinzip des *Perceptual Coding* besteht darin, das Quantisierungsrauschen in einem eben unhörbaren Maß zuzulassen.[90] Die theoretische Grundlage hierfür findet sich in der Psychoakustik, und der Schlüsselbegriff lautet nach BRANDENBURG ‚Maskierung‘ (s. Kap. 2.7.).[91]

Das Quantisierungsrauschen, das isoliert betrachtet hörbar ist, wird genau dann unhörbar, wenn es unterhalb der vom Schallsignal abhängigen Mithörschwelle verschwindet. Das *Psychoakustische Modell* errechnet in kurzen Zeitintervallen die jeweilige Mithörschwelle und entscheidet über das zulässige Rauschmaß, indem es die Quantisierung als ‚Rauschverursacher‘ über die Bitzuweisung

---

[86] Vgl. Meissner, Hansgeorg, a.a.O., S. 80: „*Kompression von Audiosignalen im Rahmen des MPEG-Standards*“.

[87] Vgl. Brandenburg, Karlheinz 1989, a.a.O., S. 2.

[88] Rumsey, Francis: *The Audio Workstation Handbook*, Oxford/Boston 1996, S. 75.

[89] dt.: gehörangepasste Quellkodierung, nach Herre, Jürgen: *Fehlerverschleierung bei spektral kodierten Audiosignalen*, Dissertation an der Universität Erlangen-Nürnberg 1995, S. 21.

[90] Brandenburg, Karlheinz 1998, a.a.O., S. 42.

[91] ebd.

steuert (s. Abb. 21). Dieses Modell arbeitet je nach Enkodiersystem[92] unterschiedlich.

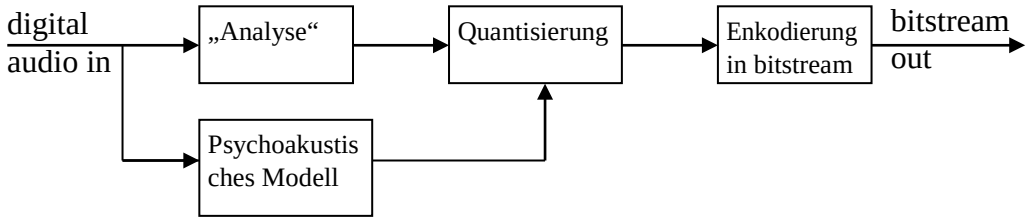


Abb. 21: Blockdiagramm eines Enkodierers mit psychakustischem Modell.

[92] Als *Enkoder* wird derjenige Teil des Kodiersystems bezeichnet, der die Audiodaten nach einem bestimmten Verfahren verschlüsselt.

5.1. Datenreduktionsverfahren der Digital Compact Cassette: PASC

Die Firma PHILIPS stellte 1991 mit der *Digital Compact Cassette (DCC)* ein digitales Audiosystem vor, das im Gegensatz zum DAT-System mit einem feststehenden Magnetkopf für das Lesen und Schreiben digitaler Informationen arbeitet. Die Geschwindigkeit von 4,76 cm/s, mit der das Band am Schreib-/Lesekopf vorbeigeführt wird, entspricht der der analogen *MusiCassette (MC)*.[93] Durch einen zusätzlichen analogen Signalweg mit eigenem Lesekopf und entsprechender Wiedergabeelektronik ist die DCC abwärts-kompatibel, so dass auf einem DCC-Recorder auch MCs abgespielt werden können. Die Aufnahmefunktion beschränkt sich jedoch auf das Digitalsystem.
Diese gewünschte Kompatibilität bedingt natürlich, dass neben der Bandgeschwindigkeit auch die Abmessungen der DCC sowie die Breite und die Länge des verwendeten Bandes mit denen der MC übereinstimmen müssen.
Aus diesen Voraussetzungen ergibt sich die maximale Speicherkapazität einer Kassette, die eine Datenrate von ca. 700 KBit/s zulässt.[94] Somit muss die Datenmenge „zusammengeschrumpft“ werden, damit der Speicherplatzbedarf der zu speichernden Informationen nicht die Kapazität des Datenträgers übersteigt.

Die Audiodaten werden vor der Speicherung auf DCC nach dem *PASC*-Algorithmus (*Precision Adaptive Subband Coding*) kodiert. Dieser Algorithmus wurde ursprünglich mit dem Namen MPEG-1 Audio Layer 1 (s. Kap. 5.3.) nach den Anforderungen des DCC-Systems entwickelt.[95] Das Audiodatenvolumen wird so auf 25% des *Vollformats* (einer CD-DA) reduziert. [96]

---

[93] Krieg, Bernhard, a.a.O., S. 211.

[94] Krieg, Bernhard, a.a.O., S. 207.

[95] Fraunhofer IIS: *Frequently Asked Questions about MPEG Audio Layer-3*, nach: http://www.iis.fhg.de/amm/techinf/layer3/layer3faq/index.html .

[96] Internetseite der Firma Philips: http://www.magnavox.com/product/pa332dcc.html .

Der PASC-Algorithmus nutzt für die Datenreduktion die Erkenntnisse der Psychoakustik über Maskierungseffekte (s. Kap. 2.7.). Dabei wird das Prinzip verfolgt, nicht die dem analogen Signal entsprechenden digitalen Informationen zu kodieren, d.h. im Vollformat zu erfassen, sondern diese nach ihrer Relevanz für den Höreindruck zu filtern.

Um die Mithörschwelle eines komplexen Signals zu berechnen, wird das Audiosignal in 32 gleich breite Frequenzbänder (*Subbänder*) unterteilt.[97] Der Prozessor ermittelt für jedes einzelne Band die Mithörschwelle und bestimmt alle maskierten Töne. Diese Methode wird als *Teilbandkodierung* bezeichnet.[98]

Die Signale oberhalb der Mithörschwelle werden nun in einer der Amplitude angepassten Auflösung nach dem Verfahren der *adaptiven Notation* kodiert. Diese Anpassung basiert auf der Kodierung der Signale in der speicherplatzsparenden *Gleitpunktdarstellung*:

Jede Zahl Z lässt sich darstellen in der Form $Z = M \cdot B^e$ (mit M=Mantisse, B=Basis, e=Exponent).[99] Die Mantisse umschreibt dabei den Wert Z des Samples, während B^e als Maßstabsfaktor dient.[100]

---

[97] Krieg, Bernhard: *Digitale Audiotechnik ohne Ballast*, München 1992, S. 92.

[98] Krieg, Bernhard, a.a.O., S. 83.

[99] vgl. Van den Enden, Ad W.M./ Verhoeckx, Niek A. M.: *Digitale Signalverarbeitung*, Braunschweig/Wiesbaden 1990, S. 283f.

[100] Zur Verdeutlichung des Prinzips der Gleitpunktdarstellung sei die Einwohnerzahl einer fiktiven Metropole betrachtet, die bei 12.400.000 E (E=Einwohner) liegt. Dient ‚E' als Grundeinheit des Rechensystems, gilt nach obiger Formel: $12.400.000\ E = 124\ E \cdot 10^5$ ($Z = M \cdot B^e$). Das Ergebnis dieser Umformung ohne besondere Berücksichtigung der Einheit ‚E' enthält zwei Variablen, den Wert M = 124 und den Exponenten e = 5, die Basis 10 hingegen ist mit dem Dezimalsystem vorgegeben. M und e sind also alle benötigten Informationen für die korrekte Darstellung der Einwohnerzahl. Mit der Vereinbarung, dass diese nach der Syntax eM kodiert werde, lautet die kodierte Einwohnerzahl 5124 und benötigt genau die Hälfte an Stellen gegenüber der nichtkodierten Einwohnerzahl, die als Festpunktzahl angesehen werden kann. Diesem Beispiel liegt das Dezimalsystem (B=10) zugrunde; die Handhabung im Dualsystem (B=2) erfolgt nach demselben Prinzip.

Im Gegensatz dazu steht die *Festpunktdarstellung*, die beispielsweise in der CD-Technologie angewendet wird.[101]

Die eigentliche Anpassung erfolgt durch die *Adaptive Allocation*. Dabei werden der Darstellung der Mantisse exakt die benötigten Bits zugewiesen. Deren Anzahl schwankt zwischen 2 und 15 und ist proportional zur Größe der Amplitude, berücksichtigt aber gleichzeitig den noch zur Verfügung stehenden Speicherplatz. So lässt sich dieser effektiv nutzen, und eine Verschwendung von Speicherplatz wird umgangen. Bei PHILIPS geht man außerdem davon aus, dass sich Mithörschwellen und Dynamikausprägungen in Relation zur Abtastfrequenz nur langsam verändern. Daher fasst man 12 Samples als einen *PASC-Rahmen* zusammen, für den die Mithörschwelle und der Exponent e für die Gleitkommadarstellung nur einmal berechnet werden.[102] Daraus resultiert eine weitere Einsparung an Speicherplatz. Sowohl die Ausnutzung der Verdeckungsphänomene als auch der Einsatz der adaptiven Notation tragen zu gleichen Teilen zur Verminderung der Datenmenge bei, so dass die Audiodatenrate des DCC-Systems bei 384 KBit/s liegt.

5.2. Datenreduktionsverfahren der MiniDisc: ATRAC

Die *MiniDisc (MD)*, 1991 von der Firma SONY vorgestellt, ist ein wiederbeschreibbares Medium und zählt zu den *magneto-optischen Medien (MO).* Der Auslesevorgang entspricht aus technischer Sicht dem von der CD bekannten: Über einen Laser und einen Photosensor wird das Reflexionsverhalten der Oberfläche der CD ausgewertet. Die Speicherung der Informationen ermöglichen Materialien, deren Reflexionsverhalten sich durch

[101] Shibazaki, Isao: *The Signal Compression Technology of Sony's Newest MD Recorder*, MJ Audio Technology Magazine Technical Report, Dezember 1995, nach: http://www.amulation.co.uk/minidisc/mj_ja3es.html .

[102] Internetseite der Firma Philips: http://www.magnavox.com/product/pa332dcc.html .

ein Magnetfeld in Abhängigkeit von ihrer Temperatur modifizieren lassen: Um die benötigte Datendichte realisieren zu können, erhitzt ein Laser einen Punkt auf dem Datenträger, so dass sich nur diese erhitzte Stelle dauerhaft nach dem Magnetfeld ausrichtet, das durch den Magnetkopf erzeugt wird. Das verwendete optische Leseverfahren nutzt Laserlicht mit der Wellenlänge λ=780 nm, dessen Lichtpunkt genauso groß ist wie der des optischen Systems eines CD-Players[103]; folglich ist auch die Datendichte einer MD identisch mit der einer CD.
Der Durchmesser einer Minidisc beträgt 64 mm, der einer CompactDisc 120 mm. Größenbedingt beträgt die Kapazität einer MD etwa 140 MB; eine (Standard-) CD fass hingegen etwa 650 MB. Trotz der großen Kapazitätsunterschiede bieten beide Technologien eine Spielzeit von 74 Minuten Audio (in Stereo[104]).[105]

Um das Speicherplatzdefizit zu kompensieren, wird in MD-Recordern das *ATRAC*-Verfahren (*Adaptive Transform Acoustic Coding*) zur Datenreduktion eingesetzt. Die Grundlage für die Reduktion durch ATRAC bildet die A/D-Wandlung der Audioinformation mit 44,1 kHz und 16 Bit-Auflösung, deren Speicherplatzbedürfnisse zunächst denen einer CD entsprechen.
Das digitale Signal wird zunächst in die Bänder low (0 kHz bis 5,5 kHz), mid (5,5 kHz bis 11 kHz) und high (11 kHz bis 22 kHz) unterteilt.[106] Diese drei Frequenzbänder werden nun, um den Signalprozessor zu entlasten, getrennt voneinander mittels der *MDCT (modifizierte diskrete Cosinus-Transformation)* aus dem Zeitbereich in den Frequenzbereich übertragen.[107]

---

[103] Bremm, Peter: Artikel *Ein Klang von Licht und Magnet · Durch nichts zu erschüttern – MiniDisc-Laufwerk*, KEYS 11/1996, S. 105.

[104] Die Spielzeit einer MD verdoppelt sich im Mono-Aufnahmemodus.

[105] Bremm, Peter, a.a.O., S. 105.

[106] Tsutsui, Kyoya/Suzuki, Hiroshi/Shimoyoshi, Oscamu u.a.: *ATRAC: Adaptive Transform Acoustic Coding For MiniDisc*, 93rd AES Convention, San Fransisco 1992, Preprint 3456, nach: http://www.amulation.co.uk/minidisc/aes _atrac.html .

[107] ebd.

Die Größe des Zeitfensters, innerhalb dessen die MDCT stattfindet, wird in zwei Modi gebildet. Die Standardgröße beträgt für die drei Frequenzbänder einheitlich 11,6 ms (*long mode*), und bei Bedarf wechselt der Prozessor in den alternativen Modus, in dem das ‚high'-Band alle 1,45 ms und das ‚mid'- und das ‚low'-Band in 2,45 ms langen Zeitblöcken untersucht wird (*short mode*).[108]

Die 512 Samples (44,1 Samples/ms · 11,6 ms) werden durch die MDCT in 512 Spektralkomponenten (low+mid+high=128+128+256, s. Abb. 22) überführt. Da das Leistungsvermögen des Prozessors eine solch hohe Anzahl an zu verarbeitenden Komponenten nicht zulässt, werden diese in maximal 52 vorgegebene Bänder unterteilt; diese Vorgabe orientiert sich an den Erkenntnissen der Psychoakustik über die Frequenzgruppen des menschlichen Hörspektrums (s. Kap. 2.3.). Dementsprechend sind die Frequenzbänder im Tiefenbereich wesentlich schmaler als im Höhenbereich.[109]

In Sonderfällen werden weniger als 52 Bänder untersucht, da der Prozessor diejenigen Bänder ausspart, die keine Signalkomponente aufweisen, was sich

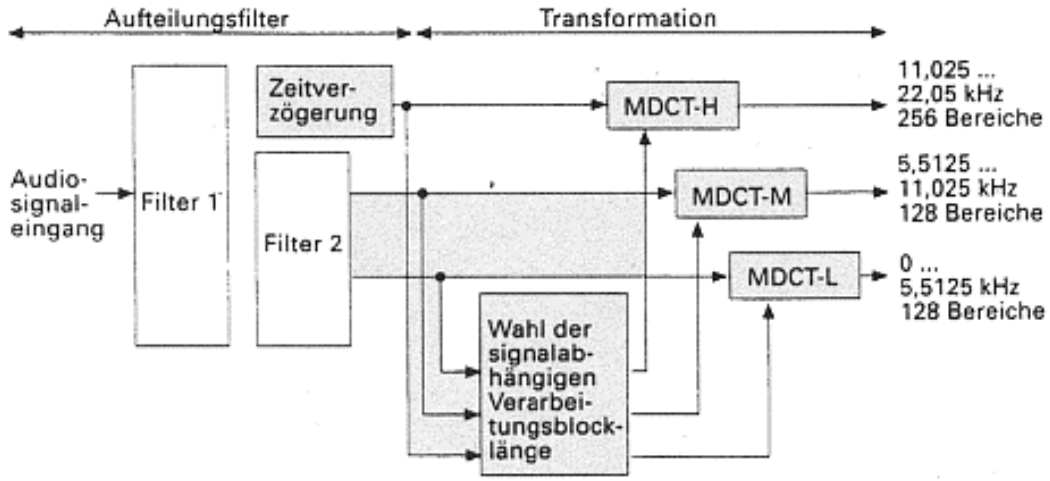


Abb. 22: ATRAC-Enkoder in Blockdarstellung

[108] ebd.

[109] ebd.

zugunsten der Speicherplatzminderung auswirkt.

Die Anzahl der Bits, die den jeweiligen Spektralkomponenten zugewiesen werden, ist variabel und richtet sich nach deren Intensität. Der Grundgedanke hinter diesem Vorgehen besteht darin, dass laute Signalanteile eine geringere Amplitudenauflösung erlauben als leise, da zwar die kleinere Wortbreite ein stärkeres Quantisierungsrauschen bewirkt, dieses jedoch durch das eigentliche Signal maskiert wird (s. Abb. 23).

Folglich findet ein ständiger Transfer an Bit-Ressourcen von lauten zu leisen Signalanteilen statt mit dem Ziel, für laute Signalkomponenten nur soviele Bits zu verwenden wie nötig und im Gegenzug leise Komponenten mit den zusätzlichen Bits so fein wie möglich aufzulösen. Die Bit-Zuweisung reguliert die Bitverteilung so, dass das Quantisierungsrauschen unhörbar bleibt und gleichzeitig der Speicherplatzbedarf in Relation zum Vollformat verringert wird.

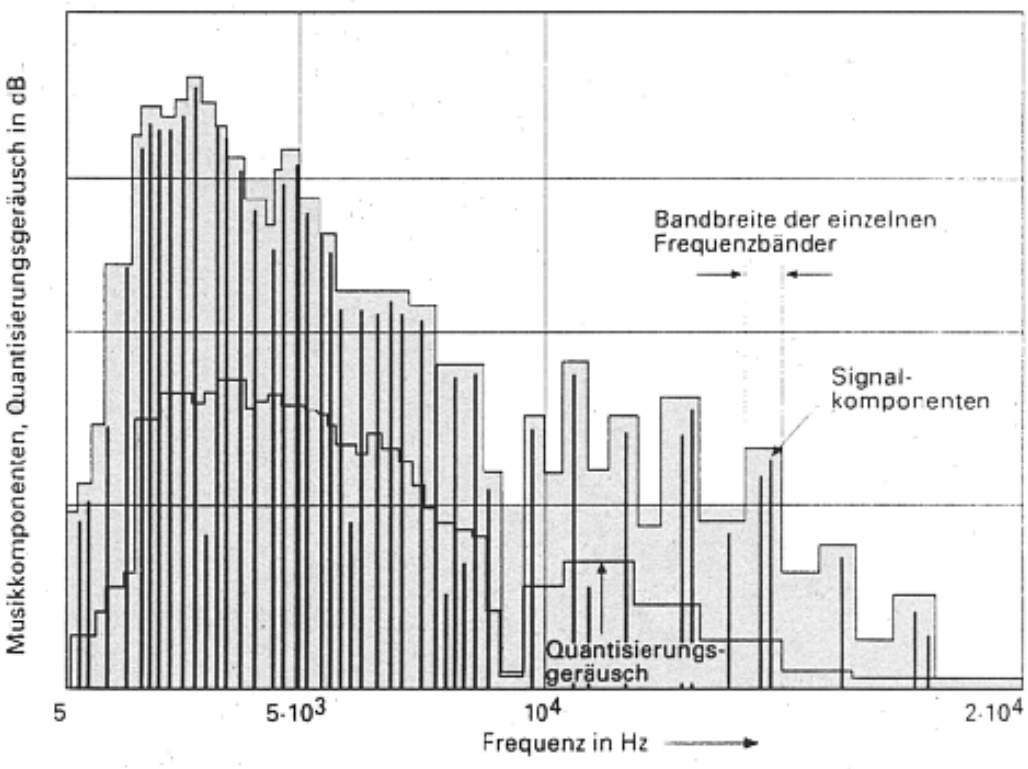


Abb. 23: Bit-Zuweisung durch ATRAC

Da dieses Quantisierungsrauschen innerhalb eines Zeitblocks von 11,6 ms immer gleich stark vorhanden ist, kann ein abrupter Lautstärkewechsel innerhalb des Zeitfensters zu Problemen führen: Der laute Signalanteil bewirkt eine verringerte Amplitudenauflösung, aus der ein stärkeres Quantisierungsrauschen resultiert, das während des lauten Schallereignisses simultan maskiert wird. Da die Bit-Zuweisung für ein Zeitfenster nur einmal vorgenommen wird, tritt dieses Rauschen auch während der ‚stillen' Phase als *pre-echo* auf, ohne jedoch verdeckt zu werden (s. Abb. 24). Deshalb schaltet der ATRAC-Prozessor in dieser Situation in den ‚short mode' um und verkleinert das Zeitfenster auf 1,45 ms bzw. 2,9 ms. Tritt dasselbe Rauschphänomen nun auf, so ist das Rauschsegment vor dem intensiven Signal so kurz, dass es in die Phase der Vorverdeckung fällt und somit maskiert wird (s. Kap 2.7.5.).[110]

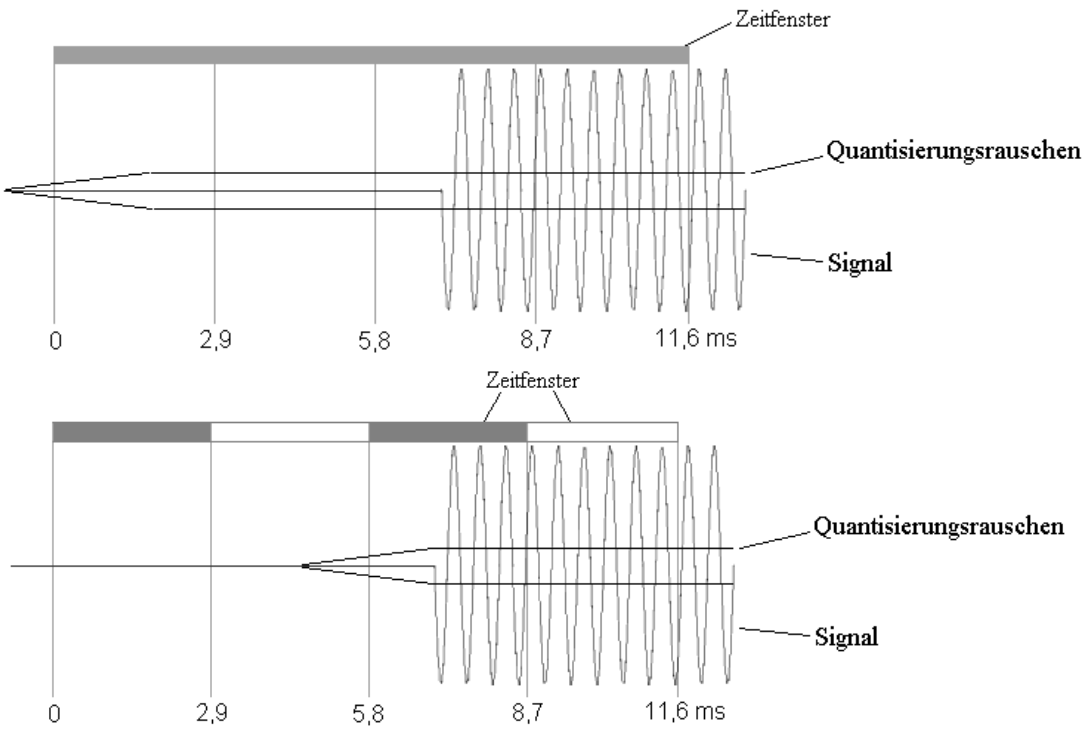


Abb. 24: ATRAC – MDCT-Zeitfenstergrößen.
In dieser Situation ist der short mode (unten) gegenüber dem long mode (oben) vorzuziehen.

[110] Tsutsui, Kyoya/Suzuki, Hiroshi/Shimoyoshi, Oscamu u.a., a.a.O..

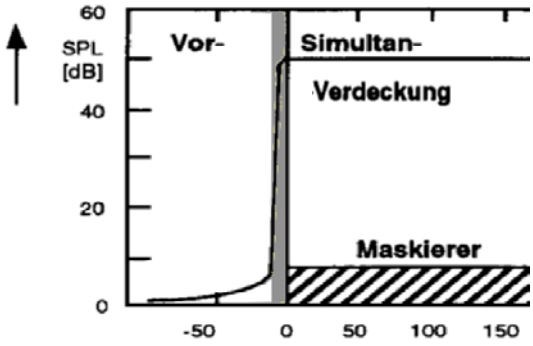


Abb. 25: Übergang Vor- zu Simultanverdeckung

Nach ZWICKER reicht der Vorverdeckungszeitraum bis rund 20 ms vor den Eintritt des Schallereignisses; der long mode mit seiner Blocklänge von 11,6 ms liegt also gänzlich innerhalb der Vorverdeckung. Allerdings erhöht sich die Mithörschwelle zunächst nur sehr langsam und steigt spät steil auf das Hörschwellenniveau der Simultanverdeckung an (s. Abb. 25). Der flache Verlauf der Mithörschwelle genügt demnach nicht immer zur vollständigen Maskierung des Rauschens, es wird also wahrnehmbar.
Im short mode hat das Quantisierungsrauschen jedoch maximal die Dauer von 2,9 ms vor dem lauten Schallsignal. Kurz vor Erreichen der Simultanverdeckung steigt die Vorverdeckungskurve sehr steil an (s. gelbe Markierung in Abb. 25), so dass die Vorverdeckung in diesem Bereich am stärksten wirkt.
Diese zusätzliche Flexibilät bedeutet zwar einen größeren Rechenaufwand für den Prozessor, eröffnet aber gleichzeitig neues Potential zur Biteinsparung.

Auch die dem ATRAC-Verfahren zugrundeliegenden Strategien lassen sich nach Datenreduktion und Datenkompression differenzieren.
Die Datenreduktion hat ihre Wurzeln in der Psychoakustik; zur Kompression werden die bereits reduzierten Informationen in der Gleitpunktdarstellung gespeichert (vgl. Kap. 5.1.). Beide Faktoren ermöglichen zusammen eine Reduktion der Daten auf 20% des Vollformates.[111]

[111] Yoshida, Tadao: *The Rewriteable MiniDisc System*, Reprinted from the Proceedings of the IEEE USA, Oktober 1994, Bd. 82, Nr. 10, S. 1492–1500,
nach: http://www.amulation.co.uk/minidisc/ieee_paper.html.

5.3. MPEG Audio - Layer 3

1988 begann eine Arbeitsgruppe namens MPEG (MOVING PICTURES EXPERTS GROUP), Kompressionsstandards für Video und Audio zu definieren. Seitdem wurden drei Verfahren zur Audiokodierung vollständig entwickelt, die der Datenreduktion unter Beibehaltung der ursprünglichen Vollformatsqualität dienen: MPEG-1 Audio (1992), MPEG-2 Audio (1994) und MPEG-2 AAC (1997).[112]

MPEG-2 Audio stellt eine Erweiterung des MPEG-1 Audio-Standards um die Samplingfrequenzen 16 kHz, 22,05 kHz und 24 kHz dar, während der gesamte Kodieralgorithmus identisch ist.[113] Daher wird im Folgenden nur auf MPEG-1 Audio näher eingegangen.

MPEG-1 Audio ist Bestandteil des in ISO 11172 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION) beschriebenen MPEG-1-Standards[114] und bezieht sich auf die Einsatzgebiete Speicherung auf Magnetband, Digitales Radio und Live-Übertragung von Audio über ISDN-Leitungen.[115]

Entsprechend den Anforderungen der einzelnen Bereiche existieren drei *Layer*. Von Layer 1 bis Layer 3 steigt die Effektivität der Kodierung; das bedeutet jedoch nicht, dass, laut MEISSNER, die Audioqualität bei sinkender Datenrate steigt.[116] Vielmehr kann durch effektivere Enkodieralgorithmen der Reduktionsfaktor gesteigert werden, ohne schwerwiegende Qualitätsverluste zu verursachen, so dass die Audioqualität annähernd erhalten bleibt.[117]

---

[112] Brandenburg, Karlheinz 1998, a.a.O., S. 75.

[113] ebd.

[114] ISO/IEC International Standard IS 11172-3: *Information Technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 MBits/s – Part 3: Audio*.

[115] ebd.

[116] Meissner, Hansgeorg, a.a.O., S. 82.

[117] Effektivere und somit komplexere Algorithmen stellen höhere Ansprüche an die Rechenleistung des Systems.

Verfahren	Kompressions-verhältnis	Datenrate (Stereosignal)	Anwendung
Vollformat	1:1	1411,2 KBit/s	Audio-CD, DAT
Layer1	1:4	384 KBit/s	DCC
Layer 2	1:6 - 1:8	256-192 KBit/s	Digitaler Rundfunk: DAB
Layer 3	1:10 - 1:12	128-112 KBit/s	ISDN (Music-on-Demand)

Tab. 3: Gängige Kompressionsraten der MPEG-1 Audio-Layer[118]

MPEG-1 Audio verarbeitet PCM-Audiosignale mit der Samplingrate 32 kHz, 44,1 kHz und 48 kHz.

Layer 1 entspricht dem PASC-Verfahren (s. Kap. 5.1.) und wird z. B. in der DCC-Technik mit einer Datenrate von 384 KBit/s pro Stereosignal eingesetzt.

Layer 2 reduziert die Datenmenge so, dass eine Datenrate von 256 KBit/s für ein Stereosignal benötigt wird. Dieses Verfahren wird z.B. in der digitalen Rundfunkübertragung *DAB (Digital Audio Broadcasting)* eingesetzt.

Layer 2 unterscheidet sich von Layer 1 durch Modifizierungen der Bitzuweisung und der Skalierungsfaktoren für die Gleitpunktdarstellung und durch die Erweiterung auf fünf Kanäle durch die *multichannel extension*.[119] Unterschiede zwischen Layer 1 und Layer 2 sind unter psychoakustischen Aspekten jedoch unbedeutend.[120]

Der komplexeste Enkoder, Layer 3, ermöglicht die digitale Audioübertragung via ISDN bei einer Übertragungsrate von 128 KBit/s (vgl. Tab. 3).[121]

Layer 1 , 2 und 3 sind nach demselben Schema konstruiert; die unterschiedliche Leistungsfähigkeit resultiert aus der jeweiligen technischen Umsetzung der einzelnen Komponenten.

---

[118] Fraunhofer IIS: *MPEG Audio Layer-3 - History*, nach: http://www.iis.fhg.de/amm/techinf/layer3/index.html.

[119] Fraunhofer IIS: *Frequently Asked Questions about MPEG Audio Layer-3*, nach: http://www.iis.fhg.de/amm/techinf/layer3/layer3faq/index.html.

[120] Brandenburg, Karlheinz 1998, a.a.O., *S. 76.*

[121] ebd.

Da zum einen Layer 1 in seiner Funktionsweise bereits in Kapitel 5.1. (DCC) beschrieben wurde, zum anderen Layer 3 am effektivsten, also ohne (deutlichen) Qualitätsverlust die Bitrate reduziert, wird im Folgenden nur das Reduktionsverfahren MPEG-1 Audio-Layer 3, das häufig aufgrund der beschriebenen Übereinstimmungen der Audiokomponenten von MPEG-1 und MPEG-2 als *MPEG Audio-Layer 3* oder *MP3*[122] bezeichnet wird.

5.3.1. Signalzerlegung - Polyphasenfilterbank

Alle drei Layer zerlegen das PCM-Signal mittels einer Filterbank (sog. *Polyphasenfilterbank)* in 32 gleich große Frequenzsubbänder.[123] Diese Unterteilung spiegelt jedoch nicht die Gehörunterteilung in kritische Bandbreiten wider, da im tiefen Frequenzbereich mehrere kritische Bandbreiten in ein einziges Subband zusammengefass sind, so dass die ersten fünf Subbänder alle Frequenzen bis etwa 5000 Hz umfassen (s. Abb. 26: Die Grenze von 18 Bark zu 19 Bark liegt nach Tab. 1, S. 19 bei 5300 Hz, sofern dieser Darstellung die Bark-Skala nach ZWICKER zugrunde liegt).
Das Maß für die Unhörbarkeit des Quantisierungsrauschens kann somit nicht pro Frequenzgruppe bestimmt werden, sondern orientiert sich an der ‚schwächsten' Frequenzgruppe innerhalb des Subbandes.
Durch diese die Filterkomplexität mindernde Unterteilung des gesamten Frequenzspektrums in gleich große Subbänder wird dem Signal ein Fehler zugeführt, der allerdings unhörbar bleibt.[124]

---

[122] Die Bezeichnung „MP3" ist angelehnt an das Suffix der entsprechenden Audiodateien, die innerhalb von Computer-Betriebssystemen die Form *.mp3 haben.

[123] Pan, Davis: *A Tutorial on MPEG/Audio Compression.* Reprinted IEEE Multimedia Journal, Ausgabe Sommer 1995,
nach: www.eskimo.com/~miyaguch/mp3info.html .

[124] Pan, Davis 1995, a.a.O..

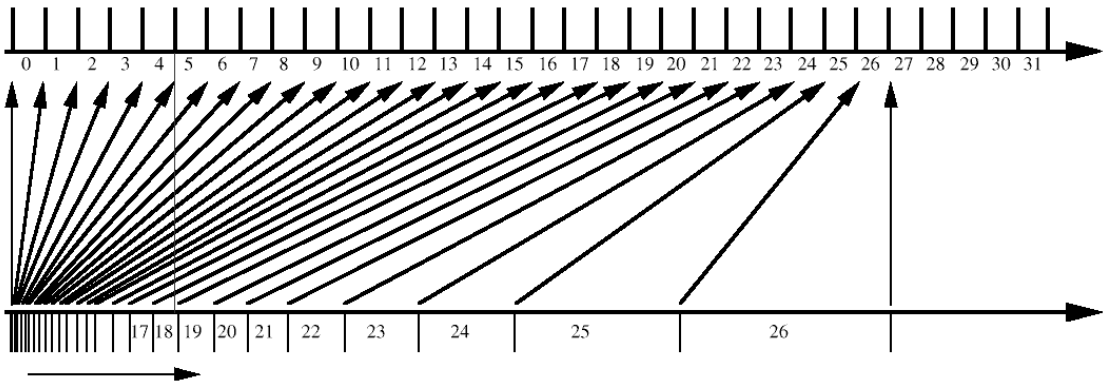


Abb. 26: Subbandunterteilung (oben) im Vergleich zur Anordnung der Frequenzgruppen (unten).

5.3.2. Hybrid-Filterbank

Hybrid-Filterbänke bestehen aus unterschiedlichen Filterbanktypen, die zur Verbesserung des Gesamtergebnisses kombiniert werden.
Um die eben beschriebenen Unzulänglichkeiten der Polyphasenfilterbank zu kompensieren, werden deren Ausgänge durch eine MDCT mit dem Resultat einer besseren Frequenzauflösung bearbeitet.[125]

Wie ATRAC verwendet Layer 3 alternative Zeitfenstergrößen:

- langes Zeitfenster: 36 samples
- kurzes Zeitfenster: 12 samples.[126]

Bei der Darstellung des ATRAC-Verfahrens wurde bereits aufgezeigt, dass das kurze Zeitfenster das Auftreten von pre-echos verhindert (s. Kap. 5.2.).

[125] Pan, Davis: *Digital Audio Compression*, in: Digital Technical Journal, Bd. 5, Nr. 2, Frühjahr 1993, nach: www.eskimo.com/~miyaguch/mp3info.html .
[126] ebd.

5.3.3. Psychoakustisches Modell

MPEG-Audio verwendet zwei unterschiedliche psychoakustische Modelle: Layer 1 und Layer 2 verwenden Modell 1, Layer 3 das komplexere Modell 2[127], das näher beschrieben werden soll:

Das psychoakustische Modell 2 analysiert das ankommende Audiosignal und bestimmt, wo im Signalspektrum Quantisierungsrauschen maskiert wird und welche maximale Stärke es haben darf, ohne wahrgenommen zu werden.

Seine wesentlichen Aufgaben lassen sich wie folgt zusammenfassen:[128]

- Mit dem Ziel eines qualitativ hochwertigeren Ergebnisses wird das Frequenzspektrum des Signals entsprechend den kritischen Frequenzbandbreiten unterteilt.
- Das psychoakustische Modell erkennt und trennt tonale und nicht-tonale Signalkomponenten. Da sich deren jeweilige Maskierungseigenschaften unterscheiden (s. Kap. 2.7.), werden ihnen *Tonalitätsindizes* zugewiesen, die als solche direkt in die Berechnung der Mithörschwellen einfließen.[129]
- Das Modell enthält eine empirisch bestimmte absolute Hörschwelle, nach der eine Vorauswahl bezüglich der Hörbarkeit der Klänge getroffen wird.[130]
- Das Modell errechnet die Mithörschwellen mit einer gegenüber der Polyphasenfilterbank stark erhöhten Frequenzauflösung, um der Tatsache gerecht zu werden, dass ein Subband im Bereich tiefer Frequenzen mehr als nur eine Frequenzgruppe beinhaltet und somit auch mehrere Maskierer aufweisen kann. Liegen die Dimensionen von Subband und Frequenzgruppe weit auseinander – was im tiefen Frequenzbereich gegeben ist – so benutzt das Modell die niedrigste Mithörschwelle für dieses Subband. Im hohen

[127] Pan, Davis 1995, a.a.O..

[128] nach: Pan, Davis 1995, a.a.O..

[129] Brandenburg, Karlheinz 1998, a.a.O., S. 62.

[130] Pan macht keine genauere Angaben über die Gestalt dieser empirisch bestimmten Hörschwelle.

Frequenzbereich, in dem die Breite der Subbänder weit geringer ist als die der entsprechenden Frequenzgruppen, wird der Durchschnittswert der Mithörschwellen ausgewählt.[131]

Abschließend errechnet das Psychoakustische Modell das Verhältnis von Signal zu Mithörschwelle (*signal-to-mask ratio)* innerhalb eines Bandes und übermittelt diesen Wert an die Bit-Zuweisung, die dann genau so viele Bits zur Verfügung stellt, dass das Quantisierungsrauschen unhörbar ist.

5.3.4. Joint Stereo Coding

Grundsätzlich ermöglichen zwei Stereoeffekte, die Datenrate noch weiter zu senken. Besonders bei Datenraten von 128 KBit/s und weniger spielt dabei die *Stereokodierung* eine wichtige Rolle.[132]

Zunächst führen Korrelationen zwischen beiden Kanälen zu *Stereoredundanzen*, da der Stereoeindruck in dem Extremfall der völligen Korrelation zwischen linkem und rechtem Kanal dem monophonen Klangbild entspricht. Daher lässt sich das Stereosignal um diese Korrelate reduzieren, um dann bei der Dekodierung im vollständigen Stereobild rekonstruiert zu werden.[133]

Der zweite Stereoeffekt basiert auf der Art und Weise, wie das Gehör Schallquellen im Raum orten kann: wie bereits in Kap. 2.6. erläutert, werden tiefe Frequenzen sowohl über die Amplituden- als auch über die Phasenbeziehung der am jeweiligen Ohr ankommenden Stimuli geortet, während Frequenzen oberhalb von 2 kHz nur noch durch Amplitudenunterschiede und nicht mehr durch die Phasenlage lokalisiert werden können.

---

[131] Zur Relation der Frequenzgruppen zur Subbandbreite sei verwiesen auf Abb. 26.

[132] Herre, Jürgen: a.a.O., S. 29.

[133] ebd.

Beide Stereoeffekte machen sich die im MPEG-Layer 3 verwendeten Verfahren, die *M/S-Stereokodierung* und die *Intensitäts-Stereokodierung* zunutze, die im Folgenden beschrieben werden.

5.3.4.1. M/S-Stereokodierung

Bei diesem Kodierungsmodus wird das „LR-Stereosignal“, bestehend aus linkem und rechtem Kanal, vor der abschließenden Quantisierung für jedes Subband in einen Summenkanal M_i (L_i+R_i) und einen Differenzkanal S_i (L_i-R_i) transformiert (i=Subbandindex 1,2,...,32).[134] Daher wird dieses Verfahren als Mitte/Seite-Stereokodierung (M/S-Stereokodierung) bezeichnet.[135]
Die meisten Stereosignale sind im Summenkanal stärker als im Differenzsignal, und je monophoner das Signal ist, desto stärker ist die Summe und desto schwächer die Differenz.[136] Betrachtet man ein Stereosignal mit zwei identischen, phasengleichen Sinustönen auf dem linken und dem rechten Stereokanal, so wird der Extremfall dieses Zusammenhanges deutlich: die Summe besteht aus der Überlagerung der Töne, die Wellen addieren sich und die Amplitude verdoppelt sich; die Differenzbildung jedoch hat zur Folge, dass beide Komponenten sich gegenseitig auslöschen. In diesem Fall muss nur der Mittenkanal anstelle der beiden Kanäle L und R kodiert werden, woraus eine Bit-Einsparung von 50% resultiert. Gibt es nur wenige Unterschiede zwischen den beiden Kanälen, so bleiben z.B. in der Differenz diese nicht in beiden Kanälen vertretenen Anteile erhalten. Es ergibt sich eine von der Signaleigenheit abhängige Stereoredundanz und somit ein Einsparungspotential von bis zu 50% (s. Abb. 27).

[134] ISO/IEC International Standard IS 11172–3, S. 144.

[135] Herre, Jürgen, a.a.O., S. 29.

[136] Rumsey, Francis, a.a.O., S. 86.

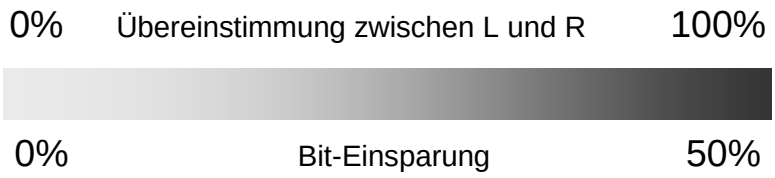


Abb. 27: Bit-Einsparungspotential in Abhängigkeit von der Stereoredundanz

Anhand von Kodierungs- und Dekodierungsmatrizen[137] werden die Stereosignale in der LR-Form in die MS-Form umgerechnet und umgekehrt.
Um weitere Bit-Ressourcen einzusparen, werden bestimmte Spektralwerte im Differenzkanal gleich Null gesetzt, sofern sie für den Stereoeindruck irrelevant sind. Dabei handelt es sich zum einen entsprechend dem ‚Subwoofer-Prinzip' um tiefe Frequenzen bis 1500 Hz (s. Kap. 2.6.), zum anderen errechnet der Kodierungsprozessor das Verhältnis der Signalstärke des Differenzkanals S_i (i=Subbandindex) und des zugehörigen Mittenkanals M_i. Unterschreitet dieses Verhältnis einen bestimmten vorgegebenen Wert, so gilt $S_i=0$. In diesem bestimmten Frequenzband i sind folglich die Stereoeigenschaften derart schwach ausgeprägt, dass sie vernachlässigt werden können.

5.3.4.2. Intensitäts-Stereokodierung

Bei der Intensitäts-Stereokodierung wird das Stereosignal pro Subband auf einen Summenkanal und je einen Intensitätsskalierungsfaktor für den linken und den rechten Stereokanal reduziert. Dabei werden der linke und der rechte Kanal zusammengeführt, während die Skalierungsinformation dafür sorgt, dass bei der Dekodierung im jeweiligen Kanal das Intensitätsniveau rekonstruiert werden kann (s. Abb. 28). Bei dieser Methode werden die ursprünglichen Wellenformen

[137] vgl. Herre, Jürgen, a.a.O., S. 30.

aufgrund der verlorenen Phaseninformationen verändert, so dass dieses Verfahren auf den Frequenzbereich oberhalb von 2 kHz beschränkt wird.[138]

Wie in Kapitel 2.6. beschrieben, basiert der räumliche Höreindruck oberhalb von 2 kHz hauptsächlich auf Intensitätsunterschieden des ankommenden Schalls an beiden Ohren. Die Information über etwaige Phasenunterschiede wird jedoch ab ca. 1500 Hz aufgrund der begrenzten Entladungsrate der Hörnervenzellen unbedeutend und muss somit nicht berücksichtigt werden.[139]

BRANDENBURG weist darauf hin, dass das Verfahren der Intensitäts-Stereokodierung daher relativ erfolgreich ist. Die dennoch existierenden wahrnehmbaren Verluste treten besonders dann zu Tage, wenn das dekodierte Signal über Kopfhörer abgespielt wird.[140]

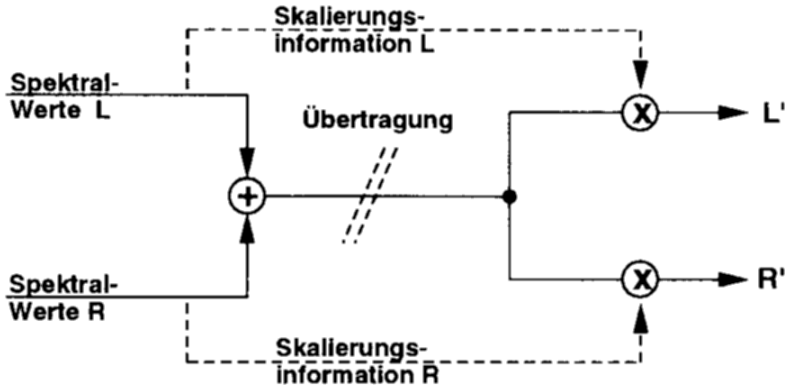


Abb. 28: Signalweg während der Kodierung und Dekodierung bei der Intensitäts-Stereokodierung

[138] Herre, Jürgen, a.a.O., S. 31.

[139] Die Entladungsrate der Nervenfasern hat auch im Zusammenhang mit der Frequenzselektivität des Hörsinns durch Periodizitätsanalyse limitierende Auswirkung (vgl. Kap.1.3.).

[140] Brandenburg, Karlheinz 1998, a.a.O., S. 71.

5.3.5. Huffman Coding

Abschließend wird das durch Perceptual Coding reduzierte Audiosignal nach dem *Huffman*-Algorithmus kodiert. Dabei werden den Audiodaten Codes mit variabler Wortlänge zugewiesen: Die am häufigsten vorkommende Bitfolge wird durch einen Code mit minimaler Bitanzahl substituiert, und mit sinkender Häufigkeit nimmt die Wortlänge zu.[141]

Allein durch dieses statistische Kodierungsverfahren sind Biteinsparungen von bis zu 20% möglich.[142] Zusätzlich findet bei MP3 ein symbiotisches Wechselspiel zwischen Huffman Coding und Perceptual Coding statt: In „lauten“ Passagen, in denen sehr viele Signalanteile verdeckt werden und das Psychoakustische Modell sehr effizient arbeitet, existieren nur wenige identische Informationen, so dass die Häufigkeitskategorisierung des Huffman-Kodierers ineffizient arbeitet. Umgekehrt enthält das „leise“ Signal, in dem nur wenige Maskierungseffekte entstehen, eine Vielzahl an redundanten[143] Bytes, die durch kürzere Codes ersetzt werden.

Diese Verflechtung von statistischen und psychoakustischen Kodierungsprinzipien verdeutlicht, weshalb es nicht ganz der Realität entspricht, wenn Audiokodierung lediglich in Hinblick auf gehörangepasste Kodierungsverfahren untersucht wird, da so ein verzerrtes Bild von deren Leistungsfähigkeit entsteht.

---

[141] vgl. *„MP3‘ Tech – Overwiew of the MP3 techniques“*, nach: http://freeflight.cockpit.be/mp3tech/tech.html .

[142] ebd.

[143] hier: mehrfach vorkommende, identische Bitfolgen.

5.4. Kritische Betrachtung der gehörangepassten Datenreduktion

Auch wenn in den vorangegangenen Kapiteln die komplexe Funktionsweise und der (quantitative) Erfolg verschiedener Reduktionsverfahren dargestellt wurden, darf nicht verschwiegen werden, dass im Umfeld dieser Technologie auch Probleme auftreten: Beispielsweise verschlechtert sich die Audioqualität erheblich, wenn das Audiomaterial mehrfach hintereinander mit Datenreduktionsverfahren bearbeitet wird. Dieser Vorgang wird als *Kaskadierung* bezeichnet und lässt sich im praktischen Einsatz kaum vermeiden, da die „Vergangenheit" eines Audiosignals nicht immer bekannt ist. Wird beispielsweise Tonmaterial von einem Hörer auf MiniDisc aufgezeichnet, das im Funkhaus mit Datenreduktion aufgenommen und schließlich im Internet oder über Digitalen Hörfunk nochmals datenreduziert ausgestrahlt wird,[144] so kann es zu einer Kaskadierung kommen, die die Klangqualität mindert. Bereits die Beeinflussung des datenreduzierten Materials durch einen Equalizer oder den Volume-Regler der Verstärkerendstufe kann das empfindliche Verhältnis von maskierten und maskierenden Signalanteilen so nachhaltig stören, dass die ursprünglich maskierten Anteile hervortreten.[145]

Die zwei weiteren Hauptprobleme im Zusammenhang mit der Datenreduktion lassen sich folgendermaßen formulieren:

1. Wie lässt sich die Qualität datenreduzierten Materials beurteilen?
2. Wie lässt sich die illegale Massenverbreitung im Internet eindämmen?

Sie betreffen zum einen die Entwicklung von derartigen Audiosystemen und zum anderen die Auswirkungen auf die Musik- und Tonträgerindustrie.

---

[144] Derzeit existieren etwa 600 Rundfunkstationen in Europa, die Internet-Radio mit Datenreduktion anbieten. Nach: Stoll, Gerhard: *Internet-Radio und excellente Tonqualität: Wunschtraum oder Wirklichkeit?*, http://www.irt.de/IRT/FuE/as/internetradio.htm.

[145] Bremm, Peter, a.a.O., S. 106.

Genaugenommen lässt sich die augenscheinliche Bewusstwerdung dieser Problematik bereits als deutliches Indiz für die Akzeptanz von reduzierten Audiodaten und somit für eine die Allgemeinheit zufriedenstellende Qualität werten. Diese beschränkt sich allerdings nicht nur auf die Klangqualität, sondern auch auf Faktoren wie Handhabung, Performance oder gar Design der Benutzeroberfläche.

5.4.1. Qualitätsbeurteilung als Grundproblem der Datenreduktion

BRANDENBURG nutzt eine sehr pragmatische, aber treffende Formulierung, um die Anforderungen an ein Datenreduktionsverfahren zu beschreiben:

> *„Bei keinem Musikstück sollen Unterschiede zwischen Original und verarbeitetem Stück wahrnehmbar sein.“*[146]

In Bereichen wie der Sprachkodierung wird der SNR[147] als Indiz für eventuelle Unterschiede und somit als Maß für Kodierungsqualität bestimmt, da er sich über eine unkomplizierte mathematische Gleichung errechnen lässt.[148] Mittels der Audiokodierung versucht man aber, das Maß an Quantisierungsrauschen so exakt zu begrenzen, dass das Rauschen nicht hörbar ist. Das ist der Fall, wenn es in seiner Lautstärke unterhalb der Ruhehörschwelle bleibt oder maskiert wird. Die Maskierungseffekte fließen in eine solche Formel zur Berechnung des SNR jedoch nicht ein, so dass beispielsweise trotz eines relativ guten SNR von über 60 dB das Quantisierungsrauschen hörbar sein kann.[149]

Um diese Schwachpunkte objektiver Qualitätsbeurteilung auszugleichen, werden Hörtests durchgeführt, in denen Versuchspersonen subjektiv ihren Höreindruck bewerten. Ein solches Verfahren bietet nach Angaben des FRAUNHOFER IIS die

---

[146] Brandenburg, Karlheinz: *Ein Beitrag zu den Verfahren und der Qualitätsbeurteilung für hochwertige Musikkodierung*, Dissertation an der Universität Erlangen-Nürnberg 1989, S. 5.

[147] SNR (signal-to-noise ratio) = Signal-Rauschabstand; vgl. Kap. 3.2. .

[148] Brandenburg, Karlheinz 1989, a.a.O., S. 16.

[149] ebd.

einzige, wenn auch sehr kostspielige Möglichkeit, aussagekräftige Qualitätsurteile über das datenreduzierte Materials zu erhalten.[150]

In den Versuchen werden direkte Vergleiche mit einer vorgegebenen Referenz durchgeführt, um sehr kleine Qualitätsbeeinträchtigungen zu erkennen und zu bewerten. Im Rahmen internationaler Standardisierung von Audiosystemen hat sich die *Triple stimulus-/hidden reference-/double blind*-Methode als Standard durchgesetzt.

5.4.1.1. Subjektive Qualitätsbewertung nach der Triple stimulus-/hidden reference-/double blind-Methode

Die Gestalt dieser Methode wurde in einer ITU-R-Richtlinie[151] festgelegt, um eine maximale Vergleichbarkeit unter verschiedenen Tests zu gewährleisten.

Während einer Testsequenz (*trial*) werden abwechselnd drei Versionen des jeweiligen Test-Exzerpts dargeboten. Aus der Folge: > reference – version A – version B < leitet sich die Bezeichnung *triple stimulus* ab.

Die einzelnen Komponenten sind:

- die offene Referenz (*reference*), die als das unbearbeitete Original kenntlich gemacht wird;
- die Versionen A und B, von denen die eine mit der Referenz identisch ist (*hidden reference*), die andere das zu bewertende Testsignal beinhaltet (*object*).

Die jeweilige Zuordnung von A und B ist der Testperson nicht bekannt und wechselt zufällig mit jeder neuen Sequenz (*blind test*). Der Versuchsleiter hat ebenfalls keine Kenntnis über die Konstellation eines trials, so dass unerwünschte Beeinflussungen der Testpersonen vermieden werden (*double blind*).

---

[150] Fraunhofer IIS: *Frequently Asked Questions about MPEG Audio Layer-3*, a.a.O..

[151] Standardisierung in Richtlinie ITU-R BS.1116: *Methods for the subjective assessment of small impairments in audio systems, including multichannel sound systems*, Genua 1994.

Durch den Vergleich mit der offenen Referenz sollen die Testpersonen nun unter den Versionen A und B das Testobjekt erkennen und bewerten. Der Test kann sowohl als Einzeltest als auch als Gruppentest durchgeführt werden. Der Einzeltest bietet dabei den Vorteil, dass die Testpersonen individuell zwischen den einzelnen Hörbeispielen umschalten können. Allein dieser Umstand kann *„die Empfindlichkeit der Testmethode wesentlich erhöhen“*[152].

Zum Zwecke der Vergleichbarkeit der Tests untereinander wurden Bewertungsskalen für den subjektiven Höreindruck von der ITU-R vorgegeben, die eine anschließende statistische Auswertung vereinfachen. (s. Abb.29).

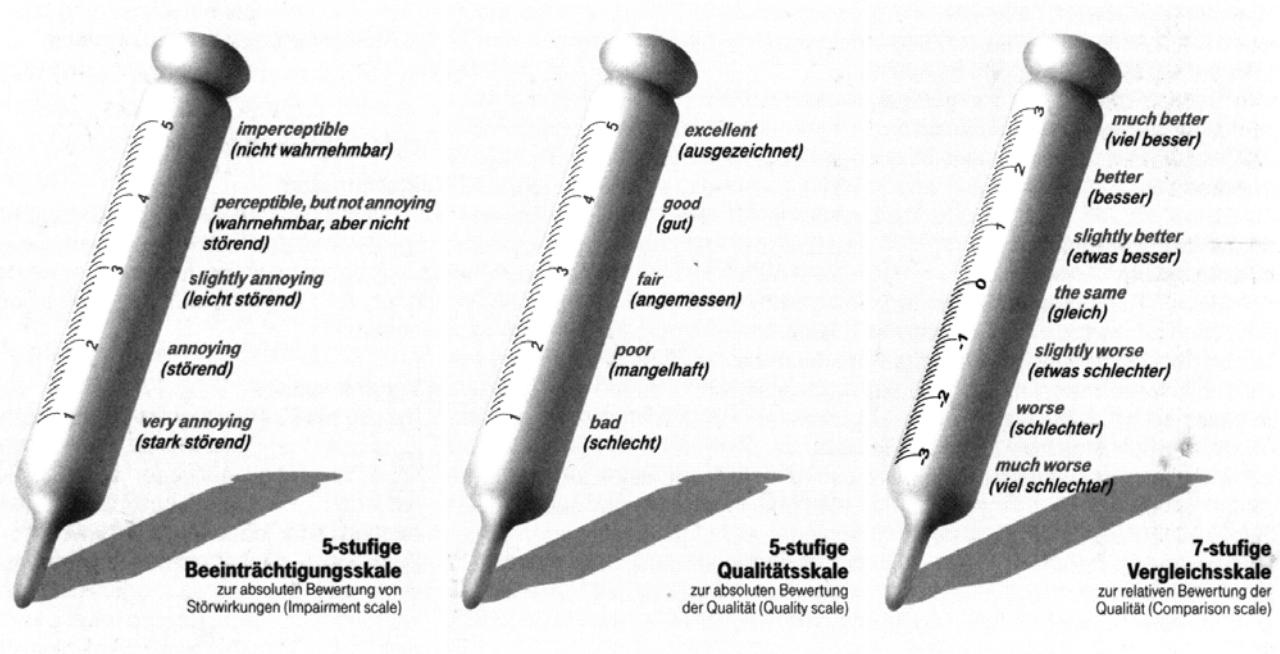


Abb. 29: Bewertungsskalen für die subjektive Qualitätsbewertung; englische Bezeichnung nach ITU-R BS.1116.

[152] Hoeg, Wolfgang/Steffen, Edith: *Methoden zur subjektiven Qualitätsbewertung von Audiosignalen*, in: Highlights aus der Forschung – Quellenkodierung und Qualitätsbeurteilung als Impulsgeber für Innovation bei Sprache, Audio und Video (Deutsche Telekom AG Hg.), Darmstadt 1996, S. 30.

Die Skalen umfassen die Kriterien:

- Wahrnehmbarkeit von Beeinträchtigungen (*impairment*);
- absolute Bewertung von Qualität (*quality*);
- Vergleich von Varianten (*comparison*).

Bei der Zusammenstellung des Testpersonenkreises liegt das Augenmerk nicht auf einer großen, den Bevölkerungsquerschnitt repräsentierenden, sondern auf einer relativ kleinen Gruppe[153] von *expert listeners*, also auf Personen mit ausreichender Urteilssicherheit, gutem Hörvermögen und Erfahrung im analytischen Hören. Unter Berücksichtigung der mit dem Alter zunehmenden natürlichen Hörverluste sollen die Probanden nicht älter als 40 Jahre sein.[154]

Da die Qualität eines Datenreduktionssystems nicht direkt, sondern nur indirekt über die Klangqualität bestimmter übertragener Hörbeispiele ermittelt werden kann, müssen diese bestimmte Anforderungen erfüllen: Die Exzerpte haben eine durchschnittliche Länge von jeweils 20 bis 30 Sekunden und sollen in ihrer Gesamtheit inhaltlich repräsentativ sein, also eine Vielzahl von Programmsituationen wie Sprache, klassische Musik, Popmusik oder Geräusche widerspiegeln. Neben der Repräsentativität haben jedoch auch andere Kriterien ihre Berechtigung: Die Grenzen eines Systems treten am deutlichsten unter *worst-case*-Bedingungen zu Tage. Daher enthalten die Programmausschnitte überwiegend repräsentative Passagen, die darüber hinaus als kritisch angesehen werden. BRANDENBURG weist in diesem Zusammenhang darauf hin, dass ein einheitlicher Katalog mit „kritischem“ Tonmaterial nicht existiert, da das kritische Moment mit kleinsten Veränderungen, z.B. der Enkoderparameter, sich ebenfalls verändert.[155] Beispielsweise können bei einem Kodierungsprozessor

[153] Hoeg, Wolfgang/Steffen, Edith, a.a.O., S. 31-34: Die Autoren geben 20 bis 30 Personen als ausreichend an (S. 31), stellen allerdings wenige Seiten später selbst fest, dass oft aufgrund erheblicher individueller Unterschiede innerhalb der subjektiven Urteile nur bei einer großen Anzahl an Testpersonen eine zuverlässige Aussage getroffen werden könne (S. 34).

[154] Hoeg, Wolfgang/Steffen, Edith, a.a.O., S. 31.

[155] Brandenburg, Karlheinz 1989, S. 91.

mit langem Zeitfenster starke Impulse zu dem bereits beschriebenen pre-echo-Effekt führen. Verändert man in diesem Kodierungsablauf nur den Parameter ‚Zeitfenstergröße' nach unten, so stellt dasselbe impulsartige Signal kein Problem mehr dar.[156]

Auch in ITU-R BS.1116 werden konkrete Anforderungen an den Hörtestraum und die Abhörvorrichtung gestellt, die eine Vergleichbarkeit mit anderen Tests und eine Reproduzierbarkeit der Testergebnisse erlauben sollen. Diese Empfehlungen in Form von Toleranzgrenzen beziehen sich auf:

- die *Betriebs-Schallpegelkurve* (auch als Übertragungskurve bezeichnet), gemessen am normierten Betriebs-Schalldruckpegel, zur Kennzeichnung der wichtigsten Eigenschaften des Direktschallfeldes,
- den *Nachhallzeitverlauf*, in Verbindung mit einer Limitierung der *Kurzzeitreflexionen*,
- die *Störgeräuschkurve*, gemessen unter verschiedenen Betriebsbedingungen (z.B. Klimaanlage ein-/ausgeschaltet). Die Einhaltung dieser Grenzen ist insbesondere dann von großer Bedeutung, wenn die zu erkennenden Störkomponenten im Bereich der Ruhehörschwelle angesiedelt sind.[157]

Abb. 30 zeigt grafisch die Teilergebnisse aus einem Versuch nach der Triple stimulus-/hidden reference-/double blind-Methode, der von der DEUTSCHEN TELEKOM mit unterschiedlichen MPEG-Konfigurationen durchgeführt wurde.
Verwendet wurde jeweils MPEG Audio-Layer 3 in stereo als Codec[158]-System B mit einer Datenrate von 128 KBit/s und als Codec-Variante C mit 256 KBit/s.

[156] vgl. Kap. 5.2./Kap. 5.3.2..

[157] Hoeg, Wolfgang/Steffen, Edith, a.a.O., S. 34.

[158] Codec = engl. Codierer und Decodierer.

Außerdem wurde die Versuchsanordnung derart modifiziert, dass nach Methode M1 die Testperson individuell den Versuchsablauf steuert und nach M2 der Ablauf fest in der Reihenfolge Referenz-A-B-Referenz-A-B-(...) vorgegeben ist. Auffällig ist die unterschiedliche Eignung der Exzerpte für das jeweilige Audiokodierungsverfahren, das in diesem Fall auf MPEG-1 Audio-Layer 3 (in stereo) basierte: während bei einer Bitrate von 256 KBit/s die Qualitätsminderung aller Hörbeispiele als nicht wahrnehmbar oder als wahrnehmbar, aber nicht störend eingestuft wurde, zeigten sich bei 128 KBit/s deutliche Unterschiede z.B. zwischen dem Sprachbeispiel (Qualitätsminderung störend) und dem Streichquartett (Qualitätsminderung wahrnehmbar, aber nicht störend).

Nr.	Beschreibung
1	Sprache deutsch, männlicher Sprecher
2	Cembalo, arpeggio
3	Rock Concert 1993 Gitarren, Violine, Klappern
4	Schlagzeug
5	Streichquartett

Tab. 4: Übersicht der verwendeten Programmausschnitte. Die Dauer eines Ausschnitts betrug jeweils ca. 25 Sekunden.

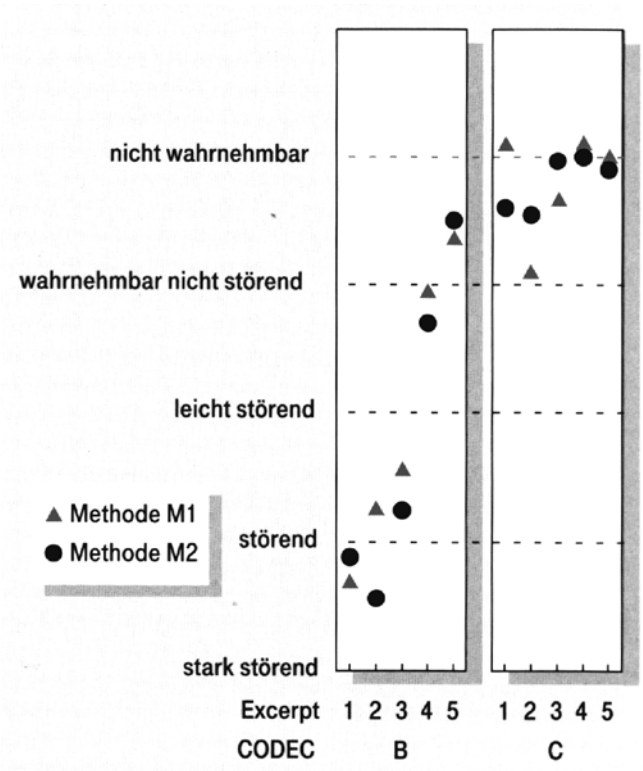


Abb. 30: Mittelwerte der Qualitätsminderung gegenüber der Referenz, verursacht durch das Codecsystem B und C, dargestellt für jedes Exzerpt 1 bis 5.

5.4.1.2. „CD-Qualität“ – ein realistisches Gütesiegel?

Mit „CD-Qualität“ wird im allgemeinen die Klangqualität eines Musik- oder Sprachsignals umschrieben. Der MP3-Standard ist bei den Online-Nutzern momentan sehr beliebt; gerade deshalb lassen sich an seinem Beispiel deutliche Unterschiede in der Verwendung des Terminus „CD-Qualität“ beobachten: Der Begriff tritt in diesem Zusammenhang in verschiedenen Mutationen auf:

- „CD-Qualität“ als Vergleichsmaßstab, als Referenz:
  „CD-like sound quality“[159] – Es werden potentielle Verluste eingeräumt, aber CD-Qualität wird annähernd erreicht.

- „CD-Qualität“ wird ersetzt durch „HiFi“:
  „Die Klangqualität hat dabei [bei der Umwandlung in MP3-Daten] *immer Hi-Fi-Niveau.“*[160]

- CD-Qualität als absolute (erreichte) Qualitätsstufe:
  „MPEG Audio Layer 3, kurz MP3 genannt, ist der etablierte Standard für Musikarchivierung und Audio on Demand in CD-Qualität bei geringem Speicherplatzbedarf.“[161]

- „CD“ wird als Synonym verwendet für datenreduziertes Tonmaterial:
  „Bei der kompletten Übertragung der Titel/CDs handelt es sich um eine T-Net-ISDN-Verbindung, welche die Musikdatenbank über zwei ISDN-Kanäle zu Ihnen aufbaut.“[162]

---

[159] Fraunhofer IIS: *Basics about MPEG Perceptual Audio Coding,* nach: http://www.iis.fhg.de/amm/techinf/basics.html .

[160] Deutsche Telekom: *Music On Demand – die neue Musik-Power im Internet,* Bonn 7/98, S. 4.

[161] http://www.mp3.de .

[162] Deutsche Telekom: *Music On Demand – Informationen für Kunden,* 2/98.

Einerseits wird der Nutzer durch die nicht zutreffende Beschreibung der Klangqualität in den Glauben versetzt, die zu erwartende „Musiklieferung“ entspräche in allen Anforderungen einer CD. Dass diese Hypothese nicht zutrifft, wurde bereits in Kap. 5.4.1.1. erläutert.
Andererseits zeigen die Hörversuchsergebnisse, dass durchaus die Möglichkeit besteht, datenreduziertes Material in der Qualität einer CD zu erhalten; der Erfolg der Reduktion hängt wie beschrieben maßgeblich vom zu reduzierenden Material ab. Daher zieht das Marketing Nutzen aus dieser Ambivalenz, da Aussagen wie „MP3 in CD-Qualität“ zwar nicht falsch sind, aber nicht die gesamte Realität offenlegen.

Die Vermarktungsstrategie zu MUSIC ON DEMAND, einer von der DEUTSCHEN TELEKOM ins Leben gerufenen und unterhaltenen Oberfläche zur Vermarktung von Musik via Internet[163], lässt vermuten, dass durch die Verwendung des Begriffs CD als Subsitut für MP3 einem neuen Musikmarkt, nämlich dem Online-Markt, die Attribute eines etablierten physischen Tonträgers verliehen werden sollen: Zunächst weckt der gängige Name „CD“ mehr Aufmerksamkeit als „MPEG“ oder „MP3“, so dass sich der Bekanntheitsgrad des neuen Produkts schneller steigern lässt. Gleiche Wirkung erzielt der Hinweis auf HiFi-Qualität, da die Zielgruppe weitgehend identisch ist.
Außerdem werden Erfahrungen, die im Umgang mit der „alten“ Technologie Compact Disc gemacht wurden, übertragen auf die „neue“ Technologie. Diejenigen Konsumenten, deren Skepsis z.B. hinsichtlich der Qualität der datenreduzierten Musik nur in geringem Maße vorhanden ist, werden möglicherweise eher dazu bewogen, sich mit MP3 auseinanderzusetzen.

Vor allem aber wird versucht, eine Verbindung herzustellen zwischen zwei unterschiedlichen Marktformen:

[163] vgl. Kap. 5.4.2.1..

> *„Der ‚Unique-Selling-Point' (USP) für physische und für nicht-physische Angebote ist einfach nicht identisch. Dies wird dazu führen, dass beide Absatzwege nebeneinander existieren können und werden."*[164]

Das Kaufverhalten des Musikkonsumenten ist ferner durch die Vergangenheit geprägt, in der Musikkonserven in Form von physischen Tonträgern gekauft wurden: Hier stehen sich die ‚greifbaren' Tonträger, angeführt von der CD, und die ‚unsichtbaren' Zahlenkolonnen der Audiofiles gegenüber. Geht man davon aus, dass sich im Kauf eines dieser Produkte in gewisser Weise das Bedürfnis nach gegenständlichen Besitztümern äußert, so ist zu erwarten, dass auf lange Sicht die CD z.B. wegen ihres individuellen Covers bevorzugt wird, das fester Bestandteil des Gesamtprodukts ist.
Die Welt der Digitaltechologie war schon immer begleitet von der Frage nach den psychologischen Auswirkungen ihrer steigenden Abstraktion. Beispielsweise wird in Zusammenhang mit den Umsatzverlusten durch illegale Raubkopien regelmäßig diskutiert, ob der Mensch den Wert eines geistigen Produkts in der virtuellen Realität überhaupt noch anerkennt.

Gegen dieses Konsumverhalten müssen sich Anbieter wie MUSIC ON DEMAND behaupten. Langfristig gesehen hängt deren Bestehen davon ab, ob sich die dahinterstehende Technologie behaupten wird, kurzfristig bedeutet das jedoch, den vorhandenen, traditionellen Markt zu nutzen, indem das Produkt „datenreduzierte Musik" mit den Attributen der CD gleichgesetzt wird.

---

[164] Zombik, Peter: *Music-on-Demand: „Tanz auf dem Vulkan" oder „Gib dem Affen Zucker"? – Chancen und Gefahren der neuen Technologie,*
nach: http://www.ifpi.de/recht/re-6.htm.

„Qualität“ beschreibt die Güte oder den Wert eines Objektes. In diesem Sinne lässt sich die unterschiedliche Anwendung des Begriffs „CD-Qualität“ folgendermaßen interpretieren:

Auf der einen Seite steht „CD-Qualität“ als Gütekennzeichen. Es wird von Annäherung des kodierten Materials an die Klangqualität der CD gesprochen, jedoch deutlich auf bestehende Defizite hingewiesen. Diese Form ist meist im wissenschaftlich orientierten Kontext wie z.B. auf entsprechenden Web-Seiten des FRAUNHOFER IIS anzutreffen.[165]

Dem gegenüber steht der wertmäßige Gebrauch des Kennzeichens „CD-Qualität“, der sich auf den Wert des Produktes bezieht, also auf werbepsychologischen Intentionen basiert. „CD-Qualität“ fungiert in diesem Fall als Werbeslogan. Deutlich wird dies ebenfalls auf den Web-Seiten des FRAUNHOFER IIS: Die entwickelten Verfahren zur Datenreduktion werden in Form von Lizenzen zur Nutzung des MP3-Kodierers verkauft; dieser Absicht entsprechen Seiten, die auf die Vermarktung des MPEG-Enkoders ausgerichtet sind und folglich der Produktpräsentation dienen – hier wird uneingeschränkt mit „*the original CD sound quality*“ geworben.[166]

Die Hörgewohnheiten und die Erwartungen des Nutzers sind sehr unterschiedlich: Auf dem CD-Markt existieren CDs mit hoher Klangqualität, in der Fachpresse auch als „Referenz-CDs“ bezeichnet, und gleichzeitig werden minderwertigere CDs angeboten, denen schlechtes Audiomaterial zugrunde liegt. Letzteres kann bedingt sein durch verrauschte Mikrophonsignale bei der Aufnahme, minderwertige Masterbänder und -discs oder auch alte, analoge Vorlagen für das Remastering. Ferner hängt der Höreindruck von der Qualität der einzelnen HiFi-Komponenten ab, über die die CD abgespielt wird. Und letztlich beeinträchtig natürlich das individuelle Hörvermögen auch das

[165] vgl. Fraunhofer IIS: *Basics about MPEG Perceptual Audio Coding*, a.a.O..

[166] Fraunhofer IIS: *MPEG Audio Layer-3*,
nach http://www.iis.fhg.de/amm/techinf/layer 3/index.html.html .

subjektive Qualitätsempfinden; ein allgemeingültiger Klangqualitätsstandard lässt sich daher nicht definieren.

Auf der Basis dieser Überlegungen sollte genau untersucht werden, worauf die jeweilige Qualitätsbeschreibung basiert und in welchem, ggf. zweckgebundenen Kontext sie zu sehen ist.

5.4.2. Schutz der Urheberrechte

Ein vor allem in der jüngsten Vergangenheit auftretendes Problemfeld stellt die Einhaltung und der Schutz von Urheberrechten dar: Mit entsprechender Dekoder-Software[167], die für jedermann via Internet zugänglich ist, lassen sich MP3-Dateien auf dem privaten Personal Computer abspielen:

> *„Es ist die Schreckensvision der Musikindustrie: das Internet als riesiger Selbstbedienungsladen. Alles kostenlos, alles verfügbar, alles außer Kontrolle.“*[168]

Der illegalen Verbreitung von Musik im Internet ist kaum Einhalt zu gebieten, da Raubkopierer nicht mehr physische Vervielfältigungsexemplare anfertigen müssen, die zudem häufig mit Qualitätseinbußen verbunden sind. Vielmehr werden Musikdateien „online“ angeboten, die sich jeder Internetnutzer verlustfrei herunterladen kann, so dass das world wide web die Rolle des unkontrollierbaren „*Multiplikators*“[169] übernimmt.

Während sich bisher MP3-Files nur über einen Computer mit entsprechender Ausstattung (Soundkarte, Playersoftware) abspielen ließen, werden zum Ende

---

[167] Unter www.mp3.de werden z.B. WinPlay 3 (Fraunhofer IIS/Freeware) und Winamp (Nullsoft/Shareware) zum kostenlosen Download angeboten.

[168] Gröbchen, Walter: *MP3 – eine Formel zur Datenkompression schockt die Musikbranche,* in: DIE ZEIT 1998 Nr. 44,
nach: http://www4.zeit.de/bda/int/zeit/print/199844.mp3_.html.

[169] Schaefer, Martin/ Braun, Thorsten: *„Verantwortlichkeit im Internet“- ein neues Schlagwort betritt die Szene,*
nach: http://www.ifpi.de/recht/re-2.htm .

des Jahres 1998 portable MP3-Player auf den Markt kommen. Die Audiodaten werden über eine Docking-Station eingelesen und gespeichert.[170] Mit der Begründung, dass die Mehrheit der MP3-Player-Nutzer das kommerzielle Musikbusiness, also die „*Ertragsquelle*“[171] für die Tonträgerindustrie, umgeht, indem sie illegale Internetseiten benutzt, um den Speicher ihres Players mit Musik zu füllen, erwirkte die RIAA (RECORDING INDUSTRY ASSOCIATION OF AMERICA) vor Gericht den Produktionsstop eines Players der Firma DIAMOND MULTIMEDIA SYSTEMS,[172] der nach wenigen Tagen gerichtlich wieder aufgehoben wurde. Da laut Urteilsbegründung das Gerät keine „*seriellen Kopien*“ erstellen kann, darf der MP3-Player weiterhin ohne Zahlung von Abgaben an Verwertungsgesellschaften (in den USA) vermarktet werden. In Deutschland sind keine derartigen Lizenzprobleme zu erwarten, „*da das Gerät der GEMA bereits vorgestellt und laut §54 des Urheberrechtsgesetzes als Tonaufzeichnungsgerät lizenziert worden ist*“.[173]

Um Anbietern von Musik die Möglichkeit zu bieten, bei einer nicht physischen Auswertung von Musik „*die wirtschaftliche und rechtliche Hoheit über die Auswertung ihrer Inhalte* [zu behalten], *die sie bei der physischen Auswertung zweifellos besitzen*“[174], wurde in Deutschland der Online-Dienst MUSIC-ON-DEMAND entwickelt, der mittlerweile im Internet angeboten wird.[175]

---

[170] Carstens, Matthias: *Musik-Speicher: Layer 3 auf Reisen – der etwas andere Walkman*, in: c’t 21/98, S. 244.

[171] Thurow, Norbert: *Sechs Thesen zum Thema: Digitales Zeitalter und Musikleben*, nach: http://www.ifpi.de /recht/re-digi.htm.

[172] Schulemann, Gaby: *Krieg den MP3-Musikpiraten*, in: c’t 22/98, S. 18.

[173] Wiedenhoff, Chris: *Abgeschmettert*, in: c’t 23/98, S. 20.

[174] Schaefer, Martin: *Vertragliche Gestaltungsmöglichkeiten bei Online-Musiknutzungen, insbesondere bei Musik-Abrufdiensten*, nach: http://www.ifpi.de/recht/re-vertr.htm.

[175] Zugang zu MUSIC-ON-DEMAND: http://www.audio-on-demand.de.

5.4.2.1. Music-on-Demand

Das von der Tonträgerindustrie favorisierte Modell des Online-Musikmarktes versucht, die aus dem Geschäft mit physischen Tonträgern bekannten ökonomischen und juristischen Strukturen zu übernehmen. So wurde zwischen den Tonträgerherstellern als *Inhalteanbietern* und der DEUTSCHEN TELEKOM als *Diensteanbieter* ein Rahmenvertrag geschlossen:

Der Tonträgerhersteller bietet seine Aufnahmen über das T-ONLINE-Netz (der TELEKOM) dem Konsumenten direkt an. Dieser wählt auf der zu MUSIC-ON-DEMAND (MOD) gehörenden Online-Plattform aus dem Musikangebot aus. Der Kunde hat die Möglichkeit, einzelne Stücke probe zu hören, um anschließend seine Auswahl direkt beim Tonträgeranbieter zu bestellen. Die in MP3 kodierte Musik wird dann in Echtzeit über 2 gekoppelte ISDN-Kanäle[176] ausgeliefert und auf dem Computer des Kunden gespeichert. Die Daten können so mit gleichbleibender Datenrate übertragen werden, wodurch Qualitätsschwankungen im Gegensatz zu einem Transfer via Internet vermieden werden.

MOD ist nur registrierten Nutzern zugänglich, die sich durch Eingabe Ihres Nutzernamens und -kennwortes identifizieren müssen. Dies bietet den Vorteil, dass zum einen der Kunde vor Mißbrauch seines Zugangs durch Fremde geschützt wird und dass andererseits den Musikanbietern durch spezielle Kodierungsverfahren ein Mittel bereitgestellt wird, illegale Nutzungen des Musikangebotes zurückverfolgen zu können. Zu diesem Zwecke werden die Musikfiles vor der Lieferung an den MOD-Kunden so kodiert, dass sie ein unhörbares ‚Wasserzeichen' enthalten, das genauen Aufschluß darüber gibt, an wen die Musik verschickt wurde. So kann beispielsweise die illegale

[176] Durch die Kopplung zweier ISDN-Kanäle wird eine Übertragungsrate von 128 KBit/s erreicht. MUSIC-ON-DEMAND stehen für die reine Musikübertragung 112 KBit/s zur Verfügung, die restlichen Kanalkapazitäten werden für „außermusikalische" Zwecke wie das Übertragungsprotokoll benötigt.

Verbreitung der MP3-Dateien via Internet zurückverfolgt werden, um gerichtliche Schritte gegen den entsprechenden Kunden einzuleiten.

Die DEUTSCHE TELEKOM hat als Diensteanbieter die Aufgabe, über ihren T-ONLINE-Dienst den Zugang zu den einzelnen Musikanbietern zu ermöglichen und *„die störungsfreie und technisch perfekte Übermittlung der Musikdaten zum Kunden zu gewährleisten.“*[177] Die TELEKOM stellt dem Kunden für die Übertragung anfallende Kosten in Höhe von derzeit 0,232 DM pro Minute Musiklieferung in Rechnung.[178]

Den eigentlichen, vom Kunden zusätzlich zu zahlenden Preis der Musikwerke legen die Tonträgerhersteller individuell fest. Allerdings wurde der TELEKOM aufgetragen, die fälligen Rechnungsbeträge über die Telefonrechnung vom Kunden einzufordern.[179]

Darüber hinaus hat sie sich wegen ihrer Rolle als Multiplikator verpflichtet, der GEMA eine monatliche Auflistung der Musikwerke zu übermitteln, die auf den Servern der DEUTSCHEN TELEKOM von Tonträgerherstellern abgelegt wurden.[180]

MOD steckt noch in den Anfängen; daher wird sich erst in Zukunft erweisen, inwieweit der Markt der nicht-physischen Musikformen expansionsfähig ist, welche neuen Absatzpotentiale freigesetzt werden und wie sich die Koexistenz von physischem und nicht-physischem Musikformat gestaltet. Es ist zu erwarten, dass beide Formate wegen ihres eigenen USP weiterhin koexistieren werden.

Da die Tonträgerindustrie jedoch die Angebots-, Preis- und Gestaltungshoheit in Absatzformen wie MOD besitzt, hängt von ihr in maßgeblicher Form ab, ob und inwiefern das Online-Musikgeschäft ein wirtschaftlicher und kultureller Erfolg im Einklang mit den urheber- und medienrechtlichen Grundsätzen wird.

---

[177] Schaefer, Martin, a.a.O..

[178] Deutsche Telekom: *Music On Demand – die neue Musik-Power im Internet,* Bonn 7/98, S. 10.

[179] Wolf, Alexander: *Urheberrechtliche Lizenzeinräumung und Kontrollrechte einer Verwertungsgesellschaft am Beispiel des Music-on-Demand-Projekts der Deutschen Telekom,* in: GEMA-Nachrichten, Ausgabe 157, 6/98, S. 83.

[180] ebd.

6. Schlussbetrachtung und Ausblick

Eine kritische Auseinandersetzung mit der Datenreduktion von Musik findet derzeit fast ausschließlich in Verbindung mit dem Thema ‚Musik im Internet' statt, da auf anderen Anwendungsgebieten der verwendete Reduktionsfaktor geringer ist und die Qualität des Audiosignals als qualitativ hochwertig wahrgenommen wird (s. Kap. 5.4.1.1.).[181] Die Audiodatenreduzierung im Internet hat jedoch nur dann Sinn, wenn ein vernünftiges Verhältnis zwischen Qualität und Übertragungskapazität besteht.
In absehbarer Zeit wird das technische Problem der begrenzten Kapazität von Datennetzen durch die Einführung von *ADSL (Asymmetric Digital Subscriber Line)* entschärft werden, da mit dieser Technik Übertragungsraten von bis zu 8 MBit/s[182] erreicht werden.[183] Aufgrund der Preisgestaltung der TELEKOM bleibt jedoch vorerst abzuwarten, wann und in welchem Ausmaß der durchschnittliche Konsument Interesse an ADSL zeigen wird.[184]

Der Heimanwender wird durch die – zweifelsohne vorhandenen – technologischen Leistungen angesprochen, die bei der „unhörbaren" Reduktion vollbracht werden. Das explosionsartige Wachstum der MP3-Gemeinde im Internet setzte erwartungsgemäß in dem Moment ein, als MP3-Enkoder und -Dekoder im Internet als Freeware zu bekommen waren. Die Zahl der

[181] Als Beispiel hierfür seien hier MP3 mit einer Datenrate von 128 KBit/s im Gegensatz zur MiniDisc mit 256 KBit/s und das *Digital Audio Broadcasting (DAB)* genannt, das sein Radioprogramm in datenreduzierter Form bei 192 KBit/s sendet. Weitere Informationen zu DAB unter http://www.dab-plattform.de.

[182] Zur Erinnerung: die Audiodatenrate einer CD liegt bei 1,411 MBit/s.

[183] nach: http://www.adsl.com/adsl_forum.html.

[184] Lt. telefonischer Auskunft des Telekom-Kundendienstes vom 11.12.1998 liegt im derzeitigen Pilotversuch die monatliche Nutzungsgebühr für den ADSL-Anschluss mit 8 MBit/s bei über 1200,– DM;
s. auch http://www.telekom.de/angebot/bba/index.htm.

Internetadressen, hinter denen sich große MP3-Bibliotheken verbargen, nahm entsprechend rapide zu.

Inzwischen hat sich der Trend umgekehrt; die meisten Links auf sogenannten MP3-Pools sind ‚verwaist', da die Strafverfolgungsmaßnahmen gegen Raubkopierer mittlerweile Wirkung zeigen, so dass den Nutzern des WWW die Grenze zur Illegalität bewusst gemacht wurde.

Kommerzielle Systeme wie portable MP3-Player können dem Handhabungskomfort einer CD oder einer MD noch nicht entsprechen, da die Speicherkapazitäten noch zu gering sind (64 MB ermöglichen eine Spielzeit in MP3 von ca. einer Stunde). Der Austausch des Musikprogramms ist nur über einen PC möglich und nimmt etwa 15 min in Anspruch.[185] Künftige Systeme könnten diese Nachteile durch den Einsatz von Miniaturfestplatten kompensieren.

Auch lässt sich ein Personal-Computer optisch schwer mit der Vorstellung von einer HiFi-Anlage und deren spezieller „Ästhetik" vereinbaren, selbst wenn dieser inzwischen standardmäßig einige Gigabyte an Festplattenspeicher zur Verfügung stellt, so dass er sich bestens als MP3-Jukebox eignet.

Außerdem können die Preise der online angebotenen Musikwerke dem Vergleich mit den Preisen einer CD oftmals nicht standhalten. Ein Beispiel: Zwei willkürlich ausgewählte Angebote wurden eingeholt, jeweils bei MUSIC-ON-DEMAND und einem großen CD-Versandhandel, der die Möglichkeit der Online-Bestellung bietet. Die gesuchten Objekte waren die Alben der DIRE STRAITS*: Brothers In Arms* (low price) und von STING: *Mercury Falling* (normal price).

Abb. 31 veranschaulicht die deutlichen Preisunterschiede: Der Preis des MoD-Angebotes ist, zumindest im oben beschriebenen Falle, nicht so attraktiv wie der

[185] Carstens, Matthias: *Musik-Speicher: Layer 3 auf Reisen – der etwas andere Walkman*, in: c't 21/98, S. 244.

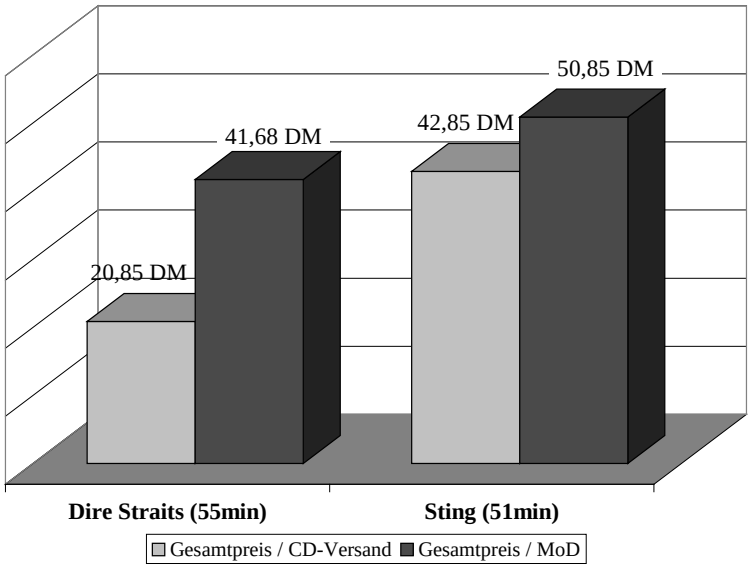


Abb. 31: Preisvergleich zwischen MUSIC-ON-DEMAND und CD-Versand anhand zweier willkürlich ausgewählter Alben.

Die ermittelten Preise setzen sich zusammen aus:

	MoD / an Plattenfirma	MoD / Übertragung 0,23DM pro min	CD-Versand / Tonträgerpreis	CD-Versand / Porto
DIRE STRAITS (55min)	28,92 DM	12,76 DM	14,95 DM	5,90 DM
STING (51min)	39,02 DM	11,83 DM	36,95 DM	5,90 DM

Kosten für die Online-Bestellung wurden nicht berücksichtigt, da beide Bestellungen über das Internet aufgegeben werden.
(MoD: http://www.audio-on-demand.de <> CD-Versand: http://www.jpc.de; Stand: 05.12.1998)

des CD-Versands. Auch der angebliche Vorteil von MOD, einzelne Titel bestellen zu können, scheidet in der Realität aus, weil häufig das Album nur in voller Länge angeboten wird – auch die beiden oben genannten CDs konnten nur komplett bestellt werden.

Die Möglichkeiten von MOD lassen sich jedoch auch anders nutzen: Es bietet sich an, Alben, die nicht mehr als Tonträger verfügbar sind, auf dem Online-Weg anzubieten. Diese Reaktivierung hätte doppelten Nutzen: Zunächst würden zusätzliche Verkäufe ermöglicht, die nicht zu Lasten bestehender Vertriebswege gingen. Außerdem würde der Konsument in die Lage versetzt, jegliches verfügbare Musikwerk erwerben zu können; wenn nicht in physischer, dann zumindest in nicht-physischer Form.

Während bei der Einführung des MPEG Audio-Layer 3 unter vermutlich werbestrategischen Gesichtspunkten oft verschwiegen wurde, dass durch die

relativ starke Reduktion Verluste in der Klangqualität hervorgerufen werden, ist zu beobachten, dass jetzt zunehmend auf diese durch die Kodierung entstehenden Verluste hingewiesen wird. Ob der durchschnittliche Rezipient datenreduzierter Musik diese Klangbeeinträchtigungen überhaupt wahrnehmen kann, ist fraglich. Das bestätigen die durchgeführten Versuche, bei denen selbst *expert listeners* nicht bei allen Hörbeispielen Beeinträchtigungen registrierten.[186] Daher existiert die Frage nach der Klangqualität tatsächlich nur im Kreise weniger kritischer Personen, für den eigentlichen Markt ist sie jedoch irrelevant. Andererseits kann die Thematisierung dieses Sachverhaltes dazu führen, dass sich eine *hypothesengeleitete Wahrnehmung*[187] einstellt, der Hörer also aufgrund seines Wissens in das Klangbeispiel Mängel hineinprojiziert. Das bedeutet umgekehrt natürlich auch, dass er als „CD-Qualität" erachtet, was ihm als solche angepriesen wird.

Nach einer MEDIA PERSPEKTIVEN-Studie, die im Rahmen des DAB-Pilotprojektes in Nordrhein-Westfalen durchgeführt wurde, lobten die befragten DAB-Nutzer vor allem die hohe Klangqualität der DAB-Sendungen.[188] Diese eindeutig positive Bewertung muss allerdings vor dem Hintergrund gesehen werden, dass heute Radio noch überwiegend in analoger Form ausgestrahlt wird. Somit hat der Vergleich von digitalem DAB mit analogem UKW-Radio nicht die gleiche Aussagekraft wie z.B. der Vergleich von MP3 und CD, die beide auf der Digitaltechnik basieren.

---

[186] vgl. Kap. 5.4.1.1. .

[187] vgl. Behne, Klaus-Ernst/Barkowsky, Johannes: *Analoge und digitale Musikwiedergabe im unmittelbaren Vergleich – Eine Studie zur hypothesengeleiteten Wahrnehmung*, Hannover 1994, S. 10.

[188] Kliment, Tibor: *Digital Radio im Urteil der Hörer – Ergebnisse einer Akzeptanzuntersuchung von DAB in Nordrhein-Westfalen*, Media Perspektiven 11/98, S. 551.

Zusammenfassend erscheint es also als sinnvoll, Datenreduktion in unmittelbarer Beziehung zum Anwendungsgebiet zu betrachten. Ab einem bestimmten Reduktionsgrad, der sowohl vom Kodierverfahren als auch von der individuellen Gestalt des Ausgangsmaterials abhängt, wird die Qualität hörbar beeinträchtigt. Daher eignen sich hohe Reduktionsraten vor allem für Anwendungen, bei denen nicht der Hörgenuß, sondern vielmehr der Informationsgehalt im Vordergrund steht.

Solche Anwendungsgebiete finden sich im gesamten Multimediasektor. Beispielsweise können Werbungen und Produktpräsentationen klanglich unterstützt werden; beispielhaft dafür ist die Probehörfunktion von MUSIC-ON-DEMAND, bei der die ersten 20 Sekunden in einer Qualität angeboten werden, die schlechter ist als die der eigentlichen Musiklieferung.[189] Ein anderer Verwendungszweck könnte darin liegen, Live-Gesprächsmitschnitte datenreduziert im Internet zugänglich zu machen.

Diese Beispiele stellen nur einen kleinen Ausschnitt des Bedarfs dar, der im heutigen Informationszeitalter herrscht; entscheidend ist, dass hier eine eindeutige Orientierung am Informationsgehalt dominiert, so dass das Audiomaterial zwar verständlich sein muss, aber niemand dessen Wert daran festmacht, ob der räumliche Eindruck oder der Klang in den tiefen Frequenzen adäquat wiedergegeben ist. Ein höherer Qualitätsverlust würde folglich toleriert.

Den Gegenpol zur Information bildet der Konsum, hier speziell der Musikkonsum. Im Vordergrund steht der Hörgenuß, der dem ständigen Vergleich mit der CD ausgesetzt ist.

Hier muss jene Qualität gewahrt werden, die häufig als „CD-Qualität“ scheinkonkretisiert wird. Bei der Entwicklung von Audiokodierungsverfahren, die auf Datenreduktion basieren, wird, wie bereits erläutert, als „CD-Qualität“

[189] Der Probehörausschnitt wird in Mono übertragen.

diejenige Schwelle angesehen, ab der der Hörer Klangbeeinträchtigungen wahrnimmt.
In diesem Extrembereich der Datenreduktion lässt sich kein umfassender, am Hörgenuß orientierter Qualitätsstandard ansiedeln. Vielmehr zeigt der derzeitige Erfolg der MD, die oft als Nachfolgerin der analogen MC gesehen wird, dass es sinnvoll ist, die Datenreduktionsrate zugunsten der Erhaltung der Klangqualität zu senken. Der Speicherplatzbedarf steigt zwar im Vergleich zu MP3 an, wird jedoch ausgeglichen durch weniger wahrnehmbare Klangbeeinträchtigungen.
Ist die Übertragungsrate jedoch aus technischen Gründen, wie z.B. durch hohe Netzauslastung im Internet, begrenzt, so bieten Verfahren zur Audiokodierung die Möglichkeit, Audioinformationen ohne schwerwiegende hörbare Verluste zu reduzieren.
Allerdings macht die kritische Auseinandersetzung mit Veröffentlichungen im Internet oder in Computermagazinen deutlich, dass mit dem hohen Diskussions- und Informationsbedarf häufig ein großer Wissensmangel hinsichtlich der technischen und klanglichen Möglichkeiten der besprochenen Verfahren einhergeht, wodurch sich möglicherweise auch der zuweilen unbeabsichtigte Mißbrauch des Gütesiegels „CD-Qualität“ erklären lässt.

Die vorliegende Arbeit legt die konstruktive Verflechtung von Wissenschaft, Technologie und Ökonomie dar. Der geleistete Überblick über die empirischen Untersuchungen zur Psychoakustik hat gezeigt, dass die Physiologie des Gehörs eine Vielzahl von Ansatzmöglichkeiten zur Entwicklung von hochwertigen Datenreduktionsverfahren bietet. Die technische Umsetzung der psychoakustischen Erkenntnisse führt ihrerseits zur Erschließung neuer Märkte.
Die Musikbranche stellt einen bedeutenden Wirtschaftszweig dar[190], deren Ziel letztlich darin besteht, Umsatz zu erzielen. Dies geschieht unter anderem durch

[190] Das Gesamtvolumen des deutschen Tonträgermarktes betrug 1997 5,375 Mrd. DM, nach: Hansen, Thomas/Zombik, Peter: *Wirtschaftsbericht 1997 – Der Markt für Compact Discs, MusiCassetten und Schallplatten*,

die Erschließung neuer Absatzmärkte. Ein solches Absatzpotential bietet der Markt im Internet, der noch in der Entstehung ist und dessen zukünftige Ausmaße noch nicht zu überschauen sind.

Die Kapazität der Netze gibt die Rahmenbedingungen für den Online-Handel vor: Der größte Anspruch an ein virtuelles Kaufhaus besteht darin, dass seine Produkte im Datennetz komfortabel und kostengünstig bewegt werden können. Diesem Anspruch können Audiodaten im Vollformat nicht gerecht werden.[191]

Daher werden auch in Zukunft neue Verfahren entwickelt werden, die durch Datenreduktion den Markt über Datennetze solch behäbigen digitalen Formaten wie z.B. dem PCM-Format zugänglich machen. Die Qualitätssicherung wird dabei auch weiterhin an erster Stelle stehen und psychoakustische Erkenntnisse bei steigenden technischen Möglichkeiten immer effektiver nutzen.

---

http://www.ifpi.de/miz/wb1997.htm.

[191] Nach der von Sony vorgeschlagenen Abtastrate von 96 kHz bei einer Auflösung von 24 Bit für die Audio-*DVD (Digital Versatile Disc)* könnte sich die zukünftige Audiodatenrate des Vollformats von 1,411 MBit auf 4,608 MBit erhöhen.
Nach: http://www.sony.de/prod_special/dvd/wh_flexible_data_allocation.html#content.

7. Literatur

Behne, Klaus-Ernst/Barkowsky, Johannes: *Analoge und digitale Musikwiedergabe im unmittelbaren Vergleich – Eine Studie zur hypothesengeleiteten Wahrnehmung*, Hannover 1994.

Borucki, Hans: *Einführung in die Akustik*, Zürich 1980.

Brandenburg, Karlheinz: *Ein Beitrag zu den Verfahren und der Qualitätsbeurteilung für hochwertige Musikkodierung*, Dissertation an der Universität Erlangen-Nürnberg 1989.

Brandenburg, Karlheinz: *Perceptual Coding Of High Quality Digital Audio*, in: Applications Of Digital Signal Processing To Audio And Acoustics (Brandenburg, Karlheinz/Kahrs, Mark Hg.), Boston/Dordrecht/London 1998, S. 39–83.

Bremm, Peter: Artikel *Ein Klang von Licht und Magnet · Durch nichts zu erschüttern- MiniDisc-Laufwerk*, in: KEYS, 1996 11/96, S. 104–107.

Carstens, Matthias: *Musik-Speicher: Layer 3 auf Reisen – der etwas andere Walkman*, in: c't 21/98, S. 244.

Deutsche Telekom AG - Breitbandige Angebote T-DSL – Deutsche Telekom realisiert breitbandige Anwendungen auf ADSL-Technologie in Feldversuch, nach: http://www.telekom.de/angebot/bba/right.htm (Stand: 14.12.98).

Deutsche Telekom: *Music On Demand – die neue Musik-Power im Internet*, Bonn 7/98.

Deutsche Telekom: *Music On Demand – Informationen für Kunden*, 2/98.

Dickreiter, Michael: *Handbuch der Tonstudiotechnik – Band 2*, München [6]1997 [[1]1977].

Feiten, Bernhard: *Beurteilung von Quellenkodierungsverfahren für Audiosignale bei Berücksichtigung der Verdeckungseigenschaften des Gehörs*, Dissertation an der Technischen Universität Berlin 1989.

Franz, Dieter: *Elektroakustik – Grundlagen der Schallverarbeitung praxisnah dargestellt*, München 1990.

Fraunhofer IIS: *Audio & Multimedia (AMM) Technology*, nach: http://www.iis.fhg.de/amm/techinf/ (Stand: 01.06.2000).

Fraunhofer IIS: *Basics about MPEG Perceptual Audio Coding*, nach: http://www.iis.fhg.de/amm/techinf/basics.html (Stand: 01.06.2000).

Fraunhofer IIS: *Frequently Asked Questions about MPEG Audio Layer-3*, nach: http://www.iis.fhg.de/amm/techinf/layer3/layer3faq/index.html (Stand: 01.06.2000).

Fraunhofer IIS: *MPEG Audio Layer-3 – History*, nach: http://www.iis.fhg.de/amm/techinf/layer3/index.html (Stand: 01.06.2000).

Gesellschaft für Konsumforschung: *GfK Online-Monitor*, Nürnberg 9/1998 nach: http://www.gfk.cube.net/website/presse/980911_1.htm (Stand: 14.12.98).

Griefahn, Barbara/Jansen, Gerd: *Schallwirkung beim Menschen und Fragen des Gehörschutzes*, in: Taschenbuch der Technischen Akustik (Heckl, Manfred/Müller, Helmut A. Hg.), 2. Aufl., Berlin/Heidelberg/New York [2]1994 [[1]1975], S. 69–85.

Gröbchen, Walter: *MP3 – eine Formel zur Datenkompression schockt die Musikbranche*, in: DIE ZEIT, 1998, Nr. 44, nach:

http://www4.zeit.de/bda/int/zeit/print/199844.mp3_.html (Stand: 14.12.98).

Hall, Donald E./Goebel, Johannes (Hg.), *Musikalische Akustik – Ein Handbuch*, Mainz 1997.

Hansen, Thomas/Zombik, Peter: *Wirtschaftsbericht 1997 – Der Markt für Compact Discs, MusiCassetten und Schallplatten*, nach: http://www.ifpi.de/miz/wb1997.htm (Stand: 14.12.98).

Helmholtz, Hermann von: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Braunschweig ^{6}1913 [^{1}1862].

Herre, Jürgen: *Fehlerverschleierung bei spektral kodierten Audiosignalen*, Dissertation an der Universität Erlangen-Nürnberg 1995.

Hoeg, Wolfgang/Steffen, Edith: *Methoden zur subjektiven Qualitätsbewertung von Audiosignalen*, in: Highlights aus der Forschung – Quellenkodierung und Qualitätsbeurteilung als Impulsgeber für Innovation bei Sprache, Audio und Video (Deutsche Telekom AG Hg.), Darmstadt 1996.

ISO/IEC International Standard IS 11172-3: *Information Technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 MBits/s – Part 3: Audio.*

Kämmer, Heinz W.: *Die Compact Disc*, 2., aktualisierte Aufl., München ^{2}1988 [^{1}1983].

Kliment, Tibor: *Digital Radio im Urteil der Hörer – Ergebnisse einer Akzeptanzuntersuchung von DAB in Nordrhein-Westfalen*, Media Perspektiven 11/1998, S. 545–558.

Kremer, Martina: *Akustik-Show – Ein interaktives Tutorial zu den Themen Schall, Gehör und Psychoakustik*,

nach: http://www.dasp.uni-wuppertal.de/akustikshow/ (Stand: 14.12.98)
aktualisiert: http://www.dasp.uni-wuppertal.de/audite/ (Stand: 01.06.2000).

Krieg, Bernhard: *Digitale Audiotechnik ohne Ballast – Moderne Verfahren, Technik, Geräte, praktische Hinweise und Tips*, München 1992.

Meissner, Hansgeorg: *Digitale Multimediasysteme*, Berlin 1994.

Meyer, Martin: *Signalverarbeitung – Analoge und digitale Signale, Systeme und Filter*, Braunschweig/Wiesbaden 1998.

Moore, Brian C. J.: *An Introduction to the Psychology of Hearing*, London/San Diego/New York ^{3}1992 [^{1}1977].

Pan, Davis: *A Tutorial on MPEG/Audio Compression.* Reprinted IEEE Multimedia Journal, Ausgabe Sommer 1995, nach: http://www.eskimo.com/~miyaguch/mp3info.html (Stand: 01.06.2000).

Pan, Davis: *Digital Audio Compression*, in: Digital Technical Journal, Bd. 5/1993, Nr. 2, nach: http://www.eskimo.com/~miyaguch/mp3info.html (Stand: 01.06.2000).

Parncutt, Richard: Art. *Leistungen des Gehörs*, in: Musikpsychologie: ein Handbuch (Bruhn, Herbert/Oerter, Rolf/Rösig, Helmut Hg.), Reinbek bei Hamburg 1993.

Pickles, James O.: *An Introduction to the Physiology of Hearing*, London 1998.

Richtlinie ITU-R BS.1116: *Methods for the subjective assessment of small impairments in audio systems, including multichannel sound systems,* Genua 1994.

Rumsey, Francis: *The Audio Workstation Handbook*, Oxford/Boston 1996.

Schaefer, Martin/ Braun, Thorsten: *„Verantwortlichkeit im Internet“– ein neues Schlagwort betritt die Szene*, nach: http://www.ifpi.de/recht/re-2.htm (Stand: 01.06.2000).

Schaefer, Martin: *Vertragliche Gestaltungsmöglichkeiten bei Online-Musiknutzungen, insbesondere bei Musik-Abrufdiensten*, nach: http://www.ifpi.de/recht/re-vertr.htm (Stand: 01.06.2000).

Schulemann, Gaby: *Krieg den MP3-Musikpiraten*, in: c’t 22/98, S. 18

Shannon, Claude E./Weaver, Warren: *The Mathematical Theory Of Communication*, Urbana/Chicago 1963.

Shibazaki, Isao: *The Signal Compression Technology of Sony’s Newest MD Recorder*, MJ Audio Technology Magazine Technical Report, Dezember 1995, nach: http://www.amulation.co.uk/minidisc/mj_ja3es.html (Stand: 01.06.2000).

Stahlknecht, Peter/Hasenkamp, Ulrich: *Einführung in die Wirtschaftsinformatik*, Berlin/Heidelberg ^{8}1997 [^{1}1983].

Steinmetz, Ralf: *Multimedia-Technologie – Einführung und Grundlagen*, Berlin/Heidelberg 1993.

Stoll, Gerhard: *Internet-Radio und excellente Tonqualität: Wunschtraum oder Wirklichkeit?*, nach: http://www.irt.de/IRT/FuE/as/internetradio.htm (Stand: 01.06.2000).

Thurow, Norbert: *Sechs Thesen zum Thema: Digitales Zeitalter und Musikleben*, nach: http://www.ifpi.de/recht/re-digi.htm (Stand: 01.06.2000).

Tsutsui, Kyoya/Suzuki, Hiroshi/Shimoyoshi, Oscamu u.a.: *ATRAC: Adaptive Transform Acoustic Coding For MiniDisc*, 93rd AES Convention, San Fransisco 1992, Preprint 3456, nach: http://www.amulation.co.uk/minidisc/aes_atrac.html (Stand: 01.06.2000)

Van den Enden, Ad W.M./Verhoeckx, Niek A. M.: *Digitale Signalverarbeitung*, Braunschweig/Wiesbaden 1990.

Webers, Johannes: *Tonstudiotechnik – Schallaufnahme und -wiedergabe bei Rundfunk, Fernsehen, Film und Schallplatte*, München [5]1989 [[1]1968].

Wiedenhoff, Chris: *Abgeschmettert*, in: c't 23/98, S. 20.

Winkler, Gernot: *Tonaufzeichnung Digital – Moderne Audiotechnik mit Computerhilfe*, Aachen 1990.

Wolf, Alexander: *Urheberrechtliche Lizenzeinräumung und Kontrollrechte einer Verwertungsgesellschaft am Beispiel des Music-on-Demand-Projekts der Deutschen Telekom*, in: GEMA-Nachrichten, Ausgabe 157, 6/1998, S.83.

Yoshida, Tadao: *The Rewriteable MiniDisc System*, Reprinted from the Proceedings of the IEEE USA, Oktober 1994, Bd. 82, Nr. 10, S. 1492–1500, nach: http://www.amulation.co.uk/minidisc/ieee_paper.html (Stand: 01.06.2000).

Zenner, Hans-Peter: *Hören: Physiologie, Biochemie, Zell- und Neurobiologie*, Stuttgart 1994.

Zölzer, Udo: *Digitale Audiosignalverarbeitung*, Stuttgart 1996.

Zombik, Peter: *Music-on-Demand: „Tanz auf dem Vulkan" oder „Gib dem Affen Zucker"? – Chancen und Gefahren der neuen Technologie*, nach: http://www.ifpi.de/recht/re-6.htm (Stand: 01.06.2000).

Zwicker, Eberhard/Fastl, Hugo: *Psychoacoustics – Facts and Models*, Berlin/ Heidelberg/New York 1990.

Zwicker, Eberhard/Feldtkeller, Richard: *Das Ohr als Nachrichtenempfänger*, 2. neubearbeitete Aufl., Stuttgart 1967.

Zwicker, Eberhard: *Psychoakustik*, Berlin/Heidelberg/New York 1982.

http://freeflight.cockpit.be/mp3tech/tech.html : *MP3‘ Tech – Overwiew of the MP3 techniques* (Stand: 14.12.98)

Aktualisiert: http://www.mp3-tech.org/tech.html (Stand: 01.06.2000).

http://www.aes.org : AUDIO ENGINEERING SOCIETY (Stand: 14.12.98).

http://www.altavista.com : Internet-Suchmaschine Altavista (Stand: 14.12.98).

http://www.audio-on-demand.de : Music-on-Demand der Deutschen Telekom (Stand: 14.12.98).

http://www.audiomap.de/infos/minidisc.htm : *Audiomap – Umfangreiche Beschreibung der MiniDisc* (Stand: 01.06.2000).

http://www.dolby.com/tech/ : Dolby Laboratories – *Technical Information Dolby Digital, Dolby AC-2* (Stand: 01.06.2000).

http://www.internetstreik.de : *‚Initiative Internetstreik‘* (Stand: 01.06.2000).

http://www.jpc.de : Versandhandel JPC (Stand: 01.06.2000).

http://www.magnavox.com/product/pa332dcc.html : Philips DCC (Stand: 04.10.98).

http://www.minidisc.de/sonymd/atrac.html : Sony MiniDisc – *Verbesserte Dynamik durch ATRAC-Technologie* (Stand: 01.06.2000)

http://www.minidisc.de/sonymd/fach.html : Sony MiniDisc - *Von A-Z (Die Fachbegriffe)* (Stand: 01.06.2000).

http://www.mp3.de : MPEG Audio-Layer 3 (Stand: 01.06.2000).

http://www.sony.de/prod_special/dvd/wh_flexible_data_allocation.html#content : Sony Deutschland – *Digital Versatile Disc – Was ist DVD?* (Stand: 01.06.2000).

http://www.vqf.com : Audioreduktionsverfahren VQF (Stand: 01.06.2000).

http://www-eu.sv.philips.com/catalogue/ : Online-Katalog der Firma Philips (Stand: 01.06.2000).

Anhang

Niels Rump | Business Relations Manager -- kd_inel@iis.fhg.de |
| Fraunhofer Institute for Integrated Circuits IIS |
| Audio & Multimedia -- http://www.iis.fhg.de/amm/ |
| Am Weichselgarten 3, D-91058 Erlangen, Germany |
| Tel. +49(0)9131 776-611; Fax. +49(0)9131 776-398 |

Hallo Michael,
Grundsätzlich bietet sich "natürlich" unsere Webseite http://www.iis.fhg.de/amm/ (insbesondere <technology>) an. Dort gibt es einige informationen zu den "basics". Die Technische Umsetzung steht in den ISO-Standards (IS 11172-3 und IS 13818-3), die man via DIN oder via ISO beziehen kann. Es gibt auch einige AES Preprints, die die Materie behandeln (siehe unser Layer-3 FAQ).

Die darunterliegende Psychoakustik ist sehr gut und verständlich im Buch "Psychoakustik" von Zwicker (1982) beschrieben. (Ich hoffe das Buch ist noch lieferbar, aber es gilt als eine Referenz)

Gruß und viel Spaß/Glück beim Examen...
Niels

07.09.1998 per Email

Deutsche Telekom AG
Music on Demand - Kundenbetreuung

Sehr geehrter Herr Cammann,

wir möchten Ihre Frage mal so beantworten, die Musiktitel werden in Echtzeit und Hi-Fi Stereo an unsere Kunden geliefert und während der Übertragung sofort abgespielt. Die Musikdateien werden dabei im MPEG1 Layer 3 Format im Verhältnis 1:12 komprimiert, dass heißt eine Minute Spielzeit = 1 Minute Übertragungszeit. Da die Daten von uns über eine direkte ISDN Verbindung mit Kanalbündelung übertragen werden (die vollen 128 Kb/sec incl. Steuerdaten) und nicht über das Internet, entstehen praktisch keine Schwankungen in der Übertragung. Minimale Abweichungen werden durch einen Buffervorlauf von 40 Kb abgefangen.

Für weitere Fragen stehen wir Ihnen gerne jederzeit zur Verfügung.
Weiterhin viel Spaß mit MoD.

Mit freundlichen Grüßen

Ihr MoD-Team

07.12.1998 per Email

Stichwortverzeichnis

A

A/D-Wandler 37, 39
Abstandsempfindung 14
Abtastfrequenz 38, 52
Abtastrate 44
Abtasttheorem 38, 44
Adaptive Allocation 52
Adaptive Transform Acoustic Coding
  ATRAC 53
Aliasing 41
Amboß 5
Asymmetric Digital Subscriber Line
  ADSL 83
Audiodatenrate
  der CD 45
  der Digital Compact Cassette 52
Auflösung 39
Außenohr 5

B

Bandbreite 22
  kritische 26
Bark-Skala 20
Basilarmembran 8, 12, 18, 34
Bit 36
Bitzuweisung 48
Bogengänge 7

C

CAV-Modus 45
CD-Qualität 75, 78, 86
CD-ROM 43
CLV-Modus 45
Cochlea 7, 10
Compact-Disc (CD) 43
Cortisches Organ 9

D

D/A-Wandler 37, 42
Datenkompression 47
Datenreduktion 14, 43, 47
Datenreduktionsverfahren
  ATRAC (MiniDisc) 52
  MPEG Audio - Layer 3 58
  PASC (Digital Compact Cassette) 50
Dekoder 79
Differenzton
  kubischer 28
Digital Audio Broadcasting (DAB) 59
Digital Audio Tape (DAT) 46
Digital Compact Cassette (DCC) 50
Digitale Signalverarbeitung 36
Digitale Tonaufzeichnung
  ohne Datenreduktion 43
  unter Datenaufzeichnung 47
Digitalisierung 43
Dualsystem 36

E

Enkoder 49, 59, 78
Entladungsrate 12
Eustachische Röhre 5

F

Festpunktdarstellung 52
Frequenz.....12
Frequenzauflösung 10
Frequenzbandbreite, kritische 18
Frequenzdispersion 10
Frequenzgruppe.....18, 22, 60
Frequenzgruppen.....21
Frequenzselektivität 9, 10, 12
Frequenzspektrum.....25

G

Gehör.....14, 23
Gehörgang.....5
Gehörknöchelchen.....5, 6
Gleichgewichtssinn 7
Gleitpunktdarstellung.....51

H

Haarzelle 9, 12
Haarzellen 12
Hammer.....5
Helicotrema.....8
Hirnstamm.....12
Höreindruck 14
Hörfähigkeit 14
Hörfläche.....16
Hörmechanismen, binaureale.....23
Hörnerv 12, 13
Hörschwelle 12, 16, 24, 26, 31
Hörspektrum.....25
Hybrid-Filterbank.....61
hypothesengeleitete Wahrnehmung 86

I

Innenohr.....5
Intensitätsunterschiede.....23

J

JND.....14
JND, frequenzbezogenen.....15
JND, lautstärkebezogen 17
Joint Stereo Coding.....63

K

Kaskadierung.....68
Klangrauhigkeit 19
Knalltrauma 6
kritische Frequenzbandbreite 22

L

Lands 44
Laufzeitunterschiede.....23
Lautheit.....21
Lautstärke 16
Lymphflüssigkeit 8

M

Maskierer.....26
Maskierton.....29
Maskierung.....24, 34, 48
„Addition“ mehrerer Maskierer 33
durch instationäre Schalllereignisse....31
Maskierung eines Tons
durch einen anderen, stationären Ton .28
durch komplexe, stationäre Töne 30
durch stationäres Schmalbandrauschen
.....26
durch stationäres Weißes Rauschen....25

MiniDisc 68
MiniDisc (MD) 52
Mithörschwelle 24, 28, 30, 48
Mithörschwellen 26
Mittelohr 5
Modifizierte diskrete Cosinus-Transformation
MDCT 53
MOVING PICTURES EXPERTS GROUP
(MPEG) 58
MP3 60
MPEG-Audio
5.3.3. Psychoakustisches Modell 62
Music on Demand 81, 84
MUSIC ON DEMAND 76

N

Nachverdeckung 31
Nachverdeckungseffekt 32
Nervenfaser 12
Nervenkanal, auditorischer 12
Nervensystem, zentrales auditorisches 13

O

Ohr 4
Ohrmuschel 5
Ortsprinzip 10, 12, 13
Ovale Fenster 8
Ovales Fenster 6

P

Paukengang 8
Paukenhöhle 5
Perceptual Coding 48
Perilymphe 6
Phasenunterschiede 23
Pits 44
Polyphasenfilterbank 60
Precision Adaptive Subband Coding
(PASC) 50
pre-echo 56
Psychoakustik 14
Psychoakustisches Modell 48
Pulse-Code-Modulation (PCM) 37

Q

Quantisierungsfehler 40
Quantisierungsrauschabstand 40
Quantisierungsrauschen ..40, 44, 48, 55, 62, 63, 69
Quantisierungsrauschens 60

R

Raubkopie 77, 79, 84
Räumliches Hören 23
Rauschen
gleichverteiltes ..... Siehe Quantisierungsrauschen
Schmalbandrauschen 26
weißes 25
Red Book-Standard 46
Redundanzen 63
Ruhehörschwelle 16, 26, 28, 30
Rundes Fenster 6

S

Samplerate 38
Schalldruckpegel 12
Schallempfindlichkeit 16
Schallfrequenz 13

Schallintensität 16
Schallintensitätsdichte 26
Schmerzgrenze 16
Schnecke 7, 11
Schwebung 19
Signal
analoges 37, 39
digitales 37
wertkontinuierliches 37
zeitkontinuierliches 37
Signal-Rauschabstand 40
signal-to-mask ratio
(SNR) 63
signal-to-noise ratio
(SNR) 69
signal-to-noise ratio (SNR) 41
Simultanverdeckung 31, 32
Sone-Skala 21
Source Coding 48
Speichermedium
magneto-optisches 52
Speichermediumn
magnetisches 46
Speichermediun
optisches 45
Steigbügel 5, 6
Steigbügelfußplatte 6
Stereokodierung
Intensitäts-Stereokodierung 65
M/S-Stereokodierung 64
Störabstand 40
Störpegel 26
Störschall 24
Subbänder 51, 60

T

Teilbandkodierung 51
Tonalitätsindex 62
Tonhöhe 14
Triple stimulus-hidden reference-double
blind 70
Trommelfell 5, 6

U

Urheberrecht 79

V

Verdeckung 24, 31
Vorhofsgang 8
Vorverdeckung 31, 56

W

Wanderwelle 9
Wortbreite 40

Z

Zeitliches Auflösungsvermögen des Gehörs
..... 21