

**Ein psychophysiologisches Gehörmodell
zur Nachbildung von Wahrnehmungsschwellen
für die Audiocodierung**

Vom Fachbereich Elektrotechnik und Informationstechnik
der Universität Hannover

zur Erlangung des akademischen Grades

Doktor-Ingenieur

genehmigte

Dissertation

von

Dipl.-Ing. Frank Baumgarte

geboren am 13. April 1964 in Bad Gandersheim

2000

Referent: Prof. Dr.-Ing. H. G. Musmann,
Korreferent: Prof. Dr. rer. nat. Dr. med. B. Kollmeier
Tag der Promotion: 29. Februar 2000

Vorwort

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung der Universität Hannover.

Dem Direktor des Instituts, Herrn Professor Dr.-Ing. H. G. Musmann, danke ich für die Anregung zum Thema der Arbeit und für die hervorragenden Arbeitsmöglichkeiten. Ihm gilt mein besonderer Dank für die Betreuung der Arbeit und für die Übernahme des Hauptreferats.

Herrn Professor Dr. rer. nat. Dr. med. B. Kollmeier danke ich für die Übernahme des Korreferats.

All meinen Kolleginnen und Kollegen, die durch Diskussionen, Anregungen und vielerlei praktische Unterstützung zum Gelingen der Arbeit beigetragen haben, danke ich herzlich. Mein ganz besonderer Dank gilt Herrn Dr.-Ing. Bernd Edler für seine vielen konstruktiven Anregungen und Diskussionen sowie seine stets unermüdliche Hilfsbereitschaft.

Hannover, im Februar 2000

Frank Baumgarte

Meinen Eltern gewidmet

Kurzfassung

In dieser Dissertation wird ein Modell des menschlichen Gehörs zur Nachbildung auditiver Wahrnehmungsschwellen vorgestellt. Die Wahrnehmungsschwelle quantifiziert die Größe einer gerade wahrnehmbaren Änderung eines Schallsignals im Vergleich zum Referenzsignal. Diese Wahrnehmungsschwellen werden in der datenreduzierenden Codierung von Audiosignalen benötigt. Sie bestimmen den maximalen Pegel der von der Quantisierung verursachten Verzerrungen, so daß diese Verzerrungen gerade noch nicht wahrnehmbar sind. Dabei wird vor allem die Verdeckung der Verzerrungen durch das Audiosignal ausgenutzt.

Das Gehörmodell besteht aus einem physiologisch motivierten Teil, der die Schallverarbeitung im peripheren Gehörs nachbildet, und einem psychoakustisch motivierten Teil zur Nachbildung der sich anschließenden neuronalen Verarbeitung im Gehör. Es wird angenommen, daß die mechanischen Schwingungen innerhalb der Cochlea des Innenohres bei geringen Schalldruckpegeln durch die äußeren Haarzellen massiv verstärkt werden. Diese Verstärkung verschwindet bei hohen Pegeln infolge einer Sättigung. Das Teilmodell der Cochlea berücksichtigt die nichtlineare Verstärkungswirkung der äußeren Haarzellen, so daß die Effekte der Nichtlinearität auf die Wahrnehmungsschwellen durch das Gehörmodell adäquat nachgebildet werden.

Im physiologischen Teil werden die wesentlichen Ursachen der spektralen Verdeckungswirkung nachgebildet. Der psychoakustisch motivierte Teil des Gehörmodells berücksichtigt vor allem die zeitabhängigen Verdeckungseffekte. Die Wahrnehmbarkeit von Signaländerungen wird am Ausgang dieses Teils mit Hilfe von Schwellendetektoren bestimmt. Das Gehörmodell wird weitgehend als elektrisches Netzwerk dargestellt. Die Simulation des Netzwerks im Zeitbereich beruht größtenteils auf Wellendigitalfiltern.

Für den Parameterabgleich des Gehörmodells werden sowohl physiologische, als auch psychoakustische Meßergebnisse herangezogen. Die Verifikation des Gehörmodells erfolgt anhand weiterer psychoakustischer Meßdaten, die über die klassischen Ergebnisse für die zeit- und frequenzabhängige Verdeckung hinausgehen. Damit wird gezeigt, daß die wesentlichen Abhängigkeiten der Wahrnehmungsschwellen von den Parametern des Schallsignals nachgebildet werden.

In einem Anwendungsbeispiel ersetzt das Gehörmodell das bislang verwendete psychoakustische Modell zur Steuerung eines Audiocoders, der dem Stand der Technik entspricht. Die damit erzielte subjektive Klangqualität wird mit dem ursprünglichen Referenzcoder verglichen. Die unterschiedlichen Ergebnisse werden anhand eines Vergleichs der Wahrnehmungsschwellen beider Modelle analysiert. Diese Anwendung zeigt, daß das Gehörmodell auch bei beliebig komplexen Audiosignalen, wie beispielsweise Sprachsignalen, adäquate Wahrnehmungsschwellen erzeugt.

Schlagworte: Wahrnehmungsmodell, Modell der Verdeckung, Irrelevanzreduktion

Abstract

A model of the auditory system to simulate threshold of detectability is presented in this thesis. This threshold is the level at which a change of an acoustic signal is detectable by a human listener. The threshold can be used to achieve high data compression in audio coding. It specifies the maximum level of introduced quantization distortion that is inaudible. In audio coding masking of distortion by the desired acoustic signal is exploited.

The auditory model consists of a physiologically motivated part, which simulates sound processing in the peripheral auditory system, and a psychoacoustically motivated part, which simulates subsequent neural processing in the central auditory system. It is assumed that in case of a low sound pressure level the mechanical vibrations inside the cochlea of the inner ear are greatly amplified by the outer hair cells. This amplification saturates and has no effect at high levels. The cochlear model as part of the auditory model takes into account this nonlinear amplification effect of the outer hair cells so that the influence of this nonlinearity on perception thresholds is simulated adequately.

The physiological part of the auditory model takes into account the fundamental mechanisms of spectral masking effects. The psychoacoustically motivated part simulates time-dependent masking effects. The output of the model is a number related to the audibility of sound changes. This number is derived by threshold detectors in the psychoacoustically motivated part. The auditory model is chiefly represented as an electrical network. The network modeling the peripheral auditory system is simulated in the time domain by wave digital filters.

The parameters of the auditory model are adjusted according to data from both physiological and psychoacoustical measurements. The auditory model is verified using additional data from psychoacoustical measurements that go beyond the classical results for time- and frequency-domain masking. The verification confirms that the fundamental dependencies of perception thresholds on the acoustic signal are reproduced.

In an example, the auditory model is used to replace the currently used psychoacoustical model in a state-of-the-art perceptual audio coder. The resulting subjective audio quality is compared to that of the coder in its original form. The results are analyzed by a comparison of the perception thresholds from both models. This example confirms that the auditory model adequately reproduces the perception thresholds of arbitrary complex audio signals, e. g. speech signals.

Keywords: Perception Model, Model of Masking, Irrelevance Reduction

Inhalt

| | |
|---|-----------|
| 1 Einleitung | 1 |
| 2 Grundlagen | 7 |
| 2.1 Anatomie und Physiologie des Gehörs | 7 |
| 2.1.1 Peripheres Gehör | 8 |
| 2.1.2 Zentrales Gehör | 13 |
| 2.2 Physikalisch meßbare Reizantworten des Gehörs | 14 |
| 2.2.1 Übertragungseigenschaften des Außen- und Mittelohres | 14 |
| 2.2.2 Schallinduzierte Schwingung der cochleären Trennwand | 15 |
| 2.2.3 Aktionspotentiale auf einzelnen Hörnervenfasern | 18 |
| 2.2.4 Otoakustische Emissionen | 19 |
| 2.3 Lautheit | 20 |
| 2.4 Psychoakustisch gemessene Wahrnehmungsschwellen | 21 |
| 2.4.1 Ruhehörschwellen | 22 |
| 2.4.2 Mithörschwellen | 23 |
| 2.4.3 Unterschiedsschwellen | 26 |
| 2.5 Physiologische Ursachen der Verdeckung | 27 |
| 2.6 Konzepte zur Vorhersage auditiver Wahrnehmungsschwellen | 29 |
| 3 Das psychophysiologische Gehörmodell | 37 |
| 3.1 Struktur des Gehörmodells | 37 |
| 3.2 Modell des Außen- und Mittelohres | 39 |
| 3.3 Physiologisches Modell der Cochlea | 40 |
| 3.3.1 Modell der Hydromechanik und der äußeren Haarzellen | 40 |
| 3.3.2 Modell der inneren Haarzellen | 50 |
| 3.4 Psychoakustisches Modell der neuronalen Verarbeitung | 51 |
| 3.4.1 Nachbildung der Vorverdeckung | 52 |
| 3.4.2 Nachbildung der Nachverdeckung | 52 |
| 3.4.3 Nachbildung der Abhängigkeit der Verdeckung von der Testsignaldauer | 55 |
| 3.4.4 Schwellendetektor | 57 |
| 3.4.5 Nachbildung der Abhängigkeit der Verdeckung von der Einhüllendenfluktuation | 58 |
| 3.5 Generierung von Wahrnehmungsschwellen | 59 |

| | |
|---|------------|
| 4 Verifikation des Gehörmodells | 61 |
| 4.1 Vergleich von Simulationsergebnissen und physiologischen Meßergebnissen | 61 |
| 4.1.1 Schallinduzierte Schwingung der cochleären Trennwand | 61 |
| 4.1.1.1 Wanderwelle | 61 |
| 4.1.1.2 Dynamikkompression | 64 |
| 4.1.1.3 Suppression | 64 |
| 4.2 Vergleich von Simulationsergebnissen und psychoakustischen Meßergebnissen | 66 |
| 4.2.1 Lautheit | 66 |
| 4.2.2 Wahrnehmungsschwellen | 67 |
| 4.2.2.1 Ruhehörschwellen | 67 |
| 4.2.2.2 Mithörschwellen für spektrale Verdeckung | 68 |
| 4.2.2.3 Mithörschwellen für zeitabhängige Verdeckung | 77 |
| 4.2.2.4 Mithörschwellen für kombinierte zeit- und frequenzabhängige Verdeckung | 80 |
| 4.2.2.5 Wahrnehmungsschwellen für Amplituden- und Frequenzmodulation | 81 |
| 4.3 Diskussion der Ergebnisse | 82 |
| 5 Einbindung des Gehörmodells in ein Audiocodierverfahren | 95 |
| 5.1 Das Codierverfahren nach ISO/IEC MPEG-2 AAC | 95 |
| 5.2 Das psychoakustische Basismodell zur Quantisierersteuerung des AAC | 97 |
| 5.3 Quantisierersteuerung des AAC mit dem Gehörmodell | 99 |
| 5.4 Bewertung der Klangqualität anhand ausgewählter Audiosignale | 104 |
| 5.5 Vergleich generierter Signal-Maskierungs-Abstände vom optimierten psychoakustischen Modell und vom psychophysiologischen Gehörmodell | 107 |
| 6 Zusammenfassung | 117 |
| 7 Literaturverzeichnis | 123 |
| Anhang A Algorithmen und Parameter des psychophysiologischen Gehörmodells | 137 |

Verzeichnis der Abkürzungen, Größen und Formelzeichen

| Abkürzung | Erläuterung |
|------------------|---|
| AAC | A dvanced A udio C oding (Teil des ISO/IEC MPEG-2 Standards) |
| ÄHZ | äußere H aarzelle |
| AM | A mpli t uden m odulation |
| AMO | A ußen- und M ittelohr |
| AP | A ktions p otential auf einer Nervenfasern |
| ATRAC | A daptive T ransform A coustic C oding for MiniDisc |
| BM | B asilarmembran |
| CD | C ompact D isc |
| CF | charakteristische F requenz |
| CMR | c o-modulation m asking r elease |
| FM | F requenz m odulation |
| HM | H ydromechanik der Cochlea |
| IHZ | innere H aarzelle |
| ISO/IEC | I nternational O rganization for S tandardization / I nternational E lectro- t echnical C ommission |
| ITU-R | I nternational T elecommunication U nion – R adiocommunication Section |
| ITU-T | I nternational T elecommunication U nion – T elecommunication Section |
| JND | j ust n oticeable d ifference — gerade wahrnehmbarer Unterschied |
| LSB | least significant b it — niederwertigstes Bit |
| MPEG | M oving P icture E xperts G roup (Untergruppe innerhalb der ISO/IEC) |
| NV | n euronale V erarbeitung |
| SL | sensation level — Schalldruckpegel, gemessen relativ zum SPL des Signals an der Ruhehörschwelle |
| SNR | s ignal-to- n oise r atio — Signal-Rausch-Abstand |

SPL sound pressure level — Schalldruckpegel, gemessen relativ zum Bezugspegel von 20 μPa .

WDF Wellendigitalfilter

Größe/Einheit Erläuterung

A Amplitude

B Hz Bandbreite

C F Kapazität

C_m m/N mechanische Nachgiebigkeit

D Ausgangsgröße des Beobachters (Signaldetektionstheorie: decision)

E W Energie

F N Kraft

G $1/\Omega$ Leitwert

I $\frac{W}{m^2}$ Schallintensität

L dB logarithmischer Pegel

L dB SPL Schalldruckpegel, bezogen auf den Referenzdruck von 20 μPa

L_S dB SL Schalldruckpegel, bezogen auf die Ruhehörschwelle des gleichen Signals

L H Induktivität

M dB Verdeckungspegel (masking): Differenz von Mithörschwelle und Ruhehörschwelle

N sone Lautheit

N_i N/A Einheitswandlerkonstante (FI -Analogie)

N' sone/Bark spezifische Lautheit

N_{Sek} Anzahl der Sektionen des Gehörmodells

Ψ von Beobachter ausgewertete Meßgröße (Signaldetektionstheorie)

Q Güte

R Ω elektrischer Widerstand

R_B Ω Bezugswiderstand (WDF)

R_m N s/m mechanischer Widerstand

| | | |
|-----------|----------|---|
| T | s | Dauer |
| T_A | s | Periodendauer eines Abtastintervalls |
| Z | Ω | (komplexe) Impedanz |
| a | | hinlaufende Welle (WDF) |
| b | | rücklaufende Welle (WDF) |
| β | | Likelihood-Verhältnis an der Entscheidungsschwelle Ψ_S (Signaldetektionstheorie) |
| d | | Grad der Wahrnehmbarkeit am Ausgang des Fluktuationsdetektors |
| d' | | Sensitivität (Signaldetektionstheorie) |
| η | % | Modulationsindex bei FM |
| f | Hz | Frequenz |
| f_A | Hz | Abtastfrequenz |
| f_B | Hz | Bezugsfrequenz für die Bilineartransformation (WDF) |
| fl | | Fluktuation der Signaleinhüllenden |
| ϕ | | normierte Frequenzvariable im Bereich zeitdiskreter WDF |
| g | | Verstärkung (gain) |
| g | | Torleitwert (WDF) |
| γ | | WDF Koeffizient |
| i | A | elektrischer Strom |
| $l(\Psi)$ | | Likelihood-Verhältnis (Signaldetektionstheorie) |
| λ | m | Wellenlänge |
| m | kg | Masse |
| m | % | Modulationsgrad bei AM |
| $m(t)$ | Pa | zeitabhängiger Schalldruck eines Maskierersignals (masker) |
| μ | | Differenz der Sektionsnummern gekoppelter Sektionen |
| $n(t)$ | Pa | zeitabhängiger Schalldruck eines Rauschsignals (noise) |
| n | W/Hz | Rauschleistungsdichte |
| n | | Index des Abtastintervalls zeitdiskreter Folgen |

| | | |
|-------------------|-----------------------------|---|
| ν | | Stufennummer des Modells der Cochlea |
| ω | 2π Hz | Kreisfrequenz ($\omega = 2\pi f$) |
| ω_N | 2π Hz | Normierungskreisfrequenz für die Bilineartransformation (WDF) |
| p | Pa | Druck, Schalldruck |
| $q(t)$ | | beobachtetes Signal (Signaldetektionstheorie) |
| r | | Torwiderstand (WDF) |
| r | | Verhältnis der kurzzeitintegrierten spezifischen Lautheiten (ratio) |
| $s(t)$ | Pa | zeitabhängiger Schalldruck eines Testsignals (signal) |
| σ | | Standardabweichung |
| t | s | Zeit |
| τ | s | Zeitkonstante |
| u | V | elektrische Spannung |
| v | $\frac{\text{m}}{\text{s}}$ | Geschwindigkeit, Schnelle |
| w | | Gewichtungsfaktor |
| x | m | Position auf der cochleären Trennwand gemessen vom ovalen Fenster |
| ξ | m | Auslenkung |
| z | Bark | Tonheit |
| Δz | Bark | Tonheitsabstand der Mittenfrequenz benachbarter Sektionen |
| z | | komplexe Variable der z -Transformation |
| z_{char} | Bark | charakteristische Tonheit |

Konstante Erläuterung

π 3,1415926...

Formelzeichen Erläuterung

| | |
|---------------|-------------------------------------|
| $\cot(\cdot)$ | Kotangensfunktion |
| $\exp(\cdot)$ | Exponentialfunktion mit der Basis e |
| $\ln(\cdot)$ | Logarithmusfunktion zur Basis e |
| $P(\cdot)$ | Wahrscheinlichkeit |

| | |
|--------------------|---|
| $P(\cdot \cdot)$ | bedingte Wahrscheinlichkeit |
| $p(\cdot)$ | Wahrscheinlichkeitsdichte |
| $Sp(n)$ | zeitdiskrete Spreizfunktion |
| Index | Erläuterung |
| A | Abtastung |
| A | Außenohr |
| Ab | Abklingfunktion (Modellierung der Nachverdeckung) |
| BM | Basilarmembran |
| C | Cochlea |
| D | Messung der D auer |
| E | Erweiterung (zweite Verstärkerstufe) |
| F | Fensterfunktion (zum Ein- und Ausblenden eines Signals) |
| H | Helicotrema |
| HF | hohe Frequenz |
| HN | Hörnerv |
| HT | high-threshold Theorie |
| IHZ | innere Haarzelle |
| Int | Kurzzeitintegrator |
| K | laterale K opplung von Modellsektionen |
| L | Pegelmessung (l evel) |
| LF | kleine Frequenz (l ow f requency) |
| M | M askierer |
| M | M ittelohr |
| N | Rauschen (n oise) |
| NA | neuronal e A ktivität |
| OF | ovales F enster am Eingang der Cochlea |
| R | R eferenzsignal |
| RC | RC -Glied |

| | |
|-----|---|
| Res | Resonanz |
| T | Testsignal |
| T | tonales Signal (im Gegensatz zu rauschartig) |
| S | Schwelle |
| Sät | Sättigung |
| Sek | Sektion |
| Sp | Spreizfunktion |
| Z | Zirkulator |
| d3 | Verzerrungsprodukt bei der kubischen Differenzfrequenz |
| eff | Effektivwert |
| l | lateral |
| mod | Modulation |

1 Einleitung

Für die Bewertung der Klangqualität technischer Kommunikationssysteme sind sehr aufwendige Hörtests erforderlich. Ein Modell der Wahrnehmung, das die Empfindungsgrößen ideal nachbildet, könnte diese Hörtests ersetzen. Darüber hinaus kann ein Wahrnehmungsmodell, welches die nicht wahrnehmbaren Anteile eines Audiosignals bestimmt, auch für die Datenreduktion genutzt werden. Diese Beispiele zeigen, daß Modelle der Wahrnehmung grundlegende Bedeutung und vielfältige Anwendungen in der Kommunikationstechnik haben.

Ein Audiocoder kann als digitales Teilsystem eines akustischen Kommunikationssystems betrachtet werden, das die zur Repräsentation eines Audiosignals erforderliche Datenmenge für eine wirtschaftliche Übertragung oder Speicherung reduziert. Eine ideale Codierung ohne Klangqualitätsverlust entfernt dazu alle aufgrund der Signalstatistik redundanten Signalanteile, die beim Decoder wieder hinzugefügt werden können, sowie irrelevante Signalanteile, die für den Empfänger, das menschliche Gehör, nicht wahrnehmbar sind. Die Wahrnehmbarkeit kann mit Hilfe von psychoakustisch gemessenen Wahrnehmungsschwellen quantifiziert werden, die eine gerade wahrnehmbare Signaländerung angeben. Stimmen die bei der Codierung verwendeten Wahrnehmungsschwellen nicht mit denen des Empfängers überein, können hörbare Unterschiede im decodierten Signal auftreten, oder es wird wegen einer unvollständigen Irrelevanzreduktion eine höhere Datenrate aufgewendet als erforderlich. Daher ist bei der Codierung eine möglichst genaue Approximation der Wahrnehmungsschwellen des Empfängers anzustreben [108].

Die wichtigsten Wahrnehmungsschwellen für diese Anwendung sind die Ruhehörschwellen und Mithörschwellen. Die Psychoakustik definiert diese Hörschwellen als Pegel eines Testsignals, bei dem dieses mit einer vorgegebenen Wahrscheinlichkeit von beispielsweise 50% wahrgenommen wird. Bei einer Mithörschwellenmessung wird zusätzlich zum Testsignal ein Maskiersignal dargeboten, welches das Testsignal teilweise oder ganz verdecken kann. Übertragen auf die Audiocodierung entspricht dem Maskiersignal das Audiosignal, welches die von der Codierung verursachten Verzerrungen verdecken kann. Dem Testsignal entsprechen somit die Codierverzerrungen, deren zeit- und frequenzabhängige Wahrnehmungsschwelle mit einem Gehörmodell nachgebildet werden soll.

Im allgemeinen stellt die Psychoakustik Meßergebnisse der bei Testpersonen erzielten Empfindung in Abhängigkeit von physikalischen Parametern des Schallreizes für einfache Signale zur Verfügung. Für komplexere Schallsignale, die in der Regel in Kommunikationssystemen verarbeitet werden, sind solche Messungen aufgrund der Vielfalt möglicher Signalvariationen nicht umfassend durchführbar. Einen Ausweg ermöglicht die vereinfachende Modellannahme, daß die Signalanalyse des Gehörs wie eine Spektralzerlegung ähnlich einer linearen Filterbank wirkt, an deren Ausgängen die Leistungen der Teilbandsignale ausgewertet werden. Mit dieser Annahme kann die zu erwartende Wahrnehmungsschwelle für komplexere Signale aus den psychoakustischen Meßergebnissen für einfache Signale mit Hilfe einer linearen Überlagerung der Teilbandleistungen bestimmt werden. Durch diese

Vereinfachung werden jedoch wesentliche nichtlineare Höreigenschaften vernachlässigt, die im folgenden erläutert werden. Es ergeben sich daher systematische Fehler durch die Extrapolation psychoakustischer Ergebnisse auf komplexere Signale.

Das Gehör besteht aus Außen-, Mittel- und Innenohr, sowie den angeschlossenen neuronalen Verarbeitungsstufen. Der für die Schallverarbeitung wesentliche Teil des Innenohres ist die schneckenförmige Cochlea. Das Schallsignal breitet sich als sogenannte Wanderwelle in dem flüssigkeitsgefüllten Schneckenkanal aus, der durch die membranartige cochleäre Trennwand longitudinal in zwei akustisch wirksame Kammern unterteilt wird. Dieses System besitzt eine ortsabhängige Resonanzfrequenz, so daß der Ort der maximalen Schwingungssamplitude der Trennwand von der Signalfrequenz abhängt und somit eine Frequenz-Orts-Transformation bewirkt wird. Physiologische Messungen an der Cochlea, sowie psychoakustische Messungen zeigen, daß die Schallverarbeitung des Gehörs im Gegensatz zu der obigen Annahme einen ausgeprägt nichtlinearen Charakter aufweist. Das ergibt sich unter anderem aus der beobachteten Dynamikkompression bei der Schallverarbeitung [144, 164], der Entstehung von Verzerrungsprodukten des Eingangssignals in der Cochlea [197] und anderen nichtlinearen Wahrnehmungseigenschaften [34, 69, 156, 165]. Die Wahrnehmungsleistungen des Gehörs werden bei einer Schädigung der nichtlinearen Verarbeitung beeinträchtigt; so ist beispielsweise sein großer Dynamikumfang nur bei Vorhandensein dieser Nichtlinearität gegeben [197].

Ein Modell der auditiven Wahrnehmung bildet die Prozesse der physiologischen Schallverarbeitung bis zur psychoakustischen Empfindung entweder implizit oder explizit nach. Für den Entwurf eines solchen Modells sind aus der Literatur Lösungen bekannt, die den Wahrnehmungsprozeß ganz oder nur teilweise nachbilden. Diese Modelle lassen sich grob in die folgenden vier Klassen einteilen:

- *Erregungsmuster-Modelle* für die Nachbildung von Wahrnehmungsschwellen komplexer Schallsignale,
- *Lautheits-Modelle* für die Nachbildung der subjektiven Bewertung der Klangqualität,
- *Black-Box-Modelle* für die Nachbildung von Wahrnehmungsschwellen spezieller, einfacher Schallsignale,
- *Cochlea-Modelle* für die physiologische Nachbildung der Schallverarbeitung in der Cochlea.

Die wesentlichen Eigenschaften dieser Modellklassen werden im folgenden kurz umrissen.

Die gemeinsame Grundlage der *Erregungsmuster-Modelle* ist das sogenannte Schwellenfunktionsschema von Zwicker und Feldtkeller [198]. In derartigen Modellen wird ein Erregungsmuster der inneren Haarzellen in der Cochlea aus dem Eingangssignal bestimmt. Die inneren Haarzellen wandeln mechanische Schwingungen der cochleären Trennwand in elektrische Impulse auf dem Hörnerv um. Die Anzahl von Impulsen pro Zeit wird als sogenannte mittlere Entladungsrate angegeben. Die Schallverarbeitung in der Cochlea wird in diesen Modellen jedoch nicht physiologisch nachgebildet, sondern die Ermittlung der Erregungsmuster erfolgt indirekt aus gemessenen Mithörschwellenverläufen schmalbandiger Maskierersignale. Die Wahrnehmungsschwelle für andere Maskierersignale ergibt sich im Schwellenfunktionsschema mit der Annahme, daß die Veränderung des Eingangssignals

wahrnehmbar ist, sobald sich die Erregung um mehr als 1 dB an einem Ort in der Cochlea erhöht hat. Diese Erhöhung kann zum Beispiel durch die Überlagerung eines Testsignals verursacht werden.

In derzeit bekannten Audiocodierverfahren, die für eine der Compact Disk vergleichbare Klangqualität ausgelegt sind und eine nennenswerte Irrelevanzreduktion erreichen [1, 28, 101, 182, 183], werden Erregungsmuster-Modelle als sogenannte psychoakustische Modelle zur Mithörschwellenbestimmung eingesetzt. Diese Modelle vereinfachen das Schwellenfunktionsschema, indem die Abhängigkeit des Erregungsmusters von der Maskierfrequenz und dem Maskiererpegel vernachlässigt wird. Mit dieser Vereinfachung können die Mithörschwellen aus den Signalleistungen am Ausgang einer FFT-Spektralzerlegung mit einer Faltung bestimmt werden.

Als *Lautheits-Modelle* werden hier die Modelle zusammengefaßt, die eine Bewertung der Klangqualität, basierend auf der nachgebildeten Lautheit oder einer vergleichbaren gehörinternen Schallsignalrepräsentation, durchführen. Die Lautheit ist ein psychoakustisches Maß für die wahrgenommene Schallintensität. Diese Lautheits-Modelle sind wie die Erregungsmuster-Modelle psychoakustisch motiviert, aber demgegenüber erweitert. Sie erlauben außer einer Nachbildung der Wahrnehmungsschwellen eine Bewertung der Klangqualitätsverminderung, die von überschwellig wahrnehmbaren Signaländerungen hervorgerufen werden. Die aufwendigen Hörtests zur Bewertung digitaler Audiocodierungen können durch diese Modelle ergänzt oder ersetzt werden. Ein grundlegendes Verfahren zur Bewertung der Klangqualität ist mit dem psychoakustischen Modell von Schroeder et al. [170] gegeben, welches das Schwellenfunktionsschema erweitert, indem es nicht nur eine Überschreitung der Mithörschwelle detektiert, sondern auch überschwellige Verzerrungen anhand der spezifischen Lautheit bewertet. Die spezifische Lautheit nähert die ortsabhängige Verteilung der von den inneren Haarzellen aufgenommenen Teillautheiten in der Cochlea an. Sie kann aus dem Erregungsmuster bestimmt werden. Die Summation der Teillautheiten ergibt die wahrgenommene Lautheit, deren Änderung schließlich für die Qualitätsbewertung herangezogen wird. Auf vergleichbaren Ansätzen beruhen erweiterte neuere Modelle [16, 17, 29, 31, 37, 149, 178, 181], die in dem ITU-R-Standard PEAQ [103] zu einem Verfahren für die Qualitätsbewertung von Audiosignalen zusammengefaßt wurden. Das gilt ebenfalls für einige Modelle zur Qualitätsbewertung von Sprachsignalen [18, 77, 78, 105], die für einen Signalqualitätsbereich ausgelegt sind, der die volle Sprachverständlichkeit voraussetzt. Alle genannten Lautheits-Modelle zur Qualitätsbewertung zeichnen sich unter anderem durch eine dem Gehör besser angepaßte Spektralzerlegung und präzisere Nachbildung der Zeitabhängigkeit der Wahrnehmung gegenüber den oben genannten psychoakustischen Modellen der Audiocodierung aus.

Unabhängig von der Audiocodierung wurden in der Psychoakustik zahlreiche Modelle zur Nachbildung von Wahrnehmungsschwellen entwickelt, die nur in sehr geringem Maß Annahmen über die Schallverarbeitung des Gehörs voraussetzen, wie zum Beispiel [95, 109, 122, 157]. Sie werden hier als *Black-Box-Modelle* bezeichnet und dienen zur mathematischen Beschreibung von Verdeckungseffekten spezieller Maskierer-Testsignal-Kombinationen. Eine Black-Box stellt einen analytischen Zusammenhang zwischen den physikalischen

Parametern des Schallreizes und einer Empfindungsgröße her. Aufgrund der Spezialisierung wird eine hohe Genauigkeit in der Regel jedoch nur für die Signale und die Empfindungsgröße erreicht, für die das jeweilige Modell entworfen wurde. Erst seit einigen Jahren gibt es Entwicklungen, die eine Nachbildung von Wahrnehmungsschwellen für einen weniger eingeschränkten Bereich von Maskierersignalen und auch für mehrere Verdeckungseffekte anhand eines einzelnen Modells mit ausreichender Genauigkeit erlauben. Dazu wird das Eingangssignal zunächst einer Spektralzerlegung und danach jede Spektralkomponente einer Black-Box zugeführt. Diese erweiterten Modelle setzen entweder schmalbandige Testsignale voraus oder erlauben aufgrund einer parametrischen Signaldarstellung keine Verarbeitung beliebig komplexer Maskierer und Testsignale [82, 132]. Wegen dieser Einschränkungen wurden diese Modelle bislang nicht in der Codierung angewendet, bei der im allgemeinen breitbandige Codierverzerrungen auftreten. Andere Verfahren zur Nachbildung von Wahrnehmungsschwellen, wie beispielsweise die drei folgenden Modelle, sind demgegenüber prinzipiell auch auf beliebige Maskierer und Testsignale anwendbar. Das erste Modell in [66, 67, 68] wurde jedoch nur für sehr wenige Maskierersignale verifiziert. Das zweite Modell [41, 42] erreicht für bestimmte spektrale und zeitliche Verdeckungseffekte eine gute Übereinstimmung mit psychoakustischen Messungen. Zur Spektralzerlegung wird in diesem Modell als Vorverarbeitungsstufe ein physiologisch motiviertes, lineares Modell der Cochlea eingesetzt. Das dritte Modell [38, 39, 40] ist gegenüber dem zweiten modifiziert und für die Nachbildung der Wahrnehmung von Amplitudenmodulation erweitert.

Die Bestimmung der Wahrnehmungsschwelle über Erregungsmuster als gemeinsame Basis der drei erläuterten Modellklassen führt zu der entscheidenden Konsequenz, daß dort nur die Verdeckungswirkung berücksichtigt wird, die infolge der Erregung durch den Maskierer verursacht wird. Für ein Testsignal unterhalb der Wahrnehmungsschwelle verhindert diese Erregung die Wahrnehmung der erzeugten Erregungszunahme. Dabei wird angenommen, daß sich die vom Testsignal hervorgerufene Erregung unabhängig überlagert. Verdeckungswirkungen, die durch eine nichtlineare Interaktion von Maskierer- und Testsignal in der Cochlea entstehen, werden daher nicht berücksichtigt. Diese Interaktion bewirkt eine Suppression, die sich in einer erheblichen Verminderung der vom Testsignal hervorgerufenen Erregung äußert und zu einer erhöhten Mithörschwelle führt [133, 148]. Eine weitere Auswirkung der Nichtlinearität auf die Mithörschwelle ist durch entstehende Verzerrungsprodukte von Maskierer- und Testsignal gegeben, die eine Erregung an weiteren Orten in der Cochlea zur Folge haben. Da es möglich ist, daß die Verzerrungsprodukte weniger verdeckt werden als das Testsignal selbst, können sie zu einer verminderten Mithörschwelle führen. Die oben genannten Modelle bilden diese nichtlinearen Eigenschaften des Gehörs nicht nach und führen daher zu fehlerhaften Wahrnehmungsschwellen bei wirksamer Suppression oder bei der Wahrnehmung von Verzerrungsprodukten. Die Nachbildungen der cochleären Schallverarbeitung der folgenden vierten Modellklasse bieten demgegenüber Ansätze zur Berücksichtigung der Nichtlinearität.

Im Bereich der Gehörforschung sind physiologische Modelle mit dem Ziel entwickelt worden, beobachtete physiologische Vorgänge zu erklären oder Hypothesen über die Funktionsweise der Cochlea zu verifizieren. Aufgrund der genannten Mängel der oben erwähnten

Modelle sollen in dieser Arbeit physiologische Modelle betrachtet werden, da sie eine Nachbildung der Cochlea mit ihren ausgeprägten nichtlinearen Eigenschaften erlauben. Diese *Cochlea-Modelle* können somit für die Nachbildung der Suppression und der Verzerrungsprodukte herangezogen werden. Mit einem Cochlea-Modell können beispielsweise die Schwingungen der cochleären Trennwand, und mit einem Modell der inneren Haarzellen auch die Entladungsraten des Hörnervs nachgebildet werden. Einige Cochlea-Modelle einschließlich Zwickers Analogmodell [195] werden in Kapitel 3 erläutert. In [195] konnte bereits die Suppressionswirkung bei der Verdeckung von Sinustönen durch Sinustöne gezeigt werden [118, 119, 120]. Eine Erweiterung des Zwickerschen Analogmodells [153, 201] ist bereits erfolgreich zur Simulation der wesentlichen Eigenschaften otoakustischer Emissionen eingesetzt worden. Diese Emissionen werden in der Cochlea erzeugt und können mit einem Mikrofon am Außenohr gemessen werden. Unter speziellen Testbedingungen können anhand der Emissionen auch Mithörschwellenverläufe bestimmt werden [202]. Dieses erweiterte Cochlea-Modell ist ein im Hinblick auf die Nachbildung von Wahrnehmungsschwellen weitgehend erprobtes nichtlineares Modell, das jedoch nur für die Mithörschwellenbestimmung von Sinustönen verifiziert wurde. Ein nichtlineares physiologisches Modell, das zur Nachbildung von Wahrnehmungsschwellen für beliebige Audiosignale als Maskierer und Codiervverzerrungen als Testsignal erweitert und systematisch untersucht wurde, ist bislang nicht bekannt.

Ziel dieser Arbeit ist die Entwicklung eines psychophysiologischen Gehörmodells zur Nachbildung von Wahrnehmungsschwellen, das gegenüber den oben erläuterten bekannten Modellen eine präzisere Nachbildung durch die Berücksichtigung der wesentlichen nichtlinearen Eigenschaften des Gehörs erreicht. Dazu soll das Gehörmodell aus einem physiologisch und einem psychoakustisch motivierten Teil zusammengesetzt werden. Aus einer vom physiologischen Teil des Gehörmodells gewonnenen gehörinternen Größe, z. B. der Entladungsrates des Hörnervs, sollen anschließend mittels des psychoakustischen Teils des Gehörmodells die Wahrnehmungsschwellen bestimmt werden. Dieser Ansatz vermeidet das Problem, daß weder die aus der Literatur bekannten psychoakustischen, noch die physiologischen Meßdaten allein für eine umfassende Modellierung von Wahrnehmungsschwellen ausreichend sind. Für den Parameterabgleich eines psychophysiologischen Modells können die Meßdaten aus beiden Gebieten verwendet werden. Ein weiterer Vorteil einer physiologischen Nachbildung besteht darin, daß diese unabhängig von der Komplexität des akustischen Eingangssignals gültig ist.

Der Entwurf des psychophysiologischen Gehörmodells führt zu den nachfolgend genannten wesentlichen Teilproblemen, deren schrittweise Lösung sich in der Gliederung dieser Arbeit widerspiegelt.

Basierend auf den aus der Literatur bekannten physiologischen Meßergebnissen soll zunächst der physiologische Teil des Gehörmodells bestehend aus den Komponenten für das Außen-, Mittel- und Innenohr entwickelt werden, der alle als relevant bewerteten physiologischen Verarbeitungsschritte beinhaltet. Die wesentlichen nichtlinearen Komponenten sind dabei im physiologischen Modell der Cochlea enthalten. Da die physiologischen Eigen-

schaften der neuronalen Verarbeitung im Gehirn nur unzureichend bekannt sind, soll diese Verarbeitung durch den psychoakustischen Teil des Gehörmodells nachgebildet werden. Um aus der Kombination beider Teile eine geeignete Nachbildung des Gehörs zu erreichen, müssen die darin enthaltenen Komponenten genau aufeinander abgestimmt werden. Der psychoakustische Teil des Gehörmodells ist deshalb derart zu entwerfen, daß er die Vorverarbeitung des Eingangssignals durch den physiologischen Teil in geeigneter Weise ergänzt. Der physiologische Teil des Gehörmodells soll mit bekannten physiologischen Meßdaten abgeglichen werden. Die Nachbildung der Wahrnehmungsschwellen beliebig komplexer Eingangssignale erfordert den Abgleich des gesamten Gehörmodells mit psychoakustischen Meßergebnissen. Die einzelnen Komponenten des Gehörmodells müssen mathematisch beschrieben und in Form numerischer Algorithmen auf einem Digitalrechner implementiert werden.

Die Leistungsfähigkeit des Gehörmodells soll anhand psychoakustisch gemessener Wahrnehmungsschwellen verifiziert werden. In Ergänzung dazu soll der physiologische Teil durch einen Vergleich mit Meßdaten aus physiologischen Untersuchungen überprüft werden. Die Auswertung soll zeigen, welche Wahrnehmungseigenschaften vom Gehörmodell nachgebildet werden und in welchem Maß eine Übereinstimmung der nachgebildeten Wahrnehmungsschwellen mit psychoakustisch gemessenen Schwellen erzielt wird.

In einem Anwendungsbeispiel soll das Gehörmodell zur Steuerung eines Audiocoders zur Datenreduktion eingesetzt werden. Dazu soll ein standardisiertes Codierverfahren herangezogen werden, bei dem das bisher verwendete psychoakustische Modell durch das neue Gehörmodell ersetzt werden soll. Dies erfordert die Entwicklung einer Codersteuerung, die geeignete Steuergrößen für den Audiocoder aus den Ausgangsgrößen des Gehörmodells generiert. Dazu muß das Eingangssignal des Coders und das Ausgangssignal des Decoders durch das Gehörmodell ausgewertet werden. Aufgrund der nichtlinearen Eigenschaften des Gehörmodells sind die Steuergrößen voraussichtlich nur iterativ zu bestimmen. Die mit Hilfe des Gehörmodells erzielte Klangqualität soll anhand von Bewertungen aus Hörtests mit dem bisher verwendeten psychoakustischen Modell verglichen und die unterschiedlichen Eigenschaften beider Modelle anhand der erzeugten Wahrnehmungsschwellen analysiert werden.

In *Kapitel 2* sind die Grundlagen der physiologischen und psychoakustischen Eigenschaften des Gehörs dargestellt. Eine Analyse bekannter Konzepte zur Erklärung der Wahrnehmungseigenschaften gibt einen Überblick und Aufschluß über die Eignung dieser Konzepte für die Zielsetzung dieser Arbeit. In *Kapitel 3* wird die Struktur des psychophysiologischen Gehörmodells und der Aufbau der darin enthaltenen Teilmodelle mit Hilfe der physiologischen und psychoakustischen Grundlagen abgeleitet und im Detail angegeben. *Kapitel 4* beinhaltet die Ergebnisse der anhand von Daten aus verschiedenen physiologischen und psychoakustischen Messungen durchgeführten Verifikation des Gehörmodells. Die Anwendung des Gehörmodells zur Steuerung eines Audiocodierverfahrens wird in *Kapitel 5* beschrieben. Die damit erzielten Codierergebnisse werden mit einem bestehenden Codierverfahren verglichen. In *Kapitel 6* werden die wesentlichen Ergebnisse der Arbeit abschließend zusammengefaßt.

2 Grundlagen

In der Physiologie und Psychoakustik geht mit der stetigen Verfeinerung der Meßmethoden eine wesentliche Erweiterung des Kenntnisstandes einher, was die Verfügbarkeit umfassenderer und besser fundierter Modelle nach sich zieht. In diesem Kapitel werden daher die für diese Arbeit wichtigsten, dem Stand der Forschung entsprechenden Ergebnisse aus der Anatomie, Physiologie und Psychoakustik zusammengefaßt. Ein Modell zur Nachbildung von Wahrnehmungsschwellen sollte die nach dem derzeitigen Stand der Gehörforschung als sicher geltenden Ergebnisse berücksichtigen. In anderen Teilbereichen, in denen noch konkurrierende Modellvorstellungen vorhanden sind, sollte die erfolgversprechendste Variante im Hinblick auf eine möglichst präzise Wahrnehmungsschwellennachbildung ausgewählt werden.

2.1 Anatomie und Physiologie des Gehörs

Das Gehör empfängt Schalldruckschwankungen aus der Umgebung und verarbeitet diese zu auditiven Wahrnehmungen. Das Hörorgan gliedert sich in das in Bild 2.1 dargestellte periphere Gehör, bestehend aus dem linken und rechten Außen-, Mittel- und Innenohr, und das zentrale Gehör (auch als Hörbahn bezeichnet). Im Rahmen dieser Arbeit werden mit dem Begriff „zentrales Gehör“ die an der Reizleitung und -verarbeitung beteiligten Nerven der aufsteigenden Hörbahn beginnend mit den Spiralganglienzellen des VIII. Hirnnervs bis zum auditorischen Cortex sowie die absteigende Hörbahn vom auditorischen Cortex bis zum Innenohr zusammengefaßt. Das Außen- und Mittelohr stellt ein Reiztransportorgan dar, das

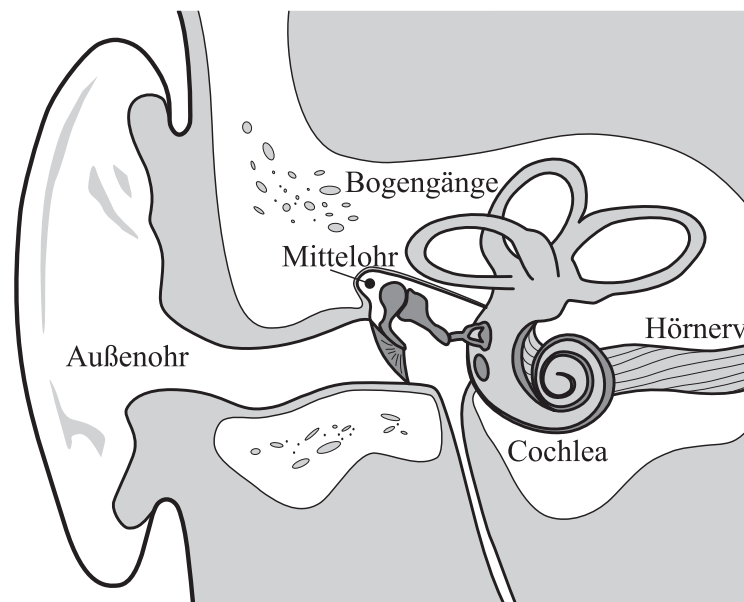


Bild 2.1 Schnitt des peripheren Gehörs. Die Bogengänge sind Teil des Gleichgewichtsorgans, das zusammen mit der Schnecke (Cochlea) das Innenohr bildet (in Anlehnung an [197], nicht maßstabsgerecht).

primär zur Aufnahme des Schallsignals und der akustischen Impedanzanpassung zwischen der umgebenden Atmosphäre und der Innenohrflüssigkeit dient. Die Transduktion des vorverarbeiteten Schallsignals in eine neuronale Repräsentation bzw. Codierung erfolgt in der Schnecke (Cochlea) des Innenohrs. Der verbleibende Teil des Innenohrs, das Gleichgewichtsorgan, ist an der Schallverarbeitung nicht beteiligt. Das zentrale Gehör verarbeitet die vom linken und rechten peripheren Gehör kommenden neuronalen Signale und ermöglicht damit u. a. die räumliche Wahrnehmung oder die Tonhöhenwahrnehmung. Eine Verknüpfung mit erlernten Mustern erlaubt darüber hinaus komplexere Wahrnehmungsleistungen, wie zum Beispiel die Spracherkennung oder die Erkennung von Schallquellen.

2.1.1 Peripheres Gehör

Das Außenohr besteht aus Ohrmuschel und äußerem Gehörgang, der durch das Trommelfell abgeschlossen wird. Der äußere Gehörgang hat Abmessungen von etwa 25 mm Länge und einen mittleren Durchmesser von 7–8 mm, so daß ein Volumen von rund 1 cm³ umschlossen wird.

Das am Trommelfell beginnende Mittelohr umfaßt im wesentlichen die Mittelohrknöchelchen Hammer, Amboß und Steigbügel. Die Steigbügel Fußplatte ist mit dem ovalen Fenster des Innenohrs verwachsen. Das Trommelfell wirkt wie ein Schalldruckempfänger, dessen Auslenkung unmittelbar die Kette der Mittelohrknöchelchen mitbewegt. Durch den unterschiedlichen Flächeninhalt des Trommelfells im Vergleich zum ovalen Fenster und die Hebelwirkung von Hammer und Amboß wird die Schallimpedanzanpassung zwischen der Luftsäule im äußeren Gehörgang und der Lymphflüssigkeit am Eingang der Cochlea erreicht. Ferner existieren verschiedene Mittelohrmuskeln, die durch Anspannung zu einer Verringerung der Beweglichkeit der Mittelohrknöchelchen führen. Dieser Mechanismus wird bei hohen Schallpegeln wirksam, um das Innenohr vor sehr großen, möglicherweise schädlichen mechanischen Beanspruchungen zu schützen.

Die Cochlea wird durch den in zweieinhalb Windungen schneckenförmig aufgerollten Teil des Innenohrs gebildet, der in entrollter Form in Bild 2.2 schematisch dargestellt ist und einen abnehmenden Durchmesser von etwa 3 mm an der Basis bis 1,5 mm an der Spitze aufweist. Sie ist in eine sehr harte Knochensubstanz eingebettet und mit Lymphflüssigkeit gefüllt. Der durch die Knochenwand gebildete schneckenförmige Kanalraum wird durch die Reissnersche Membran und die Basilarmembran (BM) in drei Kammern (Scalae) längs des Kanals geteilt (s. Bild 2.3). Das ovale Fenster grenzt die Scala vestibuli, die sich oberhalb der Reissnerschen Membran befindet, zum Mittelohr ab. Die Scala media erstreckt sich zwischen der Reissnerschen Membran und der BM. Sie enthält Endolymphe mit einem Potential von +80 mV. Das runde Fenster bildet den Abschluß der Scala tympani, die sich unter der BM befindet und an der Schneckenspitze, dem Helicotrema, mit der Scala vestibuli durch eine Membranöffnung verbunden ist. Beide Kammern enthalten Perilymphe mit einem Potential von 0 mV. Auf der BM sitzt das Cortische Organ mit den inneren Haarzellen, die als Sensoren fungieren. Das Cortische Organ wird durch die Tektorialmembran bedeckt (s. Bild 2.4). Das System aus BM, Cortischem Organ und Tektorialmembran wird als cochleäre Trennwand bezeichnet, die etwa 32 mm lang ist. Innerhalb des Cortischen Organs sind

etwa 3400 innere Haarzellen (IHZs) in einer Reihe angeordnet sowie 3 bis 4 weitere Reihen von insgesamt etwa 13000 äußeren Haarzellen (ÄHZs) vorhanden. Die IHZs werden von 90–95% der 30000–40000 afferenten Nervenfasern innerviert, wobei jede Faser mit genau einer IHZ und jede IHZ mit bis zu 20 Fasern Synapsen bildet. Dieser hohe Anteil der afferenten, zum zentralen Gehör führenden Fasern zeigt, daß die IHZs eine Sensorfunktion haben. Demgegenüber werden die ÄHZs von ca. 90% der efferenten Nervenfasern, deren Anzahl in der Größenordnung von 1000 liegt, kontaktiert, indem jede Faser mit 10–50 benachbarten ÄHZs und jede ÄHZ mit bis zu 6 verschiedenen Fasern Synapsen bildet.

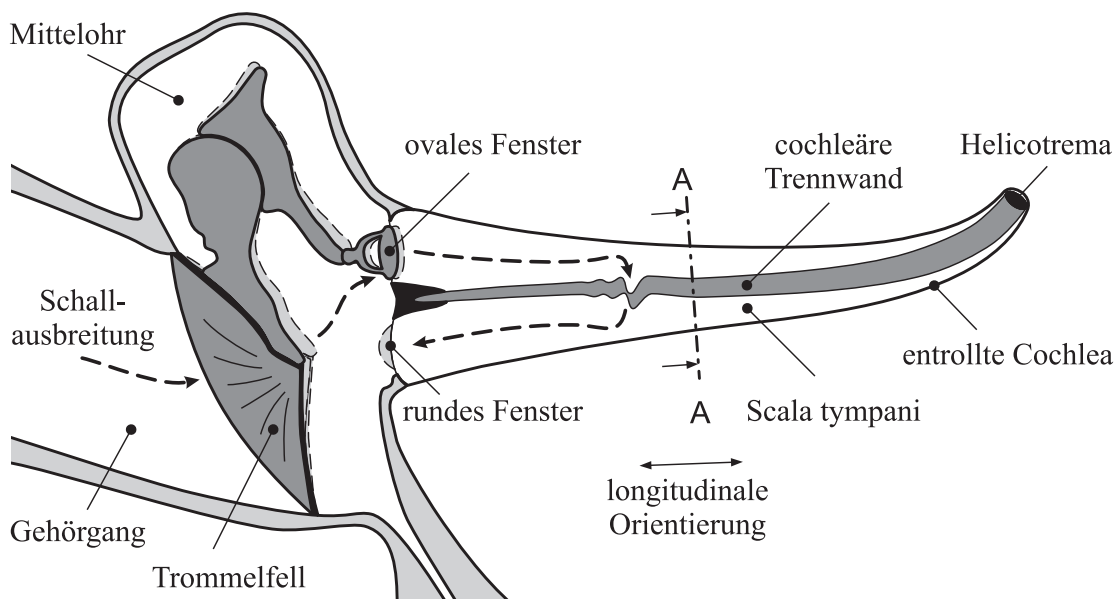


Bild 2.2 Schnitt des Mittelohres und der entrollten Cochlea (schematisch). Bei einer sinusförmigen Trommelfellschwingung prägt sich an einem Ort der cochleären Trennwand eine Resonanz aus. Vor allem an diesem Ort findet der Druckausgleich zwischen der Ober- und Unterseite der Trennwand statt. Die Auslenkung der Mittelohrknöchelchen und Fenstermembranen der Cochlea ist zur Veranschaulichung gestrichelt und stark vergrößert dargestellt.

Für die Schallverarbeitung in der Cochlea ist die Hydromechanik ausschlaggebend, die sich aus der Anatomie, insbesondere ihrer Geometrie und Materialeigenschaften ergibt. Eine Auslenkung des ovalen Fensters hat aufgrund der nahezu inkompressiblen Lympflüssigkeit eine entgegengesetzte Auslenkung des runden Fensters zur Folge. Eine sinusförmige Schwingung des ovalen Fensters bei Frequenzen im hörbaren Bereich führt in axialer Richtung zu einer Druckdifferenz zwischen der Ober- und Unterseite der cochleären Trennwand, da die Auswirkungen der Reissnerschen Membran auf die Hydromechanik zu vernachlässigen sind. Die Trennwand stellt ein schwingungsfähiges mechanisches System dar, das aufgrund seiner Steifigkeit die Druckdifferenz mit zunehmender Entfernung vom ovalen Fenster zunächst nur langsam abbaut, bis die ortsabhängige Resonanzfrequenz der Trennwand mit der Schwingungsfrequenz übereinstimmt. An diesem Ort findet ein Druckausgleich zwi-

schen der Scala vestibuli bzw. der Scala media und der Scala tympani statt, so daß jenseits dieses Ortes bis zum Helicotrema keine Druckdifferenz mehr vorhanden ist (s. Bild 2.2). Aufgrund ihrer abnehmenden Steifigkeit bei zunehmender Breite vom ovalen Fenster bis zum Helicotrema weist die cochleäre Trennwand eine über ihre Länge kontinuierlich abnehmende örtliche Resonanzfrequenz auf, die als charakteristische Frequenz (CF) bezeichnet wird. Daher erzeugen Schallsignale verschiedener Frequenz eine maximale Auslenkung der Trennwand jeweils an einem anderen Ort, so daß die Cochlea eine Spektralzerlegung des Eingangssignals, die sog. „Frequenz-Orts-Transformation“, bewirkt. In dem Bereich zwischen ovalem Fenster und dem Resonanzort wird die Rückstellkraft der cochleären Trennwand von ihrer Steifigkeit bestimmt. Die transversalen Schwingungen der Trennwand in diesem Bereich sind mit Oberflächenwellen auf einer Wasserfläche vergleichbar, die sich wesentlich langsamer ausbreiten als Kompressionswellen innerhalb des Mediums [151]. Für die sich entlang der cochleären Trennwand ausbreitende transversale Schwingung wird der Begriff „Wanderwelle“ verwendet [19]. Transversale Schwingungen der cochleären Trennwand führen zu einer Reizung der IHZs aufgrund der Auslenkung der Zilien auf der Haarzelle (s. Bild 2.5). Die IHZs führen die eigentliche Transduktion des akustischen Signals in ein neuronales Signal mit Hilfe zwischengeschalteter biochemischer Prozesse durch.

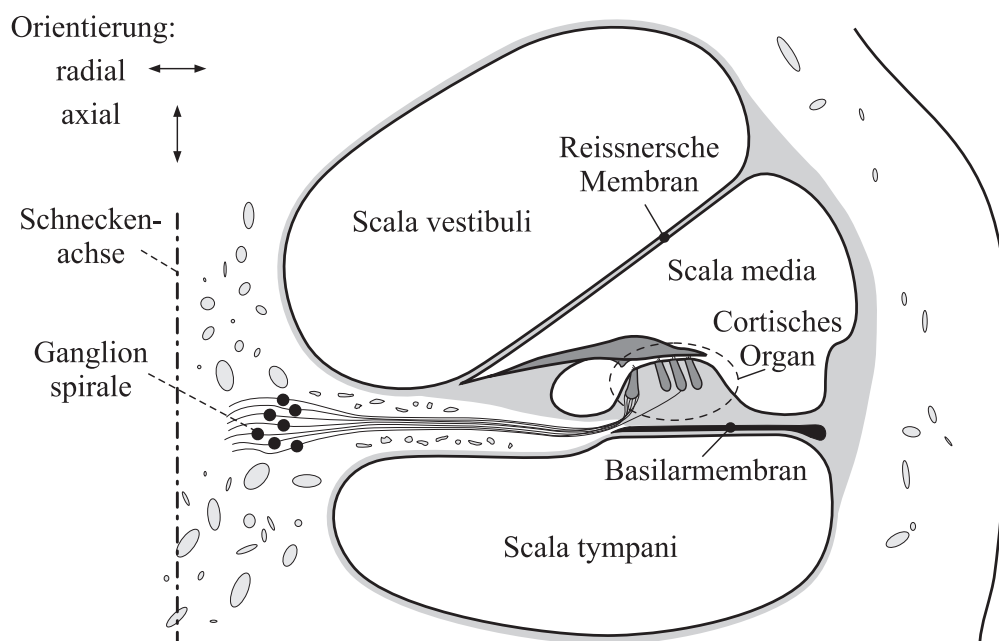


Bild 2.3 Querschnitt A – A der Cochlea aus Bild 2.2. Der um die Schneckenachse gewundene Kanal wird durch die Reissnersche Membran und die Basilarmembran in drei Kammern geteilt. Die Basilarmembran mit dem aufsitzenden Cortischen Organ, das die Sensoren beinhaltet, bildet zusammen mit der darüberliegenden Tektorialmembran die cochleäre Trennwand. Die Reissnersche Membran isoliert die unterschiedlichen Lymphflüssigkeiten von Scala vestibuli und Scala media und hat praktisch keine hydromechanische Bedeutung.

Die bisherige Betrachtung geht von einer makromechanischen Modellvorstellung der Cochlea aus, bei der die cochleäre Trennwand als homogene Membran betrachtet wird, die durch Schwingungen des ovalen Fensters zu transversalen Schwingungen angeregt wird. Für die Modellierung der Schallverarbeitung sind jedoch auch die innerhalb der cochleären Trennwand vorhandenen Relativbewegungen von Bedeutung, die durch sog. „mikromechanische“ Modelle der Trennwand beschrieben werden, z. B. [4, 61, 125, 138, 151].

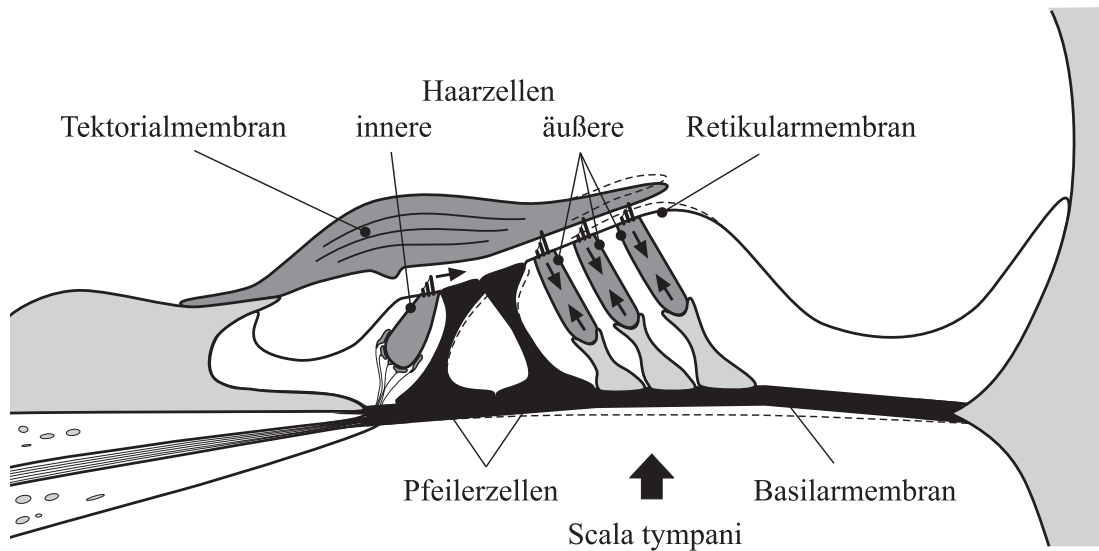


Bild 2.4 Querschnitt des Cortischen Organs. Eine transversale Schwingung (in Pfeilrichtung) der cochleären Trennwand hat eine Scherbewegung zwischen der Basilarmembran und der Tektorialmembran im Bereich des Cortischen Organs zur Folge. Die Tektorialmembran und die der Schneckenachse zugewandte Seite der Basilarmembran rotieren dabei jeweils um eine Achse längs ihrer Aufhängung. Eine entsprechende Rotation der relativ steifen Pfeilerzellen bewirkt die Scherung der Haarbündel der äußeren Haarzellen zwischen Retikular- und Tektorialmembran. Die Flüssigkeitsgeschwindigkeit am Haarbündel der inneren Haarzellen ist das mechanische Reizsignal der Sensoren. Durch Kontraktion der äußeren Haarzellen wird die gestrichelt dargestellte, von einer Schwingung des ovalen Fensters hervorgerufene Auslenkung der Basilarmembran verstärkt. Diese Kontraktion bewirkt auch eine Verschiebung der inneren Haarzelle, so daß sich die Relativgeschwindigkeit des Haarbündels gegenüber der umgebenden Flüssigkeit erhöht.

Eine transversale Auslenkung der cochleären Trennwand führt zu einer Scherbewegung zwischen Tektorialmembran und BM mit aufsitzendem Cortischen Organ (s. Bild 2.4). Diese Scherung hat einerseits die Auslenkung der Haarbündel der ÄHZs, die mit der Spitze in der Tektorialmembran verankert sind, zur Folge. Andererseits wird eine Flüssigkeitsverschiebung zwischen Tektorialmembran und Retikularmembran verursacht, die die Haarbündel der IHZs aufgrund der Reibung proportional zur Flüssigkeitsgeschwindigkeit auslenkt. Eine radiale, von der Schneckenachse nach außen gerichtete Auslenkung der Haarbündel führt zur

Öffnung von Ionenkanälen in der Zellmembran. Die damit hervorgerufenen Ionenströme depolarisieren die IHZ aufgrund der erhöhten K^+ -Konzentration. Die Depolarisation bewirkt die vermehrte Freisetzung von Transmittersubstanz der IHZ in die synaptischen Spalte der verbundenen afferenten Hörnervenfasern. Die Wahrscheinlichkeit für die Auslösung eines Aktionspotentials auf den Nervenfasern wächst mit zunehmender Transmitterkonzentration. Nach jeder Auslösung eines Aktionspotentials tritt eine Verringerung der Transmitterkonzentration ein. Im Ruhezustand werden bereits Aktionspotentiale ausgelöst, deren Häufigkeit als sog. „spontane“ Entladungsrates der Nervenfaser bezeichnet wird. Während eine Depolarisation der IHZ zu einer Erhöhung der Entladungsrates über die spontane Rate hinaus führt, bewirkt eine Hyperpolarisation durch eine Haarbündelauslenkung zur Schneckenachse hin eine Reduzierung der Rate [155, 190].

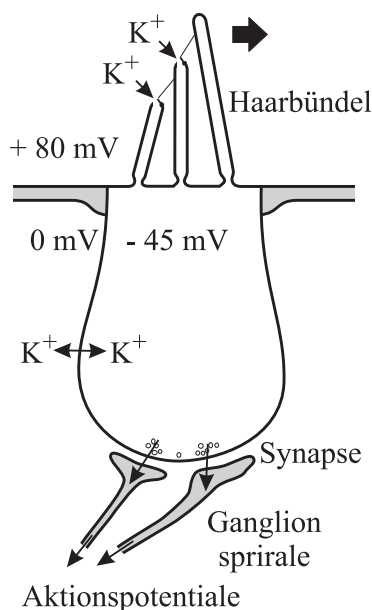


Bild 2.5 Längsschnitt einer inneren Haarzelle. Die Reizung der inneren Haarzelle durch radiale Auslenkung des Haarbündels von der Schneckenachse nach außen führt zur Öffnung von Ionenkanälen an der Spitze der Zilien. Das elektrische Potential erhöht sich daraufhin. Diese Depolarisation führt zur erhöhten Freisetzung von Transmittersubstanz in den synaptischen Spalten und damit zu einer erhöhten Entladungswahrscheinlichkeit (Aktionspotentialgenerierung) auf den Nervenfasern der Spiralganglien. (In Anlehnung an [155].)

Eine Auslenkung der BM nach oben führt auch bei den ÄHZs aufgrund der Scherung zu einer Haarbündelauslenkung nach außen und zur Depolarisation der ÄHZs. Die Depolarisation bewirkt eine aktive Kontraktion der ÄHZs in Richtung ihrer Längsachse, mit der innerhalb der Struktur der cochleären Trennwand eine Verringerung des Abstandes zwischen Retikularmembran und BM erfolgt (s. Bild 2.4) [89, 190]. Bei einer Auslenkung der BM nach unten tritt eine Hyperpolarisation der ÄHZs ein, die eine aktive Elongation der Zelle bewirkt. Durch die ÄHZ wird daher eine Kraft auf die BM ausgeübt, die eine aufgrund der Schwingung des ovalen Fensters vorhandene Auslenkung der BM noch verstärkt [141]. Eine ÄHZ wirkt daher wie ein Motor, der im folgenden als mechanischer ÄHZ-Antrieb bezeichnet wird. Durch eine radiale Verschiebung der Retikularmembran und der darin verankerten IHZs durch die Längenänderung der schräg angeordneten ÄHZs ist darüber hinaus ein Mechanismus denkbar, der zu einer verstärkten Auslenkung der Haarbündel der IHZs aufgrund einer entsprechenden Haarbündelverschiebung in der masseträgen Flüssigkeit führen kann.

Die Längenänderungen der ÄHZs können Schwingungen mit Frequenzen von mehr als 30 kHz folgen, ihre Amplitude besitzt jedoch einen kleineren Dynamikbereich verglichen mit dem Dynamikumfang des Gehörs. Die Verstärkung ist daher nur bei einem Schalldruckpegel (SPL) bis etwa 50 dB SPL in vollem Umfang wirksam, während bei einer Erhöhung des Pegels eine Begrenzung der Längenänderung eintritt, so daß die Verstärkung der BM-Auslenkung dann im Vergleich zu einem passiven System ohne die ÄHZ-Antriebe vernachlässigbar klein ist. Aufgrund der mikromechanischen Struktur der cochleären Trennwand wird vermutet, daß die Masse der Tektorialmembran und die Steifigkeit der ÄHZ-Haarbündel ein System mit Resonanzen im Audiofrequenzbereich darstellt. Die Lage der Resonanzfrequenzen dieses Feder-Masse-Systems in Bezug zu derjenigen der BM bestimmt das Ausmaß der Verstärkung oder Dämpfung der Reizung der IHZs [73, 85]. Die Krafteinkopplung auf die BM durch die ÄHZs wirkt wie eine Entdämpfung, wenn sie sich in Phase zur BM-Schwingung befindet. Die einigen Modellen zugrundeliegende Annahme, daß die ÄHZs eine Veränderung der Steifigkeit der BM oder der Steifigkeit ihrer Stereozilien bewirken und zu einer dynamischen Verschiebung der Resonanzfrequenzen der entsprechenden Feder-Masse-Systeme führen [2, 3], konnten bislang weder bewiesen noch widerlegt werden. Eine ausreichende meßtechnische Verifikation komplexerer mikromechanischer Modelle, z. B. [60, 125, 137], ist derzeit nicht gegeben. Diese Modelle sind daher im allgemeinen nicht universell für quantitative Untersuchungen einsetzbar, sondern eher als Erklärungsmodelle für bestimmte Funktionshypothesen der Cochlea anzusehen.

2.1.2 Zentrales Gehör

Bild 2.6 zeigt eine stark vereinfachte Darstellung der aufsteigenden (afferenten) Hörbahn, die von den IHZs einer Cochlea ausgeht. Die nicht gezeigte Verknüpfung der neuronalen Signale der rechten und linken Cochlea erfolgt bereits in den ersten Verarbeitungsstufen des Olivenkomplexes, die auf die Spiralganglienzellen folgen. Am Ausgang des Olivenkomplexes sind unterschiedliche Signaltypen identifizierbar, in denen bestimmte Merkmale des Reizsignals von den Spiralganglien codiert sind und parallel zu höher gelegenen Nerven weitergeführt werden [160, 173]. Die Hörbahn endet im auditorischen Cortex, der eine tonotopie (nach Tonfrequenzen geordnete) Organisation aufweist und spezialisierte rezeptive Felder beispielsweise für amplituden- oder frequenzmodulierte Schallsignale besitzt [35].

Die Charakteristika der absteigenden (efferenten) Hörbahn sind lediglich teilweise erforscht. Die nahezu ausschließlich efferente Innervation der ÄHZs deutet auf einen Steuermechanismus hin, der eine aufgrund der neuronalen Verzögerung relativ langsame Adaption der Verstärkung oder der mechanischen Eigenschaften der cochleären Trennwand durch die ÄHZs erlaubt [155].

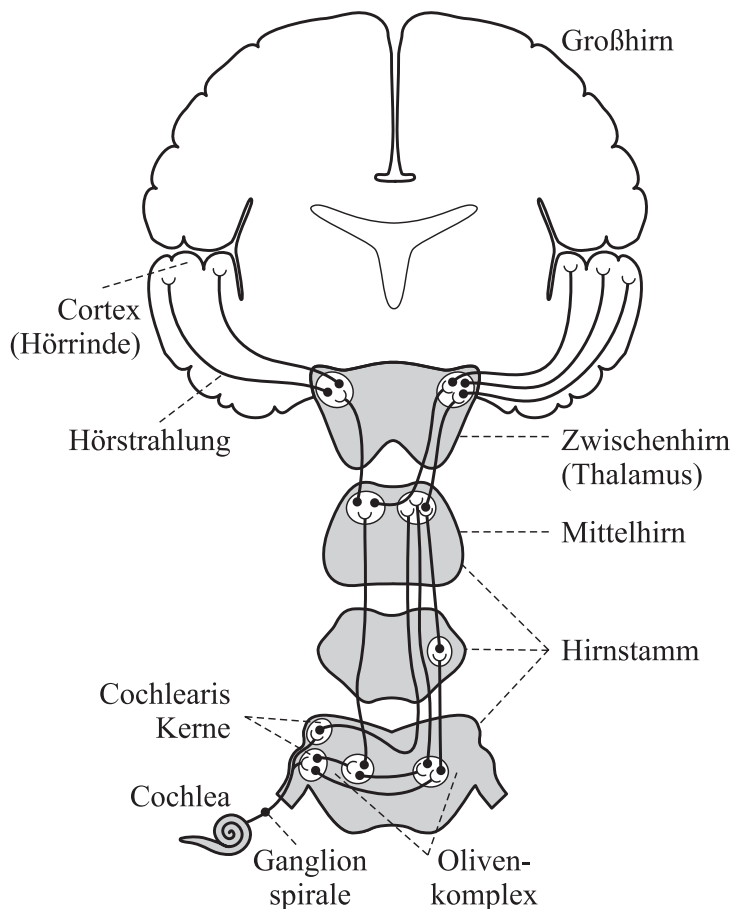


Bild 2.6 Von einer Cochlea aufsteigende Hörbahn des zentralen Gehörs (stark vereinfacht). Der durch die Spiralganglienzellen gebildete Hörnerv gabelt sich im Olivenkomplex auf, wo unterschiedliche spezialisierte Neurone bestimmte Schallsignalmerekmale extrahieren. Eine Verschaltung mit dem contralateralen Hörnerv ist bereits auf dieser Ebene vorhanden. (In Anlehnung an [190].)

2.2 Physikalisch meßbare Reizantworten des Gehörs

Die beiden wesentlichen im Gehör auftretenden Schallsignalrepräsentationen sind akustische Signale im peripheren Gehör sowie elektrische Signale im zentralen Gehör. Für die Nachbildung des Gehörs sind Messungen dieser Signale von Interesse, die vom lebenden Organismus mit einer vollen Funktionsfähigkeit der ÄHZs stammen. Die Messung von Signalen in der Cochlea erfordert daher die Eröffnung des Innenohres *in vivo*. Solche Messungen sind am Menschen in der Regel nicht durchführbar. Aufgrund des strukturell gleichen Aufbaus des Gehörs bei Säugetieren können die in Tierversuchen gewonnenen Ergebnisse jedoch weitgehend auf den Menschen übertragen werden. Der Einfluß des menschlichen Außen- und Mittelohres auf das Schallsignal kann hingegen auch durch Messungen und Modellrechnungen bestimmt werden.

2.2.1 Übertragungseigenschaften des Außen- und Mittelohres

Die Richtungsabhängigkeit der Druckübertragungsfunktion von einer Schallquelle zum Trommelfell wird weitgehend von den akustischen Eigenschaften der Ohrmuschel, des Kopfes und des Oberkörpers bestimmt [21, 22, 172]. Zusätzlich bewirkt der rohrförmige Gehörgang Resonanzen, die bei einer Übereinstimmung seiner Länge mit einem ungeradzahligem

Vielfachen eines Viertels der Wellenlänge auftreten. Die erste Resonanz liegt bei etwa 3 kHz, wie die Ergebnisse in Bild 2.7 für verschiedene Außenohren zeigen.

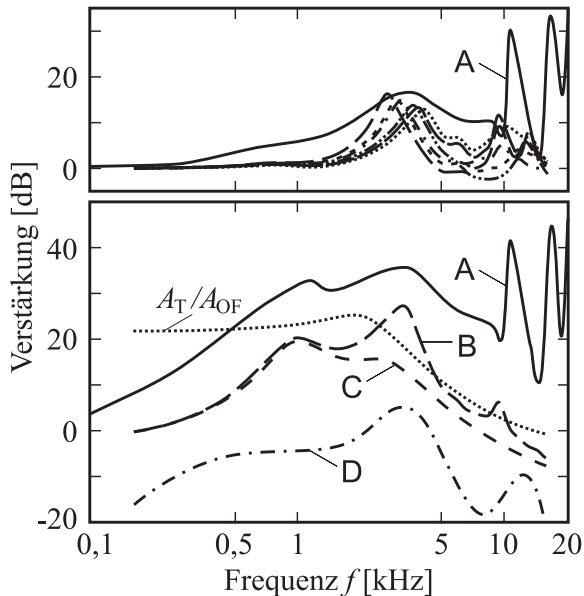


Bild 2.7 *Oben:* Berechnete Druckübertragungsfunktionen von einer externen Schallquelle zum Trommelfell. (A) nach dem Gehörmodell aus [163], übrige Kurven aus der Geometrie verschiedener Gehörgänge aus [179] (nach [94]). *Unten:* Berechnete Druckübertragungsfunktionen von einer externen Schallquelle zum Eingang der Cochlea. (A) nach dem Modell aus [91], (B) nach dem Modell aus [163]. Zum Vergleich: Druckübertragungsfunktion vom Trommelfell zum Eingang der Cochlea (C), effektives Flächenverhältnis von Trommelfell zu Steigbügelfußplatte A_T/A_{OF} und Ruhehörschwelle nach DIN (D) (unskaliert) aus [163].

Die Übertragungsfunktion des Mittelohres hat Bandpaßcharakter, da bei tiefen Frequenzen einerseits eine Dämpfung der Trommelfellschwingungen durch das angekoppelte Luftvolumen der Mittelohrröhle entsteht und andererseits die Elastizität der Mittelohrknöchelchen aufgrund der größeren Auslenkung stärker dämpfend wirkt [155]. Bei höheren Frequenzen wird die Dämpfung hauptsächlich durch die abnehmende effektive Fläche des Trommelfells bestimmt [92, 93, 94, 163]. Eine Anspannung der Mittelohrmuskeln, die bei Schalldruckpegeln von mehr als etwa 80 dB eintritt, bewirkt bei sehr großen Pegeln eine Dämpfung bis zu 10 dB im Frequenzbereich unter 1 kHz [150].

2.2.2 Schallinduzierte Schwingung der cochleären Trennwand

Die Frequenz-Orts-Transformation der Cochlea wird in Bild 2.8 anhand der ortsabhängigen Transversalgeschwindigkeit der BM bei Einspeisung von drei mit gleichen Pegeln überlagerten Sinusschwingungen am ovalen Fenster veranschaulicht. Die Einhüllende der Wanderwelle zeigt jeweils bis zu den drei ausgeprägten Maxima an den Resonanzorten einen langsam ansteigenden Verlauf und danach jeweils ein schnelles Abklingen. Bestimmt man die Einhüllende für eine einzelne Sinusschwingung, so ergibt sich bei einer (linearen) Darstellung über der Ortskoordinate wie in Bild 2.8 eine näherungsweise frequenzunabhängige Charakteristik, deren Maximum an den Ort mit übereinstimmender CF verschoben ist. Da die Einhüllendenbreite ein Maß für die Frequenzselektivität des Gehörs darstellt, wurde die wahrnehmungsbezogene Tonheitsskala mit der Einheit Bark eingeführt. Diese Skala ordnet der ortsabhängigen CF eine von der Ortskoordinate linear abhängige charakteristische Tonheit z_{char} zu, wobei 1 Bark der Filterbandbreite bzw. der Frequenzgruppenbreite entspricht.

Die Umrechnung erfolgt entsprechend [197] mit (2.1). Der in dieser Arbeit verwendete analytische Zusammenhang zwischen der CF und der charakteristischen Tonheit wird weiter unten angeben.

$$\frac{z_{\text{char}}}{[\text{Bark}]} = 0,75 \left(33,3 - \frac{x}{[\text{mm}]} \right). \quad (2.1)$$

Zum Verständnis der Spektralzerlegung des Schallsignals in der Cochlea kann jeder Ort der cochleären Trennwand, an dem eine IHZ die Bewegungsgeschwindigkeit detektiert, als ein Filterkanal aufgefaßt werden. Der Frequenzgang eines Filterkanals kann zum Beispiel durch Messen der frequenzabhängigen Geschwindigkeitsamplitude an diesem Ort bestimmt werden (vgl. Bild 2.9). Der Frequenzgang kann aber ebenfalls aus der ortsabhängigen Einhüllendencharakteristik der BM-Schnelle eines Sinustons mit der CF bestimmt werden, wenn die Einhüllendencharakteristik frequenzunabhängig ist. Es ist dann lediglich die Ortskoordinate der Einhüllendencharakteristik durch die Skala der zugeordneten CFs zu ersetzen.

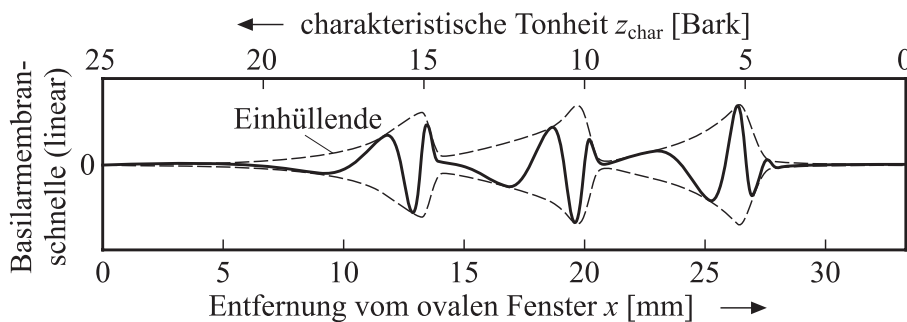


Bild 2.8 Ortsabhängige momentane Basilarmembranschnelle (berechnet) für drei überlagerte Sinustöne mit 60 dB SPL und den Tonheiten 5, 10 und 15 Bark (530, 1280 und 2700 Hz). Die Tonheitsskala steht in linearem Zusammenhang zur Position auf der Basilarmembran und entspricht der zugeordneten charakteristischen Frequenz (CF). Die berechnete Einhüllende ergibt sich aus den Maxima und Minima der Basilarmembranschnelle.

Die Verstärkung der BM-Schwingung durch die ÄHZs kann durch Variation des Schallpegels gemessen werden. In Bild 2.9 ist das Verhältnis von BM-Schnelle zum SPL einer externen Schallquelle, gemessen am Ort mit einer CF von 9 kHz, dargestellt. Bei geringem SPL ist eine ausgeprägte Resonanz erkennbar, deren Maximum 47,9 dB über dem Maximalwert für 80 dB SPL liegt. Dieses Ergebnis kann als entdämpfende Wirkung der ÄHZs bei geringem Pegel interpretiert werden, da bei einer Verringerung der Dämpfung in einem Feder-Masse-System eine vergleichbare Resonanzüberhöhung beobachtet wird. Die Frequenzselektivität des entsprechenden Filterkanals bei der CF nimmt mit zunehmendem Pegel ab.

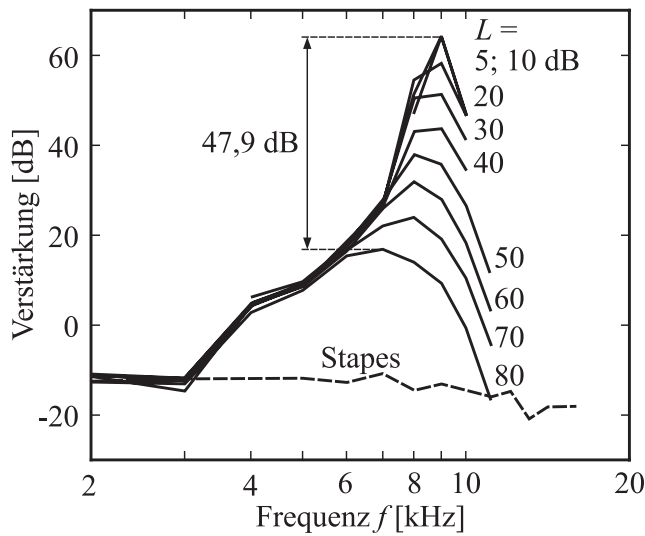


Bild 2.9 Verhältnis von Basilar-membranschnelle zu SPL in Abhän-gigkeit der Sinustonfrequenz f bei Variation des SPL L . Gemessen am Ort mit der charakteristischen Fre-quenz 9 kHz in der Cochlea eines Chinchillas. *Gestrichelt*: mittlere Schnelle des Steigbügels. Gegen-über der Basilar-membranschnelle bei $L = 80$ dB wird bei $L = 0$ dB eine Verstärkung von 47,9 dB erreicht (Daten aus [164]).

Die Wirkung der ÄHZ-Antriebe wird ebenfalls durch einen Vergleich der gemessenen BM-Schnelle in vivo und post mortem deutlich [144, 164]. Bild 2.10 zeigt Ergebnisse von zwei Chinchillas als Darstellung der bei Variation der Frequenz maximal erzielbaren BM-Schnelle für unterschiedliche konstante Schallpegel. Die Verstärkung bei aktivem ÄHZ-Antrieb in vivo beträgt dabei 69 bzw. 54 dB bei etwa 10 dB SPL. In vivo wird durch die Verstärkung bei kleinem SPL eine Dynamikkompression von Schwingungen mit Frequenzen nahe der CF erreicht, die zu einer minimalen Steigung der Eingangs-Ausgangs-Kennlinie (vgl. Bild 2.10) von ca. 0,2 dB/dB bei mittlerem SPL führt.

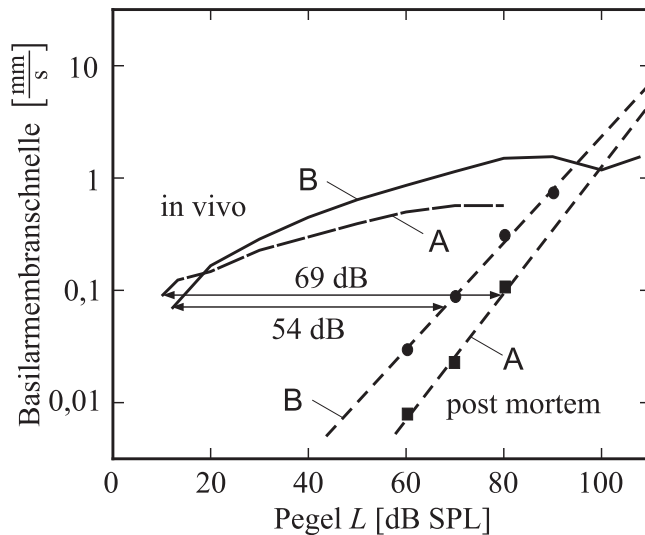


Bild 2.10 Spitzenwert der Basilar-membranschnelle für Sinusbursts mit der charakteristischen Frequenz an der Meßposition in Abhängigkeit vom SPL für zwei Chinchillas (A, B). Präparation A liegt auch Bild 2.9 zugrunde. Durch einen Verstärkungsmechanismus, der in vivo wirksam ist, erreicht A eine max. Verstärkung von 69 dB bzw. B von 54 dB im Vergleich zur post mortalen Kondition (Daten aus [164]).

Schwingungsmessungen der BM werden in der Regel mit einem Laserinterferometer durchgeführt, der von unten (vgl. Bild 2.4) auf die BM fokussiert wird [142]. Messungen von der Oberseite der cochleären Trennwand erlauben die Bestimmung der Schwingungsamplituden an verschiedenen radialen Positionen. Die ersten verfügbaren Meßergebnisse dieser Art geben Hinweise auf unterschiedliche mikromechanische Bewegungsabläufe innerhalb der Trennwand, insbesondere der Tektorialmembran, dem freiliegenden Bereich des Cortischen

Organs und der BM [73, 85]. Die Ergebnisse zeigen, daß die Reizung der IHZs nicht nur von der BM-Schnelle abhängt, sondern auch von Schwingungen der an die BM angekoppelten mikromechanischen Systeme entscheidend mitbestimmt wird.

Verschiedene Meßergebnisse deuten darauf hin, daß die Dynamikkompression der BM-Schwingung nicht durch eine pegeladaptive Verstärkungsregelung, sondern durch eine Sättigungskennlinie des ÄHZ-Antriebs verursacht wird [116]. Aufgrund dieser nichtlinearen Kennlinie entstehen Verzerrungsprodukte in der Cochlea [58, 162, 175], die sowohl in Nervenfaserableitungen nachgewiesen worden sind [65], als auch in Verzerrungsprodukt-otoakustischen-Emissionen (DPOAEs, vgl. Kapitel 2.2.4).

2.2.3 Aktionspotentiale auf einzelnen Hörnervenfaser

Die Entladungsrates einer afferenten Faser des Hörnervs gibt Aufschluß über die Reizung der angeschlossenen IHZ, die wiederum von der Relativgeschwindigkeit des Haarbündels zur umgebenden Flüssigkeit an der Position der IHZ abhängt. Diese Nervenfaserableitungen zeigen die gleiche Frequenzselektivität wie Messungen der BM-Schnelle [161]. Die IHZs sind daher als Sensoren anzusehen, die das mechanische Reizsignal in ein neuronales Signal wandeln, ohne die Frequenzselektivität zu verändern, die auf der mechanischen Ebene vorhanden ist.

Sowohl das Membranpotential der IHZ als auch die Entladungsrates der angeschlossenen Nervenfasern zeigen ein Sättigungsverhalten, das eine Beschränkung des nutzbaren Dynamikbereiches auf der Ebene des Hörnervs impliziert. Die Schallsignalverarbeitung mit wesentlich größerem Dynamikumfang ist dennoch möglich, da der ÄHZ-Antrieb eine Dynamikkompression des IHZ-Reizsignals verursacht und verschiedene Typen von Fasern vorhanden sind, deren Arbeitsbereich bei unterschiedlichen Schalldruckpegeln liegt [189]. In Bild 2.11 sind für unterschiedliche Nervenfasern typische gemessene Entladungsrates dargestellt, wobei der Dynamikbereich zwischen Reizschwelle und Sättigung bei einem Teil der IHZs nur etwa 20 bis 40 dB bezogen auf den SPL umfaßt. Der Dynamikumfang am Eingang der IHZ ist noch geringer, da bei der Messung die Kompression durch den ÄHZ-Antrieb mit-enthalten ist.

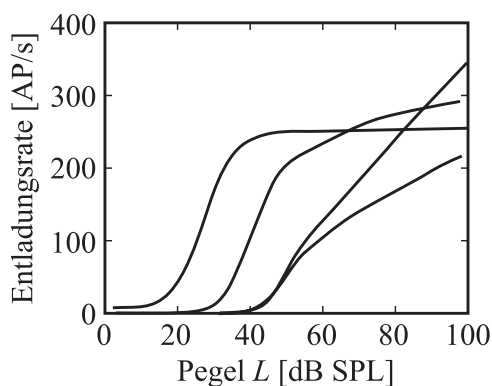


Bild 2.11 Schalldruckabhängigkeit der Entladungsrates unterschiedlicher Fasern des Hörnervs. (Daten aus [189]).

Bei einer Hyperpolarisation der IHZ ist die Entladungsrates auf den angeschlossenen Nervenfasern gegenüber der spontanen Rate im Ruhezustand etwas reduziert. Bei einer Depolarisation steigt die Rate deutlich an, bis eine Sättigung eintritt. Diese asymmetrische Eingangs-

Ausgangs-Kennlinie bewirkt einen Gleichrichtungseffekt (vgl. Bild 2.12) [107, 116]. Bei Reizung der IHZ mit einem Sinuston ist ein schnelles Einschwingen der Entladungsrate zu beobachten, das zu einem Überschwingen der Entladungsrate gegenüber dem eingeschwungenen Zustand führen kann. Die IHZ reagiert daher sehr empfindlich auf Änderungen des Eingangssignals. Nach Abschalten eines Sinustons erfolgt ein langsames Ausschwingen, bei dem die spontane Entladungsrate zeitweise unterschritten werden kann (s. Bild 2.13). Die Dauer des Ausschwingvorgangs hängt von der Adaption der IHZ an das Reizsignal ab. Nach der Adaption wird der Ruhezustand der IHZ relativ langsam wieder erreicht.

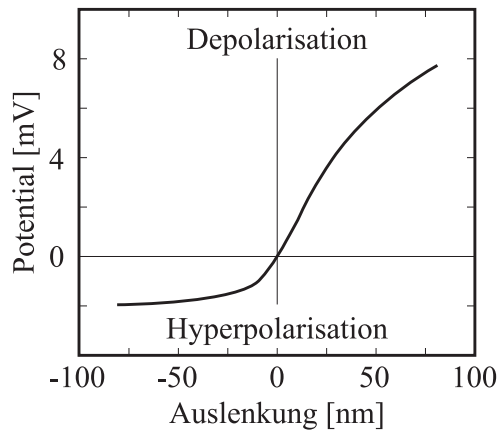


Bild 2.12 Membranpotential einer IHZ in Abhängigkeit von der Haarbündelauslenkung (Daten aus [166]).

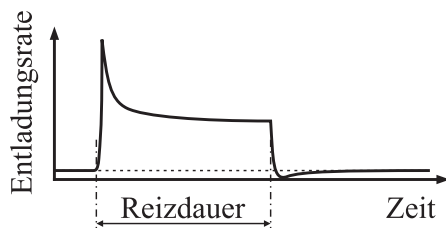


Bild 2.13 Typischer Verlauf der zeitabhängigen Entladungsrate einer Faser des Hörnervs. Die Höhe des Überschwingens beim Einschwingvorgang ist vom Schalldruckpegel abhängig.

Die Entladungsrate mehrerer Hörnervenfaser, die mit benachbarten IHZs verbunden sind, ist bei Sinustonfrequenzen unterhalb von etwa 2 kHz mit der Tonfrequenz moduliert. Aufgrund des Gleichrichtungseffekts wird dann in jeder Sinushalbwelle ein Maximum der Entladungsrate erreicht. Diese reizsynchrone Modulation kann von der nachfolgenden neuronalen Verarbeitung beispielsweise für eine Frequenzanalyse genutzt werden. In welchem Umfang dieser in der neuronalen Repräsentation teilweise erhalten bleibende Zeitverlauf des Eingangssignals für die auditive Wahrnehmung im Gehör ausgewertet wird, ist bislang nicht vollständig geklärt [117, 124, 155].

2.2.4 Otoakustische Emissionen

Otoakustische Emissionen sind Schallaussendungen des peripheren Gehörs, die sowohl spontan ohne Schallreiz vorhanden sein können, als auch durch externe Schallsignale hervorgerufen (evoziert) werden können. Sie entstehen in der Cochlea vermutlich durch aktive, vom ÄHZ-Antrieb erzeugte Schwingungen, die durch das ovale Fenster zum Mittelohr und weiter zum Trommelfell geleitet werden. Evozierte Emissionen werden u. a. zur Messung der Frequenzselektivität und Nichtlinearität der Cochlea verwendet, wobei generell eine

Übereinstimmung mit Schwingungsmessungen der BM beobachtet wird. Otoakustische Emissionen stellen eine wichtige Möglichkeit für nichtinvasive Messungen an der Cochlea des Menschen dar [110, 111].

2.3 Lautheit

Die subjektiv empfundene Lautstärke eines Schallsignals wird in der Psychoakustik als Lautheit N in der Einheit Sone angegeben. 1 Sone ist definiert als die Lautheit eines 1 kHz Sinustons mit 40 dB SPL [197]. Eine Möglichkeit zur Generierung einer Lautheitsskala ist die wiederholte Bestimmung des SPLs, der als doppelt bzw. halb so laut wie ein Referenzschall empfunden wird (Verhältnislautheit). Die zugeordnete Lautheit in Sone ergibt sich dann entsprechend mit dem Faktor 2 bzw. 0,5 aus der Lautheit des Referenzsignals. Eine Verdoppelung der Lautheit, die einer Zunahme von 3 dB entspricht, erfordert eine Erhöhung des SPLs um ca. 10 dB. Der Zusammenhang zwischen Lautheit und SPL weist daher eine Dynamikkompression auf, die in vergleichbarer Weise auch bei der BM-Schwingung festgestellt wird (vgl. Kapitel 2.2.2). So zeigen Messungen [34, 167] eine Proportionalität zwischen dem Quadrat der Basilarmembranschnelle und der Lautheit. Diese Meßergebnisse sind mit der Annahme vereinbar, daß die Lautheit proportional zur gesamten Entladungsrate aller Hörnervenfasern ist [55, 205], wenn eine quadratische IHZ-Kennlinie zugrunde gelegt wird.

Für die Entwicklung von Modellen der Lautheitswahrnehmung ist die Einführung der spezifischen Lautheit N' mit der Einheit Sone/Bark sinnvoll [197]. Diese Größe kann in Analogie zur Entsprechung der Lautheit mit der gesamten Entladungsrate als tonheitsspezifische Teillautheit aufgefaßt werden, die aufgrund der Entladungsrate auf einer einzelnen Faser verursacht wird. Die Lautheit ergibt sich aus der spezifischen Lautheit durch Integration über die Tonheit, entsprechend der oben genannten Summation der Entladungsraten aller Fasern des Hörnervs.

Ein zeitveränderlicher SPL führt zu einer entsprechenden Schwankung der Lautheit. Die Lautheitsänderung folgt aber den Änderungen des SPLs nicht beliebig schnell. So weist die Lautheitsänderung beim An- bzw. Abschalten eines Schallsignals Zeitkonstanten auf, die etwa der Vor- bzw. Nachverdeckung entsprechen [194].

Die Lautheit kann auch separat für ein Testsignal angegeben werden, das als solches neben einem überlagerten Maskierersignal wahrnehmbar ist. In diesem Fall bezieht sich die Lautheitsangabe nicht auf das insgesamt wahrgenommene Schallsignal, sondern nur auf das evtl. teilweise verdeckte Testsignal („partially masked loudness“) [84, 197]. Die Lautheit des Testsignals ist demzufolge null, wenn das Testsignal vom Maskierer vollständig verdeckt wird. Bei zunehmendem Testsignalpegel nähert sich seine Lautheit dann nach Überschreiten der Mithörschwelle der Lautheit des Testsignals ohne das überlagerte Maskierersignal an. Die pegelabhängige Lautheit eines partiell verdeckten Testsignals umfaßt daher sowohl die Mithörschwelle für das Testsignal als auch die Intensität der Testsignalwahrnehmung bei einem Pegel oberhalb der Mithörschwelle.

2.4 Psychoakustisch gemessene Wahrnehmungsschwellen

Eine psychoakustisch gemessene Wahrnehmungsschwelle berücksichtigt implizit die Eigenschaften des Gehörs als Ganzes. Demgegenüber ermöglichen physiologische Messungen von Reizschwellen beispielsweise mit Nervenfaserableitungen lediglich eine Beschreibung von Teilsystemen des Gehörs. Ein unmittelbarer Zusammenhang zwischen einer physiologischen Reizschwelle und einer Wahrnehmungsschwelle ist in der Regel nicht herzuleiten, weil von der nachfolgenden Verarbeitungsstufe weder genau bekannt ist, wie hoch ihre Leistungsfähigkeit bezüglich einer Testsignal detektion ist, noch welche Signalparameter darin ausgewertet werden. Qualitative Übereinstimmungen der physiologischen und psychoakustischen Schwellen lassen aber dennoch Rückschlüsse auf die physiologischen Mechanismen und Ursachen der Verdeckung zu (vgl. Kapitel 2.5). Auf physiologische Reizschwellen wird hier nicht weiter eingegangen, da sie für die Verifikation eines Wahrnehmungsmodells von untergeordneter Bedeutung sind.

Die psychoakustische Messung von Wahrnehmungsschwellen unterliegt dem Problem, daß eine unmittelbare Beobachtung der gesuchten Größe nicht möglich ist. Vielmehr stellt die jeweilige Testperson sowohl das Versuchsobjekt als auch den Beobachter dar, der im Experiment Angaben über die eigene Wahrnehmung macht. Daher ist das Meßergebnis mit einer Varianz behaftet, die nicht nur von der Detektionswahrscheinlichkeit des Stimulus abhängt, sondern auch vom Entscheidungskriterium, das der Beobachter verwendet. Um eine möglichst objektive (Bias-freie) Bewertung durch den Beobachter zu erzielen, können die Experimente so gestaltet werden, daß er beispielsweise aufgrund einer zufälligen Reihenfolge kein Wissen über die Art des Stimulus nutzen kann. Eine solche Messung führt zu einer psychometrischen Funktion, die beispielsweise die Wahrnehmungswahrscheinlichkeit des Stimulus in Abhängigkeit von der Stimulusintensität angibt (s. Bild 2.14). Die Wahrnehmungsschwelle wird in diesem Fall als die Intensität des Stimulus definiert, die mit einer vorgegebenen Wahrscheinlichkeit zu einer Wahrnehmung führt. In der Regel wird dabei eine 50%ige Wahrscheinlichkeit verwendet. Geeignete psychoakustische Meßverfahren sind so ausgelegt, daß die Varianz der Meßgröße minimiert wird, ohne die Versuchsdauer bzw. -anzahl gravierend zu vergrößern.

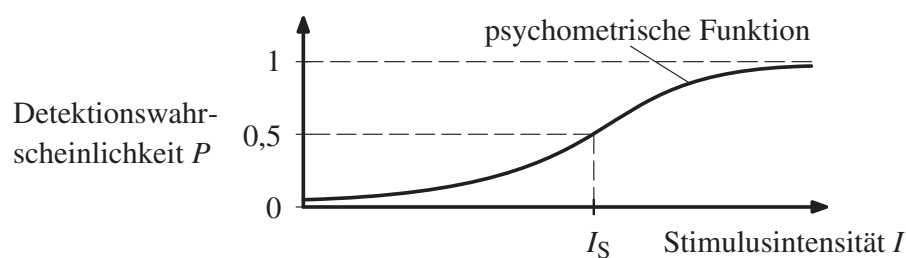


Bild 2.14 Bestimmung der Wahrnehmungsschwelle I_S anhand einer psychometrischen Funktion. Die Stimulusintensität führt an der Schwelle I_S zu einer Detektionswahrscheinlichkeit der Stimulation von 50%.

Für die Messung der Mithörschwelle kann im Experiment sequentiell das Maskierersignal im ersten Intervall und danach im zweiten Intervall das Maskierersignal mit überlagertem

Testsignal dargeboten werden. Das Maskierersignal wird dabei von der Testperson als Referenz herangezogen, mit der die Empfindung des Signals im zweiten Intervall verglichen wird. Die Empfindung des Referenzsignals ist daher für den Vergleich als „Muster“ im Gedächtnis vorhanden.

Das beschriebene Experiment ist zur Bestimmung der Schwelle für die Wahrnehmung eines beliebigen Unterschieds zwischen den beiden dargebotenen Signalen geeignet, der beispielsweise durch ein Verzerrungsprodukt von Maskierer- und Testsignal gegeben sein kann. Die Testperson könnte aber auch so instruiert werden, daß sie nur bei der Wahrnehmung des Testsignals selbst anspricht. Für dieses Experiment wäre ein Erlernen des Testsignals notwendig, das in darauf folgenden Versuchen im überlagerten Maskierersignal entdeckt werden muß. Für die Detektion eines Unterschieds wertet das Gehör die Änderung einer Wahrnehmungsgröße aus, wie zum Beispiel die Lautheit, Tonhöhe oder Rauigkeit.

In dieser Arbeit werden die Wahrnehmungsschwellen für Schallsignalunterschiede aufgrund eines überlagerten Testsignals als Mithörschwellen bezeichnet. Demgegenüber bezeichnet eine Unterschiedsschwelle die Wahrnehmungsschwelle bei der Änderung eines Schallsignalparameters, wie zum Beispiel Amplitude, Frequenz oder Modulationsindex. Während die Ruhehörschwellen und Mithörschwellen für das entstehende Quantisierungsgeräusch in der Audiocodierung benötigt werden, können die Unterschiedsschwellen zum Abgleich eines Gehörmodells zusätzlich herangezogen werden. Die Verdeckungseigenschaften bei binauralem Hören bleiben hier unberücksichtigt, da das Gehörmodell im Rahmen dieser Arbeit auf die monaurale Wahrnehmung beschränkt sein soll.

2.4.1 Ruhehörschwellen

Die Ruhehörschwelle für ein Testsignal ist frequenzabhängig und weist Unterschiede bezüglich des Schallsignalcharakters zum Beispiel für Sinustöne und Schmalbandrauschen auf [53]. Über viele Testpersonen gemittelte Werte sind in Bild 2.15 dargestellt. Individuell kann

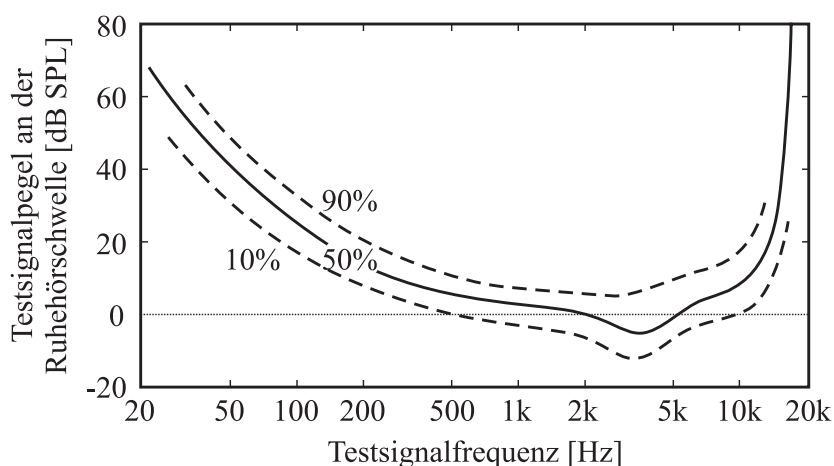


Bild 2.15 Ruhehörschwelle von 100 Personen mit normaler Hörfunktion für einen Sinuston. Die Prozentangaben geben den Anteil der Testpersonen an, deren Ruhehörschwelle unterhalb der jeweiligen Kurve liegt; dabei entspricht 50% dem Medianwert (Daten aus [198]).

die Ruhehörschwelle von Normalhörenden jedoch nennenswert davon abweichen. Als physiologische Ursache der Ruhehörschwelle wird ein gehörinternes Rauschen angesehen, das sich dem von einem Schallsignal hervorgerufenen Reizsignal überlagert und dieses verdecken kann. Als Rauschquelle kommt im Bereich tiefer Frequenzen das von der Blutzirkulation hervorgerufene Geräusch in Betracht [6, 176]. Über die gesamte Bandbreite wahrnehmbarer Frequenzen überlagert sich zusätzlich die Spontanaktivität auf den Fasern des Hörnervs wie ein gleichmäßiges „Grundrauschen“. Da die Rauschquellen im Innenohr angenommen werden, wirkt auf ein Testsignal vor der Überlagerung mit dem Rauschen der Frequenzgang des Außen- und Mittelohrs. Die Ruhehörschwelle weist daher beispielsweise auch ein Minimum bei 3 kHz auf, das von der Resonanz des Gehörgangs hervorgerufen wird (vgl. Bild 2.7).

2.4.2 Mithörschwellen

In diesem Kapitel werden die Verdeckungseigenschaften des Gehörs anhand einiger grundlegender Ergebnisse aus der Psychoakustik erläutert.

Bei der Messung von Mithörschwellen wird neben dem Testsignal ein Maskierersignal dargeboten, das die Wahrnehmbarkeit des Testsignals beeinflusst. In den klassischen Experimenten wurde die Verdeckung eines Sinustons als Testsignal von einem simultan vorhandenen maskierenden Sinuston oder Schmalbandrauschen untersucht, indem bei unterschiedlichen Testsignalfrequenzen der Testsignalpegel an der Mithörschwelle für einen festen Maskiererpegel gemessen wurde. Das in Bild 2.16 dargestellte Ergebnis zeigt die Anhebung der Wahrnehmungsschwelle gegenüber der Ruhehörschwelle des Testsignals. Die Verdeckungswirkung erstreckt sich deutlich über die Signalbandbreite des Maskierers hinaus und ist weiter zu höheren Testsignalfrequenzen ausgedehnt als zu tieferen. Außerdem ist die Wahrnehmungsschwelle bei einem maskierenden Schmalbandrauschen gegenüber einem maskierenden Sinuston im Bereich der Mittenfrequenz deutlich erhöht. Bei Testsignalfrequenzen deutlich oberhalb der Maskierfrequenz kehrt sich das Verhältnis jedoch wieder um.

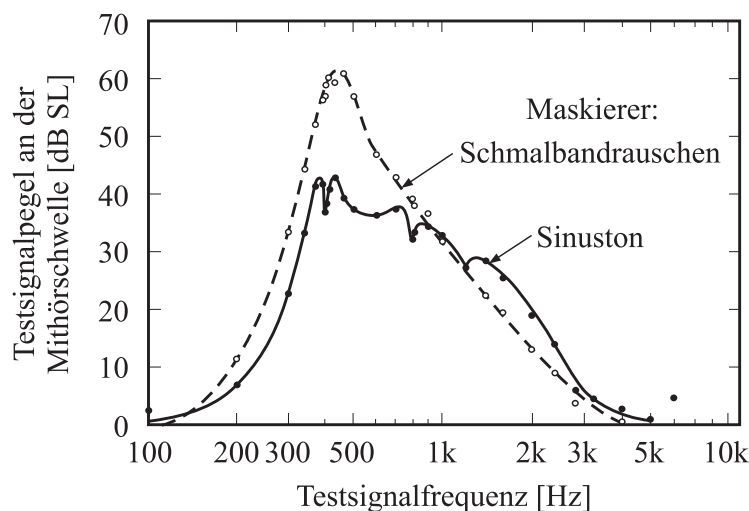


Bild 2.16 Mithörschwelle eines Sinustons, verdeckt durch ein Schmalbandrauschen (Bandbreite 90 Hz) bzw. einen Sinuston bei 400 Hz und 80 dB SPL. Die Verdeckung gibt hier den Pegel über der Ruhehörschwelle (SL) an. Daten aus [48].

Die Abnahme der Mithörschwelle mit zunehmender Testtonfrequenz oberhalb der Maskierfrequenz weist eine Steilheit auf, die vom Maskiererpegel abhängt. Beispielsweise führt

eine Erhöhung des Maskiererpegels um 10 dB zu einer Mithörschwellenzunahme um wesentlich mehr als 10 dB im Bereich von Testsignalfrequenzen oberhalb der Maskierfrequenz [54]. Dieser als „nonlinear growth of masking“ bezeichnete Effekt ist eine Folge der nichtlinearen Schallverarbeitung des Gehörs.

Die deutlich unterschiedliche Wirksamkeit der Verdeckung eines Sinustonmaskierers auf ein schmalbandiges Testrauschen im Vergleich zur Wirkung eines Schmalbandmaskierers auf einen Sinustestton wird als „asymmetry of masking“ bezeichnet [74]. Entsprechende Mithörschwellenmessungen sind u. a. in [74, 81, 84, 131, 186] veröffentlicht worden. Dabei wird die Abhängigkeit der Mithörschwelle vor allem vom Charakter des Maskierers bestimmt, wie die Ergebnisse für einen Sinuston und Schmalbandrauschen zeigen [131].

Verschiedene Studien wurde beobachtet, daß die Verdeckungswirkung eines Maskierers nicht nur von der spektralen Zusammensetzung sondern auch entscheidend vom zeitabhängigen Verlauf seiner Einhüllenden abhängt [33, 80, 81, 83, 114]. Schmalbandmaskierer mit einer Gaußschen Amplitudenverteilung erzeugen eine hohe Verdeckungswirkung für Testsignale in der Nähe des Maskierfrequenzbereiches. Demgegenüber erzeugen Rauschsignale mit gleicher Bandbreite und verringerter Einhüllendenschwankung eine reduzierte Verdeckungswirkung in diesem Frequenzbereich. Diese Eigenschaften besitzen beispielsweise die sog. „low-noise noise“-Signale [158], die aus vielen Sinussignalen mit konstantem Frequenzabstand synthetisiert werden, deren Phasen zur Erzielung einer möglichst „flachen“ zeitlichen Einhüllenden bzw. eines geringen Crest-Faktors optimiert sind. Die Verminderung der Einhüllendenschwankung führt zu einer Annäherung des Mithörschwellenverlaufs eines Schmalbandmaskierers an den eines Sinustons.

Die Verdeckungswirkung eines Rauschmaskierers mit konstanter Leistung in Abhängigkeit der Bandbreite auf einen im Rauschband zentrierten Testton zeigt eine näherungsweise konstante Mithörschwelle, solange die Maskiererbandbreite kleiner als die Filterbandbreite des Gehörs auf der gleichen Mittenfrequenz ist [27, 168]. Bei zunehmender Maskiererbandbreite verringert sich die Mithörschwelle etwa in dem Maß, wie sich die am Gehörfilterausgang erhaltene Signalleistung aufgrund der Bandpaßfilterung reduziert. Besteht das Testsignal ebenfalls aus einem Rauschsignal, wird eine verringerte Mithörschwelle beobachtet, wenn die Testsignalbandbreite größer als die Maskiererbandbreite ist [74, 191].

Die Mithörschwelle eines Testsignals erhöht sich in der Regel, wenn dem Maskierersignal ein zweites Maskierersignal hinzugefügt wird. Für schmalbandige Maskierer, wie schmalbandigem Gaußschen Rauschen oder Sinustönen, wird die Mithörschwelle der kombinierten Maskierer mit der Mithörschwelle bei einzelnen Maskierern in [20, 69, 121, 130, 147, 199] verglichen. Die Ergebnisse zeigen, daß die vom kombinierten Maskierer hervorgerufene Mithörschwelle um 10 bis 17 dB größer sein kann, als die aus der Intensitätssumme der Mithörschwellen von den beiden einzelnen Maskierern berechnete Mithörschwelle. Dieser als „nonlinear additivity of masking“ bezeichnete Effekt ist wirksam, wenn die beteiligten Maskierersignale nicht in derselben Frequenzgruppe liegen oder sich zeitlich nicht überlappen [36, 96].

Die Mithörschwelle des Sinustonmaskierers zeigt einen ungleichmäßigen Verlauf, der auf bestimmte Wahrnehmungsphänomene zurückzuführen ist. Bei Annäherung der Testsi-

gnalfrequenz an die Maskierfrequenz und deren ganzzahlige Vielfache entstehen hörbare Schwebungen. Testtonfrequenzen oberhalb des Maskierers führen aufgrund der Nichtlinearität des peripheren Gehörs zu hörbaren Verzerrungsprodukten. Insbesondere der kubische Differenzton mit einer Frequenz von $f_{d3} = 2f_M - f_T$ ist häufig wahrnehmbar, wenn dessen Frequenz unterhalb der Maskierfrequenz liegt [54].

Neben den oben erläuterten spektralen Verdeckungseffekten für quasi-stationäre Maskierer- und Testsignale werden zeitabhängige Verdeckungseffekte bei zeitveränderlichen Schallsignalen beobachtet, deren wichtigste Eigenschaften im folgenden beschrieben werden.

Für die Mithörschwellemessung mit einem zeitveränderlichen Maskierer wird in der Regel ein kurzer Testtonimpuls verwendet, bei dessen Entwurf ein geeigneter Kompromiß zwischen Bandbreite und Impulsdauer eingegangen werden muß. In Bild 2.17 ist beispielhaft die zeitabhängige Mithörschwelle für einen Testtonimpuls, maskiert durch ein breitbandiges Rauschen, qualitativ dargestellt. Der Maskierer wird zum Zeitpunkt t_0 für die Dauer T_M eingeschaltet. Der Testtonimpuls wird zum Zeitpunkt t überlagert. Bereits vor dem Einschalten des Maskierers ist die Wahrnehmungsschwelle für das Testsignal gegenüber der Ruhehörschwelle aufgrund der hier wirksamen Vorverdeckung erhöht [50, 51, 52]. Unmittelbar nach dem Einschalten des Maskierers ist bei mittleren Pegeln von 50 bis 60 dB SPL ein sog. „Overshoot-Effekt“ zu beobachten, während dessen eine gegenüber dem eingeschwungenen Zustand bei eingeschaltetem Maskierer deutlich erhöhte Mithörschwelle gemessen wird [113]. Der Overshoot-Effekt wächst mit zunehmender Maskiererbandbreite [50, 113]. Nach Abschalten des Maskierers fällt die Mithörschwelle aufgrund der hier wirksamen Nachverdeckung relativ langsam ab, bis sie nach ca. 200 ms in die Ruhehörschwelle übergeht [50, 51, 52, 109]. Im Bereich der Vor- und Nachverdeckung werden mit zunehmender Maskiererdauer erhöhte Mithörschwellen beobachtet [112, 154, 194, 187, 196]. Eine Zunahme des Maskiererpegels führt außerdem zu einer vergrößerten Steilheit der Vor- und Nachverdeckung [154, 187]. Aufgrund der kurzen Vorverdeckungsdauer sind Messungen der Vorverdeckung schwieriger durchzuführen und weisen eine wesentlich größere interindividuelle Variabilität auf als Nachverdeckungsmessungen.

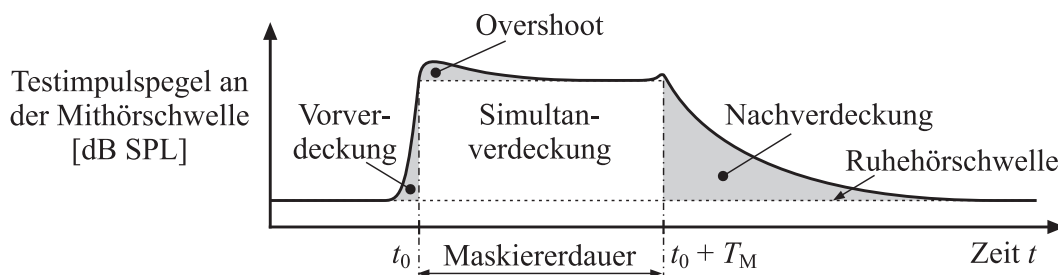


Bild 2.17 Qualitative Darstellung der zeitabhängigen Mithörschwelle eines Testsignalimpulses, verdeckt durch ein zeitlich begrenztes Breitbandrauschen mit konstantem Pegel. Der Testsignalimpuls wird zum Zeitpunkt t dargeboten. Bei der Messung der Vorverdeckung liegt dieser Zeitpunkt vor dem Maskierersignal, bei der Nachverdeckung nach Abschalten des Maskierers.

Die Wahrnehmungsschwelle hängt von der Testsignaldauer ab, wenn diese kleiner als etwa 200 ms ist. Der entsprechende Zusammenhang wird durch Messungen der Simultanverdeckung oder der Ruhehörschwelle bestimmt. Für die Ruhehörschwelle ergibt sich ein asymptotischer Anstieg mit abnehmender Testsignaldauer von 10 dB/Dekade [197].

Weitere Verdeckungseffekte können anhand von zeitlich und spektral veränderlichen Maskierersignalen nachgewiesen werden. Ein einfaches Beispiel ist die Verdeckung durch einen frequenzmodulierten Sinuston, der zeit- und frequenzabhängige Mithörschwellen verursacht. Bei gleicher Momentanfrequenz des Maskierers ist dabei eine Abhängigkeit der Verdeckungswirkung von der Modulationsphase (zunehmende oder abnehmende Maskierfrequenz) gegeben, die auf die Nachverdeckung zurückzuführen ist [197].

Bei breitbandigen Maskierern hängt die Verdeckungswirkung auch vom zeitabhängigen Verlauf der Einhüllenden ab. Eine tieffrequente Amplitudenmodulation eines breitbandigen Maskierers hat eine deutlich reduzierte Mithörschwelle gegenüber dem unmodulierten Maskierer zur Folge. Es wird daher vermutet, daß das Gehör einen Vergleich der Einhüllenden in unterschiedlichen Frequenzbändern durchführt. Aufgrund dieses Vergleichs kann ein in einem Band überlagerter Sinustestton dann leichter detektiert werden. Wegen der gleichen Modulation in mehreren Bändern wird dieser Verdeckungseffekt als „co-modulation masking release“ (CMR) bezeichnet [75, 168].

2.4.3 Unterschiedsschwellen

Unterschiedsschwellen beschreiben die Empfindlichkeit des Gehörs für Schalländerungen, die durch Variation von physikalischen Schallparametern hervorgerufen werden. Im Gegensatz zu den Mithörschwellen wird dabei kein überlagerter Maskierer verwendet. In der Literatur werden Schwellenmessungen für die Änderung von Schallsignalparametern, wie beispielsweise Intensität oder Frequenz beschrieben. Üblicherweise werden Unterschiedsschwellen im englischen Sprachraum als „just-noticeable-difference (JND)“ oder „difference limen (DL)“ bezeichnet.

Die Intensitätsunterschiedsschwelle entspricht der gerade wahrnehmbaren Lautheitsänderung und beträgt im Mittel etwa 1 dB SPL. Sie verringert sich mit zunehmendem SPL [70]. Die Frequenzunterschiedsschwelle liegt näherungsweise bei einer Frequenzänderung von 3,6 Hz für einen Sinuston unterhalb 500 Hz. Bei Frequenzen oberhalb von 500 Hz wird die Schwelle bei einer Frequenzänderung des Sinustons von etwa 0,7% erreicht [197].

Die Intensitäts- und Frequenzunterschiedsschwelle wird in der Regel mit Hilfe von Schallsignalen mit sinusförmiger Modulation der Amplitude bzw. Frequenz gemessen. Aufgrund des begrenzten zeitlichen Auflösungsvermögens des Gehörs sind diese Schwellen von der Modulationsfrequenz abhängig. Bild 2.18 zeigt den Modulationsgrad an der Wahrnehmungsschwelle einer Amplitudenmodulation (AM) von breitbandigem Rauschen in Abhängigkeit von der Modulationsfrequenz. Eine solche Darstellung der Modulationsschwelle wird auch als „zeitliche Modulationsübertragungsfunktion“ („temporal modulation transfer function (TMTF)“) bezeichnet [8, 185]. Die TMTF hat einen Tiefpaßcharakter, dessen 3 dB Grenzfrequenz für das breitbandige Rauschen bei etwa 50 Hz liegt. Die maximale Empfindlichkeit für eine AM erreicht das Gehör bei einer Modulationsfrequenz von etwa 4 bis 8 Hz.

Analog zur Intensitätsunterschiedsschwelle kann eine Modulationsunterschiedsschwelle für eine gerade wahrnehmbare Änderung des Modulationsgrads angegeben werden.

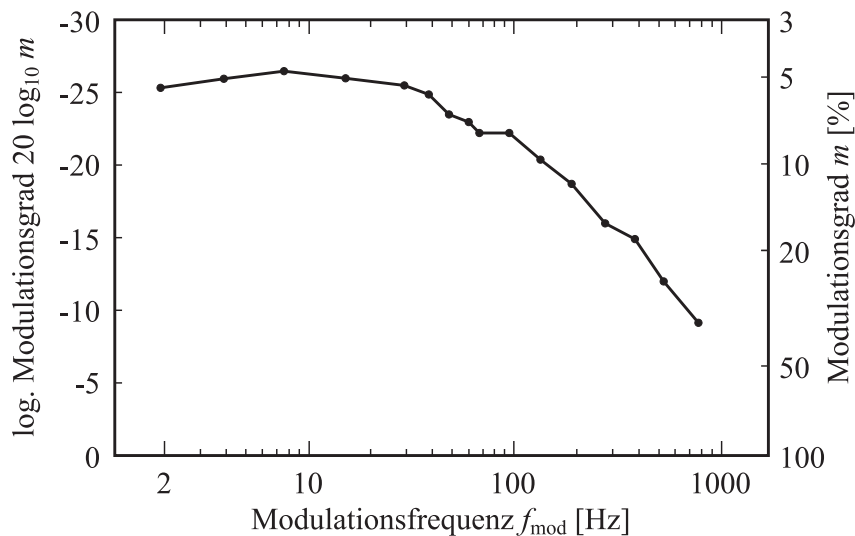


Bild 2.18 Modulationsgrad m für eine gerade wahrnehmbare sinusförmige Amplitudenmodulation eines breitbandigen Rauschens mit Modulationsfrequenz f_{mod} (Zeitliche Modulationsübertragungsfunktion TMTF). Daten aus [8].

Eine AM mit gegebener Modulationsfrequenz kann durch eine zweite überlagerte AM mit einer anderen Modulationsfrequenz verdeckt werden. Die zweite AM kann daher als maskierende Modulation und die erste AM als Testmodulation aufgefaßt werden. Analog zu Mithörschwellenmessungen treten bei Messungen der AM-Wahrnehmungsschwelle an die Stelle der Maskierer- und Testsignalfrequenzen die maskierende Modulationsfrequenz und die Testmodulationsfrequenz. Der Modulationsgrad der Testmodulation an der Wahrnehmungsschwelle wird dann für unterschiedliche Testmodulationsfrequenzen gemessen und im Modulationsfrequenzbereich analog zu den frequenzabhängigen Mithörschwellen angegeben [38, 41, 49].

2.5 Physiologische Ursachen der Verdeckung

Die Mechanismen der Schallverarbeitung des Gehörs, die einen Einfluß auf die Wahrnehmungsschwelle haben, können zum Teil durch physiologische Messungen der gehörinternen Signalrepräsentation untersucht werden. Wegen der relativ guten Zugänglichkeit des Hörnervs existieren zahlreiche Untersuchungen der Verdeckungswirkungen auf dieser Verarbeitungsstufe, z. B. [174]. Im Hörnerv sind sämtliche Informationen codiert, die dem zentralen Nervensystem für die Bildung der Hörempfindung zur Verfügung stehen. Als wesentlicher Informationsträger wird in dieser Arbeit die Entladungsrate der einzelnen Fasern angesehen, während die in den einzelnen Zeitpunkten der Aktionspotentiale enthaltene Information als sekundär eingestuft wird [44]. In einigen anderen Arbeiten wird demgegenüber angenommen, daß eine wesentliche Information in der reizsynchronen Entladungsrate für Tonfrequenzen unterhalb von 2 kHz vorhanden ist („phase locking“ [124]). Mit der Annahme, daß

die Reizunterschiedsschwelle der nächst höheren Verarbeitungsstufe durch eine Änderung der Entladungsrate um einen bestimmten Betrag erreicht wird, kann beispielsweise der SPL eines Testsignals an der Reizschwelle gemessen werden. In Bild 2.19 sind exemplarisch die Entladungsraten einer Nervenfasers mit der CF des Testsignals dargestellt, die sich für ein Testsignal im Pegelbereich von 0 bis 80 dB SPL sowohl allein als auch mit überlagertem Maskierer einstellen. Die pegelabhängige Entladungsrates des Testsignals allein ist zum Vergleich in allen Diagrammen in Bild 2.19 (A–D) identisch gepunktet eingezeichnet und entspricht dem bereits in Bild 2.11 gezeigten Zusammenhang. Die Wirkung unterschiedlicher überlagerter Maskierersignale auf die Entladungsrates ist durchgezogen eingezeichnet.

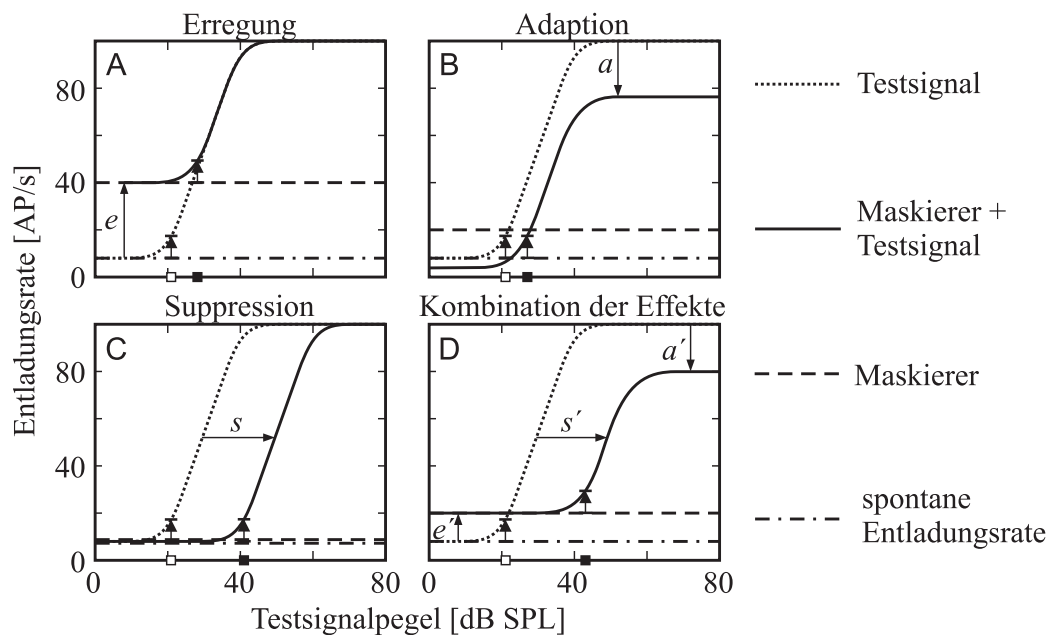


Bild 2.19 Wirkung grundlegender physiologischer Verdeckungsmechanismen auf die Entladungsrates einer Hörnervenfasers. Jedes Diagramm stellt den idealisierten Verlauf der Entladungsrates in Abhängigkeit vom Schalldruckpegel des Testsignals allein (*gepunktet*) und mit überlagertem Maskierer (*durchgezogen*) für verschiedene Verdeckungsmechanismen dar. Die Entladungsrates, die der Maskierer mit konstantem Pegel hervorruft, ist zum Vergleich *gestrichelt* eingezeichnet. Die spontane Entladungsrates der Fasers ist *strichpunktiert*. Die Detektionsschwelle für das Testsignal allein (\square) und mit Maskierer (\blacksquare) folgt aus dem Schwellenkriterium einer konstanten Erhöhung der Entladungsrates durch das Testsignal. Weitere Erklärungen s. Text. (In Anlehnung an [43, 44]).

In Diagramm A beträgt die vom Maskierer allein erzeugte Entladungsrates 40 AP/s. Aufgrund dieser um e gegenüber der spontanen Rate erhöhten Entladungsrates ist bei der Überlagerung eines Testsignals mit kleinem Pegel keine Zunahme zu verzeichnen. Dieser als „*Erregung*“ bezeichnete Effekt führt daher zu einer Erhöhung der Detektionsschwelle für das Testsignal (vgl. \blacksquare bzw. \square in Bild 2.19A). Diagramm B gibt die Situation zu einem Zeitpunkt bei abklingendem oder abgeschaltetem Maskierer wieder, in dem die IHZ mit der ange-

schlossenen Nervenfasern aufgrund der langsamen *Adaption* noch nicht an den momentanen Maskiererpegel angepaßt ist (vgl. Bild 2.13). Die eingezeichnete, vom Maskierer hervorgerufene Entladungsrate wurde vor dem Abschalten des Maskierers erzeugt. Die relativ langsame *Adaption* hat eine um a reduzierte Entladungsrate als Reaktion auf das Testsignal und eine entsprechend erhöhte Detektionsschwelle zur Folge. Diagramm C zeigt die aufgrund von *Suppression* reduzierte Entladungsrate. *Suppression* entsteht durch die vom Maskierersignal hervorgerufene Aussteuerung der nichtlinearen ÄHZ-Antriebe, deren Verstärkungsfaktor sich aufgrund der größeren Sättigung vermindert. Auch für das Testsignal ist daher nur noch eine verminderte Verstärkung wirksam, so daß die Entladungsrate an der Nervenfasern mit der CF des Testsignals entsprechend abnimmt. Um die gleiche Entladungsrate zu erreichen, muß daher der Testsignalpegel bei *Suppression* um s erhöht werden. Da die *Suppression* auch bei Maskierern mit Frequenzen deutlich unterhalb des Testsignals wirksam ist, verursacht der Maskierer allein u. U. keine erhöhte Entladungsrate in der Nervenfasern mit der CF des Testsignals. Die in Diagramm D gezeigte Kombination der drei genannten Effekte verursacht die größte Detektionsschwellenzunahme und dementsprechend die größte Verdeckung. In dieser Kombination beinhaltet das Maskierersignal eine abklingende Komponente nahe der Testsignalfrequenz wie in B, sowie eine tieferfrequente, supprimierende Komponente wie in C.

Die Verminderung der Entladungsrate auf einzelnen Fasern des Hörnervs aufgrund von *Adaption* wurde in [79, 184] mit psychoakustischen Mithörschwellen bei der Nachverdeckung verglichen. In [79] wird eine Übereinstimmung der gemessenen Zeitkonstanten festgestellt, während in [184] die Ursache der Nachverdeckung an anderer Stelle vermutet wird. Beide Untersuchungen zeigen jedoch, daß in der Synapse der IHZ bereits ein wesentlicher Anteil der Nachverdeckung nachweisbar ist.

Bild 2.19 verdeutlicht, daß wesentliche Verdeckungsphänomene mit der Schallverarbeitung in der Cochlea inklusive der Haarzellen erklärt werden können.

2.6 Konzepte zur Vorhersage auditiver Wahrnehmungsschwellen

Die im vorangegangenen Kapitel 2.5 erläuterten physiologischen Ursachen der Verdeckung erfassen nur den Einfluß eines Teilsystems des Gehörs auf die Wahrnehmungsschwellen. Demgegenüber werden psychoakustisch gemessene Wahrnehmungsschwellen vom gesamten Gehörapparat bestimmt und können darüber hinaus psychologischen Einflüssen, beispielsweise der Konzentration, unterliegen. Die zur Messung verwendeten psychoakustischen Verfahren basieren auf unterschiedlichen Theorien zur Beschreibung der Detektion eines Testsignals durch eine Testperson. Die in den Theorien getroffenen Annahmen über verschiedene Einflußfaktoren bei der Messung von Wahrnehmungsschwellen sind aber auch für die Konzeption einer Wahrnehmungsschwellenmodellierung von Bedeutung.

Das „Ja-Nein-Verfahren“ („forced-choice“) ist eine einfache Meßprozedur zur Bestimmung einer Wahrnehmungsschwelle, bei dem das Maskierersignal (Ereignis R) oder das Maskierersignal mit überlagertem Testsignal (Ereignis T) nach zufallsgesteuerter Aus-

wahl in einem Intervall präsentiert wird. Die Testperson wird instruiert mit „ja“ zu antworten (Entscheidung D_T), wenn sie das Testsignal wahrgenommen hat, sonst mit „nein“ (Entscheidung D_R). Wird in 50% der Intervalle das Maskierersignal allein präsentiert, ergibt sich eine psychometrische Funktion wie in Bild 2.14. Die Wahrnehmungsschwelle wird bei demjenigen Testsignalpegel angenommen, bei dem das Testsignal mit einer Wahrscheinlichkeit von 50% wahrgenommen wird. Gemäß der „high-threshold“-Theorie [126] wird diese Detektionswahrscheinlichkeit $P_{HT}(D_T)$ anhand der gemessenen bedingten Wahrscheinlichkeiten für die Detektion eines Testsignals bei vorhandenem Testsignal $P(D_T|T)$ sowie bei nicht vorhandenem Testsignal $P(D_T|R)$ berechnet. Dabei wird angenommen, daß ein Teil der richtig detektierten Testsignale zufällig, bzw. durch Raten zustande gekommen ist. Dieser Anteil wird so groß angenommen, wie der Anteil falsch detektierte Testsignale in Intervallen ohne Testsignal. Die Detektionswahrscheinlichkeit wird daher um die Wahrscheinlichkeit zufällig detektierte Testsignale vermindert und entsprechend der angenommenen Gesamtzahl richtig bewerteter Intervalle mit und ohne Testsignal gemäß (2.2) normiert.

$$P_{HT}(D_T) = \frac{P(D_T|T) - P(D_T|R)}{1 - P(D_T|R)}. \quad (2.2)$$

Diese Theorie erweitert klassische Meßmethoden um die Berücksichtigung der Wahrscheinlichkeit von Fehldetektionen bei der Wahrnehmungsschwellenmessung. Sie beruht jedoch noch auf der Vorstellung, daß ein zunächst nicht wahrnehmbares Testsignal beim Überschreiten eines Schwellenwertes abrupt wahrnehmbar wird. Demzufolge werden Detektionsfehler psychologischen Ursachen zugeschrieben.

Einen umfassenderen Rahmen für die quantitative Behandlung des Problems einer Testsignal-detektion stellt die Signaldetektionstheorie [71, 126] zur Verfügung, deren Ergebnisse in sehr guter Übereinstimmung mit psychoakustischen Messungen stehen. Im Gegensatz zur „high-threshold“-Theorie wird keine abrupte sondern eine kontinuierliche Erhöhung der Detektionswahrscheinlichkeit bei einer Überschreitung der Schwelle durch das Testsignal vorausgesetzt (vgl. Bild 2.14). Für die Anwendung der Signaldetektionstheorie auf eine psychoakustische Wahrnehmungsschwellenmessung wird hier ein einfaches Signalmodell zugrunde gelegt, an dem die verschiedenen Einflußgrößen erläutert werden sollen.

Es wird angenommen, daß die Detektierbarkeit eines Testsignals durch einen fiktiven Beobachter, der im zentralen Gehör realisiert sein soll, gemäß Bild 2.20 im wesentlichen durch zwei Störquellen beeinträchtigt wird. Einerseits ist dem Testsignal selbst im erweiterten Sinne ein gehöreexternes Rauschen überlagert, das zum Beispiel als Grundrauschen vorhanden ist oder bei dem Experiment als Maskierersignal $m(t)$ überlagert wird. Andererseits sind im Innenohr akustische Störsignale vorhanden, und es entstehen Störungen aufgrund der begrenzten Auflösung bei der Umsetzung des Schallsignals, das als Folgen von APs auf den Hörnervenfasern codiert wird [6, 176]. Diese internen Störungen werden als additives internes Rauschen $n(t)$ berücksichtigt. Dem zentralen Gehör steht dementsprechend nur das „verrauschte“ Testsignal $q(t)$ für die Detektion zur Verfügung.

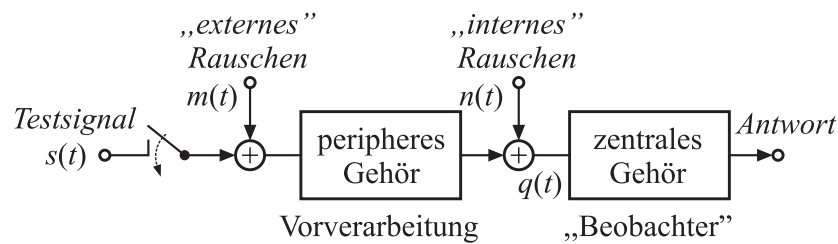


Bild 2.20 Konzept eines Wahrnehmungsmodells, bei dem die Detektionswahrscheinlichkeit des Testsignals durch den Beobachter von additiv überlagertem Rauschen beeinträchtigt wird.

Für das oben genannte Ja-Nein-Verfahren kann der Beobachter durch eine Verarbeitungseinheit für das beobachtete Signal $q(t)$ nachgebildet werden, die den Meßwert Ψ nach dem Beobachtungszeitraum ausgibt und einen Schwellwertentscheider mit dem Schwellwert Ψ_S (s. Bild 2.21). Die ausgegebene Entscheidung über das Vorhandensein eines Testsignals lautet $D = D_T \equiv$ „ja“, wenn $\Psi > \Psi_S$ ist, sonst $D = D_R \equiv$ „nein“.

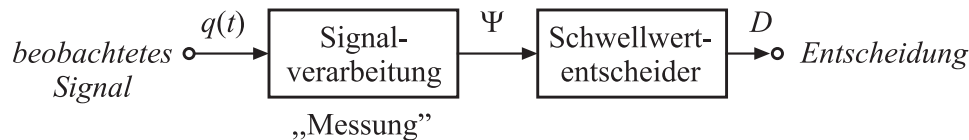


Bild 2.21 Modell des Beobachters für ein Ja-Nein-Verfahren.

In der Regel kann das Signal $q(t)$ als gehörinterne Größe nicht gemessen, sondern allenfalls durch ein Modell des peripheren Gehörs approximiert werden.

Ein im Sinne der Fehlerwahrscheinlichkeit optimaler Beobachter kann entworfen werden, wenn die Statistik des beobachteten Signals $q(t) = q_T(t)$ bei Einspeisung des Testsignals und des beobachteten Signals $q(t) = q_R(t)$ ohne Einspeisung des Testsignals zunächst als bekannt vorausgesetzt wird. Dabei wird $q_R(t)$ als „Referenzsignal“ bezeichnet. Der optimale Beobachter ist durch den Algorithmus der Signalverarbeitung zur Generierung der Meßgröße Ψ und den Schwellwert Ψ_S bestimmt.

Für den optimalen Beobachter kann eine theoretische Wahrnehmungsschwelle bestimmt werden, die Aufschluß über die Abhängigkeit der Wahrnehmungsschwelle von den Parametern der Signale $q_R(t)$ und $q_T(t)$ gibt. Dieser Zusammenhang soll anhand eines einfachen Beispiels verdeutlicht werden. Dazu sei vorausgesetzt, daß $q_R(t)$ und $q_T(t)$ unkorrelierte Rauschsignale mit einer Leistungsdichte von n_R bzw. n_T sind. Die Bandbreite B sowie die Dauer T_T der Signale seien identisch. Die Überlagerung des Testsignals $s(t)$ soll also ausschließlich zu einer Änderung der Signalleistungsdichte führen. Zur Detektion des Testsignals bestimmt ein optimaler Beobachter als Meßgröße Ψ die Energie des Signals $q(t)$. Der Beobachter wird daher als „Energiedetektor“ bezeichnet. Die Schwelle Ψ_S wird nun so bestimmt, daß die Fehlerwahrscheinlichkeit minimiert wird. Dies führt auf den sog. „Maximum-Likelihood-Entscheider“ [126]. Dazu werden die Wahrscheinlichkeitsverteilungsdichten der Energien ohne Testsignal $p(\Psi_R)$ bzw. mit überlagertem Testsignal $p(\Psi_T)$ ausgewertet. In Bild 2.22 sind die Verteilungsdichten als Gaußverteilungen mit gleicher Standard-

abweichung σ und unterschiedlichen Mittelwerten $\bar{\Psi}_R$ und $\bar{\Psi}_T$ angenommen worden. Ein Fehler tritt auf, wenn $\Psi_R > \Psi_S$ ist und kein Testsignal vorhanden war (Ereignis R), oder wenn $\Psi_T \leq \Psi_S$ ist und ein Testsignal vorhanden war (Ereignis T). Die entsprechenden bedingten Wahrscheinlichkeiten werden als $P(D_T|R)$ bzw. $P(D_R|T)$ bezeichnet. Mit der Auftretenswahrscheinlichkeit des Testsignals $P(T)$ ergibt sich die Fehlerwahrscheinlichkeit P_F aus

$$P_F = P(T) \int_{-\infty}^{\Psi_S} p(\Psi_T) d\Psi_T + [1 - P(T)] \int_{\Psi_S}^{\infty} p(\Psi_R) d\Psi_R, \quad (2.3)$$

mit
$$p(\Psi_T) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{\Psi_T - \bar{\Psi}_T}{\sigma}\right]^2\right) \quad (2.4)$$

und
$$p(\Psi_R) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{\Psi_R - \bar{\Psi}_R}{\sigma}\right]^2\right). \quad (2.5)$$

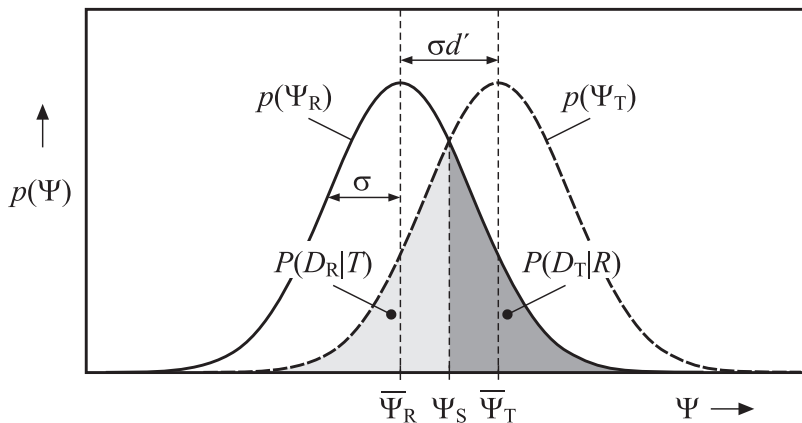


Bild 2.22 Gaußsche Wahrscheinlichkeitsdichtefunktionen der Meßgröße Ψ_R und Ψ_T mit den Mittelwerten $\bar{\Psi}_R$ bzw. $\bar{\Psi}_T$ und der Standardabweichung σ . Der Schwellwertentscheider verwendet die Entscheidungsschwelle Ψ_S . Der gemeinsame Flächeninhalt der beiden grau dargestellten Flächen entspricht der Fehlerwahrscheinlichkeit.

In der Signaldetektionstheorie wird die Sensitivität d' als normierte Größe eingeführt, die ein Maß für die Unterscheidbarkeit der beiden zugrunde liegenden Signale gemäß (2.6) darstellt.

$$d' = \frac{\bar{\Psi}_T - \bar{\Psi}_R}{\sigma} \quad (2.6)$$

Die Wahrnehmungsschwelle ist definiert als derjenige Testsignalpegel, bei dem eine Sensitivität von $d' = 1$ erreicht wird. Die Mittelwerte der vorausgesetzten Verteilungsdichtefunktionen unterscheiden sich dann genau um die Standardabweichung σ . Die Fehlerwahrscheinlichkeit an der Wahrnehmungsschwelle folgt aus (2.3) und beträgt $P_F = 31\%$ für

$P(T) = 50\%$. Dieser Wert kann mit der „high-threshold“-Theorie verglichen werden, indem eine entsprechende Korrektur nach (2.2) für das Ja-Nein-Verfahren durchgeführt wird. Mit $P(D_T|T) = 69\%$ und $P(D_T|R) = 31\%$ folgt $P_{HT}(D_T) = 55\%$, also eine geringe Abweichung der Schwellenkriterien beider Theorien.

Die Detektionsschwelle entspricht bei den hier gewählten Signalen der Intensitätsunterschiedsschwelle. Sie ist erwartungsgemäß nur von der Standardabweichung und den Mittelwerten der Wahrscheinlichkeitsdichtefunktionen nach (2.6) abhängig. Anstelle der Mittelwerte und Varianz der gemessenen Energie kann die Sensitivität auch unmittelbar in Abhängigkeit von der Rauschleistungsdichte bzw. vom „Signalenergie-Rauschleistungsdichte-Abstand“ E_T/n_R angegeben werden [70, 71]:

$$d' = \sqrt{BT_T} \frac{n_T}{n_R}. \quad (2.7)$$

Die Testsignalenergie beträgt dabei $E_T = BT_T n_T$. Durch den Energiedetektor in diesem Beispiel wird das Vorwissen über die zu erwartende Änderung des beobachteten Signals durch Überlagerung des Testsignals optimal ausgenutzt. Diese Überlegung verdeutlicht, daß die Sensitivität nicht nur von der Signalenergie sondern auch vom Vorwissen des Beobachters abhängt. Falls der Beobachter das Testsignal exakt kennt, besteht der optimale Beobachter aus einem „matched filter“ bzw. einem Korrelator, der die Kreuzkorrelation von dem beobachteten Signal und dem bekannten Testsignal bestimmt. Die Sensitivität dieses Beobachters ergibt sich mit der Testsignalenergie E_T zu [70, 71]

$$d' = \sqrt{2 \frac{E_T}{n_R}}. \quad (2.8)$$

Mit Hilfe der Signaldetektionstheorie kann die Sensitivität d' des Beobachters auch bestimmt werden, wenn das Maximum-Likelihood-Kriterium nicht verwendet wird. Eine Testperson könnte beispielsweise eine höhere Entscheidungsschwelle Ψ_S verwenden, wenn sie eine möglichst sichere Entscheidung für das Vorhandensein eines Testsignals treffen möchte, um die bedingte Fehlerwahrscheinlichkeit $P(D_T|R)$ zu vermindern. Dadurch wird jedoch die Fehlerwahrscheinlichkeit $P(D_R|T)$ erhöht, wie aus Bild 2.23 hervorgeht. Die verwendete Entscheidungsschwelle Ψ_S wird durch den Wert $\beta = a/b$ charakterisiert, der mit dem Likelihood-Verhältnis $l(\Psi) = p(\Psi_T)/p(\Psi_R)$ an der Stelle der Entscheidungsschwelle Ψ_S übereinstimmt. Das Maximum-Likelihood-Kriterium entspricht einem Wert von $\beta = 1$.

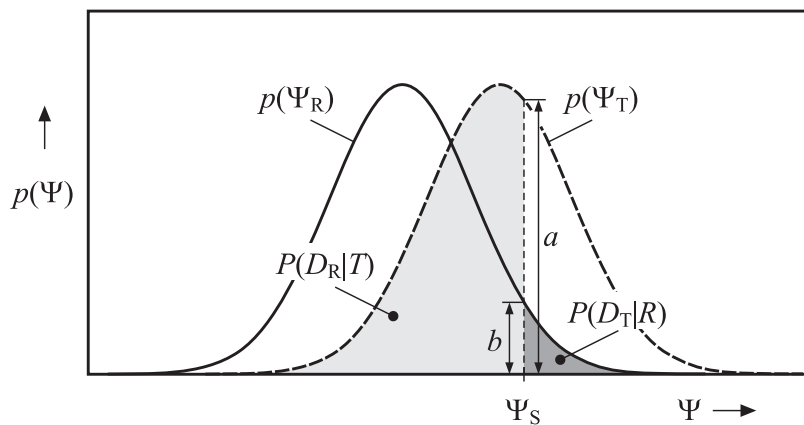


Bild 2.23 Gaußsche Wahrscheinlichkeitsdichtefunktionen der Meßgröße Ψ_R und Ψ_T . Der Schwellwertentscheider verwendet die Entscheidungsschwelle Ψ_S .

Durch eine geeignete Instruktion der Testperson können die Fehlerwahrscheinlichkeiten für Entscheidungskriterien entsprechend unterschiedlicher Werte von β gemessen werden. Die Meßergebnisse werden als eine sog. „Receiver-Operating Characteristic (ROC)“, wie in Bild 2.24, dargestellt. Dabei wird die bedingte Wahrscheinlichkeit $P(D_T|T)$ über der bedingten Wahrscheinlichkeit $P(D_T|R)$ aufgetragen. Es kann gezeigt werden, daß die Steigung der ROC identisch zum Likelihood-Verhältnis $l(\Psi) = \beta$ an der Entscheidungsschwelle Ψ_S ist [71].

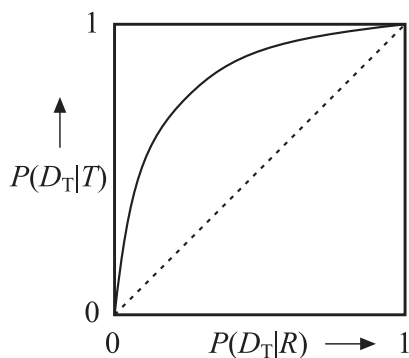


Bild 2.24 Darstellung der Wahrscheinlichkeit für eine korrekte Testsignaldetektion über der Wahrscheinlichkeit für eine fehlerhafte Detektion in einer sog. „Empfängercharakteristik“ (Receiver-Operating Characteristic ROC).

Die ROC kann für die vorgegebenen Verteilungsdichten nach (2.4) und (2.5) für den Parameter d' berechnet werden. Umgekehrt ist daher auch die Bestimmung von d' aus der ROC möglich. Eine gemessene ROC erlaubt daher eine unabhängige Bestimmung der Empfindlichkeit des Beobachters (Sensitivität d') von dem gewählten Entscheidungskriterium (Likelihood-Verhältnis) und damit auch die Trennung sensorischer und psychologischer Auswirkungen auf die Messung von Wahrnehmungsschwellen.

Die meßtechnische Bestimmung der ROC ist relativ aufwendig und kann durch einfachere Verfahren ersetzt werden, wenn nur die Sensitivität bzw. die Wahrnehmungsschwelle bestimmt werden soll. Ein geeignetes Verfahren ist die „two-alternative-forced-choice“ (2AFC) Methode, bei dem die zu bewertenden Schallsignale in jeweils zwei Intervallen präsentiert werden, wobei ein Intervall das Testsignal enthält. Die Testperson muß dabei jeweils entscheiden, in welchem der beiden Intervalle das Testsignal vorhanden ist. Die Theorie geht

davon aus, daß der Beobachter in diesem Fall einen direkten Vergleich der Likelihood-Verhältnisse durchführt und sich für das Testintervall entscheidet, in dem das größere Verhältnis beobachtet wurde. Der Beobachter bildet hier keine feste Entscheidungsschwelle Ψ_S bzw. kein festes Likelihood-Verhältnis β , sondern führt den Vergleich in jedem Intervallpaar unabhängig durch. Die Wahrscheinlichkeit für die korrekte Erkennung des Testsignalintervalls beträgt 76% für die vorausgesetzten Gaußverteilungen und für eine Sensitivität von $d' = 1$ an der Wahrnehmungsschwelle [126] und kann unabhängig von der ROC gemessen werden.

Folgende wesentliche Zusammenhänge zwischen Signalparametern und der Sensitivität eines Beobachters ergeben sich aus der Signaldetektionstheorie. Die Sensitivität des Beobachters nimmt zu, wenn der Signal-Rauschleistungsdichte-Abstand zunimmt oder die Varianz des Störsignals abnimmt. Die Sensitivität wächst außerdem mit zunehmendem Vorwissen des Beobachters über das Testsignal. Für die Konzeption eines Modells für eine Wahrnehmungsschwelennachbildung sind neben diesen Zusammenhängen auch die Annahmen bezüglich der Verdeckungsmechanismen von internem und externem Rauschen von Bedeutung und sollten in geeigneter Weise berücksichtigt werden. Die Theorie erlaubt andererseits auch Vergleiche zwischen der Sensitivität eines optimalen Beobachters und der Sensitivität des Gehörs oder anderen Wahrnehmungsmodellen.

Im folgenden wird dargelegt, daß mit dem Schwellenfunktionsschema bzw. durch die Anwendung des Weberschen Gesetzes nur ein Teilaspekt der Testsignaldetektion nachgebildet werden kann.

Das Schwellenfunktionsschema [198] beruht auf der Annahme, daß eine Erregungszunahme um 1 dB an einem beliebigen Ort in der Cochlea gerade wahrnehmbar ist. Die Erregung wird dabei durch ein lineares Modell berechnet, das vereinfachte Frequenzgruppenfilter nachbildet. Der mit diesem Modell realisierte Beobachter ist ein Energiedetektor, der lediglich die Signalenergie am Ausgang jedes Frequenzgruppenfilters auswertet. Dieses Wahrnehmungsmodell berücksichtigt daher weder eine Abhängigkeit von der Störsignalvarianz, noch von einem möglicherweise vorhandenen Vorwissen des Beobachters über das Testsignal.

Das Webersche Gesetz stellt eine Abhängigkeit zwischen der Intensitätsunterschiedsschwelle, angegeben als Schalldruckänderung Δp , und dem effektiven Schalldruck p eines Referenzsignals gemäß (2.9) her.

$$\frac{\Delta p}{p} = \text{konst.} \quad (2.9)$$

Eine Interpretation dieses Zusammenhangs im Sinne der Signaldetektionstheorie führt wie bei dem Schwellenfunktionsschema auf einen Energiedetektor. Jedoch entfällt hier die Vorverarbeitung des Schallsignals durch Frequenzgruppenfilter. Beide Modelle sagen voraus, daß sich die Empfindlichkeit proportional zur Maskierer- bzw. Störsignalleistung verhält. Dies steht im Einklang mit der Signaldetektionstheorie, die eine Proportionalität der Sensitivität mit dem Signal-Rauschleistungsdichte-Abstand nach (2.7) voraussagt.

3 Das psychophysiologische Gehörmodell

Eine präzise Nachbildung von Wahrnehmungsschwellen beliebiger Audiosignale soll vor allem durch eine möglichst weitgehende Nachbildung des Gehörs auf physiologischer Ebene erreicht werden. Insbesondere der Einfluß der nichtlinearen Schallverarbeitung des Gehörs ist als wesentliche Komponente für die Verdeckung zu berücksichtigen. Verbleibende Teile des Gehörs, wie beispielsweise weite Bereiche der neuronalen Verarbeitung, deren Physiologie nicht ausreichend bekannt ist, sollen durch psychoakustisch motivierte Teilmodelle nachgebildet werden.

Das psychophysiologische Gehörmodell soll nur die zur Nachbildung von Wahrnehmungsschwellen relevanten Eigenschaften des Gehörs berücksichtigen. Dabei erscheint die Nachbildung der Schallsignalverarbeitung im Gehör bis zu dem Detaillierungsgrad sinnvoll, der anhand physiologischer oder psychoakustischer Messungen noch eindeutig nachvollziehbar ist. Die mit der Modellkomplexität wachsende Anzahl von Parametern sollte also für einen sinnvollen Abgleich noch physiologisch oder psychoakustisch motivierbar sein.

Im folgenden wird ausgehend von den in Kapitel 2.6 vorgestellten Konzepten zur Bestimmung auditiver Wahrnehmungsschwellen die Struktur des psychophysiologischen Gehörmodells angegeben. Die Auslegung der Teilmodelle für das Außen- und Mittelohr, die Cochlea sowie die neuronale Verarbeitung erfolgt daran anschließend. Die Modellparameter werden dabei als Ergebnis des Abgleichs mit angegeben. Die zum Abgleich verwendeten Daten werden in Kapitel 4 genauer dargestellt. Beim Entwurf der Teilmodelle werden aus der Literatur bekannte Ansätze berücksichtigt. Geeignete Modelle werden ggf. ausgewählt und falls erforderlich erweitert. Die über die angegebenen Literaturquellen hinausgehenden Modellkomponenten beruhen auf eigenen Entwicklungen.

Für die Darstellung des Modells wurde größtenteils eine Repräsentation als elektrisches Netzwerk gewählt, um eine einheitliche mathematische Beschreibungsweise zu erreichen. Zur Nachbildung mechanischer Systeme wird dazu die in Anhang A.3 motivierte Transformation akustischer bzw. mechanischer Größen in elektrische Größen verwendet, die auf der Kraft-Strom Analogie beruht (*FI*-Analogie [204]).

3.1 Struktur des Gehörmodells

Die in Kapitel 2.6 genannten Konzepte zur Bestimmung auditiver Wahrnehmungsschwellen, auf denen viele bekannte Modelle basieren, werden ebenfalls hier für den grundlegenden Entwurf der Struktur des psychophysiologischen Modells herangezogen. Das zu entwickelnde Gehörmodell soll daher auf das grundlegende Blockdiagramm aus der Signaldetektionstheorie, bestehend aus einer Vorverarbeitung des Schallsignals, der Überlagerung eines „internen“ Störsignals und einem Beobachter (vgl. Bild 2.20), abgebildet werden.

Für eine didaktisch sinnvolle Darstellung des Gehörmodells soll hier ausgehend von der Struktur des Gehörmodells jedes darin enthaltene Teilmodell beschrieben werden. Die Vorgehensweise bei der Realisierung des Gehörmodells war demgegenüber umgekehrt und erfolgte durch Entwicklung und Zusammensetzung der Teilmodelle. Die Struktur des

Gehörmodell ist in Bild 3.1 in Anlehnung an die Anatomie und Physiologie angegeben und spiegelt die wesentlichen Schallverarbeitungsstufen des Gehörs wider. Das Gesamtmodell besteht aus Teilmodellen für das Außen- und Mittelohr (AMO), die Cochlea sowie die neuronale Verarbeitung. Das Teilmodell zur Nachbildung der Cochlea ist aus einzelnen, in Serie geschalteten Sektionen zusammengesetzt. Jede Sektion bildet die Eigenschaften der Hydromechanik (HM) sowie der ÄHZs und IHZs in einem kleinen longitudinalen Abschnitt der Cochlea nach. In jeder Sektion ist der IHZ ein Teilmodell der neuronalen Verarbeitung (NV) nachgeschaltet. Aus dem Schalldrucksignal am Außenohr als Eingangssignal des Gesamtmodells wird zunächst die Schnelle des ovalen Fensters bestimmt. In jeder Sektion des Modells der Cochlea resultiert daraus die ortsabhängige BM-Schnelle, die über das Modell der ÄHZ in jeder Sektion auf die Hydromechanik zurückgekoppelt wird. Die BM-Schnelle bildet zusammen mit dem Ausgangssignal der ÄHZ das Stimulationssignal der IHZ. Das Modell der IHZ bestimmt in jeder Sektion die neuronale Aktivität als Eingangssignal für das NV-Modell.

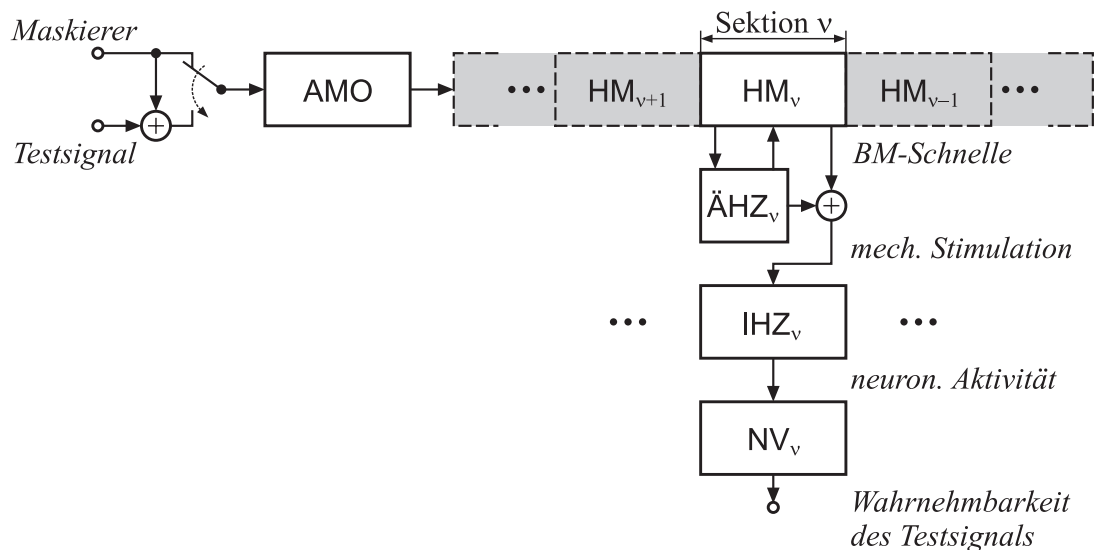


Bild 3.1 Vereinfachtes Blockdiagramm des psychophysiologischen Gehörmodells. Die einzelnen Blöcke beinhalten folgende Teilmodelle:

AMO: Außen- und Mittelohr; HM_v : Hydromechanik der Cochlea in Sektion v ; ÄHZ_v : äußere Haarzelle in Sektion v ; IHZ_v : innere Haarzelle in Sektion v ; NV_v : neuronale Verarbeitung in Sektion v . Die nicht dargestellten Sektionen des Modells besitzen die gleiche Struktur wie Sektion v .

Die Bestimmung der Wahrnehmungsschwelle für ein Testsignal erfolgt iterativ, indem zunächst das Maskierersignal in das Modell eingespeist wird. Im ersten Iterationsschritt wird auf der Ebene der neuronalen Verarbeitung die interne Repräsentation des Maskierersignals als gehörinternes Referenzsignal gespeichert. Im nächsten Iterationsschritt wird dem Maskierersignal das Testsignal überlagert. Dieses Summensignal wird vom Modell verarbeitet, so daß nunmehr die gehörinterne Repräsentation des Summensignals erzeugt wird. Diese Signalrepräsentation wird mit der gespeicherten Repräsentation des Maskierersignals in

jeder Sektion des Modells verglichen. Das Testsignal wird als wahrnehmbar angenommen, wenn die Änderung der internen Signalrepräsentation aufgrund des überlagerten Testsignals einen Schwellwert überschreitet. Die Wahrnehmungsschwelle kann nun iterativ bestimmt werden, indem der Testsignalpegel in jedem weiteren Iterationsschritt geeignet adaptiert wird, bis die Änderung der internen Signalrepräsentation in einer Sektion gerade den Schwellwert erreicht und in allen anderen Sektionen kleiner ist als der Schwellwert. Die iterative Ermittlung der Wahrnehmungsschwelle steht im Einklang mit psychoakustischen Messungen, da auch hier prinzipiell die gleiche iterative Vorgehensweise und die entsprechend generierten Maskierer- und Summensignale verwendet werden.

3.2 Modell des Außen- und Mittelohres

Die Veränderung des Schallsignals bei der Ausbreitung über das Außen- und Mittelohr kann im wesentlichen durch eine lineare Filterung beschrieben werden, wenn die Wirkung der Mittelohrmuskeln bei Pegeln oberhalb 80 dB SPL vernachlässigt wird (vgl. Kapitel 2.2.1). Da im allgemeinen die Einfallsrichtung des Schallsignals unbekannt ist, wird für die Modellierung eine mittlere Übertragungsfunktion des Außenohrschalldrucks zum Schalldruck am ovalen Fenster verwendet. Bei Frequenzen unterhalb von etwa 1 kHz wächst die Dämpfung mit etwa 6 dB/Okt. bei abnehmender Frequenz. Die erste Resonanz des Gehörgangs führt zu einer Überhöhung bei etwa 3 kHz (z. B. Bild 2.7 Kurve B). Die Resonanzen bei ungeradzahligem Vielfachen von 3 kHz werden in der Regel als weniger ausgeprägt angenommen. Die Nachbildung dieser Resonanzen ist jedoch nicht zweckmäßig, da die Längenabmessung des Gehörgangs verschiedener Personen variiert und entsprechend unterschiedliche Resonanzfrequenzen zur Folge hat. Die Übertragungsfunktion wird daher oberhalb der 3 kHz Resonanz als konstant angenommen.

Diese Übertragungsfunktion kann durch das in Bild 3.2 dargestellte passive elektrische Netzwerk nachgebildet werden. Der Strom i_{AO} entspricht dem Schalldruck am Außenohr. Die zunehmende Dämpfung bei tiefen Frequenzen wird durch die abnehmende Impedanz der Parallelschaltung von R_{M1} und L_M bewirkt. Eine Resonanz bei 3 kHz entsteht durch die Resonanz des Serienschwingkreises aus R_A , C_A und L_A . Der Phasengang der Übertragungsfunktion wurde bei der Modellierung nicht in Betracht gezogen. Die Werte der elektrischen Bauteile sind in Anhang A.6 angegeben.

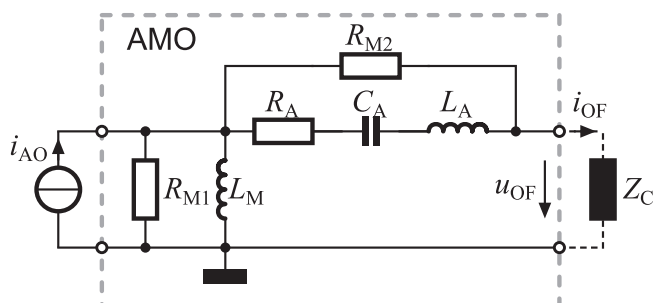


Bild 3.2 Modell des Außen- und Mittelohres, dargestellt als elektrisches Netzwerk. Der Ströme i_{AO} bzw. i_{OF} entsprechen dem Schalldruck am Außenohr bzw. am ovalen Fenster. Z_C repräsentiert die Eingangsimpedanz der nachfolgenden Teilmodelle.

In Bild 3.3 ist der gemessene Frequenzgang des Außen- und Mittelohrmodells bei Belastung mit den nachgeschalteten Teilmodellen dargestellt.

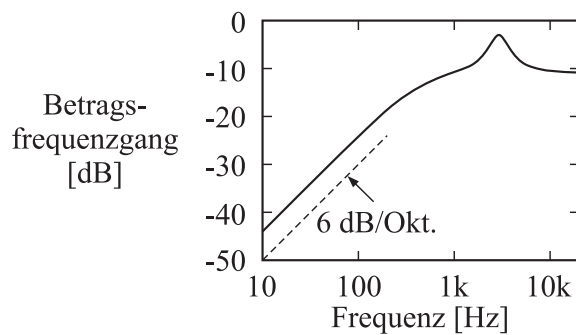


Bild 3.3 Gemessener Betragsfrequenzgang des Außen- und Mittelohrmodells aus Bild 3.2 mit Belastung durch die nachfolgenden Teilmodelle. Messung anhand der WDF-Implementierung gemäß Anhang A.6.

3.3 Physiologisches Modell der Cochlea

In der Cochlea befindet sich der Teil des Hörorgans, der die Spektralzerlegung und Transduktion des Schallsignals vollzieht (vgl. Kapitel 2.1.1). Die für eine Nachbildung von Wahrnehmungsschwellen relevanten Eigenschaften dieser Verarbeitungsschritte auf physiologischer Ebene erfordert eine physikalische Beschreibung der Hydromechanik der Cochlea, die durch eine Berücksichtigung des Verstärkungsmechanismus der ÄHZs ergänzt werden muß. Die Transduktion durch die IHZs ist ein nachfolgender Verarbeitungsschritt, bei dem die mechanischen Schwingungen der cochleären Trennwand in elektrische Aktionspotentiale auf den Hörnervenfasern umgesetzt werden. Wesentliche Eigenschaften der Transduktion sollen daher von einem Modell der IHZ nachgebildet werden.

3.3.1 Modell der Hydromechanik und der äußeren Haarzellen

Die Hydromechanik der Cochlea wird durch die physikalischen Eigenschaften ihrer Wandung, der Lymphflüssigkeiten, der Membranen und der cochleären Trennwand bestimmt. Die erreichbare Genauigkeit einer auf diesen Eigenschaften basierenden Nachbildung hängt wesentlich von der Meßgenauigkeit der physikalischen Parameter ab. Aufgrund der geringen Abmessungen der Cochlea und ihrer Empfindlichkeit gegenüber äußeren Eingriffen ist eine vollständige Vermessung der Cochlea bislang nicht gegeben. Deshalb gehen bekannte Modelle der Hydromechanik von einem vorgegebenen Strukturmodell des schwingungsfähigen mechanischen Systems aus. Zur Bestimmung der Parameter dieses Systems ist die Kenntnis der einzelnen Massen, Nachgiebigkeiten und Dämpfungen des Systems nicht im Detail notwendig, da ggf. die aus Schwingungsmessungen an der Cochlea bestimmten Resonanzfrequenzen und Resonanzgüten stattdessen herangezogen werden können.

Cochlea-Modelle werden entsprechend ihres Detailierungsgrads unterschieden. Eine Übersicht geben [90, 151, 169]. Die dreidimensionale Cochlea kann mit sinnvollen, vereinfachenden Annahmen auch mit zwei- oder eindimensionalen Modellen in ihren wesentlichen Eigenschaften nachgebildet werden. Ein zweidimensionales Modell vernachlässigt die Abhängigkeit der Schwingungsgrößen von der radialen Position. Ein eindimensionales Modell berücksichtigt außerdem keine Abhängigkeit von der axialen Position. Eine weitere Unterscheidung in makro- bzw. mikromechanische Modelle ist dadurch gegeben, daß nur mikromechanische Modelle eine Nachbildung der Strukturen innerhalb der cochleären

Trennwand beinhalten. Cochlea-Modelle mit aktiven Elementen zur Nachbildung der Wirkung der ÄHZs werden als aktiv und ggf. nichtlinear bezeichnet.

Das für das psychophysiologische Gehörmodell auszuwählende Modell der Cochlea sollte die Nachbildung derjenigen Wahrnehmungseigenschaften ermöglichen, die auch anhand gemessener Wahrnehmungsschwellen beobachtet werden können. In diesem Zusammenhang erscheint ein eindimensionales Modell als ausreichend, da mit zwei- oder dreidimensionalen Modellen, beispielsweise [23, 24], ein deutlich größerer Rechenaufwand erforderlich ist und die Ergebnisse gegenüber einem eindimensionalen Modell keine gravierend anderen Eigenschaften aufweisen [46, 90].

Eine Nachbildung der nichtlinearen Verstärkungswirkung der ÄHZs ist notwendig, um die Dynamikkompression, die Verzerrungsprodukte und die Suppression im Modell zu berücksichtigen. Die Annahme einer aktiven Verstärkungswirkung in der Cochlea wird durch folgende Beobachtungen gestützt:

- Die Energiequelle zur Generierung otoakustischer Emissionen befindet sich in der Cochlea [47].
- Eine elektrische Reizung der ÄHZs in der Cochlea führt zur Auslenkung der BM [76, 145, 188].
- Die ÄHZs zeigen eine Längenänderung infolge einer Auslenkung ihrer Stereozilien [106].
- Die Applikation bestimmter Substanzen zur Verhinderung der ÄHZ-Funktion führt zu einer reduzierten Empfindlichkeit der Cochlea und linearem, passivem Verhalten [152].
- Passive Modelle erreichen die Empfindlichkeit der Cochlea nur durch Verwendung unrealistischer Modellparameter. Die dabei erhaltene Filtergüte ist wesentlich zu groß.

Die Notwendigkeit einer Nachbildung der mikromechanischen Struktur der cochleären Trennwand hängt von mehreren Faktoren ab, die im folgenden diskutiert werden. Die mikromechanischen Strukturen der cochleären Trennwand sind als schwingungsfähige Systeme mit mehreren Freiheitsgraden anzusehen, die einen wesentlichen Einfluß auf den Transduktionsprozeß besitzen. Mit den heute verfügbaren physiologischen Daten ist aber eine Bestimmung der wichtigsten mikromechanischen Teilsysteme und ihrer Schwingungsmodi mit ausreichender Genauigkeit nicht möglich. Die meßtechnische Bestimmung physiologischer Daten der Cochlea ist u. a. aufgrund der folgenden Randbedingungen nur eingeschränkt möglich:

- Auslenkungen der Strukturen innerhalb der cochleären Trennwand liegen in der Größenordnung von Atom- bis Moleküldurchmesser: weniger als 1 nm an der Ruheschwelle.
- Nur ein kleiner Abschnitt der Cochlea ist heute bei Säugetieren für Messungen in physiologisch intaktem Zustand zugänglich.
- Die meisten Schwingungsmessungen innerhalb der Cochlea wurden an der Unterseite der BM durchgeführt. Schwingungsmessungen der näher an den Stereozilien der Sensoren gelegenen reticular Lamina sind nur in geringer Zahl verfügbar.
- Mit den üblichen Meßmethoden (Laser-Velocimeter, Mössbauer-Technik) wird in der Regel nur die Transversalschwingung der BM gemessen. Mögliche Radial- oder Longi-

tudinalkomponenten, die zum Beispiel bei der Tektorialmembran wahrscheinlich sind, werden damit nicht erfaßt. Erste Untersuchungen der Schwingungen in zwei zueinander senkrechten Richtungen mit erweiterten Meßgeräten zeigen, daß die radiale Bewegungskomponente der Tektorialmembran nicht vernachlässigt werden kann [73, 85].

Aufgrund der für eine umfassende mikromechanische Nachbildung der Cochlea unzureichenden physiologischen Meßdaten ist ein großer Teil der bekannten Cochlea-Modelle mit dem Ziel entwickelt worden, bestimmte Funktionshypothesen zu überprüfen. Die betreffenden Cochlea-Modelle werden zu diesem Zweck zusätzlich anhand der Eigenschaften der physiologischen Reizantworten auf einer höheren Ebene, wie beispielsweise den Aktionspotentialen auf einer Hörnervenfaser verifiziert. Wird eine möglichst umfassende Übereinstimmung mit physiologischen Daten erreicht, können derart verifizierte Funktionshypothesen fehlende physiologische Daten teilweise ersetzen.

Mikromechanische Modelle berücksichtigen neben der Basilarmembran auch die Tektorialmembran als Element innerhalb eines angekoppelten schwingungsfähigen Systems. Zusätzliche Annahmen über mögliche Schwingungsmoden anderer Teile der cochleären Trennwand führen zu komplexeren mikromechanischen Modellen mit zusätzlichen Freiheitsgraden. Mit zunehmender Komplexität ist jedoch eine physiologisch sinnvolle Parametereinstellung schwieriger zu erreichen. Repräsentative Beispiele solcher mikromechanischer Modelle sind publiziert in [2, 4, 60, 61, 125, 140, 136, 137]. Ein systematischer Vergleich unterschiedlicher mikromechanischer Modelle wird in [25] gegeben.

Vor dem Hintergrund der bislang nicht quantitativ ausreichend gegebenen Verifikation mikromechanischer Modelle durch physiologische Messungen erscheint es sinnvoll, soweit vertretbar auf eine Nachbildung der Mikromechanik im psychophysiologischen Gehörmodell zu verzichten. Insgesamt ergibt sich aus den oben aufgeführten Gründen, daß das Modell der Cochlea innerhalb des psychophysiologischen Modells auf einem eindimensionalen, aktiven, nichtlinearen Modell basieren sollte. Das am besten geeignete Cochlea-Modell wird im folgenden auf der Basis einer Auswahl von Cochlea-Modellen aus der Literatur, die die genannten Voraussetzungen erfüllen bestimmt.

Die Auswahl umfaßt die Cochlea-Modelle von Allen [2, 3], Nobili und Mammano [140], Giguère und Woodland [62, 63], Zwicker [195], sowie Zwicker und Peisl [201]. Die Modelle von Allen [2, 3] realisieren die Nichtlinearität durch die Annahme einer nichtlinearen Steifigkeit. Eine aktive Verstärkung ist nicht vorgesehen, so daß beispielsweise die Generierung von otoakustischen Emissionen nicht erklärt werden kann. Nobili und Mammano [140] schlagen ein detailliertes nichtlineares Modell vor, das jedoch aufgrund der Formulierung im Frequenzbereich nur auf stationäre Signale angewendet werden kann. Die erforderliche numerische Implementierung des Modells im Zeitbereich für eine Verarbeitung von Audiosignalen erscheint nicht zuletzt aufgrund der detaillierten Nachbildung der Cochlea als zu aufwendig. Das Cochlea-Modell von Giguère und Woodland [62, 63] beinhaltet als Nachbildung der Hydromechanik das Modell von Strube [180] und ist um nichtlineare Verstärker erweitert. Es entspricht daher weitgehend der Struktur des Analogmodells von Zwicker [195]. Beide letztgenannten Modelle führen jedoch bei einem hohen Verstärkungsfaktor zu unrealistisch hohen Filtergüten, die nicht im Einklang mit physiologischen Messungen ste-

hen. Das erweiterte Analogmodell von Zwicker und Peisl [201] führt eine laterale Kopplung benachbarter Verstärkerstufen zur Nachbildung der ÄHZs ein. Mit dieser Kopplung wird auch bei hohen Verstärkungsfaktoren eine adäquate Filtergüte erreicht. Dieses Modell wurde bereits erfolgreich zur Nachbildung von otoakustischen Emissionen eingesetzt [153] und zeigt die darin zu beobachtenden Suppressionswirkungen. Das erweiterte Analogmodell [153] erfüllt die hier gestellten Anforderungen am weitgehendsten und wird daher als Basis für das Modell der Cochlea innerhalb des psychophysiologischen Modells herangezogen.

Für die Implementierung eines nichtlinearen Modells sind verschiedene numerische Verfahren verfügbar, die auf einer Formulierung im Frequenz- oder Zeitbereich basieren und sich in Rechenaufwand und Stabilität unterscheiden. Eine Formulierung im Frequenzbereich ist für die Verarbeitung von Audiosignalen ungeeignet, da sie bei Vernachlässigung der Signalphase einerseits die Annahme zumindest kurzzeitig stationärer Signale erfordert. Andererseits führen die aufgrund der Nichtlinearität auftretenden Verzerrungsprodukte zu einem sehr hohen Rechenaufwand [140]. Liegt eine Modellbeschreibung in Form von Differentialgleichungen vor, ist eine Lösung mit Hilfe der bekannten Verfahren zur numerischen Integration möglich, wie beispielsweise [5]. In einem anderen Beispiel [45] wird das Verfahren finiter Differenzen verwendet, während bei einem Vergleich unterschiedlicher Verfahren in [46] ein Runge-Kutta-Verfahren die besten Stabilitätseigenschaften zeigt. Die Wahl der Schrittweiten bei der numerischen Integration erweist sich jedoch als kritischer Parameter, von dem sowohl der Approximationsfehler als auch die Stabilität des Verfahrens abhängt. Die Cochlea arbeitet bei geringen Schallpegeln bereits am Rande der Stabilität, kurzzeitige Instabilitäten können otoakustischen Emissionen hervorrufen [153]. Die numerischen Integrationsverfahren versagen insbesondere in diesem Arbeitsbereich und werden instabil. Die Diskretisierung des Cochlea-Modells im Ortsbereich ermöglicht demgegenüber eine Formulierung als System gekoppelter Differentialgleichungen, die unter Verwendung der mechanisch-elektrischen Analogie auch äquivalent durch ein elektrisches Netzwerk mit konzentrierten Elementen dargestellt werden können. Für die Simulation des elektrischen Netzwerkes stehen ebenfalls eine Reihe von Verfahren zur Verfügung. Die Simulation mit Hilfe von Wellendigitalfiltern (WDFs) [56] stellt dabei eine Methode mit hervorragenden Stabilitätseigenschaften dar. Im Gegensatz zur numerischen Integration kann unter Verwendung von WDFs die Stabilität auch bei einer hohen Verstärkung gewährleistet werden, wie Voruntersuchungen anhand von Implementierungen nach [45, 153] auch im Vergleich mit den Untersuchungen in [46] gezeigt haben.

Das von Fettweis [56] vorgeschlagene Konzept der WDF beruht auf der Wellenparametertheorie der Vierpole und ermöglicht die Simulation im Zeitbereich. WDF besitzen bei großem Dynamikbereich eine besonders geringe Empfindlichkeit gegenüber Toleranzen der Filterkoeffizienten. Mit Hilfe des Pseudoleistungskonzepts [57] konnte nachgewiesen werden, daß parasitäre Schwingungen (Grenzyklusschwingungen), die sich aufgrund der endlichen Wortlänge eines Digitalrechners ergeben und zu instabilem Verhalten führen, in einem WDF vermieden werden können. Gegenüber der üblichen, auf rekursiven Strukturen basierenden Kaskadenrealisierung [159] einer Übertragungsfunktion weisen WDFs auch bei Quantisierung der Filterkoeffizienten eine strukturinhärente Pseudopassivität und ein gerin-

geres Rundungsrauschen auf [57]. Darüberhinaus ist mit der WDF-Simulationsmethode eine einfache Implementierung der aktiven nichtlinearen Elemente möglich. Aus diesen Gründen wird das Modell der Cochlea als WDF implementiert und dementsprechend in Form eines elektrischen Netzwerkes angegeben.

Das hier zugrundeliegende Cochlea-Modell von Zwicker und Peisl [153] wurde sowohl als elektrisches Analogmodell mit diskreten Bauteilen realisiert, als auch mit WDFs als Rechnersimulation. In dieser Arbeit konnte die ursprüngliche Modellstruktur abgesehen von einer weiter unten beschriebenen Erweiterung im wesentlichen beibehalten werden. Im Rahmen des Gesamtmodells wurden die Parameter bei einem Modellabgleich leicht modifiziert. Die Darstellung des daraus resultierenden Cochlea-Modells erfolgt in zwei Stufen, wobei zunächst nur die makromechanische Nachbildung der passiven Cochlea angegeben wird. Diese wird danach um nichtlineare aktive Komponenten erweitert.

Das makromechanische Modell der passiven Cochlea wird im weiteren als Modell der Hydromechanik bezeichnet. Es geht von den mechanischen Elementen eines Strukturmodells, z. B. [90], oder einer äquivalenten Formulierung der beschreibenden Differentialgleichungen aus, z. B. [26, 146]. Beide Darstellungen können in ein elektrisches Netzwerk überführt werden, das bei örtlicher Diskretisierung des Modells mit konzentrierten Elementen aufgebaut werden kann. Das in Bild 3.4 dargestellte elektrische Netzwerk ist ein makromechanisches eindimensionales passives Modell der Hydromechanik, das auf der Basis der Kraft-Strom-Analogie (vgl. Anhang A.3) aus einem mechanischen Strukturmodell abgeleitet wurde [146]. Es besteht aus einer Serienschaltung von gleich strukturierten Sektionen, die jeweils die mechanischen Eigenschaften eines kleinen longitudinalen Abschnitts der Cochlea nachbilden. Die im Modell auftretenden Spannungen entsprechen den Volumengeschwindigkeiten, und die Ströme entsprechen den Kräften bzw. Druckdifferenzen in der Hydromechanik. Die am Eingang anliegende Spannung u_{OF} ist äquivalent zur Schnelle des ovalen Fensters. Der Parallelschwingkreis in einer Sektion v bildet durch die Induktivität L_v die örtliche Nachgiebigkeit der cochleären Trennwand nach. Die Kapazität C_v entspricht der

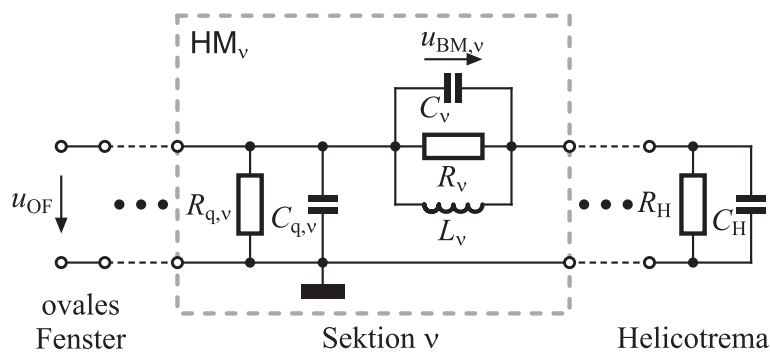


Bild 3.4 Makromechanisches eindimensionales passives Modell der Cochleahydromechanik. Alle Sektionen des Modells weisen die gleiche Struktur auf. Die Eingangsspannung u_{OF} ist äquivalent zur Schnelle des ovalen Fensters. Die Ausgangsspannungen $u_{BM,v}$ entsprechen der örtlichen Transversalschnelle der cochleären Trennwand. Die Resonanzfrequenz des Parallelschwingkreises einer Sektion nimmt mit zunehmendem Abstand vom ovalen Fenster ab.

Masse der cochleären Trennwand einschließlich der axial mitbewegten Lymphflüssigkeit in einem Abschnitt. Der Widerstand R_ν bildet Reibungsverluste nach, die zu einer endlichen Güte des schwingungsfähigen Systems führen. $C_{q,\nu}$ berücksichtigt die Masse der longitudinal bewegten Lymphflüssigkeit in einem Abschnitt und $R_{q,\nu}$ die dabei auftretenden Reibungsverluste. Die oberhalb und unterhalb der Trennwand bewegten Massen sind im Modell zusammengefaßt worden. Die Öffnung der cochleären Trennwand am Helicotrema wird durch die Masse der hindurchtretenden Lymphflüssigkeit entsprechend C_H und durch eine Dämpfung entsprechend R_H modelliert. Es wird die Inkompressibilität der Lymphflüssigkeit vorausgesetzt. Die bei hohen Frequenzen auftretenden Kompressionswellen werden daher nicht nachgebildet.

Das angegebene einfache Modell der Hydromechanik erlaubt eine Nachbildung der Frequenz-Orts-Transformation und Frequenzselektivität. Eine teilweise Berücksichtigung der dreidimensionalen Geometrie und Materialeigenschaften ist durch eine Verwendung der effektiv wirksamen Massen, Nachgiebigkeiten und Dämpfungen auch mit dem eindimensionalen Modell möglich.

Die Bestimmung der Parameter des Modells der Hydromechanik basiert auf Daten aus [146]. Die Dimensionierung des Parallelschwingkreises mit C_ν , R_ν und L_ν in jeder Sektion wird wie in [153] mit Hilfe der ortsabhängigen Resonanzfrequenz und Resonanzgüte durchgeführt. Die Resonanzfrequenzen ergeben sich mit der Festlegung einer Sektionsgröße von $133 \mu\text{m}$ longitudinaler Ausdehnung, so daß der Tonheitsabstand der Parallelresonanz benachbarter Sektionen $0,1$ Bark beträgt. Insgesamt werden 251 Sektionen vorgesehen, deren Resonanzen im Bereich zwischen $0,05$ und 25 Bark bzw. 5 Hz und $19,45$ kHz liegen. Die Wahl der Sektionsanzahl wird durch die gewünschte örtliche Auflösung bestimmt. Eine gegenüber $0,1$ Bark verringerte Auflösung kann jedoch zu unerwünschten Signalreflektionen an den Sektionsgrenzen führen, wenn die Impedanzen der gekoppelten Sektionen zu unterschiedlich sind. Die Resonanzgüten Q_ν werden basierend auf physiologischen Meßergebnissen für die passive Cochlea oder für hohe Schalldruckpegel eingestellt. Die Dimensionierung der Widerstände $R_{q,\nu}$ und Induktivitäten $L_{q,\nu}$ wird aus [153] übernommen. Die Bestimmungsgleichungen für die Werte der Bauelemente sind im Anhang A.7 angegeben. In Bild 3.5 sind die Werte in Abhängigkeit von der Sektionsnummer ν dargestellt.

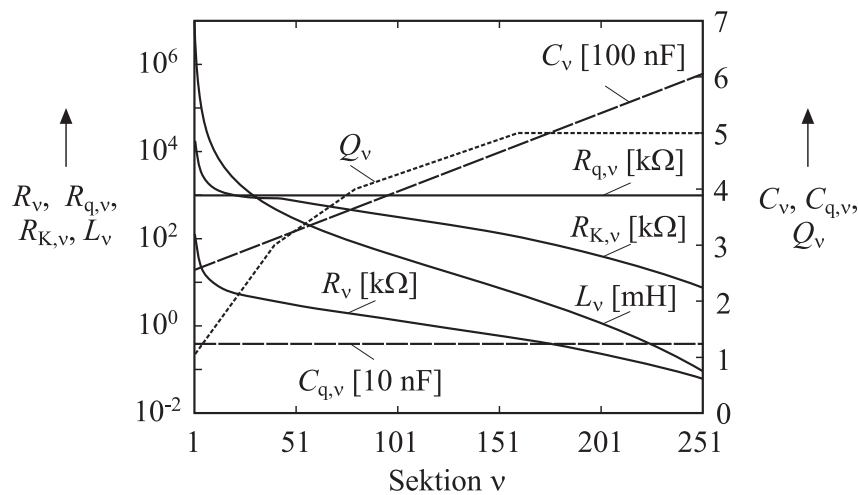


Bild 3.5 Schwingkreisgüten und Dimensionierung der Impedanzen im Cochlea-Modell.

Das Modell der Cochlea ohne IHZs ist in Bild 3.6 angegeben. Es zeigt die Erweiterung des Modells der Hydromechanik aus Bild 3.4 um die aktiven und nichtlinearen Elemente eines Modells der ÄHZ in Sektion v . Die nicht dargestellten Sektionen besitzen die gleiche Struktur wie Sektion v . Die Verstärkung der BM-Schnelle durch die ÄHZ wird mit Hilfe einer spannungsgesteuerten Spannungsquelle und einer punktsymmetrischen Sättigungskennlinie innerhalb einer Rückkopplungsschleife nachgebildet. Die Steuerspannung $u_{\text{BM},v}$ entspricht der BM-Schnelle und wird mit dem Faktor g_1 verstärkt. Die Ausgangsspannung des Verstärkers mit Sättigung wird lokal auf das Modell der Hydromechanik zurückgekoppelt und über weitere Kopplungswiderstände $R_{1a,v,\mu}$ und $R_{1b,v,\mu}$ auch auf benachbarte Sektionen gegeben. Dieser Teil des Modells entspricht dem erweiterten Analogmodell von Zwicker und Peisl [153]. Der spannungsgesteuerte Verstärker des ÄHZ-Modells setzt ein zur Geschwindigkeit der BM proportionales Eingangssignal der ÄHZ in der Cochlea voraus. Aufgrund der Anatomie ist aber von einer Proportionalität zur Auslenkung auszugehen (vgl. Kapitel 2.1.1), die jedoch wegen des höheren Rechenaufwands für die Integration der Geschwindigkeit zur Berechnung der Auslenkung nicht nachgebildet wird. Die Modellierung geht von der Annahme aus, daß die Proportionalität der Geschwindigkeit zum Produkt von Auslenkung und Frequenz bereits durch den Verstärkungsfaktor g_1 berücksichtigt wird. Bei hinreichend schmalbandiger Signalanalyse des HM-Modells ist diese Approximation ausreichend genau.

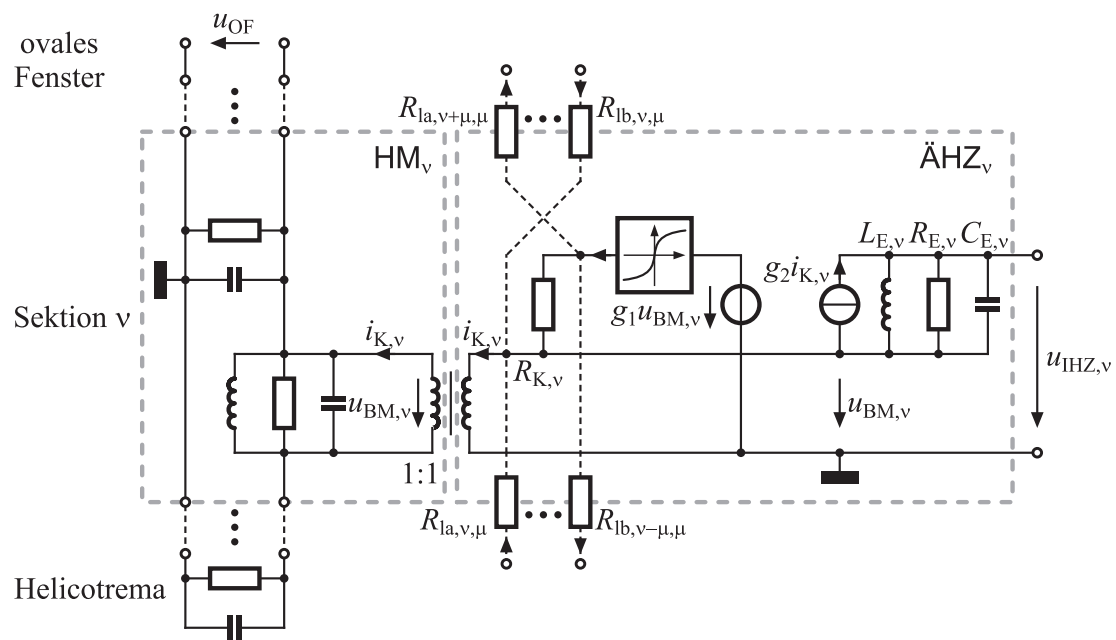


Bild 3.6 Modell der Cochlea ohne IHZs. Darstellung der ν -ten Sektion, bestehend aus einem Modell der Hydromechanik (HM_ν) und der äußeren Haarzelle ($\ddot{A}HZ_\nu$). Die nicht dargestellten Sektionen des Modells weisen die gleiche Struktur auf. Die Widerstände R_{la} und R_{lb} repräsentieren nur einen Teil der zwischen benachbarten Sektionen vorhandenen Kopplungswiderstände.

Der rückgekoppelte Verstärker führt bei einer Verstärkung der BM-Schelle von mehr als ca. 40 dB gegenüber dem passiven Modell zur Instabilität. Daher wurde zur Erzielung der in physiologischen Messungen beobachteten Verstärkung von etwa 60 dB [144, 164, 167] eine zweite Verstärkerstufe als Erweiterung des Modells von Zwicker und Peisl [153] außerhalb der Rückkopplungsschleife ergänzt. Die Stromquelle dieser Stufe wird von dem in das Modell der Hydromechanik zurückgekoppelten Strom $i_{K,\nu}$ gesteuert und mit $g_2 = 20$ verstärkt. Die Belastung der Stromquelle mit dem Parallelschwingkreis bewirkt eine Frequenzabhängigkeit des Spannungsabfalls über dieser Impedanz, ähnlich wie in der Sektion des Modells der Hydromechanik. Dieser Parallelschwingkreis ist mit den gleichen Kapazitätswerten und der gleichen Resonanzgüte wie der Parallelschwingkreis in derselben Sektion des Modells der Hydromechanik ausgelegt ($C_{e,\nu} = C_\nu$). Seine Resonanzfrequenz ist jedoch um den Faktor 0,9 geringer. Die Spannung $u_{BM,\nu}$ wird dementsprechend am Ausgang des ÄHZ-Modells um eine Spannung erhöht, die näherungsweise proportional zu der vom Rückkopplungskreis durch den Strom $i_{K,\nu}$ verursachten Teil der Spannung $u_{BM,\nu}$ ist. Die ergänzte zweite Verstärkerstufe ist physiologisch nicht unmittelbar zu rechtfertigen, da die ÄHZ nur ein verstärkendes Element enthält. Für die hier gewählte Struktur spricht jedoch, daß damit eine nur teilweise Rückkopplung der ÄHZ-Verstärkung auf die Hydromechanik angenommen wird, während die gesamte Verstärkung auf das Reizsignal $u_{IHZ,\nu}$ der IHZ wirkt. Das ursprüngliche Konzept des Zwickerschen Analogmodells sieht bereits eine nur teilweise Rückkopplung der ÄHZ-Verstärkung vor [193, 197: Fig. 3.9]. Diese Annahme stellt eine

Lösungsmöglichkeit des Stabilitätsproblems bei Verwendung physiologisch gemessener Verstärkungsfaktoren dar, das bislang auch mit anderen Modellansätzen, die eine Signalverstärkung durch die ÄHZ beinhalten, nicht gelöst werden konnte. Die Modellerweiterung durch eine Verstärkerstufe außerhalb der Rückkopplung stellt weiterhin sicher, daß die von Peisl erfolgreich durchgeführte Nachbildung von otoakustischen Emissionen [153] in gleicher Weise mit dem hier angegebenen Modell möglich ist.

Die Dimensionierung des Modells der ÄHZ erfolgt in Anlehnung an [153]. Die Sättigungskennlinie wird hier aufgrund der geringeren Parameteranzahl als punktsymmetrisch zum Ursprung angenommen. Daher entstehen bei der Aussteuerung der Kennlinie nur Verzerrungsprodukte ungeradzahlgiger Ordnung. Diese Annahme steht im Einklang mit der Auffassung, daß die im peripheren Gehör erzeugten Verzerrungen gerader Ordnung im wesentlichen auf das Mittelohr zurückzuführen sind [197]. Die Sättigungskennlinie ist in Bild 3.7 auf der Grundlage der in Anhang A.1 angegebenen Pegelnormierung dargestellt. Sie wurde so eingestellt, daß einerseits zusammen mit der Verstärkung g_1 eine mit physiologischen Messungen vergleichbare pegelabhängige Verstärkung erreicht wird und andererseits die entstehende Verzerrungsprodukte mit zunehmender Ordnung einen geringeren Pegel aufweisen und unterhalb des Pegels bei der Grundfrequenz liegen. Dieses Vorgehen führt zu vergleichbaren Ergebnissen, wie sie in psychoakustischen Kompensationsmessungen für Kombinationsöne beobachtet werden [197].

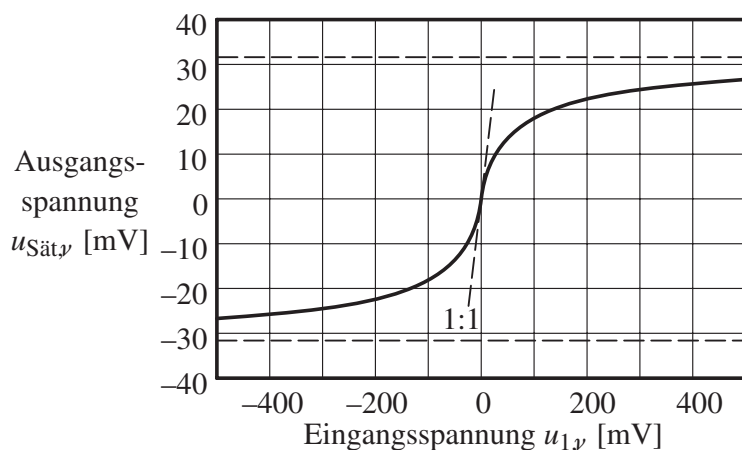


Bild 3.7 Sättigungskennlinie des Modells der ÄHZ. Die zum Ursprung punktsymmetrische Kennlinie besitzt im Ursprung die Steigung eins und nähert sich asymptotisch den angegebenen Grenzwerten von $\pm 31,6$ mV.

In Bild 3.8 sind die Pegel der Grundfrequenzkomponente und der Obertöne bis zur siebten Harmonischen dargestellt, die aufgrund der Sättigungskennlinie mit Berücksichtigung der Verstärkung $g_1 = 100$ bei Einspeisung eines Sinustons $u_{\text{BM},y} = A \sin(2\pi f_0 t)$ entstehen. Der Oberton mit dreifacher Grundfrequenz weist den höchsten Pegel der auftretenden Obertöne auf. Mit zunehmendem Eingangspegel verringert sich die Pegeldifferenz zwischen dem Grundton und den Obertönen. Diese Eigenschaft führt im Gehörmodell zu maximalen kubischen Verzerrungen bezogen auf den Grundtonpegel bei mittleren Eingangspegeln, da bei hohen Eingangspegeln das lineare passive Verhalten (gepunktete Gerade in Bild 3.8) dominiert. Damit wird der in Kompensationsmessungen bestimmte Pegel der kubischen Verzerrungen [197] qualitativ nachgebildet.

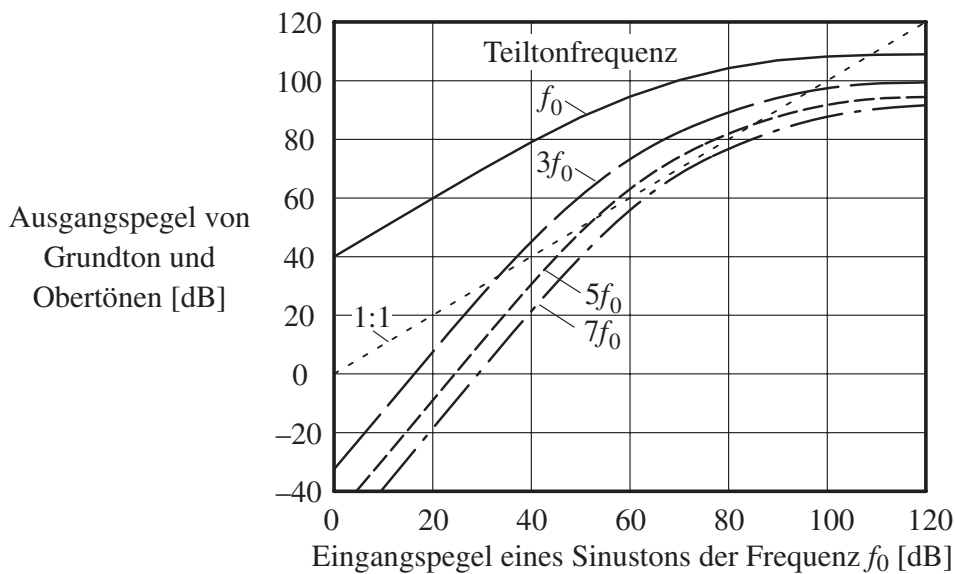


Bild 3.8 Pegel des Grundtons und der ersten 7 Obertöne, die aufgrund der Sättigungskennlinie in Bild 3.7 mit Berücksichtigung der Verstärkung von $g_1 = 100$ bei Einspeisung eines Sinustons mit der Frequenz f_0 entstehen.

Bei gegebener Sättigungskennlinie bestimmt der Verstärkungsfaktor g_1 zusammen mit dem Rückkopplungswiderstand $R_{K,\nu}$ und den lateralen Kopplungswiderständen $R_{1a,\nu,\mu}$ und $R_{1b,\nu,\mu}$ zwischen den Sektionen die resultierende Verstärkung durch den Rückkopplungszweig in einer Sektion. Die Verstärkung wird mit $g_1 = 100$ festgelegt. Die Dimensionierung der Kopplungswiderstände wurde in Vorversuchen mit dem Ziel durchgeführt, bei stabilem Gesamtsystem eine über den gesamten Ortsbereich möglichst konstante maximale Verstärkung für kleine Signalpegel zu erreichen. Dabei wurde für die lateralen Kopplungswiderstände eine Gaußförmige Gewichtungsfunktion wie in [153] verwendet. Dies führt zu einer geringer werdenden Verkopplung mit zunehmender Distanz der Sektionen. Die verwendete Gewichtungsfunktion ist in Bild 3.9 dargestellt. Die lateralen Kopplungswiderstände werden aus den in Bild 3.5 dargestellten Werten der Widerstände $R_{K,\nu}$ durch entsprechende Gewichtung ermittelt.

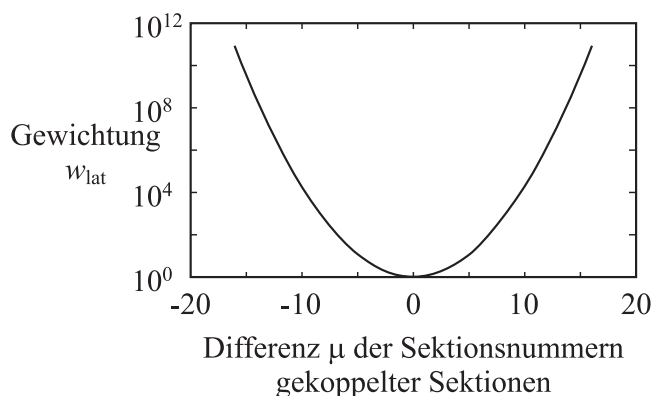


Bild 3.9 Gewichtungsfunktion zur Bestimmung der Kopplungswiderstände $R_{1a,\nu,\mu}$ und $R_{1b,\nu,\mu}$ aus den Kopplungswiderständen $R_{K,\nu}$ in Abhängigkeit von der Sektionsdifferenz.

Die im Modell der Cochlea verwendete Dimensionierung der Netzwerkelemente und die entsprechende Realisierung als WDF ist im Anhang A.7 angegeben.

3.3.2 Modell der inneren Haarzellen

Die Reizung der IHZ erfolgt näherungsweise proportional zur Relativgeschwindigkeit der Stereozilien zur umgebenden Flüssigkeit. Dieses Reizsignal wird im Gehörmodell durch die Spannung $u_{\text{IHZ},\nu}$ in jeder Sektion repräsentiert. Das Modell der IHZ in jeder Sektion enthält zur Nachbildung ihrer wesentlichen Eigenschaften einen Quadrierer und ein Tiefpaßfilter erster Ordnung gemäß Bild 3.10. Die Zeitkonstante $\tau_{\text{IHZ},\nu}$ des Tiefpasses ist in Anhang A.8 angegeben. Die mit einer IHZ verbundenen unterschiedlichen Fasertypen des Hörnervs zeigen einen Arbeitsbereich, der eine eingeschränkte Dynamik bei einem unterschiedlichen Signalpegel umfaßt (Bild 2.11). Diese eingeschränkte Dynamik wird im Modell nicht im Detail berücksichtigt. Es wird stattdessen angenommen, daß die Gesamtheit der angeschlossenen Fasern den ganzen Dynamikbereich des Eingangssignals uneingeschränkt, d. h. ohne Dynamikbegrenzung oder Kompression, überträgt. Der Gleichrichtungseffekt der IHZ wird durch den Quadrierer nachgebildet. Die Quadrierung trägt außerdem Meßergebnissen Rechnung, die eine näherungsweise lineare Abhängigkeit von Wahrnehmungsgrößen bezüglich der Signalintensität am Eingang des zentralen Gehörs zeigen [156]. Obwohl physiologische Messungen darauf hindeuten, daß die Bildung der Signalintensität erst jenseits der IHZ lokalisiert ist, wird die Quadrierung im Modell vorgezogen. Untersuchungen haben gezeigt, daß die Verlagerung der Quadrierung in das NV-Modell keinen gravierenden Einfluß auf die Ergebnisse der Mithörschwellessimulation hat, solange sie vor dem darin enthaltenen Kurzzeitintegrator erfolgt. Die Quadrierung im Modell der IHZ führt jedoch aufgrund der dadurch eingesparten Betragsfunktion zur Durchführung der Gleichrichtung zu einem reduzierten Rechenaufwand.

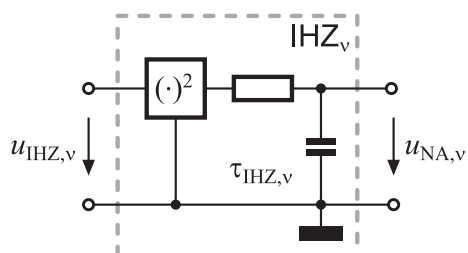


Bild 3.10 Modell der inneren Haarzelle mit Quadrierer und Tiefpaß erster Ordnung. Die Zeitkonstante des RC -Glieds $\tau_{\text{IHZ},\nu}$ ist von der Sektionsnummer ν abhängig.

Im Gehörmodell wurden auch andere aus der Literatur bekannte komplexere IHZ-Modelle auf der Grundlage von [88, 127, 128, 129, 135] erprobt. Dies führt jedoch nicht zu einer verbesserten Übereinstimmung der Simulationsergebnisse des Gesamtmodells mit psychoakustischen Messungen. Die Untersuchungen haben gezeigt, daß insbesondere die nachgebildete pegelabhängige Sättigung der Entladungsrates dieser IHZ-Modelle nicht ohne weiteres in das Gehörmodell integriert werden kann. Die IHZ-Modelle besitzen außerdem den Nachteil eines wesentlich erhöhten Rechenaufwands.

Die Eigenschaften der IHZ-Synapsen zeigen ein deutliches zeitliches Adaptionsverhalten, das im wesentlichen nach einer Stimulation der IHZ wirksam ist und dort zu einer reduzierten Empfindlichkeit führt [187]. Dieses Adaptionsverhalten trägt zur Nachverdeckung

bei und wird daher im NV-Modell gemeinsam mit dem Beitrag des zentralen Gehörs nachgebildet.

Die Ausgangsspannung $u_{NA,v}$ des IHZ-Modells repräsentiert die neuronale Aktivität der Nervenfasern, die in dem einer Sektion entsprechenden Abschnitt der Cochlea kontaktiert sind. Die neuronale Aktivität kann als kurzzeitig gemittelte Entladungsrate aufgefaßt werden. Sie stellt das Eingangssignal des NV-Modells in jeder Sektion dar.

3.4 Psychoakustisches Modell der neuronalen Verarbeitung

Die Modellierung der neuronalen Verarbeitung von der IHZ Synapse bis zum auditorischen Cortex erfolgt auf der Basis einer psychoakustischen Nachbildung, da die beteiligten physiologischen Prozesse bislang nur unzureichend bekannt sind. Psychoakustische Messungen zeigen, daß die Frequenzselektivität des Gehörs weitgehend durch die Signalverarbeitung der Cochlea bestimmt ist. Demgegenüber wird die Auflösung eines Signals im Zeitbereich nicht nur durch mechanische Ein- und Ausschwingvorgänge im peripheren Gehör, sondern auch wesentlich durch neuronale Adaptionsvorgänge in den IHZ-Synapsen und im zentralen Gehör bestimmt. Aus diesem Grund wird im Gehörmodell in jeder Sektion jeweils ein unabhängiges NV-Modell verwendet, so daß die Frequenzselektivität des Gehörmodells auf dieser Stufe beispielsweise nicht durch eine Kopplung benachbarter NV-Modelle beeinflusst wird. Beim Entwurf des NV-Modells wurde das Ziel verfolgt, mit einer möglichst geringen Zahl von Modellparametern eine geeignete Approximation zeitabhängiger Verdeckungseffekte zu erreichen. Zu diesem Zweck sind mathematische Funktionen ausgewählt worden, die neben einer einfachen Parametrierung auch einen möglichst geringen Rechenaufwand bei der Signalverarbeitung erfordern.

Das Blockdiagramm des NV-Modells in einer Sektion ist in Bild 3.11 dargestellt. Das Eingangssignal $u_{NA,v}$ wird mit einem internen, zwischen 0 und $u_{NMAX,v}$ gleichverteilten, weißen Rauschsignal überlagert, das die Möglichkeit einer Signaldetektion auf den Pegelbereich oberhalb der Ruhehörschwelle einschränkt. Die Werte von $u_{NMAX,v}$ sind in Anhang A.9.1 angegeben. Das Summensignal wird mit $u_{HN,v}$ bezeichnet. Jedes NV-Modell beinhaltet zur Nachbildung der wesentlichen zeitlichen Verdeckungseigenschaften mehrere Funktionseinheiten. Im einzelnen sind das die Vor- und Nachverdeckung, sowie die Abhängigkeit der Verdeckung von der Testsignaldauer. Darüber hinaus wird ein Vergleich der internen Signalrepräsentation von dem Maskierer allein und dem Maskierer mit überlagertem Testsignal durch einen Speicher ermöglicht, der den hierfür relevanten Teil des Gedächtnisses repräsentiert. Die Entscheidung über das Vorhandensein eines Testsignals wird mit Hilfe eines Schwellendetektors in den NV-Modellen jeder Sektion gewonnen und durch eine Maximumbildung über die Ausgänge aller Sektionen ausgewertet. Die Funktionseinheiten werden im folgenden erläutert. Im Anhang A.4 und A.9 sind weitere Implementierungsdetails des NV-Modells dargelegt.

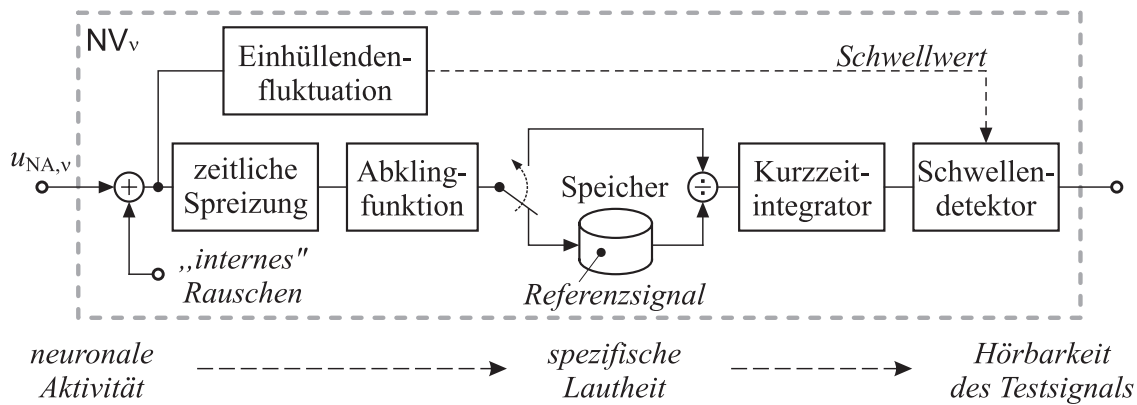


Bild 3.11 Vereinfachtes Blockdiagramm der Nachbildung der neuronalen Verarbeitung in einer Sektion. Für die Bestimmung der Wahrnehmungsschwelle eines Testsignals wird zunächst die spezifische Lautheit des Referenzsignals im Speicher, der die Nachbildung des Gedächtnisses darstellt, abgelegt. In nachfolgenden Iterationsschritten wird der Testsignalpegel so eingestellt, daß die Signaländerung in einer Sektion als gerade wahrnehmbar detektiert wird.

3.4.1 Nachbildung der Vorverdeckung

Die mechanischen Einschwingvorgänge des peripheren Gehörs verlaufen schneller als die bei der Vorverdeckung gemessenen Dauern. Zur Nachbildung der Vorverdeckung wird daher eine zusätzliche zeitliche Spreizung (Verschmierung) vorgesehen, die durch eine Faltung mit einer Spreizfunktion realisiert wird. Die Größe am Ausgang der Spreizung wird mit $u_{Sp,v}$ bezeichnet. In der Literatur werden vergleichbare Ansätze zur Nachbildung der zeitabhängigen Verdeckung verwendet, beispielsweise mit einer Gewichtsfunktion [147]. Die Spreizfunktion $Sp(t)$ wird in Bild 3.12 gezeigt. Für die Darstellung ist die Gaußförmige Funktion auf den Maximalwert von eins normiert und zentriert worden. Die Spreizfunktion wird in Anhang A.9.2 analytisch beschrieben.

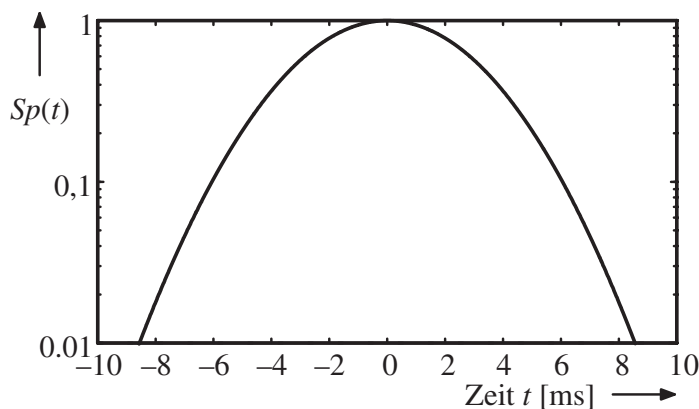


Bild 3.12 Spreizfunktion in normierter zentrierter Darstellung.

3.4.2 Nachbildung der Nachverdeckung

Die Nachverdeckungsschwelle für Maskiererdauern über 200 ms weist näherungsweise eine lineare Pegelabnahme auf einer logarithmierten Zeitachse nach Abschalten des Maskierers

auf [109, 197]. Der Übergang vom Simultanverdeckungspegel zur Ruhehörschwelle vollzieht sich innerhalb der Nachverdeckungsdauer von etwa 200 ms. Dieser Abklingvorgang kann effizient mit additiv überlagerten Exponentialfunktionen angenähert werden [79]. Die Steilheit der Abklingfunktion wird dann mit Hilfe von unterschiedlichen Zeitkonstanten und Gewichtungen mit dem System eingestellt, das in Bild 3.13 gegeben ist. Die Implementierung der verwendeten RC -Glieder als zeitdiskrete Systeme wird in Anhang A.4 beschrieben. Die Ausgangsspannung $u_{N',v}$ wird als eine der spezifischen Lautheit entsprechende Größe angenommen (vgl. [192]).

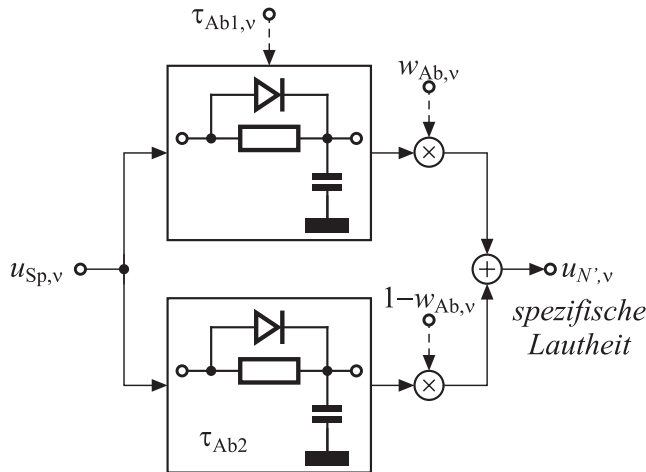


Bild 3.13 Blockdiagramm des Systems zur Erzeugung der Abklingfunktion für die Modellierung der Nachverdeckung.

Die größere Zeitkonstante beträgt $\tau_{Ab2} = 50$ ms. (3.1)

Die Nachverdeckung weist eine pegelabhängige Steilheit auf, da unabhängig vom Simultanverdeckungspegel der Übergang in die Ruhehörschwelle nach etwa der gleichen Dauer eintritt. Die Steilheit der Abklingfunktion muß daher pegelabhängig gesteuert werden. Dies wird über eine Steuerung der relativen Gewichtung beider Exponentialfunktionen mit dem Zusammenhang gemäß (3.2) erreicht. Der momentane Signalpegel in einer Sektion wird anhand der Ausgangsspannung $u_{L,v}$ eines Tiefpasses, wie in Bild 3.14 dargestellt, gewonnen.

$$w_{Ab,v}(n) = 1 - 0,001 \left(\frac{u_{L,v}(n)}{50 \text{ V}} \right)^{-0,6} \quad (3.2)$$

Bei Verringerung der Maskiererdauer tritt ein schnelleres Abklingen der Nachverdeckung ein, so daß die Steilheit der Nachverdeckungsfunktion zunimmt. Dieser Effekt wird nachgebildet, indem die kleinere der beiden Zeitkonstanten $\tau_{Ab1,v}$ in Abhängigkeit von der approximierten Maskiererdauer gesteuert wird. Die Approximation der Dauer wird mit Hilfe einer Quotientenbildung der Ausgangssignale zweier Tiefpaßfilter gemäß dem Blockdiagramm in Bild 3.14 und (3.3) durchgeführt. Bei einem konstanten Eingangssignal ist der Quotient zunächst eins. Wenn der Signalpegel anschließend zunimmt, verringert sich der Quotient, da der Divisor aufgrund der Maximumbildung zwischen Eingang und Ausgang des RC -Gliedes (vgl. Anhang A.4) dem Eingangssignal unmittelbar folgt. Falls der Signalpegel danach konstant bleibt, nähert sich der Quotient während einer Übergangszeit von etwa 200 ms wieder

dem stationären Wert eins an. Während dieser Übergangszeit wird die Zeitkonstante $\tau_{Ab1,v}$ gemäß (3.4) bis maximal $\tau_{Ab1,min}$ reduziert.

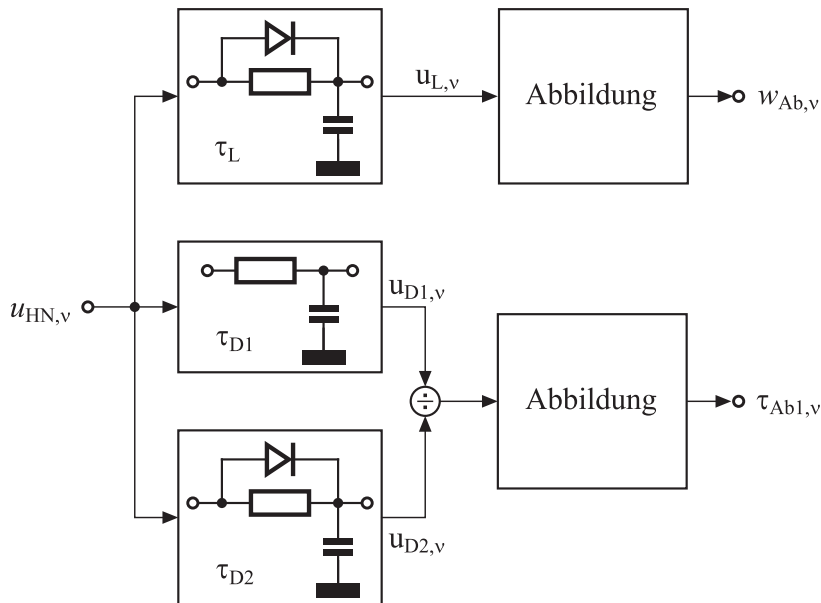


Bild 3.14 Blockdiagramm des Systems zur Approximation des Signalpegels und der Signaldauer zur Steuerung der Abklingfunktion des Systems in Bild 3.13.

$$\tau_L = 0,2 \text{ s} ; \tau_{D1} = 0,3 \text{ s} ; \tau_{D2} = 0,3 \text{ s} \quad (3.3)$$

$$\tau_{Ab1,v}(t) = \text{Min} \left(\tau_{Ab1,max} , k_1 \frac{u_{D1,v}(t)}{u_{D2,v}(t)} + \tau_{Ab1,min} \right) \quad (3.4)$$

mit $\tau_{Ab1,min} = 5 \text{ ms} ; \tau_{Ab1,max} = 20 \text{ ms} ; k_1 = 23,7 \text{ ms} . \quad (3.5)$

Die Diagramme in Bild 3.15 veranschaulichen die Wirkung der unterschiedlichen Gewichtung der Exponentialfunktionen sowie der Adaption einer Zeitkonstanten bezüglich der Abklingfunktion.

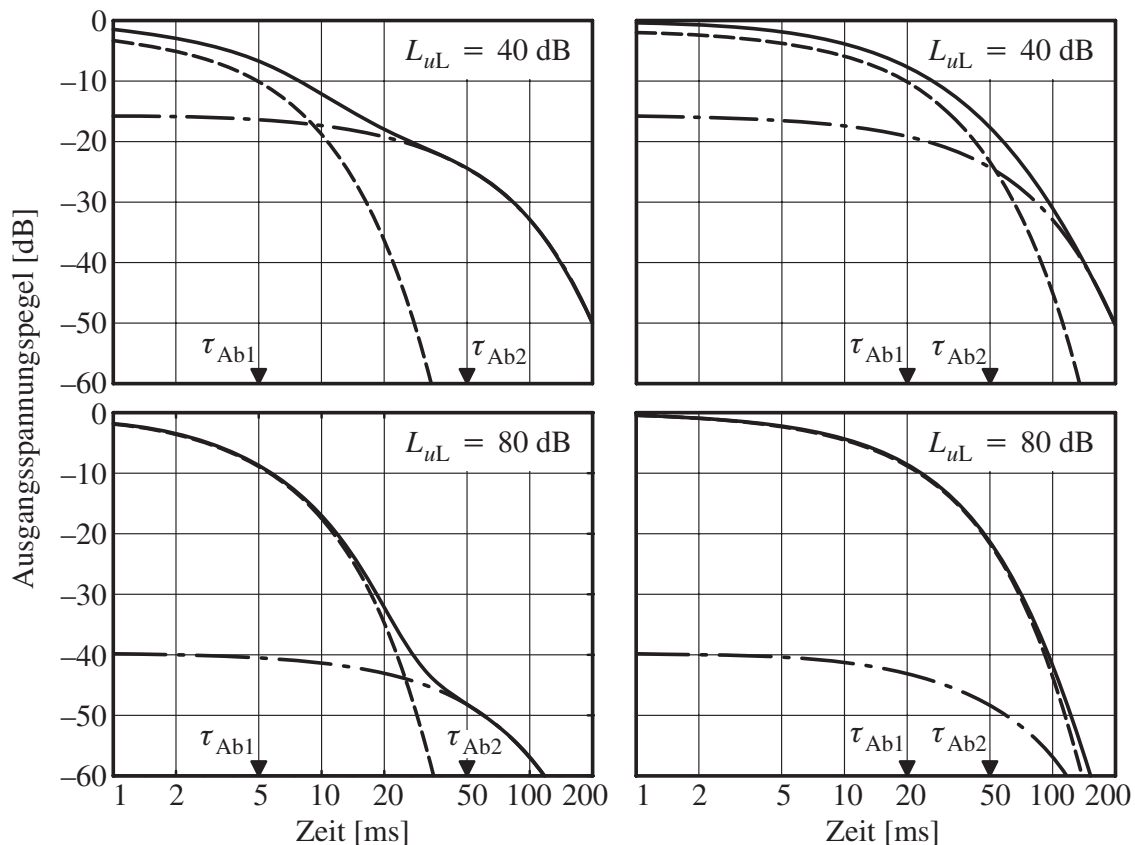


Bild 3.15 Abklingfunktion für unterschiedliche Gewichtungen und Zeitkonstanten des Systems aus Bild 3.13 nach Abschalten des Eingangssignals mit einem Pegel von 0 dB. Der Signalpegel (*oben* 40 dB, *unten* 80 dB) führt zu einer unterschiedlichen Gewichtung der beiden *gestrichelt* dargestellten Exponentialfunktionen. Die Maskiererdauer (*links* klein, *rechts* groß) steuert die Zeitkonstante τ_{Ab1} einer Exponentialfunktion.

3.4.3 Nachbildung der Abhängigkeit der Verdeckung von der Testsignaldauer

Bei Verwendung von Testsignalen unterschiedlicher Dauer verringert sich die Ruhehörschwelle bzw. die Mithörschwelle bei identischem Maskierersignal, wenn die Testsignaldauer bis etwa 200 ms erhöht wird [197]. Diese Zeitkonstante kann daher als Integrationszeit des Gehörs aufgefaßt werden, die der Dauer eines Zeitfensters entspricht, innerhalb dessen die Signalenergie summiert wird. Die Integrationswirkung ist bezüglich der von einem Testsignal hervorgerufenen Signaländerung zu beobachten. Daher ist vor der Kurzzeitintegration die Änderung der spezifischen Lautheit des Maskierers durch die in Bild 3.11 dargestellte Quotientenbildung von der spezifischen Lautheit des Eingangssignals und der intern gespeicherten spezifischen Referenzlautheit des Maskierers zu bestimmen. Die Pegelabnahme der Verdeckung kann ähnlich wie bei der Nachverdeckung als linear bezüglich einer logarithmierten Zeitachse angenähert werden. Aus diesem Grund werden für die Nachbildung ent-

sprechend die unterschiedlich gewichteten Ausgänge von zwei RC -Gliedern verwendet, wie im oberen Teil von Bild 3.16 gezeigt, die eine Abnahme von 10 dB/Dekade bezüglich der Zeitachse approximieren. Die Eingangsgröße ist der Quotient der spezifischen Lautheiten $r_{N',v}$. Die zugeordneten Parameter des Kurzzeitintegrators sind in (3.6) angegeben.

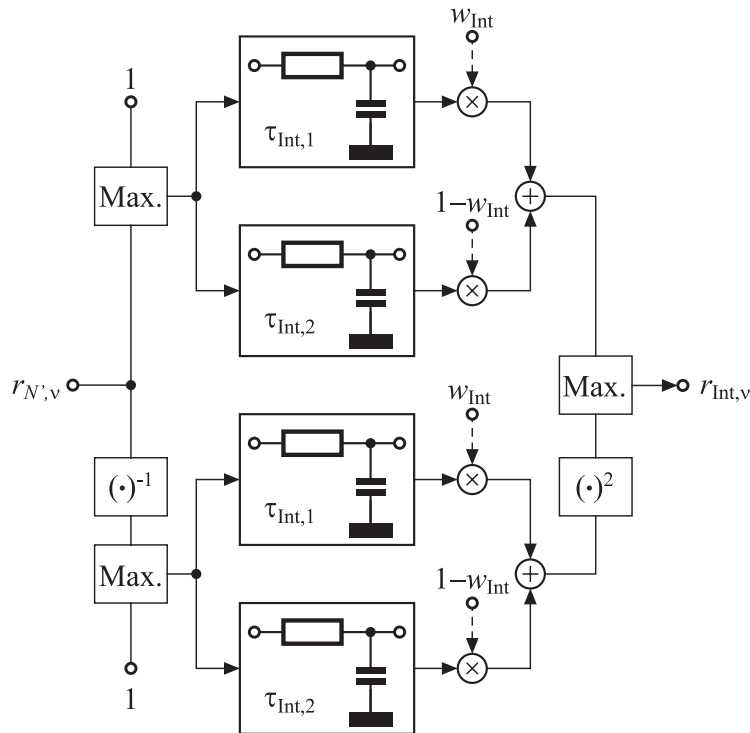


Bild 3.16 Blockdiagramm des Kurzzeitintegrators zur Nachbildung der Abhängigkeit von der Testsignaldauer. Das obere Paar von RC -Gliedern bildet den Kurzzeitintegrator für relative Änderungen der spezifischen Lautheit größer als eins. Das untere Paar bildet den Kurzzeitintegrator für Änderungen der spezifischen Lautheit, die kleiner als eins sind (Abnahme der spezifischen Lautheit).

$$w_{\text{Int}} = 0,9; \tau_{\text{Int},1} = 50 \text{ ms}; \tau_{\text{Int},2} = 2 \text{ ms} \quad (3.6)$$

Vorversuche zur Simulation von Verdeckungsexperimenten haben gezeigt, daß in der Regel die *Zunahme* der spezifischen Lautheit durch das überlagerte Testsignal zum Erreichen der Ruhe- bzw. Mithörschwelle führt. Bei Anwendungen in der Audiocodierung ist aber ebenfalls die *Verringerung* der spezifischen Lautheit von Bedeutung, da nicht übertragene (codierte) Spektralbereiche als fehlend wahrgenommen werden können. Daher ist zusätzlich ein Kurzzeitintegrator für eine gegenüber der Referenz reduzierte spezifische Lautheit vorgesehen. Dieser Integrator ist identisch aufgebaut, aber er wird mit dem Kehrwert der spezifischen Lautheitsänderung $1/r_{N',v}$ gespeist, wie in Bild 3.16 unten dargestellt. Die Ausgangsgröße dieses Integrators wird zusätzlich durch eine Quadrierung noch verstärkt, so daß eine Verringerung der spezifischen Lautheit stärker gewichtet wird als eine Erhöhung.

Die Sprungantwort eines Kurzzeitintegrators, bestehend aus zwei unterschiedlich gewichteten Ausgangssignalen zweier Tiefpässe erster Ordnung, ist in Bild 3.17 für die Parameter aus (3.6) dargestellt.

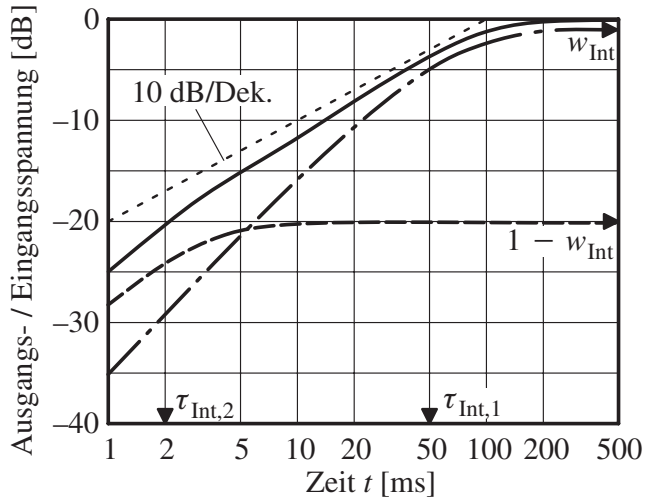


Bild 3.17 Verhältnis von Ausgangs- zu Eingangsspannung des Kurzzeitintegrators bei Einspeisung eines Spannungssprungs von 0 V auf einen konstanten Wert zum Zeitpunkt $t = 0$ s. Die Spannungsverhältnisse der RC -Glieder sind *strichpunktiert* ($\tau_{\text{Int},1}$) bzw. *gestrichelt* ($\tau_{\text{Int},2}$) dargestellt.

3.4.4 Schwellendetektor

Die Bestimmung der Wahrnehmbarkeit einer Signaländerung wird mit Hilfe eines Schwellendetektors in jeder Modellsektion durchgeführt. Der jeweilige Schwellwert wird auf $r_{\text{Int},\nu}$, dem kurzzeitintegrierten Verhältnis der spezifischen Lautheiten, angewendet. Die Änderung der spezifischen Lautheit wird in den Sektionen als wahrnehmbar angenommen, in denen der Schwellwert überschritten wird. Im Sinne der Signaldetektionstheorie ist dabei das Ausmaß der Über- bzw. Unterschreitung des Schwellwerts für eine Abschätzung der Detektionswahrscheinlichkeit von Bedeutung. Demzufolge beträgt die Detektionswahrscheinlichkeit gemäß der Mithörschwellddefinition 50%, wenn $r_{\text{Int},\nu}$ den Schwellwert erreicht. Bei einer Unterschreitung ist die Wahrscheinlichkeit geringer, bei einer Überschreitung des Schwellwerts größer als 50%.

Für die Wahrnehmungsschwellegenerierung werden die Ausgänge der Schwellendetektoren parallel ausgewertet. Die Wahrnehmungsschwelle wird dabei durch die Modellsektion bestimmt, in welcher der Schwellwert zuerst erreicht wird.

Bei der Generierung von Wahrnehmungsschwellen wird der Grad der Wahrnehmbarkeit d bestimmt, wie durch (3.7) und in Bild 3.18 angegeben. Der Grad der Wahrnehmbarkeit von Signaländerungen bei Verwendung identischer Eingangssignale führt auf die Werte $r_{\text{Int},\nu} = 1$ und $d_\nu = 0$. An der Wahrnehmungsschwelle unterschiedlicher Signale gilt $r_{\text{Int},\nu} = r_{\text{S},\nu}$ und $d_\nu = 1$. Bei Überschreitung der Wahrnehmungsschwelle ergeben sich Werte von $d_\nu > 1$. Der Grad der Wahrnehmbarkeit d ist nicht zu verwechseln mit der Sensitivität d' aus Kapitel 2.6.

$$d_\nu = \frac{r_{\text{Int},\nu} - 1}{r_{\text{S},\nu} - 1} \quad (3.7)$$

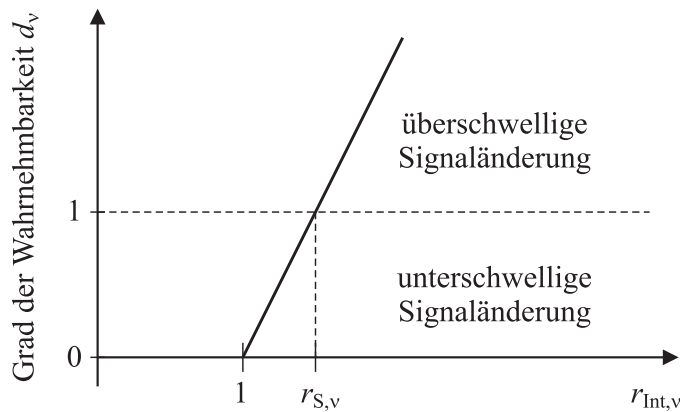


Bild 3.18 Charakteristik des Schwellendetektors einer Sektion mit dem Schwellwert $r_{S,v}$ und der Eingangsgröße $r_{Int,v}$.

3.4.5 Nachbildung der Abhängigkeit der Verdeckung von der Einhüllendenfluktuation

Die Wahrnehmbarkeit einer Signaländerung hängt u. a. von dem Signalcharakter des Maskierers ab (vgl. Kapitel 2.4.2). Viele Verdeckungswirkungen sind dabei auf eine unterschiedliche Fluktuation der Signaleinhüllenden zurückzuführen. Diese Beobachtung steht im Einklang mit der Signaldetektionstheorie (vgl. Kapitel 2.6), die eine Verdeckungszunahme mit zunehmender Varianz der Signalleistung vorhersagt. Im Gehörmodell werden die Fluktuationen der Signaleinhüllenden in der neuronalen Verarbeitungsstufe jeder Modellsektion durch einen Fluktuationsdetektor gemessen. Der in Bild 3.19 dargestellte Detektor einer Sektion wird von dem Eingangssignal des NV-Modells mit überlagertem internen Rauschen $u_{HN,v}$ gespeist. Die zeitabhängigen Minima und Maxima werden über den jeweiligen Detektor aus dem Signal herausgefiltert und klingen aufgrund des Tiefpasses erster Ordnung in Abhängigkeit vom Eingangssignal näherungsweise exponentiell mit den Zeitkonstanten $\tau_{DMIN,v}$ bzw. $\tau_{DMAX,v}$ ab. Das Verhältnis der Ausgangsgrößen des Minimum- und Maximumdetektors wird von der Konstanten eins subtrahiert und bildet nach anschließender Tiefpaßfilterung die gemessene Fluktuation $f_{l,v}$. Der verwendete Tiefpaß wird durch eine Diode kurzgeschlossen, wenn das Ausgangssignal vom Eingangssignal überschritten wird. Daher folgt der Ausgang einem abnehmenden Minimum-zu-Maximum Verhältnis unmittelbar,

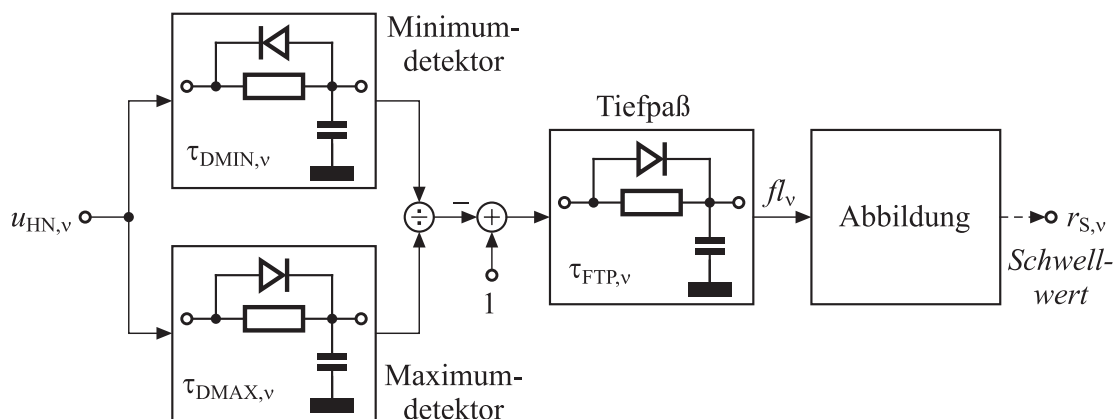


Bild 3.19 Fluktuationsdetektor einer Modellsektion

während bei einer Zunahme die Zeitkonstante $\tau_{FTP,v}$ wirkt. Mögliche Werte der Fluktuation fl_v liegen im Bereich von $0 \leq fl_v \leq 1$, bei Signalen ohne Fluktuation bzw. mit maximaler Fluktuation.

Die Einhüllendenfluktuation wird für die Adaption des vom Schwellendetektor verwendeten Schwellwertes in jeder Sektion bestimmt, um die daraus resultierenden unterschiedlichen Empfindlichkeit nachzubilden. Die entsprechende Abbildungsfunktion ist in Bild 3.20 grafisch dargestellt. Der entsprechende analytische Zusammenhang und die verwendeten Parameter sind in Anhang A.9.3 angegeben.

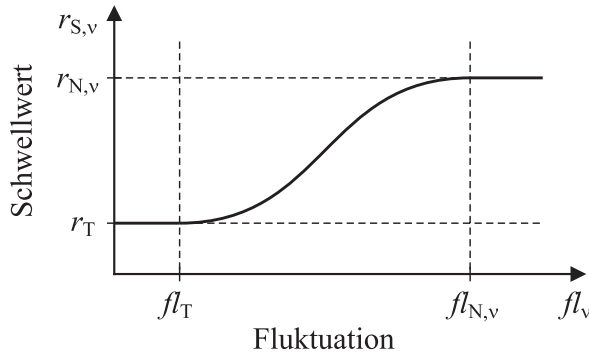


Bild 3.20 Abbildungsfunktion der gemessenen Fluktuation fl_v auf den Schwellwert $r_{S,v}$.

3.5 Generierung von Wahrnehmungsschwellen

Die von einem Testsignal verursachte Signaländerung befindet sich an der Wahrnehmungsschwelle, sobald der Eingangswert eines Schwellendetektors den zugeordneten Schwellwert erreicht hat. Da von dem modellinternen Schwellwert aufgrund der Nichtlinearität nicht ohne weiteres auf den dazu notwendigen Testsignalpegel zurückgeschlossen werden kann, wird der Testsignalpegel durch ein Iterationsverfahren bestimmt. Bei diesem Verfahren wird in jedem Iterationsschritt zunächst die Sektion \hat{v} ermittelt, in welcher der Grad der Wahrnehmbarkeit $d_{\hat{v}}$ gemäß (3.7) zu einem beliebigen Zeitpunkt während der Signaldauer maximal ist. Diese Sektion ist der Kandidat für die Erreichung des Schwellwerts durch das Eingangssignal im nächsten Iterationsschritt, wobei eine Unterschreitung der Schwelle in allen anderen Sektionen erwartet wird. Demzufolge wird die Testsignalamplitude A_T in geeigneter Weise entsprechend der Distanz zwischen dem Eingangssignal und dem Schwellwert des Schwellendetektors für den nächsten Iterationsschritt $i + 1$ eingestellt. Das Maskiersignal muß dabei in jedem Iterationsschritt identisch sein. Eine geeignete Formel zur iterativen Pegeleinstellung ist mit (3.8) gegeben:

$$A_T(i + 1) = A_T(i) d_{\hat{v}}^{-0,7} \quad (3.8)$$

Dieser Mechanismus ist nicht als Teil des Gehörmodells anzusehen. Er ist hingegen mit einer Apparatur zur automatischen Testsignalpegeleinstellung in psychoakustischen Hörtests vergleichbar. Ein sinnvolles Abbruchkriterium für die Iteration ist erfüllt, wenn das Eingangssignal des Schwellendetektors der Sektion \hat{v} mit dem Schwellwert innerhalb eines vorgegebenen Toleranzbereichs übereinstimmt.

4 Verifikation des Gehörmodells

Für die Verifikation des Gehörmodells werden Daten aus psychoakustischen Messungen als Referenz herangezogen. Darüber hinaus dienen physiologische Meßdaten der Überprüfung des physiologischen Teils des Gehörmodells. Die zum Parameterabgleich verwendeten Meßdaten bilden eine Teilmenge der in diesem Kapitel wiedergegebenen Ergebnisse. Auf die zum Abgleich verwendeten Meßdaten wird daher jeweils hingewiesen.

4.1 Vergleich von Simulationsergebnissen und physiologischen Meßergebnissen

Physiologische Messungen am Innenohr sind in der Regel nicht am Menschen in vivo durchführbar. Die Meßergebnisse von Säugetieren können aber aufgrund der vergleichbaren Gehörphysiologie weitgehend auf den Menschen übertragen werden. Diese Meßergebnisse, die zum Teil in Kapitel 2.2 dargestellt sind, werden hier für die Überprüfung des physiologischen Teils des Gehörmodells verwendet.

4.1.1 Schallinduzierte Schwingung der cochleären Trennwand

Eine Möglichkeit zur Schwingungsmessung der cochleären Trennwand ist der Einsatz eines Laserinterferometers an einem Ort der Cochlea. Durch Variation der Frequenz und des Pegels eines Sinustons am Eingang des Gehörs können somit die grundlegenden Übertragungseigenschaften der Cochlea bestimmt werden. Damit sind Rückschlüsse über die Ausprägung der Wanderwelle und die Dynamikkompression möglich. Wird zusätzlich ein zweiter Sinuston mit fester Frequenz und festem Pegel verwendet, können auch die entstehenden Verzerrungsprodukte sowie die Suppression nachgewiesen werden [165].

4.1.1.1 Wanderwelle

Die ortsabhängige Einhüllende der von einem Sinuston erzeugten Wanderwelle in der Simulation ist in Bild 4.1 für eine Frequenz von 1 kHz dargestellt. Dazu wird die IHZ-Stimulation $u_{\text{IHZ},\nu}$ parallel in allen Modellsektionen bestimmt. Diese Stimulation korrespondiert zur örtlichen Schnelle der cochleären Trennwand und ist daher mit der physiologischen Messung aus Bild 2.9 vergleichbar. Die in Bild 2.9 dargestellte BM-Schnelle wurde jedoch an einem festen Ort für eine variable Sinustonfrequenz gemessen, so daß sich prinzipiell ein horizontal gespiegelter Verlauf ergibt. Außerdem führt dort der gemeinsame Bezugspegel aller Kurven am Helicotrema zu einer vertikalen Verschiebung gegenüber der Darstellung in Bild 4.1. Die Ausbreitung der Wanderwelle bezüglich der Darstellung in Bild 4.1 erfolgt von rechts nach links, da die im Bild oben angegebene Sektion 1 das Helicotrema nachbildet.

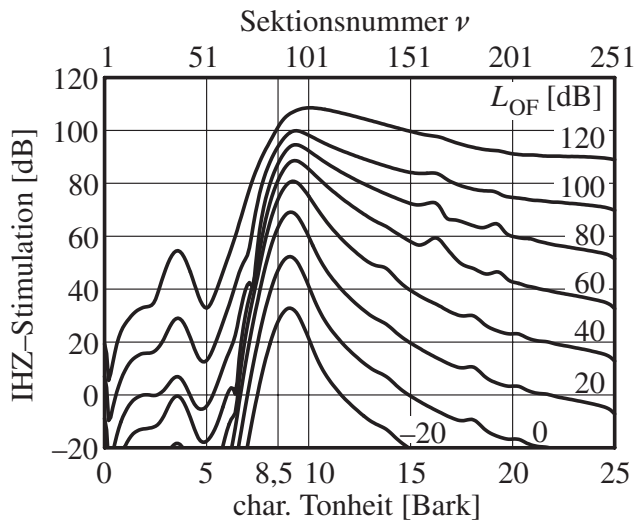


Bild 4.1 IZH-Stimulationspegel am Eingang des IZH-Modells in jeder Sektion über der charakteristischen Frequenz, die als Tonheitsskala dargestellt ist. Am Eingang des HM-Modells wird ein 1 kHz (8,5 Bark) Sinuston mit dem Pegel L_{OF} eingespeist.

Die Frequenzselektivität der Cochlea wird durch die Breite des Resonanzmaximums bestimmt, die sowohl in Abhängigkeit von der Sektionsnummer als auch in Abhängigkeit von der zugeordneten charakteristischen Frequenz (CF) angegeben werden kann. Im Gehörmodell wird die Resonanzfrequenz des Parallelschwingkreises $f_{Res,\nu}$ in jeder Sektion des HM-Modells als CF definiert, der eine charakteristische Tonheit gemäß (A.41) zugeordnet ist. Mit zunehmendem Schalldruckpegel nimmt die Frequenzselektivität ab, da die Verstärkung durch die ÄHZs zunehmend in den Sättigungsbereich kommt und damit zu einer verringerten Resonanzgüte führt. Dieser Mechanismus bewirkt am Ort der Resonanz eine Dynamikkompression, die, wie am Beispiel in Bild 4.1 erkennbar, zu einer Reduzierung der Eingangssignaldynamik von 140 dB auf einen Dynamikbereich von ca. 80 dB am Ausgang der Modellsektion mit maximaler Stimulation führt. Die kleineren Resonanzen oberhalb des Maximums werden durch die kubischen Verzerrungen hervorgerufen, die bei ungeradzahligem Vielfachen der Eingangsfrequenz auftreten. Verzerrungen mit Frequenzen oberhalb der Nyquistfrequenz führen zu Aliasing-Artefakten, die bei Frequenzen unterhalb des Maximums beobachtet werden. Für Eingangspegeln von weniger als 100 dB liegt der Pegel dieser Artefakte jedoch unterhalb der Ruheshwelle, so daß keine nennenswerte Auswirkung auf die Wahrnehmungsschwellennachbildung vorhanden ist.

In Bild 4.2 ist der Pegel der ortsabhängigen IZH-Stimulation aus dem Gehörmodell für unterschiedliche Sinustonfrequenzen und Pegel von 0 dB bzw. 100 dB dargestellt. Daraus geht hervor, daß die Filtercharakteristik über einen weiten mittleren Frequenzbereich unabhängig vom Ort bzw. von der Signalfrequenz ist. Erst bei sehr niedrigen und sehr hohen Frequenzen ist eine leichte Pegelzunahme bei 100 dB, bzw. eine Abnahme des Ausgangspegels bei 0 dB Eingangspegel festzustellen. Insbesondere bei hohen Frequenzen ist die Verstärkung der ÄHZs wegen dem inhomogenen Abschluß am Helicotrema des HM-Modells weniger wirksam. Eine unrealistisch hohe Resonanzgüte bei geringem Signalpegel wird durch die laterale Kopplung der Modellsektionen vermieden, so daß eine Übereinstimmung mit Meßergebnissen der BM-Schwingung erreicht wird.

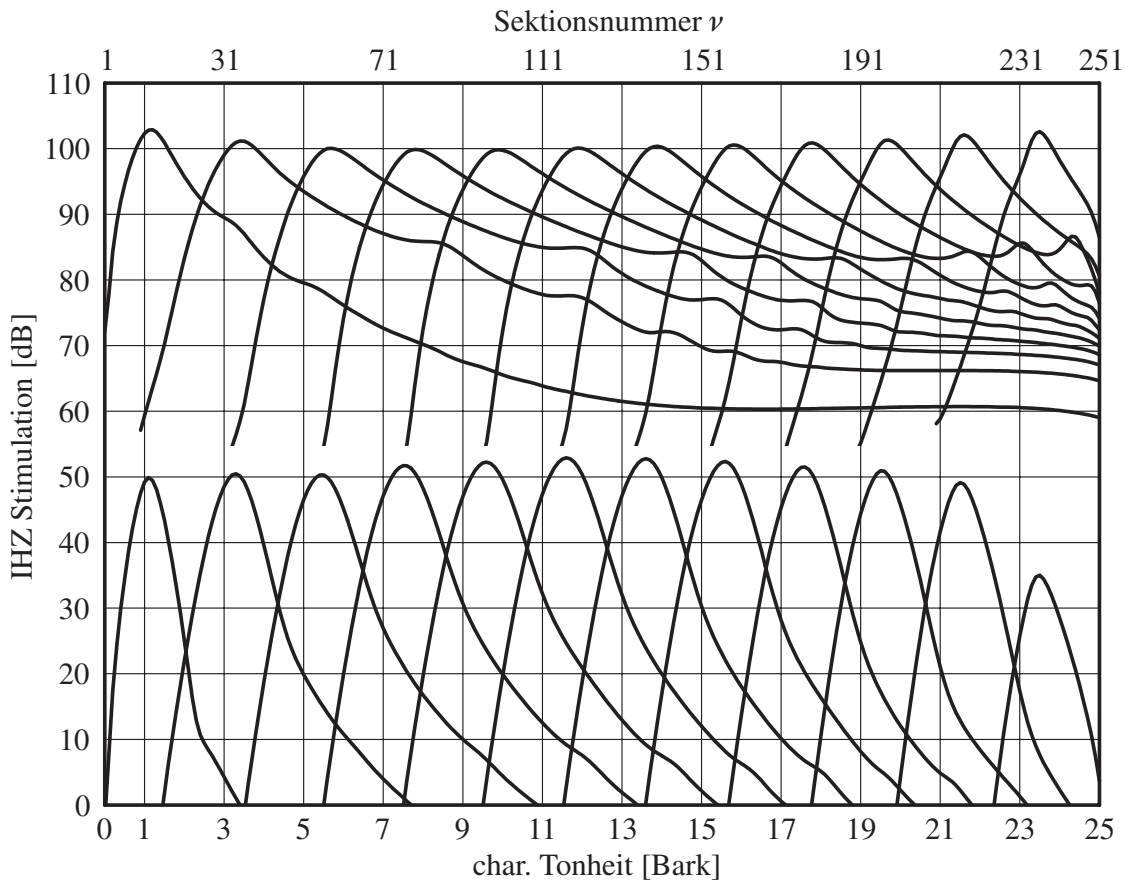


Bild 4.2 IZH-Stimulationspegel aller Modellsektionen für eine Sinusschwingung am Eingang des HM-Modells. Die Tonheit des Sinus beträgt 1, 3, 5, ..., 23 Bark und bildet sich in dem Maximum der jeweiligen Kurve ab, das in einer Modellsektion mit einer etwas größeren Resonanztonheit (Abszisse) auftritt. Die *obere* Kurvenschar ergibt sich bei 100, die *untere* bei 0 dB Eingangssignalpegel am ovalen Fenster.

Die Resonanzgüte der IZH-Stimulation läßt sich näherungsweise aus den Kurven in Bild 4.2 ermitteln. Dazu wird der Abstand der charakteristischen Tonheiten, bei denen der Stimulationspegel um 10 dB gegenüber dem Maximum abgefallen ist, als Bandbreite zur Berechnung der Resonanzgüte $Q_{10\text{ dB}}$ verwendet. Der Wert der Resonanzgüte ergibt sich aus der Division der Eingangssignalfrequenz durch die Bandbreite in Hz. In Bild 4.3 sind die Werte im Vergleich zu Meßdaten aus Tuningkurven der Entladungsraten auf den Hörnervenfasern von

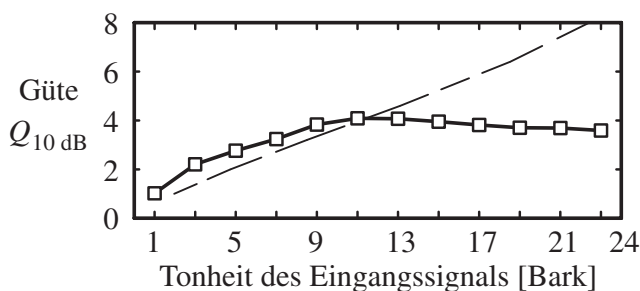


Bild 4.3 Resonanzgüte (\square) der simulierten IZH-Stimulation, gemessen an der Bandbreite des Frequenzgangs des IZH-Stimulationspegels bei unterschiedlichen Eingangssignaltonheiten und 0 dB SPL. — — Meßdaten [155: Fig. 4.5A]

Katzen dargestellt [155: Fig. 4.5A]. Die Meßdaten sind aufgrund der nichtlinearen Kennlinie der inneren Haarzellen (vgl. Bild 2.11) nicht unmittelbar vergleichbar, aber es ist zu erwarten, daß die Resonanzgüte der IHZ-Stimulation in der gleichen Größenordnung liegen muß. Darüber hinaus sind die Meßdaten mit einer großen Streuung behaftet. Die Resonanzgüten des Gehörmodells weisen Werte auf, die innerhalb des Streubereichs der Messung liegen.

4.1.1.2 Dynamikkompression

Die Eingangs-/Ausgangsfunktion (I/O-Funktion) der Cochlea ergibt sich durch Auswertung der Maximalpegel der Kurven in Bild 4.1 in Abhängigkeit vom Eingangspegel. Der so bestimmte Zusammenhang ist in Bild 4.4 dargestellt. Zum Vergleich ist die lineare Abhängigkeit für das passive Modell (für $g_1 = g_2 = 0$) der Cochlea mit der Steigung 1:1 gestrichelt eingezeichnet. An der Differenz der beiden Kurven kann bei sehr kleinen Eingangspegeln die maximale Verstärkung im Vergleich zum passiven Cochlea-Modell abgelesen werden, die hier etwa 62 dB beträgt. Die Kompression der Eingangssignaldynamik schlägt sich in der Steigung der I/O-Funktion nieder, die bei mittleren Eingangspegeln ein Minimum von ca. 0,33 dB/dB aufweist (vgl. Gerade mit Steigung 1/3). Zum Vergleich sind in Bild 4.4 die gemessenen Schnellepegel der BM von verschiedenen Versuchstieren [164] mit eingezeichnet. Die beobachtete Kompression in der Messung und Simulation stimmt weitgehend überein. Die Meßkurven repräsentieren eine mittlere Steigung, während in einigen anderen Messungen [144, 189] eine minimale Steigung von etwa 0,2 dB/dB beobachtet wurde.

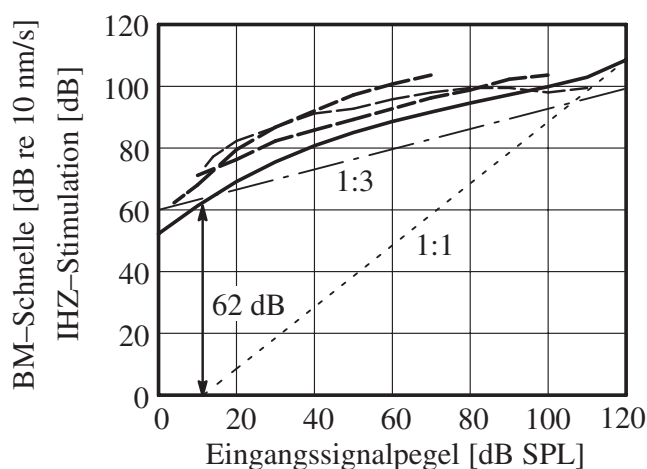


Bild 4.4 Pegelabhängigkeit der IHZ-Stimulation am Ort der BM-Resonanz eines 1 kHz Sinustones im Vergleich zu Messungen der BM-Schnelle verschiedener Versuchstiere [164].

— — gemessene BM-Schnelle
 ——— simulierte IHZ Stimulation
 ····· passives Verhalten

4.1.1.3 Suppression

Die von einem Sinustestton verursachte maximale IHZ-Stimulation am Ort seiner Resonanz im Gehörmodell kann durch einen überlagerten Suppressorton vermindert werden [7, 143, 171]. Dieser Suppressionseffekt ist auf die nichtlineare Interaktion von Suppressor und Sinustestton zurückzuführen. Er resultiert im Cochlea-Modell aus der vergrößerten Sättigung der ÄHZ-Verstärkungen bei überlagertem Suppressor, die eine verminderte Verstärkung des Sinustesttons bewirkt [14, 13]. Suppressionsmessungen werden in der Regel als sog. „Suppressionstuningkurven“ angegeben, die den Suppressorpegel als Funktion der Suppressorfrequenz zur Erzielung einer konstanten Testtonsuppression bezeichnen. Die

Suppression wird dabei an dem Ort gemessen, an dem der Testton allein zu der maximalen IHZ-Stimulation führt. Die Signale werden am Eingang des Gehörmodells eingespeist. Bild 4.5 zeigt die entsprechende Darstellung für einen Testton mit 3 kHz und 50 dB SPL aus der Simulation. In den beiden äußeren grau dargestellten Bereichen wird die IHZ Stimulation durch den Suppressor auf weniger als 80% reduziert. Für Suppressorfrequenzen nahe der Testtonfrequenz erhöht sich demgegenüber die Stimulation, so daß der Suppressor zum Erreger wird. Suppressoren innerhalb der mittleren grauen Fläche führen zu einer Zunahme der Stimulation auf mehr als 200%, daher wird dieser Bereich mit „Erregung“ bezeichnet. Das Ergebnis zeigt, daß beide Ursachen der Simultanverdeckung, die nach Kapitel 2.5 in der Erregung und Suppression bestehen, vom Modell berücksichtigt werden. Die nachgebildete Suppression steht im Einklang mit physiologischen Meßergebnissen an der IHZ [171] und dem Hörnerv [7].

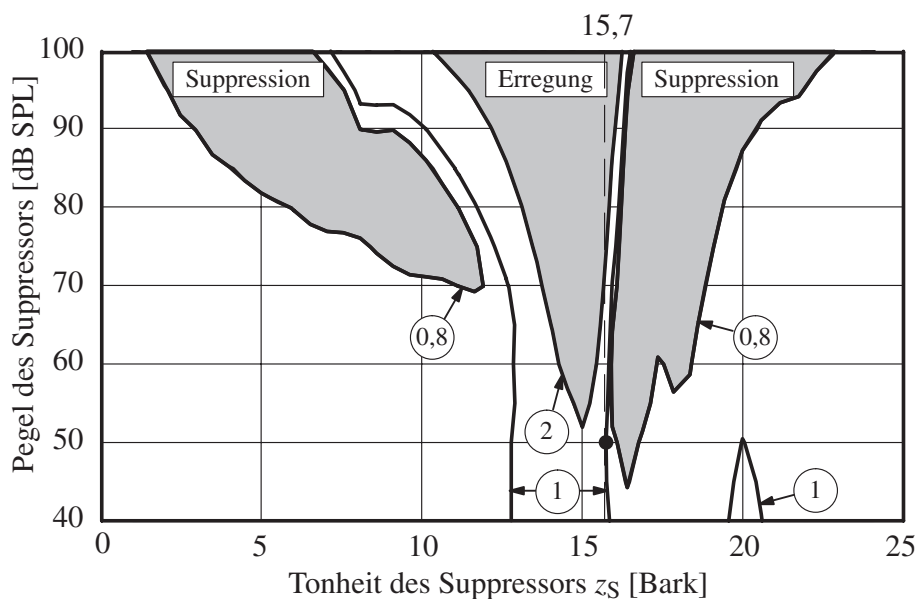


Bild 4.5 Simulierte Suppressionstuningkurven für einen Sinustestton bei 3 kHz (15,7 Bark) und $L_T = 50$ dB SPL. Die durchgezogenen Kurven gleicher Testtonsuppression sind mit den Suppressionsfaktoren gekennzeichnet. Ein Faktor von 0,8 entspricht einer Verminderung der IHZ-Stimulation auf 80% durch den Suppressor, ein Faktor von 2 die Zunahme auf 200%.

In Bild 4.6 ist eine gemessene Suppressionstuningkurve der Entladungsrate einer Hörnerfaser gemäß [7] dargestellt. Die grau dargestellten Suppressionsbereiche werden wie bei den Simulationsergebnissen für Suppressorfrequenzen unterhalb und oberhalb des Testsignals beobachtet. Die Simulation (vgl. Bild 4.5) führt zu Suppressionsbereichen, deren Größe und Form mit den Meßergebnissen vergleichbar ist.

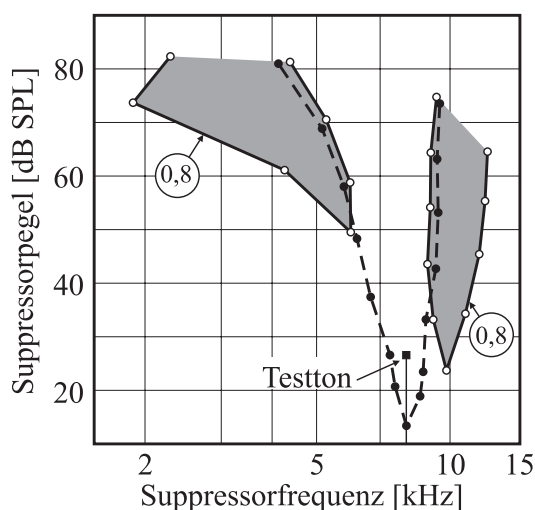


Bild 4.6 Gemessene Suppressionstuningkurven der mittleren Entladungsrate der Faser des Hörnervs mit der CF des Testtons nach [7]. Die mit dem Faktor 0,8 gekennzeichnete Kurve gibt den Pegel und die Frequenz des Suppressortons an, der zu einer Verminderung der Entladungsrate auf 80% führt. Innerhalb der grau dargestellten Flächen ist die Suppression größer, d. h. die Entladungsrate vermindert auf weniger als 80%. Die Tuningkurve der Faser ist *gestrichelt* dargestellt.

4.2 Vergleich von Simulationsergebnissen und psychoakustischen Meßergebnissen

Die Leistungsfähigkeit des Gehörmodells in Bezug auf die Wahrnehmungsschwellennachbildung soll im folgenden mit Hilfe von psychoakustischen Referenzdaten beurteilt werden. Der Vergleich von Simulationsergebnissen mit gemessenen Wahrnehmungsschwellen zeigt, welche Wahrnehmungseigenschaften vom Gehörmodell qualitativ nachgebildet werden und in welchem Maß eine quantitative Übereinstimmung erzielt wird.

4.2.1 Lautheit

Im Gehörmodell wird die spezifische Lautheit N' innerhalb des NV-Modells jeder Sektion mit der Nummer ν durch die Größe $u_{N',\nu}$ nachgebildet. Durch Summation dieser Größe über alle Sektionen kann die Lautheit eines Eingangssignals bestimmt werden. Die Lautheit wird in Sone angegeben, wobei 1 Sone üblicherweise als Lautheit eines 1 kHz Sinustons bei 40 dB SPL definiert wird. Für einen 1 kHz Sinuston bzw. ein breitbandiges Rauschen sind die Simulationsergebnisse im Vergleich zu Messungen in Bild 4.7 angegeben. Die bei gleichem Signalpegel größere wahrgenommene Lautheit des Rauschens wird qualitativ, aber nicht in vollem Umfang nachgebildet. Die Kompression bei der Lautheitswahrnehmung wird vom Gehörmodell ebenfalls nicht ganz erreicht, was sich durch die größere Steilheit der simulierten Lautheitskurven gegenüber der Messung zeigt. Durch eine erhöhte Dynamikkompression des Gehörmodells mit einer entsprechend vergrößerten Maximalverstärkung durch die ÄHZ Modelle könnte die Übereinstimmung hier verbessert werden. Die Übereinstimmung mit anderen Messungen, wie zum Beispiel der I/O-Funktion (vgl. Bild 4.4), würde dadurch jedoch vermindert. Zur Erzielung einer erhöhten Dynamikkompression müßte die maximale Verstärkung der ÄHZ-Modelle über 60 dB hinaus vergrößert werden, was ebenfalls nicht im Einklang mit physiologischen Meßergebnissen steht. Die Abweichung in den Simulationsergebnissen sollte jedoch nicht überbewertet werden, da die Ergebnisse von Lautheitsmessungen signifikant von der Meßmethode abhängen und mit einer nennenswerten Streuung behaftet sind.

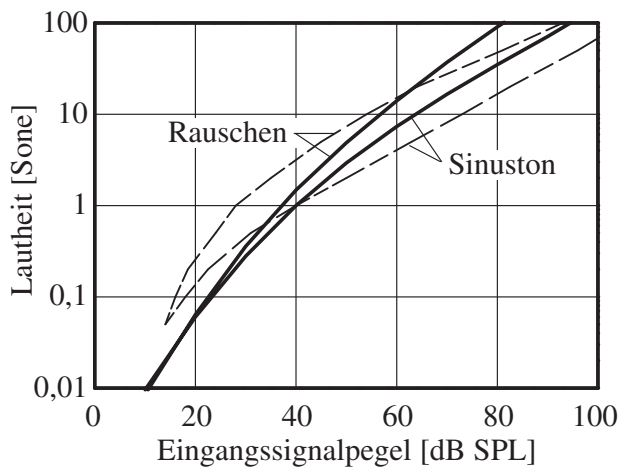


Bild 4.7 Lautheit eines 1 kHz Sinustons und eines gleichmäßig verdeckenden Rauschens mit einer Bandbreite von 16 kHz (vgl. [197]).

—————
Gehörmodell

Messung [197: Fig. 8.4]

4.2.2 Wahrnehmungsschwellen

Die nachgebildeten Wahrnehmungsschwellen umfassen die für die Audiocodierung wesentlichen Ruhehörschwellen und Mithörschwellen. Darüber hinaus werden Detektionsschwellen für die Änderung eines Parameters des Eingangssignals nachgebildet, wie den Parameter Modulationsindex bzw. Modulationsgrad eines Trägersignals. Die Verifikation erfolgt anhand von psychoakustischen Meßergebnissen aus der Literatur und eigenen Messungen.

4.2.2.1 Ruhehörschwellen

Die simulierte Ruhehörschwelle eines Sinustons ist in Bild 4.8 unterschiedlichen Meßdaten gegenübergestellt. Die durch ein Kreuz (\times) gekennzeichneten Werte stellen den Medianwert der eigenen Messungen von 8 Testpersonen dar. Das verwendete Meßverfahren ist in Anhang A.10 beschrieben. Die Streuung der Meßwerte wird mit dem als vertikaler Balken gekennzeichneten Interquartilbereich angegeben, der den Pegelbereich umfaßt, in dem sich 50% der Meßwerte befinden [197]. Dabei sind jeweils 25% der Meßwerte größer bzw. kleiner als der Interquartilbereich. Die Meßdaten wurden zum Abgleich der Ruhehörschwelle verwendet. Zum Vergleich sind auch einige Daten aus der Literatur mit dargestellt. Die Einstellung der Ruhehörschwelle anhand des „internen“ Rauschpegels erfolgte mit der Maßgabe, daß die Meßdaten angenähert, aber nicht überschritten werden. Die Simulation zeigt insgesamt einen glatteren Verlauf als die gemessenen Schwellen. Die relativ hohen gemessenen Werte zwischen 18 und 20 Bark treten in der Simulation jedoch nicht auf. Ruhehörschwellen eines Rauschsignals mit einer Bandbreite bis zur Frequenzgruppenbreite weisen gegenüber Sinustönen kaum abweichende Werte in der Messung [53] bzw. Simulation auf und werden hier daher nicht weiter betrachtet. Die in den Simulationen verwendeten Fensterfunktionen zur zeitlichen Begrenzung der Eingangssignale sind im Anhang A.2 erläutert.

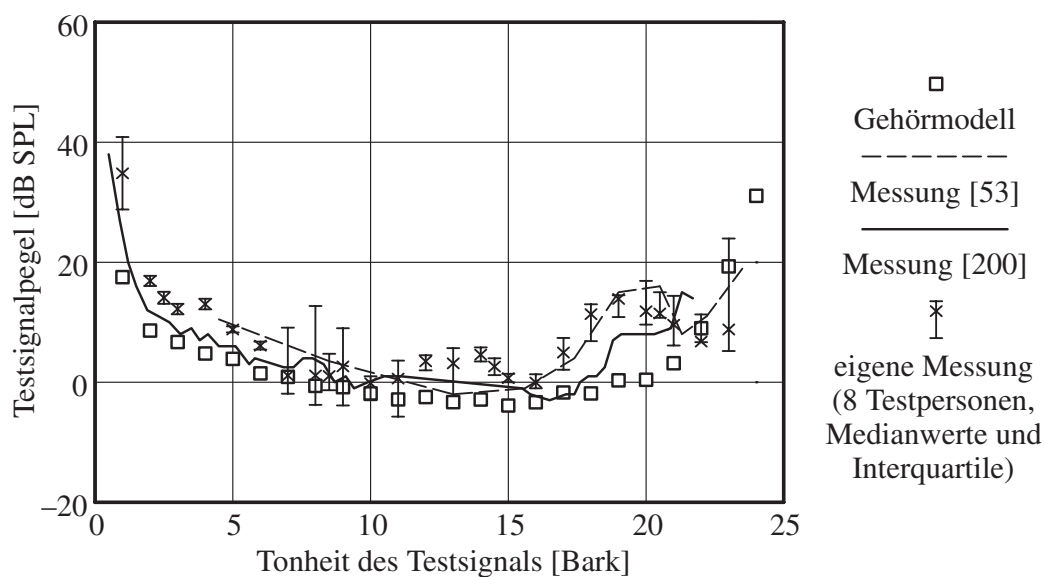


Bild 4.8 Ruhehörschwelle eines Sinustestsignals. Parameter der Simulation: $T_T = 200$ ms, $T_{F,T} = 5$ ms.

4.2.2.2 Mithörschwellen für spektrale Verdeckung

Im folgenden sind Mithörschwellen aus Messungen der Simultanverdeckung im Vergleich zu Simulationsergebnissen angegeben. Die Zeitabhängigkeit der Verdeckung wird dabei nicht betrachtet, sondern ausschließlich die Frequenzabhängigkeit. Zum Parameterabgleich des Gehörmodells, insbesondere dem Teilmodell der Cochlea, wurden Mithörschwellen von einem maskierenden frequenzgruppenbreiten (1 Bark breiten) Rauschen bei der Mittenfrequenz von 1 kHz verwendet. Gegenüber den Mithörschwellen eines Sinustonmaskierers zeichnen sich diese Meßergebnisse durch eine größere Konsistenz der Meßergebnisse mehrerer Versuchspersonen aus. Die in Bild 4.9 dargestellten Simulationsergebnisse zeigen den typischen asymmetrischen Verlauf der Mithörschwelle über der Tonheit des Sinustesttons. Die pegelabhängige Flankensteilheit der flacheren Flanke oberhalb der Maskiererfrequenz zeigt gerade in dem für die Audiocodierung wichtigen Bereich mittlerer Maskiererpegel eine sehr gute Übereinstimmung mit der Messung (vgl. [12]).

Bei der Messung [197] wurde die Wahrnehmung von Verzerrungsprodukten zur Detektion eines vorhandenen Testsignals („*off-frequency listening*“) unterbunden. Das Gehörmodell bildet diese Einschränkung dadurch nach, daß nur in denjenigen Modellsektionen eine Signaländerung detektiert wird, die sich in der Nähe des Resonanzmaximums des Testsignals befinden. Diese Nachbildung wird im folgenden als *selektive Detektion* bezeichnet. Dabei werden nur die Modellsektionen für die Detektion einer Signaländerung herangezogen, deren Resonanztonheit (charakteristische Tonheit) im Bereich von 1 Bark unterhalb bis 2,5 Bark oberhalb der Testsignaltonheit liegt.

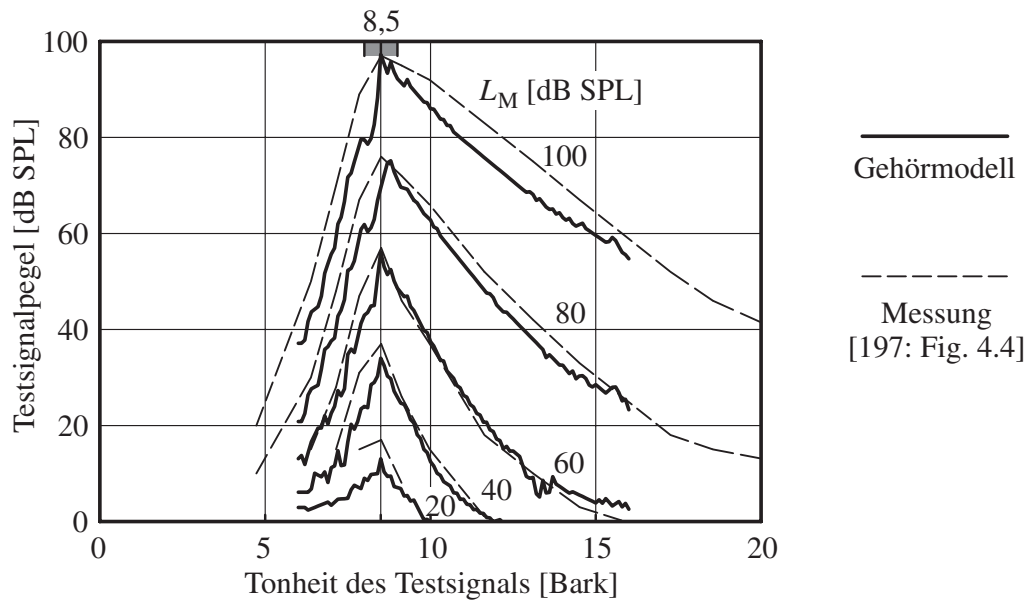


Bild 4.9 Mithörschwelle eines Sinustestsignals verdeckt durch Gaußsches frequenzgruppenbreites Rauschen der Mittenfrequenz 1 kHz (8,5 Bark). Parameter der Simulation: $T_M = 300$ ms, $T_T = 200$ ms, $T_{F,T} = 5$ ms, $B_M = 160$ Hz, selektive Detektion.

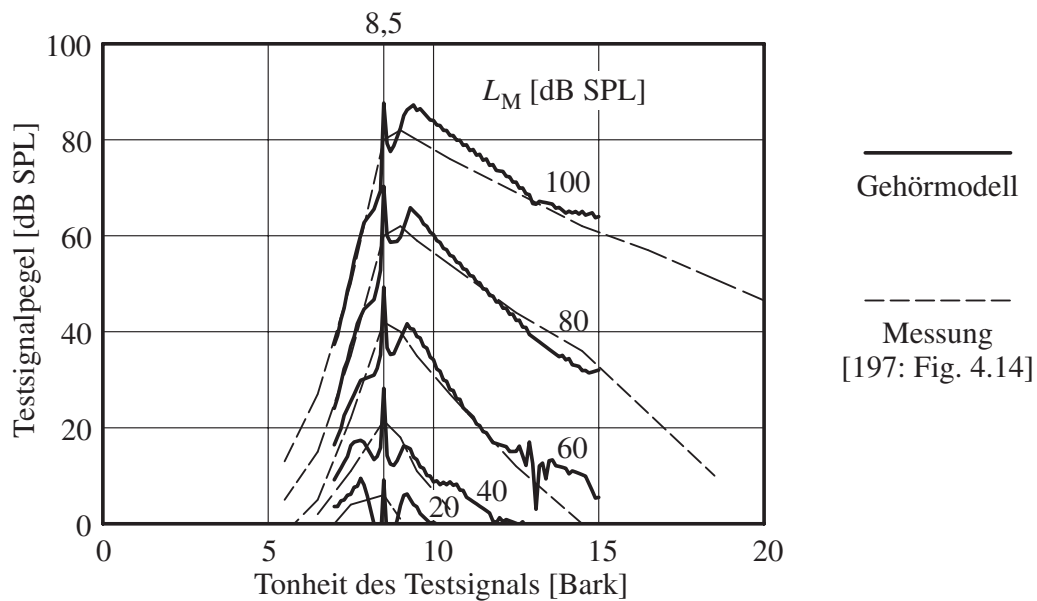


Bild 4.10 Mithörschwelle eines Sinustestsignals verdeckt durch einen Sinuston bei 8,5 Bark (1 kHz). Parameter der Simulation: $T_M = 300$ ms, $T_T = 200$ ms, $T_{F,T} = 10$ ms, selektive Detektion. Die Meßergebnisse wurden durch Mittelung von Daten aus verschiedenen Frequenzbereichen gewonnen.

Zum Abgleich der Schwellwertsteuerung durch den Fluktuationsdetektor wurden die Mithörschwellen eines Sinustestsignals, verdeckt durch einen Sinustonmaskierer verwendet

(Bild 4.10). Gegenüber den Mithörschwellen von Gaußschem Rauschen (Bild 4.9) weist die Messung wesentlich geringere Mithörschwellen bei Sinustonmaskierern gleichen Pegels in der Nähe der Mittenfrequenz auf. Die reduzierten Mithörschwellen werden entsprechend nachgebildet. Aufgrund hörbarer Schwebungen bei Testsignalfrequenzen nahe der Maskierfrequenz bricht die Mithörschwelle in diesem Bereich ein. Bei den dargestellten Meßdaten ist dieser Effekt jedoch vernachlässigt worden.

Die vom Gehörmodell nachgebildeten Mithörschwellen für Sinustonmaskierer wurden anhand von Meßdaten für unterschiedliche Maskierfrequenzen und -pegel verifiziert. Die Ergebnisse sind in Bild 4.11 dargestellt. Bis auf den 13 Bark Maskierer bei 80 dB SPL liegen dabei Maskiererpegel von 20, 40 und 60 dB SPL zugrunde. Das Gehörmodell wurde hier ohne selektive Detektion betrieben, d. h. für die Detektion einer Signaländerung werden alle Modellsektionen ausgewertet. Die vom Gehörmodell detektierten Verzerrungsprodukte, insbesondere bei der kubischen Differenzfrequenz $f_{d3} = 2f_M - f_T$, und Schwebungen („beating“) führen deshalb zu Einbrüchen des Mithörschwellenverlaufs bei hohen Maskiererpegeln. Dieser Effekt tritt bei dem 8,5 Bark (1 kHz) Maskierer mit 60 dB SPL deutlich in Erscheinung. Im Vergleich zu der Simulation in Bild 4.10 mit selektiver Detektion treten hier zusätzliche Minima oberhalb der Maskierfrequenz auf.

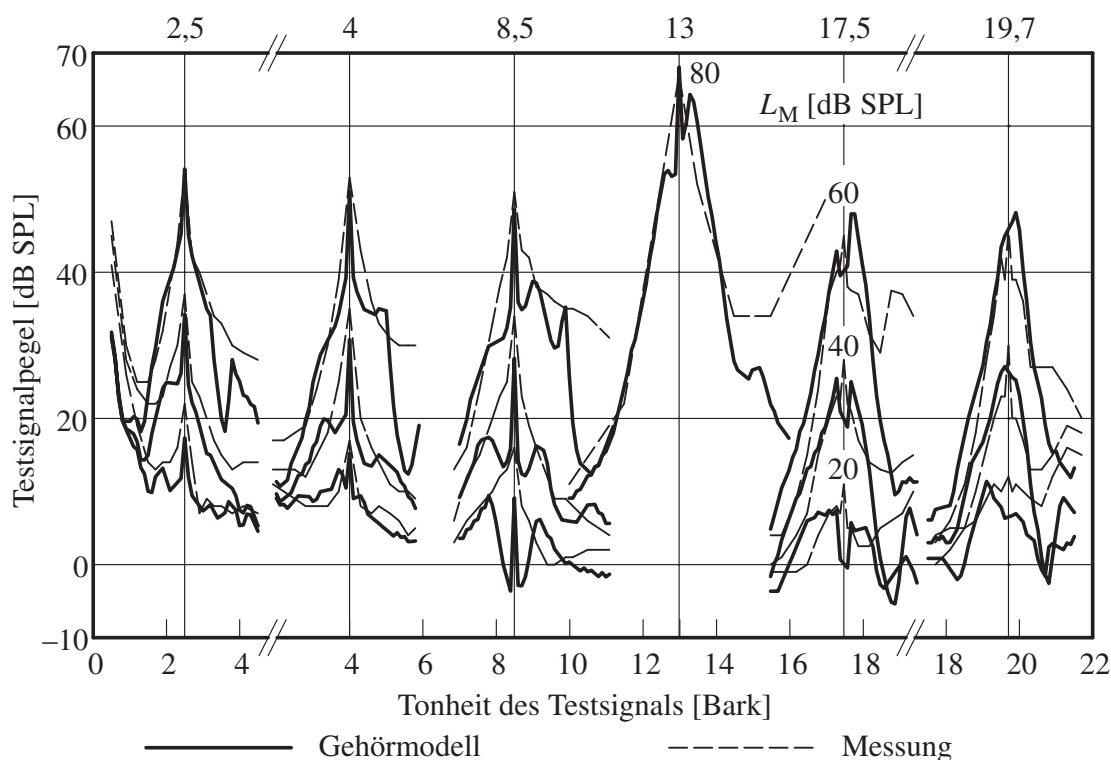


Bild 4.11 Mithörschwelle eines Sinustestsignals verdeckt durch einen Sinuston. Verwendete Maskierertonheit: 2,5; 4; 8,5; 13; 17,5; 19,7 Bark (250, 400, 1k, 2k, 4k, 6 kHz). Maskiererpegel: 80 dB SPL bei 13 Bark, sonst 20, 40, bzw. 60 dB SPL. Parameter der Simulation: $T_M = 300$ ms, $T_T = 200$ ms, $T_{F,T} = 10$ ms. Psychoakustische Daten aus [200], bzw. [197 Fig. 4.37d] für den 13 Bark Maskierer.

Eine Übereinstimmung der Simulationsergebnisse mit den Messungen zeigt sich vor allem für Testsignalfrequenzen unterhalb und bei der Maskierfrequenz. Dies führt zu einer weitgehend übereinstimmenden Flankensteilheit des Mithörschwellenverlaufs in diesem Bereich. Bei größeren Testsignalfrequenzen sind Abweichungen vorhanden, die im wesentlichen von der Detektion von Verzerrungsprodukten im Gehörmodell beeinflusst werden und in der Regel zu verringerten Mithörschwellen führen. Bei der Interpretation ist zu berücksichtigen, daß alle Meßdaten mit einer nennenswerten Varianz behaftet sind, deren Größe jedoch in der Literatur [197, 200] nicht angegeben ist.

Die Verdeckung durch Sinustöne wird im folgenden für ein frequenzgruppenbreites Rauschen als Testsignal dargestellt, das durch Bandpaßfilterung von Gaußschem Rauschen generiert wurde. Für dieses Testsignal treten gegenüber Messungen mit sinusförmigem Testsignal im Bereich der Maskierfrequenz keine so stark ausgeprägten Schwebungen auf. Die Wahrnehmung von Verzerrungsprodukten ist aber nach wie vor gegeben. Die Referenzdaten wurden aus eigenen psychoakustischen Messungen nach der im Anhang A.10 beschriebenen Methode gewonnen. Die Meßdaten des Maskierers mit 80 dB SPL wurden anhand von 8 Testpersonen ermittelt. Die Meßdaten bei 60 dB SPL stammen von 3 und bei 40 dB SPL von 2 Testpersonen. Im Gegensatz zu traditionellen Messungen wurde dabei jeweils ein Testsignal fester Mittenfrequenz und ein Maskierer variabler Frequenz in der Mithörschwellenmessung verwendet. Die Mittenfrequenz des Testsignals betrug dabei 2,5; 8,5; 14,5; bzw. 20,5 Bark. Die Testpersonen sollten beliebige, vom Testsignal hervorgerufene Unterschiede detektieren. In der Messung wurden keine zusätzlichen Signale zur Verdeckung von Verzerrungsprodukten eingesetzt. Die Meßergebnisse sind in Bild 4.12 und nochmals in Bild 4.13 mit den Interquartilen dargestellt. Bei einem Maskiererpegel von 80 dB SPL und Testsignaltonheiten von 14,5 und 20,5 Bark sind deutliche Einbrüche der gemessenen Mithörschwelle für Maskierfrequenzen unterhalb der Testsignalfrequenz zu beobachten, die entsprechend der großen Interquartile eine hohe interindividuelle Streuung aufweisen. Insgesamt stehen die Meßergebnisse im Einklang mit Daten aus der Literatur [131].

Die Nachbildung der Mithörschwellen wurde sowohl mit selektiver Detektion als auch ohne durchgeführt (Bild 4.12 bzw. 4.13). Bei der selektiven Detektion werden nur Modellsektionen mit einer Resonanz nahe der Testsignalfrequenz für die Detektion einer Signaländerung herangezogen. Der Vergleich zeigt, daß sich die selektive Detektion nur bei Maskierfrequenzen unterhalb der Testsignalfrequenz auswirkt und mit zunehmendem Pegel und zunehmender Frequenz größeren Einfluß hat.

Wie bei den vorangegangenen Mithörschwellensimulationen wird eine weitgehende Übereinstimmung mit der Messung für Maskierfrequenzen oberhalb und bei der Testsignalfrequenz festgestellt. Unterhalb der Testsignalfrequenz liegt die simulierte Mithörschwelle bei hohen Maskiererpegeln oberhalb der gemessenen, wenn keine selektive Detektion verwendet wird. Bei selektiver Detektion (Bild 4.13) wird die gemessene Mithörschwelle in diesem Frequenzbereich nicht mehr überschritten. Bei Maskiererpegeln von 80 dB SPL ist das Gehörmodell gegenüber Verzerrungsprodukten sehr empfindlich und führt bei Testsignaltonheiten von 8,5; 14,5 und 20,5 Bark zu einer deutlichen Unterschreitung der gemessenen Mithörschwellen in diesem Bereich.

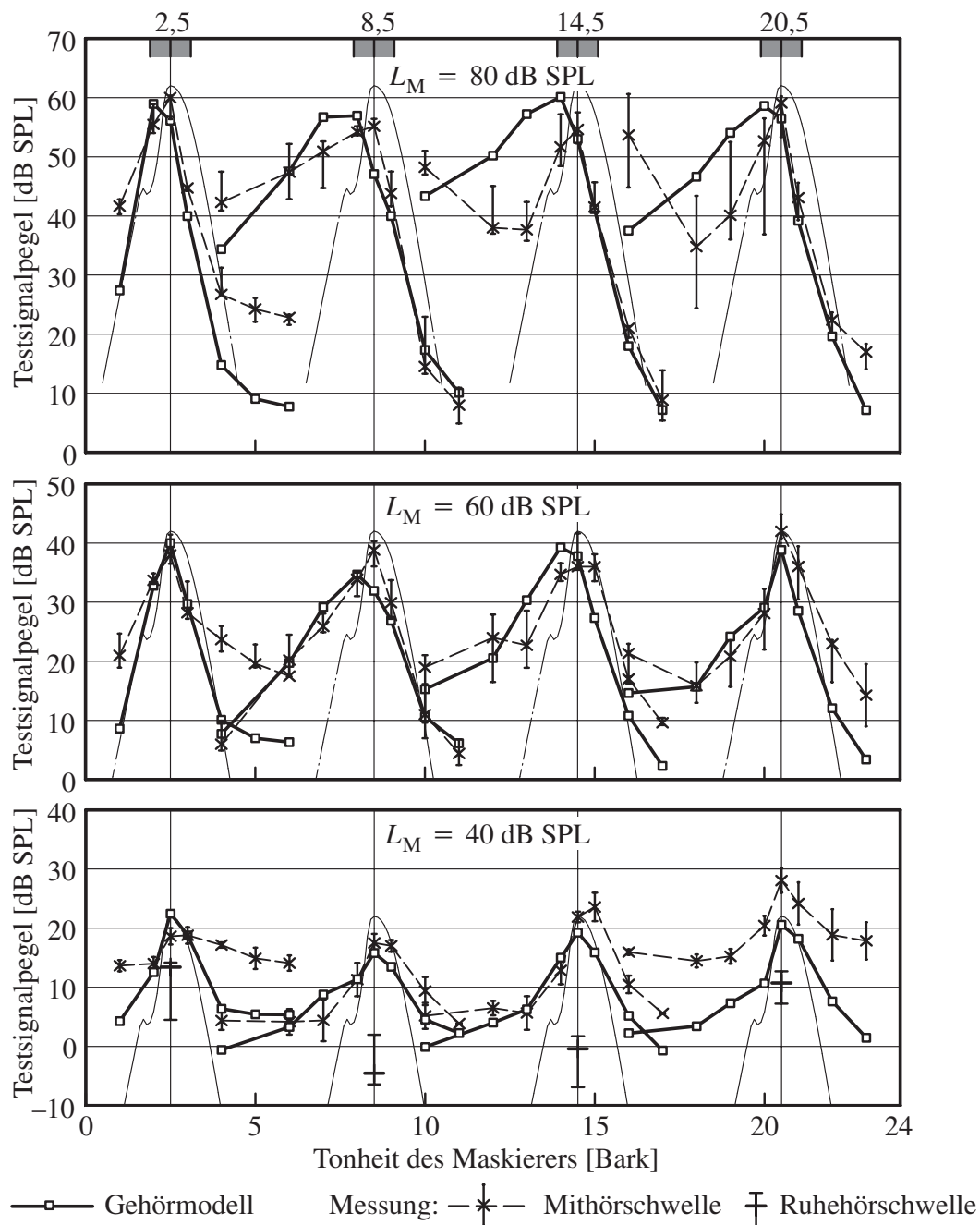


Bild 4.12 Mithörschwelle eines frequenzgruppenbreiten Gaußschen Rauschens verdeckt durch einen Sinuston. Parameter: $z_T = [2,5; 8,5; 14,5; 20,5 \text{ Bark}]$ (250; 1k; 2,5k; 7 kHz), $B_T = 1 \text{ Bark}$ (100, 160, 380, 1300 Hz), $T_T = 400 \text{ ms}$, $T_{F,T} = 10 \text{ ms}$, $L_M = [40, 60, 80 \text{ dB SPL}]$, $T_M = 500 \text{ ms}$. Simulation mit selektiver Detektion. Meßwerte mit Angabe der Medianwerte und Interquartile. Die Ruhehörschwellen des Rauschens sind im unteren Diagramm angegeben. Berechnete Mithörschwellen eines psychoakustischen Modells [101] sind mit einer *dünn durchgezogenen* Linie eingezeichnet.

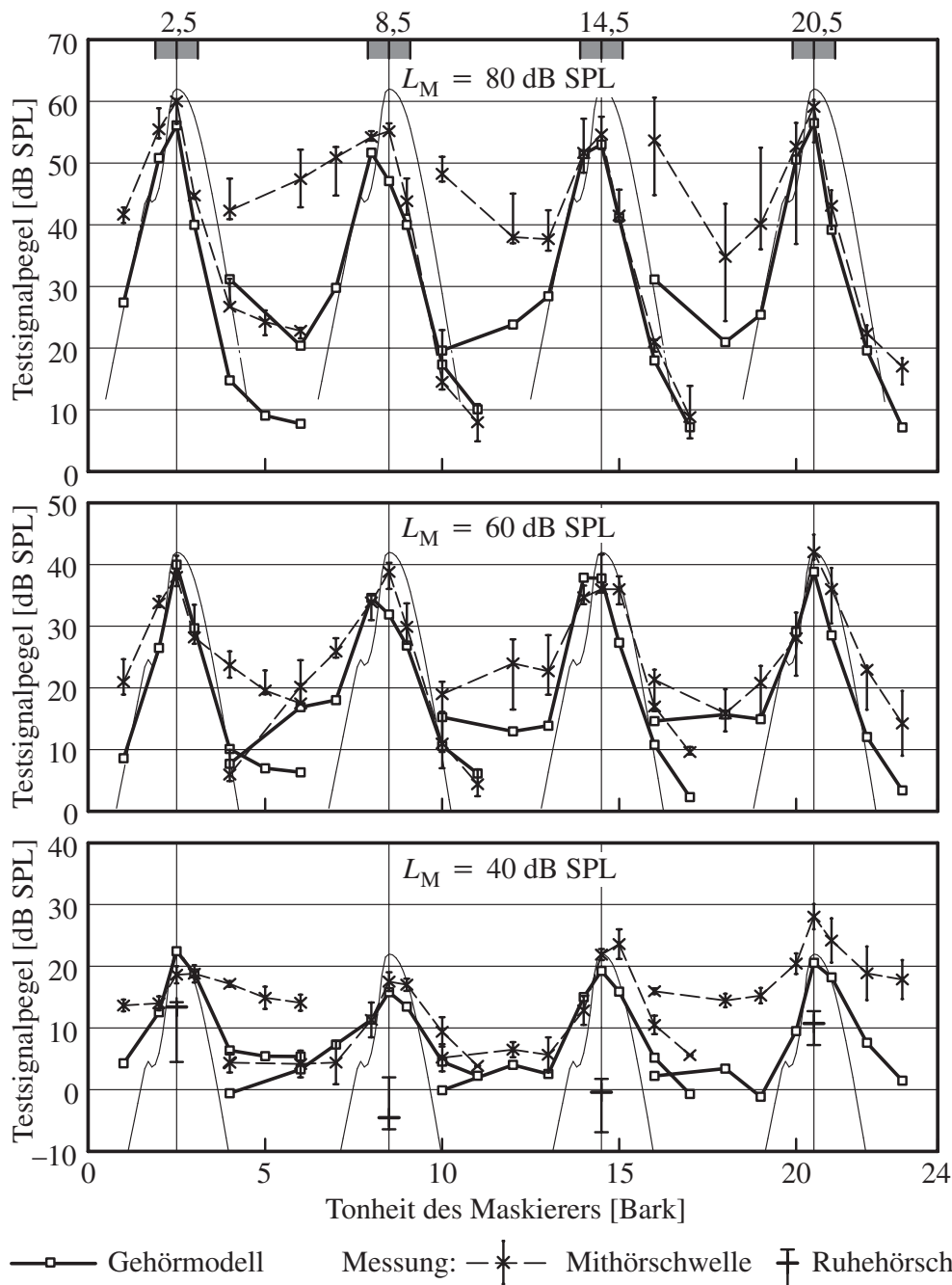


Bild 4.13 Mithörschwelle eines frequenzgruppenbreiten Gaußschen Rauschens verdeckt durch einen Sinuston. Parameter und Meßdaten wie in Bild 4.12. Simulation *ohne* selektive Detektion.

Auf die in Bild 4.12 und Bild 4.13 zusätzlich angegebene Mithörschwelle eines psychoakustischen Modells aus einem Codierverfahren wird in Kapitel 5 eingegangen. Die Mithörschwellen für die Verdeckung eines schmalbandigen Testrauschens unterscheiden sich von den in Bild 4.11 dargestellten Ergebnissen für die Verdeckung eines Sinustesttons im Bereich der Maskierfrequenz. Der Sinustestton führt zu einem ausgeprägten Maximum der Mithörschwelle, wenn seine Frequenz mit der Maskierfrequenz übereinstimmt. In diesem Fall

wird lediglich eine Lautheitsänderung des Maskierersignals durch den überlagerten Sinustestton hervorgerufen, die erst bei höheren Testtonpegeln wahrnehmbar ist, als die andernfalls entstehenden Schwebungen und Verzerrungsprodukte von Maskierer und Testsignal. Bei dem schmalbandigen Testsignal sind diese Verzerrungskomponenten nicht zu vermeiden, so daß die Mithörschwelle im Bereich der Maskierfrequenz dieses ausgeprägte Maximum nicht aufweist.

Aus psychoakustischen Messungen [33, 64, 80, 81, 83, 114] ist bekannt, daß die spektrale Verdeckungswirkung eines Maskierers signifikant vom Zeitverlauf seiner Einhüllenden abhängt. Zur Untersuchung dieses Phänomens sind aus der Literatur spezielle Maskierersignale bekannt, wie Pseudorauschen („*low-noise noise*“ [158]) oder sog. „multipliziertes Rauschen“. Ein Pseudorauschen vorgegebener Bandbreite wird aus einzelnen Sinusschwingungen mit gleicher Amplitude und gleichmäßigem Frequenzabstand innerhalb der Bandbreite synthetisiert. Die Phasen der Sinusschwingungen werden so eingestellt, daß sich eine minimale Einhüllendenfluktuation, bzw. ein minimaler Crest-Faktor des Pseudorauschens ergibt. Ein multipliziertes Rauschen wird hier durch Modulation eines Gaußschen Rauschens erzeugt. Als Modulationssignal wird beispielsweise ein Sinussignal [83] oder ein schmalbandiges Tiefpaßrauschen [75] verwendet. Während das Pseudorauschen einen geringeren Crest-Faktor als Gaußsches Rauschen aufweist, besitzt multipliziertes Rauschen einen größeren Crest-Faktor.

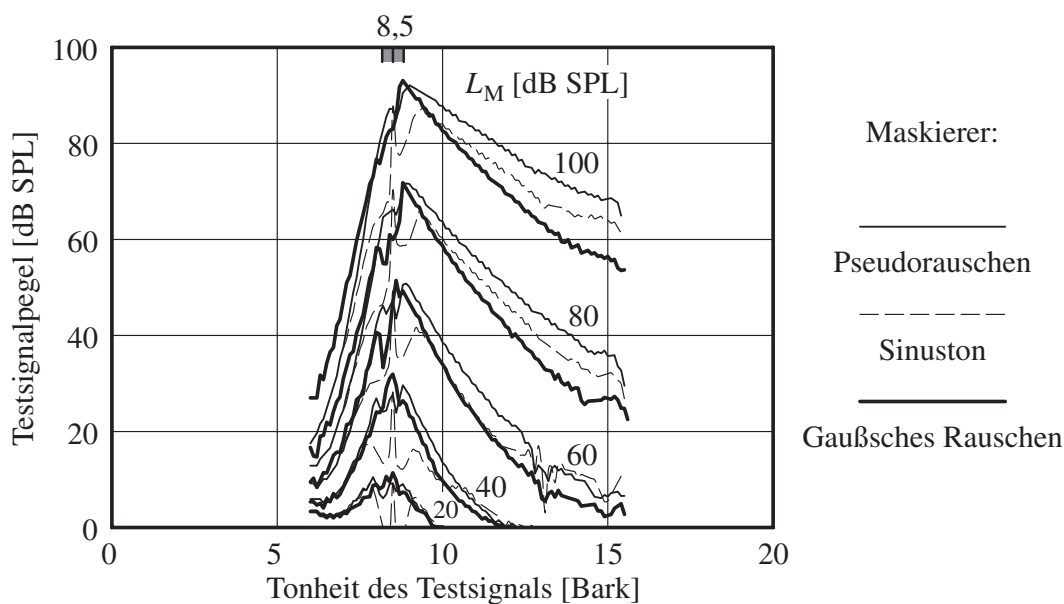


Bild 4.14 Simulierte Mithörschwelle eines Sinustesttons verdeckt durch einen Sinuston, Gaußsches Rauschen, bzw. Pseudorauschen mit geringer Einhüllendenfluktuation. Parameter: $T_M = 300$ ms, $T_T = 200$ ms, $T_{F,T} = 10$ ms, Bandbreite des Rauschens $B_M = 100$ Hz. Simulation mit selektiver Detektion.

In Bild 4.14 sind die Simulationsergebnisse der Mithörschwellen eines Sinustesttons, verdeckt durch Pseudorauschen, im Vergleich zu Gaußschem Bandpaßrauschen bzw. einem Sinuston dargestellt. Die Maskiererbandbreite beträgt 100 Hz bei einer Mittenfrequenz von

1 kHz. Das Pseudorauschen wurde mit dem Algorithmus aus [158] generiert, wobei für den konstanten Frequenzabstand der Sinusschwingungen 1 Hz gewählt wurde. Abgesehen von unregelmäßigen Einbrüchen der Mithörschwelle innerhalb der Bandbreite der Rauschmaskierer ergibt sich ein systematischer Unterschied für Testsignalfrequenzen oberhalb der Maskierfrequenz. In diesem Frequenzbereich zeigt das Pseudorauschen eine höhere Verdeckungswirkung als der Sinuston. Die Mithörschwelle des Sinustonmaskierers liegt bei hohen Pegeln in diesem Frequenzbereich bereits deutlich über derjenigen von Gaußschem Rauschen.

Psychoakustische Messungen der Mithörschwelle eines 2 kHz Sinustesttons, dessen Frequenz oberhalb des Maskierers liegt, wurden in [83] für einen maskierenden Sinuston, Gaußsches bzw. multipliziertes Bandpaßrauschen dokumentiert. Das multiplizierte Rauschen wurde dabei durch Multiplikation eines Tiefpaßrauschens der halben Bandbreite mit einem Sinus bei der Mittenfrequenz von 1,3 kHz erzeugt. Dieses Bandpaßrauschen weist daher ein um die Mittenfrequenz symmetrisches Spektrum auf und besitzt gegenüber dem Gaußschen Bandpaßrauschen eine veränderte Verteilungsdichte der Einhüllendenamplitude [83]. Es wurden Rauschbandbreiten von 20 und 100 Hz verwendet. Die Simulationsergebnisse sind für den Anstieg der Mithörschwelle bei zunehmendem Maskiererpegel („*growth of masking*“) unter Verwendung der in [83] angegebenen Signale in Bild 4.15 zusammen mit den Meßdaten dargestellt. Zum Vergleich wurde bei der Simulation zusätzlich auch maskierendes Pseudorauschen mit entsprechenden Bandbreiten als Maskierer verwendet. Das Pseudorauschen wurde mit dem Algorithmus aus [80] generiert, wobei für den konstanten Frequenzabstand der Sinusschwingungen 1 Hz gewählt wurde.

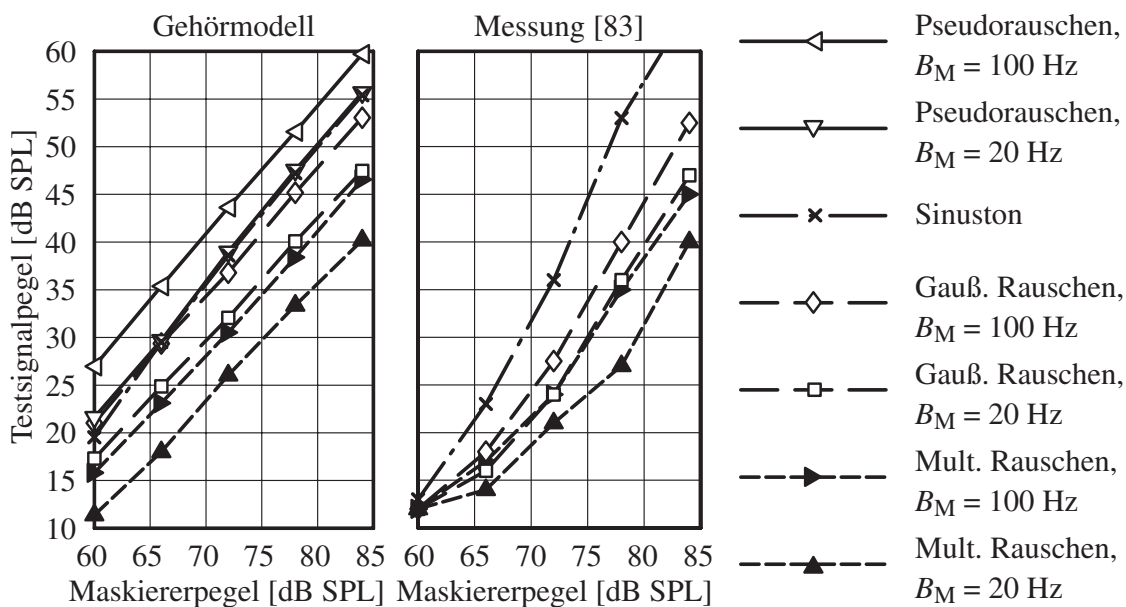


Bild 4.15 Mithörschwelle eines 2 kHz Sinustesttons verdeckt durch die rechts angegebenen Maskierer mit einer Mittenfrequenz von 1,3 kHz. Parameter: $T_M = 500$ ms, $T_T = 400$ ms, $T_{F,T} = 20$ ms, $L_M = [60, 66, 72, 78, 84$ dB SPL], Simulation mit selektiver Detektion.

Es zeigt sich eine gute quantitative Übereinstimmung mit der Messung [83: Fig. 1] für die Rauschmaskierer bei hohen Pegeln. Die Mithörschwelle des Sinustonmaskierers liegt in der Simulation jedoch etwa 6 bis 10 dB unter den Meßwerten. Bei höheren Pegeln nähern sich die Mithörschwellen des 100 Hz breiten Pseudorauschens den Meßergebnissen [83] für den Sinustonmaskierer an. Der Anstieg der Mithörschwellen mit dem Maskiererpegel in der Simulation ist insgesamt geringer als in der Messung. Die Unterschiede zwischen den Verdeckungseigenschaften der Rauschmaskierer werden jedoch weitgehend auch quantitativ nachgebildet. Dabei ist zu beachten, daß der Unterschied der Mithörschwellen von multipliziertem Rauschen und Pseudorauschen bei 20 Hz Bandbreite in der Simulation bis zu etwa 15 dB betragen kann. Vergleichbare psychoakustische Referenzdaten für Pseudorauschen in der verwendeten Konfiguration liegen in der Literatur bislang nicht vor.

Ein maskierendes Gaußsches Bandpaßrauschen erzeugt eine praktisch konstante Mithörschwelle bei einer darin zentrierten Testsignalfrequenz, wenn die Bandbreite bei konstanter Rauschleistungsdichte über die Frequenzgruppenbreite hinaus vergrößert wird. Bei der Verdeckung durch ein multipliziertes Rauschen, sinkt die Mithörschwelle jedoch mit zunehmender Bandbreite ab. Die Mithörschwelle dieses multiplizierten Rauschens liegt bei einer Bandbreite von 1 kHz und Mittenfrequenz von 1 kHz etwa 10 dB unter derjenigen des Gaußschen Rauschens gleicher Bandbreite [75]. Diese Abnahme der Verdeckungswirkung wird im Englischen als „*co-modulation masking release*“ (CMR) bezeichnet.

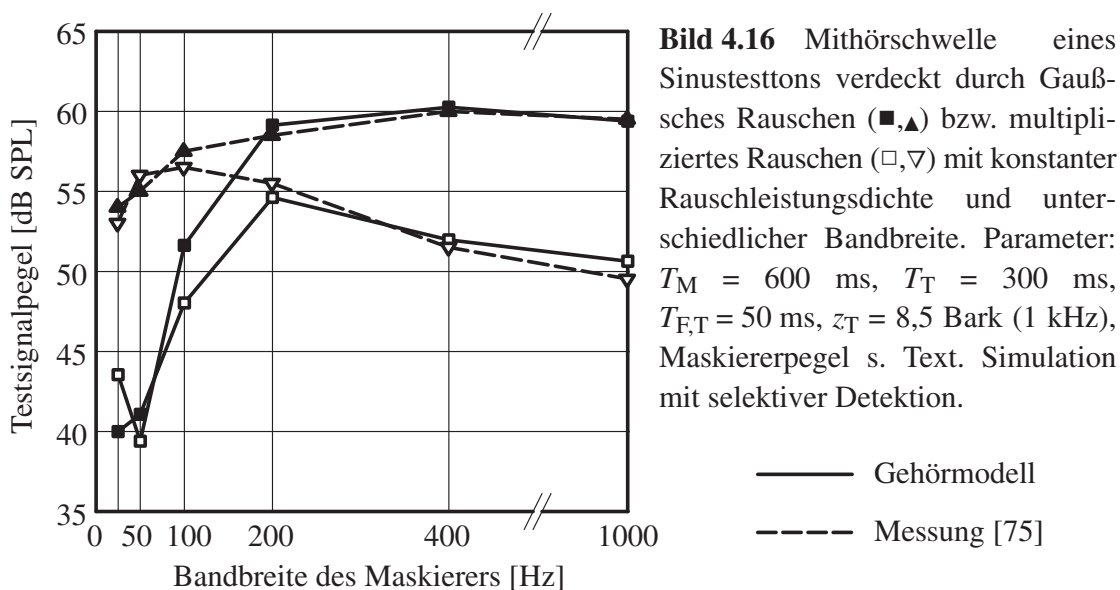


Bild 4.16 Mithörschwelle eines Sinustesttons verdeckt durch Gaußsches Rauschen (■,▲) bzw. multipliziertes Rauschen (□,▽) mit konstanter Rauschleistungsdichte und unterschiedlicher Bandbreite. Parameter: $T_M = 600$ ms, $T_T = 300$ ms, $T_{F,T} = 50$ ms, $z_T = 8,5$ Bark (1 kHz), Maskiererpegel s. Text. Simulation mit selektiver Detektion.

In Bild 4.16 sind die simulierten und die gemessenen Mithörschwellen eines 1 kHz Testtons für unterschiedliche Bandbreiten eines verdeckenden Gaußschen Rauschens bzw. multiplizierten Rauschens dargestellt. Das Gaußsche Rauschen wurde durch Bandpaßfilterung eines 10 kHz breiten Tiefpaßrauschens mit 80 dB SPL bei der Mittenfrequenz 1 kHz erzeugt. Das multiplizierte Rauschen wurde aus dem Gaußschen Rauschen durch Multiplikation mit 50 Hz breitem Tiefpaßrauschen gewonnen. Die simulierte Mithörschwelle steigt für Bandbreiten des Gaußschen Rauschens bis zur Frequenzgruppenbreite von ca. 160 Hz erwartungsge-

mäß an. Für größere Bandbreiten ist die Mithörschwelle praktisch konstant. Das multiplizierte Rauschen führt demgegenüber oberhalb der Frequenzgruppenbreite zu abnehmenden Mithörschwellen, die im Einklang mit der Messung in [75] stehen. Das Gehörmodell detektiert das Testsignal in einem Minimum der Einhüllenden des modulierten Rauschens. Aufgrund der Modulation sind diese Minima wesentlich ausgeprägter und ermöglichen damit die Testsignaldetektion schon bei geringeren Pegeln [134]. Während der Minima der Einhüllenden ist die Testsignalsuppression außerdem verringert, so daß die Detektierbarkeit zunimmt. Die Änderung des Eingangssignals durch den überlagerten Testton in den Minima kann mit dem Gehörmodell sehr einfach durch einen Vergleich mit dem Maskierersignal ohne Testsignal detektiert werden, da in beiden Fällen ein identisches Maskierersignal verwendet wird.

Die Kombination mehrerer Maskierersignale führt in der Regel zu einer erhöhten gemessenen Mithörschwelle des Testsignals im Vergleich zu den einzelnen Maskierern [69, 121, 130, 199]. Die Mithörschwelle des kombinierten Maskierers kann dabei deutlich größer sein als die lineare Summe der Testsignalintensitäten an der Mithörschwelle der einzelnen Maskierer. Dieser Effekt („*nonlinear additivity of masking*“ oder „*excess masking*“) ist besonders bei zeitlich oder spektral nicht überlappenden Maskierern zu beobachten [36, 96, 147]. Bei zwei Maskierern kann beispielsweise die Mithörschwelle um etwa 10 bis 17 dB über der Intensitätssumme liegen [121].

Das Gehörmodell bildet diese hohe Mithörschwellenzunahme bei der Simultanverdeckung qualitativ nach [11]. Wie bei psychoakustischen Messungen [96] ist ebenfalls eine erhöhte Zunahme mit größerem spektralen Abstand der Maskierer festzustellen. Diese Nachbildung ist für die Audiocodierung von Bedeutung, da besonders bei komplexeren Audiosignalen eine entsprechende Zunahme der Mithörschwelle zu erwarten ist. Aufgrund der diversen möglichen Maskiererkombinationen und der hohen Variabilität der psychoakustischen Meßergebnisse, wird hier von der Darstellung einzelner Beispiele abgesehen.

4.2.2.3 Mithörschwellen für zeitabhängige Verdeckung

Der Parameterabgleich des Gehörmodells bezüglich der zeitabhängigen Verdeckungseffekte erfolgte anhand der psychoakustischen Meßdaten, die in den Bildern 4.17, 4.18, 4.19 und 4.20 dargestellt sind. Die berücksichtigten Effekte umfassen die Vorverdeckung, Nachverdeckung und die Abhängigkeit der Simultanverdeckung von der Testsignaldauer.

Die Vorverdeckung wird in den NV-Modellen durch die zeitliche Spreizung nachgebildet. In Bild 4.17 sind die Vorverdeckungsergebnisse für ein gleichmäßig verdeckendes Rauschen dargestellt. Die verwendeten Sinustestsignale besitzen eine Gaußförmige Einhüllende mit einer Dauer von 1 ms. Bei der psychoakustischen Messung [50: Fig. 11] betrug die Testsignalfrequenz 8 kHz. Die Simulation wurde für die Frequenzen 7,84 kHz (21 Bark), 8,72 kHz (21,5 Bark) und 9,74 kHz (22 Bark) durchgeführt. In der Simulation ergeben sich Vorverdeckungsschwellen, die etwa 20 dB unter den gemessenen Schwellen liegen. Dieser Unterschied ist vor allem auf den Overshoot-Effekt zurückzuführen, der beim Einschalten eines breitbandigen Maskierers zu beobachten ist (vgl. Kapitel 2.4.2). Im Vergleich dazu ist die Mithörschwelle bei Simultanverdeckung etwa 30 dB geringer als das Maximum der Vorver-

deckung. Da das Gehörmodell den Overshoot-effekt nicht nachbildet, geht die simulierte Vorverdeckungsschwelle unmittelbar in die Simultanverdeckungsschwelle über. Gegenüber der Messung ist die Ruhehörschwelle in der Simulation etwa 20 dB geringer. Die Parametereinstellung orientierte sich daher an der Steilheit und der zeitlichen Ausdehnung der Vorverdeckung. Die Mithörschwellen für die Simultanverdeckung, etwa 20 ms nach Einschalten des Maskierers, stimmen aus der Simulation und Messung überein.

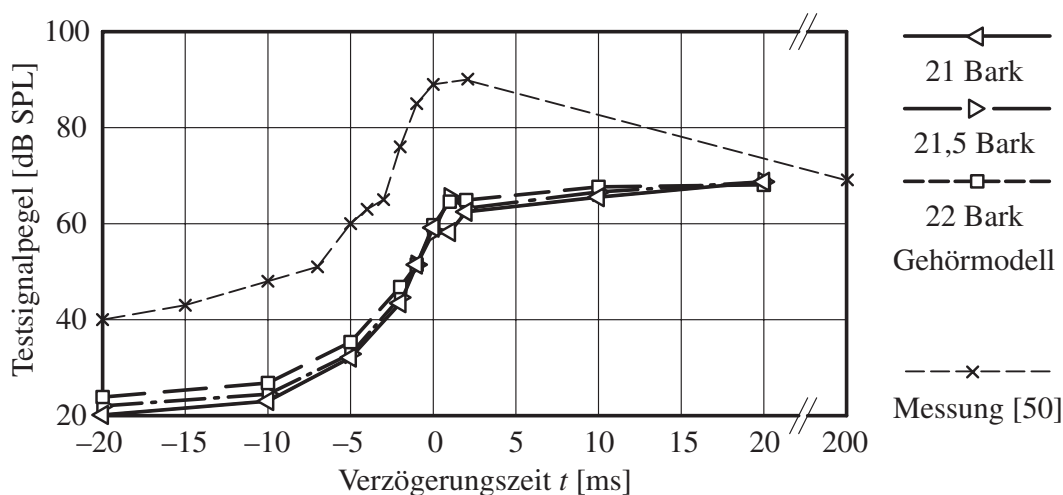


Bild 4.17 Wahrnehmungsschwelle eines Sinustesttons verdeckt durch ein einsetzendes gleichmäßig verdeckendes Rauschen. Parameter: $T_M = 300$ ms, $T_{F,M} = 0,25$ ms, $T_T = 1$ ms, $T_{F,T} = 0,25$ ms, $B_M = 16$ kHz, $L_M = 60$ dB SPL. Messung: $f_T = 8$ kHz. Simulation: $z_T = [21; 21,5; 22$ Bark] (7,84; 8,72; 9,74 kHz). Die Verzögerungszeit t ist die Zeitdifferenz zwischen Maskiererbeginn und Testsignalbeginn.

Die Nachverdeckung wird durch die Abklingfunktion in den NV-Modellen nachgebildet, die in Abhängigkeit des Pegels und der Dauer des Maskierers gesteuert wird. Die pegelabhängige Steilheit der Nachverdeckung geht aus der Messung [197] in Bild 4.18 hervor, bei der ein maskierendes weißes Rauschen und ein 20 μ s langer Testimpuls verwendet wurde. Die Simulationsergebnisse zeigen eine mit dem Maskiererpegel zunehmende Steilheit der Nachverdeckung, so daß unabhängig von der Simultanverdeckungsschwelle ($t = 0$) die konstante Ruhehörschwelle nach Abschalten des Maskierers in etwa 200 ms erreicht wird (vgl. [109]). Bei hohen Maskiererpegeln tritt eine signifikante Abweichung der simulierten Nachverdeckungsschwelle bis zu einer Verzögerung von etwa 10 ms auf. Außerdem ist die Ruhehörschwelle in der Simulation um etwa 10 dB erhöht.

Bei Maskiererdauern von weniger als 200 ms wird eine abnehmende Nachverdeckungsdauer beobachtet [50, 51, 52, 154, 194, 197]. In Bild 4.19 sind Meßergebnisse für ein maskierendes, gleichmäßig verdeckendes Rauschen mit einer Dauer von 5 bzw. 200 ms dargestellt [197: Fig. 4.23]. Das Testsignal besteht aus einem 2 kHz Sinuston mit 5 ms Dauer. Die beiden gemessenen Nachverdeckungskurven erscheinen auf der logarithmischen Zeitachse horizontal verschoben. Die Adaption der Nachverdeckung anhand der Maskiererdauer im

Gehörmodell bildet diesen Effekt nach. Dabei treten allerdings Abweichungen für Verzögerungszeiten von weniger als 10 ms auf, und die Ruhehörschwelle weicht um etwa 7 dB ab.

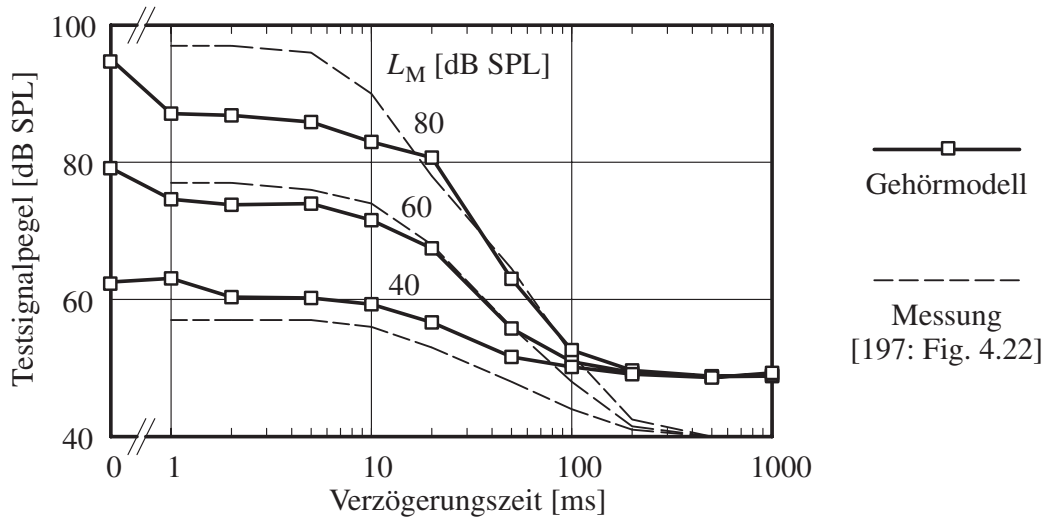


Bild 4.18 Wahrnehmungsschwellen eines Testimpulses verdeckt durch ein abgeschaltetes weißes Gaußsches Rauschen. Parameter der Simulation: $T_M = 250$ ms, $T_{FM} = 1$ ms, $B_M = 16$ kHz, $T_T = 20$ μ s, $T_{F,T} = 10$ μ s. Die Verzögerungszeit t ist die Zeitdifferenz zwischen dem Abschalten des Maskierers und dem Maximum des Testimpulses.

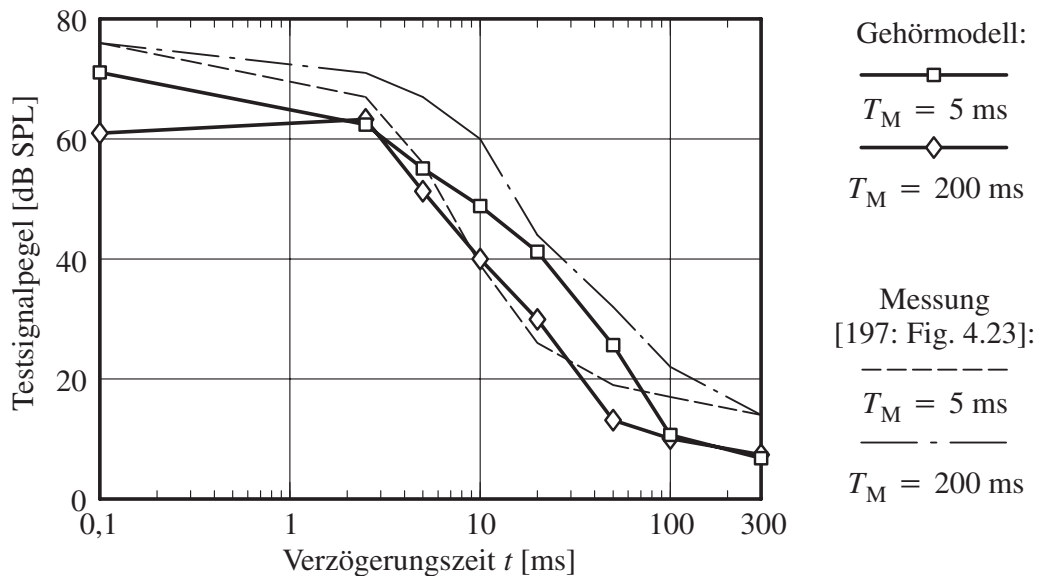


Bild 4.19 Wahrnehmungsschwellen eines Sinustesttons verdeckt durch ein gleichmäßig verdeckendes Rauschen. Parameter: $T_{F,M} = 1$ ms, $L_M = 80$ dB SPL, $B_M = 16$ kHz, $T_T = 5$ ms, $T_{F,T} = 1$ ms, $z_T = 13$ Bark (2 kHz). Die Verzögerungszeit t ist die Zeitdifferenz zwischen dem Abschalten des Maskierers und dem Ende des Testsignals.

Die Simultanverdeckungsschwelle eines Sinustesttons hängt von seiner Dauer ab. Bei einem maskierenden, gleichmäßig verdeckenden Rauschen steigt die Mithörschwelle bei abnehmender Testtondauer von weniger als 200 ms mit etwa 10 dB/Dekade an (s. Bild 4.20). Diese Abhängigkeit wird von dem Kurzzeitintegrator in den NV-Modellen nachgebildet, dessen Parameter entsprechend der in Bild 4.20 dargestellten Meßergebnisse abgeglichen wurden. Die gemessene Steilheit wird in der Simulation sehr gut angenähert, während die Mithörschwellen der Simulation generell um wenige dB unter den Meßwerten liegen.

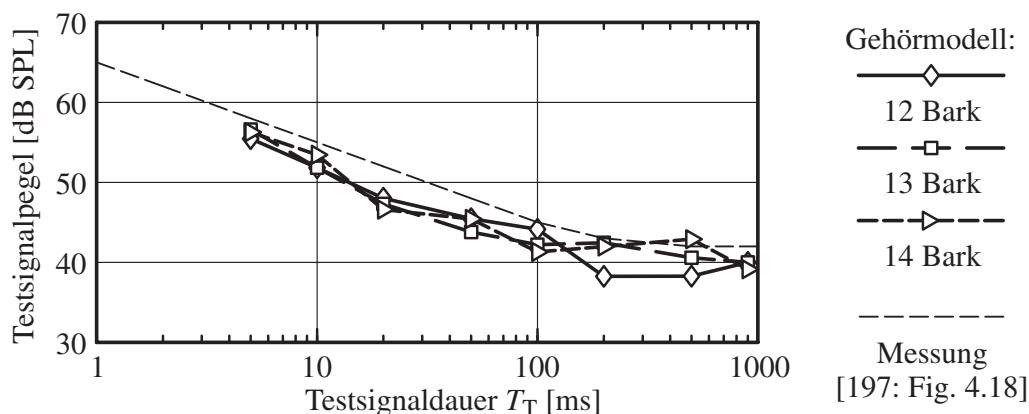


Bild 4.20 Simultane Mithörschwelle eines Sinustesttons verdeckt durch gleichmäßig verdeckendes Rauschen. Parameter: $L_M = 60$ dB SPL, $B_M = 16$ kHz, $T_{F,T} = 5$ ms.

4.2.2.4 Mithörschwellen für kombinierte zeit- und frequenzabhängige Verdeckung

Die Kombination der zeit- und frequenzabhängigen Verdeckungseffekte tritt bei Maskierersignalen auf, deren Verlauf der spektralen Einhüllenden zeitveränderlich ist. Beispielsweise erzeugt ein frequenzmodulierter Sinustonmaskierer eine Mithörschwelle, deren spektraler Verlauf bei gleicher momentaner Maskierfrequenz davon abhängt, ob die Maskierfrequenz zu- oder abnimmt (z. B. [197: Fig. 4.28], [9]). Bei zunehmender Maskierfrequenz bewirkt die Nachverdeckung eine höhere Mithörschwelle bei Testtonfrequenzen unterhalb der Maskierfrequenz im Vergleich zu einem unmodulierten Sinustonmaskierer. Bei abnehmender Frequenz ist das Verhältnis umgekehrt. Durch die Nachbildung beider Verdeckungseffekte im Gehörmodell werden diese kombinierten Verdeckungswirkungen qualitativ simuliert.

Bei sehr schneller Frequenzmodulation sind die Wirkungen der Signallaufzeit in der Cochlea zu bemerken, da tieffrequente Eingangssignale eine größere Laufzeit bis zu ihrem Resonanzort besitzen als hochfrequente. Bei einer dreieckförmigen FM eines Sinustonmaskierers kann daher bei ansteigender Frequenz eine näherungsweise gleichzeitige Resonanzwirkung über einen großen Bereich der Cochlea erzielt werden. Diese erzeugt eine entsprechend kurzzeitige breitbandige Verdeckungswirkung. Bei Verwendung einer abnehmenden Signalfrequenz wandert der Resonanzort hingegen mit zunehmender Verzögerung vom ovalen Fenster zum Helicotrema. In diesem Fall ist daher eine stärkere frequenzselektive Verdeckung vorhanden. Aufgrund der realistischen Nachbildung der Signallaufzeiten im Modell

der Cochlea wird dieser Effekt vom Gehörmodell nachgebildet. In der Literatur wurde dieses Phänomen anhand von sog. „Schröder-Phasen-Signalen“ psychoakustisch untersucht [115].

4.2.2.5 Wahrnehmungsschwellen für Amplituden- und Frequenzmodulation

Die Wahrnehmungsschwellen für eine sinusförmige Amplituden- (AM) bzw. Frequenzmodulation (FM) eines 1 kHz Sinusträgers sind in Bild 4.21 dargestellt. Die Schwellen beschreiben die Größe des Modulationsgrades bzw. Modulationsindex bei gerade wahrnehmbarer Modulation. Bei einer kleinen Modulationstiefe (kleiner als 30%) ist das FM-Betragspektrum näherungsweise identisch mit dem AM-Betragspektrum, das aus einer Spektrallinie bei der Trägerfrequenz und der oberen und unteren Seitenlinie im Abstand der Modulationsfrequenz von der Trägerfrequenz besteht. Die Signale unterscheiden sich dann im wesentlichen durch die Phasen. Bei Modulationsfrequenzen oberhalb der halben Frequenzgruppenbreite (ca. 80 Hz) werden die Seitenlinien als separate Töne wahrgenommen, da ihr Frequenzabstand dann größer als die spektrale Auflösung des Gehörs ist. Bei kleineren Modulationsfrequenzen wird jedoch bei FM eine Tonhöhenchwankung und bei AM eine Lautheitschwankung wahrgenommen. Die Modulation der Amplitude wird aufgrund der Nichtlinearität bei der cochleären Schallverarbeitung in der Umgebung des Resonanzortes komprimiert. Die AM wirkt sich jedoch zunehmend unkomprimiert an den Orten aus, die weiter vom Resonanzort entfernt in Richtung zum ovalen Fenster liegen (vgl. Bild 4.1). Daher ist eine Detektion der AM in diesem Bereich bereits möglich, bevor am Resonanzort der Träger-schwingung die Modulation oberhalb der Wahrnehmungsschwelle liegt. Bei kleinen Modulationsfrequenzen ist die Detektionsschwelle von AM daher nahezu unabhängig von der Modulationsfrequenz. Bei einer langsamen FM verschiebt sich anschaulich die Einhüllende der Wanderwelle mit der Modulationsfrequenz longitudinal in der Cochlea, wobei das

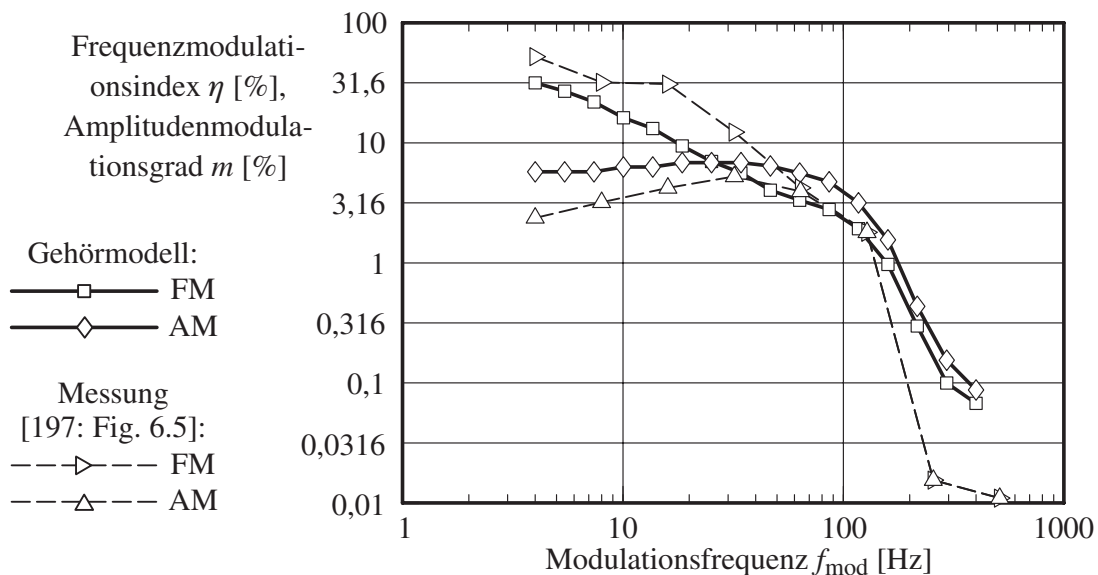


Bild 4.21 Detektionsschwelle einer sinusförmigen Frequenz- (FM) bzw. Amplitudenmodulation (AM) eines Sinusträgers. Parameter: $f_M = 1$ kHz, $L_M = 80$ dB SPL, $T_M = 500$ ms.

Maximum dem Resonanzort der momentanen Signalfrequenz folgt. Die Detektierbarkeit dieser Verschiebung wächst mit zunehmender Flankensteilheit der Einhüllenden, bzw. des Erregungsmusters der Cochlea (vgl. Bild 4.1). Daher wird die FM im Bereich des steilsten Verlaufs des Erregungsmusters detektiert. Unter der Voraussetzung eines konstanten Frequenzhubs Δf und einer gegebenen Steilheit des Erregungsmusters ist daher bei langsamer Modulation die Detektierbarkeit von FM unabhängig von der Modulationsfrequenz. Die Detektionsschwelle in Bild 4.21 steigt für FM zu kleineren Modulationsfrequenzen jedoch an, da dort der Modulationsindex $\eta = \Delta f / f_{\text{mod}}$ aufgetragen ist.

Die simulierte Detektionsschwelle von AM liegt insgesamt über den Referenzmeßergebnissen. Die simulierte Detektionsschwelle für eine FM liegt bei Modulationsfrequenzen bis 100 Hz unterhalb der Meßwerte und für höhere Modulationsfrequenzen darüber. Die Abhängigkeit von der Modulationsfrequenz stimmt bei AM und FM im wesentlichen mit den Meßkurven überein. Der Übergang von der Wahrnehmung von AM bzw. FM zur Wahrnehmung der einzelnen Spektralkomponenten wird ebenfalls nachgebildet.

4.3 Diskussion der Ergebnisse

Die Überprüfung des Gehörmodells anhand physiologischer Meßdaten zeigt, daß die Frequenzselektivität und Dynamikkompression innerhalb der Nachbildung der mechanischen Schallverarbeitung durch das Gehörmodell in Übereinstimmung mit den Messungen steht. Als grundlegende Mechanismen der spektralen Simultanverdeckung werden die Erregung und Suppression ebenfalls in qualitativer Übereinstimmung nachgebildet. Damit sind die zur Nachbildung der Simultanverdeckung wesentlichen Effekte im Gehörmodell bereits am Ausgang der IHZ-Modelle vorhanden.

Die Verifikation des Gehörmodells zur Nachbildung von Wahrnehmungsschwellen für beliebig komplexe Eingangssignale ist aufgrund ihrer Vielfalt nicht umfassend möglich. Daher ist zu diesem Zweck eine Auswahl von psychoakustischen Meßdaten verwendet worden, in denen sich die wesentlichen Abhängigkeiten der Verdeckungswirkung von den physikalischen Parametern des Eingangssignals widerspiegeln. Die hier betrachteten Parameter des Maskierersignals umfassen: Pegel, Frequenz, Bandbreite, Dauer, Einhüllendenfluktuation, Modulationsgrad sowie Modulationsindex. In der Regel bestand das Testsignal aus einem Sinuston, für einige Messungen wurde davon abweichend ein Gaußimpuls oder ein frequenzgruppenbreites Rauschen verwendet. Die betrachteten Testsignalparameter sind: Frequenz, Dauer sowie Verzögerungszeit gegenüber dem Maskierer.

Für die genannten Abhängigkeiten der Wahrnehmungsschwellen von den physikalischen Parametern wird eine qualitative und zum großen Teil auch quantitative Übereinstimmung der Simulation mit psychoakustischen Meßdaten festgestellt. Das bedeutet, daß die Variation eines Parameters sowohl in der Simulation als auch bei der Messung zu einer gleichartigen Änderung der Wahrnehmungsschwelle führt. Darüber hinaus werden bestimmte Eigenschaften gemessener Mithörschwellenverläufe nachgebildet, wie beispielsweise die pegelabhängige Steilheit der Mithörschwelle über der Testsignalfrequenz für Frequenzen oberhalb des Maskierers (Bild 4.9) oder die pegelabhängige Steilheit der Nachverdeckungsschwelle über der Testsignalverzögerung (Bild 4.18). Eine gute quantitative Über-

einstimmung wird zwischen den simulierten und gemessenen simultanen Mithörschwellen im Bereich mittlerer Eingangssignalpegel und für Testsignalfrequenzen innerhalb einiger Bark um die Maskierfrequenz erreicht. Aufgrund der hohen Varianz psychoakustischer Meßergebnisse der Mithörschwelle eines Sinustonmaskierers ist die erreichbare Übereinstimmung mit der Simulation von vornherein begrenzt. Die durch hörbare Schwebungen und Verzerrungsprodukte bei der Messung verursachten Einbrüche der Mithörschwelle unterliegen interindividuellen Schwankungen, die mit dem Gehörmodell allenfalls für eine einzelne Testperson nachgebildet werden könnten.

Bei der Überlagerung mehrerer Maskierersignale wird in vielen Fällen eine überproportionale Zunahme der Mithörschwelle beobachtet. Diese sog. „nichtlineare Additivität“ der Verdeckung wird vom Gehörmodell qualitativ nachgebildet.

Für die Analyse und das Verständnis der Zusammenhänge zwischen der simulierten spezifischen Lautheit und den Mithörschwellen ist die folgende vereinfachte qualitative Darstellung hilfreich. Es wird angenommen, daß die spezifische Lautheit eines Sinustons einen idealisierten pegelabhängigen Verlauf gemäß Bild 4.22A aufweist. Bei zunehmendem Eingangssignalpegel L_M nimmt die Verstärkung am Ort der Resonanz ab, so daß im Bereich der Resonanz eine (pegelunabhängige) Dynamikkompression von $1/3$ erreicht wird. Die idealisierte Approximation durch drei Geradenstücke berücksichtigt einige wesentliche Eigenschaften der simulierten Verläufe der spezifischen Lautheit, die bei stationären Eingangssignalen mit den Kurven in Bild 4.1 vergleichbar sind.

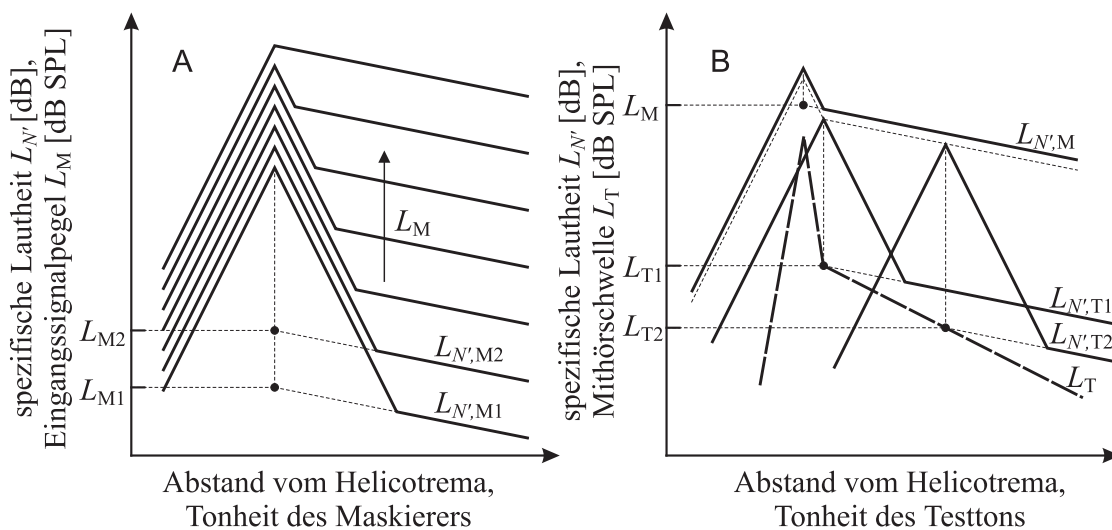


Bild 4.22 Idealisierte Verläufe der spezifischen Lautheit. (A) Spezifische Lautheit eines Sinustons mit unterschiedlichem Pegel L_M . Die Dynamikkompression am Resonanzort ist dabei als $1/3$ angenommen worden. Die Pegel von zwei Eingangssignalen sind als „•“ eingezeichnet. (B) Konstruktion des Mithörschwellenverlaufs L_T anhand der idealisierten spezifischen Lautheit unter Vernachlässigung der Suppression. Die Flankensteilheit der Mithörschwelle gegenüber der spezifischen Lautheit des Maskierers ist um den Faktor 3 erhöht.

Mit der bekannten spezifischen Lautheit von Maskierer und Testsignal kann die Mithörschwelle konstruiert werden, wenn eine lineare Intensitätssummutation der spezifischen Lautheiten vorausgesetzt wird. Bei einem konstanten Schwellwert für die Detektion einer Signaländerung ist ein Testsignal demnach unterhalb der Mithörschwelle, wenn seine spezifische Lautheit die Maskiererlautheit mit einem konstanten Pegel-Offset nicht überschreitet. Diese Schwelle für die spezifische Lautheit des Testsignals ist in Bild 4.22B als gestrichelte Parallele zu $L_{N',M}$ angegeben. In Bild 4.22B sind die spezifischen Lautheiten für zwei Sinustestsignale T_1 bzw. T_2 an der Mithörschwelle des Sinustonmaskierers M eingezeichnet. Anhand der zugehörigen Testsignalpegel kann die Mithörschwelle L_{T1} bzw. L_{T2} abgelesen werden. Aufgrund der Dynamikkompression mit dem Faktor $1/3$ ergibt sich eine Steilheit der Mithörschwelle, die gegenüber der spezifischen Lautheit des Maskierers verdreifacht ist. Da diese Betrachtungen von einer linearen Intensitätssummutation ausgehen, wird u. a. die Testsignalsuppression vernachlässigt. Der folgende Vergleich dieser idealisierten Betrachtung mit Simulationsergebnissen soll daher Aufschluß über die von den vereinfachenden Annahmen verursachten Abweichungen geben.

Bild 4.23 zeigt die simulierte spezifische Lautheit von einem Sinustonmaskierer M, sowie von einzelnen Sinustestsignalen T_1 bzw. T_2 , deren Pegel gleich dem Mithörschwellenpegel ist. Das Sinustestsignal wurde jeweils ohne überlagerten Maskierer eingespeist. Der entsprechende Mithörschwellenverlauf ist mit $L_T(z)$ bezeichnet. Gemäß der oben genannten idealisierten Betrachtung wäre ein dem Maskierer überlagertes Testsignal gerade wahrnehmbar, wenn dessen spezifische Lautheit gerade die Maskiererlautheit mit einem Pegel-Offset erreicht. Der Pegel-Offset errechnet sich aus dem hier als konstant angenommenen internen Schwellwert der Schwellendetektoren (für Signale ohne Einhüllendenfluktuation gilt $r_S = 1,045$). Der Pegel-Offset ist damit $10 \log(r_S - 1) = -13,5$ dB). Dieser Pegel wird jedoch deutlich überschritten (vgl. Δ_1, Δ_2). Im Vergleich zur idealisierten Betrachtung zeigt die Simulation anhand des Gehörmodells, daß die Suppressionswirkung und ggf. andere nichtlineare Effekte, wie beispielsweise Verzerrungsprodukte, einen wesentlichen Einfluß auf die nachgebildete Mithörschwelle haben.

Für die Nachbildung von Mithörschwellen mit dem Gehörmodell sind aufgrund dieser Ergebnisse somit die nichtlinearen Effekte, die durch die Interaktion von Maskierer und Testsignal hervorgerufen werden, nicht zu vernachlässigen. Die idealisierte Betrachtung bietet jedoch den Vorteil, daß die Mithörschwelle bei bekannter spezifischer Lautheit des Testsignals für verschiedene Signalpegel aus der spezifischen Lautheit des Maskierers ohne Iteration in einem Schritt bestimmt werden kann. Die Simulationsergebnisse in Bild 4.23 zeigen jedoch, daß dieses Verfahren zur Approximation der im Gehörmodell auftretenden Größen für die Mithörschwellenbestimmung ungeeignet ist.

Im Gegensatz zur Nachbildung von Mithörschwellen erlaubt die Idealisierung eine anschauliche Erklärung der unterschiedlichen Detektionsschwellen für eine AM bzw. FM. In Bild 4.24 wird die spezifische Lautheit des unmodulierten Trägersignals im Vergleich zur momentanen spezifischen Lautheit des modulierten Trägers dargestellt. Bei AM tritt die größte Änderung der spezifischen Lautheit außerhalb des Resonanzbereiches auf, innerhalb dessen eine Dynamikkompression erfolgt. Bei FM bewirkt die Änderung der Trägerfrequenz

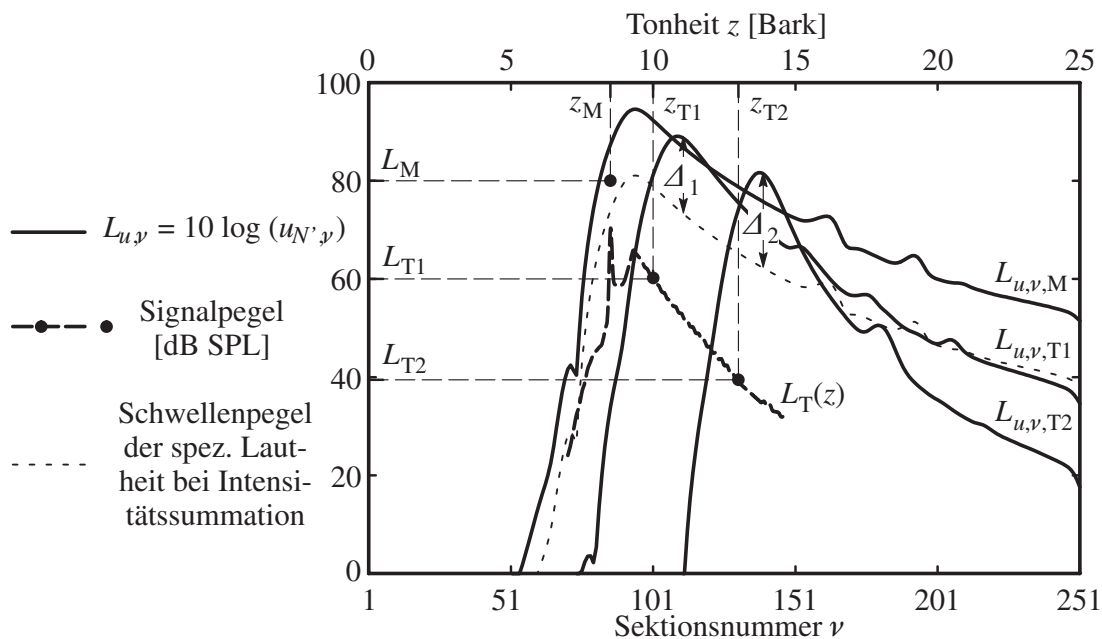


Bild 4.23 Simulierte spezifische Lautheit $L_{u,y}$ eines Sinustonmaskierers sowie der Sinustestsignale T_1 bzw. T_2 an der Mithörschwelle. Die Eingangssignalpegel und -tonheiten sind durch „•“ gekennzeichnet. Die simulierte Mithörschwelle ist durch $L_T(z)$ gekennzeichnet. Der *dünn gestrichelt* eingezeichnete Pegel gibt die obere Grenze der spezifischen Lautheit für unterschwellige Testsignale unter der Voraussetzung an, daß eine lineare Intensitätsaddition der spezifischen Lautheiten von Maskierer und Testsignal bei deren Überlagerung erfolgt. Aufgrund der Testsignalsuppression erzeugt das Testsignal T_1 bzw. T_2 an der Mithörschwelle jedoch eine wesentlich größere spezifische Lautheit (vgl. Δ_1, Δ_2). Parameter: $L_M = 80$ dB SPL, $L_{T1} = 60$ dB SPL, $L_{T2} = 39$ dB SPL, $z_M = 8,5$ Bark, $z_{T1} = 10$ Bark, $z_{T2} = 13$ Bark.

eine horizontale „Verschiebung“ der spezifischen Lautheit. Die größte Änderung tritt daher im Bereich der größten Steilheit des Verlaufs der spezifischen Lautheit auf. Im Gehörmodell ist das auf der apikalen (dem Helicotrema zugewandten) Seite des Resonanzmaximums der Fall.

Die unterschiedlichen Wahrnehmungsschwellen von AM und FM bei kleinen Modulationsfrequenzen werden durch das Gehörmodell nachgebildet. Dabei bewirkt die Nichtlinearität des Modells der Cochlea, daß eine AM in Modellsektionen, die entfernt von dem Resonanzort der Trägerschwingung liegen, unkomprimiert erscheint. Die Detektion ist daher in diesen Sektionen eher möglich, als in der Umgebung des Resonanzorts. Die Detektierbarkeit hängt im Bereich kleiner Modulationsfrequenzen nur vom Modulationsgrad ab. Eine FM führt demgegenüber zu einer Verschiebung der Einhüllenden der Wanderwelle in der Cochlea, die dem Resonanzort der momentanen Eingangssignalfrequenz folgt. Die Detektierbarkeit einer FM ist daher im Bereich kleiner Modulationsfrequenzen nur vom Frequenzhub abhängig. Der in Bild 4.21 dargestellte Modulationsindex η fällt gegenüber dem Frequenzhub mit der Modulationsfrequenz um 20 dB/Dekade ab.

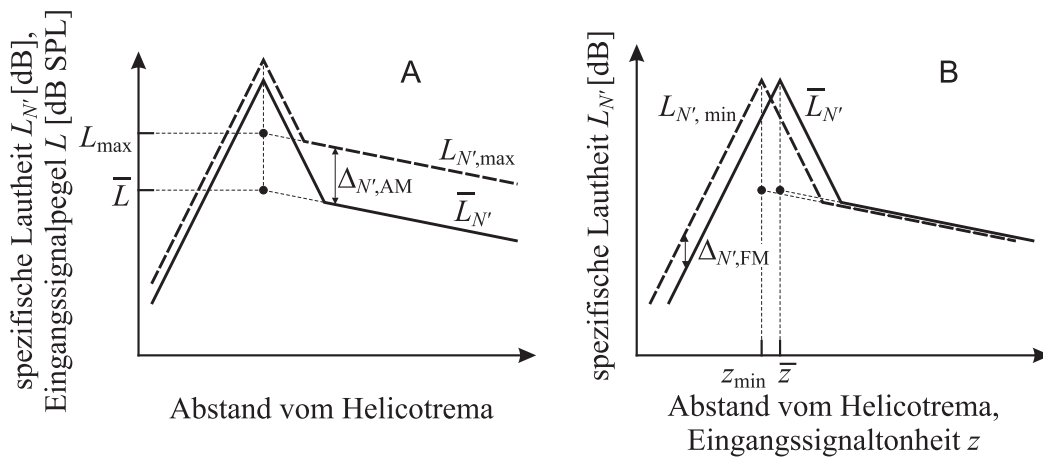


Bild 4.24 Idealisierter Verlauf der spezifischen Lautheit bei AM (A) bzw. FM (B). Die spezifische Lautheit des unmodulierten Trägers ist $\bar{L}_{N'}$. Der maximale Eingangsspiegel bei AM ist L_{max} . Die Tonheit des unmodulierten Trägers beträgt \bar{z} . Die minimale Tonheit bei FM ist z_{min} . Bei AM tritt die maximale Änderung der spezifischen Lautheit $\Delta_{N',AM}$ basal vom Resonanzort auf. Bei FM tritt die maximale Änderung im Bereich der Resonanz auf.

Die im Gehörmodell nachgebildete spezifische Lautheit eines unmodulierten Sinusträgers bei 1 kHz ist in Bild 4.25 logarithmiert entsprechend $10 \log(u_{N',v})$ dargestellt. Der Verlauf der spezifischen Lautheit über den Sektionen ist qualitativ mit den Kurven in Bild 4.1 vergleichbar. Eine AM oder FM des Trägersignals an der Wahrnehmungsschwelle ($m = 7,2\%$ bzw. $\eta = 18,3\%$) entsprechend Bild 4.21 ist jedoch anhand dieser Darstellungsweise nicht

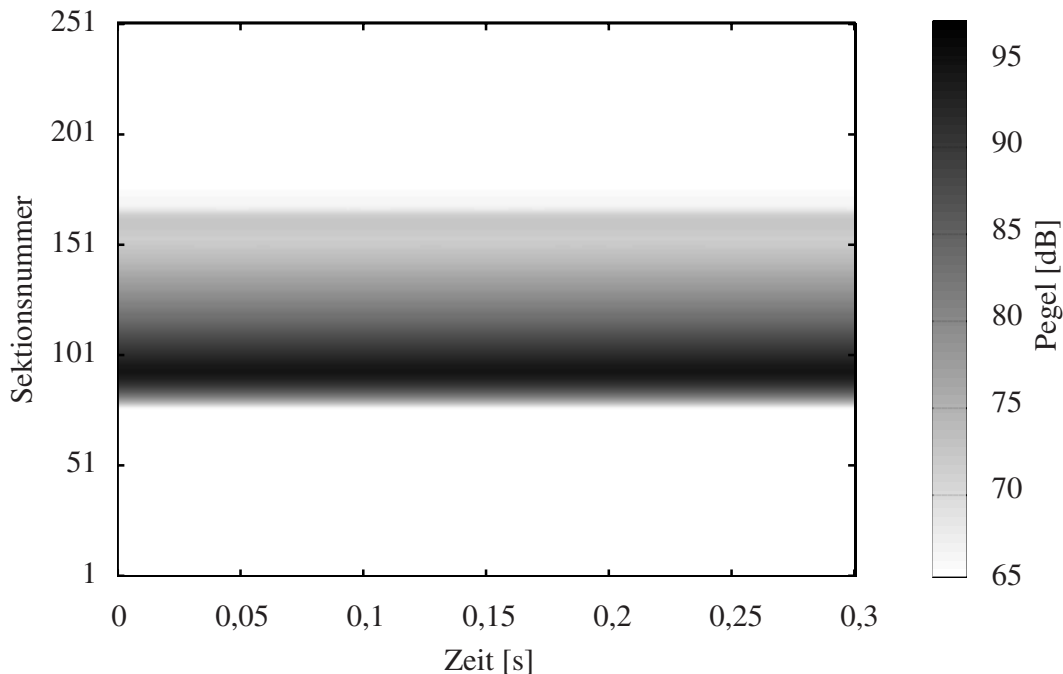


Bild 4.25 Simulierte spezifische Lautheit $u_{N',v}$ eines Sinustons. Logarithmierte Darstellung von $10 \log(u_{N',v})$. Parameter: $f_M = 1$ kHz, $L_M = 80$ dB SPL.

erkennbar. In der Darstellung des Verhältnisses der spezifischen Lautheiten des modulierten und unmodulierten Trägers werden die vom Gehörmodell ausgewerteten Signalunterschiede jedoch erkennbar. Das Verhältnis wird als logarithmierter Wert $L_{r,N',\nu}$ entsprechend (4.1) dargestellt. Dabei bezeichnet $u_{N',\nu,M+T}$ die nachgebildete spezifische Lautheit eines modulierten Trägersignals bzw. eines Maskierers mit überlagertem Testsignal. $u_{N',\nu,M}$ bezeichnet die nachgebildete spezifische Lautheit des unmodulierten Trägers bzw. des Maskierers allein.

$$L_{r,N',\nu} = 10 \log\left(\frac{u_{N',\nu,M+T}}{u_{N',\nu,M}}\right) \quad (4.1)$$

Bild 4.26 zeigt, daß die größten detektierbaren Signaländerungen durch die AM in Sektionen mit Resonanzen deutlich oberhalb der Testsignalresonanz auftreten. Im Gegensatz dazu geht aus Bild 4.27 hervor, daß die Änderungen der spezifischen Lautheit in den Sektionen mit Resonanzen etwas unterhalb der Testsignalfrequenz am größten sind. Dieses Beispiel verdeutlicht die unterschiedlichen Detektionsmechanismen von AM bzw. FM, wie oben beschrieben (vgl. Bild 4.24).

Die im Gehörmodell ausgewertete Einhüllendenfluktuation der internen Modellsignale zur Steuerung des Schwellwerts in jeder Sektion erweist sich aufgrund der Simulationsergebnisse als ein geeigneter Mechanismus zur Berücksichtigung der Verdeckungswirkungen

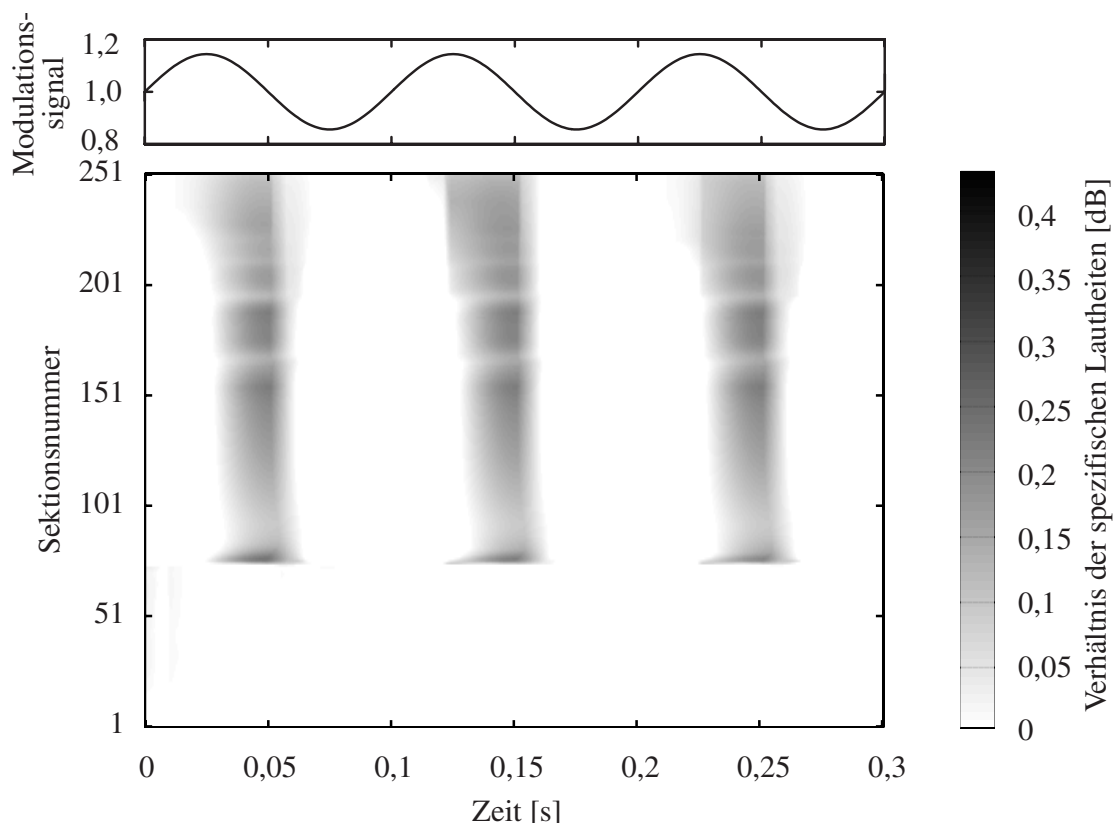


Bild 4.26 Verhältnis der spezifischen Lautheiten $L_{r,N',\nu}$ eines sinusförmig amplitudenmodulierten Sinusträgers und dem unmodulierten Sinusträger. Parameter: $f_M = 1$ kHz, $L_M = 80$ dB SPL, $f_{\text{mod}} = 10$ Hz, $m = 7,2$ %.

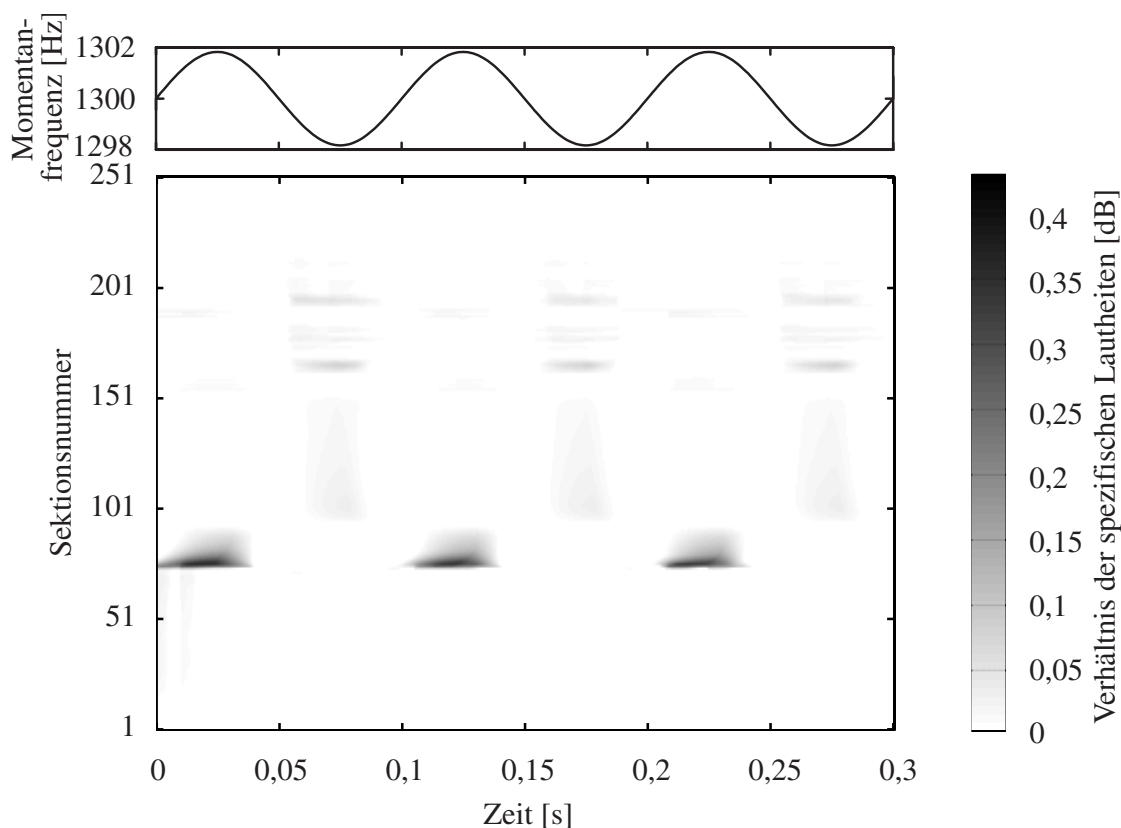


Bild 4.27 Verhältnis der spezifischen Lautheiten $L_{r,N',y}$ eines sinusförmig frequenzmodulierten Sinusträgers und dem unmodulierten Träger. Parameter: $f_M = 1$ kHz, $L_M = 80$ dB SPL, $f_{\text{mod}} = 10$ Hz, $\eta = 18.3\%$.

unterschiedlicher Rauschmaskierer. Die signifikanten Mithörschwellenunterschiede, die aufgrund der verschiedenen Maskierereinhüllenden für Testsignale oberhalb der Maskierereinfrequenz entstehen, werden daher in der Simulation entsprechend nachgebildet.

Für die abnehmende Verdeckungswirkung eines multiplizierten Rauschens bei zunehmender Bandbreite (Bild 4.16) wird als eine mögliche Ursache ein im Gehör stattfindender Vergleich der „Filterausgänge“ über mehrere Frequenzgruppen hinweg angesehen [75]. Entsprechend dieser Vorstellung wird die Testsignaldetektion bei multipliziertem Rauschen dadurch erleichtert, daß die Unterschiede in den Minima der Einhüllenden nur innerhalb derjenigen Frequenzgruppe vorhanden sind, in der die Testsignalfrequenz liegt, und durch einen Vergleich mit anderen Frequenzgruppen erkannt werden können. Im Gehörmodell wird eine vergleichbare gemeinsame Auswertung über die Ausgänge verschiedener Sektionen hinweg nicht vorgenommen. Deshalb ist die verminderte Verdeckung multipliziertem Rauschens in der Simulation auf die veränderte Einhüllende zurückzuführen, die ausgeprägtere Minima gegenüber Gaußischem Rauschen aufweist. Da diese Minima in allen Frequenzgruppen praktisch gleichzeitig auftreten, ist die über mehrere Frequenzgruppen wirkende Testsignalsuppression zu diesem Zeitpunkt sehr klein (vgl. Bild 4.5). Es wird daher vermutet, daß die unterschiedliche Verdeckungswirkung durch das Zusammenwirken des veränderten Einhüllendenverlaufs mit der verminderten Suppressionswirkung zustande kommt. Bei Gauß-

schem Rauschen ist demgegenüber in der Regel eine supprimierende Maskiererkomponente in mindestens einer benachbarten Frequenzgruppe vorhanden. Die Erklärung der verminder- ten Verdeckung ist daher besser mit der Vorstellung vereinbar, daß vom Gehör die Minima der Maskierereinhüllenden für die Detektion effizient genutzt werden (vgl. [33]).

Die bei der Detektion von AM bzw. FM beobachteten Mechanismen können als Bei- spiele für die Verdeckungsmechanismen bei den unterschiedlichen schmalbandigen Rausch- maskierern angesehen werden. Während ein multipliziertes Rauschen ein amplitudenmodu- liertes Signal darstellt, wird Pseudoräuschen eher als frequenzmoduliertes Signal wahrgen- ommen. Dieser unterschiedliche Klangcharakter mit einer größeren Rauigkeit des multi- plizierten Rauschens wird auch in den Zeitverläufen der spezifischen Lautheit beider Signale deutlich. In Bild 4.28 ist die durch $u_{N',\nu}$ nachgebildete spezifische Lautheit für ein multipliziertes Rauschen bei einer Mittenfrequenz von 1,3 kHz dargestellt. Bild 4.29 enthält die ent- sprechende Darstellung für Pseudoräuschen. Beiden Maskierersignalen ist das gleiche 2 kHz Testsignal überlagert. Während das multiplizierte Rauschen eine deutliche zeitabhängige Schwankung der spezifischen Lautheit aufweist, führt das Pseudoräuschen vor allem zu einer zeitabhängigen Verschiebung des Verlaufs der spezifischen Lautheit über die Sektio- nen hinweg. In beiden Bildern ist die Verzerrungskomponente bei der kubischen Differenz- frequenz deutlich erkennbar. Die Frequenzschwankungen dieser Komponente resultieren aus der Variation der momentanen Frequenzen des Rauschens.

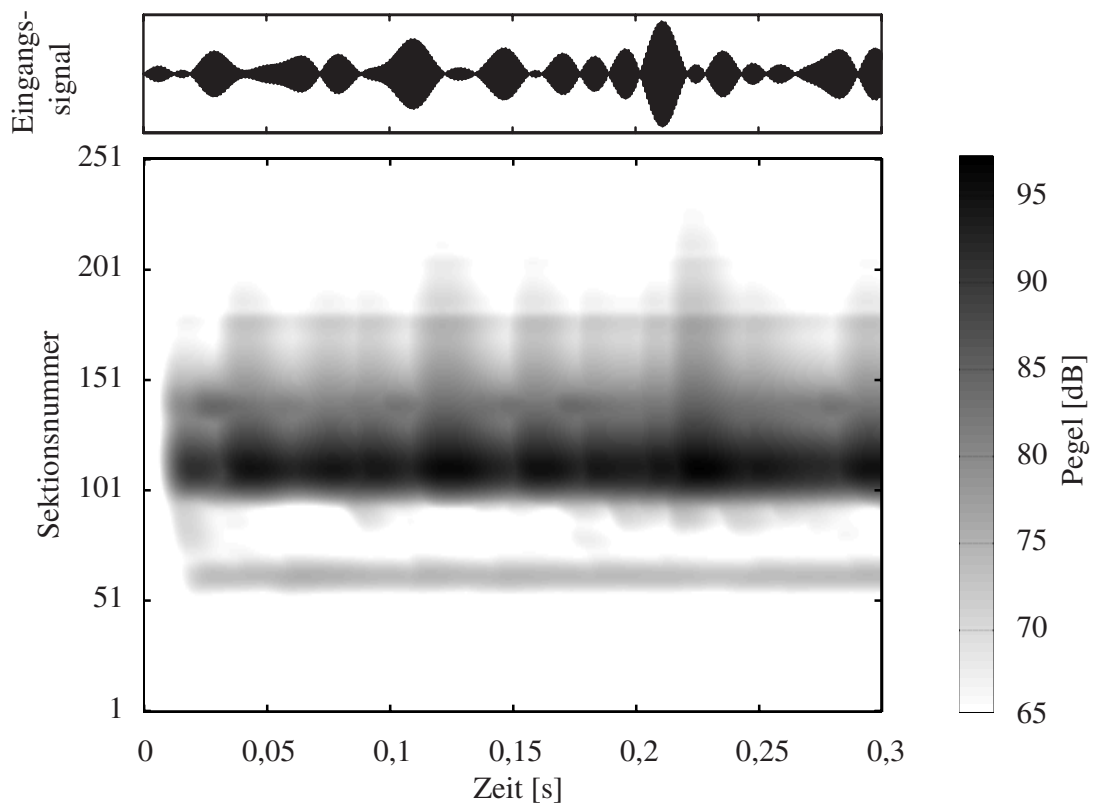


Bild 4.28 Simulierte spezifische Lautheit $u_{N',\nu}$ von multipliziertem Rauschen mit überla- gertem Sinustestsignal. Logarithmierte Darstellung von $10 \log(u_{N',\nu})$. Parameter: $f_M = 1,3$ kHz, $B_M = 100$ Hz, $L_M = 80$ dB SPL, $f_T = 2$ kHz, $L_T = 50$ dB SPL.

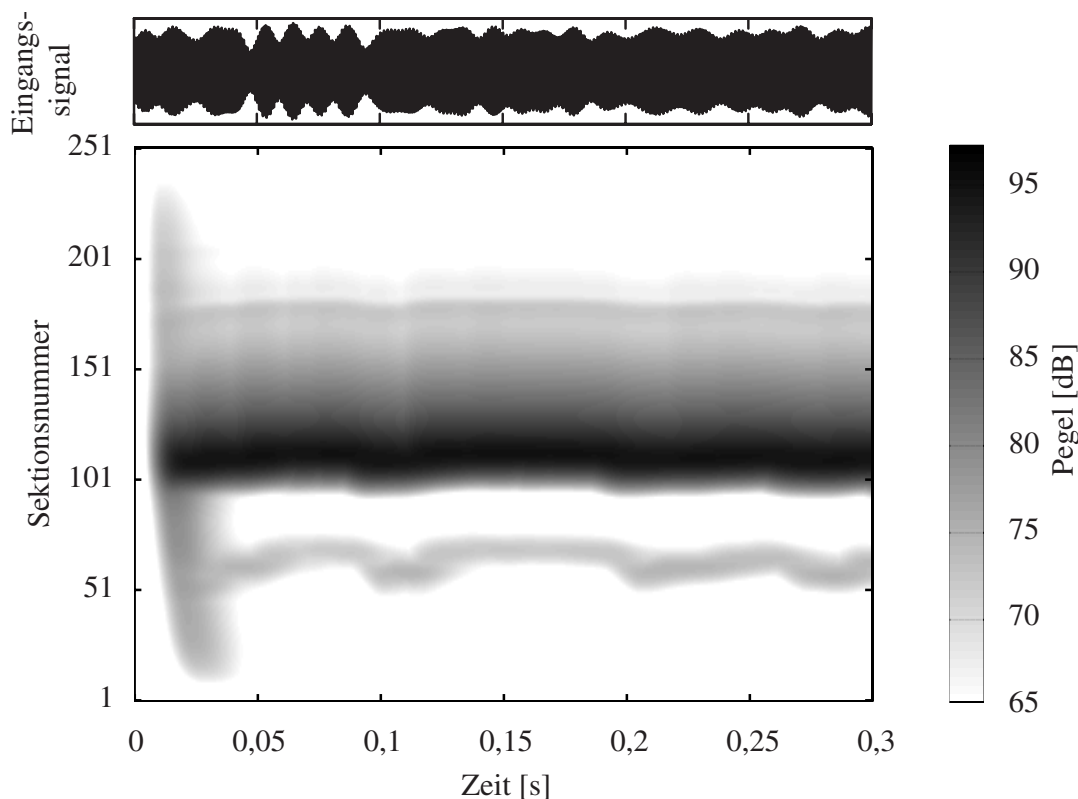


Bild 4.29 Simulierte spezifische Lautheit $u_{N',v}$ von Pseudoräuschen mit überlagertem Sinustestsignal. Logarithmierte Darstellung von $10 \log(u_{N',v})$. Parameter: $f_M = 1,3$ kHz, $B_M = 100$ Hz, $L_M = 80$ dB SPL, $f_T = 2$ kHz, $L_T = 50$ dB SPL.

Das Verhältnis der spezifischen Lautheiten ist für das multiplizierte Rauschen in Bild 4.30 und in Bild 4.31 für das Pseudoräuschen dargestellt. Die verwendeten Maskierer und Testsignale stimmen mit denen aus Bild 4.28 bzw. Bild 4.29 überein. Aufgrund der selektiven Detektion des Gehörmodells, die auf die Sektionen nahe der Testsignalresonanz beschränkt ist, wird eine deutliche Überschreitung des internen Schwellwertes bei multipliziertem Rauschen in den Minima der spezifischen Lautheit des Maskierers beobachtet. Das Pseudoräuschen zeigt demgegenüber eine wesentlich stärkere Verdeckung des Testsignals, da hier keine ausgeprägten Minima vorhanden sind. Die kubischen Verzerrungskomponenten führen in beiden Fällen zu einer deutlichen Überschreitung des Schwellwerts. Dies wird aber aufgrund der selektiven Detektion nicht ausgewertet.

Das vom Gehörmodell verwendete Maskierersignal für die Bestimmung der spezifischen Lautheit als Referenz ist mit dem Maskierersignal identisch, dem das Testsignal zur Bestimmung der Mithörschwelle überlagert wird. In psychoakustischen Messungen werden abweichend davon beispielsweise bei Rauschmaskierern häufig unterschiedliche, zufällig ausgewählte Rauschsignale von einem Rauschgenerator verwendet. Dadurch ist eine Variabilität des Maskierers in aufeinander folgenden Testintervallen bedingt, die im allgemeinen die Detektierbarkeit eines Testsignals aufgrund des erschwerten Vergleichs zwischen den Intervallen vermindert. Durch die Variabilität ändert sich auch die zeitliche Position der Minima der Maskierereinhüllenden in jedem Intervall, so daß dort die Testsignal detektion

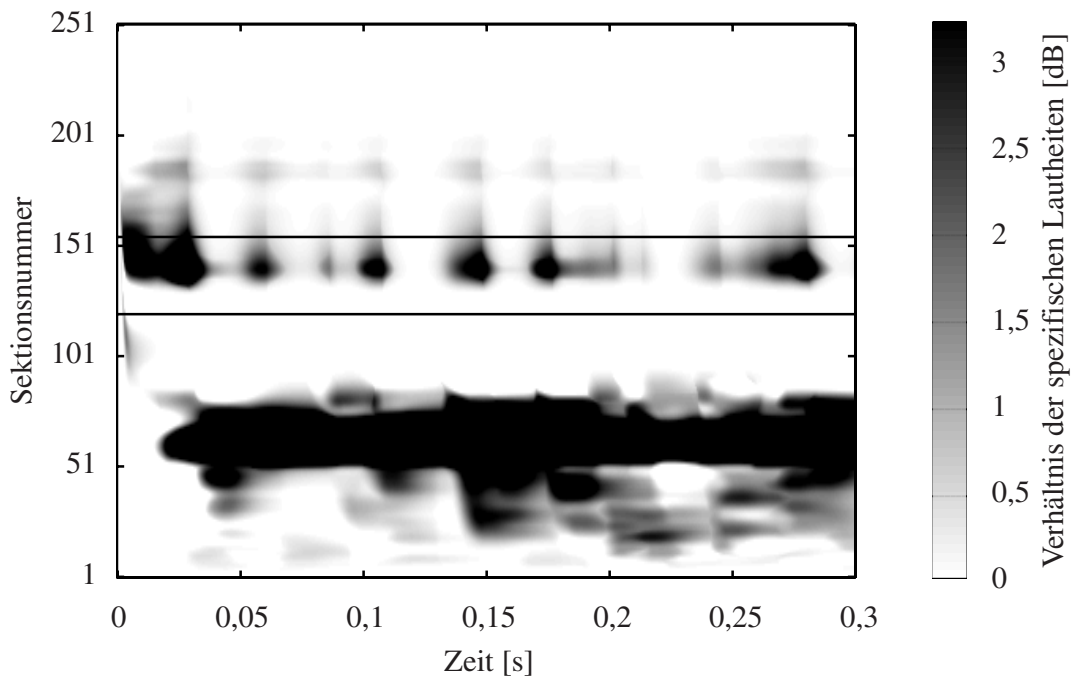


Bild 4.30 Verhältnis der spezifischen Lautheiten $L_{r,N',y}$ von multipliziertem Rauschen (80 dB SPL) allein und mit überlagertem 2 kHz Sinuston (50 dB SPL). Die selektive Detektion einer Signaländerung erfolgt in den Sektionen zwischen den horizontal eingezeichneten Linien.

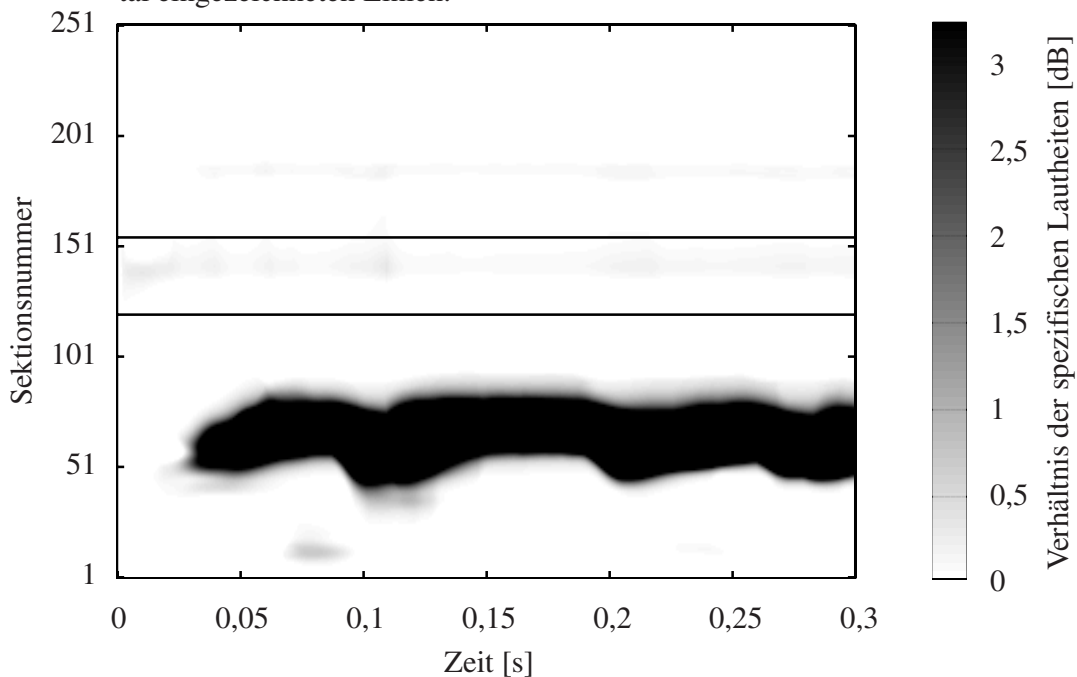


Bild 4.31 Verhältnis der spezifischen Lautheiten $L_{r,N',y}$ von Pseudoräuschen (80 dB SPL) allein und mit überlagertem 2 kHz Sinuston (50 dB SPL). Die selektive Detektion einer Signaländerung erfolgt in den Sektionen zwischen den horizontal eingezeichneten Linien.

nicht durch Vergleiche verbessert werden kann. Die von der Maskierervariabilität verursachte erhöhte Verdeckung („*informational masking*“ [123, 139, 157]) wird mit dem Gehörmodell derzeit nicht nachgebildet.

Die nachgebildeten zeitabhängigen Verdeckungseffekte umfassen die Vor- und Nachverdeckung sowie den Einfluß der Testsignaldauer auf die Simultanverdeckung. Die herangezogenen psychoakustischen Referenzdaten wurden zum Parameterabgleich der NV-Modelle verwendet. Die NV-Modellparameter sind dabei weitgehend unabhängig voneinander einstellbar, so daß beispielsweise die Parameter zum Abgleich der Vorverdeckung einen vernachlässigbaren Einfluß auf die Nachverdeckung haben und umgekehrt. Die Adaption der Zeitkonstanten und Gewichtsfaktoren an den Pegel und die Dauer des Maskierers gestatten die Berücksichtigung des Einflusses dieser Parameter auf die Nachverdeckung. Für die Nachbildung der Nachverdeckung wurde die bei der Entladungsrate des Hörnervs beobachtete kurzzeitige Abnahme unterhalb der spontanen Rate nach Abschalten eines Schallsignals im Gehörmodell nicht berücksichtigt (vgl. Bild 2.13). Obwohl der Zeitverlauf der Entladungsrate in diesem Fall abweicht, werden dennoch die Eigenschaften der Nachverdeckung richtig wiedergegeben. Eine mögliche Begründung für dieses Verhalten besteht darin, daß die Unterdrückung der spontanen Rate nach Abschalten des Maskierers in gleichem Maß auf das Testsignal und damit auf die Verdeckung wirkt. Im Gehörmodell wird statt dieser Unterdrückung eine Verdeckung aufgrund der abklingenden Erregung durch den Maskierer angenommen, die offenbar einen vergleichbaren Zeitverlauf der Verdeckung aufweist.

Die Nachbildung der Vorverdeckung ist aus folgenden Gründen problematisch. Psychoakustische Messungen der Vorverdeckung zeigen im Vergleich zur Nachverdeckung höhere interindividuelle Abweichungen. Die Messung der Zeitkonstanten der Vorverdeckung wird zusätzlich durch die sehr kritische Wahl des geeigneten Maskierers und Testsignals erschwert. Die Nachbildung durch das Gehörmodell berücksichtigt keinen Overshoot-Effekt, der im Einschaltmoment breitbandiger Maskierer zu einer signifikant erhöhten Wahrnehmungsschwelle führt. Der Abgleich der Vorverdeckung beschränkt sich daher auf die Steilheit des Zeitverlaufs bzw. die angenommene Zeitkonstante der Vorverdeckung. Eine Nachbildung des Overshoot-Effekts erscheint grundsätzlich durch den Einsatz eines komplexeren Modells der IHZ möglich, das auch die auf den Nervenfasern beobachteten schnellen Adaptionsvorgänge bei einsetzendem Reizsignal nachbildet, z. B. [127, 128, 129]. Experimente mit solchen Modellen führten jedoch bislang zu einer insgesamt ungenaueren Nachbildung der Wahrnehmungsschwellen.

Der frequenzabhängige Verlauf der Ruhehörschwelle wird bei Berücksichtigung der Abweichungen verschiedener Meßdaten durch die Simulation zufriedenstellend angenähert. Der Verlauf wird im wesentlichen durch den Frequenzgang des AMO-Modells und die frequenzabhängige Verstärkung des Cochlea-Modells bei kleinen Eingangspegeln bestimmt (vgl. Bild 4.2). Für kürzere Testsignale, die bei der Simulation der Vor- bzw. Nachverdeckung verwendet wurden, sind dennoch Abweichungen zu beobachten. Die Testsignale mit einer Dauer von 5 bzw. 1 ms zeigen dort eine um etwa 18 bzw. 7 dB geringere Ruhehörschwelle. Für den Testimpuls mit 20 μ s Dauer ergibt sich in der Simulation eine um etwa 10 dB höhere Ruhehörschwelle.

Die in der Audiocodierung eingesetzten psychoakustischen Modelle sind für eine möglichst präzise Nachbildung der Simultanverdeckung ausgelegt, während die Vor- und Nachverdeckung in der Regel nicht explizit berücksichtigt wird. Diese Realisierungsform beruht auf der Erkenntnis, daß die Irrelevanzreduktion vieler Audiosignale durch eine präzisere Nachbildung der Vor- und Nachverdeckung nur verhältnismäßig wenig verbessert werden kann. Die wesentlichen Codierungsgewinne bei der Irrelevanzreduktion werden in der Regel durch die Ausnutzung der Simultanverdeckung erzielt. Insbesondere bei Audiosignalen mit ausgeprägt transienten Komponenten sind jedoch Vorteile durch die bessere Nachbildung der zeitabhängigen Verdeckung durch das Gehörmodell zu erwarten.

Das Gehörmodell erreicht insbesondere bei der Simultanverdeckung eine gute Übereinstimmung der Wahrnehmungsschwellen mit den Referenzmessungen. Damit erfüllt es eine wesentliche Bedingung für den Einsatz in der Audiocodierung oder anderen Anwendungen in der Kommunikationstechnik. In Verbindung mit der expliziten Nachbildung zeitabhängiger Verdeckungseffekte werden auch kombinierte zeit- und frequenzabhängige Verdeckungseffekte adäquat nachgebildet. Solche Effekte treten beispielsweise bei Gesang oder Sprache auf, wenn die Grundfrequenz moduliert ist. Die Verdeckungswirkung hängt im allgemeinen von der Grundfrequenzänderung ab, d. h. bei gleicher Momentanfrequenz ist die Wirkung aufgrund der Nachverdeckung bei zunehmender oder abnehmender Grundfrequenz unterschiedlich.

Bei der binauralen Wahrnehmung kann eine verminderte Verdeckung eines Testsignals beispielsweise eintreten, wenn es gegenüber dem Maskierer an einer anderen räumlichen Position lokalisiert wird. Diese reduzierte Verdeckungswirkung („binaural masking level difference (BMLD)“) kann durch das Gehörmodell aufgrund der nur monauralen Nachbildung bislang nicht simuliert werden. Zur Erfassung dieser Effekte müßte das Modell um eine Einheit zur Nachbildung der binauralen Wahrnehmung erweitert werden.

5 Einbindung des Gehörmodells in ein Audiocodierverfahren

Das Gehörmodell wird in einem Anwendungsbeispiel zur Datenkompression in einem Audiocodierverfahren eingesetzt. Das dazu ausgewählte Referenzcodierverfahren ist als ISO/IEC MPEG-2 Advanced Audio Coding (AAC) standardisiert [101]. Im Vergleich zu anderen Verfahren benötigt es derzeit die geringste Datenmenge zur Repräsentation von allgemeinen Audiosignalen mit hoher Klangqualität, bei der Unterschiede zwischen dem Original- und dem decodierten Audiosignal praktisch nicht wahrnehmbar sind [30, 177]. Die erreichbare Klangqualität wurde in subjektiven Hörtests im Zuge der Standardisierung gemessen und dokumentiert [28, 30, 97, 98, 99, 100, 102].

Bislang basiert die Irrelevanzreduktion dieses Codierverfahrens auf einem psychoakustischen Wahrnehmungsmodell, das durch das Gehörmodell ersetzt werden soll. Der so modifizierte Coder erlaubt damit eine implizite Verifikation des Gehörmodells für beliebig komplexe Eingangssignale. Auf der Grundlage eines Vergleichs der Codiererergebnisse mit dem Referenzcoder kann gezeigt werden, ob auch für komplexe Signale die Wahrnehmungsschwellen richtig nachgebildet werden. Die Verifikation der Wahrnehmungsschwellen komplexer Eingangssignale anhand psychoakustischer Messungen wäre demgegenüber wesentlich aufwendiger und ist daher im allgemeinen nicht durchführbar.

Die mit dem modifizierten Coder, bei dem das psychoakustische Modell durch das Gehörmodell ersetzt ist, erzielte Klangqualität wird anhand von Bewertungen aus Hörtests mit dem Referenzcodierverfahren verglichen. Die unterschiedlichen Eigenschaften des psychoakustischen Modells bzw. des Gehörmodells werden durch einen Vergleich des erzeugten Signal-Maskierungs-Abstands beider Modelle analysiert.

5.1 Das Codierverfahren nach ISO/IEC MPEG-2 AAC

Gegenstand des AAC-Standards [101] ist das Bitstromformat und das Decodierverfahren. Damit ist der Standard für Encodermodifikationen offen, soweit sie nicht zu Veränderungen bei den standardisierten Teilen führen. Sinnvolle Modifikationen zur Optimierung der Klangqualität bestehen beispielsweise in Verbesserungen des Wahrnehmungsmodells oder der Bitzuteilungsstrategie, die unten erläutert wird. Eine beispielhafte Encoderrealisierung ist im Standard mit der Funktionsweise der darin enthaltenen Einheiten angegeben. Die im folgenden gegebene Beschreibung beschränkt sich im wesentlichen auf die in dieser Arbeit relevanten Einheiten des Encoders bzw. Decoders. Eine ausführlichere Darstellung enthält [28].

Der Encoder (s. Bild 5.1) verarbeitet aufeinander folgende Abtastwerte innerhalb eines Blocks des PCM-Audiosignal gemeinsam. Ein Block umfaßt 2048 Abtastwerte. Aufeinander folgende Blöcke besitzen einen Vorschub von 1024 Abtastwerten und damit eine 50%ige Überlappung. Ein Block wird am Eingang des Encoders mit Hilfe einer Filterbank in Teilbandsignale zerlegt. Vor der Filterbank wird eine geeignete Fensterung des Blocks mit einem

langen Fenster oder acht kurzen Fenstern durchgeführt. Bei der Verwendung des langen Fensters, das 2048 Abtastwerte umfaßt, wird am Ausgang der Filterbank jeweils ein spektraler Abtastwert in 1024 Teilbändern erzeugt. Bei einer Abtastfrequenz von $f_A = 44,1$ kHz beträgt der Abstand der Teilbandmittenfrequenzen dann 21,5 Hz, und der Fenstervorschub ist 23 ms. Ein kurzes Fenster umfaßt 256 Abtastwerte und hat eine 50%ige Überlappung mit dem benachbarten kurzen Fenster innerhalb des Blocks. Bei Verwendung kurzer Fenster wird die Filterbank mit reduzierter Teilbandanzahl betrieben, so daß die sequentielle Verarbeitung der acht gefensterten Teilfolgen eines Blocks zu jeweils acht zeitlich aufeinander folgenden spektralen Abtastwerten in 128 Teilbändern führt. Bei $f_A = 44,1$ kHz ergibt sich damit ein Abstand der Teilbandmittenfrequenzen von 172 Hz und ein Fenstervorschub von 2,9 ms. Bei der Umschaltung zwischen einem langen Fenster und kurzen Fenstern werden geeignete „Umschaltfenster“ verwendet.

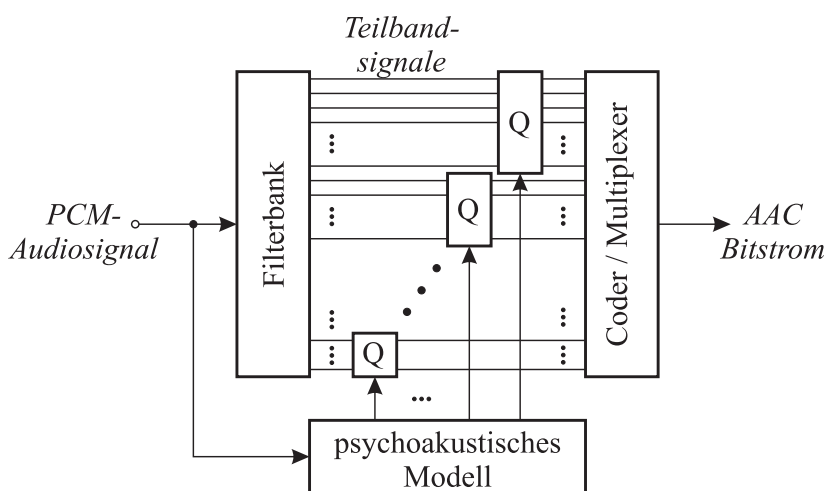


Bild 5.1 Vereinfachtes Blockdiagramm des AAC Encoders. Der Quantisierer (Q) in einem Skalenfaktorband wird individuell auf jeden Abtastwert in den Teilbänder des Skalenfaktorbandes angewendet.

Auf die momentanen spektralen Abtastwerte der Teilbänder kann eine Prädiktionstechnik angewendet werden, die spektrale Abtastwerte aus den zeitgleichen Abtastwerten in anderen Teilbändern prädiziert. Diese Prädiktion kann auch als Filterung im Spektralbereich interpretiert werden, die aufgrund der Dualität von Spektral- und Zeitbereich nach einer Quantisierung und Rekonstruktion zu einer entsprechenden Formung des Quantisierungsrauschens im Zeitbereich führt. Diese Rauschformung („temporal noise shaping (TNS)“) kann für eine verbesserte dynamische Anpassung des Rauschens an die Wahrnehmungsschwelle genutzt werden [86].

Die Prädiktion zeitlich aufeinander folgender Abtastwerte innerhalb einzelner Teilbänder wird ebenfalls unterstützt, um evtl. vorhandene zeitliche Korrelationen auszunutzen [59].

Die Teilbänder am Ausgang der Filterbank werden in sog. „Skalenfaktorbändern“ zusammengefaßt, deren Bandbreite an die spektrale Auflösung des Gehörs angepaßt ist. Auf

die zeitgleichen spektralen Abtastwerte wird ein Quantisierer angewendet, der für alle Teilbänder innerhalb des Skalenfaktorbandes identisch ist. Die Quantisierer besitzen eine ungleichförmige Kennlinie, die eine der Wahrnehmungsschwelle besser angepaßte spektrale Rauschformung bewirkt als eine gleichförmige Quantisierung. Dies wird durch eine größere Quantisierung größerer Werte erreicht. Die Quantisierungsrauschleistung in einem Skalenfaktorband wird durch die Quantisiererstufengröße bestimmt.

Die nach der Quantisierung verbleibende Redundanz wird durch nachfolgende Huffmancodierung reduziert. Die für eine Decodierung notwendigen Seiteninformationen werden ebenfalls in geeigneter Weise codiert. Die codierten spektralen Abtastwerte werden zusammen mit den Seiteninformationen am Ausgang des Encoders zu einem standardgemäßen Bitstrom formatiert und ausgegeben.

Falls eine Codierung ohne Vorgabe einer festen Zielbitrate erfolgt, kann eine sinnvolle Einstellung der Quantisiererstufengröße allein aus der vom psychoakustischen Modell generierten Mithörschwelle bestimmt werden. Die Quantisiererstufengröße wird in der Regel so eingestellt, daß die Wahrnehmungsschwelle von der zu erwartenden maximalen Quantisierungsrauschleistung in einem Skalenfaktorband gerade erreicht wird. Die Audiocodierung soll dementsprechend in dem Anwendungsbeispiel des Gehörmodells mit variabler Datenrate erfolgen. Damit wird erreicht, daß die Quantisiererstufengröße nicht von einer bei konstanter Datenrate evtl. unzureichenden Anzahl zur Verfügung stehender Bits mitbestimmt wird. Aufgrund der Codierung mit variabler Datenrate und der Einschränkung auf einkanalige Audiosignale werden die folgenden vorhandenen Funktionseinheiten des Coders nicht genutzt, sie sind daher hier von untergeordneter Bedeutung.

Bei Vorgabe einer festen Zielbitrate, zum Beispiel einer konstanten Rate, kann es zu einer Überschreitung der generierten Mithörschwelle durch das Quantisierungsrauschen kommen, wenn der Bedarf an Bits größer ist, als aufgrund der vorgegebenen Bitrate zur Verfügung stehen. In diesem Fall werden die als wahrnehmbar angenommenen Störungen durch eine geeignete, von einem Bitzuteilungsalgorithmus gesteuerte Quantisierereinstellung zeitlich und spektral verteilt, um den störenden Charakter dieser Anteile bei der Wahrnehmung zu reduzieren.

Für die Codierung von mehrkanaligen Audiosignalen können Verfahren zur Ausnutzung der in den Kanälen ggf. vorhandenen Redundanz bzw. Irrelevanz eingesetzt werden. Die Verfahren umfassen das sog. „Intensitäts-Stereo“ und das Mitte/Seite (M/S)-Stereo.

Zur Unterstützung kleinerer Abtastfrequenzen unterhalb 32 kHz kann eine Signalvorverarbeitung verwendet werden. Diese sog. „Gain Control“ führt in diesem Fall zu einer erhöhten Codierungseffizienz.

5.2 Das psychoakustische Basismodell zur Quantisierersteuerung des AAC

Der Standard beschreibt ein psychoakustisches Basismodell, das die Grundlage vieler bekannter Implementierungen bildet. An diesem Basismodell können daher die prinzipiellen

Eigenschaften bezüglich der Wahrnehmungsschwellennachbildung gezeigt werden. Die folgende Darstellung des psychoakustischen Modells ist an dieses Basismodell angelehnt.

Für das psychoakustische Modell wird eine separate Spektralzerlegung des Audiosignals mit Hilfe von unterschiedlichen FFTs durchgeführt. Die lange FFT verwendet ein 2048 Abtastwerte umfassendes Fenster, während die kurze FFT eine Fensterlänge von 256 Abtastwerten des Audiosignals verwendet und auf 8 überlappende Teile des Blocks angewendet wird. Die Fenstergrößen entsprechen denen der Filterbank und erlauben eine hohe spektrale Auflösung bei dem langen Fenster, sowie eine feine zeitliche Auflösung bei kurzen Fenstern. Durch die Anwendung der FFT wird eine spektrale Darstellung in Form von Betrag und Phase ermöglicht, die für die nachfolgende Tonalitätsschätzung verwendet wird. Für die Tonalitätsschätzung werden die aus zwei vorangehenden FFTs prädizierten Werte des Betrags und der Phase mit den berechneten Werten der aktuellen FFT verglichen. Bei einer Übereinstimmung wird die entsprechende Spektralkomponente als tonal angenommen, während bei größer werdender Abweichung auf einen zunehmend rauschartigen Charakter geschlossen wird. Das derart bestimmte, sog. „Nichtvorhersagbarkeitsmaß“ wird mit der Leistung der FFT-Spektralkomponenten gewichtet.

Die Leistung der Spektralkomponenten wird genauso wie das gewichtete Nichtvorhersagbarkeitsmaß in sog. „Partitionen“ zusammengefaßt, die eine Bandbreite von etwa $1/3$ Bark besitzen. Die derart „partitionierte“ Leistung wie auch das partitionierte Nichtvorhersagbarkeitsmaß wird dann mit einer Spreizfunktion („spreading function“) gefaltet, die an gemessene Mithörschwellenverläufe schmalbandiger Maskierer angelehnt, bzw. als eine Erregungsfunktion betrachtet werden kann. Für die nachfolgenden Berechnungen werden die Ergebnisse in geeigneter Weise normiert, so daß sie später auf die Teilbandabtastwerte der Filterbank angewendet werden können. Das Nichtvorhersagbarkeitsmaß wird in einen Tonalitätsindex für jede Partition umgerechnet. Der Tonalitätsindex bestimmt die Pegeldifferenz zwischen der gespreizten Leistung und der gesuchten Wahrnehmungsschwelle, die mit zunehmender Tonalität wächst. Die Wahrnehmungsschwelle wird daher mit dem Tonalitätsindex für jede Partition bestimmt. Zur Anwendung der Wahrnehmungsschwelle auf die Teilbandsignale der Filterbank werden die auf Partitionen bezogenen Werte auf die Bandbreiten der Skalenfaktorbänder normiert. Die derart umgerechnete Wahrnehmungsschwelle bestimmt dann zusammen mit den Leistungen in den Skalenfaktorbändern den sog. „Signal-Maskierungs-Abstand“ (SMR), der letztlich zur Quantisierersteuerung verwendet wird. Bild 4.12 und Bild 4.13 zeigen Beispiele von entsprechend berechneten Wahrnehmungsschwellen für Sinustonmaskierer vom psychoakustischen Basismodell.

Das psychoakustische Modell nimmt zusätzlich eine geeignete Steuerung des SMR zur Vermeidung sog. „Vorechos“ vor, die bei Audiosignalen mit kurzzeitig ansteigender Leistung auftreten können. Vorechos können aufgrund der begrenzten Zeitauflösung des Coders entstehen, da sich das Quantisierungsrauschen bei der Decodierung auf die Gesamtdauer der Fensterfunktion auch dann ausbreitet, wenn ein Audiosignal erst beispielsweise während der zweiten Hälfte der Fensterfunktion einsetzt. Vorechos sind wegen der kurzen Vorverdeckungsdauer besonders kritisch. Bei schnellen Änderungen der Audiosignalenergie wird die Filterbank in der Regel durch das psychoakustische Modell auf kurze Fensterfunktionen

umgeschaltet. Die damit erzielte feinere Zeitauflösung bei der Codierung wird zur besseren Anpassung des Quantisierungsrauschens an die sich schnell ändernde Wahrnehmungsschwelle genutzt.

Das psychoakustische Modell berücksichtigt im Gegensatz zum Gehörmodell einige wesentliche Wahrnehmungseigenschaften nicht. So wird die Abhängigkeit der Verdeckungswirkung von der Einhüllendenfluktuation nicht explizit vom psychoakustischen Modell nachgebildet. Die Pegelabhängigkeit der Verdeckung und Suppressionseffekte werden nicht berücksichtigt. Die nichtlineare Additivität überlagerter Maskierersignale wird ebenfalls nicht nachgebildet.

Der hier verwendete Referenzcoder beinhaltet ein optimiertes psychoakustisches Modell, das gegenüber dem Basismodell erweitert und verbessert ist. Die für einen Vergleich mit dem Gehörmodell relevanten Eigenschaften des optimierten psychoakustischen Modells werden jedoch weitgehend bereits durch die Eigenschaften des Basismodells bestimmt. Auf eine Darstellung der Erweiterungen des optimierten Modells wird daher in dieser Arbeit verzichtet.

5.3 Quantisierersteuerung des AAC mit dem Gehörmodell

Die in Kapitel 3.5 erläuterte Generierung einer Wahrnehmungsschwelle mit dem Gehörmodell geht von einem bekannten Maskierer und Testsignal aus. Das durch die Audiocodierung erzeugte Störsignal, für das die Wahrnehmungsschwelle bestimmt werden soll, ist jedoch erst nach erfolgter Codierung und Decodierung genau bekannt. Da das Test- bzw. Störsignal in dieser Anwendung somit nicht vordefiniert ist, wurden zwei Ansätze zur Lösung des Problems verfolgt. Der erste Ansatz besteht in einer Abschätzung des Störsignals, das nach erfolgter Codierung und Decodierung mit einer gegebenen Quantisierereinstellung zu erwarten ist. Das zu erwartende Störsignal wird dann auf dieser Grundlage generiert. Bei dem zweiten Ansatz wird die Codierung und Decodierung mit einer vorgegebenen initialen Quantisierereinstellung durchgeführt. Das decodierte Audiosignal beinhaltet dann das gesuchte Störsignal.

In Voruntersuchungen wurde festgestellt, daß mit dem ersten Ansatz bei vertretbarem Aufwand im Vergleich zum zweiten Ansatz nur eine unzureichende Approximation des tatsächlich auftretenden Störsignals zu erreichen ist. Daher wurde der zweite Ansatz gewählt, der auch aufgrund der exakten Störsignalgenerierung vorteilhaft ist.

Die erforderliche initiale Quantisierereinstellung des Coders wird durch ein einfaches Verfahren bestimmt, das von einer groben Näherung der Wahrnehmungsschwelle ausgeht. Diese Näherung wird aus den Leistungen der Skalenfaktorbänder nach einem Verfahren bestimmt, das in Bild 5.2 veranschaulicht ist. Ausgehend von den Maxima der Skalenfaktorbandleistungen werden zwei Flankensteilheiten angesetzt, die nach einer Absenkung um 10 dB und einer Berücksichtigung der Ruhehörschwelle die approximierte Wahrnehmungsschwelle ergeben. Die Flankensteilheit auf der Seite kleinerer Frequenzen beträgt 15 dB pro Skalenfaktorband, auf der Seite höherer Frequenzen -3 dB pro Skalenfaktorband. Der SMR ist in der Darstellung nach Bild 5.2 die Pegeldifferenz von Leistung und Wahrnehmungsschwelle.

Der Coder stellt die Stufengröße des Quantisierers in jedem Skalenfaktorband so ein, daß die Quantisierungsfehlerleistung des betragsmäßig größten Teilbandabstastwertes gerade die Wahrnehmungsschwelle erreicht. Nach der Codierung wird das decodierte Audiosignal dem Gehörmodell zugeführt. Das originale Audiosignal wird vorher als Referenzsignal in das Gehörmodell eingespeist. Der Grad der Wahrnehmbarkeit von Signaländerungen d_v wird dann vom Gehörmodell durch einen Vergleich der spezifischen Lautheiten beider Signale ermittelt (vgl. Kapitel 3.4.4). Falls der Grad der Wahrnehmbarkeit in einer Sektion über der Schwelle liegt ($d_v > 1$), war die Quantisierung in mindestens einem Skalenfaktorband zu grob. Befindet sich der Grad der Wahrnehmbarkeit unterhalb der Schwelle ($d_v < 1$), so war die Quantisierung evtl. zu fein. Die Quantisierereinstellung ist daher im nächsten Iterationsschritt so zu modifizieren, daß möglichst in allen Sektionen der Grad der Wahrnehmbarkeit die Schwelle erreicht ($d_v = 1$). Das Blockdiagramm des Encoders mit der Gehörmodellsteuerung ergibt sich aus diesem Ansatz gemäß Bild 5.3.

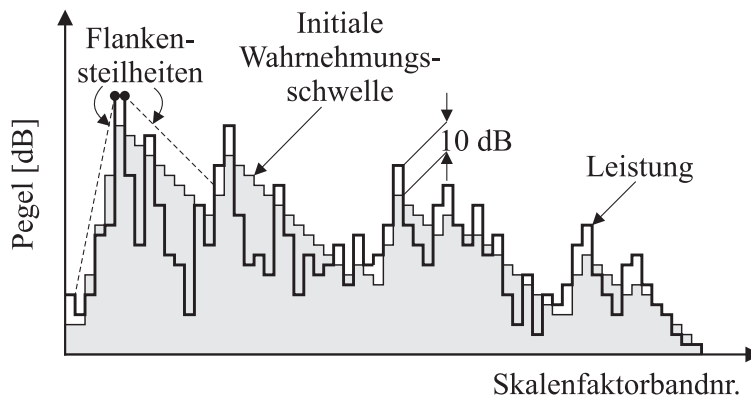


Bild 5.2 Initiale Bestimmung der Wahrnehmungsschwelle und des Signal-Maskierungs-Abstandes (SMR) für den ersten Iterationsschritt der Quantisierereinstellung.

Die Quantisierereinstellung mit Hilfe des Gehörmodells erfolgt über eine Steuereinheit. Diese Steuereinheit stellt eine Erweiterung des iterativen Verfahrens zur Bestimmung der Wahrnehmungsschwelle aus Kapitel 3.5 dar. Sie bestimmt aus dem Grad der Wahrnehmbarkeit als Ausgangsgröße des Gehörmodells eine geeignete Modifikation der aktuellen Quantisierereinstellung für den nächsten Iterationsschritt. Der Grad der Wahrnehmbarkeit wird vom Gehörmodell für alle 251 Sektionen als zeitabhängige Größe bestimmt. Das zu lösende Problem besteht nun darin, bei einer Über- oder Unterschreitung der Schwelle zu einem bestimmten Zeitpunkt in einer Sektion, denjenigen codierten Block und das Skalenfaktorband zu bestimmen, in dem diese Abweichung zur Schwelle verursacht wird. Nur wenn das gelingt, kann diese Abweichung durch eine Modifikation der Einstellung des entsprechenden Quantisierers gezielt verringert werden.

Ein praktikabler Ansatz zur Lösung des Problems ist die Annahme, daß die Quantisierung in einem Skalenfaktorband nur zu Änderungen der spezifischen Lautheit in den Sektionen führt, deren Resonanzfrequenz innerhalb des Skalenfaktorbandes liegt. Damit ist festgelegt, daß Abweichungen vom Schwellwert ggf. von dem Quantisierer des entsprechenden Skalenfaktorbandes verursacht worden sind. Zur Festlegung des Blocks, in dem diese

Abweichung verursacht wurde, werden die unterschiedlichen Signallaufzeiten im Coder und im Gehörmodell berücksichtigt. Während die Laufzeit im Coder unabhängig vom betrachteten Skalenfaktorband ist, nimmt die Laufzeit im Gehörmodell für Sektionen mit tieferen Resonanzfrequenzen zu. Demzufolge wird in der Sektion mit der niedrigsten Resonanzfrequenz eine Abweichung von der Schwelle mit der größten Verzögerung gegenüber der verursachenden Quantisierung festgestellt. Die maximale Laufzeit im Gehörmodell beträgt etwa 10 ms [10] und liegt damit unterhalb einer Blocklänge. Die oben genannten Voraussetzungen gestatten eine eindeutige Zuordnung des vom Gehörmodell generierten Wahrnehmbarkeitsgrads jeder Sektion und für jeden Zeitpunkt zu einem bestimmten Skalenfaktorband und einem Block bzw. Teilblock bei kurzen Fenstern des Audiosignals.

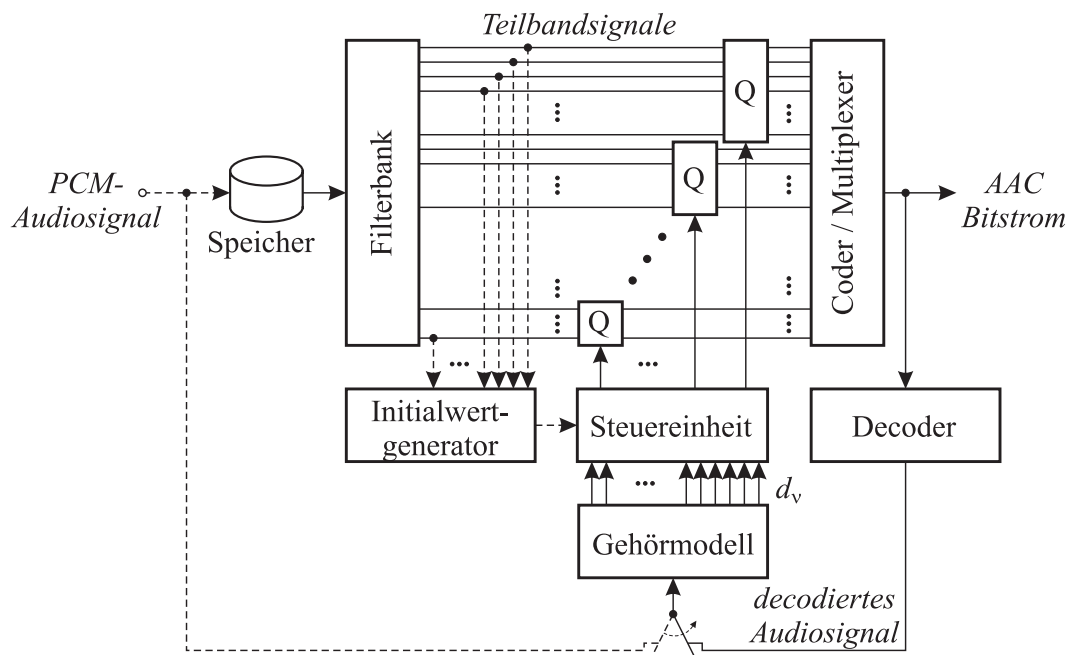


Bild 5.3 Vereinfachtes Blockdiagramm eines AAC Encoders mit der Quantisieresteuerung durch das Gehörmodell. Das iterative Verfahren beginnt im ersten Schritt mit einer initialen Quantisierereinstellung durch den Initialwertgenerator. In den nachfolgenden Iterationsschritten wird die Quantisierereinstellung durch das Gehörmodell modifiziert, wobei die *gestrichelt* dargestellten Signale inaktiv sind.

Die für die Quantisieresteuerung getroffene Zuordnung der einzelnen Sektionen des Gehörmodells zu den Skalenfaktorbändern anhand ihrer Resonanzfrequenz ist jedoch für die Verminderung von wahrnehmbaren Signaländerungen aufgrund von evtl. auftretenden Verzerrungsprodukten ungeeignet. Das von der Codierung verursachte Quantisierungsrauschen kann mit den Spektralkomponenten des Audiosignals Verzerrungsprodukte im Gehör bilden. Das in der Regel am ausgeprägtesten wahrnehmbare Verzerrungsprodukt tritt bei der kubischen Differenzfrequenz auf. Die Frequenz dieser Verzerrung ergibt sich aus dem Frequenzabstand von zwei sich überlagernden Spektralkomponenten, die das Verzerrungsprodukt bil-

den. Aufgrund der schmalbandigen Teilbandzerlegung der Filterbank im AAC hat das von einer Spektralkomponente in einem Teilband verursachte Quantisierungsrauschen näherungsweise die gleiche Bandbreite wie das Teilband. Der Frequenzabstand der Spektralkomponente zum Quantisierungsrauschen ist daher durch die Bandbreite des Teilbands beschränkt. Diese Spektralkomponente bildet aber Verzerrungsprodukte mit anderen Spektralkomponenten des Audiosignals wie das Quantisierungsrauschen in dem gleichen Teilband. Daher ist bei der kubischen Differenzfrequenz immer ein Verzerrungsprodukt vorhanden, das allein aufgrund der Überlagerung von zwei Spektralkomponenten des Audiosignals entsteht und eine größere Amplitude hat, als das Verzerrungsprodukt, welches mit dem Quantisierungsrauschen gebildet wird. Das Verzerrungsprodukt mit dem Quantisierungsrauschen wird daher in der Regel verdeckt.

Zur Bestimmung einer verbesserten Quantisierereinstellung wird die maximale Abweichung des Wahrnehmbarkeitsgrads vom Schwellwert ($d_v = 1$) für jeden Block durch Auswertung der Ausgänge der zugeordneten Sektionen des Gehörmodells ermittelt. Dieses Maximum bestimmt die Änderung des SMR, aus der die Quantisiererstufengröße im Coder bestimmt wird. Die Änderung des SMR gegenüber der vorangegangenen Einstellung erfolgt nach (5.1). $\text{SMR}_{\text{SFB}}(i)$ bezeichnet darin den (nicht logarithmierten) Signal-Maskierungs-Abstand im aktuellen Iterationsschritt i . \hat{d}_{SFB} ist die maximale Abweichung des Wahrnehmbarkeitsgrads d_v vom Schwellwert in den Sektionen, die einem Skalenfaktorband zugeordnet sind. Für unterschwellige Signaländerungen ($\hat{d}_{\text{SFB}} < 1$) führt die Verwendung der Quadratwurzel in (5.1) nur zu einer geringeren Erhöhung des SMR im Vergleich zu der Absenkung des SMR bei überschwelligen Änderungen. Damit wird ein „Überschwingen“ im Verlauf der Iteration vermieden.

$$\text{SMR}_{\text{SFB}}(i + 1) = \begin{cases} \text{SMR}_{\text{SFB}}(i) / (\hat{d}_{\text{SFB}} + 0,05) & ; \text{ für } \hat{d}_{\text{SFB}} \geq 1 \\ \text{SMR}_{\text{SFB}}(i) / \sqrt{\text{Max}(\hat{d}_{\text{SFB}}, 0, 1)} & ; \text{ für } \hat{d}_{\text{SFB}} < 1 \end{cases} \quad (5.1)$$

Mit der geänderten Quantisierereinstellung wird eine erneute Codierung durchgeführt. Nach der Decodierung wird der Grad der Wahrnehmbarkeit von Signaländerungen gegenüber dem Referenzsignal durch Einspeisen des decodierten Audiosignals in das Gehörmodell bestimmt. Mit (5.1) kann daraus entweder wieder eine verbesserte Quantisierereinstellung bestimmt werden, oder die Iteration wird beendet, wobei die codierten Daten als Endergebnis übernommen werden.

In Bild 5.4 ist der Grad der Wahrnehmbarkeit von Signaländerungen eines codierten Audiosignals in 5 Iterationen dargestellt. In der ersten Iteration treten deutliche Überschreitungen der Wahrnehmungsschwelle auf, die von der initialen Quantisierereinstellung hervorgerufen werden. In den nachfolgenden Iterationen werden diese Überschreitungen schrittweise reduziert. Da bei deutlicher Unterschreitung der Wahrnehmungsschwelle für

die nächste Iteration auch eine gröbere Quantisierung erfolgen kann, sind auch temporär in einer Iteration auftretende Überschreitungen der Schwelle möglich.

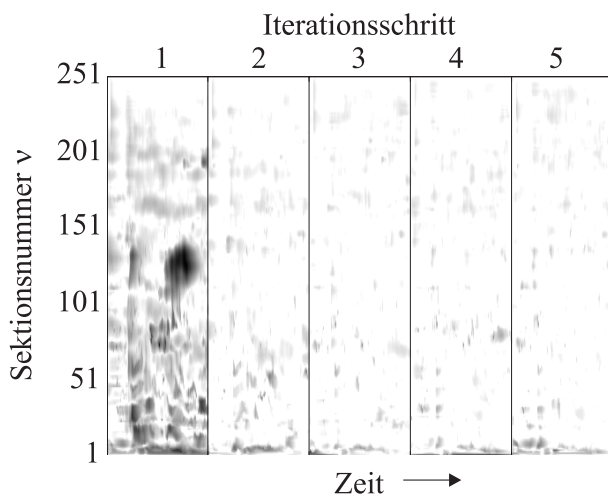


Bild 5.4 Qualitative Darstellung des Grads der Wahrnehmbarkeit in aufeinander folgenden Iterationsschritten. Große Werte erscheinen dunkler, in weißen Bereichen wird kein wahrnehmbarer Unterschied vom Hörmodell detektiert. Das verwendete Audiosignal ist „Deutscher Sprecher 1“. Die Abszisse repräsentiert in jedem Iterationsschritt die Dauer des Eingangssignals.

Die iterative Quantisierereinstellung mit dem Hörmodell wurde für die Codierung der hier gezeigten Audiosignale so implementiert, daß innerhalb einer Iteration das gesamte Audiosignal verarbeitet wird. Dieses Verfahren ist jedoch wegen des zunehmenden Speicheraufwands und der erhöhten Signalverzögerung für Audiosignale mit größerer Dauer ungeeignet. In diesem Fall ist die iterative Quantisierereinstellung individuell für jeden Block des Audiosignals durchzuführen. Gegenüber dem Referenzcoder wäre dabei im allgemeinen mit einer zusätzlichen Codiervverzögerung von einem Block zu rechnen, die von der Signallaufzeit innerhalb des Hörmodells verursacht wird.

Zur Vermeidung zu großer zeitlicher Schwankungen des SMR wird eine zeitliche Glättung der SMR-Werte vorgenommen, die jeweils zwei vorangehende und zwei nachfolgende Werte für die Glättung einbezieht (gewichtete Mittelung). Falls die Iterationen bei der Codierung für jeden Block individuell durchgeführt werden sollen, wäre demgegenüber eine rekursive Glättung angebracht, die keine zusätzliche Verzögerung bei der Codierung erfordert.

Bei der Codierung mit dem modifizierten Coder wurde die Steuerung der Fensterlängen vom Referenzcoder übernommen, die dort vom psychoakustischen Modell gesteuert wird. Bei beiden Codern wurde die zeitliche Prädiktion verwendet. TNS wurde nicht eingesetzt, da die vorliegende Encoderimplementierung mit TNS noch einige deutliche Codierungsartefakte zeigte. Innerhalb der Codierung wurde eine Bandbegrenzung auf 15,5 kHz durchgeführt. Für die Audiosignale wurde eine Zahlendarstellung mit 32 Bit Festkomma verwendet, da bei reduzierter Wortlänge unter Umständen keine sinnvolle Einstellung der Quantisierer bei Audiosignalen nahe der Ruheschwelle mit dem Hörmodell möglich ist. Dieses Problem tritt auf, wenn die Amplitude des Audiosignals kleiner als das LSB der Zahlendarstellung ist. Die Vorechokontrolle des psychoakustischen Modells entfällt bei der Hörmodellsteuerung, da die Vorechos vom Hörmodell wie andere Signaländerungen detektiert und weitgehend „ausgeregelt“ werden.

Die Quantisierersteuerung berücksichtigt nicht explizit die Vor- bzw. Nachverdeckung, da nur die Signallaufzeit im Gehörmodell zur Bestimmung des Zeitpunkts für die Modifikation der Quantisierereinstellung verwendet wird. Durch zeitliche Glättung der SMR-Werte wird jedoch zum Teil berücksichtigt, daß die zeitliche Änderung des SMR aufgrund zeitabhängiger Verdeckungswirkungen relativ langsam erfolgt.

5.4 Bewertung der Klangqualität anhand ausgewählter Audiosignale

Für die Bewertung und den Vergleich der Klangqualität wurde zunächst eine größere Anzahl von Audiosignalen mit dem Referenzcoder bei einer konstanten Bitrate von 64 bzw. 56 kbit/s codiert. Die Audiosignale, die nach der Codierung die größte Qualitätseinbuße aufwiesen, wurden für die weiteren Untersuchungen ausgewählt. Es zeigte sich, daß insbesondere bestimmte Einzelinstrumentensignale und Sprachsignale eine Klasse kritischer Audiosignale bilden. Dieses Ergebnis steht im Einklang mit anderen subjektiven Testergebnissen [97, 102]. Als Einzelinstrumentensignale wurden ein Kastagnettensignal und ein auf einem Cembalo arpeggio gespielter Akkord ausgewählt. Die Sprachsignale umfassen zwei Proben eines deutschen Sprechers und zwei Proben unterschiedlicher englischer Sprecher. Zusätzlich wurde ein Sologesangsstück ausgewählt, das noch als kritisch bezüglich der Codierung einzustufen ist.

Die ausgewählten sieben Audiosignale wurden sowohl mit dem Referenzcoder als auch mit dem modifizierten Encoder mit variabler Datenrate codiert. Um eine Vergleichbarkeit der Ergebnisse herzustellen, sollten beide Encoder die gleiche Datenmenge zur Repräsentation aller sieben codierten Audiosignale verwenden. Dazu wurde zunächst die Datenmenge anhand einer Codierung mit dem modifizierten Encoder bestimmt. Zur Erzielung der gleichen Datenmenge mit dem Referenzcoder wurden die vom psychoakustischen Modell generierten SMR-Werte anschließend um 5,05 dB vergrößert. Dies entspricht der Vorgehensweise eines Standard-Bitzuteilungsalgorithmus, der zusätzlich verfügbare Bits zur Erzielung einer konstanten Pegelerhöhung des SMR nutzt. Mit den vergrößerten SMR-Werten erfolgte dementsprechend eine feinere Quantisierung, die zu einer deutlichen Klangqualitätsverbesserung im Vergleich zu den unveränderten Werten führt. Die unveränderten SMR-Werte des psychoakustischen Modells führen bei einer Codierung mit variabler Datenrate zu deutlich wahrnehmbaren Artefakten, so daß damit keine hohe Klangqualität erreicht wird. Für die veröffentlichten subjektiven Tests wurden bislang nur konstante Datenraten verwendet, wobei die durch den Bitzuteilungsalgorithmus vergrößerten SMR-Werte bei ausreichend hoher Datenrate zu einer verbesserten Qualität führen.

Die so codierten Audiosignale wurden nach der Decodierung in einem Hörtest durch sieben Testpersonen bewertet. Der Test wurde mit Kopfhörern (STAX SR- Λ -Pro) durchgeführt. Der Wiedergabepegel konnte dabei von der Testperson frei gewählt werden. Die verwendete Hörtestprozedur basiert auf der ITU-R Empfehlung BS.1116 [104] und wird als „triple stimulus / double blind / hidden reference“ bezeichnet. Für die Bewertung wird einer Testperson eine Folge von drei Audiosignalen dargeboten, die immer mit dem Originalsignal

beginnt und beliebig oft wiederholt werden kann. Die letzten beiden Audiosignale des Tripels werden durch das Originalsignal und das decodierte Signal in unbekannter, zufälliger Reihenfolge gebildet. Die Qualität der letzten beiden Audiosignale wird jeweils von der Testperson anhand der sog. „ITU-R 5-point impairment scale“ im Vergleich zum ersten, dem Originalsignal, bewertet. Diese in Bild 5.5 gezeigte Skala enthält Absolutwertungen. Es können Zwischenwertungen abgegeben werden, d. h. die Skala ist kontinuierlich zu handhaben. Die Dauer der Audiosignale wurde jeweils auf maximal 10 Sekunden begrenzt. Diese relativ kurze Dauer erleichtert in der Regel den detaillierten Vergleich der Signale eines Tripels, da erfahrungsgemäß subjektive Vergleiche über einen längeren Dauer hinweg mit zunehmender Unsicherheit behaftet sind.

Für die Darstellung der Ergebnisse wird die Differenz der Bewertungen für das decodierte Signal und für das Originalsignal (Referenzsignal) eines Tripels bestimmt und als sog. „difference grading“ angegeben. Durch Mittelung und statistische Auswertung der Standardabweichung der difference gradings mehrerer Testpersonen werden dann die Mittelwerte und 95%-Konfidenzintervalle bestimmt. Die Konfidenzintervalle umfassen den Bereich, in dem statistisch die difference gradings von 95% der Testpersonen liegen. Eine signifikante Veränderung der Klangqualität gilt als nachgewiesen, wenn die 95%-Konfidenzintervalle von Original- und decodiertem Audiosignal nicht überlappen.

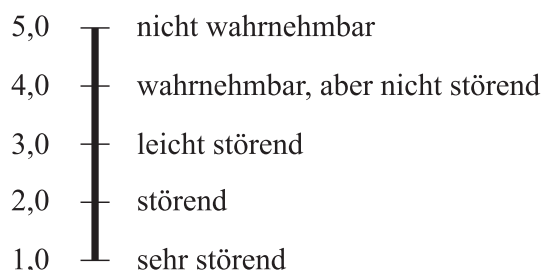


Bild 5.5 Bewertungsskala für Unterschiede zwischen dem Referenzsignal und einem Vergleichssignal. („ITU-R 5-grade impairment scale“)

In Bild 5.6 sind die Ergebnisse der subjektiven Bewertung für jedes Audiosignal individuell und die gemittelte Bewertung aller Audiosignale dargestellt. Die der Datenmenge entsprechende mittlere Datenrate ist ebenfalls für den Referenz- und modifizierten Coder als Balkendiagramm zum Vergleich angegeben.

Die gemittelte Bewertung über alle Audiosignale zeigt, daß der modifizierte Coder die gleiche Qualität wie der Referenzcoder erreicht. Bei den individuellen Signalen ergeben sich jedoch teilweise subjektive Qualitätsunterschiede beider Coder. Während bei zwei Sprachsignalen und dem Gesang eine Qualitätsverbesserung erzielt wird, führt der modifizierte Coder bei der Cembalosequenz zu einer verminderten Qualität. Bei alleiniger Auswertung der überlappenden Konfidenzintervalle, deren großes Ausmaß auf die relativ kleine Anzahl von Testpersonen zurückzuführen ist, müßte die Qualitätsänderung als nicht signifikant eingestuft werden. Eine Analyse der individuellen Wertungen zeigt jedoch, daß bei den genannten Audiosignalen mit den größten Qualitätsunterschieden die einzelnen Testpersonen eine konsistente Präferenz für das entsprechende Audiosignal haben. So wurde beispielsweise das Audiosignal „deutscher Sprecher 2“ für den modifizierten Coder von keiner Testperson schlechter bewertet als für den Referenzcoder.

Die verminderte Qualität der Cembalosequenz vom modifizierten Coder ist auf Artefakte bei dem tiefsten angeschlagenen Ton in dem Audiosignal zurückzuführen. Der entsprechende Signalabschnitt wird im folgenden Kapitel 5.5 noch genauer untersucht. Die verminderte Qualität spiegelt sich in einer reduzierten mittleren Datenrate dieser Sequenz für den modifizierten Coder wider.

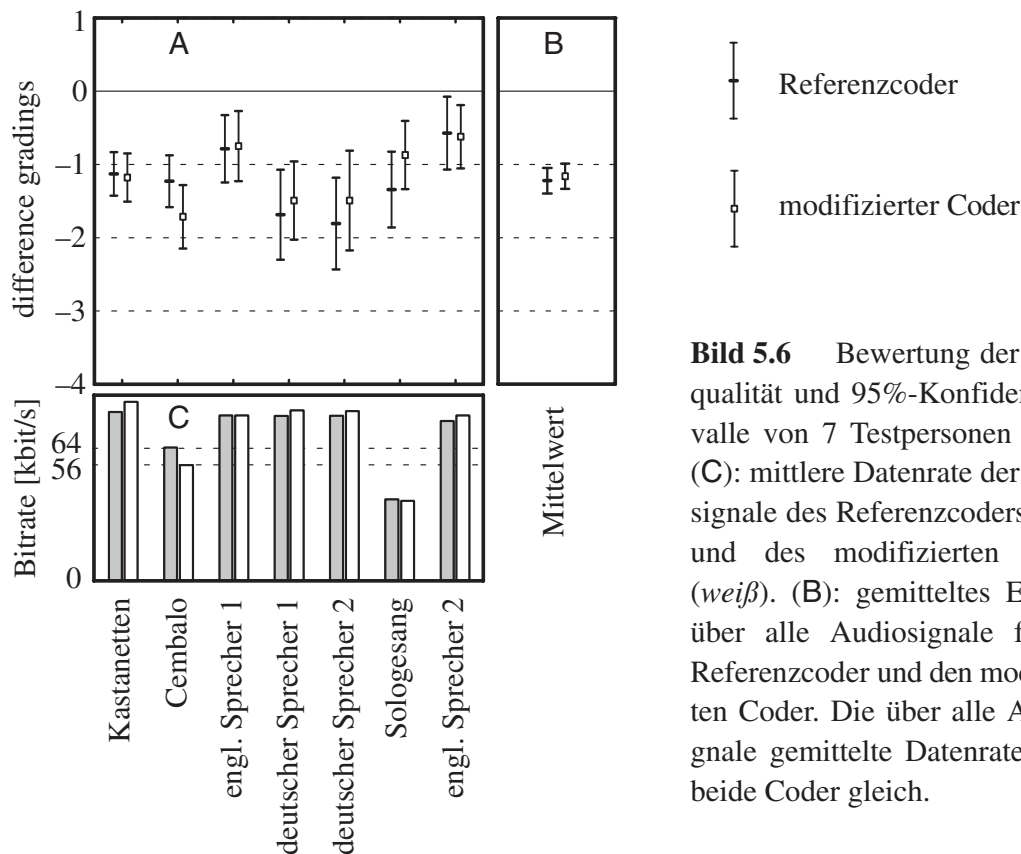


Bild 5.6 Bewertung der Klangqualität und 95%-Konfidenzintervalle von 7 Testpersonen (A, B). (C): mittlere Datenrate der Audiosignale des Referenzcoders (*grau*) und des modifizierten Coders (*weiß*). (B): gemittelttes Ergebnis über alle Audiosignale für den Referenzcoder und den modifizierten Coder. Die über alle Audiosignale gemittelte Datenrate ist für beide Coder gleich.

Die verbesserte Qualität der beiden Audiosignale des deutschen Sprechers für den modifizierten Coder lassen darauf schließen, daß insbesondere die schnellen Änderungen der Signalstatistik dieser Sprachsignale vom Gehörmodell im Vergleich zum psychoakustischen Modell besser aufgelöst werden. Im Gegensatz zur Cembalosequenz ist in diesem Fall die gehörangepaßte Zeitauflösung und Frequenzselektivität des Gehörmodells von besonderer Bedeutung. Sprachsignale enthalten oft eine modulierte Grundfrequenz bei stimmhaften Lauten, die mit einem frequenzmodulierten Signal verglichen werden kann. Diese oft als tonal wahrgenommenen stimmhaften Laute werden vom psychoakustischen Modell aufgrund der Prädiktionstechnik zur Vorhersage der Tonalität jedoch eher als rauschartig bewertet und führen somit zu einer zu hoch angenommenen Wahrnehmungsschwelle. Diese Annahme wird im folgenden Kapitel 5.5 durch eine Untersuchung spezieller synthetischer Audiosignale überprüft.

5.5 Vergleich generierter Signal-Maskierungs-Abstände vom optimierten psychoakustischen Modell und vom psychophysiologicalen Gehörmodell

Die relevanten Eigenschaften des optimierten psychoakustischen Modells und des psychophysiologicalen Gehörmodells für die Audiocodierung werden im folgenden analysiert und verglichen. Bei einer Codierung mit hoher Klangqualität sind ggf. wahrnehmbare Artefakte in der Regel in einer Spektrogrammdarstellung der Audiosignale nicht notwendigerweise lokalisierbar. Die Unterschiede in der Codierung sind wesentlich besser mit Hilfe des SMR nachzuweisen, der unmittelbar zur Quantisierersteuerung verwendet wird.

In Bild 5.7 sind verschiedene Darstellungen für einen 800 ms langen Ausschnitt eines der verwendeten Sprachsignale enthalten. Die linke Spalte enthält die Spektrogramme des Originalsignals und des decodierten Signals vom Referenzcoder bzw. modifizierten Coder. Die rechte Spalte stellt die Signalleistung des Originalsignals sowie den vom Referenzcoder und modifizierten Coder verwendeten SMR in den Skalenfaktorbändern für die einzelnen Blöcke dar. Durch einen Vergleich der unterschiedlichen SMR-Werte beider Coder in Bild 5.7e und Bild 5.7f wird deutlich, daß der modifizierte Coder zu einer wesentlich ausgeprägteren Abhängigkeit des SMR vom Audiosignal führt, während der Referenzcoder nur eine geringfügige Zeitabhängigkeit des SMRs aufweist.

Die durch eine vertikale Linie in Bild 5.7 gekennzeichneten Signalleistungen und SMR-Werte sind in Bild 5.8 herausgezogen und als Balkendiagramm dargestellt. Die untere Begrenzung des in jedem Balken schwarz dargestellten SMR ist die vom jeweiligen Modell angenommene Wahrnehmungsschwelle. Es wird deutlich, daß die im modifizierten Coder verwendete Wahrnehmungsschwelle über den Skalenfaktorbändern eine größere Variabilität aufweist als bei dem Referenzcoder. Die Breite der einzelnen Balken wurde proportional zur Bandbreite der Skalenfaktorbänder gewählt. Daher ist der gesamte schwarze Flächeninhalt der SMR-Darstellung näherungsweise proportional zur Anzahl der für die Codierung dieses Blocks benötigten Bits.

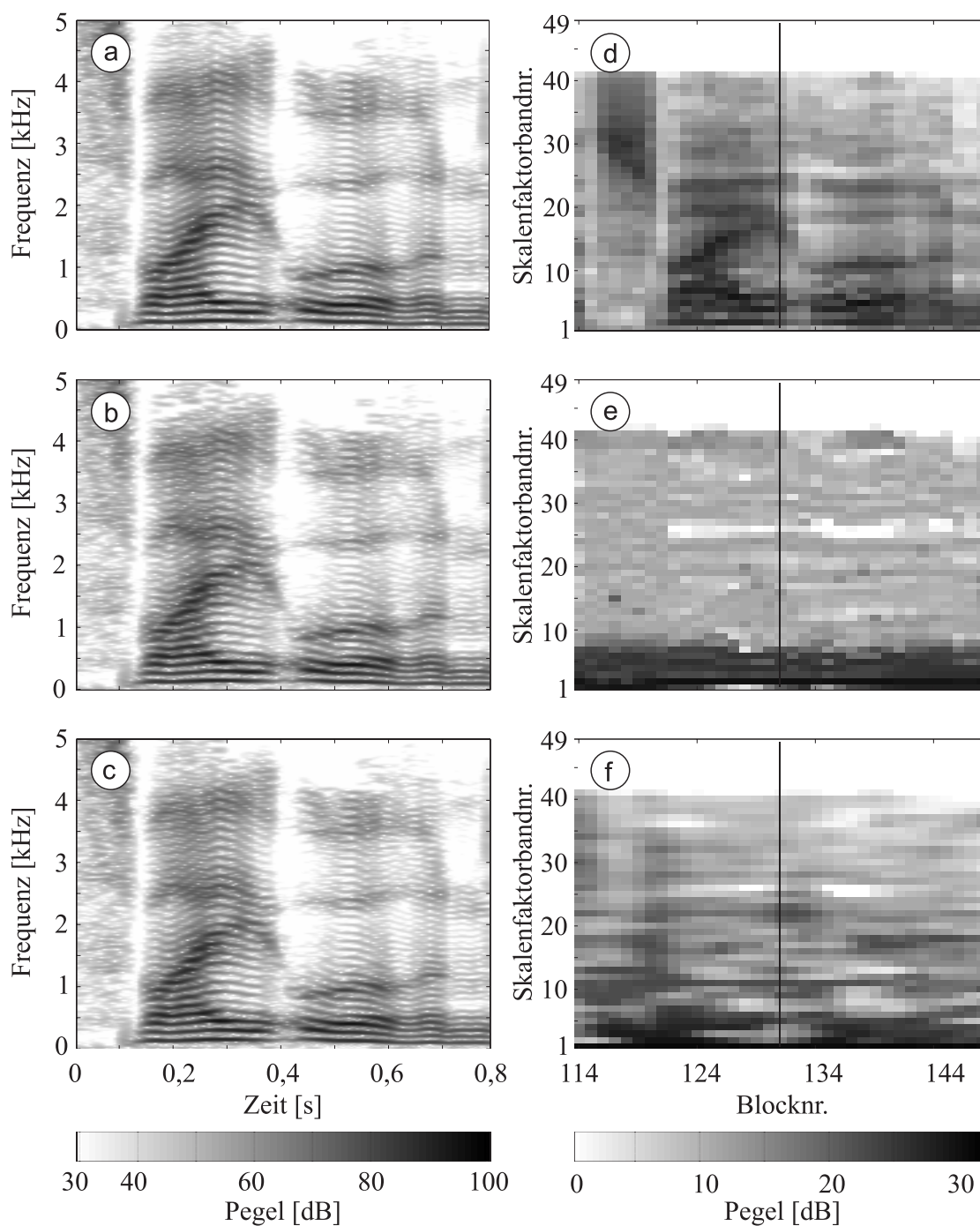


Bild 5.7 Ausschnitt des Audiosignals „deutscher Sprecher 1“ mit den Worten „zwei Ohren“. *Links:* Spektrogramm des Originalsignals (a) und des decodierten Signals vom Referenzcoder (b) bzw. vom modifizierten Coder (c). *Rechts:* Skalenfaktorbandleistungen des Originalsignals (d). Im Referenzcoder (e) bzw. im modifizierten Coder (f) verwendeter SMR. Die Zuordnung der Grauwerte zum Pegel gilt rechts nur für die Darstellungen (e) und (f).

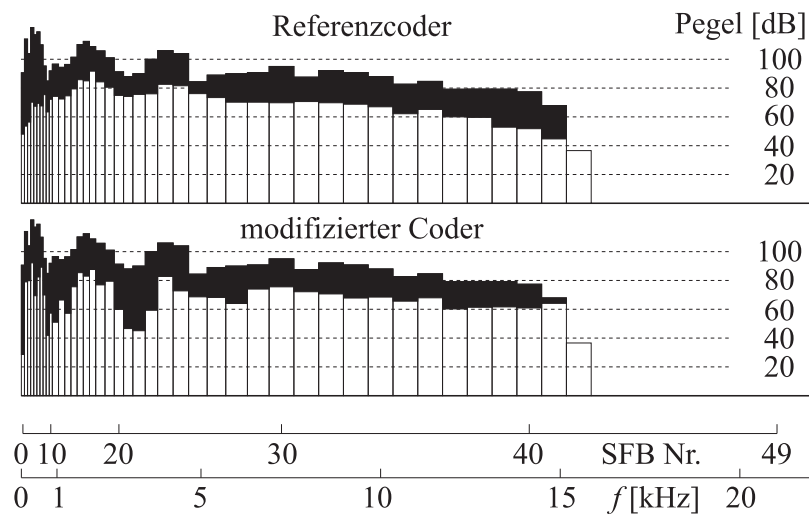


Bild 5.8 Signalleistungen und Signal-Maskierungs-Abstand für die gekennzeichneten Blöcke in Bild 5.7. Jeder Balken repräsentiert die Leistung in einem Skalenfaktorband. Der grau dargestellte Teil repräsentiert den SMR. Die Breite ist proportional zur Bandbreite in Hertz.

Die verwendete Bitanzahl zur Codierung der einzelnen Blöcke des Audiosignals „deutscher Sprecher 1“ ist in Bild 5.9A angegeben. Der modifizierte Coder zeigt dabei eine größere Schwankung der Bitanzahl, wobei insbesondere bei einer Zunahme der Audiosignalleistung oder bei der Codierung mit kurzen Fensterfunktionen kurzzeitig eine hohe Bitanzahl erreicht wird. Bild 5.9B zeigt einen vergrößerten Ausschnitt für den Audiosignalabschnitt, der in den Spektrogrammen des Bildes 5.7 gezeigt ist. Bei diesem Audiosignal ist die mittlere Bitrate des modifizierten Encoders im Vergleich zum Referenzcoder um 4% höher.

Eine Darstellung des generierten SMR für einen Ausschnitt der Cembalosequenz ist in Bild 5.10 zusammen mit den entsprechenden Spektrogrammen und der Signalleistung dargestellt. Der Ausschnitt beinhaltet den tiefsten gespielten Ton in dieser Sequenz. Der Anschlag des Tons führt in Block 4 zur Umschaltung auf kurze Fenster. In diesem Block, wie auch bei dem Anschlag des nächsten Tons in Block 37, verwendet der Coder insgesamt nur 14 breitere Skalenfaktorbänder aber eine Unterteilung des Blocks in 8 Zeitabschnitte. Die reduzierte Anzahl von Skalenfaktorbändern bei kurzen Fenstern gegenüber den 49 Bändern bei langen Fenstern ist der Übersichtlichkeit halber nicht als Achsenbezeichnung in Bild 5.10 mit angegeben.

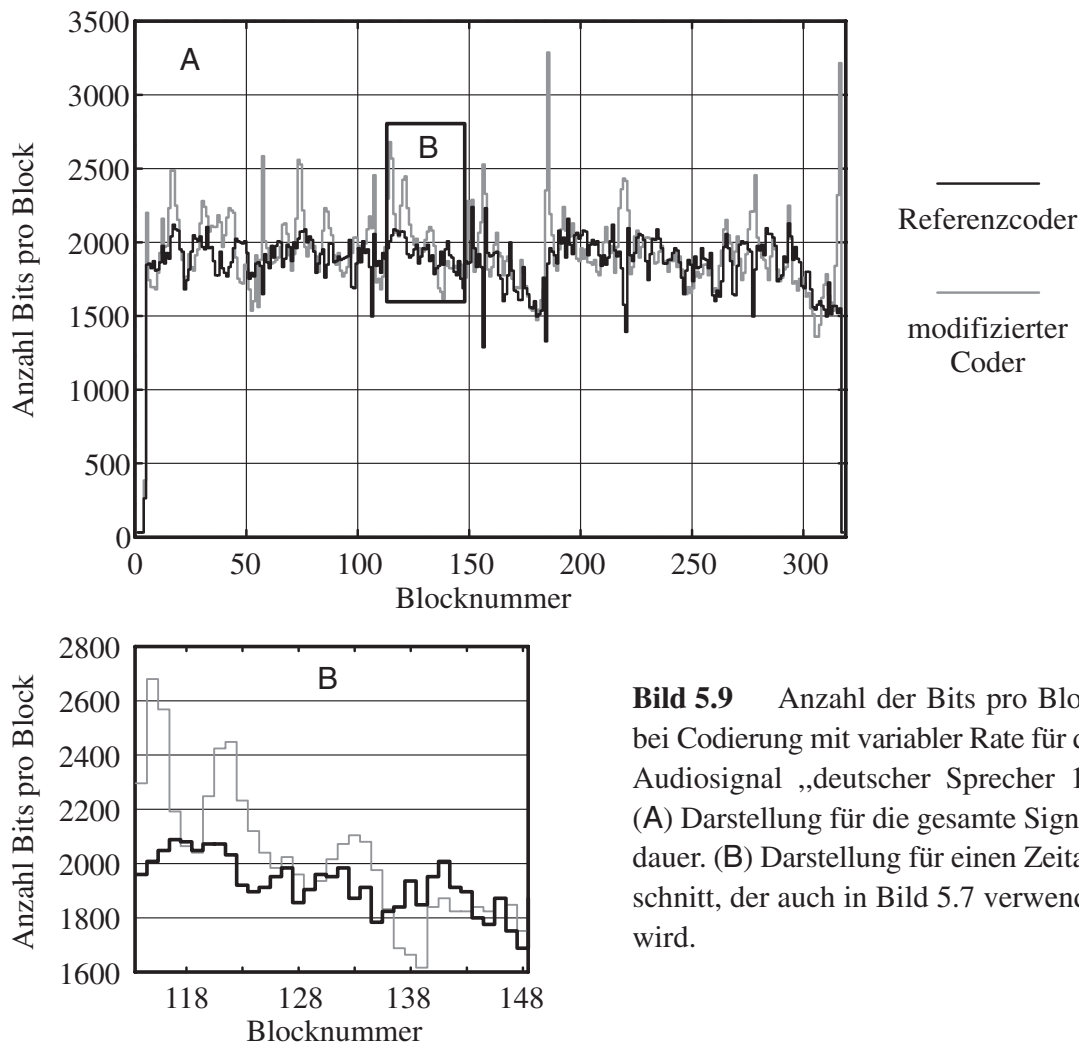


Bild 5.9 Anzahl der Bits pro Block bei Codierung mit variabler Rate für das Audiosignal „deutscher Sprecher 1“. (A) Darstellung für die gesamte Signaldauer. (B) Darstellung für einen Zeitabschnitt, der auch in Bild 5.7 verwendet wird.

Das Spektrogramm in Bild 5.10a zeigt ein ausgeprägtes Obertonspektrum, das noch über den dargestellten Bereich bis 5 kHz hinausgeht. Die Skalenfaktorbandleistungen in Bild 5.10d zeigen dementsprechend nur eine geringe Abnahme zu höheren Frequenzen. Der SMR des Referenzcoders in Bild 5.10e weist im Bereich des ausklingenden Tons bis zum Skalenfaktorband 24 hohe Werte auf. Daraus ist erkennbar, daß in diesem Bereich das Audiosignal als tonal angenommen und entsprechend fein quantisiert wird. Der SMR des modifizierten Coders weist im wesentlichen im Bereich der Skalenfaktorbänder 1 bis 7 hohe Werte auf. Bei Frequenzen oberhalb des Skalenfaktorbandes 7 ist davon auszugehen, daß mehrere Obertöne in die Filterbandbreiten des Gehörmodells fallen. Die Überlagerung mehrerer Obertöne führt dann zu Schwebungen, die aufgrund der damit verbundenen Einhüllendenfluktuation zu einer erhöhten Wahrnehmungsschwelle bzw. einem verminderten SMR führen.

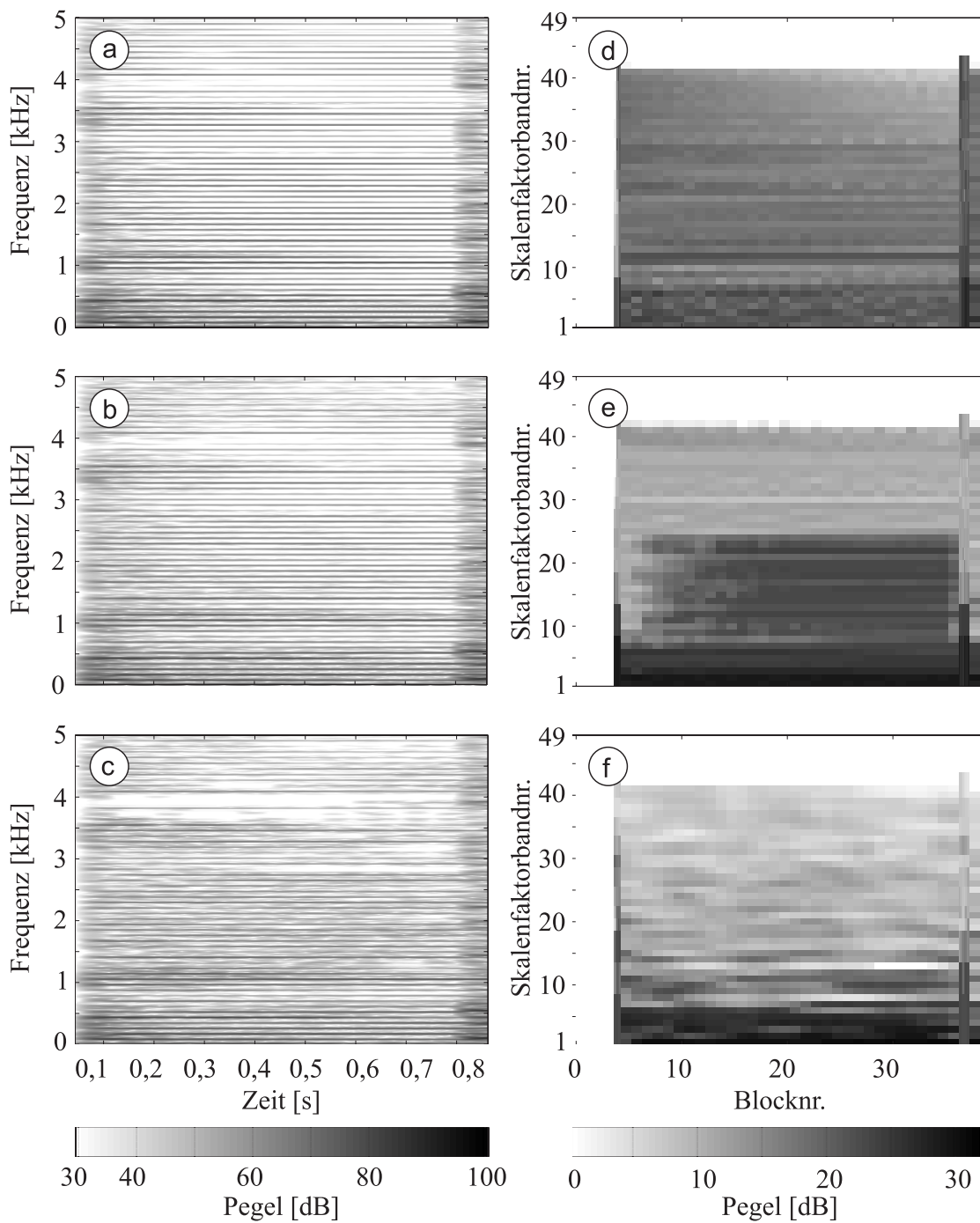


Bild 5.10 Ausschnitt des Signals „Cembalo“. Links: Spektrogramm des Originalsignals (a) und des decodierten Signals vom Referenzcoder (b) bzw. vom modifizierten Coder (c). Rechts: Skalenfaktorbandleistungen des Originalsignals (d). Im Referenzcoder (e) bzw. im modifizierten Coder (f) verwendeter SMR. Die Zuordnung der Grauwerte zum Pegel gilt rechts nur für die Darstellungen (e) und (f).

Die als schlechter bewertete Klangqualität der Cembalosequenz, die vor allem auf Artefakte bei dem tiefsten Ton zurückzuführen ist, kann als ein Hinweis auf eine unzureichende Berücksichtigung der Modulationswahrnehmung im Gehörmodell gewertet werden. Die Schwebungseffekte beim Ausklingen des Tons können auch als Amplitudenmodulation aufgefaßt werden [32], wobei die Einhüllende weitgehend periodisch ist. Geht man davon aus, daß eine Störung dieser Periodizität vom Gehör leicht wahrgenommen werden kann, so müßte die Wahrnehmungsschwelle in diesem Fall entsprechend reduziert werden. Ein solcher Mechanismus könnte im Gehörmodell durch eine „Modulationsfilterbank“ innerhalb des Teilmodells der neuronalen Verarbeitung in jeder Sektion verwirklicht werden. Die Schwellendetektoren wären dann an den Ausgängen der Filterbank anzubringen. Eine vergleichbare Realisierung von Modulationsfilterbänken wurde in [39, 40, 87] bereits verwirklicht. Wegen des hohen Aufwands an Rechenleistung wurden diese Filterbänke jedoch bislang im Gehörmodell nicht erprobt.

Bei Sprachsignalen treten häufig Frequenzmodulationen der Grundfrequenz bei stimmhaften Lauten auf, die zu einer fehlerhaften Tonalitätsschätzung im psychoakustischen Modell des Referenzcoders führen können. Zur Untersuchung dieses vermuteten Zusammenhangs wurden zwei synthetische frequenzmodulierte Audiosignale verwendet, die genauso wie die obigen Audiosignale codiert und von Testpersonen bewertet wurden (vgl. Bild 5.11). Das erste synthetische Audiosignal ist ein mit 8 Hz sinusförmig frequenzmodulierter Sinuston zwischen 600 und 1900 Hz, dem ein Rosa Rauschen überlagert ist. Bild 5.12 zeigt die entsprechenden Spektrogramme, sowie die Leistungen und SMR-Werte in den Skalenfaktorbändern. Die Bewertung der Klangqualität in Bild 5.11 zeigt, daß der modifizierte Coder bei vergleichbarer mittlerer Datenrate zu einer höheren Klangqualität führt als der Referenzcoder. Diese unterschiedliche Bewertung wird insbesondere auf die Verzerrungen im Bereich der Flanken des FM-Signals im Spektrogramm zurückgeführt, die in Bild 5.12b deutlich erkennbar sind. Ein Vergleich der SMR-Werte der Bilder 5.12e und f zeigt, daß vom Gehörmodell bis zum Skalenfaktorband 12 hohe SMR-Werte eingestellt werden, während das psychoakustische Modell in diesem Bereich eine teilweise Verdeckung von Spektralkomponenten annimmt. Das FM-Signal wird vom psychoakustischen Modell im wesentlichen als nicht tonal eingestuft, obwohl es als tonal wahrgenommen wird.

Als zweites synthetisches Signal wurde ein frequenzmodulierter Tonkomplex verwendet, der ähnlich wie der Vokal „ä“ wahrgenommen wird. Das Spektrogramm ist in Bild 5.13a dargestellt. Das Audiosignal wurde mit einer Impulsfolge variabler Rate als „Erregungssignal“ eines Sprachtraktfilters erzeugt. Die drei konstanten Resonanzfrequenzen des Filters sind als Maxima der spektralen Einhüllenden im Spektrogramm erkennbar. Die Impulsfolge erzeugt ein harmonisches Linienspektrum, wobei die Grundfrequenz der Impulsrate entspricht. Im Verlauf des Signals nimmt die Änderungsgeschwindigkeit der Grundfrequenz zu.

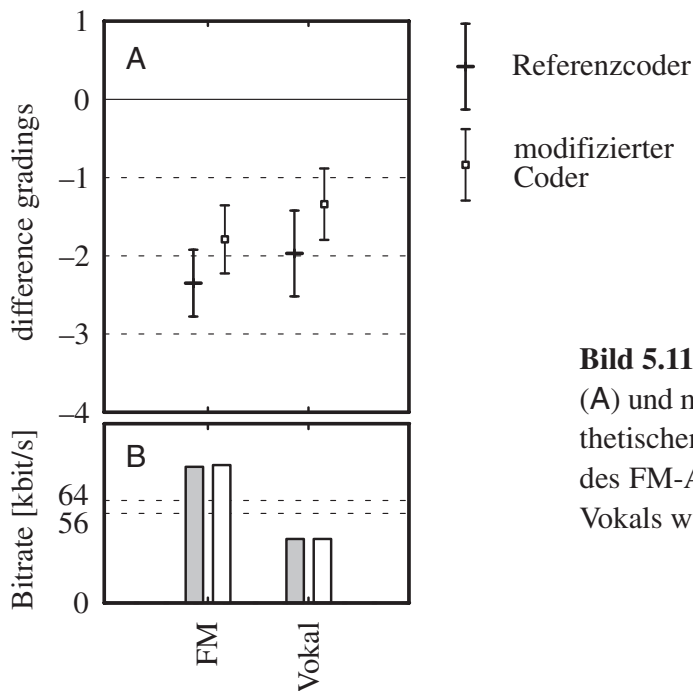


Bild 5.11 Bewertung der Klangqualität (A) und mittlere Bitraten (B) der beiden synthetischen Audiosignale. Die Klangqualität des FM-Audiosignal und des synthetischen Vokals wurde von 5 Testpersonen beurteilt.

Die Klangqualität des decodierten Audiosignals wird für den modifizierten Coder bei gleicher mittlerer Datenrate höher bewertet. Diese unterschiedliche Bewertung steht im Gegensatz zu den geringen Unterschieden der entsprechenden Spektrogramme. Die SMR-Werte des Referenzcoders bestätigen jedoch die obige Annahme, daß ein harmonisches Audiosignal während einer relativ langsamen Grundfrequenzänderung nicht als tonal erkannt wird. Lediglich bei kurzzeitig konstanter Frequenz, wie in Block 25, treten hohe SMR-Werte auf. Durch das Gehörmodell wird der SMR demgegenüber besser an die Verdeckungseigenschaften angepaßt. Die hohen SMR-Werte des modifizierten Coders am Anfang und Ende des Audiosignals sind auf „Randeffekte“ zurückzuführen, die sich bei der iterativen Quantisierungseinstellung durch das Gehörmodell auswirken können.

Die Ergebnisse des Vergleichs vom Referenzcoder und dem modifizierten Coder lassen sich wie folgt zusammenfassen. Bei einem harmonischen Audiosignal mit ausgeprägtem Obertonspektrum führt das Gehörmodell noch zu wahrnehmbaren Artefakten bei der Codierung. Diese Artefakte sind wahrscheinlich auf eine unzureichende Berücksichtigung der Wahrnehmungseigenschaften bei einer periodischen Amplitudenmodulation zurückzuführen. Bei einigen Sprachsignalen wird eine Qualitätsverbesserung erreicht. Dies wird auf die adäquate Nachbildung der Wahrnehmungsschwellen sowohl bei frequenzmodulierten Signalen, als auch bei einem zeitlich schnell veränderlichen Signalcharakter durch das Gehörmodell zurückgeführt (vgl. [15]).

Die generierten SMR-Werte zeigen bei der Gehörmodellsteuerung eine größere Schwankung und eine höhere Korrelation zum Audiosignal als bei dem psychoakustischen Modell. Dies führt zu einer stärker variierenden Datenrate, die insbesondere bei der Codierung mit kurzen Fensterfunktionen kurzzeitige Spitzen aufweist.

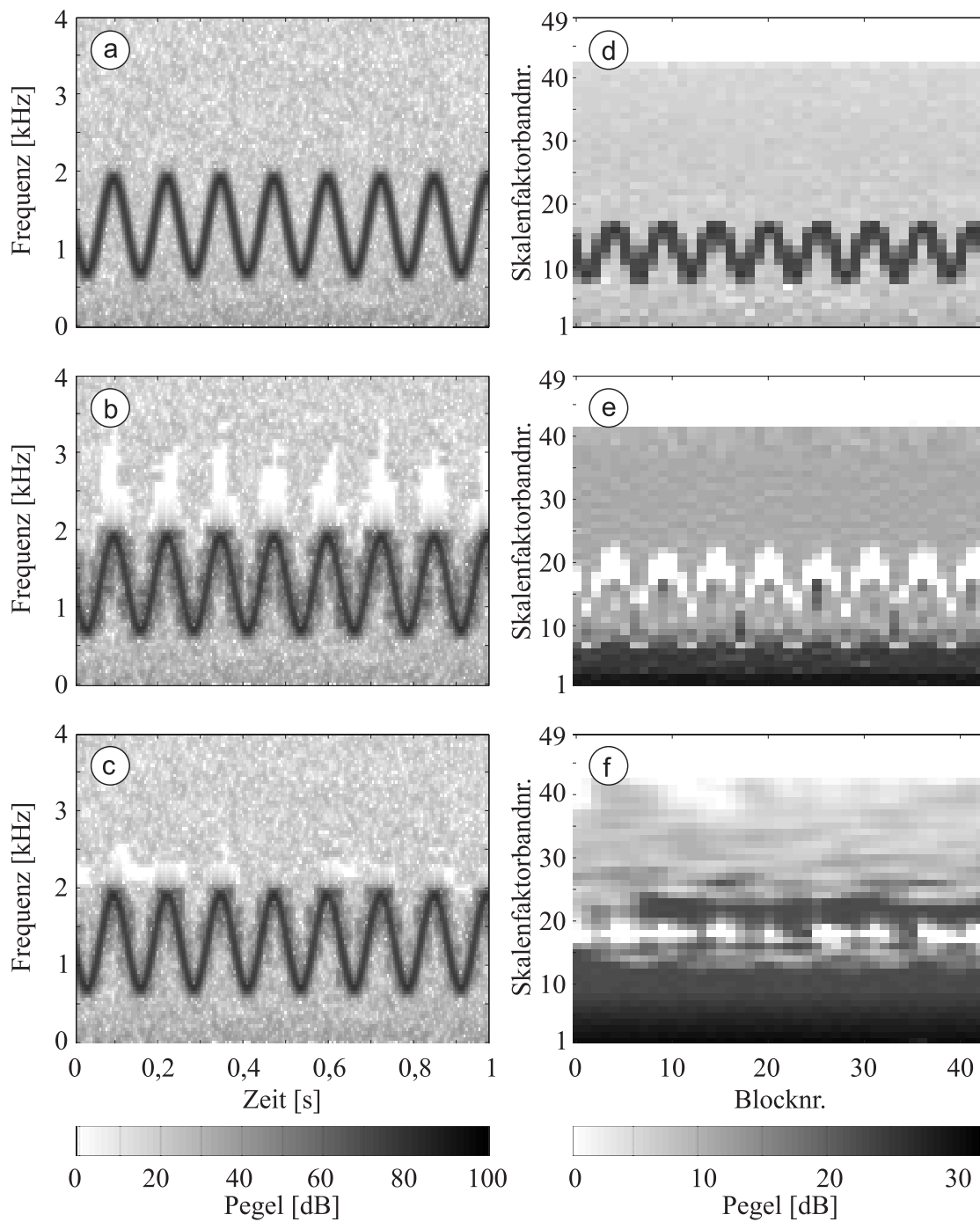


Bild 5.12 Sinusförmig frequenzmodulierter Sinuston überlagert mit Rosa Rauschen. *Links:* Spektrogramm des Originalsignals (a) und des decodierten Signals vom Referenzcoder (b) bzw. vom modifizierten Coder (c). *Rechts:* Skalenfaktorbandleistungen des Originalsignals (d). Im Referenzcoder (e) bzw. im modifizierten Coder (f) verwendeter SMR. Die Zuordnung der Grauwerte zum Pegel gilt rechts nur für die Darstellungen (e) und (f).

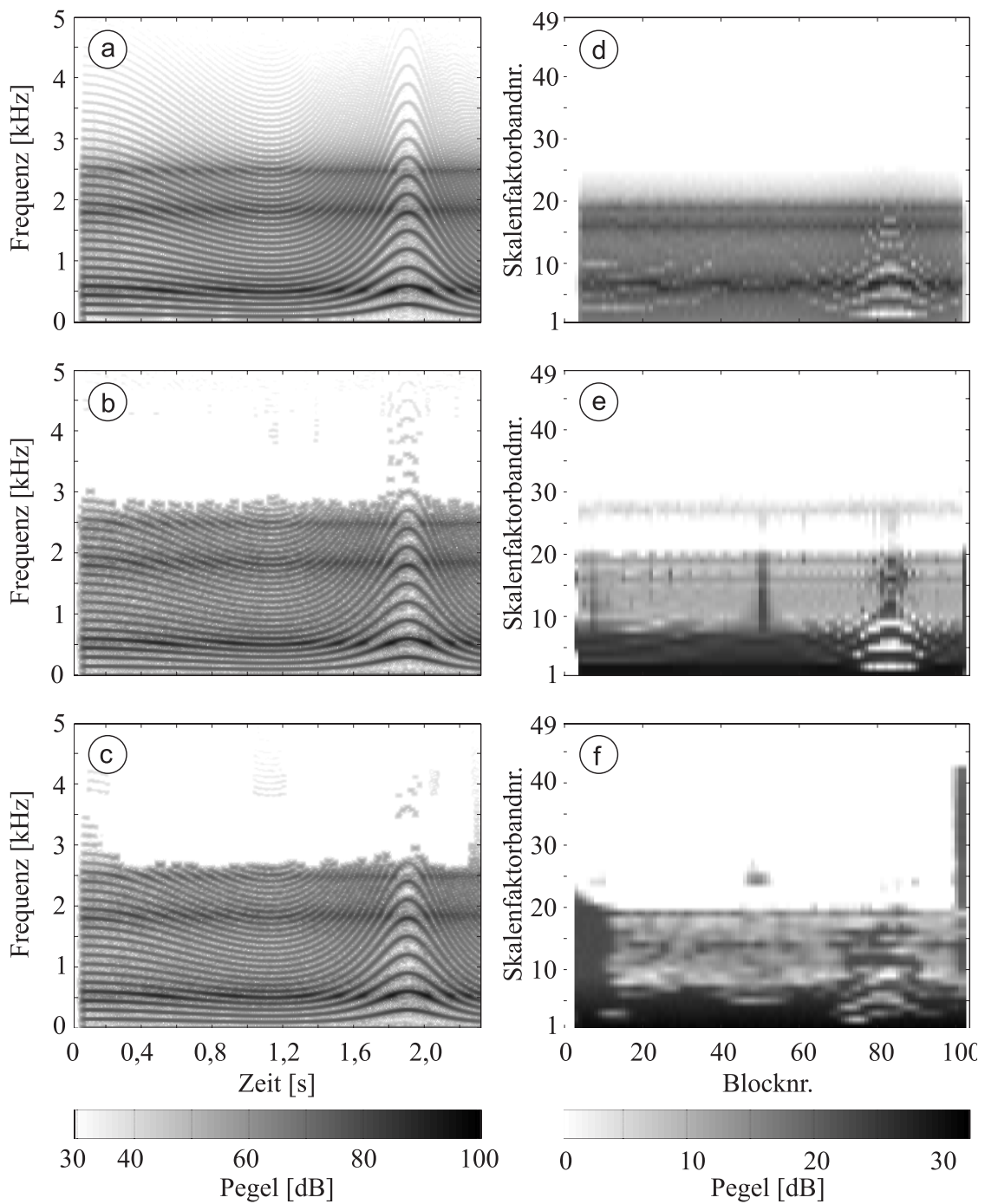


Bild 5.13 Synthetisierter Vokal „ä“. *Links:* Spektrogramm des Originalsignals (a) und des decodierten Signals vom Referenzcoder (b) bzw. vom modifizierten Coder (c). *Rechts:* Skalenfaktorbandleistungen des Originalsignals (d). Im Referenzcoder (e) bzw. im modifizierten Coder (f) verwendeter SMR. Die Zuordnung der Grauwerte zum Pegel gilt rechts nur für die Darstellungen (e) und (f).

6 Zusammenfassung

Die Zielsetzung der Arbeit bestand in einer möglichst genauen numerischen Nachbildung der von beliebig komplexen Audiosignalen hervorgerufenen Wahrnehmungsschwellen. Das dafür entwickelte Gehörmodell besteht aus der Kombination eines physiologischen Teils mit einem psychoakustisch motivierten Teil. Es waren im wesentlichen drei Probleme zu lösen:

- Zur Berücksichtigung nichtlinearer Eigenschaften des Gehörs sollte ein Gehörmodell entwickelt werden, das soweit wie möglich physiologisch begründet ist.
- Zur Verifikation der erarbeiteten Lösung sollte das Gehörmodell im zweiten Schritt anhand von bekannten physiologischen und psychoakustischen Meßdaten überprüft werden.
- Die Vorzüge des psychophysiologischen Gehörmodells im Vergleich zu einem rein psychoakustischen Modell sollten anhand eines Anwendungsbeispiels in der Audiocodierung demonstriert werden.

Als Grundlage für die Entwicklung des Gehörmodells wurden zunächst die für die Verdeckungseffekte relevanten physiologischen Schritte der Schallverarbeitung des Gehörs durch eine Auswertung der einschlägigen Literatur bestimmt. Dabei ergab sich, daß die wesentlichen physiologischen Ursachen der Simultanverdeckung durch die Erregungs- und Suppressionswirkung eines Maskierers begründet sind. Die Ursache der zeitlichen Nachverdeckung wird im wesentlichen auf Adaptionsvorgänge bei der neuronalen Verarbeitung zurückgeführt. Für die Bestimmung der Wahrnehmungsschwellen wurden die grundlegenden Konzepte der Signaldetektionstheorie herangezogen. Diese Theorie stellt eine quantitative Beschreibung der Detektierbarkeit bzw. Wahrnehmbarkeit eines Signals zur Verfügung.

Der Aufbau des Gehörmodells orientiert sich an der Anatomie des Gehörs mit dem Ziel einer möglichst weitgehenden physiologischen Nachbildung der relevanten Verarbeitungsschritte des Schallsignals unter Berücksichtigung der Nichtlinearität. Die Komponenten des Gehörmodells sind dabei relativ einfach gehalten, um die Anzahl der Modellparameter zu beschränken. Das Gehörmodell gliedert sich in die Teilmodelle für das Außen- und Mittelohr, die Hydromechanik der Cochlea, die äußeren Haarzellen, die inneren Haarzellen und die neuronale Verarbeitung. Die Modellierung des peripheren Gehörs, das sich vom Außenohr bis zu den inneren Haarzellen erstreckt, erfolgt auf Basis der Physiologie, während das Teilmodell der neuronalen Verarbeitung aus psychoakustischen Meßergebnissen abgeleitet ist.

Die Nachbildung des Außen- und Mittelohres approximiert die zwischen Außenohr und ovalem Fenster am Eingang der Cochlea gemessene Betragsübertragungsfunktion eines Schalldrucksignals. Die Nachbildung der Cochlea basiert auf dem Modell von Zwicker und Peisl, da es die wesentlichen geforderten Eigenschaften erfüllt. Insbesondere erweist sich die numerische Simulation des Teilmodells der Cochlea im Zeitbereich mit Wellendigitalfiltern als vorteilhaft. Diese Methode erlaubt sowohl die Simulation von Filtern hoher Güte unter Vermeidung numerischer Instabilität als auch die Implementierung von Elementen mit nicht-

linearer Kennlinie. Das Teilmodell der Cochlea bildet die sog. makromechanischen Eigenschaften des flüssigkeitsgefüllten Schneckenkanals mit der schwingungsfähigen cochleären Trennwand nach, die den Kanal in zwei akustisch wirksame Kammern unterteilt. In Ergänzung dazu wird die Verstärkungswirkung der äußeren Haarzellen, die auf die Schwingung der cochleären Trennwand einwirkt, berücksichtigt. Die Sättigung der Verstärkungswirkung stellt dabei ein stark nichtlineares Element dar. Für die Zielsetzung dieser Arbeit war eine Erweiterung des Teilmodells der Cochlea um eine zweite Verstärkerstufe erforderlich. Damit werden neuere physiologische Meßergebnisse berücksichtigt, die eine größere Dynamikkompression zeigen, als sie mit dem ursprünglichen Modell erreicht wurde. Die Mikromechanik der cochleären Trennwand wird nicht im Detail nachgebildet, da bislang die physiologischen Meßergebnisse für eine präzise Modellierung nicht ausreichen. Das Teilmodell der Cochlea wird aus 251 in Reihe geschalteten Sektionen gebildet. In jeder Sektion bildet ein einfaches Teilmodell der inneren Haarzelle die Transduktion des Schallsignals in ein elektrisches Signal nach. Am Ausgang dieses Teilmodells wird somit die neuronale Aktivität der Fasern des Hörnervs nachgebildet, die mit dem korrespondierenden Ort in der Cochlea verbunden sind.

Das nachfolgende Teilmodell der neuronalen Verarbeitung wird für alle Sektionen parallel realisiert. Es beinhaltet die psychoakustische Nachbildung zeitlicher Verdeckungseffekte. Innerhalb des Teilmodells der neuronalen Verarbeitung wird die wahrgenommene spezifische Lautheit nachgebildet. Änderungen der spezifischen Lautheit werden nach einer Kurzzeitintegration durch einen Schwellendetektor ausgewertet. Die Änderung eines Eingangssignals, die beispielsweise durch ein überlagertes Testsignal hervorgerufen wird, befindet sich demnach an der Wahrnehmungsschwelle, wenn sie in einer Sektion gerade den Schwellwert des Schwellendetektors erreicht. Dabei wird berücksichtigt, daß die Wahrnehmbarkeit der Signaländerung auch vom zeitlichen Verlauf der Einhüllenden des Maskierersignals abhängt. Da sich die Wahrnehmbarkeit einer Signaländerung mit zunehmender Einhüllendenfluktuation reduziert, wird der modellinterne Schwellwert entsprechend adaptiert.

Mit dem modellinternen Schwellwert ist es jedoch aufgrund der Nichtlinearität des Gehörmodells nicht möglich, den Schalldruckpegel eines dem Maskierer überlagerten Testsignals an der Wahrnehmungsschwelle unmittelbar mit Hilfe eines invertierten Gehörmodells zu berechnen. Aus diesem Grund wird die Wahrnehmungsschwelle iterativ bestimmt. In aufeinander folgenden Iterationsschritten wird dabei die Größe der Eingangssignaländerung sukzessive verändert, bis der oben genannte interne Schwellwert erreicht ist. Diese Vorgehensweise ist mit psychoakustischen Messungen vergleichbar, in denen der Testsignalpegel von der Versuchsperson variiert wird, bis der Pegel an der Wahrnehmungsschwelle gefunden ist. Der bislang im Gehörmodell verwendete Schwellendetektor erfordert, daß in den einzelnen Iterationsschritten ein identisches Maskierersignal verwendet wird. Zwischen zwei Iterationen darf daher keine Phasenverschiebung des Maskierers auftreten.

Das Gehörmodell wird weitgehend in Form eines elektrischen Netzwerkes repräsentiert, um eine einheitliche mathematische Beschreibungsweise zu ermöglichen. Die entsprechenden elektrischen Größen der Netzwerke des Außen- und Mittelohres sowie der Hydro-

mechanik und äußeren Haarzellen werden mit Wellendigitalfiltern numerisch bestimmt. Die Teilmodelle der inneren Haarzellen und der neuronalen Verarbeitung werden im wesentlichen durch einfache rekursive Digitalfilter implementiert, um den erforderlichen Rechenaufwand klein zu halten.

Der Parameterabgleich des Gehörmodells wurde anhand von bekannten physiologischen und psychoakustischen Daten durchgeführt, die wesentliche Eigenschaften der Schallverarbeitung und Verdeckungswirkungen widerspiegeln. Die herangezogenen physiologischen Ergebnisse, die aus Tierexperimenten bekannt sind, umfassen:

- Frequenzselektivität der Basilmembranschwingung an einem spezifischen Ort,
- Dynamikkompression der Basilmembranschwingung am Ort der Resonanz.

Die zum Abgleich verwendeten psychoakustischen Meßergebnisse beinhalten:

- Simultanverdeckung durch schmalbandige Maskierer,
- zeitabhängige Verdeckung einfacher Maskierer-Testsignal-Kombinationen.

Zur Verifikation des Gehörmodells wurde seine Leistungsfähigkeit anhand einer größeren Auswahl von bekannten Meßergebnissen aus der Literatur bewertet. Dabei wurde gezeigt, daß das Gehörmodell eine Reihe unterschiedlicher Wahrnehmungsphänomene nachbildet. Dazu gehören:

- Eigenschaften der wahrgenommenen Lautheit:
 - Dynamikkompression bei der Lautheitswahrnehmung,
 - unterschiedliche Lautheit von Sinuston und Rauschen bei gleichem Schalldruckpegel.
- Eigenschaften der Simultanverdeckung:
 - pegelabhängige Flankensteilheit der Mithörschwelle („growth of masking“),
 - Abhängigkeit von Verzerrungsprodukten nichtlinearer Verarbeitung („off-frequency listening“),
 - nichtlineare Additivität der Verdeckung („nonlinear additivity“),
 - unterschiedliche Verdeckungswirkung tonaler und rauschartiger Signale („asymmetry of masking“),
 - Abhängigkeit von der Signaleinhüllenden.
- Eigenschaften der zeitabhängigen Verdeckung:
 - Vorverdeckung,
 - Nachverdeckung,
 - Abhängigkeit von der Testsignaldauer,
 - Abhängigkeit von der Maskiererdauer.
- Weitere Wahrnehmungsschwellen:
 - Ruhehörschwellen,
 - unterschiedliche Detektionsschwellen für Frequenz- bzw. Amplitudenmodulation.

Zusätzlich wurde das Gehörmodell auch anhand physiologischer Meßergebnisse überprüft. Dazu gehört:

- Nichtlineare Eigenschaft der cochleären Trennwand
 - Suppression der Schwingungsamplitude durch ein Suppressorsignal.

Damit konnte gezeigt werden, daß die Suppression als eine wesentliche physiologische Ursache der Verdeckung im Gehörmodell berücksichtigt wird.

Aufgrund des erforderlichen, unverhältnismäßig hohen Rechenaufwandes werden die folgenden Eigenschaften der Verdeckung bislang nicht berücksichtigt. Die erhöhte Mithörschwelle durch den Overshoot-Effekt im Einschaltmoment eines breitbandigen Maskierers wird nicht nachgebildet. Ebenso kann die verminderte Verdeckungswirkung aufgrund der binauralen Wahrnehmung im monauralen Gehörmodell nicht nachgebildet werden, die beispielsweise bei unterschiedlich lokalisierten Schallquellen auftreten kann („binaural masking level difference (BMLD)“).

Einige Anwendungen erfordern eine präzise quantitative Nachbildung von Ruhehörschwellen und Mithörschwellen. Die Simulationsergebnisse zeigen, daß eine weitgehende quantitative Übereinstimmung dieser Wahrnehmungsschwellen mit psychoakustischen Meßergebnissen erreicht wird, die für Kombinationen einfacher Maskierer und Testsignale vorliegen. Die bekannten Meßergebnisse wurden durch eigene Mithörschwellenmessungen ergänzt, die ein Schmalbandrauschen als Testsignal verwenden. Dieses Testsignal wurde gewählt, um eine genauere Bewertung des bei der Audiocodierung entstehenden Quantisierungsrauschens zu erreichen.

Für die Anwendung des Gehörmodells zur Steuerung eines dem ISO/IEC MPEG-2 AAC Standard entsprechenden Audiocoders wurde eine Steuereinheit entworfen. Der AAC-Standard enthält eine Teilbandcodierung mit anschließender Quantisierung der Teilbandabtwerte. Die Steuereinheit bewirkt eine geeignete Einstellung der in den Teilbändern verwendeten Quantisiererstufenbreiten derart, daß der Pegel der Quantisierungsfehler an der vom Gehörmodell bestimmten Wahrnehmungsschwelle liegt. Die Wahrnehmungsschwelle wird dabei nicht als Absolutwert für die Einstellung verwendet, sondern in Bezug auf den Pegel des entsprechenden Teilbandsignals angegeben und als Signal-Maskierungs-Abstand (SMR) bezeichnet. Die Einstellung der Quantisiererstufenbreiten erfolgt aus den oben genannten Gründen iterativ und beginnt mit einer initialen Quantisierereinstellung, die aus den Teilbandleistungen des Eingangssignals berechnet wird. Mit diesen Initialwerten wird innerhalb weniger Iterationsschritte bereits eine Quantisierereinstellung gefunden, die durch zusätzliche Iterationen nicht mehr signifikant verbessert wird.

Zur Bewertung der Klangqualität der AAC-Codierung mit dem entwickelten psycho-physiologischen Gehörmodell wurde als Referenzverfahren die AAC-Codierung mit einem psychoakustischen Modell herangezogen, das in den internationalen Hörtests während der Standardisierung des AAC eingesetzt wurde [97, 98, 99]. Für den Vergleich beider Coder wurden besonders kritische Audiosignale aus diesen Hörtests verwendet. Dazu gehören zwei Einzelinstrumentsignale, vier Sprachsignale und ein Gesangssignal. Um die Wahrnehmungsschwellen für die Quantisierung voll auszunutzen, wurden beide Coder mit zeitlich variabler Datenrate, aber gleicher, über alle Audiosignale gemittelter Datenrate, betrieben. Die Klangqualität wurde von sieben Testpersonen nach dem „triple stimulus — double blind — hidden reference“-Verfahren anhand einer 5-Punkte-Skala bewertet.

Die über alle verwendeten Audiosignale gemittelten Ergebnisse des Hörtests zeigen für beide Coder die gleiche Klangqualität. Bei den einzelnen Audiosignalen treten jedoch Unterschiede in der Klangqualität der beiden Coder auf. Die wesentlichen Ursachen für die beobachteten Unterschiede werden im folgenden genannt.

Das psychophysiologische Gehörmodell führt bei einem der beiden Einzelinstrumentalsignale zu einer Qualitätsverschlechterung. Dies ist vermutlich auf eine unzureichende Nachbildung der Wahrnehmung periodischer Amplitudenmodulationen im Gehörmodell zurückzuführen. Eine Überprüfung dieser Annahme konnte aus Zeitgründen leider nicht mehr durchgeführt werden. Bei einigen Sprachsignalen und dem Gesangssignal wird hingegen eine Qualitätsverbesserung erzielt. Die Ursachen dieser Verbesserung wurden mit Hilfe von synthetischen Audiosignalen genauer analysiert. Als synthetische Signale wurden frequenzmodulierte Signale verwendet, die häufig in stimmhaften Lauten von Sprachsignalen vorhanden sind. Dabei konnte gezeigt werden, daß die synthetischen Audiosignale bei dem psychoakustischen Modell des Referenzcoders zu einer fehlerhaften, überhöhten Wahrnehmungsschwelle führen, die eine deutlich verminderte Klangqualität zur Folge haben.

Ferner wurde beobachtet, daß das psychophysiologische Gehörmodell größere Schwankungen und eine stärkere Korrelation des SMR zum Audiosignal zeigt. Die Unterschiede in der Klangqualität sind demgegenüber geringer als aufgrund der signifikant unterschiedlichen SMR-Werte beider Coder zu erwarten ist. Desweiteren führt die zeitabhängige Schwankung der SMR-Werte zu einer entsprechend größeren Schwankung der variablen Datenrate des Coders mit dem Gehörmodell.

Eine Analyse des psychoakustischen Modells im Referenzcoder zeigt, daß dieses im Gegensatz zum psychophysiologischen Gehörmodell die folgenden wesentlichen Wahrnehmungseigenschaften nicht berücksichtigt. Das psychoakustische Modell bildet die Abhängigkeit der Verdeckungswirkung von der Einhüllendenfluktuation nicht explizit nach. Die Pegelabhängigkeit der Verdeckung und Suppressionseffekte werden nicht berücksichtigt. Die nichtlineare Additivität der Verdeckung durch mehrere Maskiersignale wird ebenfalls nicht nachgebildet. Eine quantitative Analyse der Auswirkungen dieser nicht nachgebildeten Wahrnehmungseigenschaften auf die Codierergebnisse wird ebenfalls Gegenstand zukünftiger Untersuchungen sein. Es ist jedoch zu erwarten, daß in umfassenderen Hörtests mit einer größeren Anzahl von Audiosignalen für das psychoakustische Modell hörbare Artefakte nachgewiesen werden können, die auf die fehlende Nachbildung der oben genannten Wahrnehmungseigenschaften zurückzuführen sind.

Die Implementierung des psychophysiologischen Gehörmodells auf einem Standard-PC mit 400-MHz-Prozessor benötigt aufgrund der Komplexität des Gehörmodells eine um den Faktor 240 größere Rechenzeit, als für die Codierung in Echtzeit notwendig ist. Für die Echtzeitrealisierung sind daher geeignete Ansätze zur Parallelisierung der Algorithmen und zur Komplexitätsverringerung des Gehörmodells noch zu prüfen.

Insgesamt zeigen die Ergebnisse der Verifikation und subjektiven Beurteilung der erzielten Klangqualität, daß das psychophysiologische Gehörmodell sowohl für einfache, als auch für komplexere Audiosignale die Wahrnehmungsschwellen adäquat nachbildet. Prinzipiell ist das Gehörmodell damit zur Beurteilung der Wahrnehmbarkeit von Signaländerun-

gen beliebig komplexer Audiosignale geeignet. Die einheitliche Nachbildung verschiedener Wahrnehmungsphänomene wird vom Gehörmodell vor allem durch eine geeignete Berücksichtigung der Nichtlinearität erzielt. Die daraus abgeleiteten Entstehungsursachen dieser Phänomene tragen auf dem Gebiet der Psychoakustik zum besseren Verstehen der Wahrnehmungsvorgänge bei.

Neben der Anwendung zur Steuerung eines Audiocoders ist der Einsatz des Gehörmodells für eine Nachbildung der subjektiven Qualitätsbewertung von Audiosignalen denkbar. Dazu ist eine Erweiterung des Gehörmodells um ein Modul notwendig, das eine geeignete Bewertung mehrerer überschwellig wahrnehmbarer Signaländerungen erlaubt und zu einer einzelnen Bewertungsgröße für die Klangqualität zusammenfaßt. Aufgrund der Generierung der spezifischen Lautheit als Zwischengröße im Gehörmodell ist prinzipiell auch eine Anwendung für die Lautheitsbestimmung von Audiosignalen möglich.

Ferner ist die Nachbildung von otoakustischen Emissionen mit dem Teilmodell der Cochlea in gleicher Weise möglich wie mit dem ursprünglichen Modell von Zwicker und Peisl, weil die daran beteiligten Modellkomponenten des ursprünglichen Modells im wesentlichen unverändert geblieben sind. Aufgrund der physiologischen Nachbildung des peripheren Gehörs können auch Gehörschädigungen in diesem Teil des Gehörs durch geeignete Modifikationen der Modellparameter simuliert werden. Die Auswirkungen unterschiedlicher Schädigungen auf die Wahrnehmung könnten somit durch das Gehörmodell nachgebildet und beispielsweise für die Optimierung von Hörhilfen genutzt werden.

7 Literaturverzeichnis

- [1] Advanced Television Systems Committee (ATSC) (1995). “Digital Audio Compression (AC-3) Standard,” *ATSC Document A-52*.
- [2] Allen, J. B. (1980). “Cochlear micromechanics — A physical model of transduction,” *J. Acoust. Soc. Am.*, Vol. 68, No. 6, 1660–1670.
- [3] Allen, J. B. (1993). “A second cochlear-frequency map that correlates distortion product and neural tuning measurements,” *J. Acoust. Soc. Am.*, Vol. 94 (1), No. 2, 809–816.
- [4] Allen, J. B.; Neely, S. T. (1992). “Micromechanical models of the cochlea,” *Physics Today*, July, 40–47.
- [5] Allen, J. B.; Sondhi, M. M. (1979). “Cochlear macromechanics: Time domain solution,” *J. Acoust. Soc. Am.*, Vol. 66, No. 1, 123–132.
- [6] Anderson, C. M. B.; Whittle, L. S. (1971). “Physiological noise and the missing 6 dB,” *Acustica*, Vol. 24, 261–272.
- [7] Arthur, R. M.; Pfeiffer, R. R.; Suga, N. (1971). “Properties of two-tone inhibition in primary auditory neurons,” *J. Physiol*, Vol. 212, 593–609.
- [8] Bacon, S. P.; Viemeister, N. F. (1985). “Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners,” *Audiology*, Vol. 24, 117–134.
- [9] Baumgarte, F. (1997). “A physiological ear model for auditory masking applicable to perceptual coding,” *Beitrag der 103rd AES Conv.*, New York, NY, Sept. 1997, Preprint 4511.
- [10] Baumgarte, F. (1997). “A physiological ear model for specific loudness and masking,” *Proc. IEEE Workshop on Applications of Sig. Proc. to Audio and Acoustics*, New Paltz, NY, Okt. 1997.
- [11] Baumgarte, F. (1998). “Evaluation of a physiological ear model considering masking effects relevant to audio coding,” *Beitrag der 105th AES Conv.*, San Francisco, CA, Sept. 1998, Preprint 4789.
- [12] Baumgarte, F. (1998). “Generierung von Mithörschwellen durch ein physiologisches Gehörmodell,” *Fortschritte der Akustik DAGA '98*, Zürich, Schweiz, März 1998.
- [13] Baumgarte, F. (1999). “Evaluation of a physiological ear model for the simulation of nonlinear masking effects,” *Collected papers from the joint meeting “Berlin 99”* (ASA, EAA und DAGA), Berlin, März 1999.
- [14] Baumgarte, F. (1999). “A physiological ear model for the emulation of masking,” *J. Oto-Rhino-Laryngology*, Vol. 61, No. 5.
- [15] Baumgarte, F. (1999). “Application of a physiological ear model to irrelevance reduction in audio coding,” *Proc. AES 17th Int. Conf. on High Quality Audio Coding*, Signa, Italien, Sept. 1999.

-
- [16] Beerends, J. G. (1995). "Measuring the quality of speech and music codecs, an integrated psychoacoustic approach," *Beitrag der 98th AES Conv.*, Paris, Februar 1995, Preprint 3945.
- [17] Beerends, J. G.; Stemerding, J. A. (1992). "A perceptual audio quality measure based on a psychoacoustic sound representation," *J. Audio Eng. Soc.*, Vol. 40, No. 12, 963–978.
- [18] Beerends, J. G.; Stemerding, J. A. (1994). "A perceptual speech-quality measure based on a psychoacoustic sound representation," *J. Audio Eng. Soc.*, Vol. 42, No. 3, 115–123.
- [19] Békésy, G. v. (1960). *Experiments in hearing*, McGraw-Hill, New York u. a.
- [20] Bilger, R. C. (1959). "Additivity of different types of masking," *J. Acoust. Soc. Am.*, Vol. 31, No. 8, 1107–1109.
- [21] Blauert, J. (1974). *Räumliches Hören*, Hirzel Verlag, Stuttgart.
- [22] Blauert, J. (1985). *Räumliches Hören – Nachschrift. Neue Ergebnisse und Trends seit 1982*, Hirzel Verlag, Stuttgart.
- [23] Boer, E. d. (1995). "The 'inverse problem' solved for a three-dimensional model of the cochlea. I. Analysis," *J. Acoust. Soc. Am.*, Vol. 98 (1), No. 2, 896–903.
- [24] Boer, E. d. (1995). "The 'inverse problem' solved for a three-dimensional model of the cochlea. II. Application to experimental data sets," *J. Acoust. Soc. Am.*, Vol. 98 (1), No. 2, 904–910.
- [25] Boer, E. d. (1995). "On equivalence of locally active models of the cochlea," *J. Acoust. Soc. Am.*, Vol. 98, No. 3, 1400–1409.
- [26] Boer, E. d. (1996). "Mechanics of the cochlea: modeling efforts," in: Dallos, P.; Popper, A. N.; Fay R. R. (Hrsg.). *The cochlea*, Springer-Verlag, New York u. a.
- [27] Bos, C. E.; Boer, E. d. (1966). "Masking and discrimination," *J. Acoust. Soc. Am.*, Vol. 39, No. 4, 708–715.
- [28] Bosi, M.; Brandenburg, K.; Quackenbush, S.; Fielder, L.; Akagiri, K.; Fuchs, H.; Dietz, M.; Herre, J.; Davidson, G.; Oikawa, Y. (1997). "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, Vol. 45, No. 10, 789–814.
- [29] Brandenburg, K. (1987). "Evaluation of quality for audio encoding at low bit rates," *Beitrag der 82nd AES Conv.*, London, 1987, Preprint 2433.
- [30] Brandenburg, K.; Bosi, M. (1997). "Overview of MPEG Audio: Current and future standards for low-bit-rate audio coding," *J. Audio Eng. Soc.*, Vol. 45, No. 1/2, 4–21.
- [31] Brandenburg K.; Sporer, T. (1992). "'NMR' and 'Masking Flag': Evaluation of quality using perceptual criteria," *Proc. 11th Int. AES Conf.*, Portland, Oregon, Mai 1992.
- [32] Buunen, T. J. F. (1975). "Two hypotheses on monaural phase effects," *Acustica*, Vol. 34, 98–105.

-
- [33] Buus, S. (1985). "Release from masking caused by envelope fluctuations," *J. Acoust. Soc. Am.*, Vol. 78, No. 6, 1958–1965.
- [34] Buus, S.; Florentine, M.; Poulsen, T.; Müsch, H. (1999). "On the relation between peripheral compression and the growth of loudness," *J. Acoust. Soc. Am.*, Vol. 105, No. 2, 1022.
- [35] Clarey, J. C.; Barone, P.; Imig, T. J. (1992). "Physiology of thalamus and cortex," in: Popper, A. N.; Fay R. R. (Hrsg.). *The mammalian auditory pathway: Neurophysiology*, Springer-Verlag, New York u. a.
- [36] Cokeley, C. G.; Humes, L. E. (1993). "Two experiments on the temporal boundaries for the nonlinear addition of masking," *J. Acoust. Soc. Am.*, Vol. 94, No. 5, 2553–2559.
- [37] Colomes, C.; Lever, M.; Rault, J. B.; Dehery, Y. F. (1995). "A perceptual model applied to audio bit-rate reduction," *J. Audio Eng. Soc.*, Vol. 43, No. 4, 233–239.
- [38] Dau, T. (1996). *Modeling auditory processing of amplitude modulation*, Dissertation, Bibliotheks- und Informationssystem Univ. Oldenburg.
- [39] Dau, T.; Kollmeier, B.; Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *J. Acoust. Soc. Am.*, Vol. 102, No. 5 (1), 2892–2905.
- [40] Dau, T.; Kollmeier, B.; Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.*, Vol. 102, No. 5 (1), 2906–2919.
- [41] Dau, T.; Püschel, D.; Kohlrausch, A. (1996). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Am.*, Vol. 99, No. 6, 3615–3622.
- [42] Dau, T.; Püschel, D.; Kohlrausch, A. (1996). "A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements," *J. Acoust. Soc. Am.*, Vol. 99, No. 6, 3623–3631.
- [43] Delgutte, B. (1990). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," *J. Acoust. Soc. Am.*, Vol. 87, No. 2, 791–809.
- [44] Delgutte, B. (1996). "Physiological models for basic auditory percepts," in: Hawkins, H. L.; McMullen, T. A.; Popper, A. N.; Fay R. R. (Hrsg.). *Auditory computation*, Springer-Verlag, New York u. a.
- [45] Deng, L. (1992). "Processing of acoustic signals in a cochlear model incorporating laterally coupled suppressive elements," *Neural Networks*, Vol. 5, 19–34.
- [46] Diependaal, R. J. (1988). *Nonlinear and active cochlear models. Analysis and solution methods*. Dissertation, TU Delft.
- [47] Diependaal, R. J.; Boer E. de; Viergever, M. A. (1987). "Cochlear power flux as an indicator of mechanical activity," *J. Acoust. Soc. Am.*, Vol. 82, No. 3, 917–926.

-
- [48] Egan, J. P.; Hake, H. W. (1950). "On the masking pattern of a single auditory stimulus," *J. Acoust. Soc. Am.*, Vol. 22, 622–630.
- [49] Fassel, R. (1994). *Experimente und Simulationsrechnungen zur Wahrnehmung von Amplitudenmodulation im menschlichen Gehör*, Dissertation, Univ. Göttingen, Verlag Shaker, Aachen.
- [50] Fastl, H. (1976). "Temporal masking effects: I. Broad band noise masker," *Acustica*, Vol. 35, No. 5, 287–302.
- [51] Fastl, H. (1976/77). "Temporal masking effects: II. Critical band noise masker," *Acustica*, Vol. 36, No. 5, 317–331.
- [52] Fastl, H. (1979). "Temporal masking effects: III. Pure tone masker," *Acustica*, Vol. 43, No. 43, 282–294.
- [53] Fastl, H.; Baumgartner, H. (1975). „Ruhehörschwellen für Sinustöne bzw. Schmalbandrauschen," *Acustica*, Vol. 34, 111–114.
- [54] Fletcher, H. (1953). "Masking effects," Chapter 10 in: Fletcher, H., *Speech and hearing in communication*, van Nostrand, Toronto u. a.
- [55] Fletcher, H.; Munson, W. A. (1933), "Loudness, its definition, measurement, and calculation," *J. Acoust. Soc. Am.*, Vol. 9, 1–10.
- [56] Fettweis, A. (1971). "Digital filter structures related to classical filter networks," *Arch. Elektronik Übertragungstechnik*, Vol. 25, 79–89.
- [57] Fettweis, A. (1986). "Wave digital filters: theory and practice," *Proc. IEEE*, Vol. 74, 270–327.
- [58] Frank, G.; Kössl, M. (1996). "The acoustic two-tone distortions $2f_1-f_2$ and f_2-f_1 and their possible relation to changes in the operating point of the cochlear amplifier," *Hear. Res.*, Vol. 98, 104–115.
- [59] Fuchs, H. (1995). "Improving MPEG audio coding by backward adaptive linear stereo prediction," *Beitrag der 99th AES Conv.*, New York, NY, Okt. 1995, Preprint 4086.
- [60] Geisler, C. D. (1993). "A realizable cochlear model using feedback from motile outer hair cells," *Hear. Res.*, Vol. 68, 253–262.
- [61] Geisler, C. D.; Sang, C. (1995). "A cochlear model using feed-forward outer-hair-cell forces," *Hear. Res.*, Vol. 86, No. 1/2, 132–146.
- [62] Giguère, C.; Woodland, P. C. (1994). "A computational model of the auditory periphery for speech and hearing research. I. Ascending paths," *J. Acoust. Soc. Am.*, Vol. 95, No. 1, 331–342.
- [63] Giguère, C.; Woodland, P. C. (1994). "A computational model of the auditory periphery for speech and hearing research. II. Descending paths," *J. Acoust. Soc. Am.*, Vol. 95, No. 1, 343–349.

-
- [64] Glasberg, B. R.; Moore, B. C. J. (1994). "Growth-of-masking functions for several types of maskers," *J. Acoust. Soc. Am.*, Vol. 96, No. 1, 134–144.
- [65] Goldstein, J. L.; Kiang, N. Y. S. (1967). "Neural correlates of the aural combination tone," *Proc. IEEE*, Vol. 56, 981–992.
- [66] Gralla, G. (1991). *Wahrnehmungskriterien bei Mithörschwellenmessungen und deren Simulation in Rechnermodellen*, Dissertation, TU München.
- [67] Gralla, G. (1992). „Wahrnehmung bei Simultan- und Nachhörschwellenmessungen,“ *Acustica*, Vol. 77, 243–251.
- [68] Gralla, G. (1993). „Modelle zur Beschreibung der Wahrnehmungskriterien bei Mithörschwellenmessungen,“ *Acustica*, Vol. 78, 233–245.
- [69] Green, D. M. (1967). "Additivity of masking," *J. Acoust. Soc. Am.*, Vol. 41, No. 6, 1517–1525.
- [70] Green, D. M. (1993). "Auditory intensity discrimination," Kap. 2 in: Yost, W. A.; Popper, A. N.; Fay, R. R. (Hrsg.) *Human psychophysics*, Springer-Verlag, New York u. a.
- [71] Green, D. M.; Swets, J. A. (1966). *Signal detection theory and psychophysics*, Wiley, New York.
- [72] Gulick, W. L.; Gescheider, G. A.; Frisna, R. D. (1989). *Hearing. Physiological acoustics, neural coding, and psychoacoustics*, Oxford University Press, New York u. a.
- [73] Gummer, A. W.; Hemmert, W.; Zenner, H.-P. (1996). "Resonant tectorial membrane motion in the inner ear: Its crucial role in the frequency tuning," *Proc. Natl. Acad. Sci. USA*, Vol. 93, 8727–8732.
- [74] Hall, J. L. (1997). "Asymmetry of masking revisited: Generalization of masker and probe bandwidth," *J. Acoust. Soc. Am.*, Vol. 101, No. 2, 1023–1033.
- [75] Hall, J. W.; Haggard, M. P.; Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," *J. Acoust. Soc. Am.*, Vol. 76, 50–60.
- [76] Hallworth, R.; Evans, B. N.; Dallos, P. (1993). "The location and mechanism of electromotility in guinea pig outer hair cells," *J. Neurophysiology*, Vol. 70, No. 2, 549–558.
- [77] Hansen, M. (1998). *Assessment and prediction of speech transmission quality with an auditory processing model*, Dissertation, Universität Oldenburg.
- [78] Hansen, M.; Kollmeier, B. (1997). "Using a quantitative psychoacoustical signal representation for objective speech quality measurement," *Proc. ICASSP '97*, Vol. II, München, April 1997, 1387–1390.
- [79] Harris, D. M.; Dallos, P. (1979). "Forward masking of auditory nerve fiber responses," *J. Neurophys.*, Vol. 42, No. 4, 1083–1107.
- [80] Hartmann, W. M.; Pumplin, J. (1988). "Noise power fluctuations and the masking of sine signals," *J. Acoust. Soc. Am.*, Vol. 83, No. 6, 2277–2289.

-
- [81] Heijden, M. v. d. (1995). *A comparison of masking by tones and noise*, Dissertation, Techn. Universität Eindhoven.
- [82] Heijden, M. v. d.; Kohlrausch, A. (1994). "Using an excitation pattern model to predict auditory masking," *Hear. Res.*, Vol. 80, 38–52.
- [83] Heijden, M. v. d.; Kohlrausch, A. (1995). "The role of envelope fluctuations in spectral masking," *J. Acoust. Soc. Am.*, Vol. 97, No. 3, 1800–1807.
- [84] Hellmann, R. P. (1972). "Asymmetry of masking between noise and tone," *Perception & Psychophysics*, Vol. 11, No. 3, 241–246.
- [85] Hemmert, W. (1997). *Studie zur Aufklärung des aktiven Verstärkungsmechanismus im Hörorgan mittels Vibrationsmessungen auf zellulären Strukturen in der Meer-schweinchencochlea*, Dissertation, Ruhr-Universität Bochum.
- [86] Herre, J.; Johnston, J. D. (1996). "Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS)," *Beitrag der 101th AES Conv.*, Los Angeles, CA, Nov. 1996, Preprint 4384.
- [87] Hewitt, M. J.; Meddis, R. (1994). "A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus," *J. Acoust. Soc. Am.*, Vol. 95, No. 4, 2145–2159.
- [88] Hewitt, M. J.; Meddis, R. (1991). "An evaluation of eight computer models of mammalian inner hair-cell function," *J. Acoust. Soc. Am.*, Vol. 90 (1), No. 2, 904–917.
- [89] Holley, M. C. (1996). "Outer hair cell motility," in: Dallos, P.; Popper, A. N.; Fay R. R. (Hrsg.). *The cochlea*, Springer-Verlag, New York u. a.
- [90] Hubbard, A. E.; Mountain, D. C. (1996). "Analysis and synthesis of cochlear mechanical function using models," in: Hawkins, H. L.; McMullen, T. A.; Popper, A. N.; Fay R. R. (Hrsg.). *Auditory computation*, Springer-Verlag, New York u. a.
- [91] Hudde, H. (1998). „Grundeigenschaften des Mittelohres auf der Basis modellbezogener Messungen," *Fortschritte der Akustik – DAGA 98*, 320–321.
- [92] Hudde, H.; Engel, E. (1998). "Measuring and modeling basic properties of the human middle ear and ear canal. Part I: Model structure and measuring techniques," *Acustica – acta acustica*, Vol. 84, No. 4, 720–738.
- [93] Hudde, H.; Engel, E. (1998). "Measuring and modeling basic properties of the human middle ear and ear canal. Part II: Ear canal, middle ear cavities, eardrum and ossicles," *Acustica – acta acustica*, Vol. 84, No. 5, 894–913.
- [94] Hudde, H.; Engel, E. (1998). "Measuring and modeling basic properties of the human middle ear and ear canal. Part III: Eardrum impedances, transfer functions and model calculations," *Acustica – acta acustica*, Vol. 84, No. 6, 1091–1109.
- [95] Humes, L. E.; Jesteadt, W. (1989). "Models of the additivity of masking," *J. Acoust. Soc. Am.*, Vol. 85, No. 3, 1285–1294.

-
- [96] Humes, L. E.; Lee, L. W.; Jesteadt, W. (1992). "Two experiments on the spectral boundary conditions for nonlinear additivity of simultaneous masking," *J. Acoust. Soc. Am.*, Vol. 92, No. 5, 2598–2606.
- [97] ISO/IEC JTC1/SC29/WG11 N1279 (1996). "NBC Reference Model 3 monophonic subjective tests: overall results," Tampere, Finland.
- [98] ISO/IEC JTC1/SC29/WG11 N1280 (1996). "NBC Reference Model 4 stereophonic and multichannel subjective tests: overall results," Tampere, Finland.
- [99] ISO/IEC JTC1/SC29/WG11 N1419 (1996). "Report on the formal subjective listening tests of MPEG-2 NBC multichannel audio coding," Maceió, Brasilien.
- [100] ISO/IEC JTC1/SC29/WG11 N1420 (1996). "Overview of the report on the formal listening tests of MPEG-2 AAC multichannel audio coding," Maceió, Brasilien.
- [101] ISO/IEC JTC1/SC29/WG11 (1997). "Coding of moving pictures and audio – MPEG-2 Advanced Audio Coding," *ISO/IEC 13818-7 international standard*.
- [102] ISO/IEC JTC1/SC29/WG11 N2006 (1998). "Report on the MPEG-2 AAC stereo verification tests," San Jose, CA.
- [103] ITU-R task group 10/4(1998). "Method for objective measurement of perceived audio quality," *Draft new recommendation ITU-R BS.PEAQ* (Doc. 10/20).
- [104] ITU-R (1994). "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems", *Recommendation BS.1116*, Geneva.
- [105] ITU-T (1996). "Objective quality measurement of telephoneband (300–3400 Hz) speech codecs," *ITU-T recommendation P.861*, August 1996.
- [106] Iwasa, K. H.; Chadwick, R. S. (1992). "Elasticity and active force generation of the cochlear outer hair cells," *J. Acoust. Soc. Am.*, Vol. 92, No. 6, 3169–3173.
- [107] Javel, E. (1986). "Basic response properties of auditory nerve fibers," in: Altschuler, R. A.; Hoffman, D. W.; Bobbin, R. P. (Hrsg.). *Neurobiology of hearing: The cochlea*, Raven-Press, New York.
- [108] Jayant, N.; Johnston, J.; Safranek, R. (1993). "Signal compression based on models of human perception," *Proc. IEEE*, Vol. 81, No. 10, 1385–1421.
- [109] Jesteadt, W.; Bacon, S. P.; Lehman, J. R. (1982). "Forward masking as a function of frequency, masker level, and signal delay," *J. Acoust. Soc. Am.*, Vol. 71, No. 4, 950–962.
- [110] Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.*, Vol. 64, No. 5, 2156–2165.
- [111] Kemp, D. T. (1979). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," *Arch. Otorhinolaryngo.*, No. 224, 37–45.
- [112] Kidd, G. Jr.; Feth, L. L. (1982). "Effects of masker duration in pure-tone forward masking," *J. Acoust. Soc. Am.*, Vol. 72, No. 5, 1384–1386.

-
- [113] Klitzing, R. v.; Kohlrausch, A. (1994). "Effect of masker level on overshoot in running- and frozen-noise maskers," *J. Acoust. Soc. Am.*, Vol. 95, No. 4, 2192–2201.
- [114] Kohlrausch, A.; Fassel, R.; Heijden, M. v. d.; Kortekaas, R.; Par, S. v. d.; Oxenham, A. J.; Püschel, D. (1997). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," *Acustica – acta acustica*, Vol. 4, 659–669.
- [115] Kohlrausch, A.; Sander, A. (1995). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," *J. Acoust. Soc. Am.*, Vol. 97, No. 3, 1817–1829.
- [116] Kros, C. J. (1996). "Physiology of mammalian cochlear hair cells," in: Dallos, P.; Popper, A. N.; Fay R. R. (Hrsg.). *The cochlea*, Springer-Verlag, New York u. a.
- [117] Langner, G. (1992). "Periodicity coding in the auditory system," *Hear. Res.*, Vol. 60, 115–142.
- [118] Lumer, G. (1985). *Nachbildung nichtlinearer simultaner Verdeckungseffekte bei schmalbandigen Schallen mit einem Rechnermodell*, Dissertation, TU München.
- [119] Lumer, G. (1987). "Computer model of cochlear preprocessing (steady state condition). I. Basics and results for one sinusoidal input signal," *Acustica*, Vol. 62, 282–290.
- [120] Lumer, G. (1987). "Computer model of cochlear preprocessing (steady state condition). II. Two-tone suppression," *Acustica*, Vol. 63, 17–25.
- [121] Lutfi, R. A. (1983). "Additivity of simultaneous masking," *J. Acoust. Soc. Am.*, Vol. 73, No. 1, 262–267.
- [122] Lutfi, R. A. (1985). "A power-law transformation predicting masking by sounds with complex spectra," *J. Acoust. Soc. Am.*, Vol. 77, No. 6, 2128–2136.
- [123] Lutfi, R. A. (1990). "How much masking is informational masking?" *J. Acoust. Soc. Am.*, Vol. 88, No. 6, 2607–2610.
- [124] Lyon, R.; Shamma, S. (1996). "Auditory representations of timbre and pitch," in: Hawkins, H. L.; McMullen, T. A.; Popper, A. N.; Fay R. R. (Hrsg.). *Auditory computation*, Springer-Verlag, New York u. a.
- [125] Mammano, F.; Nobili, R. (1993). "Biophysics of the cochlea: Linear approximation," *J. Acoust. Soc. Am.*, Vol. 93, No. 6, 3320–3332.
- [126] McNicol, D. (1972). *A primer of signal detection theory*, Allen & Unwin, London u.a.
- [127] Meddis, R. (1986). "Simulation of mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.*, Vol. 79, No. 3, 702–711.
- [128] Meddis, R. (1988). "Simulation of auditory-neural transduction: Further studies," *J. Acoust. Soc. Am.*, Vol. 83, No. 3, 1056–1063.
- [129] Meddis, R.; Hewitt, M. J.; Shackleton, T. M. (1990). "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse," *J. Acoust. Soc. Am.*, Vol. 87, No. 4, 1813–1816.

-
- [130] Moore, B. C. J. (1985). "Additivity of simultaneous masking, revisited," *J. Acoust. Soc. Am.*, Vol. 78, No. 2, 488–494.
- [131] Moore, B. C. J.; Alcántara, J. I.; Dau, T. (1998). "Masking patterns for sinusoidal and narrow-band noise maskers," *J. Acoust. Soc. Am.*, Vol. 104, No. 2 (1), 1023–1038.
- [132] Moore, B. C. J.; Glasberg, B. R.; Baer, T. (1997), "A model for the prediction of threshold, loudness, and partial loudness," *J. Audio Eng. Soc.*, No. 4, Vol. 45, 224–240.
- [133] Moore, B. C. J.; Vickers, D. A. (1997). "The role of spread excitation and suppression in simultaneous masking," *J. Acoust. Soc. Am.*, Vol. 102, No. 4, 2284–2290.
- [134] Mott, J. B.; McDonald, L. P.; Sinex, D. G. (1990). "Neural correlates of psychophysical release from masking," *J. Acoust. Soc. Am.*, Vol. 88, No. 6, 2682–2691.
- [135] Mountain, D. C.; Hubbard, A. E. (1996). "Computational analysis of hair cell and auditory nerve processes," in: Hawkins, H. L.; McMullen, T. A.; Popper, A. N.; Fay R. R. (Hrsg.). *Auditory computation*, Springer-Verlag, New York u. a.
- [136] Neely, S. T.; Kim, D. O. (1983). "An active cochlear model showing sharp tuning and high sensitivity," *Hear. Res.*, Vol. 9, 123–130.
- [137] Neely, S. T.; Kim, D. O. (1986). "A model for active elements in cochlear biomechanics," *J. Acoust. Soc. Am.*, Vol. 79, No. 5, 1472–1480.
- [138] Neely, S. T. (1993). "A model of cochlear mechanics with outer hair cell motility," *J. Acoust. Soc. Am.*, Vol. 94, No. 1, 137–146.
- [139] Neff, D. L. (1995). "Signal properties that reduce masking by simultaneous, random-frequency maskers," *J. Acoust. Soc. Am.*, Vol. 98, No. 4, 1909–1920.
- [140] Nobili, R.; Mammano, F. (1996). "Biophysics of the cochlea II: Stationary nonlinear phenomenology," *J. Acoust. Soc. Am.*, Vol. 99 (1), No. 4, 2244–2255.
- [141] Nobili, R.; Mammano, F.; Ashmore, J. (1998). "How well do we understand the cochlea?" *Trends in Neuroscience (TINS)*, Vol. 21, No. 4, 159–167.
- [142] Nutall, A. L.; Dolan, D. F.; Avinash, G. (1990). "Measurement of basilar membrane tuning and distortion with laser doppler velocimetry," in: Dallos, P. et al. (Hrsg.). *The mechanics and biophysics of hearing*, Springer-Verlag, New York.
- [143] Nutall, A. L.; Dolan, D. F. (1993). "Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig," *J. Acoust. Soc. Am.*, Vol. 93, No. 1, 390–400.
- [144] Nutall, A. L.; Dolan, D. F. (1996). "Steady-state sinusoidal responses of the basilar membrane in guinea pig," *J. Acoust. Soc. Am.*, Vol. 99, No. 3, 1556–1565.
- [145] Nutall, A. L.; Ren, T. (1995). "Electromotile hearing: evidence from basilar membrane motion and otoacoustic emissions," *Hear. Res.*, Vol. 92, 170–177.
- [146] Oetinger, R.; Hauser, H. (1961). „Ein elektrischer Kettenleiter zur Untersuchung der mechanischen Schwingungsvorgänge im Innenohr,“ *Acustica*, Vol. 11, 161–177.

-
- [147] Oxenham, A. J.; Moore, B. C. J. (1994). "Modeling the additivity of nonsimultaneous masking," *Hear. Res.*, Vol. 80, 105–118.
- [148] Oxenham, A. J.; Plack, C. J. (1998). "Suppression and the upward spread of masking," *J. Acoust. Soc. Am.*, Vol. 104, No. 6, 3500–3510.
- [149] Paillard, B.; Mabillean, P.; Morissette, S.; Soumagne, J. (1992). "PERCEVAL: Perceptual evaluation of the quality of audio signals," *J. Audio Eng. Soc.*, Vol. 40, No. 1, 21–31.
- [150] Pascal, J.; Bourgeade, A.; Lagier, M.; Legros, C. (1998). "Linear and nonlinear model of the human middle ear," *J. Acoust. Soc. Am.*, Vol. 104 (1), No. 3, 1509–1516.
- [151] Patuzzi, R. (1996). "Cochlear micromechanics and macromechanics," in: Dallos, P.; Popper, A. N.; Fay R. R. (Hrsg.). *The cochlea*, Springer-Verlag, New York u. a.
- [152] Patuzzi, R. B.; Yates, G. K.; Johnstone, B. M. (1989). "Outer hair cell receptor current and sensorineural hearing loss," *Hear. Res.*, Vol. 42, 47–72.
- [153] Peisl, W. (1990). *Beschreibung aktiver nichtlinearer Effekte der peripheren Schallverarbeitung des Gehörs durch ein Rechnermodell*, Dissertation, TU München.
- [154] Penner, M. J. (1974). "Effect of masker duration and masker level on forward and backward masking," *J. Acoust. Soc. Am.*, Vol. 56, No. 1, 179–182.
- [155] Pickles, J. O. (1988). *An introduction to the physiology of hearing – 2nd ed.*. Academic Press, London u. a.
- [156] Plack, C. J.; Oxenham, A. J. (1998). "Basilar-membrane nonlinearity and the growth of forward masking," *J. Acoust. Soc. Am.*, Vol. 103, No. 3, 1598–1608.
- [157] Pollack, I. (1975). "Auditory informational masking," *J. Acoust. Soc. Am.*, Vol. 57, Suppl. 1, S5.
- [158] Pumplin, J. (1985). "Low-noise noise," *J. Acoust. Soc. Am.*, Vol. 78, No. 1, 100–104.
- [159] Rabiner, L. R.; Gold, B. (1975). *Theory and application of digital signal processing*, Prentice-Hall, Englewood Cliffs, NJ.
- [160] Rhode, S. R.; Greenberg, S. (1992). "Physiology of the cochlear nucleus," in: Popper, A. N.; Fay R. R. (Hrsg.). *The mammalian auditory pathway: Neurophysiology*, Springer-Verlag, New York u. a.
- [161] Robles, L.; Ruggero, M. A.; Rich, N. C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," *J. Acoust. Soc. Am.*, Vol. 80, No. 5, 1364–1374.
- [162] Robles, L.; Ruggero, M. A.; Rich, N. C. (1990). "Two-tone distortion products in the basilar membrane of the chinchilla cochlea," in: Dallos, P. et al. (Hrsg.). *The mechanics and biophysics of hearing*, Springer-Verlag, New York.
- [163] Rosowski, J. J. (1996). "Models of external- and middle-ear function," in: Hawkins, H. L.; McMullen, T. A.; Popper, A. N.; Fay R. R. (Hrsg.). *Auditory computation*, Springer-Verlag, New York u. a.

-
- [164] Ruggero, M. A.; Rich, N. C.; Recio, A.; Narayan, S. S.; Robles, L. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.*, Vol. 101, No. 4, 2151–2163.
- [165] Ruggero, M. A.; Robles, L.; Rich, N. C.; (1992). "Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression," *J. Neurophysiology*, Vol. 48, No. 4, 1087–1099.
- [166] Russell, I. J.; Cody, A. R.; Richardson, G. P. (1986). "The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and mouse cochlea grown in vitro," *Hear. Res.*, Vol. 22, 199–216.
- [167] Schlauch, R. S.; DiGiovanni, J. J.; Ries, D. T. (1998). "Basilar membrane nonlinearity and loudness," *J. Acoust. Soc. Am.*, Vol. 103, No. 4, 2010–2020.
- [168] Schooneveldt, G. P.; Moore, B. C. J. (1989). "Comodulation masking release as a function of masker bandwidth, modulator bandwidth and signal duration," *J. Acoust. Soc. Am.*, Vol. 85, 273–281.
- [169] Schroeder, M. R. (1975). "Models of hearing," *Proc. IEEE*, Vol. 63, No. 9, 1332–1350.
- [170] Schroeder, M. R.; Atal, B. S.; Hall, J. L. (1979). "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Am.*, Vol. 66, No. 6, 1647–1652.
- [171] Sellick, P. M.; Russell, I. J. (1980). "The responses of inner hair cells to basilar membrane velocity during low-frequency auditory stimulation in the guinea-pig," *Hear. Res.*, Vol. 2, 439–445.
- [172] Shaw, E. A. G. (1974). "The external ear," in: Keidel, W. D.; Neff, W. D. (Hrsg.). *Handbook of sensory physiology*, Vol. 5/1, 455–490, Springer-Verlag, Berlin.
- [173] Shepherd, G. M. (1993). *Neurobiologie*, Springer-Verlag, Berlin u. a.
- [174] Smith, R. L. (1979). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," *J. Acoust. Soc. Am.*, Vol. 65, No. 1, 166–178.
- [175] Smoorenburg, G. F. (1972), "Combination tones and their origin," *J. Acoust. Soc. Am.*, Vol. 52, 615–631.
- [176] Soderquist, D. R.; Lindsey, J. W. (1972). "Physiological noise as a masker of low frequencies: The cardiac cycle," *J. Acoust. Soc. Am.*, Vol. 52 (2), No. 4, 1216–1220.
- [177] Souloudre, G. A.; Grusec, T.; Lavoie, M.; Thibault, L. (1998). "Subjective evaluation of state-of-the-art two-channel audio codecs," *J. Audio Eng. Soc.*, No. 3, Vol. 46, 164–177.
- [178] Sporer, T. (1997). "Objective audio signal evaluation – applied psychoacoustics for modeling the perceived quality of digital audio," *Beitrag der 103rd AES Conv.*, New York, October 1997, Preprint 4512.

-
- [179] Stinson, M. R.; Lawton, B. W. (1989). "Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution," *J. Acoust. Soc. Am.*, Vol. 85, No. 6, 2492–2503.
- [180] Strube, H. W. (1985). "A computational efficient basilar-membrane model," *Acustica*, Vol. 58, 207–214.
- [181] Thiede, T.; Kabot, E. (1996). "A new perceptual quality measure for bit rate reduced audio," *Beitrag der 100th AES Conv.*, Copenhagen, Mai 1996, Preprint 4280.
- [182] Todd, C. C.; Davidson, G. A.; Davis, M. F.; Fielder, L. D.; Link, B. D.; Vernon, S. (1994). "AC-3: Flexible perceptual coding for audio transmission and storage," *Beitrag der 96th AES Conv.*, Amsterdam, Februar 1994, Preprint 3796.
- [183] Tsutsui, K.; Suzuki, H.; Shimoyoshi, O.; Sonohara, M.; Akagiri, K.; Heddle, R. M. (1992). "ATRAC: Adaptive Transform Acoustic Coding for MiniDisc," *Beitrag der 93rd AES Conv.*, San Francisco, Oktober 1992, Preprint 3456.
- [184] Turner, C. W.; Relkin, E. M. (1994). "Psychophysical and physiological forward masking studies: Probe duration and rise-time effects," *J. Acoust. Soc. Am.*, Vol. 96 (1), No. 2, 795–800.
- [185] Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.*, Vol. 66, No. 5, 1364–1380.
- [186] Weber, D. L.; Patterson, R. D. (1984). "Sinusoidal and noise maskers in simultaneous and forward masking," *J. Acoust. Soc. Am.*, Vol. 75, No. 3, 925–931.
- [187] Widin, G. P.; Viemeister, N. F. (1979). "Intensive and temporal effects in pure-tone forward masking," *J. Acoust. Soc. Am.*, Vol. 66, No. 2, 388–395.
- [188] Yates, G. K.; Kirk, D. L. (1984). "Electrically evoked travelling waves in the guinea pig cochlea," *Symposium Biophysics*, 296–302.
- [189] Yates, G. K.; Winter, I. M.; Robertson, D. (1990). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," *Hear. Res.*, Vol. 45, 203–220.
- [190] Zenner, H.-P. (1994), *Hören. Physiologie, Biochemie, Zell- und Neurobiologie*, Thieme Verlag, Stuttgart u. a.
- [191] Zwicker, E. (1954). "Die Verdeckung von Schmalbandgeräuschen durch Sinustöne," *Acustica*, Vol. 4, 415–420.
- [192] Zwicker, E. (1977). "Procedure for calculating loudness of temporally variable sounds," *J. Acoust. Soc. Am.*, Vol. 62, No. 3, 675–682.
- [193] Zwicker, E. (1979). "A model describing nonlinearities in hearing by active processes with saturation at 40 dB," *Biological Cybernetics*, Vol. 35, 243–250.
- [194] Zwicker, E. (1984). "Dependence of post-masking on masker duration and its relation to temporal effects in loudness," *J. Acoust. Soc. Am.*, Vol. 75, No. 1, 219–223.

-
- [195] Zwicker, E. (1986). "A hardware cochlear nonlinear preprocessing model with active feedback," *J. Acoust. Soc. Am.*, Vol. 80, No. 1, 146–152.
- [196] Zwicker, E.; Fastl, H. (1972). „Zur Abhängigkeit der Nachverdeckung von der Störimpulsdauer," *Acustica*, Vol. 26, No. 2, 78–82.
- [197] Zwicker, E.; Fastl, H. (1999). *Psychoacoustics. Facts and models*, 2nd Ed., Springer-Verlag, New York u. a.
- [198] Zwicker, E.; Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger*, Hirzel Verlag, Stuttgart.
- [199] Zwicker, E.; Herla, S. (1975). „Über die Addition von Verdeckungseffekten," *Acustica*, Vol. 34, 89–97.
- [200] Zwicker, E.; Jaroszewski, A. (1982). "Inverse frequency dependence of simultaneous tone-on-tone masking patterns at low levels," *J. Acoust. Soc. Am.*, Vol. 71, No. 6, 1508–1512.
- [201] Zwicker, E.; Peisl, W. (1990). "Cochlear preprocessing in analog models, in digital models and in human inner ear," *Hear. Res.*, Vol. 44, 209–216.
- [202] Zwicker, E.; Stecker, M.; Hind, J. (1987). "Relation between masking, otoacoustic emissions, and evoked potentials," *Acustica*, Vol. 64, 102–109.
- [203] Zwicker, E.; Wright, H. N. (1963). "Temporal summation for tones in narrow-band noise," *J. Acoust. Soc. Am.*, Vol. 35, 691–699.
- [204] Zwicker, E.; Zollner, M. (1987). *Elektroakustik*, Springer-Verlag, Berlin u. a.
- [205] Zwislocki, J. J. (1969). "Temporal summation of loudness: An analysis," *J. Acoust. Soc. Am.*, Vol. 46 (2), No. 2, 431–441.
- [206] Zwislocki, J. (1962). "Analysis of middle-ear function. Part I: Input impedance," *J. Acoust. Soc. Am.*, Vol. 34 (2), No. 8, 1514–1523.

Anhang A Algorithmen und Parameter des psychophysiologischen Gehörmodells

Inhalt

| | |
|---|-----|
| A.1 Pegelnormierung der Modellsignale | 137 |
| A.2 Signalfensterung | 138 |
| A.3 Kraft-Strom-Analogie | 139 |
| A.4 Implementierung des Tiefpasses erster Ordnung | 140 |
| A.5 Grundlagen der Wellendigitalfilter | 142 |
| A.6 Implementierung des Außen- und Mittelohrmodells | 146 |
| A.7 Implementierung des Modells der Hydromechanik und der äußeren Haarzelle | 148 |
| A.8 Implementierung des Modells der inneren Haarzelle | 157 |
| A.9 Parameter des Modells der neuronalen Verarbeitung | 156 |
| A.10 Psychoakustisches Meßverfahren zur Bestimmung der Wahrnehmungsschwelle | 160 |

A.1 Pegelnormierung der Modellsignale

Die in dieser Arbeit verwendeten Schalldrucksignale (aufgezeichnete Mikrofonaufnahmen, synthetische Signale) besitzen in der Regel keine Normierung bezüglich des physikalischen Schalldruckpegels (SPL) am Mikrofon oder für die Wiedergabe. Für Vergleiche mit anderen Daten ist aber eine Definition der Abhängigkeit zwischen einem vorliegenden dimensionslosen Schalldrucksignal und dem SPL sinnvoll. Der Schalldruckpegel eines reellen mittelwertfreien Schalldrucksignals $x(t)$ wird in dieser Arbeit definiert als:

$$L \text{ [dB SPL]} = 10 \log_{10} \mathcal{E}\{x^2\}. \quad (\text{A.1})$$

Damit ergibt sich der Zusammenhang zwischen der physikalischen Wechselschalldruckamplitude $p(t)$ in Pa und dem (aufgezeichneten) dimensionslosen Schalldrucksignal $x(t)$ als

$$p(t) = p_0 x(t). \quad (\text{A.2})$$

Mit der Beziehung (A.2) wird der Zusammenhang zu der üblichen SPL-Definition über den Effektivwert p des Schalldrucks $p(t)$ und dem Referenzdruck $p_0 = 20 \mu\text{Pa}$ hergestellt:

$$L \text{ [dB SPL]} = 20 \log_{10} \frac{p}{p_0}. \quad (\text{A.3})$$

In Übereinstimmung mit [153] wird für die im Modell der Cochlea auftretenden Spannungen u und Ströme i folgende Pegelzuordnung über deren Effektivwerte definiert:

$$L [\text{dB}] = 20 \log_{10} \frac{u_{\text{eff}}}{0,1 \text{ mV}} \quad (\text{A.4})$$

$$L [\text{dB}] = 20 \log_{10} \left(i_{\text{eff}} \frac{274 \Omega}{0,1 \text{ mV}} \right) \quad (\text{A.5})$$

In dieser Arbeit werden häufig Signale endlicher Dauer verwendet, die in der Regel durch die Anwendung einer Fensterfunktion auf das zugrundeliegende Signal gewonnen werden. Der angegebene Pegel des so gewonnenen Signals endlicher Dauer berücksichtigt wegen der besseren Vergleichbarkeit nicht den Verlauf der Fensterfunktion, sondern entspricht dem Pegel des zugrundeliegenden, ungefensterten Signals.

A.2 Signalfensterung

Für die Verarbeitung von Eingangssignalen definierter Dauer wurde eine geeignete Fensterfunktion festgelegt, deren Dauer mit einem Parameter eingestellt werden kann. Die zeitdiskrete Fensterfunktion hat die Form einer Gaußfunktion gemäß (A.6).

$$w(n) = \exp\left(-\frac{n^2}{2\sigma^2}\right) \quad (\text{A.6})$$

Der Parameter σ wird in Abhängigkeit von der Fensterdauer T_F und der Abtastfrequenz f_A berechnet:

$$\sigma = \frac{T_F f_A}{\sqrt{\pi}} \quad (\text{A.7})$$

Die Fensterdauer T_F wird als diejenige Dauer in Sekunden bezogen auf die Abtastfrequenz f_A definiert, in der die Gaußfunktion (A.6) größere Werte als 0,675 liefert. In Bild A.1 ist der prinzipielle Verlauf der Fensterfunktionen dargestellt, die am Anfang des Signals zum Einblenden bzw. am Ende des Signals zum Ausblenden verwendet werden. Zwischen den Fensterfunktionen wird das Signal nicht modifiziert. Die Signaldauer T_M bzw. T_T ist definiert als die Summe der beiden Fensterdauern und der Dauer zwischen den Fensterfunktionen. Die Fensterfunktionen werden bis zu einem Abfall von -100 dB gegenüber dem Maximalwert angewendet. Außerhalb dieses Bereichs vor dem Einblendfenster und nach dem Ausblendfenster ist das Signal identisch null.

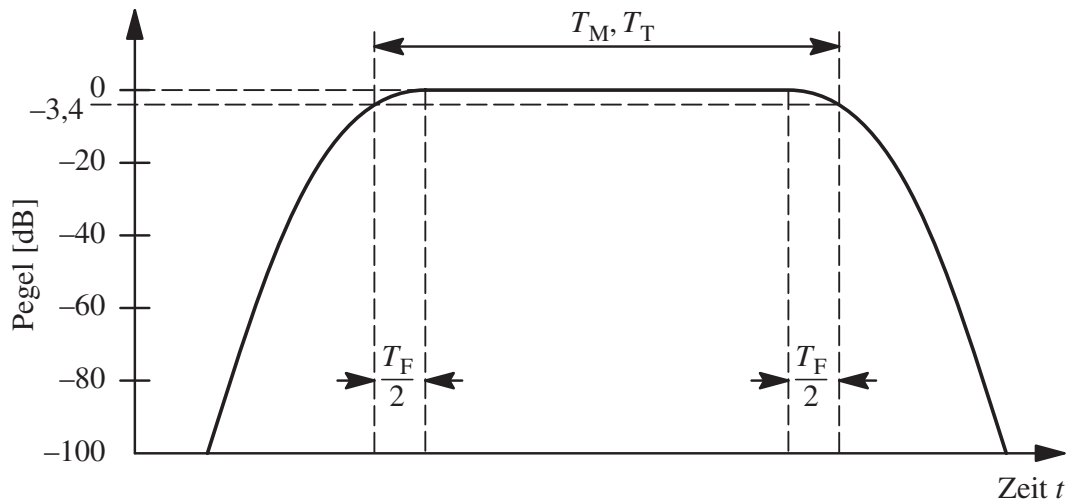


Bild A.1 Zur Definition der Parameter der Fensterfunktion.

A.3 Kraft-Strom-Analogie

Die Analogie zwischen mechanischen, akustischen und elektrischen Größen ermöglicht eine äquivalente Beschreibung eines Systems aus einem dieser Bereiche unter Verwendung der analogen Größen eines der anderen Bereiche [204]. Die weit entwickelten Verfahren zur Analyse und Simulation von elektrischen Netzwerken können auf dieser Grundlage auch auf mechanische und akustische Netzwerke angewendet werden, wenn diese unter Verwendung der Analogie in das äquivalente elektrische Netzwerk transformiert worden sind.

Alternativ zur Kraft-Strom-Analogie mechanischer und elektrischer Netzwerke, die im folgenden motiviert und in dieser Arbeit verwendet wird, existiert die Kraft-Spannungs-Analogie. Die Grundgleichungen der Mechanik zur Berechnung der Kraft F gemäß dem Newtonschen, Stokesschen und Hookeschen Gesetz können durch die Gleichungen (A.8) als Funktion der Schnelle v angegeben werden.

$$F = m \frac{dv}{dt}; \quad F = R_m v; \quad F = \frac{1}{C_m} \int v dt \quad (\text{A.8})$$

Die entsprechenden Grundgleichungen gemäß der Kraft-Strom-Analogie für elektrische Größen lauten:

$$i = C \frac{du}{dt}; \quad i = G u; \quad i = \frac{1}{L} \int u dt. \quad (\text{A.9})$$

Aus einem „Koeffizientenvergleich“ von (A.8) und (A.9) folgen die entsprechend der Kraft-Strom-Analogie äquivalenten Größen, die in den ersten beiden Spalten von Tabelle A.1 gegenübergestellt sind.

| mechanische Größe | elektrische Größe | Zusammenhang |
|----------------------|-------------------|-------------------|
| Kraft F | Strom i | $i = F / N_i$ |
| Schnelle v | Spannung u | $u = N_i v$ |
| Masse m | Kapazität C | $C = m / N_i^2$ |
| Widerstand R_m | Leitwert G | $G = R_m / N_i^2$ |
| Nachgiebigkeit C_m | Induktivität L | $L = N_i^2 / C_m$ |

Tab. A.1 Analogie mechanischer Größen zu elektrischen Größen.

Der quantitative Zusammenhang zwischen den sich entsprechenden Größen ist aufgrund der Gleichungen (A.8) und (A.9) nur bis auf eine frei wählbare Proportionalitätskonstante N_i festgelegt. Diese Konstante stellt den Zusammenhang zwischen der Kraft F und dem Strom i her:

$$F = N_i i. \quad (\text{A.10})$$

Unter der Voraussetzung gleicher Leistungen ergeben sich die Zusammenhänge zwischen mechanischen und elektrischen Größen wie in der dritten Spalte von Tabelle A.1 angegeben.

Die Analogiebeziehungen zwischen akustischen und elektrischen Netzwerken können dementsprechend abgeleitet werden. Üblicherweise wird in diesem Bereich die Druck-Strom-Analogie, oder alternativ die Druck-Spannungs-Analogie verwendet.

A.4 Implementierung des Tiefpasses erster Ordnung

In den Blockdiagrammen zur Darstellung des IHZ- und NV-Modells in Kapitel 3.3.2 und Kapitel 3.4 wird wegen der Konsistenz zu der Beschreibung der Modelle zur Nachbildung des peripheren Gehörs in Form von elektrischen Netzwerken ein RC -Glied zur Symbolisierung eines Tiefpasses erster Ordnung verwendet. Für eine hypothetische Realisierung dieser Teilmodelle mit elektrischen Widerständen und Kondensatoren ist zu beachten, daß eine ausreichende Entkopplung zusammenschalteter RC -Glieder durch Pufferverstärker gewährleistet sein muß.

Aufgrund der geringen Anforderungen an die Rechenleistung wird für die zeitdiskrete digitale Realisierung eines Filters erster Ordnung kein WDF sondern das rekursive Filter in Bild A.2 (*mitte*) verwendet. Der Filterparameter k wird dabei so eingestellt, daß sich die gleiche Zeitkonstante für das RC -Glied (Bild A.2 *links*) und das rekursive Filter ergibt.

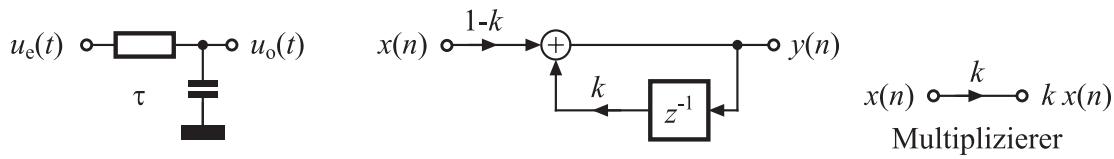


Bild A.2 Tiefpaß erster Ordnung: RC -Glied mit der Zeitkonstanten τ (*links*) und zeitdiskrete digitale Nachbildung als rekursives Filter mit dem Parameter k (*mitte*). Durch z^{-1} wird die Verzögerung um ein Abtastintervall symbolisiert. Die Multiplikation einer Folge mit einem Skalar wird durch einen Pfeil symbolisiert (*rechts*).

Die Ausgangsfolge $y(n)$ des rekursiven Filters berechnet sich aus der Eingangsfolge $x(n)$ und die um ein Abtastintervall verzögerte Ausgangsfolge $y(n - 1)$ gemäß:

$$y(n) = k y(n - 1) + (1 - k) x(n). \quad (\text{A.11})$$

Die Bestimmung des Filterparameters k wird in Abhängigkeit von der Zeitkonstanten τ des RC -Gliedes und der Abtastfrequenz f_A des zeitdiskreten Filters zunächst gemäß (A.12) vorausgesetzt:

$$k = \exp\left(-\frac{1}{\tau f_A}\right). \quad (\text{A.12})$$

Die Bestimmung der Zeitkonstanten wird anhand der Impulsantworten abgeleitet. Die Impulsantwort $h_{RC}(t)$ eines RC -Gliedes mit dem Widerstand R und der Kapazität C ist proportional zu einer Exponentialfunktion:

$$h_{RC}(t) \sim \exp\left(-\frac{t}{RC}\right) \quad (\text{A.13})$$

Die Zeitkonstante des RC -Gliedes ist durch $\tau = RC$ definiert. Im Vergleich dazu lautet die Impulsantwort des zeitdiskreten Filters nach (A.11):

$$h(n) = (1 - k) k^n. \quad (\text{A.14})$$

Mit (A.12) folgt

$$h(n) = (1 - k) \exp\left(-\frac{n}{\tau f_A}\right). \quad (\text{A.15})$$

Mit $\tau = RC$ und $t = n/f_A$ folgt aus (A.13) und (A.15):

$$h_{RC}\left(\frac{n}{f_A}\right) \sim h(n) \quad (\text{A.16})$$

Mit (A.16) ist aufgrund der Proportionalität gezeigt, daß die Zeitkonstanten von einem RC -Glied und dem entsprechenden rekursiven Digitalfilter nach (A.12) übereinstimmen. Diese Eigenschaft ist unabhängig vom Impedanzniveau des Widerstands bzw. der Kapazität, da nur das Produkt $\tau = RC$ in die Formel eingeht.

Weitere Kenngrößen der Tiefpaßfilter sind die 3 dB Eckfrequenz und die Integrationsgrenzdauer. Die 3 dB Eckfrequenz des RC -Gliedes beträgt:

$$f_{RC,g} = \frac{1}{2\pi \tau}. \quad (\text{A.17})$$

Im Vergleich dazu beträgt die 3 dB Eckfrequenz des zeitdiskreten Tiefpaßfilters:

$$f_g = \frac{f_A}{2\pi} \arccos\left(1 - \frac{(1-k)^2}{2k}\right) = \frac{f_A}{2\pi} \arccos\left(2 - \cosh\left(\frac{1}{\tau f_A}\right)\right) \approx \frac{1}{2\pi \tau}. \quad (\text{A.18})$$

Die Näherung in (A.18) ist für $\tau f_A > 1$ gültig. Die Integrationsgrenzdauer T_g , gemessen als 3 dB Wert unterhalb des Grenzwertes für $n \rightarrow \infty$ des zeitdiskreten digitalen Tiefpasses bei Verwendung einer Einheitsprungfunktion als Eingangssignal, beträgt:

$$T_g = -\tau \ln\left(1 - \frac{1}{\sqrt{2}}\right) - \frac{1}{f_A} \approx 1,23 \tau \quad (\text{A.19})$$

Verschiedene Blöcke des NV-Modells enthalten zur Minimum- oder Maximumbildung zwischen Eingangs- und Ausgangsspannung eine zusätzliche Diode in der symbolischen Darstellung des RC -Glieds. Die entsprechende zeitdiskrete digitale Nachbildung erfolgt bei Durchlaßrichtung der Diode vom Eingang zum Ausgang mit:

$$y(n) = \text{Max}(k y(n-1) + (1-k)x(n), x(n)). \quad (\text{A.20})$$

Bei umgekehrter Durchlaßrichtung vom Ausgang zum Eingang erfolgt die Nachbildung mit:

$$y(n) = \text{Min}(k y(n-1) + (1-k)x(n), x(n)). \quad (\text{A.21})$$

Die Minimum- und Maximumfunktion sind definiert als:

$$\text{Max}(a, b) = \begin{cases} a; & a > b \\ b; & a \leq b \end{cases} \quad (\text{A.22})$$

$$\text{Min}(a, b) = \begin{cases} a; & a < b \\ b; & a \geq b. \end{cases} \quad (\text{A.23})$$

A.5 Grundlagen der Wellendigitalfilter

In der Regel werden Wellendigitalfilter angewendet, um ein zeitkontinuierliches, sog. „Referenzfilter“ als zeitdiskretes Filter digital nachzubilden. Das vorgegebene Referenzfilter wird im Frequenzbereich üblicherweise mit Hilfe der komplexen Frequenzvariablen p beschrieben. Für die Nachbildung als zeitdiskretes WDF wird eine Bilineartransformation verwendet, die den Zusammenhang zwischen der komplexen Frequenzvariable p des zeit-

kontinuierlichen Referenzfilters und der komplexen Frequenzvariable ψ des zeitdiskreten WDFs mit der Abtastperiode $T_A = 1/f_A$ gemäß (A.24) herstellt.

$$\psi = \frac{z-1}{z+1} = \tanh\left(\frac{pT_A}{2}\right); \quad z = e^{pT_A} \quad (\text{A.24})$$

Unter der Voraussetzung, daß für die komplexe Frequenz gilt $p = j\omega$, folgt der Zusammenhang zwischen den reellen Frequenzvariablen im zeitkontinuierlichen und zeitdiskreten Bereich ω bzw. ϕ wie in (A.25) angegeben. Dieser Zusammenhang führt zu einer umkehrbaren Abbildung von Frequenzen des Referenzfilters bis zur Nyquistfrequenz $0 < \omega < \pi f_A$ auf den Bereich $0 < \phi < \infty$. Diese nichtlineare Abbildung der Frequenzachse bewirkt eine ungleichförmige Verzerrung der Frequenzskala.

$$\phi = \tan\left(\frac{\omega T_A}{2}\right); \quad p = j\omega; \quad \psi = j\phi \quad (\text{A.25})$$

WDFs beruhen auf der Einführung von Wellengrößen, wobei eine in ein Eintor hineinlaufende Welle (hinlaufende Welle) a und eine herauslaufende Welle (rücklaufende Welle) b gemäß (A.26) definiert werden. Die darin verwendeten Größen sind die Spannung u , der Strom i sowie der reelle Torwiderstand R .

$$a = u + R i; \quad b = u - R i \quad (\text{A.26})$$

Aus (A.26) folgen die entsprechenden Gleichungen zur Bestimmung des Stroms und der Spannung aus den Wellengrößen:

$$u = \frac{a+b}{2}; \quad i = \frac{a-b}{2R}. \quad (\text{A.27})$$

Die Darstellung der Wellengrößen erfolgt so, wie in Bild A.3 (*links*) für das Eintor μ beispielhaft gezeigt ist. Im zugehörigen Wellenflußdiagramm (Bild A.3 *rechts*), das zur Darstellung von WDFs verwendet wird, treten die Spannungen und Ströme nicht mehr auf.

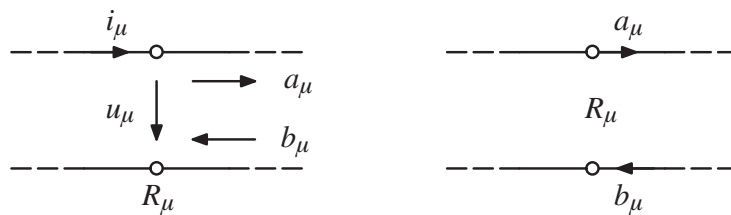


Bild A.3 *Links*: elektrisches Eintor μ mit Torwiderstand R_μ und hin- und rücklaufender Welle a_μ bzw. b_μ . *Rechts*: Zugehöriges Wellenflußdiagramm im WDF Bereich.

Die Repräsentation von elektrischen reaktiven Elementen und Spannungsquellen als WDF ist in Bild A.4 für einige Grundelemente wiedergegeben [57]. Die serielle oder parallele Zusammenschaltung von elektrischen Eintoren wird im zugeordneten WDF durch einen entsprechenden Adapter repräsentiert, der im wesentlichen die Wechselwirkung der unterschiedlichen zusammenschalteten Torwiderstände auf die Wellengrößen berücksichtigt.

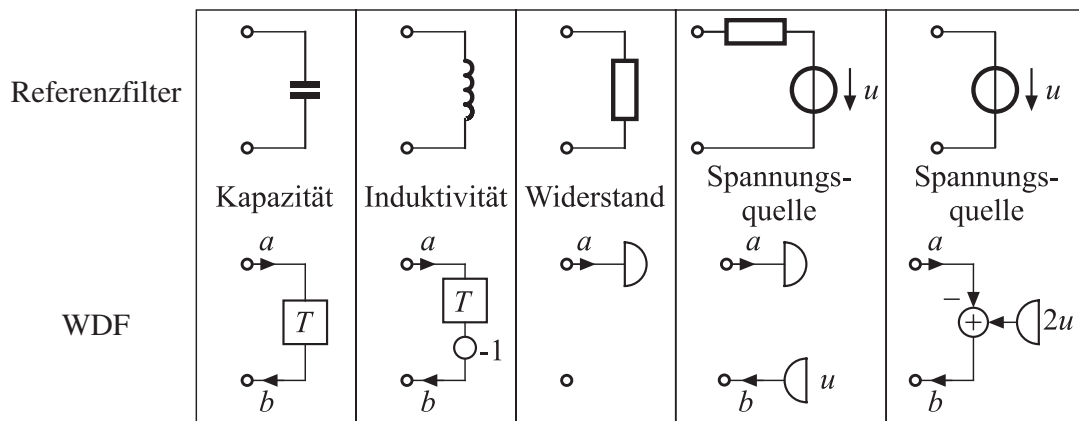
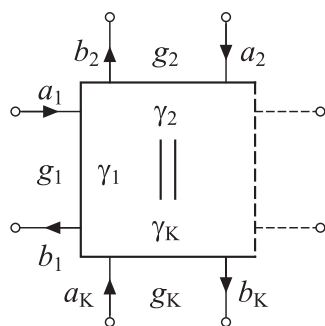


Bild A.4 Eintorige elektrische Grundelemente und deren zugeordnete Realisierung im WDF Bereich dargestellt als Wellenflußdiagramm. Der mit T bezeichnete Block symbolisiert die Signalverzögerung um ein Abtastintervall. Ein offener Kreis symbolisiert die Multiplikation mit der angegebenen Konstanten. Halbkreise zeigen Sinken bzw. Quellen für eine Wellengröße.

Die Symbole für Paralleladaptoren, die eine Parallelschaltung elektrischer Elemente repräsentieren, sind in Bild A.5 und Bild A.6 dargestellt. Bei bekannten Torleitwerten g_μ , die von den Elementen des Referenzfilters abhängen, folgen die WDF-Koeffizienten γ_μ aus den angegebenen Gleichungen. Mit Hilfe der WDF-Koeffizienten können dann die rücklaufenden Wellen b_μ aus den hinlaufenden Wellen a_μ bestimmt werden. Der Paralleladapter in Bild A.6 besitzt im Gegensatz zu dem Adapter in Bild A.5 ein reflektionsfreies Tor K . Das bedeutet, daß die rücklaufende Welle b_K nur von den hinlaufenden Wellen a_μ abhängt. Diese Eigenschaft ermöglicht die Vermeidung von verzögerungsfreien Schleifen, die in WDFs nicht realisierbar sind [57].



$$\gamma_\mu = 2 g_\mu \left[\sum_{v=1}^K g_v \right]^{-1} ; \text{ für } \mu = 1, \dots, K$$

$$b_K = a_K - \sum_{v=1}^{K-1} \gamma_v (a_K - a_v)$$

$$b_\mu = b_K + (a_K - a_\mu) ; \text{ für } \mu = 1, \dots, K - 1$$

Bild A.5 Links: Symbol des WDF-Paralleladapters mit K Toren. Rechts: Formeln zur Bestimmung der WDF-Koeffizienten γ_μ (γ_K ist redundant) und der hin- und rücklaufenden Wellen a_μ bzw. b_μ .

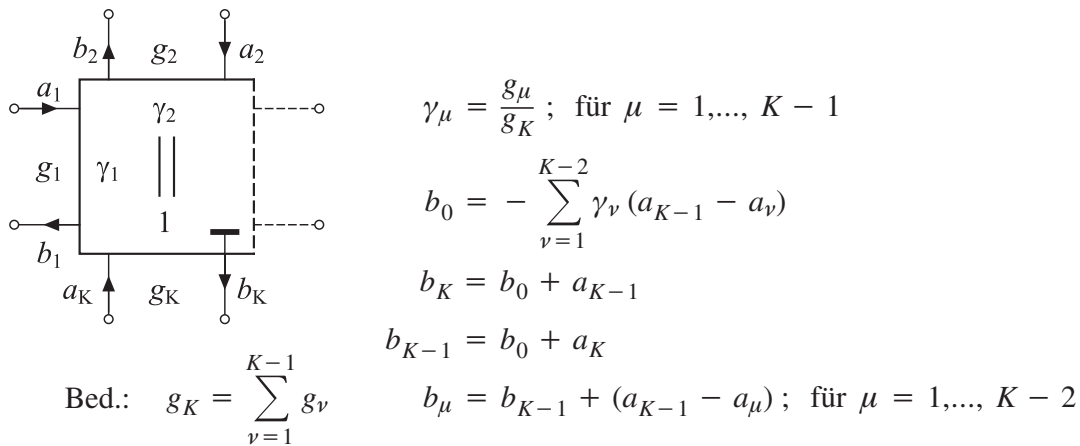


Bild A.6 Links: Symbol des WDF-Paralleladapters mit K Toren, wobei das Tor K reflektionsfrei ist. Rechts: Formeln zur Bestimmung der WDF-Koeffizienten γ_μ (γ_{K-1} ist redundant) und der hin- und rücklaufenden Wellen a_μ bzw. b_μ .

Der im Gehörmodell verwendete Serienadapter zur Repräsentation einer elektrischen Reihen- oder Serienschaltung ist in Bild A.7 angegeben. Das Tor K des Serienadapters ist reflektionsfrei. Die Bestimmung der WDF-Koeffizienten erfolgt mit der angegebenen Gleichung aus den Torwiderständen r_μ . Insgesamt ergeben sich $K - 2$ unabhängige Koeffizienten, die zur Bestimmung der Wellengrößen verwendet werden.

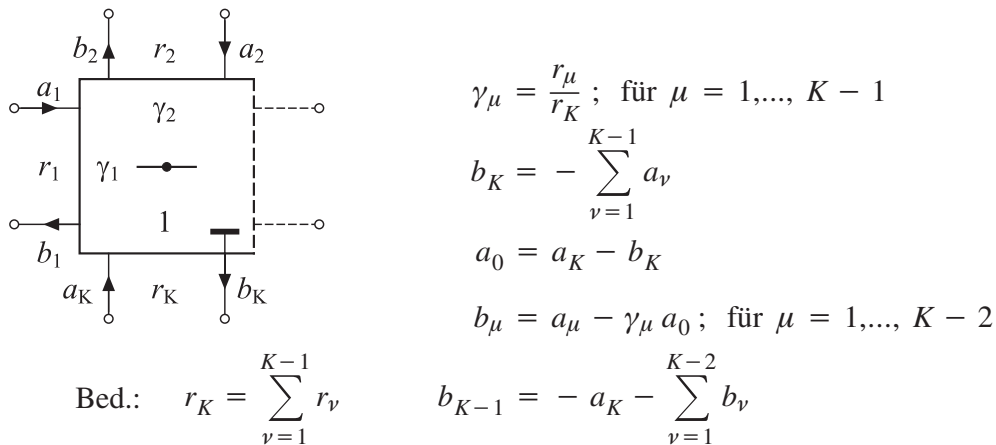
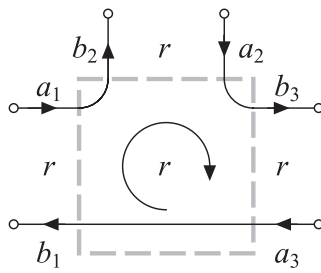


Bild A.7 Links: Symbol des WDF Serienadapters mit K Toren, wobei das Tor K reflektionsfrei ist. Rechts: Formeln zur Bestimmung der WDF Koeffizienten γ_μ (γ_{K-1} ist redundant) und der hin- und rücklaufenden Wellen a_μ bzw. b_μ .

Als weiteres WDF Element wird im Gehörmodell ein Zirkulator verwendet, dessen WDF-Symbol mit den zugehörigen Zusammenhängen der Wellengrößen in Bild A.8 angegeben ist. Der WDF-Zirkulator verwendet keine Filterkoeffizienten. Er wird lediglich mit dem identischen Torwiderstand r aller Tore gekennzeichnet.



$$b_1 = a_K$$

$$b_\mu = a_{\mu-1}; \text{ für } \mu = 2, \dots, K$$

Bild A.8 Wellenflußdiagramm des Zirkulators für $K = 3$. Beziehungen der Wellengrößen eines allgemeinen Zirkulators mit K Toren.

Für eine eindeutige Zuordnung eines WDFs zu einem Referenzfilter ist die Festlegung von zwei Bezugsgrößen erforderlich, die die Skalierung der Frequenzen bzw. Impedanzen im zeitdiskreten WDF bezüglich des zeitkontinuierlichen Referenzfilters bestimmen.

Die Verzerrung der Frequenzskala aufgrund der Bilineartransformation erlaubt die Übereinstimmung der Frequenzen in der WDF-Simulation und dem Referenzfilter nur bei einer vorgegebenen Bezugsfrequenz f_B . Für die Bestimmung der WDF-Größen wird daher die Normierungskreisfrequenz ω_N im zeitkontinuierlichen Referenzfilterbereich zugrunde gelegt, die in (A.28) angegeben ist. Dabei geht die verwendete Abtastfrequenz f_A mit ein.

$$\omega_N = 2\pi f_B \cot\left(\pi \frac{f_B}{f_A}\right) \quad (\text{A.28})$$

Die Bezugsfrequenz wird, wenn nicht anders angegeben, sinnvollerweise bei einer mittleren Frequenz im wahrnehmbaren Frequenzbereich gewählt:

$$f_B = 1000 \text{ Hz} . \quad (\text{A.29})$$

Zur Bestimmung der Filterkoeffizienten der WDF muß ein Bezugswiderstand R_B gewählt werden, mit dem die Umsetzung des Impedanzniveaus vom Referenzfilter zum WDF festgelegt wird:

$$R_B = 1000 \Omega . \quad (\text{A.30})$$

A.6 Implementierung des Außen- und Mittelohrmodells

Die Dimensionierung des Netzwerks zur Nachbildung des Außen- und Mittelohrmodells aus Bild 3.2 ist durch die folgenden Werte in (A.31) bis (A.37) gegeben. Das Impedanzniveau des Netzwerks wurde an die Eingangsimpedanz des nachfolgenden Modells angepaßt.

$$R_{M1} = 12,7 \Omega \quad (\text{A.31})$$

$$R_{M2} = 31,75 \Omega \quad (\text{A.32})$$

$$R_A = 6,35 \Omega \quad (\text{A.33})$$

$$L_M = 3,175 \text{ mH} \tag{A.34}$$

$$L_A = 1,7875 \text{ mH} \tag{A.35}$$

$$L_A = 1,7875 \text{ mH} \tag{A.36}$$

$$C_A = 1,575 \text{ }\mu\text{F} \tag{A.37}$$

Für die Implementierung des Modells als WDF wird das elektrische Netzwerk aus Bild 3.2 zunächst in das Wellenflußdiagramm gemäß Bild A.9 überführt. Die darin enthaltenen Filterkoeffizienten γ werden mit Hilfe von Torwiderständen r bzw. Torleitwerten g bestimmt. Die Stromquelle mit dem Strom i_{AO} wird aufgrund des geringeren Realisierungsaufwands durch eine Spannungsquelle mit hohem Innenwiderstand ($R_i = 6,35 \text{ k}\Omega$) ersetzt.

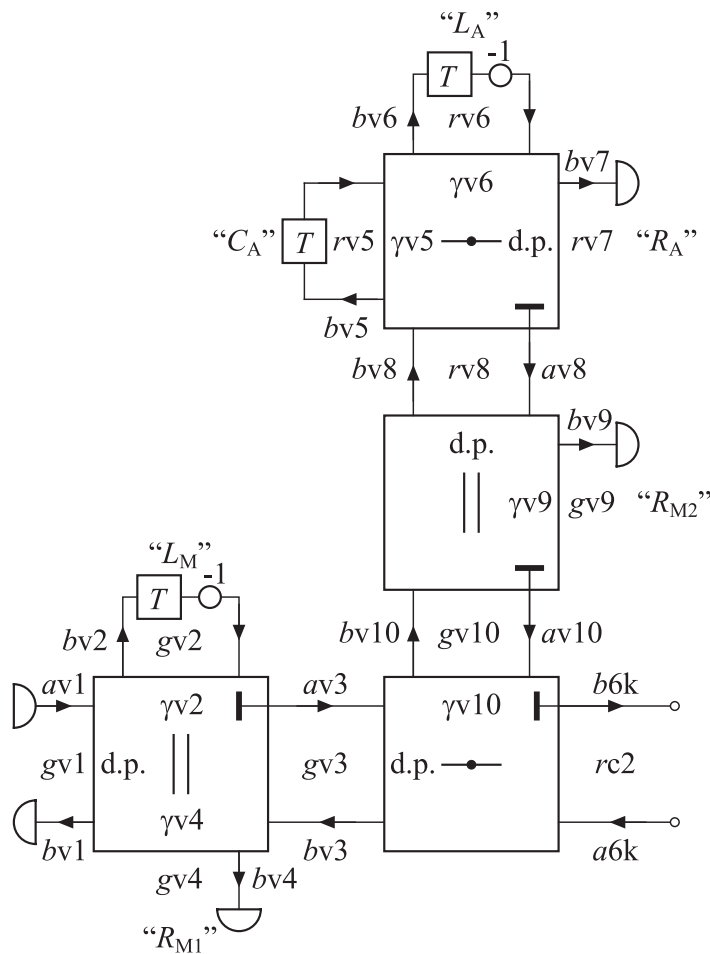


Bild A.9 Wellenflußdiagramm des Außen- und Mittelohrmodells aus Bild 3.2. Die Stromquelle i_{AO} ist durch eine Spannungsquelle im WDF ersetzt worden. Die entsprechenden Elemente des Referenzfilters sind im Wellenflußdiagramm durch Hochkommata gekennzeichnet.

Die WDF-Torwiderstände und -leitwerte des AMO-Modells berechnen sich aus den elektrischen Netzwerkelementen wie folgt:

$$\begin{aligned}
gv1 &= \frac{R_B}{R_i} & rv6 &= \frac{\omega_N L_A}{R_B} \\
gv2 &= \frac{R_B}{\omega_N L_M} & rv7 &= \frac{R_A}{R_B} \\
gv4 &= \frac{R_B}{R_{M1}} & rv8 &= rv5 + rv6 + rv7 \\
gv3 &= gv1 + gv2 + gv4 & gv9 &= \frac{R_B}{R_{M2}} \\
rv5 &= \frac{1}{\omega_N C_A R_B} & gv10 &= \frac{1}{rv8} + gv9 \\
& & rc2 &= \frac{1}{gv3} + \frac{1}{gv10}
\end{aligned}$$

Die WDF-Koeffizienten werden wie folgt aus den Torwiderständen bzw. -leitwerten berechnet:

$$\begin{aligned}
\gamma v2 &= \frac{gv2}{gv3} & \gamma v6 &= \frac{rv6}{rv8} \\
\gamma v4 &= \frac{gv4}{gv3} & \gamma v9 &= \frac{gv9}{gv10} \\
\gamma v5 &= \frac{rv5}{rv8} & \gamma v10 &= \frac{rc2}{gv10}
\end{aligned}$$

Der WDF-Algorithmus wird im Zusammenhang mit dem Modell der Hydromechanik und der äußeren Haarzellen in Kapitel A.7 angegeben.

A.7 Implementierung des Modells der Hydromechanik und der äußeren Haarzelle

Der Abstand der Resonanztonheit benachbarter Sektionen des Modells der Hydromechanik wird mit $\Delta z = 0,1$ Bark festgelegt. Die Anzahl der Sektionen beträgt $N_{\text{Sek}} = 251$. Die Resonanzfrequenzen $f_{\text{Res},\nu}$ der Parallelschwingkreise im Modell der Hydromechanik berechnen sich gemäß (A.38) bis (A.40).

$$f_{\text{Res},1} = 5 \text{ Hz} \frac{\Delta z}{0,1 \text{ Bark}} \quad (\text{A.38})$$

$$f_{\text{Res},2} = 10 \text{ Hz} \frac{\Delta z}{0,1 \text{ Bark}} \quad (\text{A.39})$$

$$f_{\text{Res},\nu} = f_{\text{Res},\nu-1} + \frac{\Delta z}{\text{Bark}} \left[25 + 75 \left[1 + 1,4 \left(\frac{f_{\text{Res},\nu-1}}{\text{kHz}} \right)^2 \right]^{0,69} \right] \quad (\text{A.40})$$

für $\nu = 3, 4, \dots, N_{\text{Sek}}$

Die Resonanzfrequenzen $f_{\text{Res},\nu}$ werden als sog. „charakteristische Frequenz (CF)“ der cochleären Trennwand am Ort der Sektion zugrundegelegt. Die entsprechende charakteristische Tonheit z_ν ergibt sich aus

$$z_\nu = (\nu - 1) \Delta z. \quad (\text{A.41})$$

Durch die Korrespondenz zwischen der charakteristischen Frequenz nach (A.38) bis (A.40) und der charakteristischen Tonheit nach (A.41) wird die in dieser Arbeit gültige Umrechnung zwischen einer Frequenz in Hz und einer Tonheit in Bark definiert .

Die Dimensionierung des elektrischen Netzwerks des Modells der Hydromechanik gemäß Bild 3.4 geht aus den folgenden Bestimmungsgleichungen für alle Sektionen $\nu \in [1, N_{\text{Sek}}]$ hervor. Dabei wird die Güte des Parallelschwingkreises Q_ν verwendet.

$$C_{q,\nu} = 12,5 \text{ nF} \frac{\Delta z}{0,1 \text{ Bark}} \quad (\text{A.42})$$

$$R_{q,\nu} = 1,0 \text{ M}\Omega \frac{0,1 \text{ Bark}}{\Delta z} \quad (\text{A.43})$$

$$C_\nu = \left(255 \text{ nF} + 335 \text{ nF} \frac{z_\nu}{24 \text{ Bark}} \right) \frac{0,1 \text{ Bark}}{\Delta z} \quad (\text{A.44})$$

$$L_\nu = \frac{1}{(2\pi f_{\text{Res},\nu})^2 C_\nu} \quad (\text{A.45})$$

$$Q_\nu = \begin{cases} 1 + 0,5 z_\nu ; & 0 < z_\nu \leq 4 \\ 2 + 0,25 z_\nu ; & 4 < z_\nu \leq 8 \\ 3 + 0,125 z_\nu ; & 8 < z_\nu \leq 16 \\ 5 ; & 16 < z_\nu \end{cases} \quad (\text{A.46})$$

$$R_\nu = Q_\nu \sqrt{\frac{L_\nu}{C_\nu}} \quad (\text{A.47})$$

Den Abschluß des Helicotremas bilden:

$$R_{\text{H}} = 281,1 \text{ k}\Omega \quad (\text{A.48})$$

$$C_{\text{H}} = 56,6 \text{ nF} . \quad (\text{A.49})$$

Die Nachbildung der ÄHZs gemäß Bild 3.6 wird mit folgenden Bestimmungsgleichungen dimensioniert:

$$R_{\text{K},\nu} = 237,29 \text{ k} R_\nu \quad (\text{A.50})$$

$$\text{wobei} \quad k = \begin{cases} 1,72 - 0,16 z_\nu ; & z_\nu < 4 \\ 1 ; & 4 < z_\nu \leq 10 \\ 0,77 + 0,023 z_\nu ; & 10 < z_\nu \leq 17 \\ 0,09 + 0,063 z_\nu ; & 17 < z_\nu \leq 20 \\ -1,11 + 0,123 z_\nu ; & 20 < z_\nu \end{cases} \quad (\text{A.51})$$

Die Werte der elektrischen Elemente sind in Bild 3.5 in Abhängigkeit von der Sektionsnummer grafisch dargestellt.

Die Werte der lateralen Kopplungswiderstände jeder Sektion werden mit Hilfe der in Bild 3.9 dargestellten Gewichtungsfunktion als Vielfache des Kopplungswiderstandes $R_{K,\nu}$ der entsprechenden Sektion angegeben. Der Gewichtungsfunktion gemäß (A.52) liegt der Kehrwert einer Gaußfunktion zugrunde. Die Anordnung der Widerstände zur lateralen Verkopplung der ÄHZ-Modelle ist in Bild A.10 für die Sektion ν und ihre unmittelbaren Nachbarsektionen vollständig dargestellt. Laterale Verkopplungswiderstände, die mit Sektionen außerhalb des Bereichs der Sektionsnummern 1 bis N_{Sek} liegen, entfallen. Die dadurch verursachte Impedanzänderung im Randbereich wird durch die im folgenden abgeleiteten Bestimmungsgleichungen für Spannungen und Ströme geeignet berücksichtigt.

$$R_{\text{la},\nu,\mu} = R_{\text{lb},\nu,\mu} = R_{K,\nu} \exp\left(3,125 \pi (\mu \Delta z)^2\right) \quad (\text{A.52})$$

mit $\mu \in [1, 16]$

Der Verstärkungsfaktor der gesteuerten Spannungsquelle beträgt in jeder Sektion $g_1 = 100$. Die Sättigungskennlinie innerhalb der Rückkopplungsschleife ist in jeder Sektion identisch. Der in Bild 3.7 dargestellte Zusammenhang zwischen Eingangsspannung $u_{1,\nu} = g_1 u_{\text{BM},\nu}$ und Ausgangsspannung $u_{\text{Sät},\nu}$ lautet:

$$u_{\text{Sät},\nu} = \frac{u_{1,\nu}}{\sqrt{1 + 20 |u_{1,\nu}| + 10 u_{1,\nu}^2}} \quad (\text{A.53})$$

Der Verstärkungsfaktor der Stromquelle in jeder Sektion beträgt $g_2 = 20$. Der angekoppelte Parallelschwingkreis besitzt die gleiche Güte wie der Schwingkreis des Modells der Hydromechanik. Die Resonanzfrequenz ist aber demgegenüber um den Faktor 0,9 verstimmt:

$$C_{E,\nu} = C_\nu \quad (\text{A.54})$$

$$L_{E,\nu} = \frac{1}{(2\pi \cdot 0,9 f_{\text{Res},\nu})^2 C_{E,\nu}} \quad (\text{A.55})$$

$$R_{E,\nu} = Q_\nu \sqrt{\frac{L_{E,\nu}}{C_{E,\nu}}} \quad (\text{A.56})$$

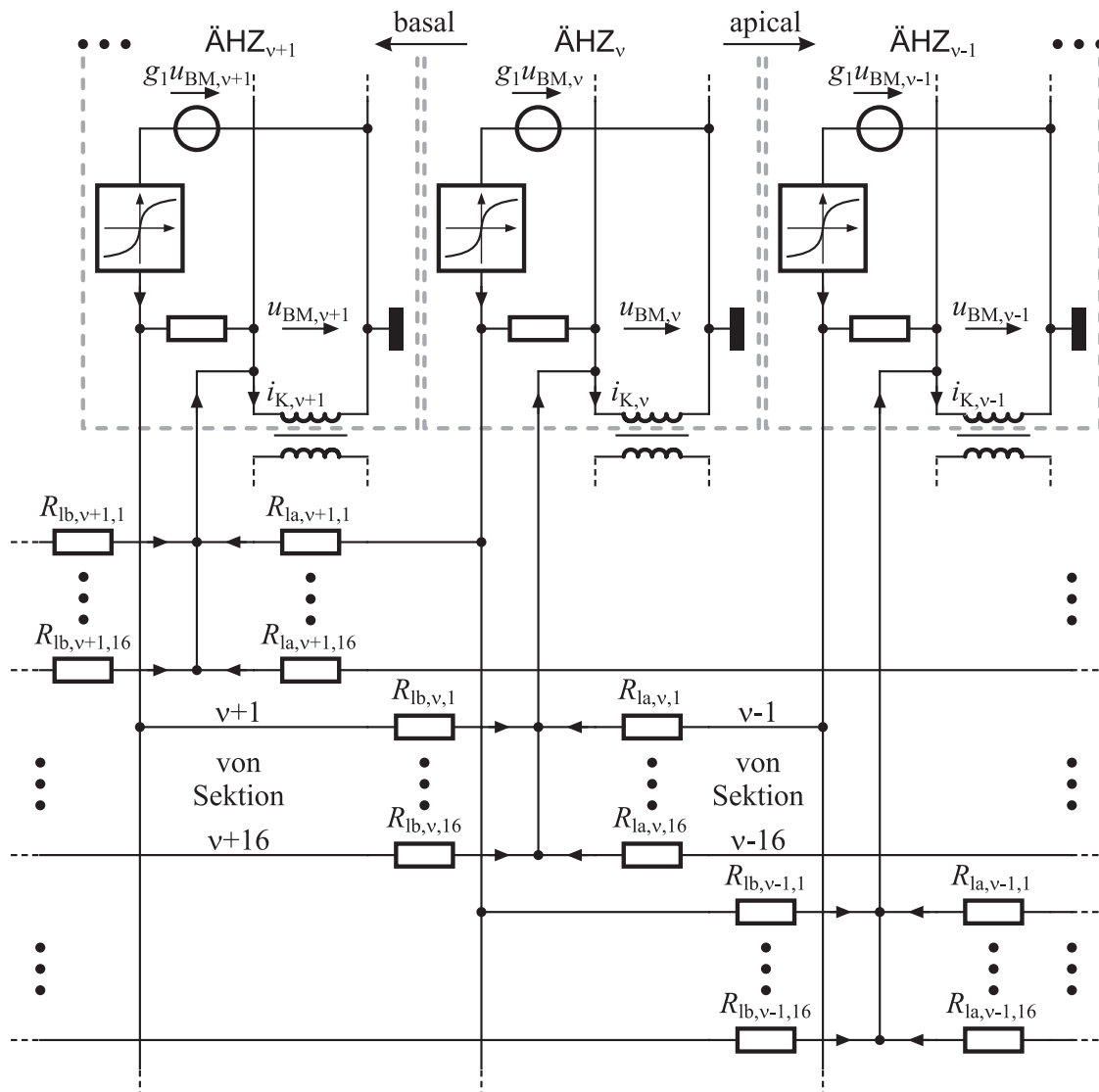


Bild A.10 Laterale Kopplung der ÄHZ-Modelle. Jedes ÄHZ-Modell ist über jeweils einen Widerstand mit den 16 benachbarten Sektionen in apicaler Richtung und 16 benachbarten Sektionen in basaler Richtung gekoppelt.

Die unmittelbare Anwendung von WDFs zur Simulation des elektrischen Kettenleiters mit den nichtlinearen Rückkopplungskreisen aus Bild 3.6 hat gravierende Konsistenzprobleme zur Folge. Diese entstehen aufgrund der nichtlinearen geschlossenen Rückkopplungskreise in den ÄHZ-Modellen. Die zwei wesentlichen Probleme bestehen darin, daß einerseits die geschlossenen Rückkopplungsschleifen nicht realisiert werden können und andererseits eine nichtlinear gesteuerte Spannungsquelle zu zeitvarianten Torwiderständen führt, die in der Theorie der WDF aber als konstant vorausgesetzt werden. Ein möglicher Ausweg wird in [153] angegeben und in dieser Arbeit übernommen. Durch Einfügen eines Verzögerungselements in jede Rückkopplungsschleife wird die Simulation als WDF ermöglicht, deren Approximationsverhalten dann jedoch von der Abtastfrequenz abhängt. Zur Verringerung

der zeitvarianten Impedanzen wird in [153] ein zusätzlicher Zirkulator in die Rückkopplungskreise integriert, der die Auswirkung der nichtlinear gesteuerten Spannungsquelle auf die angeschlossenen Torimpedanzen vermindert. Die Beschaltung des Zirkulators geht aus Bild A.11 hervor. Bei hinreichend kleinem Zirkulatorinnenwiderstand entspricht dem Zirkulator die Parallelschaltung der an den Toren angeschlossenen Netzwerke.

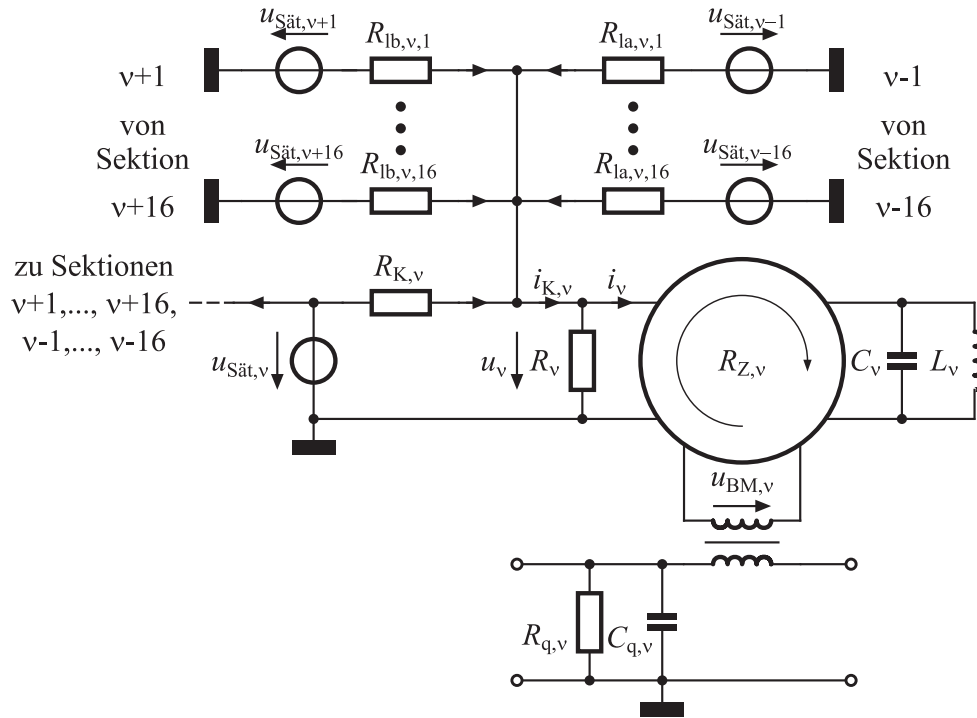


Bild A.11 Einbindung des Zirkulators in jede Sektion des Modells der Cochlea. Bei vernachlässigbarem Innenwiderstand entspricht dem Zirkulator die Parallelschaltung der an den Toren angeschlossenen Netzwerke.

Der in jeder Rückkopplungsschleife auftretende Strom $i_{K,v}$ berechnet sich aus den Spannungen und Widerständen nach (A.57) mit den Abkürzungen aus (A.58).

$$i_{K,v} = \frac{u_{Sät,v} - u_v}{R_{K,v}} + \sum_{\mu=1}^{Ma} \frac{u_{Sät,v-\mu} - u_v}{R_{la,v,\mu}} + \sum_{\mu=1}^{Mb} \frac{u_{Sät,v+\mu} - u_v}{R_{lb,v,\mu}} \quad (\text{A.57})$$

$$\text{mit} \quad Ma = \text{Min}(v - 1, 16) \quad \text{und} \quad Mb = \text{Min}(N_{\text{Sek}} - v, 16) \quad (\text{A.58})$$

Für den Strom i_v am Zirkulatoreingang gilt:

$$i_v = i_{K,v} - \frac{u_v}{R_v}. \quad (\text{A.59})$$

Durch Einsetzen von (A.57) in (A.59) kann der Strom $i_{K,v}$ eliminiert werden, so daß i_v in Abhängigkeit von der momentanen Quellenspannungen sowie der Kopplungswiderstände gemäß (A.60) angegeben werden kann.

$$i_v = \frac{u_{\text{Sät},v}}{R_{K,v}} + \sum_{\mu=1}^{Ma} \frac{u_{\text{Sät},v-\mu}}{R_{\text{la},v,\mu}} + \sum_{\mu=1}^{Mb} \frac{u_{\text{Sät},v+\mu}}{R_{\text{lb},v,\mu}} -$$

$$- u_v \left[\frac{1}{R_v} + \frac{1}{R_{K,v}} + \sum_{\mu=1}^{Ma} \frac{1}{R_{\text{la},v,\mu}} + \sum_{\mu=1}^{Mb} \frac{1}{R_{\text{lb},v,\mu}} \right] \quad (\text{A.60})$$

Mit der Beziehung

$$u_v = b7_v + i_v R_{Z,v} \quad (\text{A.61})$$

und (A.60) folgt die Abhängigkeit der Spannung u_v von den Ausgangsspannungen der Verstärker mit nichtlinearer Kennlinie:

$$u_v = \frac{b7 + R_{Z,v} \left[\frac{u_{\text{Sät},v}}{R_{K,v}} + \sum_{\mu=1}^{Ma} \frac{u_{\text{Sät},v-\mu}}{R_{\text{la},v,\mu}} + \sum_{\mu=1}^{Mb} \frac{u_{\text{Sät},v+\mu}}{R_{\text{lb},v,\mu}} \right]}{1 + R_{Z,v} \left[\frac{1}{R_v} + \frac{1}{R_{K,v}} + \sum_{\mu=1}^{Ma} \frac{1}{R_{\text{la},v,\mu}} + \sum_{\mu=1}^{Mb} \frac{1}{R_{\text{lb},v,\mu}} \right]} \quad (\text{A.62})$$

Die in (A.62) auftretenden Spannungen $u_{\text{Sät},v}$, $u_{\text{Sät},v-\mu}$, $u_{\text{Sät},v+\mu}$ hängen gemäß (A.53) mit der momentanen Spannung u_{BM} und dementsprechend mit der Spannung u_v zusammen. (A.62) ist eine Gleichung mit insgesamt $Ma + Mb + 1$ Unbekannten, die für jede Sektion gelöst werden muß. Dazu könnte beispielsweise das Newton-Raphson-Verfahren verwendet werden. Simulationen bei Abtastfrequenzen im Bereich von 88 kHz bis 100 kHz zeigen jedoch, daß eine einfache lineare Approximation bereits stabile und ausreichend genaue Ergebnisse liefert [153]. Diese lineare Approximation verwendet die um ein Abtastintervall verzögerten Spannungen $u_{\text{Sät},v}^*$, $u_{\text{Sät},v-\mu}^*$, $u_{\text{Sät},v+\mu}^*$ zur Bestimmung von u_v gemäß (A.63), so daß auf die vergleichsweise aufwendige Lösung von Gleichungssystemen verzichtet werden kann.

$$u_v = \frac{b7 + R_{Z,v} \left[\frac{u_{\text{Sät},v}^*}{R_{K,v}} + \sum_{\mu=1}^{Ma} \frac{u_{\text{Sät},v-\mu}^*}{R_{\text{la},v,\mu}} + \sum_{\mu=1}^{Mb} \frac{u_{\text{Sät},v+\mu}^*}{R_{\text{lb},v,\mu}} \right]}{1 + R_{Z,v} \left[\frac{1}{R_v} + \frac{1}{R_{K,v}} + \sum_{\mu=1}^{Ma} \frac{1}{R_{\text{la},v,\mu}} + \sum_{\mu=1}^{Mb} \frac{1}{R_{\text{lb},v,\mu}} \right]} \quad (\text{A.63})$$

In der an einem Eingang des Zirkulators wirkenden Spannung u_v sind die vom Modell der ÄHZ verursachte Spannungsverstärkung sowie die Beiträge der benachbarten Sektionen zusammengefaßt. Diese Spannung wird daher in der WDF-Nachbildung wie eine Spannungsquelle mit der Ausgangsspannung u_v in jeder Sektion behandelt. Das Wellenflußdiagramm einer Sektion des HM-Modells mit ÄHZ-Modell ist in Bild A.12 dargestellt. Diese Nachbildung beruht auf dem in Bild A.11 angegebenen Netzwerk mit eingefügtem Zirkulator.

$$\begin{aligned}
g4_\nu &= \frac{R_B}{R_{q,\nu}} \\
g5_\nu &= \frac{1}{r1_\nu} + g2_\nu + g4_\nu \\
r8_\nu &= \frac{rL_\nu rC_\nu}{rL_\nu + rC_\nu} \\
r9_\nu &= \frac{r8_\nu rL_\nu}{rC_\nu}
\end{aligned}
\left. \vphantom{\begin{aligned} g4_\nu \\ g5_\nu \\ r8_\nu \\ r9_\nu \end{aligned}} \right\} \text{ mit }
\begin{aligned}
rL_\nu &= \frac{\omega_{N,\nu} L_\nu}{R_B} \\
rC_\nu &= \frac{1}{\omega_{N,\nu} C_\nu R_B}
\end{aligned}$$

$$\begin{aligned}
r1_{\nu-1} &= \frac{1}{g5_\nu} + r8_\nu \\
ge2_\nu &= \omega_{N,\nu} C_{E,\nu} R_B \\
ge3_\nu &= \frac{R_B}{\omega_{N,\nu} L_{E,\nu}} \\
ge1_\nu &= g2_\nu R_B \\
rh1 &= r1_0 \\
gh2 &= \omega_{N,1} C_H R_B \\
gh3 &= \frac{R_B}{R_H}
\end{aligned}$$

Der entnormierte Zirkulatorinnenwiderstand $R_{Z,\nu}$ in Bild A.11 wird bestimmt durch:

$$R_{Z,\nu} = r8_\nu R_B$$

Aus den Torwiderständen und -leitwerten sind die WDF-Koeffizienten wie folgt zu bestimmen:

$$\begin{aligned}
\gamma1_\nu &= \frac{1}{r1_\nu g5_\nu} \\
\gamma4_\nu &= \frac{g4_\nu}{g5_\nu} \\
\gamma7_\nu &= \frac{r8_\nu}{r1_{\nu-1}} \\
\gamma8_\nu &= \frac{r8_\nu - r9_\nu}{r8_\nu + r9_\nu} \\
\gamma e1_\nu &= 2 \frac{ge1_\nu}{ge1_\nu + ge2_\nu + ge3_\nu} \\
\gamma e2_\nu &= 2 \frac{ge2_\nu}{ge1_\nu + ge2_\nu + ge3_\nu} \\
\gamma h1 &= 2 \frac{1}{1 + gh2 rh1 + gh3 rh1}
\end{aligned}$$

$$\gamma h3 = 2 \left(1 + \frac{1}{rh1 gh3} + \frac{gh2}{gh3} \right)^{-1}$$

Der gesamte WDF-Algorithmus der Modelle vom Außen- und Mittelohr sowie der Hydromechanik mit den äußeren Haarzellen ist durch die folgenden Gleichungen gegeben. Zunächst werden die in Ausbreitungsrichtung laufenden (hinlaufenden) Wellen bestimmt. Die mit * bezeichneten Wellengrößen sind um ein Abtastintervall gegenüber dem aktuellen Abtastintervall verzögert.

$$av1 = 600 u_{AO}$$

$$av3 = av1 - \gamma v4 av1 - \gamma v2 (av1 + bv2^*)$$

$$av8 = bv6^* - bv5^*$$

$$av10 = av8 - \gamma v9 av8$$

$$b6_{N_{Sek}} = b6k = -av10 - av3$$

Die folgenden Gleichungen sind rekursiv von $\nu = N_{Sek}, \dots, 1$ anzuwenden:

$$a5_\nu = -(1 - \gamma 1_\nu - \gamma 4_\nu) b2_\nu^* - \gamma 1_\nu b6_\nu$$

$$a7_\nu = -b9_\nu^* - \gamma 8_\nu (a8_\nu^* + b9_\nu^*)$$

$$b9_\nu = a7_\nu + a8_\nu^* + b9_\nu^*$$

$$b6_{\nu-1} = -a5_\nu - a7_\nu$$

Die Wellengrößen am Helicotrema sind bestimmt durch:

$$bh2 = (1 - \gamma h3) bh2^* - \gamma h1 (bh2^* - b6_0)$$

Die rücklaufenden Wellen ergeben sich aus:

$$a6_0 = bh1 = bh2 + bh2^* - b6_0$$

Die folgenden Gleichungen sind rekursiv von $\nu = 1, \dots, N_{Sek}$ anzuwenden:

$$b7_\nu = a7_\nu - \gamma 7_\nu (a6_{\nu-1} - b6_{\nu-1})$$

$$a8_\nu = 2 u_\nu - b7_\nu \quad \text{mit } u_\nu \text{ aus (A.63)}$$

$$b2_\nu = a6_{\nu-1} + b7_\nu - a5_\nu - b2_\nu^*$$

$$a6_\nu = b2_\nu + b2_\nu^* - b6_\nu$$

Die Wellengrößen der zweiten Verstärkerstufe sind bestimmt durch:

$$ae1_\nu = g_2 i_{K,\nu} R_{E,\nu} \quad \text{mit } i_{K,\nu} \text{ aus (A.57)}$$

$$be3_\nu = \gamma e1_\nu (be3_\nu^* + ae1_\nu) + \gamma e2_\nu (be3_\nu^* + be2_\nu^*) - be3_\nu^*$$

$$be2_\nu = be3_\nu - be3_\nu^* - be2_\nu^*$$

$$u_{IHZ,\nu} = u_\nu + 0,5 (be2_\nu + be2_\nu^*)$$

Die rücklaufenden Wellen des Mittel- und Außenohrmodells ergeben sich aus:

$$bv10 = av10 - \gamma v10 (a6_{N_{Sek}} - b6_{N_{Sek}})$$

$$bv3 = -bv10 - a6_{N_{Sek}}$$

$$bv8 = bv10 - \gamma v9 av8$$

$$bv5 = bv5^* - \gamma v5 (bv8 - av8)$$

$$bv6 = -bv6^* - \gamma v6 (bv8 - av8)$$

$$bv1 = bv3 - \gamma v4 av1 - \gamma v2 (av1 + bv2^*)$$

$$bv2 = bv1 + av1 + bv2^*$$

Die Wahl einer geeigneten Abtastfrequenz f_A für die Implementierung der WDFs hängt von unterschiedlichen Faktoren ab. Die Nyquistfrequenz, die der doppelten Signalbandbreite entspricht, ist nicht ausreichend, da aufgrund der Nichtlinearität innerhalb des Cochlea-Modells Signalkomponenten bei Vielfachen der Signalfrequenz entstehen. Diese Komponenten können bei zu geringer Abtastfrequenz zu Aliasing-Artefakten führen. Bei gegebener Bandbreite des Eingangssignals kann die Frequenzverzerrung aufgrund der Bilineartransformation durch eine Erhöhung der Abtastfrequenz vermindert werden. Für die vorliegenden Ergebnisse der Wahrnehmungsschwellennachbildung wurde eine Abtastfrequenz von $f_A = 100$ kHz gewählt. Für die Codersteuerung wurde als Abtastfrequenz die doppelte Abtastfrequenz des Audiosignals $f_A = 2 \cdot 44,1$ kHz verwendet.

Die Verarbeitung eines Audiosignals auf einem Standard-PC mit 400-MHz Prozessor mit dem gesamten Gehörmodell ist um den Faktor 40 langsamer als Echtzeit. Die Rechenzeit vervielfacht sich darüber hinaus entsprechend der Iterationsanzahl bei der Wahrnehmungsschwellenberechnung.

A.8 Implementierung des Modells der inneren Haarzelle

Das Modell der inneren Haarzelle ist in Bild 3.10 dargestellt. Es wird durch die Zeitkonstante $\tau_{\text{IHZ},\nu}$ bestimmt, die in Tab. A.2 angegeben ist.

| Anwendung | $\tau_{\text{IHZ},\nu}$ |
|----------------------|--------------------------------|
| Schwellennachbildung | 2 ms |
| Codersteuerung | $10/(2\pi f_{\text{Res},\nu})$ |

Tab. A.2 Zeitkonstanten der Tiefpaßfilter im Fluktuationsdetektor bei Anwendung des Gehörmodells zur Wahrnehmungsschwellennachbildung bzw. Codersteuerung.

Der Tiefpaß wird entsprechend der Angaben in Anhang A.4 als Digitalfilter implementiert. Die Quadrierung am Eingang des IHZ-Modells führt bei Einspeisung einer Sinusschwingung zu einem Gleichanteil und einer Spektralkomponente bei der doppelten Frequenz gemäß (A.64). Durch die Wahl der Zeitkonstanten liegt die Eckfrequenz des Tiefpasses immer deutlich unterhalb der doppelten Resonanzfrequenz, bei der die größte Signalleistung am Ausgang der entsprechenden Sektion des HM-Modells erwartet wird. Die Wirkung des IHZ-Modells ist daher mit einem Hüllkurvendetektor zur Demodulation von AM-Signalen

in der Rundfunktechnik vergleichbar, der jedoch anstelle der Quadrierung im IHZ-Modell eine Betragsbildung verwendet.

$$(A \sin(\omega_0 t))^2 = \frac{A^2}{2}(1 + \cos(2\omega_0 t)) \quad (\text{A.64})$$

A.9 Parameter des Modells der neuronalen Verarbeitung

A.9.1 Parameter des internen Rauschens

Das interne Rauschen bestimmt die vom Gehörmodell nachgebildeten Ruhehörschwellen. Der Schalldruckpegel bei der Wiedergabe eines decodierten Audiosignals ist im Gegensatz zu den Bedingungen einer psychoakustischen Messung in der Regel nicht festgelegt. Daher kann die Wiedergabe bei einem höheren Pegel dazu führen, daß Verzerrungen wahrnehmbar werden, die sonst unterhalb der Ruhehörschwelle liegen. Dies wird durch eine Absenkung der Ruhehörschwelle im Gehörmodell bei der Anwendung zur Codersteuerung vermieden. Die Maximalwerte des gleichverteilten Rauschens sind in (A.65) für die Wahrnehmungsschwellennachbildung angegeben.

$$u_{\text{NMAX},\nu} = 1,5 \text{ mV} \quad (\text{A.65})$$

Für die Codersteuerung wird der Maximalwert gemäß (A.66) abgesenkt und entsprechend Bild A.13 in Abhängigkeit von der Sektionsnummer eingestellt.

$$u_{\text{NMAX},\nu} = 0,1 \cdot 10^{g(\nu)/20} \text{ mV} \quad (\text{A.66})$$

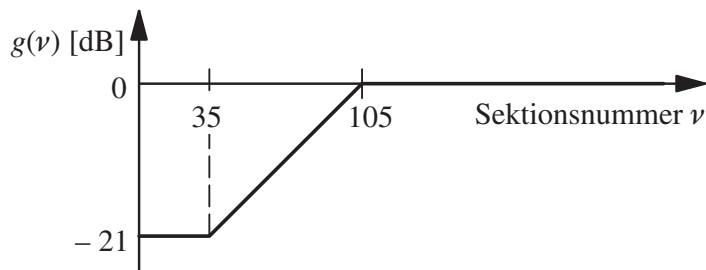


Bild A.13 Abhängigkeit der maximalen Rauschamplitude von der Sektionsnummer bei Anwendung des Gehörmodells zur Codersteuerung.

A.9.2 Parameter der Spreizfunktion

Die in Kapitel 3.4.1 eingeführte Spreizfunktion zur Nachbildung der Vorverdeckung wird im Gehörmodell in Form einer Folge zeitdiskreter Abtastwerte $Sp(n)$ nach (A.67) implementiert. Die Länge der Folge wird auf den Bereich von 0 bis $2N$ beschränkt, innerhalb dessen die Abtastwerte der Gaußfunktion größer als -40 dB bezogen auf den Maximalwert sind. Die Summe der Abtastwerte wird mittels der Konstanten S nach (A.68) auf eins normiert. Das Ausmaß der zeitlichen Spreizung ist durch die in (A.69) gegebene Zeitkonstante τ_{Sp} festgelegt.

$$Sp(n) = \frac{1}{S} \exp\left[-\frac{(n-N)^2}{2\sigma_{Sp}^2}\right]; \quad n \in [0, 2N] \quad (\text{A.67})$$

$$\text{mit} \quad S = \sum_{n=0}^{2N} \exp\left[-\frac{(n-N)^2}{2\sigma_{Sp}^2}\right] \quad (\text{A.68})$$

$$\text{und} \quad \sigma_{Sp} = \tau_{Sp} \frac{f_A}{\sqrt{\pi}}; \quad \tau_{Sp} = 5 \text{ ms}; \quad N = 1 + 2\sigma_{Sp} \sqrt{\ln(10)}. \quad (\text{A.69})$$

Die Faltung mit dem Eingangssignal erfolgt gemäß (A.70).

$$u_{Sp,\nu}(n) = \sum_{k=0}^{2N} u_{HN,\nu}(n-k) Sp(k) \quad (\text{A.70})$$

A.9.3 Parameter der Fluktuationsdetektoren

Für die in Kapitel 3.4.5 angegebenen Fluktuationsdetektoren werden zwei unterschiedliche Parametersätze angegeben, die auf die jeweilige Anwendung des Gehörmodells zur Wahrnehmungsschwellennachbildung bzw. Codersteuerung hin optimiert sind. Die unterschiedlichen Parameter sind darauf zurückzuführen, daß bei der Audiocodierung eine Detektionswahrscheinlichkeit von 50% gemäß der Definition einer Wahrnehmungsschwelle für eine der CD vergleichbare Klangqualität nicht ausreichend ist. Die Zeitkonstanten wurden daher entsprechend angepaßt und die verwendeten Schwellwerte für die Schwellendetektoren wurden für Codersteuerung geeignet verringert.

Die verwendeten Zeitkonstanten der Tiefpaßfilter sind in Tabelle A.3 und (A.71) zusammengefaßt.

| Anwendung | $\tau_{\text{DMIN},\nu}$ | $\tau_{\text{FTP},\nu}$ |
|----------------------|--|-------------------------|
| Schwellennachbildung | 50 ms | 20 ms |
| Codersteuerung | $5 \sqrt{0,001 \text{ Max}(f_{\text{Res},\nu}/\text{Hz}, 100)}$ ms | 5 ms |

Tab. A.3 Zeitkonstanten der Tiefpaßfilter im Fluktuationsdetektor bei Anwendung des Gehörmodells zur Wahrnehmungsschwellennachbildung bzw. Codersteuerung.

$$\tau_{\text{DMAX},\nu} = \tau_{\text{DMIN},\nu} \quad (\text{A.71})$$

Bei der Berechnung von $\tau_{\text{DMIN},\nu}$ geht die Resonanzfrequenz $f_{\text{Res},\nu}$ gemäß (A.40) ein. Die verwendeten Abbildungsparameter zur Steuerung der Schwellwerte $r_{S,\nu}$ in Abhängigkeit der vom Fluktuationsdetektor gemessenen Fluktuation fl_{ν} in Bild 3.20 sind in den Gleichungen (A.72) bis (A.74) und Tabelle A.4 angegeben.

$$fl_{N,\nu} = 0,5 + 0,25 \frac{\nu - 1}{N_{\text{Sek}}} \quad (\text{A.72})$$

$$fl_T = 0,2 \quad (\text{A.73})$$

$$r_{N,\nu} = r_{N,LF} + (r_{N,HF} - r_{N,LF}) \frac{\nu - 1}{N_{\text{Sek}}} \quad (\text{A.74})$$

| Anwendung | r_T | $r_{N,LF}$ | $r_{N,HF}$ |
|---------------------|-------|------------|------------|
| Schwelennachbildung | 1,045 | 1,4 | 1,4 |
| Codersteuerung | 1,03 | 1,15 | 1,1 |

Tab. A.4 Parameter für die Abbildung der gemessenen Fluktuation auf den verwendeten Schwellwert bei Anwendung des Gehörmodells zur Wahrnehmungsschwellennachbildung bzw. Codersteuerung.

Die Einhüllendenfluktuation wird für die Adaption des vom Schwellendetektor verwendeten Schwellwerts in jeder Sektion bestimmt, um die daraus resultierenden unterschiedlichen Verdeckungseigenschaften nachzubilden. Die Abbildungsfunktion der gemessenen Fluktuation fl_ν auf den Schwellwert $r_{S,\nu}$ aus Bild 3.20 ist in (A.75) analytisch angegeben.

$$r_{S,\nu} = r_T + \frac{1}{2}(r_{N,\nu} - r_T) \left(1 - \cos \left(\pi \frac{fl_{\text{lim},\nu} - fl_T}{fl_{N,\nu} - fl_T} \right) \right) \quad (\text{A.75})$$

Die in (A.75) verwendete Fluktuation $fl_{\text{lim},\nu}$ wird durch Begrenzung der gemessenen Fluktuation fl_ν auf den Bereich von fl_T bis $fl_{N,\nu}$ bestimmt:

$$fl_{\text{lim},\nu} = \begin{cases} fl_{N,\nu} & fl_\nu \geq fl_{N,\nu} \\ fl_\nu & \text{für } fl_{T,\nu} < fl_\nu < fl_{N,\nu} \\ fl_{T,\nu} & fl_\nu \leq fl_{T,\nu} \end{cases} \quad (\text{A.76})$$

A.10 Psychoakustisches Meßverfahren zur Bestimmung der Wahrnehmungsschwelle

Für die eigenen psychoakustischen Messungen der Ruhehörschwellen und Mithörschwellen wurde ein modifiziertes Ja-Nein-Verfahren verwendet. Bei der Mithörschwellenmessung wird dazu jeder Testperson in bekannter Reihenfolge das Maskierersignal und nach einer 200 ms langen Pause das Maskierersignal mit überlagertem Testsignal dargeboten. Innerhalb eines Versuchs wird diese Sequenz noch einmal wiederholt, bevor die Testperson entscheidet, ob sie einen Unterschied zwischen den beiden Signalen wahrgenommen hat.

In Abhängigkeit von der abgegebenen Antwort wird der Testsignalpegel für den nächsten Versuch eingestellt. Für die Einstellung wird eine Schrittweite von 16 dB zugrundege-

legt. Falls ein Unterschied wahrgenommen wurde, wird der Testsignalpegel zufallsgesteuert entweder um die doppelte oder die einfache Schrittweite verringert, oder nicht verändert. Andernfalls, wenn kein Unterschied wahrgenommen wurde, wird der Testsignalpegel zufallsgesteuert entweder um die doppelte oder einfache Schrittweite erhöht, oder er bleibt unverändert. Die Schrittweite in dB wird im Verlauf der Messungen halbiert, wenn die Testperson mehrmals konsistent mit „Ja“ bzw. „Nein“ bei konstanter Schrittweite geantwortet hat. Wenn eine Schrittweite von 2 dB erreicht ist, werden 6 Antworten für die Auswertung übernommen. Die Mithörschwelle wird daraus durch Interpolation des 50%-Wahrscheinlichkeitswertes der psychometrischen Funktion berechnet.

Die Messung der Ruhehörschwellen erfolgt nach der gleichen Methode. Da hier jedoch kein Maskierersignal verwendet wird, beinhaltet ein Versuch nur die Darbietung des Testsignals mit einer Wiederholung.

Die psychoakustischen Messungen wurden mit Kopfhörern (Sennheiser HDA 200) durchgeführt. Die Darbietung der verwendeten Audiosignale erfolgte monaural. Die Audiosignale wurden von einem Rechner erzeugt, bzw. von der Festplatte gelesen und ausgegeben. Die verwendeten Rauschsignale waren in jedem Versuch einer Wahrnehmungsschwellenmessung abgesehen vom Verstärkungsfaktor identisch.

Lebenslauf

Frank Baumgarte

13. April 1964 geboren in Bad Gandersheim
als Sohn des Tischlermeisters Adolf Baumgarte
und seiner Ehefrau Ute Baumgarte, geb. Burmeister,
wohnhaft in Winzenburg
- 1971 – 1975 Besuch der Grundschule Everode /Winzenburg
- 1975 – 5/83 Besuch des Gymnasiums Alfeld /L.
Abschluß: Abitur
- 8/83 – 9/83 Grundpraktikum bei der C. Behrens AG, Alfeld /L.
- 10/83 – 12/89 Studium der Elektrotechnik an der Universität Hannover
Schwerpunkt: Übertragungstechnik
Abschluß: Diplom
- ab 1/90 selbständige Tätigkeit: Entwicklung von Hardware und
Software für die Signalverarbeitung
- 4/90 – 6/91 Ableistung des zivilen Ersatzdienstes
- 3/92 – 9/99 wissenschaftlicher Mitarbeiter am Institut für Theoretische
Nachrichtentechnik und Informationsverarbeitung
der Universität Hannover
- ab 10/99 Consultant bei Lucent Technologies Bell Laboratories, New Jersey