

Der Zwicker-Ton —
Ein Phänomen der auditorischen
Informationsverarbeitung

Diplomarbeit
von
Jan-Moritz Peter Franosch

Der Zwicker-Ton —
Ein Phänomen der auditorischen
Informationsverarbeitung

Diplomarbeit
von
Jan-Moritz Peter Franosch

Physik-Department der Technischen Universität München
Institut für Theoretische Physik
Prof. Dr. J. Leo van Hemmen
7. Juli 1998

Vorwort

Diese Arbeit verfolgt folgende Ziele:

- Sie soll eine kritische Einführung in die Psychoakustik geben als Grundlage für folgende Kapitel, eigene Arbeiten und leichteres Verständnis der verfügbaren Literatur (Kap. 1).
- Überblick über das periphere (Kap. 2.1) und zentrale (Kap. 2.2) auditive System.
- Genaue Beschreibung und Implementation eines Gehörmodells von der Cochlea (Kap. 3.1.1) über die inneren Haarzellen (Kap. 3.1.2) bis zum Hörnerv (Kap. 3.1.3), das nicht allzu viel Rechenzeit erfordert und in weiteren Projekten wieder verwendet werden kann.
- Beleuchtung der Auswirkungen lateraler Inhibition in einem einfachen linearen Modell (Kap. 4.2) sowie mittels stochastischer Datenanalyse in einem Modell mit spikenden Neuronen (Kap. 4.3).
- Vorstellung von verschiedenen neuronalen Modellen zur Erklärung des Zwicker-Tons (Kap. 6).

Im Ggs. zum visuellen System [Buser 1992] scheint beim auditorischen System über Sinn und Zweck der zahlreichen Verarbeitungsstufen sowie der Kopplungen und Rückkopplungen zwischen diesen nur relativ wenig Konkretes bekannt zu sein. Es existieren zwar recht genaue psychoakustische „Modelle“ [Zwicker 1990, Sottek 1993], die auditorische Wahrnehmungen quantitativ beschreiben, die neurologische Ursache der beschriebenen Phänomene ist jedoch unklar. Für das visuelle System verfügen wir über Begriffe (Kante, Orientierung, Bewegung, Farbe, stereoskopische Tiefe), die sowohl zur Beschreibung von Wahrnehmungen dienen können als auch zur Beschreibung der Tätigkeit oder des Antwortverhaltens verschieden spezialisierter Neurone. Auditorische Wahrnehmungen werden durch Begriffe wie Lautstärke, Tonhöhe, Rauschen, Tonhöhenmodulation und Klang beschrieben. Zur Charakterisierung der Aufgaben von auditorischen Neuronen sind diese jedoch nur teilweise geeignet. Auf Fragen wie „Was ist die neuronale Ursache dieser und jener auditorischen Wahrnehmung?“ oder „Wie wird diese und jene auditorische Wahrnehmung neuronal repräsentiert?“ wissen wir oft keine treffende Antwort.

Vorliegende Arbeit soll solche Fragen angehen, indem über neuronale Verarbeitungsmechanismen nachgedacht wird, die mit den bekannten Beobachtungen verträglich sind und das psychoakustische Phänomen des Zwicker-Tons erklären. Der Zwicker-Ton kann folgendermaßen hervorgerufen werden: Tiefpassgefiltertes weißes Rauschen wird abgeschaltet. Nach dem Abschalten hört die Versuchsperson einen leisen, fast reinen Ton, den Zwicker-Ton. Da der Zwicker-Ton nach dem Abschalten eines akustischen Stimulus auftritt, bezeichnet man ihn auch als akustischen Nacheffekt. Man kann ihn als eine Art akustische Täuschung auffassen, da der Zwicker-Ton eine Schallwahrnehmung ist, die in „Wirklichkeit“ nicht vorhanden ist. Da optische Täuschungen oft Aufschluss über (neuronale) Verarbeitungsmechanismen der optischen Signale geben, kann man hoffen, dass auch akustische Täuschungen wertvolle Hinweise über die Verarbeitung akustischer Stimuli liefern. Jedenfalls stellt sich im Laufe der Analyse des Phänomens des Zwicker-Tons in dieser Arbeit heraus, dass es keine allzu einfache Erklärung dieser Wahrnehmungstäuschung gibt. Vermutlich spielen hier mehrere Mechanismen und Eigenschaften der auditorischen Verarbeitung zusammen: Die periphere Verarbeitung von Schallsignalen bildet die Grundlage für jegliches Verständnis psychoakustischer Phänomene. Sie wird deshalb ausführlich behandelt und in Modellen implementiert. In den Modellen zur Erklärung des Zwicker-Tons spielt laterale Inhibition zwischen den Neuronen einer auditorischen Verarbeitungsstufe eine zentrale Rolle. Ihre Auswirkungen im Hinblick auf Zwicker-Ton erzeugende Schalle werden deshalb untersucht. Schließlich werden hierauf basierende Modelle zur Erklärung des Zwicker-Tons vorgestellt.

Zum besseren Verständnis der psychoakustischen Grundlagen dieser Arbeit kann [Zwicker 1982] herangezogen werden. Als weiterführende Literatur zur Psychoakustik sei [Zwicker 1990] empfohlen. Konkrete Formeln und Modelle gibt das sehr gute Büchlein [Sottek 1993] an, das jedoch weitgehende Kenntnisse psychoakustischer Grundlagen voraussetzt. Eine kurze Einführung in die Psychoakustik und eine gut verständliche und übersichtliche Einführung in das auditorische System gibt [Yost 1994], auch [Pickles 1982]. Eine detailreiche Darstellung des auditorischen Systems findet man in [Popper 1992], einen kritischen Überblick in [Zenner 1994]. Psychoakustische Messungen zum Zwicker-Ton sind umfassend in [Krump 1993] beschrieben und bilden die Grundlage jeglicher in dieser Arbeit entwickelten Modelle. Eine ausführliche Beschreibung des verwendeten Spike-Response-Neuronenmodells findet man in [Gerstner 1993, Domany 1994]. Als Grundlage zur Erstellung eigener Simulationen in C++ unter Linux bzw. Unix empfiehlt der Autor die Lektüre von [Stallman XEmacs 1994, Stroustrup 1991, Loukides 1997, Stallman XEmacs 1995, Stallman GDB 1995] in etwa dieser Reihenfolge.

Dieses Dokument wurde mit L^AT_EX 2.0.9 und T_EX Version 3.14159 erstellt [Kopka I 1994, Kopka II 1995, Kopka III 1997, Knuth 1993]. Die Bilder wurden mit Hilfe des T_EX-Macropakets PiCT_EX Version 1.1 bzw. mit XFig 3.1 erstellt. Als Editor dienten Emacs und XEmacs [Stallman XEmacs 1994, Stallman XEmacs 1995].

Folgende Programme wurden zur Programmierung der Simulationen unter dem Betriebssystem Linux bzw. Unix verwendet: GNU C++-Compiler gcc 2.7.2.1 und 2.8.1 [Stallman GCC 1995], GNU Debugger

gdb 4.16 [Stallman GDB 1995] und der Parsergenerator GNU Bison 1.25 [Donnelly 1995]. Programmiersprache ist C++ [Stroustrup 1991, C++ 1996] mit STL-Erweiterungen [Glass 1996].

Die Programmdokumentation wurde mit Texinfo Version 3 [Chassell 1996] erstellt. Sie steht in den Formaten GNU-Info, DVI, Postscript, HTML, ASCII sowie als Texinfo-Quellcode zur Verfügung.

Für analytische Berechnungen und zur Darstellung einiger Graphen war Maple V, Release 4, hilfreich [Maple 1, Maple 2, Maple 3].

Die Rechtschreibung in diesem Dokument richtet sich nach der ab 1. August 1998 eingeführten neuen deutschen Rechtschreibung. Zu Rate gezogen wurde [Duden 1996], sowie die ebenfalls in [Duden 1996] abgedruckte amtliche Neuregelung der deutschen Rechtschreibung.

Inhaltsverzeichnis

1	Psychoakustik	13
1.1	Psychophysik	13
1.2	Lautstärke	14
1.2.1	Schalldruckpegel und Schallintensitätspegel	15
1.2.2	Lautstärkepegel	16
1.2.3	Lautheit	16
1.3	Tonhöhe	17
1.3.1	Verhältnistönhöhe	18
1.3.2	Tonheit	19
1.4	Verdeckung	21
1.5	Funktionsschemata	22
1.5.1	Anregung und Erregung	22
1.5.2	Schwellenfunktionsschema für langsame Schalländerungen	23
1.5.3	Funktionsschema der Lautheit	25
2	Auditorisches System	26
2.1	Peripheres auditorisches System	26
2.1.1	Außenohr	26
2.1.2	Mittelohr	27
2.1.3	Innenohr	27
2.1.4	Hörnerv	29
2.2	Zentrales auditorisches System	31
2.2.1	Nucleus cochlearis	32
2.2.2	Olivenkomplex	33
2.2.3	Nuclei lemnisci laterales	33
2.2.4	Colliculus inferior	34
2.2.5	Corpus geniculatum mediale	34
2.2.6	Auditorischer Cortex	35

3	Modelle für das auditorische System	38
3.1	Modelle für das periphere auditorische System	38
3.1.1	Cochlea	38
3.1.2	Innere Haarzellen	47
3.1.3	Hörnerv	51
3.2	Tinnitusmodell	55
4	Rolle der lateralen Inhibition	56
4.1	Physiologie der lateralen Inhibition	56
4.2	Laterale Inhibition in einem linearen Modell	57
4.2.1	Konstantes Eingabesignal an der Neuronenschicht	58
4.2.2	Stufenfunktion als Eingabesignal der Neuronenschicht	59
4.2.3	Konstantes Eingabesignal mit „Lücke“ an der Neuronenschicht	60
4.3	Laterale Inhibition mit spikenden Neuronen	64
4.3.1	Stochastische Datenanalyse	64
4.3.2	Das verwendete Neuronenmodell und die Verschaltung der Neurone	66
4.3.3	Die in der Simulation verwendeten Stimuli	67
4.3.4	Simulationsergebnisse	68
4.3.5	Zusammenfassung	70
5	Zwicker-Ton	75
5.1	Zwicker-Ton Erzeugerschalle	75
5.1.1	Rauschen	75
5.1.2	Andere Erzeugerschalle	76
5.2	Psychoakustische Messungen	76
5.2.1	Verbreitung	77
5.2.2	Otoakustische Emissionen	77
5.2.3	Bandbreite des Erzeugerschalls	79
5.2.4	Bandbreite der spektralen Lücke	80
5.2.5	Lautstärke des Erzeugerschalls	81
5.2.6	Tiefe der spektralen Lücke	81
5.2.7	Andere Einflüsse	81
5.3	Berechnung der Tonhöhe	81
5.4	Offene Fragen	82
6	Modelle für den Zwicker-Ton	83

6.1	Habituationsmodell	84
6.1.1	Beschreibung des Habituationsmodells	84
6.1.2	Beschreibung der Simulationen des Habituationsmodells	86
6.1.3	Simulationsergebnisse im Habituationsmodell	92
6.1.4	Diskussion der Simulationsergebnisse im Habituationsmodell	96
6.1.5	Erweiterungen des Habituationsmodells	97
6.2	Adaptionsmodell	99
6.2.1	Beschreibung des Adaptionsmodells	99
6.2.2	Beschreibung der Simulationen des Adaptionsmodells .	100
6.2.3	Simulationsergebnisse im Adaptionsmodell	105
6.2.4	Diskussion der Simulationsergebnisse im Adaptionsmodell	108
7	Diskussion und Ausblick	110
7.1	Rückblick	110
7.2	Weitere psychoakustische Experimente	111
7.3	Weitere neurophysiologische Experimente	112
A	Farbtafeln	115
A.1	Habituationsmodell	115
A.2	Adaptionsmodell	117
	Literaturverzeichnis	119
	Dank	123

Abbildungsverzeichnis

1.1	Breite von 1 mel in Hz	18
1.2	Frequenz als Funktion der Verhältnistönhöhe	19
1.3	Bestimmung der Frequenzgruppenbreite	21
1.4	Erregung durch einen reinen Ton	24
2.1	Zentrales auditorisches System	32
3.1	Impulsantwort der linearen cochleären Filter vierter Ordnung .	41
3.2	Vergleich cochleärer Filter	42
3.3	Vergleich cochleärer Filter als Pegel	42
3.4	Vergleich der cochleären Erregung	43
3.5	Vergleich der cochleären Erregung auf einer Barkskala	43
3.6	Cochleogramm von Tiefpassrauschen	45
3.7	Cochleogramm von einem Sinuston	46
3.8	„State-Partition“-Modell	47
3.9	Meddis-Modell	49
3.10	Neurotransmitterkonzentration	51
3.11	Basilmembranauslenkung, Neurotransmitterkonzentration und postsynaptisches Potential	52
3.12	Aktionspotentiale im Hörnerv bei Tiefpassrauschen	53
3.13	Aktionspotentiale im Hörnerv bei einem Sinuston	54
4.1	Membranpotential bei Stufenfunktion	59
4.2	Membranpotential bei Stufenfunktion mit Verschiebung	60
4.3	Verbiegung des Membranpotentials in Abhängigkeit der Verschiebung	61
4.4	Eingabesignal mit Lücke	62
4.5	Membranpotential bei Eingabesignal mit Lücke	62
4.6	Verbiegung in Abhängigkeit der Breite der Lücke	63
4.7	Netz mit lateraler Inhibition	66

4.8	Fehlerrate in Abhängigkeit der feed-forward-Kopplung	69
4.9	Wahrscheinlichkeitsverteilungen mit variierender feed-forward-Kopplung	70
4.10	Fehlerrate in Abhängigkeit der Stärke der lateralen Inhibition	71
4.11	Wahrscheinlichkeitsverteilungen mit variierender lateraler Inhibition	71
4.12	Fehlerrate in Abhängigkeit der lateralen Inhibition	72
4.13	Fehlerrate in Abhängigkeit der feed-forward-Kopplung	72
4.14	Feuerraten der Neuronen bei unterschiedlicher lateraler Inhibition	73
4.15	Aktionspotentiale der Neuronen bei unterschiedlicher lateraler Inhibition	74
5.1	Bandpassbegrenztes Rauschen als Zwicker-Ton Erzeugerschall	79
5.2	Bandpassbegrenztes Rauschen mit Hochpassanteil als Zwicker-Ton Erzeugerschall	79
5.3	Tiefpassrauschen mit bandpassbegrenztem Rauschen als Zwicker-Ton Erzeugerschall	80
5.4	Rauschen mit spektraler Lücke als Zwicker-Ton Erzeugerschall	80
5.5	Modell zur Berechnung der Tonhöhe des Zwicker-Tons	81
6.1	Entstehung des Zwicker-Tons im Habituationsmodell	85
6.2	Bestandteile des Gehörmodells	86
6.3	Simulation der Basilarmembran	87
6.4	Simulation einer inneren Haarzelle	88
6.5	Simulation einer Hörnervenfaser	89
6.6	Simulation des Nucleus cochlearis im Habituationsmodell	89
6.7	Spike-Response-Neuronenmodell	90
6.8	Modell für ermüdende Neurone	91
6.9	Tiefpassrauschen im Habituationsmodell	92
6.10	Reiner Ton mit Rauschen im Habituationsmodell	94
6.11	Hochpassrauschen im Habituationsmodell	95
6.12	Simulation des Nucleus cochlearis im Adaptionmodell	104
6.13	Tiefpassrauschen im Adaptionmodell	105
6.14	Reiner Ton mit Rauschen im Adaptionmodell	107
7.1	Tiefpassrauschen mit reinem Ton als Zwicker-Ton Erzeugerschall im Adaptionmodell	111
7.2	steil ansteigendes und abfallendes Schmalbandrauschen als Zwicker-Ton Erzeugerschall im Adaptionmodell	112

A.1 Farbcodierung 115

A.2 Zeitlicher Verlauf der Feuerraten bei Tiefpassrauschen
im Habituationsmodell 115

A.3 Zeitlicher Verlauf der Feuerraten bei einem reinen Ton
mit Rauschen im Habituationsmodell 116

A.4 Zeitlicher Verlauf der Feuerraten bei Hochpassrauschen
im Habituationsmodell 116

A.5 Zeitlicher Verlauf der Feuerraten bei Tiefpassrauschen
im Adaptionsmodell 117

A.6 Zeitlicher Verlauf der Feuerraten der Rauschunterdrückerschicht
bei Tiefpassrauschen im Adaptionsmodell 117

A.7 Zeitlicher Verlauf der Feuerraten bei einem reinen Ton
mit Rauschen im Adaptionsmodell 118

A.8 Zeitlicher Verlauf der Feuerraten der Rauschunterdrückerschicht
bei einem reinen Ton mit Rauschen im Adaptionsmodell . . . 118

Tabellenverzeichnis

1.1	Umrechnung zwischen Tonheit, Tonunterschiedsschritten, Tonhöhe und Zahl der Haarzellen	20
3.1	Parameter einer Nervenfasernach dem Meddis-Modell	50
4.1	Konstanten der Neuronenschicht mit lateraler Inhibition	67
5.1	Qualität des Zwicker-Tons und symbolische Bezeichnungen	77
6.1	Qualitativer Vergleich des Habituationsmodells mit der Realität	97
6.2	Zeitkonstanten und Schwellen der Neuronen des Adaptionsmodells	102
6.3	Konstanten für die Verschaltung der Featuredetektoren des Adaptionsmodells	104

Kapitel 1

Psychoakustik

Im Folgenden sollen einige psychoakustische Grundlagen vorgestellt werden. Sie stammen größtenteils aus [Zwicker 1982] und [Zwicker 1990].

1.1 Psychophysik

Die Psychophysik¹ interessiert sich für die Zusammenhänge zwischen der physikalischen Beschaffenheit des Reizes (*stimulus*) und den entsprechenden Wahrnehmungen (*sensations*). Dabei sollen die Wahrnehmungen auch möglichst quantitativ beschrieben werden.

Die Psychoakustik als Teilgebiet der Psychophysik beschäftigt sich also mit dem Zusammenhang zwischen den physikalischen Eigenschaften des Schalls (z.B. Frequenz und Schalldruck) und den Wahrnehmungen² des Gehörs (z.B. Tonhöhe und Lautstärke).

Um Beziehungen zwischen den physikalischen Eigenschaften des Reizes und den entsprechenden Wahrnehmungen finden zu können, müssen die Wahrnehmungen gemessen werden.

Dazu teilt man die Wahrnehmungen zunächst in „Modalitäten“ ein. Alle Wahrnehmungen, die ihre Ursache in der Reizung eines bestimmten Sinnesorgans haben, gehören zu einer Modalität³. „Sehen“ und „Hören“ sind also

¹Der Begriff stammt von GUSTAV THEODOR FECHNER 1860. Er unterschied zwischen äußerer und innerer Psychophysik. Die äußere Psychophysik beschäftigt sich mit den Zusammenhängen zwischen den physikalischen Eigenschaften des Reizes und der Wahrnehmung. Mit einer Wahrnehmung sind jedoch auch physiologische Vorgänge in den Sinnesorganen und im Nervensystem verbunden. Die innere Psychophysik beschäftigt sich mit den Beziehungen zwischen diesen und der Wahrnehmung. [Campenhausen 1981]

²Man muss hier begrifflich streng unterscheiden. Einige Begriffe der Akustik werden jedoch oft doppeldeutig verwendet, z.B. Lautstärkepegel und Schalldruckpegel. [Campenhausen 1981]

³Diese Unterteilung ist problematisch, denn welche Sinneszellen gehören zu einem Sinnesorgan? Gehört etwa „Kälte“ und „Wärme“ zur selben Modalität oder nicht? Wir sind geneigt, diese beiden Empfindungen zur selben Modalität zu zählen, obwohl sie von verschiedenen Sinneszellen registriert werden. Blinde orientieren sich nachweislich (DALLENBACH 1942, I. KOHLER 1966, C. E. RICE) auch anhand des Echos selbsterzeugter Schalle, das von Gegenständen zurückgeworfen wird. Die Versuchspersonen behaupten allerdings meist, sie nähmen die Gegenstände nicht mit dem Gehör, sondern mit dem „Gesicht“

verschiedene Modalitäten. Modalitäten unterteilt man wiederum in „Qualitäten“. So sind „Helligkeit“, „Tonhöhe“ und „Lautstärke“ Qualitäten.

Quantitativ messen kann man nur die *Intensität* einer Qualität. Hierzu gibt es mehrere Methoden:

- Methode der Vergleichswerte: Die Versuchsperson stellt einen Vergleichsreiz so ein, dass eine bestimmte Qualität des Reizes und des Vergleichsreizes als gleich empfunden wird. Die Versuchsperson stellt beispielsweise den Schalldruckpegel eines Sinustons so ein, dass sie den Sinuston als gleich laut im Vergleich zu einem vorgegebenen weißen Rauschen empfindet. Der Schalldruckpegel eines Sinustons ist dann ein Maß für die Intensität der Lautstärkeempfindung⁴.
- Methode der Verhältnswerte: Die Versuchsperson stellt einen Vergleichsreiz so ein, dass eine bestimmte Qualität des Vergleichsreizes als doppelt oder halb so intensiv wie beim Reiz selbst empfunden wird⁵.

Die Methode der Vergleichswerte ergibt eine Messgröße, die genau dann wächst, wenn auch die entsprechende Empfindungsstärke zunimmt. Die Methode der Verhältnswerte ergibt ein Maß für die subjektive Stärke der Empfindung.

1.2 Lautstärke

Die subjektive Lautstärkenempfindung wird hier einfach als „Lautstärke“ bezeichnet. Gesucht ist nun eine physikalische Messgröße, die diese subjektive Lautstärkenempfindung möglichst gut beschreibt. Von einer „möglichst guten Beschreibung“ ist mindestens zu fordern, dass die Messgröße genau dann

wahr. Hier steht also offenbar noch eine zusätzliche Qualität oder Modalität der Empfindung zur Verfügung, die sich von der Empfindung des Hörens wesentlich unterscheidet. [Campenhausen 1981, Moore 1982]

⁴Hierbei wird vorausgesetzt, dass die (subjektive) Relation „lauter“ transitiv ist, d.h. falls Schall A lauter als Schall B und Schall B wiederum lauter als Schall C ist, muss Schall A auch lauter als Schall C sein. Dies ist keineswegs sichergestellt, zumal die Schalle qualitativ ja ganz unterschiedlich sein können.

⁵Dieser Ansatz erscheint ungleich problematischer. So muss auch in der Regel über viele Versuchspersonen und über viele Versuche gemittelt werden, um vernünftige Daten zu erhalten. „Mitteln“ kann man auf verschiedene Weisen, es gibt das arithmetische und das geometrische Mittel sowie den Zentralwert. Der Zentralwert ist bei einer ungeraden Anzahl von Messungen der Wert oberhalb und unterhalb dessen gleich viele Messwerte liegen, bei einer geraden Anzahl von Messungen das arithmetische Mittel der mittleren beiden Messwerte. Der Zentralwert ist eine Abschätzung für den Median, dem Wert, der bei einer Messung mit Wahrscheinlichkeit 1/2 über- bzw. unterschritten wird. Nur der Median ist invariant gegenüber einer nichtlinearen Transformation der Messskala, d.h. bei einer nichtlinearen Transformation der Messskala wird der Median auf den Median abgebildet. Da in der Psychophysik physikalische Messgrößen i.d.R. nichtlinear transformiert werden müssen, um „empfindungsrichtige“ Messgrößen zu erhalten, gibt man hier gerne den Zentralwert an. Der Modalwert, d.i. das Maximum der Wahrscheinlichkeitsverteilung, ist ebenfalls gegenüber nichtlinearen Transformationen invariant, er ist aber bei multimodalen Verteilungen nicht eindeutig und wesentlich schwieriger zu schätzen als der Median.

ansteigt, wenn die wahrgenommene Lautstärke ansteigt. Diese Forderung wird durch das psychoakustische Maß des Lautstärkepegels erfüllt.

Außerdem kann man noch fordern, dass, falls die Messgröße doppelt so groß ist, auch ein doppelt so lauter Schall wahrgenommen wird. Diese Forderung erfüllt die Lautheit.

Bevor die Größen Lautstärkepegel und Lautheit besprochen werden, wird zunächst auf die physikalischen Größen Schalldruckpegel und Schallintensitätspegel eingegangen, von denen die empfundene Lautstärke in erster Linie abhängt. Daneben ist die Lautstärke eines Sinustons auch stark von der Frequenz des Tons abhängig.

1.2.1 Schalldruckpegel und Schallintensitätspegel

Der Schalldruck (*sound pressure*) $p(t)$ an einer bestimmten Stelle ist der Luftdruck zur Zeit t , den ein Schall an dieser Stelle hervorruft.

Die Schalldruckamplitude (*sound pressure amplitude*) \hat{p} eines Sinustons ist die Amplitude des Schalldrucks dieses Sinustons.

Der Schalldruckpegel (*sound pressure level*) L eines Sinustons ist definiert als

$$(1.1) \quad L := 20 \log(\hat{p}/p_0) \text{ dB},$$

wobei $p_0 = 2 \cdot 10^{-5}$ Pa ein willkürlich in etwa auf die Hörschwelle⁶ festgelegter Bezugsdruck ist. Als Einheit für den Schalldruckpegel wird auch dB SPL (von „sound pressure level“) verwendet⁷.

Die Schallintensität (*sound intensity*) ist die akustische Energie, die eine Fläche pro Zeiteinheit durchdringt. Schalldruckpegel L und Schallintensität I eines Sinustons hängen nach

$$(1.2) \quad L = 10 \log(I/I_0) \text{ dB}$$

zusammen, wobei $I_0 = 10^{-12} \text{ W/m}^2$ ist. Der so berechnete Schalldruckpegel wird auch als Schallintensitätspegel (*sound intensity level*) bezeichnet⁸.

Die Schallintensität pro Frequenzbereich heißt Schallintensitätsdichte (*sound intensity density*), das entsprechende logarithmische Maß ist der Schallintensitätsdichtepegel (*sound intensity density level*).

⁶Die Hörschwelle ist der Schalldruck, ab dem man einen Sinuston gerade hört. Sie hängt natürlich von der Frequenz des Tons ab.

⁷Die „Einheit“ Bel, Kurzzeichen B, ist nach ALEXANDER BELL (1847–1922) benannt [dtv-Atlas 1987].

⁸Der Schallintensitätspegel wird oft mit dem WEBER-FECHNER'schen Gesetz, benannt nach ERNST WEBER (1795–1878) und GUSTAV FECHNER (1801–1887), in Zusammenhang gebracht [Campenhausen 1981, dtv-Atlas 1987, Grimsehl 1987]. Es lautet: „Die Empfindungsstärke wächst proportional zum Logarithmus der (physikalischen) Intensität der Reizgröße“. Bei der Lautstärkeempfindung trifft dies jedoch nicht zu (siehe Kap. 1.2.3 und Kap. 1.5.3), der Schallintensitätspegel ist also nicht proportional zur Empfindungsstärke. Schlüsse über die Lautheit wie folgender (aus [Grimsehl 1987]) treffen daher nicht zu: „10 Hupen von je 90 Phon ergeben eine Lautstärke von 100 Phon; die Ausschaltung von 9 Hupen vermindert den Lärm also nur um 10 %.“ (Zur Definition des Phons siehe Kap. 1.2.2.)

Eine Zunahme des Schalldruckpegels um 3 dB bedeutet eine Zunahme der Schallintensität um etwa den Faktor 2.

1.2.2 Lautstärkepegel

Der Lautstärkepegel (*loudness level*) eines beliebigen Schalls wird in Phon⁹ gemessen. Dazu wird der Schall mit einem Sinuston der Frequenz 1000 Hz verglichen. Der Schalldruckpegel des Sinustons wird angeglichen, bis beide Schalle gleich laut erscheinen. Der Zahlenwert des Lautstärkepegels in Phon wird dann gleich dem Zahlenwert des Schalldruckpegels des Sinustons in dB gesetzt¹⁰.

Lautstärkepegel und Schalldruckpegel von Sinustönen der Frequenz 1000 Hz stimmen also überein. Außerdem werden Schalle mit dem gleichen Lautstärkepegel als gleich laut empfunden.

Der Lautstärkepegel eines gerade noch wahrnehmbaren Sinustons wird als Ruheshchwelle (*threshold in quiet*) bezeichnet. Er liegt bei etwa 3 phon. Zwischen 2 und 3 kHz liegt der empfindlichste Bereich des menschlichen Gehörs, hier können noch Schalldruckpegel bis zu etwa -5 dB detektiert werden¹¹.

1.2.3 Lautheit

Die Lautheit (*loudness*) eines Sinustons der Frequenz 1000 Hz mit einem Lautstärkepegel von 40 phon ist definiert als 1 sone. Die Lautheit eines als doppelt bzw. halb so laut empfundenen Sinustons ist 2 bzw. $1/2$ sone etc. Schalle mit dem selben Lautstärkepegel, d.i. Schalle die gleich laut erscheinen, haben definitionsgemäß die selbe Lautheit.

Damit ist die Lautheit von beliebigen Schallen festgelegt. Ein Schall mit Lautheit N wird doppelt so laut empfunden wie ein Schall mit Lautheit $1/2 N$. Ein Sinuston der Lautheit 0 sone liegt an der Wahrnehmungsgrenze, hat also einen Lautstärkepegel von etwa 3 phon.

Für die Lautheit N eines Sinustons der Frequenz 1 kHz mit der Schallintensität I gilt ab Pegeln von etwa 30 dB die Näherungsformel¹²

⁹von gr. phoné = Stimme

¹⁰Die Kurven gleicher Lautstärke (Isophonen) hat man auch durch sog. „Bewertungskurven“ angenähert. Die A-Kurve gilt für Lautstärkepegel bis 55 phon, die B-Kurve für Lautstärkepegel von 55 phon bis 85 phon und die C-Kurve für Lautstärkepegel über 85 phon, denn der Verlauf der Isophonen ist für verschiedene Lautstärkepegel unterschiedlich. Man gibt jedoch oft einfach die nach der A-Kurve über das Spektrum gewichtete Schallintensität in dB(A) an [Profos 1994]. Eine bessere Messvorschrift für die (subjektive) Lautheit beliebiger Schalle ist in Kap. 1.2.3 bzw. in Kap. 1.5.3 beschrieben.

¹¹D.h. bei einem wirksamen Querschnitt des Ohres von $0,5 \text{ cm}^2$ wird nach (1.2) eine Schalleistung von $1,6 \cdot 10^{-17} \text{ W}$ eben noch wahrgenommen. Dies entspricht mit etwa 40 Photonen der Wellenlänge 510 nm pro Sekunde ungefähr der Lichtleistung, die vom Auge eben noch wahrgenommen wird. [Grimsehl 1987]

¹²Dies ist ein Spezialfall der STEVENS'schen Potenzfunktion, benannt nach STANLEY S. STEVENS (1906–1973): $E = kI^\beta$, wobei E die Empfindungsstärke und I die (physikalische) Reizintensität bedeutet. Der Exponent β hängt von der Art des Reizes ab.

$$(1.3) \quad N \approx \frac{1}{16} \left(\frac{I}{I_0} \right)^{0,3} \text{ sone.}$$

1.3 Tonhöhe

Die subjektive Tonhöhenempfindung wird hier einfach als „Tonhöhe“ bezeichnet. Gesucht ist nun eine physikalische Messgröße, die diese subjektive Tonhöhenempfindung möglichst gut beschreibt. Von einer „möglichst guten Beschreibung“ ist mindestens zu fordern, dass die Messgröße genau dann ansteigt, wenn die wahrgenommene Tonhöhe ansteigt. Außerdem kann man noch fordern, dass, falls die Messgröße doppelt so groß ist, auch ein doppelt so hoher Ton wahrgenommen wird. Dasselbe gilt auch für die oben als Maß für die Lautstärke eingeführte Lautheit. Es sollte sich dann herausstellen, dass die gefundene Messgröße auch gut zur Beschreibung psychoakustischer Phänomene geeignet ist, die mit der Wahrnehmung der Tonhöhe zusammenhängen.

Einen Sinuston bezeichnet man auch als reinen Ton oder einfach als Ton (*tone*). Einen Schall, der aus einer Überlagerung von reinen Tönen besteht, bezeichnet man als Klang (*complex tone*). Wenn die Frequenzen alle ganzzahlige Vielfache einer Grundfrequenz (*basic or fundamental frequency*) sind, bezeichnet man diesen Klang als harmonisch (*harmonic complex tone*).

Nicht nur reine Töne rufen eine Tonhöhenempfindung hervor, sondern auch etwa Klänge, Glockentöne¹³ und sogar bandpassbegrenztes Rauschen. Außerdem ist die Tonhöhe eines reinen Tons nicht nur von dessen Frequenz, sondern auch in geringem Maße von der Lautstärke abhängig. Gesucht ist also weiter die Abhängigkeit der Tonhöhe von der physikalischen Beschaffenheit des Schalls.

Es gibt zwei psychoakustische Messgrößen für die Tonhöhe (*pitch*): Die Verhältnistönhöhe (*ratio pitch*) misst man in mel¹⁴, wobei die Skala der hörbaren Tonhöhen von 0 bis 2400 mel reicht. Die Tonheit (*critical-band rate*) misst man in Bark¹⁵, wobei deren Skala von 0 bis 24 Bark reicht.

Zunächst interessiert man sich nur für die Tonhöhe von reinen Tönen. Sie hängt in erster Linie von der Frequenz des reinen Tons ab. Andere Abhängigkeiten werden hier vernachlässigt¹⁶.

[Campenhausen 1981]

¹³Ein Glockenton wird mit einer Tonhöhe gehört, der keine Frequenz im Spektrum entspricht. Eine Obertonreihe *ohne* Grundton wird mit der Tonhöhe gehört, die der Frequenz des fehlenden Grundtons entspricht. Dieses Phänomen entdeckte A. SEEBECK 1841. Die Grundfrequenz entsteht nicht im Ohr, denn mit ihr lässt sich keine Schwebung erzeugen. Die gehörte Tonhöhe heißt Residuum. Die Tonhöhe eines beliebigen Schalls heißt auch virtuelle Tonhöhe (*virtual pitch*). [Campenhausen 1981]

¹⁴von „Melodie“

¹⁵Benannt nach HEINRICH BARKHAUSEN (1881–1956), der auch das Phon einführte.

¹⁶Die Tonhöhe eines reinen Tons hängt auch in geringem Maße von dessen Schalldruckpegel ab [Zwicker 1982].

1.3.1 Verhältnistönhöhe

Der Tonhöhe, die ein reiner Ton der Frequenz 125 Hz bei 40 dB hervorruft, ordnet man eine Verhältnistönhöhe von 125 mel zu. Ausgehend von diesem Referenzpunkt bestimmt man die Frequenz, die als die doppelte, vierfache etc. Tonhöhe empfunden wird. Außerdem wird diejenige Frequenz bestimmt, die die halbe Tonhöhenempfindung hervorruft etc.¹⁷

Man gelangt so zu einer Skala, die jeder Frequenz f eine bestimmte Verhältnistönhöhe H_v zuordnet¹⁸.

$$(1.4) \quad \frac{df(H_v)}{dH_v} \approx \begin{cases} 1 \text{ Hz/mel} & \text{für } 0 < f \leq 500 \text{ Hz} \\ 0,002f/\text{mel} & \text{für } f > 500 \text{ Hz} \end{cases}$$

$$(1.5) \quad f(H_v) \approx \begin{cases} H_v \text{ Hz/mel} & \text{für } 0 < f \leq 500 \text{ Hz} \\ 500 e^{0,002(H_v/\text{mel}-500)} \text{ Hz} & \text{für } f > 500 \text{ Hz} \end{cases}$$

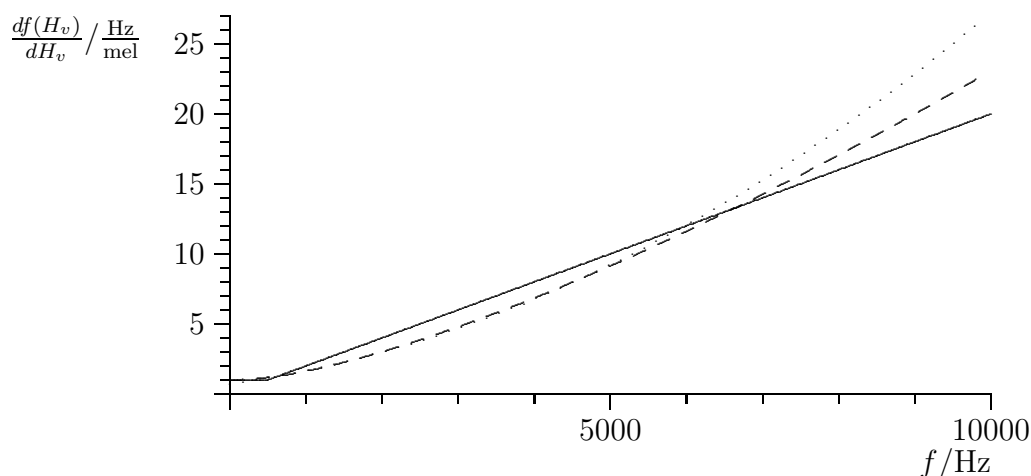


Abb. 1.1: Breite von 1 mel in Hz bei verschiedenen Frequenzen nach (1.4). Gestrichelt: Näherung nach ZWICKER, gepunktet: Näherung nach TRAUNMÜLLER (siehe Fußnote 21 auf Seite 20).

Es stellt sich heraus, dass der Ort der maximalen Auslenkung der Basilarmembran in der Cochlea, vom Helicotrema aus gemessen (siehe Kap. 2.1.3), proportional zur Verhältnistönhöhe ist, d.h. 0,7 mel entsprechen $9 \mu\text{m}$ oder einer inneren Haarzelle auf der Basilarmembran. [Zwicker 1982]

¹⁷Nach [Zwicker 1982] neigen die Versuchspersonen nach ausgiebigem Genuss abendländischer Musik dazu, einfach eine Oktave einzustellen, wenn sie nach der doppelten Tonhöhenempfindung gefragt werden. Man nimmt deshalb zur Erzeugung der Ausgangstonhöhe keinen reinen Ton, sondern Schmalbandrauschen. Außerdem benutzt man die Abfragemethode, d.h. man läßt die Versuchsperson die Tonhöhe nicht selbst einstellen, sondern fragt nur, ob die Tonhöhenempfindung bei einer vorgegebenen Vergleichstonhöhe mehr oder weniger als doppelt so hoch ist.

¹⁸Durch ihre Nichtlinearität weicht diese Skala von der üblichen Vorstellung ab, dass ein doppelt so hoher Ton auch eine doppelt so hohe Frequenz habe. Nach [Zwicker 1982] ist die abendländische Musik, jedenfalls die ältere, allerdings hauptsächlich auf einen Frequenzbereich unterhalb von 1 kHz beschränkt, in dem sich die Nichtlinearitäten noch nicht allzu stark bemerkbar machen.

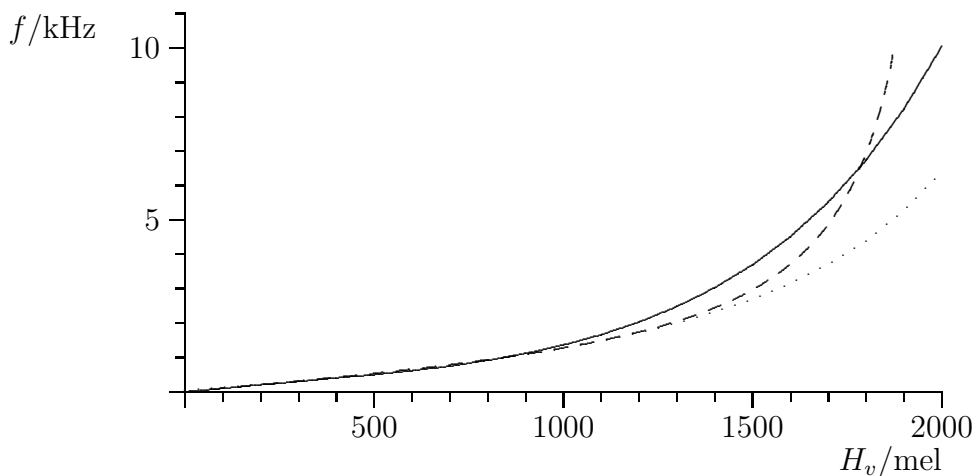


Abb. 1.2: Frequenz f als Funktion der Verhältnistönhöhe H_v nach (1.5). Gestrichelte und gepunktete Linien wie Abb. 1.1.

Die gerade noch wahrnehmbare Tonhöhendifferenz oder Tonunterschiedsschwelle¹⁹ (*just-noticeable pitch difference*) beträgt etwa 1 mel, was nur 1,5 Haarzellen entspricht. Man misst die gerade noch wahrnehmbare Tonhöhendifferenz, indem man zwei reine Töne mit Pause kurz nacheinander vorspielt. Die Versuchsperson soll entscheiden, ob die Tonhöhen der beiden Töne gleich oder verschieden waren. Je länger die Töne dauern, desto kleiner ist die gerade noch wahrnehmbare Tonhöhendifferenz. Sie erreicht bei einer Dauer von 200 ms allerdings fast schon ihren geringsten Wert.

Die gerade noch wahrnehmbare Tonhöhenmodulation (*just-noticeable pitch modulation*) beträgt etwa 3 mel. Man misst sie, indem man einen reinen Ton in der Tonhöhe moduliert. Die Versuchsperson soll entscheiden, ab welcher Amplitude sie einen Unterschied zum unmodulierten reinen Ton wahrnehmen kann. Die optimale Modulationsfrequenz ist hierbei 4 Hz.²⁰

Zum Vergleich: Das Frequenzverhältnis zweier aufeinanderfolgender Halbtöne beträgt $\sqrt[12]{2} \approx 1,059$. Bei Frequenzen kleiner als 500 Hz entspricht dies einer Tonhöhendifferenz von etwa 30 mel.

1.3.2 Tonheit

Die Tonheit (*critical-band rate*) eines reinen Tons ist

$$(1.6) \quad z \approx 100 H_v \text{ Bark/mel},$$

wobei H_v die Verhältnistönhöhe bedeutet.

¹⁹Im Unterschied zur Tonunterschiedsschwelle ist die Tonunterscheidungsschwelle diejenige Tonhöhendifferenz, ab der die Versuchsperson auch angeben kann, *welcher* der beiden Töne höher ist. Die Tonunterscheidungsschwelle ist größer als die Tonunterschiedsschwelle.

²⁰Dass die gerade noch wahrnehmbare Tonhöhenmodulation bei schnelleren Modulationsfrequenzen ansteigt, liegt wahrscheinlich an den Integrationszeiten des Gehörs. Der Anstieg bei kleinen Modulationsfrequenzen ist nach [Zwicker 1982] auf unser begrenztes Erinnerungsvermögen zurückzuführen, d.h. wenn sich die Tonhöhe nur langsam ändert, erinnern wir uns nicht mehr daran, wie hoch der Ton vorher war.

$\Delta z/\text{Bark}$	Schritte	$\Delta H_v/\text{mel}$	Haarzellen	$\Delta l/\text{mm}$
24	640	2400	3600	32
1	27	100	150	1,3
	1	3,8	5,6	0,05
		1	1,5	0,013
			1	0,009

Tab. 1.1: Umrechnung zwischen verschiedenen zur Tonhöhe proportionalen Größen: Tonheit in Bark, die Zahl der unterscheidbaren Tonhöhen (Schritte) im Sinne der gerade noch wahrnehmbaren Tonhöhenmodulation, Tonhöhe in mel, die Zahl der Haarzellen und die Länge entlang der Basilarmembran.

Der Frequenzabstand, der 1 Bark entspricht, ist nach (1.5) und (1.6) für Frequenzen kleiner 500 Hz ungefähr 100 Hz und für höhere Frequenzen ungefähr $0,2f$.²¹ Er heißt auch Frequenzgruppenbreite (*critical-bandwidth*). Alle Frequenzen, die innerhalb von einem Bark liegen, heißen auch Frequenzgruppe (*critical-band*).

Die Tonheit bzw. die Frequenzgruppe haben ihre eigene psychoakustische Berechtigung neben der Verhältnistönhöhe, da sie das geeignete Maß für einige psychoakustische Phänomene darstellen. Im Folgenden sollen einige Methoden zur Bestimmung der Frequenzgruppenbreite diskutiert werden.

Die erste Methode arbeitet mit der Lautstärkeschwelle, ab der man eine Reihe von reinen Tönen in isofrequentem Abstand, z.B. 20 Hz, gerade noch hört. „Lautstärke“ meint dabei die Schallintensität aller Töne zusammen. Liegen alle Töne innerhalb von einem Bark, so ist unabhängig von der Anzahl der Töne immer eine konstante Schallintensität notwendig, um die Töne zu hören. Liegen einige Töne außerhalb der Frequenzgruppe, so muss die Schallintensität entsprechend höher eingestellt werden. Verallgemeinert bedeutet das, ein Schall wird genau dann gehört, wenn innerhalb einer Frequenzgruppe eine ausreichende Schallintensität auf das Gehör wirkt.

²¹In [Zwicker 1982] und [Zwicker 1990] werden genauere Näherungsformeln als (1.4) und (1.5) angegeben:

$$\begin{aligned} z/\text{Bark} &= 13 \arctan(0,76f/\text{kHz}) + 3,5 \arctan(f/7,5 \text{ kHz})^2 \\ \Delta f_G/\text{Hz} &= 25 + 75 [1 + 1,4(f/\text{kHz})^2]^{0,69} \end{aligned}$$

Hierbei bedeuten z die Tonheit, f die Frequenz eines reinen Tons und Δf_G die Frequenzgruppenbreite.

TRAUNMÜLLER (1990) gibt einfachere, umkehrbare Näherungsformeln für die Tonheit an [Sotttek 1993]:

$$\begin{aligned} z/\text{Bark} &= \frac{26,81f/\text{kHz}}{1,96 + f/\text{kHz}} - 0,53 \\ f/\text{kHz} &= 1,96 \frac{z/\text{Bark} + 0,53}{26,28 - z/\text{Bark}} \\ \Delta f_G/\text{kHz} &= \frac{(1,96 + f/\text{kHz})^2}{52,5476} \end{aligned}$$

Eine zweite Methode zur Bestimmung der Frequenzgruppenbreite arbeitet analog zur ersten (siehe Abb. 1.3). Jedoch wird nun neben den reinen Tönen weißes Rauschen eingespielt. Das weiße Rauschen maskiert die reinen Töne genau dann, wenn die Schallintensität der reinen Töne innerhalb von einem Bark nicht ausreicht. D.h. erst wenn die Schallintensität der reinen Töne, die in ein Bark fallen, $1/4$ bis $1/2$ der Schallintensität des weißen Rauschens innerhalb eines Barks erreicht hat, sind die reinen Töne hörbar. Dies entspricht einem Schallpegelunterschied von 3 dB (bei niedrigen Frequenzen) bis 6 dB (bei hohen Frequenzen).

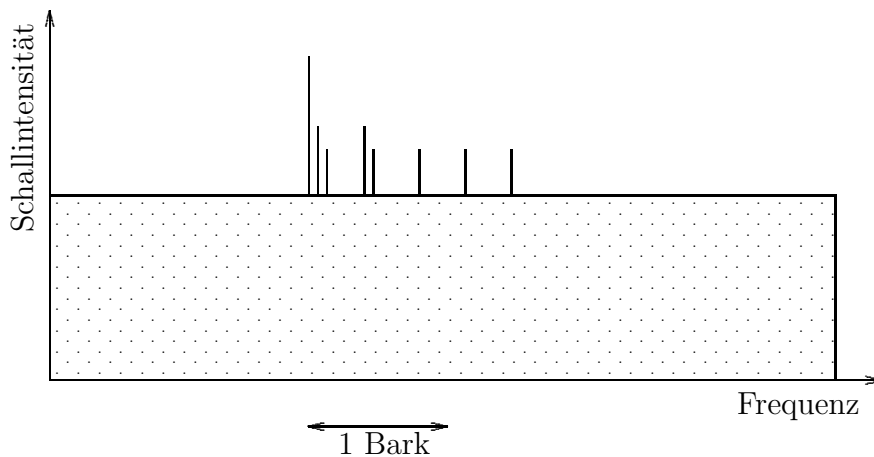


Abb. 1.3: Bestimmung der Frequenzgruppenbreite. Die Schallintensität des lautesten reinen Tons genügt gerade, um ihn trotz des Rauschens hörbar zu machen. Die beiden mittellauten Töne addieren sich zur selben Schallintensität. Bei drei oder mehr Tönen nimmt die Laustärke eines Tons nicht mehr ab, da bereits die Grenze von 1 Bark überschritten wird: Nur die Töne, die in ein Bark fallen, tragen dazu bei, das Rauschen zu übertönen.

Eine dritte Methode arbeitet mit Hilfe von Verdeckung, die im nächsten Abschnitt besprochen wird.

1.4 Verdeckung

Ein (leiser) Testschall kann durch einen anderen (lauten) Schall, den Störschall, verdeckt werden. D.h. es wird nur der Störschall gehört und nicht der Testschall. Man bezeichnet dieses Phänomen als Verdeckung (*masking*)²².

Die Mithörschwelle (*masked threshold*) ist der Schalldruckpegel eines reinen Tons (*masked tone*), den dieser haben muss, damit er trotz des Störschalls (*masker*) gerade noch wahrgenommen wird.

Als Störschall kann z.B. Schmalbandrauschen (*narrow band noise*) verwendet werden. Wichtig sind die Mithörschwellen für frequenzgruppenbreites Schmalbandrauschen (*threshold masked by critical-band wide noise*). Je näher sich die Frequenz des Testtons an der Mittenfrequenz des Schmalbandrau-

²²Dieses Phänomen ist monaural, d.h. werden Test- und Störschall verschiedenen Ohren dargeboten, so ist keine Verdeckung mehr nachweisbar [Campenhausen 1981].

schens befindet, desto größer wird die Mithörschwelle. Der Anstieg bei niedrigen Frequenzen ist dabei steiler als der Abfall bei höheren Frequenzen. Der Abfall nach höheren Frequenzen wird umso flacher, je lauter das Schmalbandrauschen ist. Man bezeichnet dies als nichtlineare Auffächerung der oberen Flanke der Mithörschwelle.

Es gibt auch das Phänomen teilweiser Verdeckung (*partial masking*). Hier wird die Lautstärkeempfindung eines Testtons in Anwesenheit eines Störschalls reduziert.

Die typische Skala, auf der Verdeckung stattfindet, ist die Barkskala: Man misst die Mithörschwelle eines reinen Tons, der von je einem Schmalbandrauschen oberhalb und unterhalb der Frequenz des reinen Tons flankiert wird. Die Mithörschwelle ist abhängig von den beiden Frequenzabständen des Schmalbandrauschens vom reinen Ton. Bei Abständen kleiner als etwa ein Bark bleibt die Mithörschwelle konstant, bei größeren Abständen nimmt sie um etwa 10 dB/Bark ab.

Interessant sind auch Vorverdeckung (*pre- or backward masking*) und Nachverdeckung (*post- or forward masking*): Nach dem Abschalten von weißem Rauschen wird ein kurzer Tonimpuls gegeben. Der Schalldruckpegel, bei dem man den Tonimpuls gerade noch hört, wird in Abhängigkeit von der Zeitdifferenz zwischen Rauschen und Tonimpuls aufgetragen. Man sieht, dass der Tonimpuls umso lauter sein muss, je geringer der Zeitabstand ist. Dieser Effekt heißt Nachverdeckung, da ein (lauter) Schall auch nach dem Abschalten einen leisen (Test-) Schall verdecken kann. Die Nachverdeckung dauert bis etwa 200 ms. Der Effekt, bei dem der Testton vor dem maskierenden Rauschen eingespielt wird, heißt Vorverdeckung. Diese dauert allerdings nur etwa 20 ms.

1.5 Funktionsschemata

Psychoakustische Funktionsschemata oder Funktionsmodelle (*models*) beschreiben den Zusammenhang zwischen den physikalischen Eigenschaften der Reizgröße (z.B. Frequenz, Schalldruck) und den Empfindungsgrößen (z.B. Tonheit, Lautheit) in Form einer Vorschrift, die es gestattet, die Empfindungsgrößen aus den Reizgrößen zu berechnen.

Funktionsschemata können zur Konstruktion von Messgeräten für Empfindungsgrößen verwendet werden und um das Zustandekommen einer bestimmten gemessenen Empfindung besser zu verstehen.

1.5.1 Anregung und Erregung

Anregung und Erregung sind Zwischengrößen, die in Funktionsschemata des Gehörs eine Rolle spielen.

Die in eine Frequenzgruppe fallende Schallintensität heißt Anregung oder Frequenzgruppenintensität (*incitation or critical-band intensity*). Für die Anregung I_G einer Frequenz f gilt

$$(1.7) \quad I_G(f) := \int_{f-\frac{1}{2}\Delta f_G(f)}^{f+\frac{1}{2}\Delta f_G(f)} \frac{dI(f)}{df} df,$$

wobei $\Delta f_G(f)$ die Frequenzgruppenbreite an der Frequenz f und $\frac{dI(f)}{df}$ die Schallintensitätsdichte ist.

Das entsprechende logarithmische Maß L_G heißt Anregungs- oder Frequenzgruppenpegel (*critical-band level*).

$$(1.8) \quad L_G := 10 \log(I_G/I_0) \text{ dB}$$

Ein frequenzgruppenbreites Schmalbandrauschen hat auf der Tonheitsskala eine dreieckförmige Anregung, ein reiner Ton hat eine rechteckige Anregung.

Zur Berechnung der Anregung kann man die Schallintensitätsdichte mit einem rechteckigen Filter falten. Dies entspricht jedoch nur ganz grob den Eigenschaften des Gehörs mit seiner beschränkten aber sehr guten Frequenzselektivität und den Mithörschwellen mit ihren Flanken. Deshalb wird das Maß der Erregung (*excitation*) eingeführt. Um die Erregung $E(z)$ zu erhalten, wird die Schallintensitätsdichte $I(z)$ mit einer Funktion $M(z)$ gefaltet, die in erster Näherung der Mithörschwelle eines reinen Tons²³ nach Kap. 1.4 entspricht, d.h.

$$(1.9) \quad E(z) = \int_0^{24 \text{ Bark}} \frac{dI(z')}{dz'} M(z' - z) dz',$$

wobei

$$M(z) \approx \begin{cases} 10^{(-27 \text{ dB/Bark } z)/(10 \text{ dB})} & \text{für } z \leq 0 \\ 10^{(-10 \text{ dB/Bark } z)/(10 \text{ dB})} & \text{für } z > 0 \end{cases}$$

Dies ist eine Vereinfachung der ausführlichen Darstellung in [Sottek 1993]. Darüberhinaus wird die nichtlineare Auffächerung der Mithörschwellen vernachlässigt.

Das entsprechende logarithmische Maß heißt Erregungspegel (*excitation level*).

Man beachte, dass die in Abb. 1.4 gezeigte Erregung in Übereinstimmung mit Kap. 1.4 zu tiefen Frequenzen hin schneller (etwa 27 dB/Bark) als zu hohen abfällt (etwa 10 dB/Bark).

1.5.2 Schwellenfunktionsschema für langsame Schalländerungen

Die Annahme dieses Modells ist, dass zwei Schalle als verschieden empfunden werden, wenn sich deren Erregungen an irgendeiner Stelle der Tonheitsskala um den Faktor 0,25 unterscheiden. Dies entspricht einer Differenz in den Erregungspegeln um 1 dB.

²³Die Mithörschwelle eines reinen Tons ist schwer zu messen, da die Versuchspersonen Schwebungen und den durch nichtlineare Verzerrungen verursachten Differenzton als Hinweis auf die Existenz des Testtons verwenden [Zwicker 1982, Zwicker 1990].

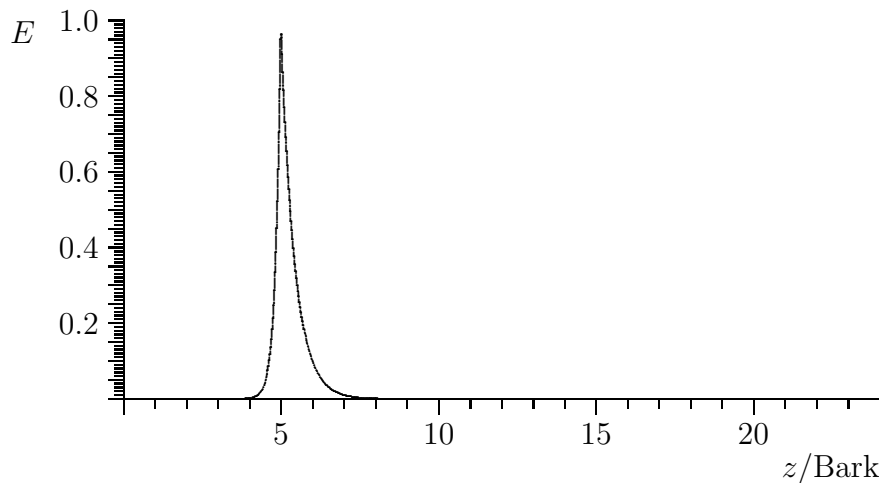


Abb. 1.4: Erregung $E(z) = M(z - 5 \text{ Bark})$ des Gehörs durch einen reinen Ton mit der Tonheit 5 Bark.

Gerade wahrnehmbare Amplitudenmodulation

Gäbe es keine nichtlineare Auffächerung, so wäre die gerade wahrnehmbare Amplitudenmodulation eines reinen Tons unabhängig vom Schallpegel genau 1 dB. Wegen der nichtlinearen Auffächerung ändert sich jedoch die Erregung an der oberen Flanke stärker als die Gesamterregung. Deshalb können bei großen Pegeln mit starker nichtlinearer Auffächerung auch Schallpegeländerungen registriert werden, die wesentlich kleiner als 1 dB sind.

Gerade wahrnehmbare Frequenzmodulation

Eine Frequenzänderung eines reinen Tons ist genau dann gerade noch wahrnehmbar, wenn sie an irgendeiner Stelle der Tonheitskala eine Erregungspegeländerung von 1 dB hervorruft. Die größte Änderung des Erregungspegels findet an der unteren Flanke statt. Die Steigung der unteren Flanke ist 27 dB/Bark. Somit gilt für die gerade noch wahrnehmbare Tonhöhenmodulationsamplitude $\Delta z \cdot 27 \frac{\text{dB}}{\text{Bark}} = 1 \text{ dB}$ und somit

$$(1.10) \quad \Delta z = \frac{1}{27} \text{ Bark} = 3,7 \text{ mel}$$

Dies stimmt gut mit der gemessenen gerade noch wahrnehmbaren Tonhöhenmodulation von etwa 3 mel überein.

Ruhehörschwelle und Mithörschwelle

Ein reiner Ton wird neben einem Störschall genau dann gehört, wenn er an irgendeiner Stelle der Tonheitskala eine Erregungspegeländerung von 1 dB hervorruft.

Die Ruhehörschwelle ist durch ein internes Rauschen erklärbar, etwa in der neuronalen Verarbeitung oder bei tiefen Frequenzen durch Muskelbewegung und Puls hervorgerufen.

1.5.3 Funktionsschema der Lautheit

Zwei Töne mit großem Frequenzabstand addieren sich in der Lautheit. Zwei Töne der gleichen Frequenz addieren sich natürlich in der Schallintensität und führen zu einer Änderung der Lautheit entsprechend (1.3). Dazwischen gibt es einen fließenden Übergang, der durch das Konzept der Erregung erklärt werden kann. Man nimmt an, dass die Lautheit N eines Schalls das Integral über die spezifische Lautheit N' (*specific loudness*)

$$(1.11) \quad N = \int_0^{24 \text{ Bark}} N' dz,$$

ist, wobei die spezifische Lautheit aus der Erregung des Schalls berechnet werden kann nach

$$(1.12) \quad N' = 0,08 \left(\frac{E_{RHS}}{E_0} \right)^{0,23} \left[\left(\frac{1 + E/E_{RHS}}{2} \right)^{0,23} - 1 \right] \frac{\text{sone}}{\text{Bark}},$$

wobei E_{RHS} die Erregung an der Ruhehörschwelle und E_0 die Erregung durch einen Ton der Schallintensität I_0 bedeuten²⁴. [Sottek 1993, Zwicker 1982]

Man beachte, dass sich für eine gerade noch unhörbare Erregung $E = E_{RHS}$ $N' = 0$ ergibt. Der Vorfaktor 0,08 ist so gewählt, dass die Lautheit eines Sinustons mit dem Lautstärkepegel 40 phon gerade $N = 1$ sone ist, wie in Kap. 1.2.3 definiert.

Das hier kurz vorgestellte Funktionsschema der Lautheit gilt nur für stationäre, d.h. nicht stark zeitabhängige Schalle. Differenzierte Darstellungen finden sich in [Sottek 1993].

²⁴Irrige Vorstellungen über die Lautheit werden in [dtv-Atlas 1987] verbreitet.

Kapitel 2

Auditorisches System

Der das Ohr erreichende Schall wird in einen neuronalen Code umgewandelt. Dieser Code wird weiterverarbeitet, mit Eingaben von anderen Sinnen und der Erfahrung kombiniert und führt letztlich zu einer Reaktion des Individuums. Diesen gesamten Vorgang bezeichnet man als „hören“.

Die für das Hören notwendigen Strukturen bezeichnet man als auditorisches System (*auditory system*). Die unten angegebenen Informationen stammen größtenteils aus [Popper 1992, Zenner 1994, Zwicker 1990].

2.1 Peripheres auditorisches System

Das periphere auditorische System (*peripheral auditory system*) wandelt den Schall in einen neuronalen Code um. Es besteht aus Außenohr, Mittelohr und Innenohr. Es wird auch einfach als „Ohr“ bezeichnet.

2.1.1 Außenohr

Das Außenohr (*outer ear*) sammelt den Schall und leitet ihn zum Mittelohr weiter. Es besteht aus der Ohrmuschel (*pinna*) und dem Gehörgang (*external auditory canal*), der am Trommelfell endet. Die zahlreichen Einbuchtungen und Erhebungen in der Ohrmuschel haben auf die Umwandlung bzw. Weiterleitung des Schalls keinen Einfluss, da sie kleiner sind als die typischen Wellenlängen hörbaren Schalls. (Ein Sinuston der Frequenz 16 kHz hat eine Wellenlänge von etwa 2 cm.)¹ Einen größeren Einfluss haben der Kopf und die Ohrmuschel selbst: Sie wirken als linearer Filter, der das Spektrum des freien Schallfeldes richtungsabhängig verändert, denn tiefere Frequenzen werden stärker um den Kopf herum gebeugt als hohe². Die Übertragungsfunktion zwischen dem Schalldruck im freien Schallfeld und dem Schalldruck

¹Dies ist jedenfalls die gängige Lehrmeinung. Nach [Spektrum 4/1997] spielen allerdings die Reflexionen an den Erhebungen evtl. beim Richtungshören eine Rolle.

²Dies kann zur Bestimmung der Richtung einer Schallquelle ausgenutzt werden, ohne den Kopf zu drehen. Nur Säugetiere, die auch relativ hohe Frequenzen noch hören können, haben bewegliche Ohrmuscheln

unmittelbar am Trommelfell heißt auch Übertragungsfunktion des Kopfes³ (*head-related transfer function, HRTF*).

2.1.2 Mittelohr

Das Mittelohr (*middle ear*) beginnt am Trommelfell (*ear drum or tympanic membrane*) und endet an der Cochlea. Es dient zur Umwandlung der Schalldruckänderungen am Trommelfell in Schwingungen der Flüssigkeit in der Cochlea. Dabei sollte möglichst die gesamte Schallenergie absorbiert werden⁴. Dazu werden die Schwingungen des Trommelfells zunächst auf die Gehörknöchelchen (*middle ear bones*) Hammer (*malleus*), Amboss (*incus*) und Steigbügel (*stapes*) übertragen, wobei letzterer die Schwingungen auf das ovale Fenster der Cochlea überträgt. Der Druck auf das ovale Fenster wird dabei etwa um den Faktor 44 verstärkt, wobei 17-fache Verstärkung auf das Flächenverhältnis zwischen Trommelfell und ovalem Fenster, 1,3-fache Verstärkung auf die Hebelwirkung der Gehörknöchelchen und 2-fache Verstärkung auf die Tendenz des Trommelfells sich zu wölben zurückzuführen ist. Dies gilt bei etwa 2500 Hz, die Übertragungsfunktion ist wieder frequenzabhängig⁵.

2.1.3 Innenohr

Das Innenohr besteht aus den Bogengängen (*semicircular canal*), dem Vestibularapparat (*vestibule*) und der Cochlea (*cochlea*). Bogengänge und Vestibularapparat dienen dem Gleichgewichtssinn. Die Cochlea führt, grob gesagt, eine Spektralanalyse des auf das ovale Fenster (*oval window*) übertragenen Schalls durch⁶.

³Diese muss bei psychoakustischen Messungen mit Kopfhörer berücksichtigt werden, indem ein sog. Freifeldentzerrer vorgeschaltet wird, der den Einfluss des Kopfes simuliert, so dass die Versuchsperson den Schall so hört, als hörte sie ihn im „freien Schallfeld“.

⁴Die Impedanz von Wasser ist wesentlich größer als die Impedanz von Luft ($1,5 \cdot 10^6$ bzw. $420 \text{ kg/m}^2\text{s}$). Dementsprechend hat die flüssigkeitsgefüllte Cochlea eine hohe Eingangsimpedanz (frequenzabhängig $6 \cdot 10^4$ – $1,5 \cdot 10^5 \text{ kg/m}^2\text{s}$ bei der Katze). Träfe der Schall somit direkt auf die Flüssigkeit in der Cochlea, würde ein Großteil der Schallenergie reflektiert werden. Das Mittelohr sorgt nun dafür, dass das System aus Trommelfell, Gehörknöchelchen und Flüssigkeit eine der Luft vergleichbare Impedanz hat. Man spricht von Impedanzanpassung durch das Mittelohr. Patienten ohne Gehörknöchelchen haben aufgrund der fehlenden Impedanzanpassung einen Hörverlust von 60 dB [Zenner 1994]. Die genaue Abhängigkeit der Impedanz des Trommelfells von der Frequenz, auch ihres imaginären Anteils, findet man in [Keidel 1995].

⁵Die Gehörknöchelchen sind bei höheren Frequenzen ohne Bedeutung, da diese durch Knochenleitung auf das Innenohr übertragen werden [Campenhausen 1981].

⁶Diese Tatsache bezeichnet man auch als OHM'sches Gesetz der Akustik, benannt nach GEORG SIMON OHM (1789–1845). Insbesondere spielen nach diesem Gesetz die Phasenlagen der Teilschwingungen keine Rolle. Außerdem sollten sich reine Töne gegenseitig nicht beeinflussen. Die Phänomene Schwebung und Rauigkeit sowie das von GEORG ANDREAS SORGE 1745 und GUISEPPE TARTINI 1754 beschriebene Hören von Kombinationstönen widersprechen jedoch dem OHM'schen Gesetz. Überlagert man etwa zwei Töne mit den Frequenzen f_1 und f_2 , wobei $f_2 = 2f_1 + \Delta f$ und $\Delta f \ll f_1$, so ändert sich deren Phasenlage periodisch. Diesen periodischen Wechsel kann man hören. Periodische Änderungen der Phasenlage können auch zur Wahrnehmung zusätzlicher Töne führen: Überlagert man die

Cochlea

Die Cochlea ist ein eingerollter, etwa 35 mm langer, flüssigkeitsgefüllter Kanal. Er ist an seinem Ende, dem Apex, verschlossen. Der Kanal ist durch Membranen in mehrere Unterkanäle eingeteilt: Die Scala tympani wird durch die Basilarmembran (*basilar membrane*) von der Scala media getrennt, die Scala media von der Scala vestibuli durch die REISSNER'sche Membran. Scala vestibuli und Scala tympani sind am Apex verbunden. Diese Verbindung heißt Helicotrema. Das Ende des Kanals am Apex heißt auch apikal, das andere Ende basal.

Das ovale Fenster stellt die Verbindung zwischen Steigbügel und Scala tympani her, das runde Fenster (*round window*) trennt die Scala tympani vom Mittelohr. Bringt der Steigbügel das ovale Fenster zum Schwingen, so überträgt sich diese Schwingung auf die Flüssigkeit in der Scala vestibuli und somit auf die Basilarmembran. Die Basilarmembran ist am apikalen Ende breiter (0,52 mm), elastischer und nicht unter Spannung, während sie am Ende mit dem ovalen Fenster steifer, schmaler (0,16 mm) und unter leichter Spannung ist. Sie wird außerdem von vielen Querfasern durchzogen. Betrachtet man jede dieser Fasern als unabhängigen Schwinger, so sieht man aufgrund der oben beschriebenen Eigenschaften der Basilarmembran, dass diese Schwinger vom basalen zum apikalen Ende hin abnehmende Eigen- bzw. Resonanzfrequenzen⁷ haben. Es ist daher zu erwarten, dass tiefe Frequenzen zu starken Auslenkungen der Basilarmembran am apikalen Ende und hohe Frequenzen zu starken Auslenkungen am basalen Ende führen. Beobachtungen, mechanische Modelle und Simulationen bestätigen dies. Tatsächlich zeigt sich, dass bei Stimulation des ovalen Fensters mit einem Sinuston eine Wanderwelle vom basalen Ende zum apikalen Ende der Basilarmembran läuft⁸. Die Einhüllende dieser Wanderwelle hat ein Maximum, wobei die Stelle dieses Maximums in der beschriebenen Weise von der Frequenz des Sinustons abhängt. Die Basilarmembran wirkt somit als Frequenzanalysator.

beiden Frequenzen f_1 und f_2 mit $f_2 = 2/3f_1$, so ist die Überlagerung mit der Frequenz $f_W = 1/2f_1$ periodisch. Diese Frequenz hört man als zusätzlichen Ton. Diese Art von Kombinationstönen kommt erst durch neuronale Verarbeitung im Gehirn zustande, denn das Phänomen tritt auch dann auf, wenn jedes Ohr mit nur einer Frequenz gereizt wird. Außerdem lässt sich der zusätzliche Ton durch Rauschen nicht verdecken. [Campenhausen 1981]

⁷ Die Resonanztheorie des Hörens stammt von HERMANN LUDWIG FERDINAND VON HELMHOLTZ 1863. Durch reine Resonanz lässt sich die Übertragungsfunktion der Basilarmembran allerdings nicht erklären, denn Resonanz wäre zu träge und Resonanzkurven wären zu breit. Das Gehör kann bereits die Tonhöhe eines Tonreizes mit nur zwei bis drei Perioden bestimmen, während ein Resonator viel mehr Zeit für den Einschwingvorgang bräuchte. Auch die Phasenverzögerungen auf der Basilarmembran verhalten sich anders als bei einem Resonator zu erwarten wäre [Campenhausen 1981]. Siehe auch Kap. 3.1.1.

⁸Die Bewegung der Basilarmembran wurde von GEORG VON BÉKÉSY (1899–1972) mit dem Mikroskop unter stroboskopischer Beleuchtung beobachtet. Er baute auch mechanische Cochlea-Modelle. [Campenhausen 1981]

Cortisches Organ

Auf der Basilarmembran, auf der Seite der Scala media, sitzt das Cortische Organ⁹ (*organ of corti*). Es besteht im Wesentlichen aus einer Reihe innerer Haarzellen vom basalen zum apikalen Ende, die von einer Membran, der Tektorial- oder Deckmembran (*tectorial membrane*), abgedeckt werden. Die Stereocilien der inneren Haarzellen berühren die Tektorialmembran nicht. Jedoch gibt es drei Reihen äußerer Haarzellen, deren Stereocilien mit der Tektorialmembran verwachsen sind. Die Bewegung der Flüssigkeit zwischen Tektorial- und Basilarmembran verbiegt die Stereocilien der inneren Haarzellen.

Innere Haarzellen

Die etwa 3600 inneren Haarzellen (*inner hair cells*) reagieren auf eine Bewegung ihrer Stereocilien mit einem elektrochemischen Potential. Es heißt Rezeptorpotential (*receptor potential*). Das Rezeptorpotential erregt die afferenten auditorischen Nervenfasern. Jede innere Haarzelle wird von etwa 8 Hörnervenfasern kontaktiert.

Äußere Haarzellen

Das Cortische Organ enthält etwa 20000 äußere Haarzellen, die in drei Reihen angeordnet sind. Die Stereocilien der äußeren Haarzellen sind im Ggs. zu denen der inneren Haarzellen fest mit der Tektorialmembran verwachsen. Sie reagieren auf eine Bewegung ihrer Stereocilien mit einer Kontraktion¹⁰, verstärken dadurch die Schwingung der Basilarmembran und somit indirekt die Erregung der inneren Haarzellen¹¹.

Efferente Nervenfasern, die an den äußeren Haarzellen enden¹², reduzieren deren motorische Aktivität [Shepherd 1993]. Man vermutet deshalb, dass die äußeren Haarzellen dazu dienen, so etwas wie den optimalen „Arbeitspunkt“ des Gehörs einzustellen.

2.1.4 Hörnerv

Hier soll beschrieben werden, wie das periphere auditorische System Schall letztlich in Nervenimpulse im auditorischen Nerv umwandelt. Insbesondere müssen Schallintensität und Frequenz eines Sinustons codiert werden.

⁹Von ALFONSO CORTI 1851 beschrieben.

¹⁰Die Zeitkonstante hierfür beträgt etwa 240 μ s, d.h. die äußeren Haarzellen können einer Bewegung der Basilarmembran sehr gut folgen.

¹¹Die äußeren Haarzellen führen auch zu einer schärferen Frequenzabstimmung in der Antwort der inneren Haarzellen. Die Frequenzselektivität der Cochlea und damit des Gehörs hängt also entscheidend von aktiven Prozessen in den äußeren Haarzellen ab. [Shepherd 1993]

¹²20 äußere Haarzellen innervieren zusammen auch eine afferente Nervenfasern. Diese Afferenzen dienen jedoch wahrscheinlich nur zur Kontrolle der äußeren Haarzellen und nicht der Repräsentation von Schall.

Codierung der Intensität

Die spontane Aktivität von auditorischen Nervenfasern reicht von wenigen Aktionspotentialen pro Sekunde bis zu etwa 100 pro Sekunde. Die Maximale Feuerrate sind etwa 500 Aktionspotentiale pro Sekunde.

Die Schwellintensität eines Neurons kann definiert werden als der Stimulus, der die Feuerrate über die spontane Feuerrate hinaus erhöht. Auditorische Nervenfasern haben eine umso niedrigere Schwellintensität, je größer ihre spontane Aktivität ist.

Der Zusammenhang zwischen der Intensität des Stimulus und der Feuerrate heißt Raten-Pegel-Funktion (*input-output-function, intensity function or rate-level function*). Je intensiver der Stimulus, desto größer wird i.d.R. die Feuerrate, bis sie auf einem maximalen Wert sättigt oder sogar wieder ein wenig abnimmt. Je größer die spontane Feuerrate der Nervenfaser, desto größer ist die Feuerrate bei Stimulation und desto niedriger ist die Schallintensität, bei der die Feuerrate dieser Nervenfaser sättigt. Das Verhältnis zwischen der größten Schallintensität, die noch ein Anwachsen der Feuerrate bewirkt, und der Schwellintensität heißt Dynamikbereich (*dynamic range*) der Nervenfasers. Der Dynamikbereich auditorischer Nervenfasern liegt zwischen 20 und 50 dB.

Das Gehör hat einen Dynamikbereich von bis zu 130 dB. Eine einzelne Nervenfasers kann diesen großen Dynamikbereich nicht codieren. Jedoch können mehrere Nervenfasern mit unterschiedlichen Schwellintensitäten diesen Dynamikbereich abdecken. Tatsächlich werden bei zunehmender Erregung einer Nervenfasers auch benachbarte Nervenfasern erregt.

Codierung der Frequenz

Ein Neuron im auditorischen System reagiert normalerweise auf einen Sinuston mit einer bestimmten Frequenz am intensivsten. Diese Frequenz heißt charakteristische Frequenz (*characteristic frequency*) des Neurons.

Das Neuron reagiert i.d.R. aber auch auf andere Frequenzen, nur weniger stark. Man kann nun die Feuerrate als Funktion der Frequenz auftragen. Ein solches Diagramm heißt Isointensitätskurve (*isolevel or iso-intensity curve*), denn die Intensität des Stimulus bleibt konstant.

Man kann jedoch auch auftragen, welche Intensität des Stimulus bei welcher Frequenz notwendig ist, um das Neuron gerade soundso oft feuern zu lassen. Ein solches Diagramm heißt Tuningkurve (*tuning curve or isorate curve*). Oft trägt man einfach die Schwellintensität über der Frequenz auf.

Tuningkurven im auditorischen Nerv zeigen folgende Charakteristik: Auf einen relativ langsamen Abfall der Stimulusintensität bis zur charakteristischen Frequenz folgt zu höheren Frequenzen hin ein sehr schneller Anstieg der notwendigen Stimulusintensität.

Zweitonsuppression

Bisher war nur von Codierung der Lautstärke und Frequenz eines Sinustons die Rede. Interessante nichtlineare Phänomene können auftreten, wenn zwei Sinustöne gleichzeitig als Stimulus angeboten werden: Wird das Gehör nur mit einem reinen Ton gereizt, so wird die Nervenfaser mit der entsprechenden charakteristischen Frequenz gut auf diesen Reiz reagieren. Wird nun gleichzeitig ein anderer reiner Ton in einem bestimmten Frequenzabstand mit bestimmter Lautstärke angeboten, so kann die Feuerrate der Nervenfaser auch wieder abnehmen. Diesen Vorgang nennt man Zweitonsuppression (*two-tone suppression*).

Zweitonsuppression tritt schon auf der Ebene der Basilarmembran auf. Verantwortlich sind nichtlineare Effekte, bei denen wohl auch die äußeren Haarzellen eine Rolle spielen. [Zenner 1994]

2.2 Zentrales auditorisches System

Das zentrale auditorische System (*central auditory system*) sind die Strukturen im Gehirn, die zur Weiterverarbeitung des vom Hörnerven zur Verfügung gestellten neuronalen Codes dienen. Sie reichen vom Hörnerven über mehrere Kerne bis zum auditorischen Cortex. Welche Strukturen sind aber an der Verarbeitung des Schalls beteiligt? Stimulation des Ohrs mag ja zur Erregung fast sämtlicher neuronaler Systeme beitragen und letztlich zu einer Muskelaktivität etc. führen. Als zum zentralen auditorischen System gehörig werden deshalb nur solche Strukturen angesehen, die besonders stark oder in erster Linie durch akustische Reize stimuliert werden.

Als Kern (*nucleus*) bezeichnet man eine Ansammlung von Neuronen, die sich anatomisch oder physiologisch von den Neuronen in ihrer Umgebung abheben.

Die Kerne des auditorischen Systems sind durch Nervenfasern verbunden. Man nummeriert diese Nervenfasern nach der Zahl der synaptischen Verbindungen, die zwischen der Faser und der Cochlea liegen. Die Nervenfasern des Hörnervs werden somit als Fasern erster Ordnung bezeichnet, die den ersten Kern verlassenden Fasern als Fasern zweiter Ordnung u.s.w. Die den auditorischen Cortex erreichenden Fasern sind Fasern etwa vierter bis fünfter Ordnung.

Innerhalb der Kerne liegen Interneuronen, die Verbindungen zwischen den Neuronen innerhalb der Kerne herstellen. Manche Nervenfasern passieren auch nur auditorische Kerne ohne synaptische Verbindungen zu ihren Neuronen herzustellen.

Die Nervenfasern des Hörnervs kontaktieren nach Verlassen der Cochlea zunächst den Nucleus cochlearis. Vom Nucleus cochlearis führen Nervenfasern sowohl zum ipsilateralen als auch zum kontralateralen Olivenkomplex. Von hier führen Nervenfasern durch den Lemniscus lateralis zum Colliculus inferior, von hier wiederum zum Corpus geniculatum mediale und letztlich zum auditorischen Cortex. Dieser Überblick ist jedoch stark vereinfacht,

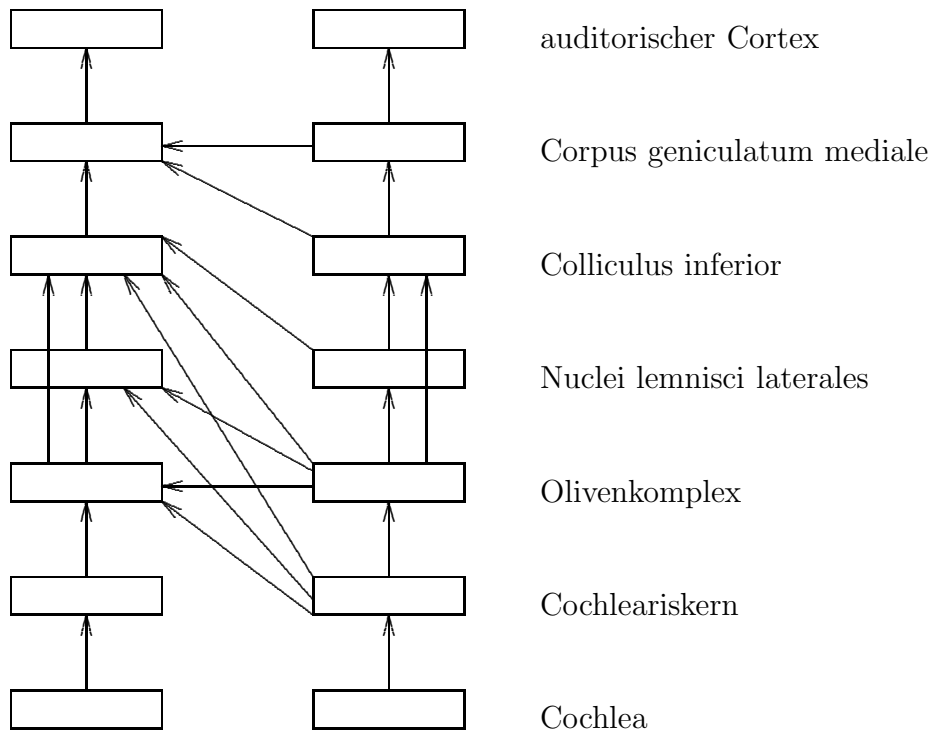


Abb. 2.1: Das zentrale auditorische System im Überblick (vereinfacht). Rechtecke bedeuten Kerne, Pfeile Nervenfaserbündel. Jedes Faserbündel, das von einem Kern der rechten Hälfte in einen Kern der linken Hälfte läuft, ist auch symmetrisch von links nach rechts vorhanden (nicht eingezeichnet). Es sind nicht alle Verbindungen eingetragen, insbesondere Verbindungen zu nicht-auditorischen Teilen des Gehirns sind weggelassen. Auch efferente Fasern sind nicht aufgeführt.

tatsächlich sind die auditorischen Kerne viel stärker untereinander verwoben (Abb. 2.1), was in den folgenden Kapiteln deutlicher werden sollte.

2.2.1 Nucleus cochlearis

Der Nucleus cochlearis (*cochlear nucleus*) bezieht seine neuronale Eingabe vom auditorischen Nerv.

Er ist etwa 3×5 mm groß und lässt sich in drei Abschnitte unterteilen, den Nucleus cochlearis anteroventralis, den Nucleus cochlearis posteroventralis und den Nucleus cochlearis dorsalis. Jeder dieser Abschnitte ist in Streifen oder Linien von Neuronen organisiert. Neuronen jedes dieser Streifen haben die selbe charakteristische Frequenz. Man spricht auch von tonotoper Organisation.

Im Nucleus cochlearis gibt es verschiedene Typen von Zellen mit jeweils anderen charakteristischen Antworten auf akustische Stimuli: „Primary-like“ Neuronen reagieren auf einen kurzen Sinuston ähnlich wie die Nervenfasern des Hörnervs mit einer zunächst kräftigen, dann schwächer werdenden Antwort. „On“-Neuronen reagieren nur auf das Anschalten des Stimulus, „Pauser“-Neuronen haben eine gewisse Latenzzeit, „Chopper“-Neuronen zeigen

eine oszillatorische Antwort. Diese Neuronentypen unterscheiden sich jeweils auch anatomisch. [Shepherd 1993]

Die Neuronen im Nucleus cochlearis dorsalis sind lateral inhibitorisch miteinander verknüpft [Zenner 1994, Møller 1983, Pickles 1982]. D.h. die Neuronen eines Frequenzbandes werden von Neuronen benachbarter Frequenzbänder gehemmt. Ein solches laterales inhibitorisches Netzwerk könnte helfen, das Spektrum eines Schalls zu verschärfen und Änderungen im Spektrum vor Hintergrundgeräuschen hervorzuheben. Die Inhibition wird vermutlich über Interneuronen im Nucleus cochlearis dorsalis selbst vermittelt [Pickles 1982].¹³ Inhibition ist von der Zweitonsuppression (siehe Kap. 2.1.4) sowohl von der Ursache als auch von der Wirkung klar zu unterscheiden: Zweitonsuppression tritt schon auf Basilmembranebene auf, Inhibition dagegen ist ein neuronaler Effekt. Zweitonsuppression setzt das Zusammenwirken zweier Töne voraus und kann niemals die spontane Feuerrate eines Neurons herabsetzen, während dies bei lateraler Inhibition durchaus möglich ist [Pickles 1982].

Die Nervenfasern aus dem Nucleus cochlearis enden sowohl im ipsilateralen als auch im kontralateralen Olivenkomplex, außerdem im kontralateralen Colliculus inferior.

Wozu der Nucleus cochlearis genau dient, ist wohl noch nicht endgültig geklärt, es findet jedoch sicherlich eine Vorverarbeitung der Daten aus dem Hörnerv statt [Zenner 1994].

2.2.2 Olivenkomplex

Der Olivenkomplex (*olivary complex*) wird von Nervenfasern aus beiden Cochleariskernen erregt. Hier wird also bereits Information aus beiden Ohren zusammen verarbeitet. Manche Zellen sprechen auf charakteristische Verzögerungen von Schall aus beiden Ohren an und dienen somit wohl der Richtungslokalisierung.

Der ipsilaterale Olivenkomplex projiziert sowohl in den ipsilateralen als auch in den kontralateralen Colliculus inferior.

2.2.3 Nuclei lemnisci laterales

Die Nuclei lemnisci laterales beziehen Nervenfasern aus den ipsi- und kontralateralen Cochleariskernen sowie aus beiden Olivenkomplexen [Møller 1983, Zenner 1994]. Sie werden als „verstreute Zellgruppen im Verlaufe des Lemniscus“ beschrieben [Kahle 1979].

Der Lemniscus lateralis (laterale Schleifenbahn) selbst ist ein Nervenfaserbündel, das die Fasern von den kontralateralen Cochleariskernen und dem ipsi- und kontralateralen Olivenkomplex vereint und in den Colliculus inferior mündet.

¹³Gerade die Inhibition wird durch eine Anästhesie leider stark unterdrückt. Deshalb ist sie u.U. in stark anästhesierten Tieren nur schwer nachzuweisen [Pickles 1982, Møller 1983]

Die Nuclei lemnisci laterales projizieren in den ipsi- und kontralateralen Colliculus inferior.

2.2.4 Colliculus inferior

Der Colliculus inferior (*inferior colliculus*) enthält Nervenendigungen sowohl von Nervenfasern aus dem ipsilateralen Olivenkomplex als auch aus dem kontralateralen Olivenkomplex sowie aus dem kontralateralen Cochleariskern.

Er ist tonotop organisiert. Außerdem sind einige Zellen nur monaural, werden also nur durch Stimulation eines Ohres erregt, während andere binaural sind, d.h. auf Stimulation beider Ohren reagieren. Dies ist plausibel, da im Colliculus inferior ja sowohl Nervenfasern aus dem binauralen Olivenkomplex als auch aus dem monauralen cochleären Kern ankommen.

Der Colliculus inferior projiziert in das Corpus geniculatum mediale.

2.2.5 Corpus geniculatum mediale

Der mittlere Kniehöcker oder Corpus geniculatum mediale (*medial geniculate body*) erhält Nervenfasern aus dem ipsilateralen Colliculus inferior und projiziert in den auditorischen Cortex. Das Corpus geniculatum mediale ist auch mit dem kontralateralen Corpus geniculatum mediale verbunden [Kahle 1979].

Die Antworten der Neuronen im Corpus geniculatum mediale werden als „sehr komplex“ bezeichnet, d.h. sie reagieren in komplizierter, noch nicht genau bekannter Weise auf akustische Stimuli.

Codierung der Frequenz

In den tonotop organisierten Teilen des Corpus geniculatum mediale reagieren viele Zellen sehr scharf auf eine bestimmte Frequenz. Manche zeigen jedoch auch breitere Tuningkurven oder Tuningkurven mit mehreren Maxima. Viele Neuronen reagieren auch inhibitorisch auf Frequenzbereiche nahe ihrer charakteristischen Frequenz.

Nur wenige Zellen zeigen phasenkorrelierte Antworten bei niedrigen Frequenzen.

Codierung der Schallintensität

Die Feuerrate eines Neurons in Abhängigkeit des Schalldruckpegels heißt, wie oben schon erwähnt, Raten-Pegel-Funktion. Die Raten-Pegel-Funktion nimmt zunächst mit steigendem Pegel zu und sättigt dann bei einer maximalen Rate, oder sie nimmt nach Erreichen eines Maximums wieder ab. In letzterem Fall heißt der Schalldruckpegel mit maximaler Rate der für das Neuron charakteristische Schalldruckpegel. Manche Neuronen zeigen jedoch auch mehrere Maxima in ihrer Raten-Pegel-Funktion. Typische Feuerraten

sind 10–100 Hz. Neuronen mit nicht monotoner Raten-Pegel-Funktion sind sehr häufig, etwa 30–70 %.

Veränderungen der Schallintensität können das zeitliche Antwortverhalten der Neuronen auf komplizierte Weise beeinflussen.

Neuronen mit monotoner Raten-Pegel-Funktion reagieren eher auf Breitbandrauschen. Ihre Latenzzeit verändert sich monoton mit der Schallintensität.

Neuronen mit nicht monotoner Raten-Pegel-Funktion dagegen reagieren weniger stark auf Breitbandrauschen. Ihre Latenzzeit ist länger und verändert sich nicht monoton mit der Schallintensität.

Nicht monotone Raten-Pegel-Funktionen müssen ihre Ursache in inhibitorischen Mechanismen haben, da alle auditorischen Nervenstränge monotone Raten-Pegel-Funktionen haben.

2.2.6 Auditorischer Cortex

Die meisten Studien wurden an Hauskatzen durchgeführt. Man kann den auditorischen Cortex in mehrere Felder unterteilen. Einige sind tonotop organisiert. Vermutlich sind die verschiedenen Felder auf verschiedene Wahrnehmungsbereiche spezialisiert. Wie die Spezialisierung genau aussieht, ist jedoch weitgehend unbekannt.

Ein noch nicht ausreichend untersuchtes Problem ist der Einfluss der Narkose auf das Antwortverhalten auditorischer Neuronen. Zusammenfassend kann man jedoch sagen, dass mit zunehmender Narkosetiefe die Antwortintensität abnimmt, die Latenzzeit zunimmt, die Schärfe der Tuningkurven abnimmt und die tonotope Organisation ausgeprägter wird [Yost 1994, Shepherd 1993]. Letzteres hat vermutlich seine Ursache darin, dass der Cortex unter Narkose überhaupt relativ wenig in die auditorische Schallverarbeitung eingreift.

Codierung der Frequenz

Der auditorische Cortex (*auditory cortex*) ist im Wesentlichen tonotop organisiert, wobei mehrere tonotope Felder existieren. Die Tuningkurven von Neuronen gleicher charakteristischer Frequenz unterscheiden sich allerdings in ihrer Breite, wobei manche so scharf wie diejenigen im Hörnerv sind.

Die tonotopen Felder sind bei den meisten Tieren von einem peripheren auditorischen Band umgeben, das nicht oder fast nicht tonotop organisiert ist. Das Antwortverhalten der Zellen hier wird auch als „schwach“ und „unsicher“ beschrieben. Außerdem haben die Zellen i.d.R. längere Latenzzeiten (50 ms im Ggs. zu weniger als 20 ms im primären auditorischen Cortex).

Die Neuronen des primären auditorischen Cortex reagieren exzitatorisch auf einen relativ engen Frequenzbereich. Einige Zellen reagieren darüberhinaus inhibitorisch auf angrenzende Frequenzbereiche.

In tonotop organisierten Feldern des primären auditorischen Cortex wird

eine Frequenz auf ein ganzes Isorefrequenzband abgebildet. Es liegt daher nahe, dass irgendein anderer Parameter entlang des Isorefrequenzbandes codiert wird. Im primären auditorischen Cortex nimmt die Schärfe der Tunigkurven entlang des Isorefrequenzbandes ab. Gleichzeitig nimmt die Antwort auf Breitbandrauschen zu. Dies wird dadurch erklärt, dass der charakteristischen Frequenz benachbarte Frequenzen inhibitorisch auf das Neuron wirken. Diese Inhibition ist umso stärker, je schärfer die Tuningkurve des Neurons ist.

Codierung der Schallintensität

Auch im auditorischen Cortex gibt es Neuronen mit monotoner und nicht monotoner Raten-Pegel-Funktion.

Zellen mit nicht monotoner Raten-Pegel-Funktion reagieren oft weniger stark auf Breitbandrauschen als auf reine Töne. Dies wird durch inhibitorische Bereiche für Frequenzen nahe der charakteristischen Frequenz sowie durch Inhibition oberhalb einer bestimmten Schallintensität erklärt.

Theorien über die Codierung der Schallintensität

Das Gehör hat einen Dynamikbereich von ca. 100 dB.¹⁴ Es gibt zwei Hypothesen, wie dieser große Dynamikbereich im Gehör codiert wird:

Zellen mit monotoner Raten-Pegel-Funktion haben einen Dynamikbereich von etwa 40 dB, in Einzelfällen allerdings auch größer. Solche Zellen haben jedoch verschiedene Schwellwerte, d.h. verschiedene Schallintensitäten, ab denen die Zelle merklich zu feuern beginnt. Eine niedrige Schallintensität würde also nur Zellen mit kleinem Schwellwert anregen, während höhere Schallintensitäten Zellen mit immer größeren Schwellwerten erregen würden. Zellen mit kleinem Schwellwert wären dann zwar schon in Sättigung, Zellen mit größerem Schwellwert würden jedoch innerhalb ihres Dynamikbereichs arbeiten. Auf diese Weise könnte ein viel größerer Dynamikbereich codiert werden als von einer Zelle allein.

Zellen mit nicht monotoner Raten-Pegel-Funktion reagieren auf eine bestimmte Schallintensität optimal. Diese optimalen Schallintensitäten sind über den gesamten Dynamikbereich des Gehörs von 0 bis 100 dB verteilt. Somit ist eine Ortskodierung der Schallintensität denkbar. Eine solche „ampliotope“ Organisation wurde allerdings außer in der Schnurrbartfledermaus in keinem anderen Säuger beobachtet¹⁵.

Temporale Codierung und komplexe Reize

Viele Neuronen im auditorischen Cortex sind besonders sensitiv für Veränderungen entweder der Frequenz oder der Lautstärke eines Sinustons. Manche reagieren etwa nur auf eine Erhöhung der Frequenz, andere nur auf eine

¹⁴Berücksichtigt man die unterschiedliche Empfindlichkeit des Gehörs bei verschiedenen Frequenzen, so kommt man auf einen Dynamikbereich von etwa 130 dB, uns interessiert hier jedoch nur die Codierung der Schallintensität einer Frequenz.

¹⁵Man findet jedoch beispielsweise in der Schleiereule eine ampliotope Organisation.

Erniedrigung, manche werden nur beim Einschalten eines Sinustons aktiv, andere nur beim Ausschalten etc. Ein Drittel der Neuronen kann nur durch solche Veränderungen im Stimulus erregt werden. [Popper 1992, Yost 1994]

Eine bloße Unterscheidung verschiedener Frequenzen ist Versuchspersonen bzw. -tieren auch ohne auditorischen Cortex möglich. Dies gilt jedoch nicht für Aufgaben, die ein vorheriges Lernen erfordern. Viele Neuronen reagieren auf charakteristische Merkmale komplexer Schalle wie etwa bestimmte Tierlaute. Verarbeitung von Sprache erfordert deshalb den auditorischen Cortex. Der auditorische Cortex enthält vermutlich auch ein auditorisches Kurzzeitgedächtnis, d.h. akustische Stimuli werden gespeichert und mit nachfolgenden zusammen ausgewertet. Somit dient der auditorische Cortex auch der Verarbeitung komplexer Zeitmuster in Schallen. Weiter unterstützt er vermutlich die Schallortung und steuert die selektive Aufmerksamkeit zu bestimmten Schallquellen. [Zenner 1994]

Kapitel 3

Modelle für das auditorische System

Hier sollen einige Modelle für Komponenten des auditorischen Systems vorgestellt werden.

3.1 Modelle für das periphere auditorische System

3.1.1 Cochlea

Die Cochlea korrekt zu modellieren ist sehr aufwändig. Vermutlich gelingt dies letztlich nur mit dreidimensionalen Modellen der beteiligten Membranen, evtl. ist sogar eine Modellierung der Flüssigkeit notwendig [Keidel 1995, Holmes, Lester 1981]. Auch sind die aktiven Prozesse, verursacht durch die äußeren Haarzellen, zu berücksichtigen. Stellt man geringere Ansprüche an die Exaktheit der Modellierung, so genügt es, die Cochlea als linearen Filter aufzufassen. Im Folgenden werden einige lineare Cochleamodelle vorgestellt.

Erregung

Es scheint zunächst naheliegend, die Cochlea einfach anhand der in Kap. 1.5.1 definierten Erregung zu modellieren. Das Amplitudenquadrat der Auslenkung der Basilarmembran an der Stelle x wäre dann die entsprechende Erregung $E(z(x))$, wobei $z(x)$ die der Stelle x entsprechende Tonheit ist. Dies ist jedoch aus folgenden Gründen problematisch:

- Um die Erregung zu bestimmen, bräuchte man die Schallintensitätsdichte $I(z)$. Hierzu müsste man in regelmäßigen Zeitabschnitten eine Fouriertransformation des Eingangssignals berechnen. Dies wäre sehr aufwändig. Wie groß sollten außerdem die Zeitabschnitte sein?
- Nach obigem System hätte man nur in relativ großen Zeitabschnitten eine Erregung zur Verfügung. Ein abruptes Ein- und Ausschalten der

Erregung an den verschiedenen Orten der Cochlea wäre die Folge. Um dies zu vermeiden, müsste man in jedem Zeitschritt (etwa 1 ms) eine Fouriertransformation über den letzten Zeitabschnitt durchführen. Das wäre ein noch größerer Aufwand.

- Die exakte zeitliche Struktur der Auslenkung der Basilarmembran, wie man sie z. B. als Eingangssignal für die daran anschließenden Haarzellenmodelle benötigt, wäre noch in keiner Weise modelliert.

Falls das cochleäre Eingabesignal jedoch einfach ist, z. B. nur Tiefpassrauschen o. ä., so ist der oben beschriebene Weg durchaus gangbar. In diesem Fall kann man allerdings auch auf eine Modellierung der inneren Haarzellen weitgehend verzichten, denn deren Feuerwahrscheinlichkeit lässt sich dann ebenfalls im Vorhinein leicht berechnen¹.

Lineare Schwinger

Die einzelnen Orte der Basilarmembran werden als unabhängige gedämpfte lineare Schwinger modelliert. Die Gleichung für die Auslenkung $y(x, t)$ an einer Stelle x der Basilarmembran ist dann

$$\ddot{y}(x, t) + 2\beta(x)\dot{y}(x, t) + \omega_0(x)^2 y(x, t) = \frac{1}{m(x)} F(t),$$

wobei β der Abklingkoeffizient, ω_0 die Eigenfrequenz des ungedämpften Schwingers, m die Masse des Schwingers und F die den Schwinger anregende Kraft (proportional zum Schalldruck des äußeren Schalls), bedeuten [Kuchling 1988].

Die Impulsantwort, d.h. die auf $F(t)/m = \delta(t)$, ist

$$y(t) = \frac{1}{\sqrt{\omega_0^2 - \beta^2}} e^{-\beta t} \sin\left(\sqrt{\omega_0^2 - \beta^2} t\right) H(t).$$

Hierbei bedeutet $H(t)$ die HEAVISIDE'sche Sprungfunktion.

Die Antwort auf einen Cosinus, d.h. die Antwort auf $F(t)/m = \cos(\omega t)$, ist im stationären Fall

$$y(t) = A_{\omega_0}(\omega) \cos(\omega t - \phi),$$

wobei die Amplitude

$$A_{\omega_0}(\omega) = \frac{1}{\sqrt{4\beta^2\omega^2 + (\omega^2 - \omega_0^2)^2}}$$

und die Phasenverschiebung

$$\phi(\omega) = \begin{cases} \arctan \frac{2\beta\omega}{\omega_0^2 - \omega^2} & \text{falls } \omega < \omega_0 \\ \pi/2 & \text{falls } \omega = \omega_0 \\ \arctan \frac{2\beta\omega}{\omega_0^2 - \omega^2} + \pi & \text{falls } \omega > \omega_0 \end{cases}$$

¹Man könnte etwa die Feuerrate der inneren Haarzellen gleich dem Erregungspegel setzen.

betragen.

Die Amplitude $A(\omega)$ ist bei der Resonanzfrequenz $\omega_R = \sqrt{\omega_0^2 - 2\beta^2}$ maximal.

Wählt man $\beta = \frac{1}{2\text{ms}} = 500\text{ Hz}$,² so ergibt sich als Halbwertsbreite³ des Filters für $\beta \ll \omega_0$ ungefähr $\beta/2\pi \approx 80\text{ Hz}$. Die Halbwertsbreite der cochleären Filter beträgt jedoch nur etwa 10–30 Hz,⁴ d.h. die Frequenzauflösung der so wie oben beschrieben modellierten Cochlea wäre viel zu schlecht.

Lineare Filter vierter Ordnung

Die Cochlea wird nach [Sottek 1993] durch lineare Filter vierter Ordnung⁵ mit der Impulsantwort

$$(3.1) \quad y(t) = \frac{t^3}{3\tau^4} e^{-t/\tau} \cos(\omega_0 t) H(t)$$

beschrieben (siehe Abb. 3.1).

Die Antwort auf einen Cosinus ist im stationären Fall

$$y(t) = A_{\omega_0}(\omega) \cos(\omega t - \phi),$$

wobei die Amplitude

$$A_{\omega_0}(\omega) \approx \frac{1}{[1 + \tau^2(\omega - \omega_0)^2]^2}$$

beträgt⁶.

Die Halbwertsbreite dieses Filters ist ungefähr $0,07/\tau$. Mit $\tau = 2\text{ ms}$ ergibt sich eine Halbwertsbreite von 35 Hz. Das Gehör hat zwar eine noch bessere Leistung, dieser Wert ist jedoch akzeptabel.

²Man könnte eine beliebig gute Frequenzauflösung erreichen, indem man die Dämpfung β beliebig klein macht. Man muss jedoch auch die schnelle Reaktionszeit und damit die gute Zeitauflösung des Gehörs berücksichtigen. Deshalb sollten die Anstiegszeiten der verwendeten Filter relativ klein sein. Fußnote 7 auf Seite 28 spricht ebenfalls für sehr kurze Einschwingzeiten.

³Der Abstand zur Resonanzfrequenz $f_R = 2\pi\omega_R$, bei dem das Amplitudenquadrat $A(\omega)^2$ auf die Hälfte des Wertes bei der Resonanzfrequenz abgefallen ist.

⁴Nach Kap. 1.5.1 ist die Steigung der unteren Flanke des Erregungspegels etwa 27 dB/Bark, d.h. 3 dB pro 0,1 Bark. Die untere Halbwertsbreite der Erregung beträgt also etwa 0,1 Bark oder 10 Hz bei Frequenzen bis 500 Hz. Die Steigung der oberen Flanke beträgt etwa 10 dB/Bark, was zu einer oberen Halbwertsbreite von etwa 30 Hz führt.

⁵Vierter Ordnung deshalb, da die Impulsantwort durch eine Differenzialgleichung vierter Ordnung beschrieben werden kann. Es handelt sich um einen sogenannten minimalphasigen Bandpassfilter vierter Ordnung [Unbehauen 1997]. Der Betrag der Übertragungsfunktion dieser Filter stimmt im Durchlassbereich gut mit den von R. D. PATTERSON 1982 angegebenen ROEX-Filtern (*rounded exponential filter*) zur Beschreibung der Erregung im Spektralbereich überein [Sottek 1993]. Das Betragsquadrat der Übertragungsfunktion der ROEX-Filter lautet $\tilde{y}(f) = (1 + 4|f - f_0|/\Delta f) e^{-4|f - f_0|/\Delta f}$, wobei $\Delta f = \Delta f(f_0)$ die frequenzabhängige „äquivalente Rechteckbandbreite“ $\tilde{y}(f_0)^{-1} \int_{-\infty}^{\infty} \tilde{y}(f) df$ der Filter ist. PATTERSON gab allerdings keine Phase für die ROEX-Filter an, so dass eine Modellierung im Zeitbereich mit Hilfe von ROEX-Filtern nicht ohne weiteres möglich ist.

⁶Die Fouriertransformierte von (3.1) ist $\tilde{y}(\omega) = \frac{1}{[1 + i\tau(\omega - \omega_0)]^4} + \frac{1}{[1 + i\tau(\omega + \omega_0)]^4}$, wobei der zweite Summand gegenüber dem ersten im Durchlassbereich ($\omega \approx \omega_0$) vernachlässigt werden kann, da $\tau\omega_0 \gg 1$.

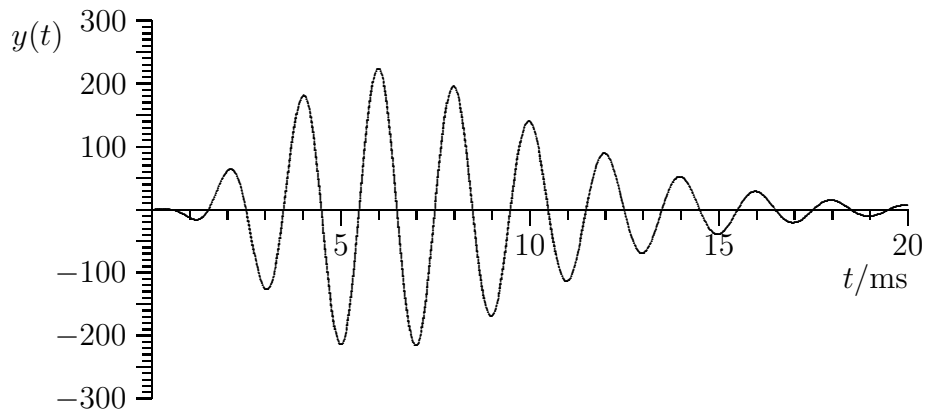


Abb. 3.1: Impulsantwort der Stelle der Cochlea mit $f_0 = \omega_0/2\pi = 500$ Hz, wobei $\tau = 2$ ms.

Je höher die Bestfrequenz ω_0 der cochleären Filter, desto breiter wird deren Durchlassbereich auf der Frequenzskala, wenn sie auch auf der Barkskala gleich breit bleiben (vgl. Kap. 1.5.1). Dies wird durch die Beziehung

$$(3.2) \quad \tau(\omega_0) = \begin{cases} 2 \text{ ms} & \text{falls } f_0 \leq 500 \text{ Hz} \\ 2 \text{ ms} \cdot 500 \text{ Hz}/f_0 & \text{falls } f_0 > 500 \text{ Hz} \end{cases}$$

modelliert, wobei $f_0 = \omega_0/2\pi$.

Abb. 3.2–3.5 vergleichen die drei oben beschriebenen Cochleamodelle im Frequenzbereich. Man beachte, dass die cochleären Filter in der Realität asymmetrisch sind, während sie im Modell des linearen Filters vierter Ordnung symmetrisch sind⁷.

⁷Die cochleäre Erregung ist allerdings auch im Modell der linearen Filter vierter Ordnung etwas asymmetrisch. Dies kommt daher, dass die Bandbreite der Filter wegen (3.2) mit steigender Bestfrequenz ω_0 zunimmt [Sottek 1993]. Die Asymmetrie ist allerdings nicht so stark wie in der Realität, wie Abb. 3.4 zeigt. Hier irrt [Sottek 1993].

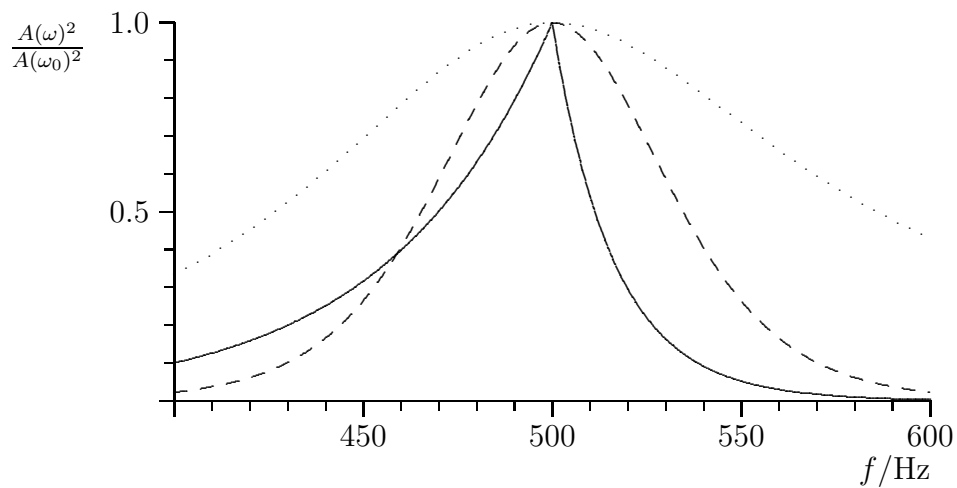


Abb. 3.2: Verlauf des cochleären Filters mit der Bestfrequenz $f_0 = 500$ Hz in verschiedenen Cochleamodellen. Aufgetragen ist das auf die Bestfrequenz bezogene Amplitudenquadrat $A(\omega)^2/A(\omega_0)^2$ in Abhängigkeit von der Frequenz $f = \omega/2\pi$.

Durchgezogen: Erregungsmodell nach Kap. 1.5.1.

Gestrichnet: Lineare Filter vierter Ordnung.

Punktiert: Lineare Schwinger nach Kap. 3.1.1.

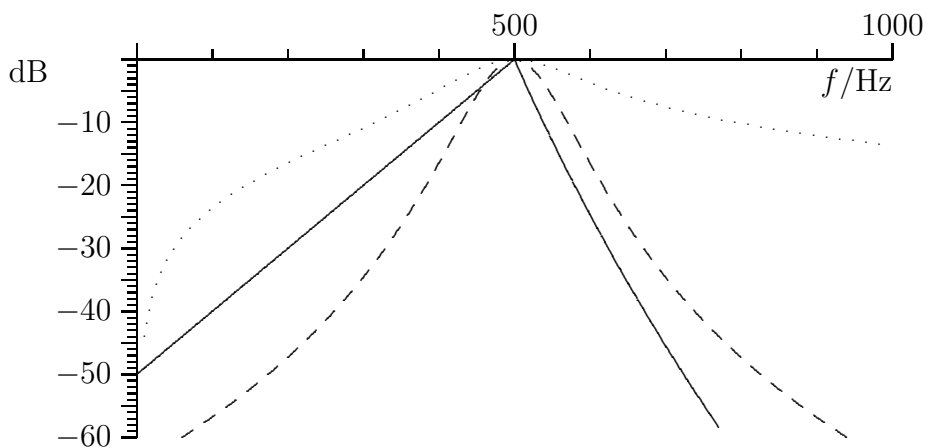


Abb. 3.3: Wie Abb. 3.2, nur als Pegel.

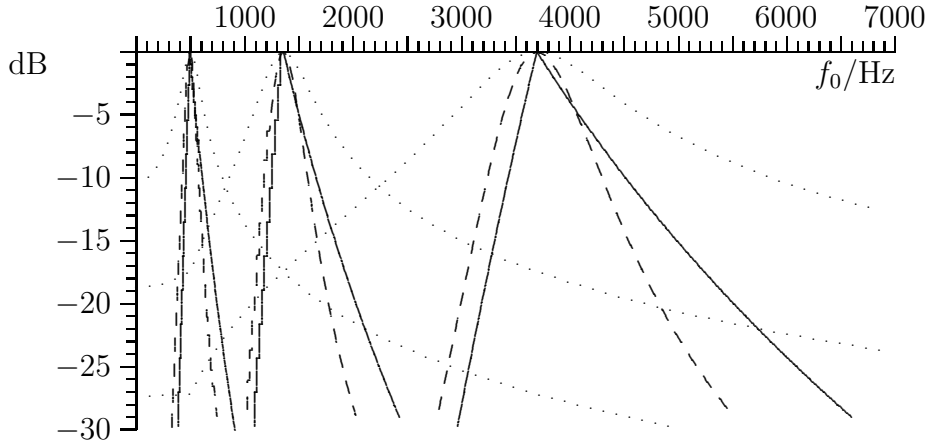


Abb. 3.4: Verlauf der cochleären Erregung in verschiedenen Cochleamodellen. Aufgetragen ist der Erregungspegel $10 \log A_{\omega_0}(\omega)^2 / A_{\omega_0}(\omega_0)^2$ der cochleären Filter bei Stimulation mit einem reinen Ton der Tonheit 5, 10 bzw. 15 Bark in Abhängigkeit der Bestfrequenz $f_0 = \omega_0 / 2\pi$ der Filter.

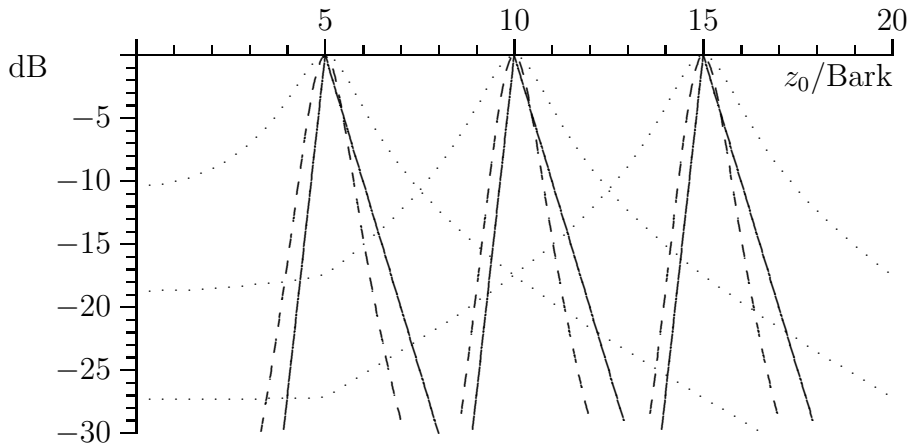


Abb. 3.5: Wie Abb. 3.4, nur ist die Bestfrequenz z_0 in der Tonheit-Skala aufgetragen. Die cochleären Erregungsmuster erscheinen somit gleich breit.

Um die Auslenkung der Basilarmembran zu berechnen, müsste man nun das Eingangssignal der Cochlea, d.h. den Schall, mit der Impulsantwort falten. Dies wäre für jede Stelle x der Cochlea durchzuführen. Alternativ könnte man das zur Impulsantwort führende Differenzialgleichungssystem integrieren. Günstiger ist jedoch folgender Weg: Berechne aus der Antwort $y(t)$ zur Zeit t und einigen Hilfsvariablen zur Zeit t die Antwort zur Zeit $t + \Delta t$.

Dieses Konzept führt zu folgendem effizienten Algorithmus⁸:

⁸Mit $\bar{y}(t) := \frac{1}{3\tau^4} t^3 e^{ct}$, wobei $c := -1/\tau + i\omega_0$, ist $y(t) = \Re(\bar{y})$. Weiter ist

$$\begin{aligned} \bar{y}(t + \Delta t) &= \frac{1}{3\tau^4} (t^3 + 3t^2\Delta t + 3t\Delta t^2 + \Delta t^3) e^{c(t+\Delta t)} \\ &= \left(\bar{y}(t) + h_1(t)\Delta t + h_2(t)\Delta t^2 + \frac{1}{3}h_3(t)\Delta t^3 \right) e^{c\Delta t}, \end{aligned}$$

wobei $h_1(t) := \frac{1}{\tau^4} t^2 e^{ct}$, $h_2(t) := \frac{1}{\tau^4} t e^{ct}$ und $h_3(t) := \frac{1}{\tau^4} e^{ct}$.
Man sieht dann, dass

$$\begin{aligned}
h_3 &:= h_3 + \frac{1}{\tau^4} p(t) \Delta t \\
\bar{y} &:= \bar{y} e^{(-1/\tau + i\omega_0)\Delta t} \\
h_1 &:= h_1 e^{(-1/\tau + i\omega_0)\Delta t} \\
h_2 &:= h_2 e^{(-1/\tau + i\omega_0)\Delta t} \\
h_3 &:= h_3 e^{(-1/\tau + i\omega_0)\Delta t} \\
\bar{y} &:= \bar{y} + \left[\left(\frac{1}{3} h_3 \Delta t + h_2 \right) \Delta t + h_1 \right] \Delta t \\
h_1 &:= h_1 + (h_3 \Delta t + 2h_2) \Delta t \\
h_2 &:= h_2 + h_3 \Delta t \\
y &:= \Re(\bar{y})
\end{aligned}$$

Hierbei ist $p(t)$ der Schalldruck des Schalls am Trommelfell. \Re bezeichnet den Realteil. Die Hilfsvariablen h_1 bis h_3 und \bar{y} werden zur Zeit $t = 0$ auf Null gesetzt. Wird h_3 zur Zeit $t = 0$ auf $1/\tau^4$ gesetzt, so erhält man in y die Impulsantwort. Aktualisiert man daher in jedem Zeitschritt Δt die Hilfsvariable h_3 wie angegeben, so integriert der Algorithmus die Impulsantworten automatisch auf.

Beispiele für die Reaktion einer cochleären Filterbank im Modell der linearen Filter vierter Ordnung zeigen Abb. 3.6 für einen Tiefpass und 3.7 für einen Sinuston als Stimulus. Bei 100 ms sieht man einen Abschalteneffekt, bei dem alle Filter kurz erregt werden. Dieser ist als „Klicken“ wahrnehmbar. Man sieht auch, dass die Reaktionszeit der cochleären Filter mit zunehmender Bestfrequenz schneller wird (vgl. (3.1) und 3.2), d.h. Filter mit hoher Bestfrequenz relaxieren nach Erregung schneller als solche mit niedrigerer Bestfrequenz.

$$\begin{aligned}
h_1(t + \Delta t) &= (h_1(t) + 2h_2(t)\Delta t + h_3(t)\Delta t^2) e^{c\Delta t} \\
h_2(t + \Delta t) &= (h_2(t) + h_3(t)\Delta t) e^{c\Delta t} \\
h_3(t + \Delta t) &= h_3(t) e^{c\Delta t}.
\end{aligned}$$

Es ist $\bar{y}(0) = h_1(0) = h_2(0) = 0$ und $h_3(0) = 1/\tau^4$. Initialisiert man also $h_3(t)$ mit $1/\tau^4$, so erhält man durch Auswertung der angegebenen Gleichungen in $\Re(\bar{y}(t))$ die Impulsantwort $y(t)$ in Zeitschritten Δt . Da $y(t)$ linear vom Startwert in $h_3(t)$ abhängt, addiert das Verfahren Impulsantworten automatisch auf. Einige numerische Vereinfachungen (HORNER-Schema zur Berechnung der Polynome, geschickte Reihenfolge der Berechnungen) ergeben den angegebenen Algorithmus. Dass sich der Algorithmus in der oben beschriebenen Weise erstellen lässt, liegt daran, dass es sich bei der Impulsantwort um ein Polynom in t mal $e^{i\omega_0 t}$ handelt, also um die Lösung einer linearen Differenzialgleichung mit konstanten Koeffizienten. Eine direkte Lösung dieser Differenzialgleichung hat sich als numerisch instabil erwiesen.

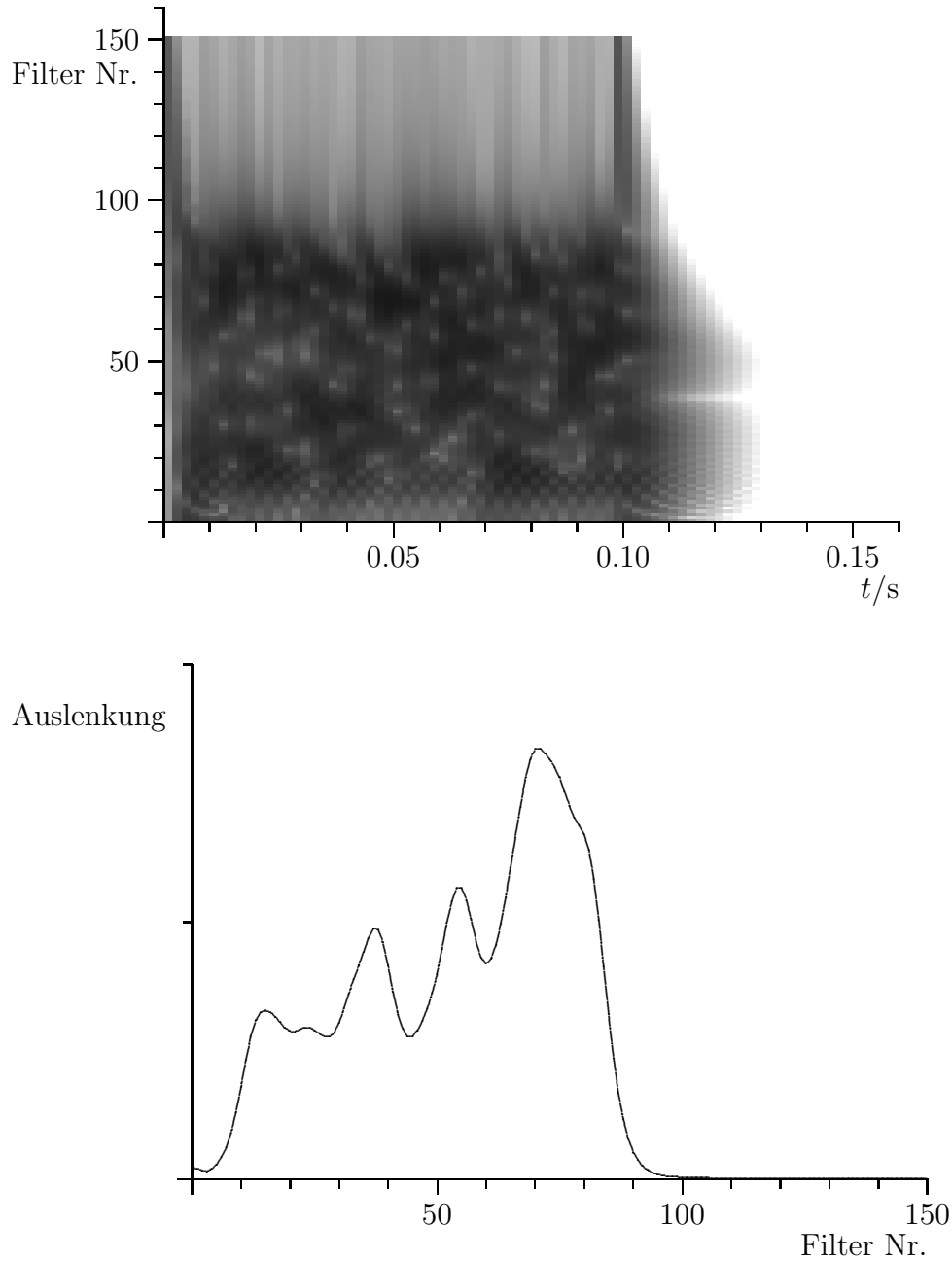


Abb. 3.6: Die obere Grafik zeigt den Logarithmus der jeweils über 20 ms gemittelten quadratischen Auslenkung der Cochlea im Modell linearer Filter vierter Ordnung. Nach rechts ist die Zeit in s aufgetragen, nach oben die Nummer der cochleären Filter. Simuliert wurden 150 cochleäre Filter mit Bestfrequenzen von 0 bis 15 Bark in Abständen von 0,1 Bark. Die untere Grafik zeigt die über die gesamte Zeit von 150 ms gemittelte quadratische Auslenkung der cochleären Filter. Stimulus war Tiefpassrauschen von 100 bis 1000 Hz.

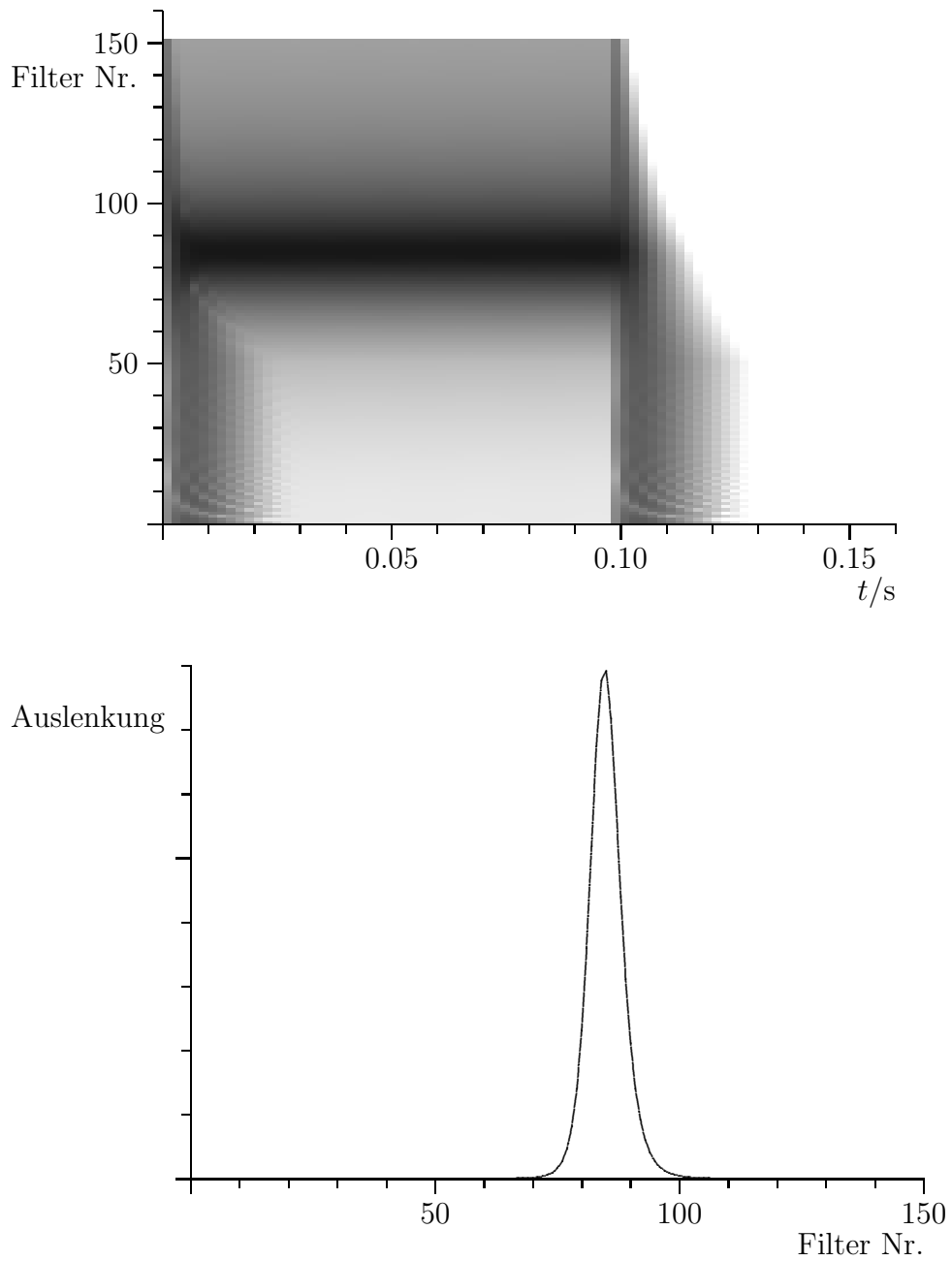


Abb. 3.7: Wie Abb. 3.6, nur war der Stimulus diesmal ein Sinuston mit Frequenz 1000 Hz.

3.1.2 Innere Haarzellen

„State-Partition“-Modell

Die folgende Beschreibung dieses Modells lehnt sich an [Cooke 1993] an. Das „State-Partition“-Modell geht davon aus, dass es in den Haarzellen viele Vorratsbehälter (*release sites*) gibt, die Neurotransmitter ausschütten können (siehe Abb. 3.8).

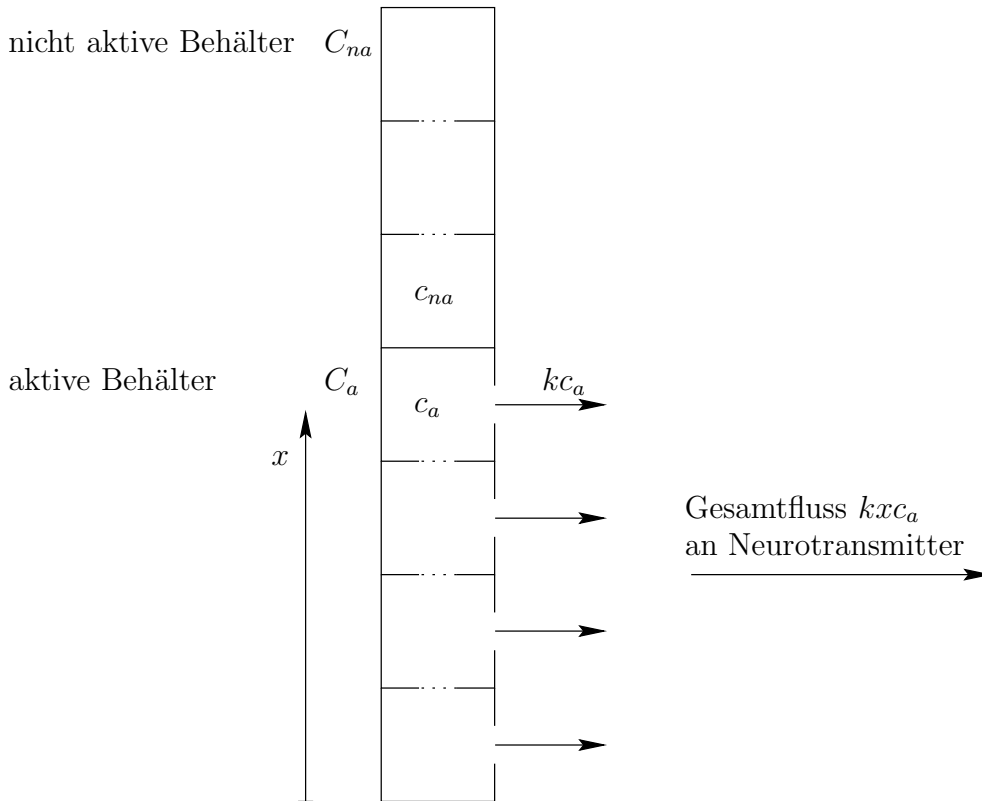


Abb. 3.8: Das „State-Partition“-Modell für innere Haarzellen. Kästen bedeuten Behälter mit Neurotransmitter. Die nach rechts offenen Behälter sind aktiv und schütten Neurotransmitter aus. Die von den Behältern ausgehenden Pfeile bedeuten Neurotransmitterströme. Im eigentlichen Modell sind alle Behälter voneinander getrennt, d.h. die gepunkteten Sperren sind geschlossen. In der im Text beschriebenen Näherung denkt man sich die gepunkteten Sperren jedoch als offen. c_a sind die Füllmengen an Neurotransmitter in den aktiven Behältern, c_{na} die Füllmengen in den nicht aktiven Behältern. Die Füllmengen sind bei geschlossenen gepunkteten Sperren in jedem Behälter verschieden, andernfalls gleichen sich die Füllmengen aus. Der Ausstrom aus jedem aktiven Behälter ist proportional zur Füllmenge des Behälters. Die Gesamtmenge an Neurotransmitter in den aktiven bzw. nicht aktiven Behältern ist C_a bzw. C_{na} .

Jeder dieser Vorratsbehälter hat eine Reizschwelle, ab der er Neurotransmitter ausschüttet. Da die Vorratsbehälter alle verschiedene Schwellen haben, hängt der Anteil x der Vorratsbehälter, die gerade Neurotransmitter ausschütten, von der Reizintensität ab. Beispielsweise kann $x(t)$ proportional

zur Einhüllenden der Auslenkung der Basilarmembran sein. Diejenigen Vorratsbehälter, die Neurotransmitter ausschütten, heißen aktiv. Die Rate des ausgeschütteten Transmitters ist außerdem proportional zum Füllzustand c_a der aktiven Behälter. Die Rate des ausgeschütteten Neurotransmitters ist also proportional zu $xc_a(t)$.

Sowohl die aktiven als auch die nicht aktiven Behälter werden mit einer Rate proportional $1 - c$ aufgefüllt, wobei 1 der willkürlich festgelegte Soll-Füllzustand ist.

Falls die Reizintensität x konstant ist, lauten die Gleichungen für die Veränderung der Mengen an Neurotransmitter $C_a(t)$ bzw. $C_{na}(t)$ in den aktiven bzw. nicht aktiven Behältern

$$\begin{aligned}\dot{C}_a(t) &= -kxc_a(t) + lx(1 - c_a(t)) \\ \dot{C}_{na}(t) &= l(1 - x)(1 - c_{na}(t)).\end{aligned}$$

Hierbei ist x der Anteil der aktiven Behälter, $c_a(t)$ die Füllmenge in den aktiven Behältern und $c_{na}(t)$ die Füllmenge in den nicht aktiven Behältern. Die Größen k und l sind Proportionalitätskonstanten.

Die Menge an Neurotransmitter $C_a(t)$ bzw. $C_{na}(t)$ in den aktiven bzw. nicht aktiven Behältern ist

$$(3.3) \quad C_a = xc_a$$

$$(3.4) \quad C_{na} = (1 - x)c_{na}.$$

Dies gilt alles für den Fall, dass der Reiz und damit x konstant ist. Was passiert nun, wenn x steigt, d.h. wenn der Reiz zunimmt und damit mehr Behälter in den aktiven Zustand übergehen? Die „frisch“ in den aktiven Zustand gekommenen Behälter haben den Füllzustand c_{na} , da sie ja gerade vorher noch inaktiv waren. D.h. man müsste nun jeden dieser Behälter einzeln modellieren, da jeder Behälter seinen individuellen Füllzustand hat. Um die Modellierung nun zu vereinfachen wird angenommen, dass im Wesentlichen nur der durchschnittliche Füllzustand in den aktiven und nicht aktiven Behältern ausschlaggebend ist (Fall offener gepunkteter Sperren in Abb. 3.8). Erhöht sich also der Anteil x der aktiven Behälter, so kommt der Neurotransmitter in den Behältern, die aktiv geworden sind, einfach zum Pool aktiver Behälter hinzu. Mit Berücksichtigung einer Änderung von x gilt daher für die Änderungen der Mengen an Neurotransmitter in den Behältern:

$$\begin{aligned}\dot{C}_a(t) &= -kxc_a(t) + lx(1 - c_a(t)) + \begin{cases} \dot{x}(t)c_a & \text{für } \dot{x}(t) < 0 \\ \dot{x}(t)c_{na} & \text{für } \dot{x}(t) \geq 0 \end{cases} \\ \dot{C}_{na}(t) &= l(1 - x)(1 - c_{na}(t)) + \begin{cases} -\dot{x}(t)c_a & \text{für } \dot{x}(t) < 0 \\ -\dot{x}(t)c_{na} & \text{für } \dot{x}(t) \geq 0 \end{cases}\end{aligned}$$

Wegen (3.3) und deshalb

$$\begin{aligned}\dot{C}_a &= \dot{x}c_a + x\dot{c}_a \\ \dot{C}_{na} &= -\dot{x}c_{na} + (1-x)\dot{c}_{na}\end{aligned}$$

lauten die Gleichungen für die Füllzustände c_a und c_{na} letztlich

$$\begin{aligned}\dot{c}_a &= -kc_a + l(1 - c_a) + H(\dot{x}) \frac{\dot{x}}{x} (c_{na} - c_a) \\ \dot{c}_{na} &= l(1 - c_{na}) + (1 - H(\dot{x})) \frac{\dot{x}}{1 - x} (c_{na} - c_a).\end{aligned}$$

Hierbei bedeutet H die HEAVISIDE-Funktion.

Das State-Partition-Modell wurde entwickelt, um die Antwort der Haarzellen richtig wiederzugeben, falls $x(t)$ die Einhüllende des Stimulus, d.h. die Einhüllende der Auslenkung der Basilarmembran, ist. Man kann sich beispielsweise leicht überlegen, dass bei einem schlagartig eingeschalteten reinen Ton (d.h. $x(t)$ macht einen Sprung) die Antwort $kxc_a(t)$ der Haarzelle ebenfalls einen Sprung macht und dann auf einen konstanten Wert relaxiert. Die genaue Zeitstruktur der Antwort wird nicht modelliert. Insbesondere kann nicht modelliert werden, dass die Antwort der Haarzellen bei niedrigen Frequenzen noch phasenrichtig ist, während sie bei höheren Frequenzen zunehmend „verschmiert“.

Meddis-Modell

In [Sottek 1993] wird ein Modell für innere Haarzellen vorgestellt. Es beruht auf [Meddis 1986].

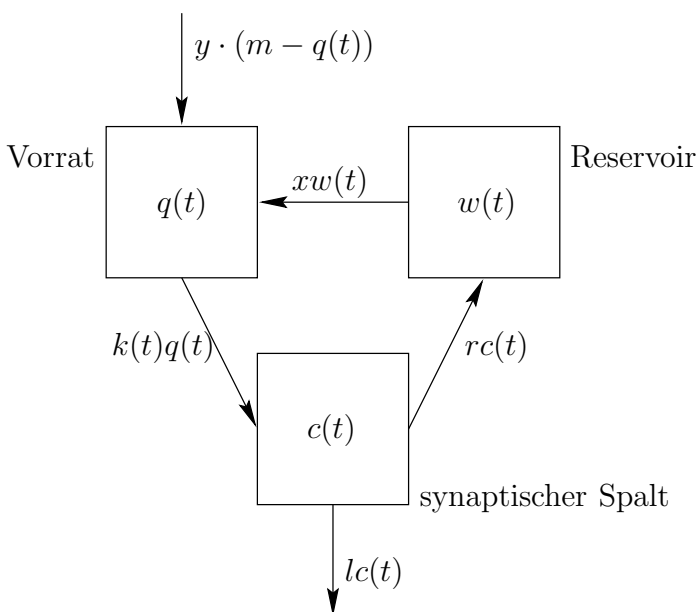


Abb. 3.9: Das Meddis-Modell für innere Haarzellen. Kästen bedeuten Behälter mit Neurotransmitter, Pfeile Neurotransmitterströme.

Siehe Abb. 3.9. Die innere Haarzelle besitzt einen Vorrat an Neurotransmittersubstanz $q(t)$. Neurotransmitter aus diesem Vorrat kann infolge der Permeabilität $k(t)$ der Zellmembran in den synaptischen Spalt zwischen Zelle und Nervenfaser diffundieren. Die Konzentration $c(t)$ des Transmitters im synaptischen Spalt bestimmt die Wahrscheinlichkeit für das Auftreten eines Aktionspotentials im postsynaptischen Nerv. Die Permeabilität $k(t)$ hängt vom Stimulus $s(t)$, d.h. von der Auslenkung der Basilarmembran, ab. Ein Teil des Neurotransmitters im synaptischen Spalt wird zersetzt bzw. verbraucht, ein anderer Teil dringt wieder in die Zelle ein und füllt ein „Reservoir“ $w(t)$. Der Vorrat $q(t)$ wird auch aus diesem Reservoir ergänzt.

Im einzelnen lauten die Gleichungen für die Veränderung der Transmitterkonzentrationen:

$$\begin{aligned} \dot{q}(t) &= y \cdot (m - q(t)) + xw(t) - k(t)q(t) \\ \dot{c}(t) &= k(t)q(t) - lc(t) - rc(t) \\ \dot{w}(t) &= rc(t) - xw(t) \\ k(t) &= \begin{cases} g \frac{s(t)+A}{s(t)+A+B} & \text{für } s(t) + A > 0 \\ 0 & \text{für } s(t) + A \leq 0 \end{cases} \end{aligned}$$

Hierbei sind $q(t)$ der Vorrat an Neurotransmitter, $c(t)$ die Neurotransmitterkonzentration im synaptischen Spalt, $w(t)$ der Neurotransmitter im Reservoir, $y \cdot (m - q(t))$ die Auffüllrate des Vorrats aus der Erzeugung von Neurotransmitter, $xw(t)$ die Auffüllrate des Vorrats aus dem Reservoir, $k(t)q(t)$ die Diffusionsrate durch die Zellmembran aus dem Vorrat in den synaptischen Spalt, $k(t)$ die Permeabilität der Zellmembran, $lc(t)$ der Verbrauch an Transmitter und $rc(t)$ die Diffusionsrate durch die Zellmembran aus dem synaptischen Spalt in das Reservoir.

Die Konstanten haben etwa die Werte in Tab. 3.1.

A	5
B	300
g	2000
y	5,05
l	2500
x	66,3
r	6580
h	2,5
m	1,0

Tab. 3.1: Parameter einer Nervenfaser mit hoher Spontanaktivität nach [Sottek 1993]. Zeiten werden in Sekunden gemessen, Neurotransmitterkonzentrationen in willkürlichen Einheiten.

Bekannt sind der Stimulus $s(t)$ sowie die Anfangsbedingungen $q(0)$, $c(0)$, $w(0)$. Hieraus lässt sich dann die Konzentration $c(t)$ des Neurotransmitters

im synaptischen Spalt berechnen. Die Wahrscheinlichkeit für das Auftreten eines Aktionspotentials ist dann nach [Sottek 1993]⁹

$$P = hc(t).$$

Man kann sich auch vorstellen, dass das postsynaptische Potential $h_{HNF}(t)$ in der entsprechenden Hörnervenfaser proportional zur Konzentration des Neurotransmitters im synaptischen Spalt ansteigt, etwa

$$(3.5) \quad \dot{h}_{HNF}(t) = hc(t) - \frac{1}{\tau_{HNF}} h_{HNF}(t),$$

wobei τ_{HNF} die Abklingzeit des postsynaptischen Potentials ist.

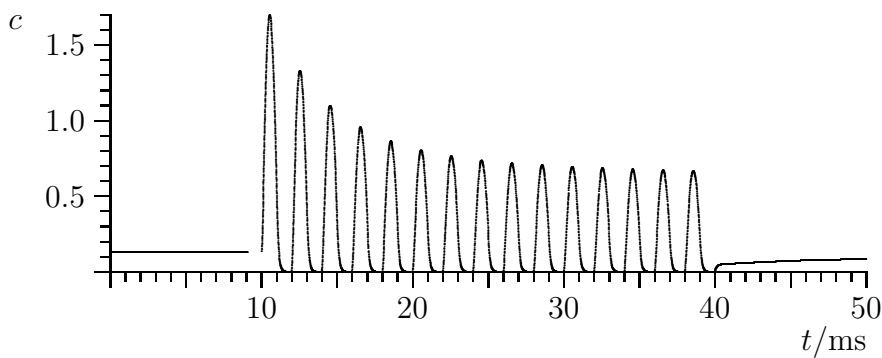


Abb. 3.10: Neurotransmitterkonzentration $c(t)$ nach dem Meddis-Modell bei einem von 10–40 ms dauernden reinen Ton der Frequenz $f = 500$ Hz mit Amplitude 50 als Stimulus $s(t)$.

3.1.3 Hörnerv

In der die Haarzelle kontaktierenden Hörnervenfaser bildet sich genau dann ein Aktionspotential, wenn das postsynaptische Potential h_{HNF} im Hörnerv nach (3.5) größer wird als eine Schwelle θ_{HNF} , d.h.

$$h_{HNF}(t) > \theta_{HNF}.$$

Jedes Aktionspotential zieht ein entsprechendes Refraktärpotential nach sich, so dass die Hörnervenfaser nicht gleich wieder feuern kann (siehe Kap. 4.3.2).

Abb. 3.11 fasst die bisher besprochenen Modelle zusammen.

Abb. 3.12 und 3.13 zeigen die Aktionspotentiale im simulierten Hörnerv bei einem Tiefpassrauschen bzw. einem reinen Ton als Stimulus wie in Abb. 3.6 und 3.7. Die Cochlea wurde als Filterbank mit linearen Filtern vierter Ordnung (Kap. 3.1.1), die Haarzellen mit dem Meddis-Modell (Kap. 3.1.2) simuliert.

⁹ P kann als Wahrscheinlichkeit, dass der Hörnerv im Zeitabschnitt $[t, t + dt]$ feuert interpretiert werden, wobei dt sehr klein ist. D.h. die Feuerrate des Hörnervs ist proportional zur Neurotransmitterkonzentration und die Aktionspotentiale sind poissonverteilt.

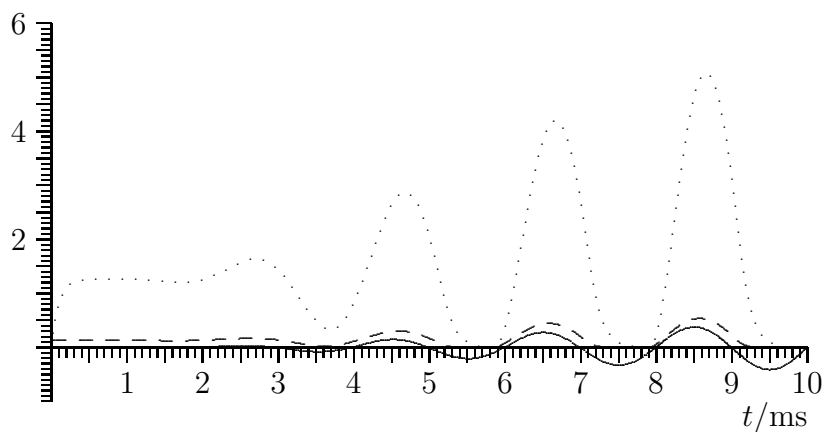


Abb. 3.11: Durchgezogen: Auslenkung $y(t)$ der Basilarmembran an der Stelle mit Bestfrequenz 500 Hz. Das Cochleamodell sind lineare Filter vierter Ordnung wie in Kap. 3.1.1 beschrieben.

Gestrichelt: Neurotransmitterkonzentration $c(t)$ nach dem Meddis-Modell (Kap. 3.1.2). Der die Haarzelle reizende Stimulus ist $s(t) = 50 y(t)$.

Gepunktet: Postsynaptisches Potential $h_{HNF}(t)$ in der angeschlossenen Hörnervenfaser nach (3.5) ($h = 250$, $\tau_{HNF} = 0,1 \text{ ms}$).

Stimulus ist ein reiner Ton der Frequenz 500 Hz, der bei $t = 0$ einsetzt.

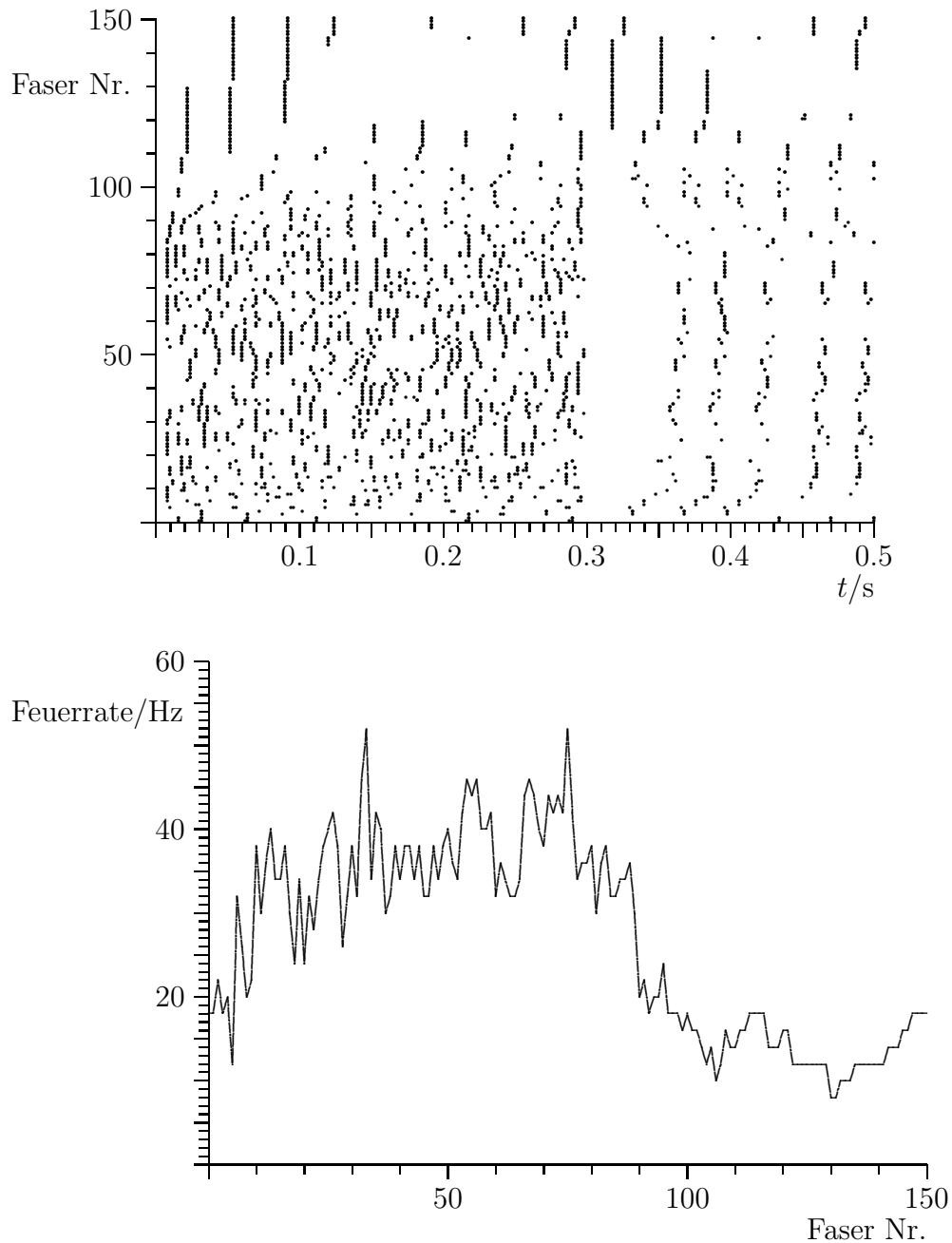


Abb. 3.12: Die obere Grafik zeigt die Aktionspotentiale in 150 Hörnervenfaseren mit Bestfrequenzen von 0 bis 15 Bark in Abständen von 0,1 Bark. Nach rechts ist die Zeit in s aufgetragen, nach oben die Nummer der Hörnervenfaser. Jeder Punkt bedeutet ein Aktionspotential zum jeweiligen Zeitpunkt in der jeweiligen Hörnervenfaser.

Die untere Grafik zeigt die mittlere Feuerrate jeder Hörnervenfaser in Hz. Simuliert wurde nach dem Meddis-Modell (Kap. 3.1.2) und nach (3.5) ($h = 250$, $\tau_{HNF} = 0,1 \text{ ms}$).

Stimulus war Tiefpassrauschen von 100 bis 1000 Hz.

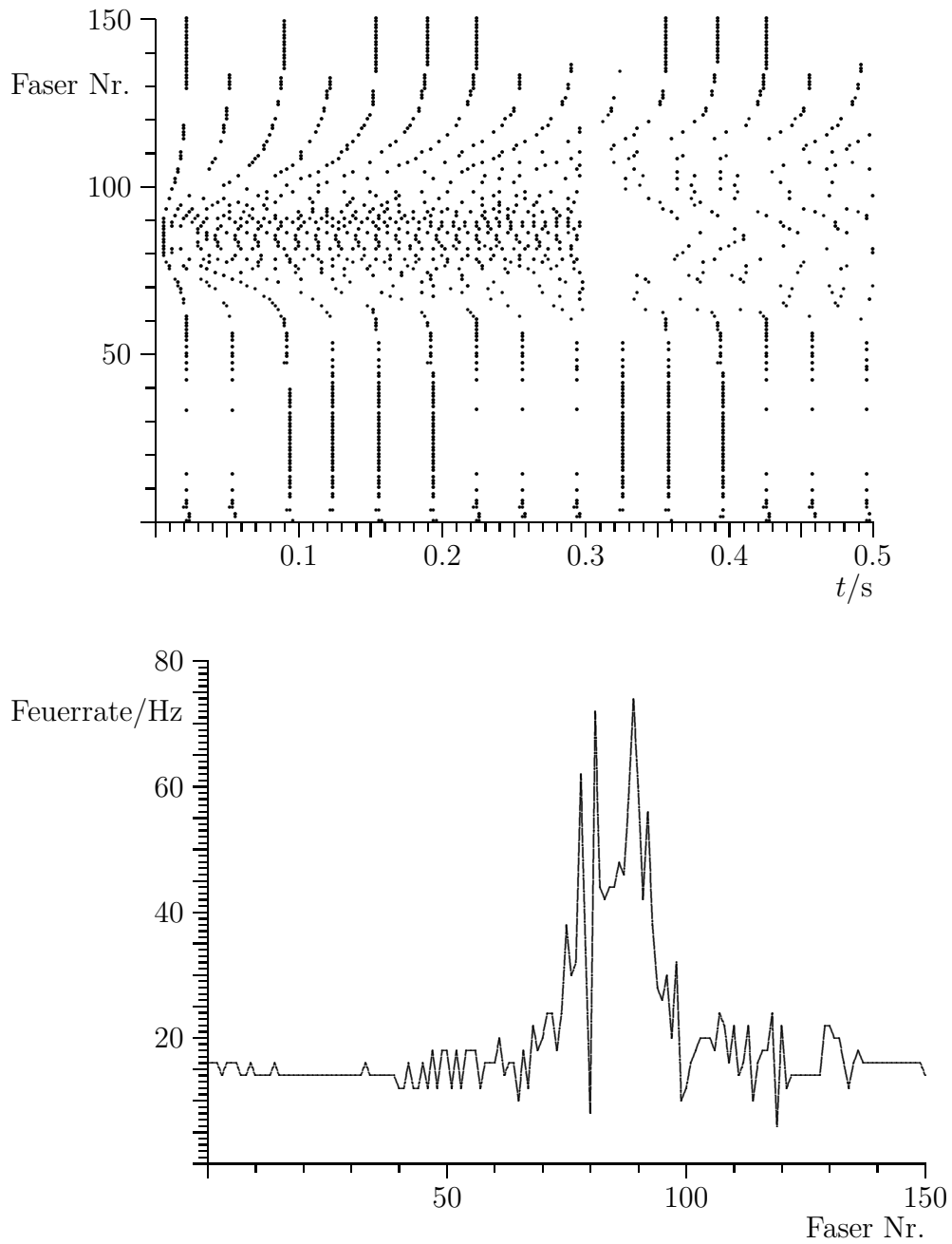


Abb. 3.13: Wie Abb. 3.12, nur war der Stimulus diesmal ein Sinuston mit Frequenz 1000 Hz.

3.2 Ein Tinnitusmodell für das zentrale auditorische System

Das hier vorgestellte Tinnitus-Modell stammt aus [Gerken 1996].

Unter Tinnitus versteht man einen Höreindruck, der ohne einen auf das Ohr einwirkenden Schall auftritt.

Tinnitus wird meist nicht durch otoakustische Emissionen (siehe Kap. 5.2.2) verursacht, denn bei nur 5% aller Personen, die unter einem Tinnitus leiden, können entsprechende otoakustische Emissionen festgestellt werden [Yost 1994]. Der Tinnitus entsteht deshalb vorwiegend erst im zentralen auditorischen System, obwohl seine Ursachen schon im peripheren auditorischen System liegen können. Tinnitus kann z. B. bei einem Gehörschaden auftreten, der seine Ursache in geschädigten Haarzellen in der Cochlea hat.

Ein Gehörschaden geht mit reduzierter spontaner Aktivität der Gehörnerven aus der Cochlea einher. Ist also das Gehör etwa ab einer Frequenz von 4 kHz durch Einwirkung von Schall mit Frequenzen hauptsächlich über 4 kHz geschädigt, so weisen die Hörnervenfasern mit einer charakteristischen Frequenz ab 4 kHz eine reduzierte spontane Aktivität auf. Nach Kap. 2.1.4 sind diese Nervenfasern dann auch gegenüber akustischer Stimulation weniger empfindlich, wie es bei einem Gehörschaden der Fall sein sollte.

In [Gerken 1996] wird nun davon ausgegangen, dass in einem der Kerne des auditorischen Systems laterale Inhibition auftritt. Diese würde jede Unebenheit in der spontanen Aktivität des Hörnervs noch verstärken. In obigem Beispiel eines Gehörschadens tritt bei 4 kHz eine Kante in der spontanen Aktivität der Hörnervenfasern auf. Diese würde durch laterale Inhibition zu einer Spitze verstärkt werden. Diese Spitze in der Erregung des Kerns bei einer charakteristischen Frequenz von 4 kHz könnte eine Erklärung für das Auftreten eines Tinnitus bei eben dieser Frequenz sein.

Darüber hinaus haben Messungen in [Gerken 1996] ergeben, dass die Empfindlichkeit der Kerne im zentralen auditorischen System gegenüber elektrischen Reizen nach einem Hörverlust zunimmt. Dies betrifft zumindest den Nucleus cochlearis, den Olivenkomplex, den Colliculus inferior und das Corpus geniculatum medialis. Man kann annehmen, dass diese Zunahme der Empfindlichkeit in höheren Regionen die Entstehung eines Tinnitus durch den oben beschriebenen Mechanismus noch fördert.

Kapitel 4

Rolle der lateralen Inhibition im auditorischen System

Schon in Kap. 2.2.1 haben wir die laterale Inhibition im Nucleus cochlearis erwähnt sowie einige Funktionen, die diese laterale Inhibition erfüllen könnte. Nach einer etwas genaueren Beschreibung der Physiologie der lateralen Inhibition in Kap. 4.1 sollen in Kap. 4.2 die Auswirkungen der Inhibition auf das Antwortverhalten der Neuronen an einem einfachen Modell quantitativ untersucht werden. Danach wird in Kap. 4.3 die Auswirkung lateraler Inhibition auf die Zuverlässigkeit der Detektion eines Signals in verrauschter Umgebung untersucht.

4.1 Physiologie der lateralen Inhibition

In diesem Abschnitt soll die Physiologie der lateralen Inhibition im Nucleus cochlearis, soweit bekannt, kurz vorgestellt werden (vgl. auch Kap. 2.2.1). [Young, Brownell 1976] unterteilen die Neuronen des Nucleus cochlearis in verschiedene Typen, abhängig von der Struktur ihres rezeptiven Feldes. Typ I-Neuronen haben hörnervähnliches Antwortverhalten und werden nicht inhibiert. Typ II und Typ III-Neuronen haben V-förmige zentrale exzitatorische Bereiche und inhibitorische Seitenbänder. Typ II-Neuronen reagieren nicht oder wenig auf weißes Rauschen und haben geringe spontane Aktivität (bis ca. 2,5/s). Typ III-Neuronen reagieren auf weißes Rauschen und haben eine höhere spontane Aktivität (bis ca. 95/s). Typen I, II und III haben monotone Raten-Pegel-Funktionen (zum Begriff der Raten-Pegel-Funktion siehe Kap. 2.1.4 und 2.2.5). Typ IV-Neuronen haben eine exzitatorische Region bei geringen Schallintensitäten und weit (über mehrere Bark) ausgedehnte inhibitorische Bereiche bei etwas höheren Intensitäten. Sie haben dementsprechend nicht-monotone Raten-Pegel-Funktionen, da bei geringen Intensitäten der zentrale exzitatorische Bereich, bei höheren Intensitäten allerdings dann der darüberliegende inhibitorische Bereich des rezeptiven Feldes angesprochen wird. Die Untersuchungen in [Voigt, Young 1980, Voigt, Young 1982, Voigt, Young 1990] stützen die darin ausgesprochene Hypothese, dass Typ IV-Neuronen von Typ II- und Typ III-Neuronen inhibiert werden. Hierdurch kann man die Struktur der rezeptiven Felder von Typ IV-Neuronen verstehen.

4.2 Laterale Inhibition in einem linearen Modell

Wir stellen uns eine Reihe von Neuronen vor, die tonotop entlang der x -Achse angeordnet sind. Das Membranpotential des Neurons an der Stelle x sei $h(x)$. Die Änderung $\frac{\partial h(x,t)}{\partial t}$ des Membranpotentials des Neurons an der Stelle x sei eine lineare Funktion der Membranpotentiale der anderen Neuronen, linear im Eingangssignal $E(x)$ und außerdem linear im Membranpotential $h(x, t)$ des Neurons selbst, d.h.

$$(4.1) \quad \frac{\partial h(x, t)}{\partial t} \frac{1}{C} = \int I(x - x') h(x', t) dx' + E(x) - h(x, t).$$

Dabei ist C eine reelle Konstante und $I(x - x')$ eine nur vom Abstand $x - x'$ der Neuronen abhängige Wechselwirkung zwischen den Neuronen x und x' . Eventuelle andere Konstanten seien bereits in der Wechselwirkung $I(x - x')$ bzw. im Eingangssignal $E(x)$ enthalten.

Ziel ist es nun, das Membranpotential $h(x)$ im stationären Zustand zu berechnen. Der stationäre Zustand ist erreicht, wenn $\frac{\partial h(x,t)}{\partial t} = 0$, d.h.

$$h(x) = \int I(x - x') h(x') dx' + E(x).$$

Fouriertransformation ergibt

$$\tilde{h}(k) = \tilde{I}(k) \tilde{h}(k) + \tilde{E}(k).$$

Nach $\tilde{h}(k)$ aufgelöst:

$$\tilde{h}(k) = \frac{\tilde{E}(k)}{1 - \tilde{I}(k)}$$

Die rechte Seite lässt sich i.d.R. nicht analytisch Fourier-rücktransformieren, wohl aber die Reihenentwicklung

$$\tilde{h}(k) = \tilde{E}(k) \sum_{n=0}^{\infty} \tilde{I}(k)^n.$$

Die Reihe konvergiert für alle k , falls $|\tilde{I}(k)| < 1$ für alle k .

Im Folgenden gehen wir von einer gaußförmigen Inhibition

$$I(x - x') = -I e^{-\frac{1}{2} \frac{(x-x'-\Delta x)^2}{d^2}}$$

von Neuron x' auf Neuron x aus. Hierbei bezeichnet I die Stärke und d die Reichweite der lateralen Inhibition. Neuron x' inhibiert das Neuron $x = x' + \Delta x$ maximal, d.h. wir betrachten auch den Fall einer asymmetrischen Inhibition. Der Parameter Δx gibt dabei den Grad der Asymmetrie an.

Die Fouriertransformierte $\tilde{I}(k)$ von $I(x)$ ist

$$\tilde{I}(k) = -\sqrt{2\pi} Id e^{-\frac{1}{2}\left[\left(dk+i\frac{\Delta x}{d}\right)^2 + \frac{\Delta x^2}{d^2}\right]}.$$

Somit lässt sich auch die Fourier-Rücktransformierte von \tilde{I}^n bzw. \tilde{h} berechnen:

$$h(x) = E(x) + \sum_{n=1}^{\infty} \mathcal{F}^{-1}(\tilde{E}\tilde{I}^n)(x)$$

Dabei bedeutet \mathcal{F} die Fouriertransformierte und \mathcal{F}^{-1} die Fourier-Rücktransformierte.

Wegen

$$\tilde{I}^n = \mathcal{F} \left[\sqrt{2\pi}^n (-Id)^n \frac{1}{\sqrt{2\pi n} d} e^{-\frac{(x-n\Delta x)^2}{2nd^2}} \right]$$

lassen sich die Summanden als Faltung schreiben:

$$(4.2) \quad h(x) = E(x) + \sum_{n=1}^{\infty} (-Id)^n \sqrt{2\pi}^{n-1} \frac{1}{\sqrt{n}} \int E(x-x') e^{-\frac{(x'-n\Delta x)^2}{2nd^2}} dx'$$

Die oben angegebene Konvergenzbedingung wird zu

$$\sqrt{2\pi} Id < 1,$$

d.h. das Produkt aus Stärke und Reichweite der Inhibition darf nicht zu groß werden.

Im Folgenden betrachten wir das stationäre Membranpotential bei verschiedenen an den Neuronen anliegenden Eingabesignalen $E(x)$.

4.2.1 Konstantes Eingangssignal an der Neuronenschicht

Es sei nun

$$E(x) = E.$$

Somit nach Gleichung (4.2)

$$\begin{aligned} h(x) &= E + E \sum_{n=1}^{\infty} \sqrt{2\pi}^n (-Id)^n \\ &= \frac{E}{1 + \sqrt{2\pi} Id}. \end{aligned}$$

Je größer also das Produkt Id aus Stärke und Reichweite der lateralen Inhibition, desto geringer wird das Membranpotential $h(x)$. Dies war zu erwarten.

Falls $\lim_{x \rightarrow \infty} E(x) = E$ ergibt sich für $\lim_{x \rightarrow \infty} h(x)$ ebenfalls der oben angegebene Wert.

4.2.2 Stufenfunktion als Eingabesignal der Neuronenschicht

Bei einer stufenförmigen Eingabe („Tiefpass“ bzw. „Hochpass“) ist zu erwarten, dass die Kanten durch die laterale Inhibition „aufgebogen“ werden.

Stufenförmige Eingabe bedeutet

$$E(x) = E H(x),$$

wobei H die HEAVISIDE-Funktion ist.

$$(4.3) \quad h(x) = E(x) + \frac{1}{2}E \sum_{n=1}^{\infty} \left(-\sqrt{2\pi} Id\right)^n \left(1 - \operatorname{erf}\left(\frac{n\Delta x - x}{\sqrt{2n}d}\right)\right).$$

Abb. 4.1 zeigt $h(x)$ für verschiedene Produkte Id aus Stärke I und Reichweite d der lateralen Inhibition. Es ist o.B.d.A. $d = 1$, denn andere Werte für d führen nach (4.3) nur zu einer anderen Skalierung der x -Achse.

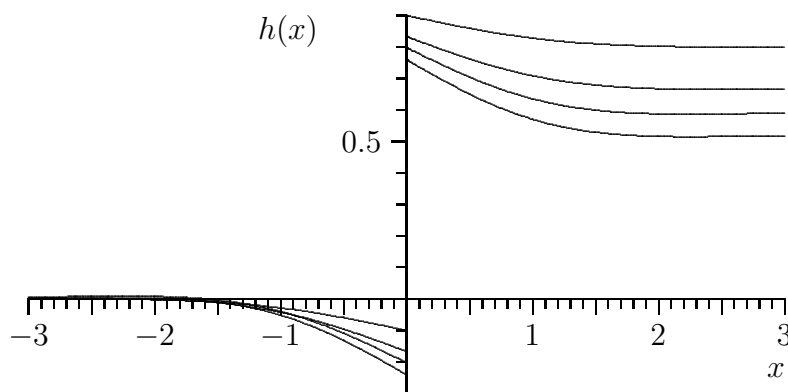


Abb. 4.1: Gezeigt ist das Membranpotential $h(x)$ für die Produkte $Id = 0,1; 0,2; 0,3; 0,37$. Je größer Id , desto kleiner ist $\lim_{x \rightarrow \infty} h(x)$. Eingabesignal ist $E(x) = H(x)$. Es ist $d = 1$ und $\Delta x = 0$.

Verbiegung für $\Delta x = 0$

Für $\Delta x = 0$ ist die Verbiegung

$$\frac{\lim_{x \downarrow 0} h(x)}{\lim_{x \rightarrow \infty} h(x)} - 1 = \sqrt{\pi/2} Id.$$

D.h. je größer die Stärke I und die Reichweite d der lateralen Inhibition, desto größer wird die Verbiegung, was auch in Abb. 4.1 zu sehen ist. Diese Verbiegung bewirkt die in Kap. 2.2.1 erwähnte Hervorhebung von Veränderungen im Spektrum entlang der Frequenzachse.

Verbiegung für $\Delta x \neq 0$

Für $\Delta x > 0$ ist die Verbiegung davon abhängig, ob $E(x) = EH(x)$ (Hochpass) oder $E(x) = EH(-x)$ (Tiefpass), denn die Inhibition wirkt dann ja stärker von links nach rechts als umgekehrt. Dies zeigt Abb. 4.2.

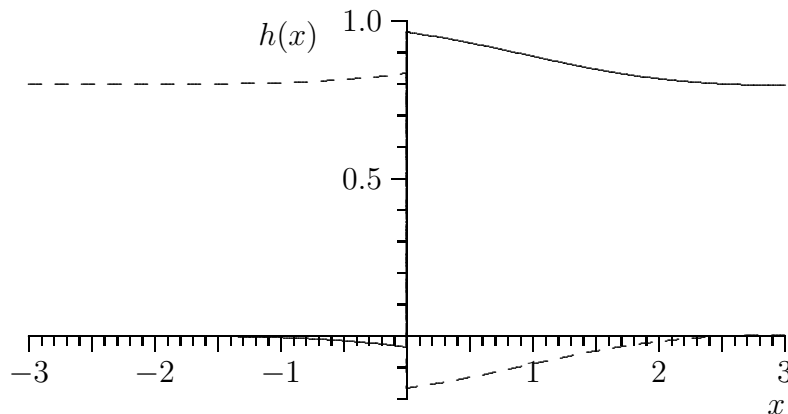


Abb. 4.2: Gezeigt ist das Membranpotential $h(x)$ für die Stärke $I = 0,1$ der Inhibition, Reichweite $d = 1$ und Verschiebung $\Delta x = 1$. Eingabesignal ist $E(x) = H(x)$ (durchgezogen) bzw. $E(x) = H(-x)$ (gestrichelt).

Für $Id \ll 1$ und $\Delta x \ll 1$ gilt für die Verbiegung¹.

$$\frac{\lim_{x \downarrow 0} h(x)}{\lim_{x \rightarrow \infty} h(x)} - 1 \approx I \cdot (\sqrt{\pi/2} d + \Delta x).$$

Abb. 4.3 zeigt die Verbiegung bei verschiedenen Reichweiten d der lateralen Inhibition in Abhängigkeit der Verschiebung Δx .

Ist also die Verschiebung $\Delta x \approx d$, d.h. ungefähr gleich der Reichweite der lateralen Inhibition, so tritt bei einem Tiefpass als Eingabesignal nurmehr eine geringe Verbiegung auf, bei einem Hochpass eine stärkere Verbiegung als mit $\Delta x = 0$. Beachte, dass ein Hochpass mit $\Delta x > 0$ einem Tiefpass mit $\Delta x < 0$ entspricht.

4.2.3 Konstantes Eingabesignal mit „Lücke“ an der Neuronenschicht

Als Eingabesignal verwenden wir nun (siehe Abb. 4.4)

$$E(x) = \begin{cases} E & \text{für } |x| > B/2 \\ (1 - T)E & \text{sonst} \end{cases}$$

B heißt hierbei Breite der Lücke, $0 \leq T \leq 1$ heißt Tiefe der Lücke.

Nach Gleichung (4.2) ergibt sich für das Membranpotential $h(x)$

¹Alle weiter unten angegebenen Näherungsformeln für die Verbiegung erhält man, indem man die unendliche Summe in der jeweiligen Gleichung für $h(x)$ durch endlich viele Terme, meist nur den ersten, ersetzt und die Fehlerfunktion nach den „kleinen“ Größen entwickelt.

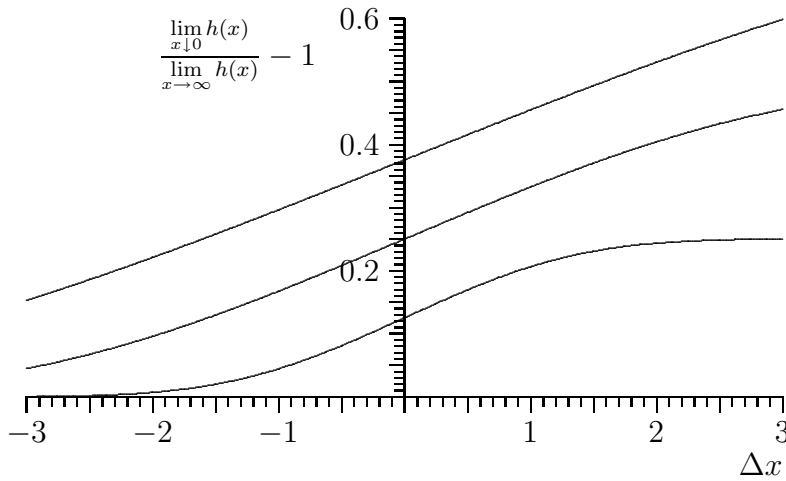


Abb. 4.3: Gezeigt ist die Verbiegung $\frac{\lim_{x \downarrow 0} h(x)}{\lim_{x \rightarrow \infty} h(x)} - 1$ in Abhängigkeit von der Verschiebung Δx bei verschiedenen Reichweiten $d = 1; 2; 3$ der lateralen Inhibition. Die Stärke der lateralen Inhibition ist $I = 0,1$. Je größer die Reichweite d , desto stärker ist die Verbiegung und desto langsamer sättigt die Verbiegung mit zunehmendem $|\Delta x|$.

$$h(x) = E(x) + \frac{1}{2}E \sum_{n=1}^{\infty} (-\sqrt{2\pi} Id)^n \left[2 + T \left(\operatorname{erf} \left(\frac{2x - 2n\Delta x - B}{2\sqrt{2nd}} \right) - \operatorname{erf} \left(\frac{2x - 2n\Delta x + B}{2\sqrt{2nd}} \right) \right) \right]$$

Abb. 4.5 zeigt hierfür ein Beispiel.

Für $\Delta x = 0$, $Id \ll 1$ und $B/d \ll 1$ gilt für die Verbiegung

$$\frac{\lim_{x \downarrow B/2} h(x)}{\lim_{x \rightarrow \infty} h(x)} - 1 \approx TIB \left(1 - \frac{B^2}{6d^2} \right).$$

Je größer also die Breite B und die Tiefe T der Lücke, desto stärker ist die Verbiegung, und zwar bei kleinen Breiten B unabhängig von der Reichweite d der lateralen Inhibition (siehe Abb. 4.6). Bei größeren Breiten B gilt allerdings: Je größer die Reichweite d der lateralen Inhibition, desto größer die Verbiegung.

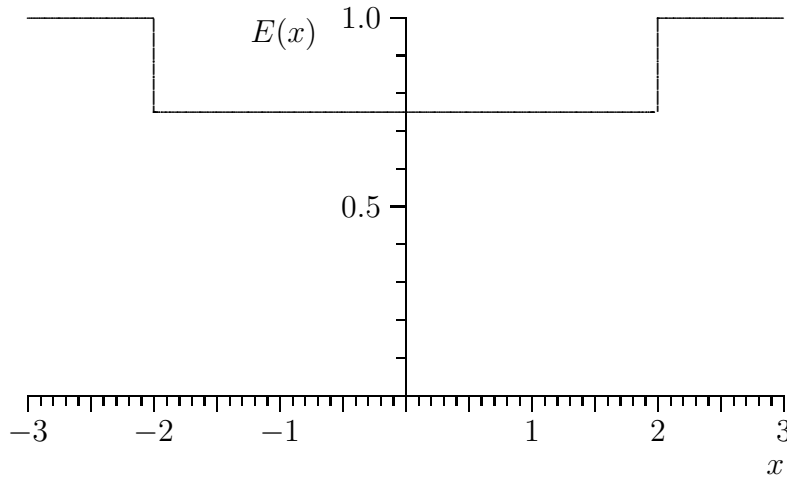


Abb. 4.4: Eingangssignal $E(x)$ mit Lücke. Breite der Lücke ist $B = 4$, Tiefe ist $T = 1/4$.

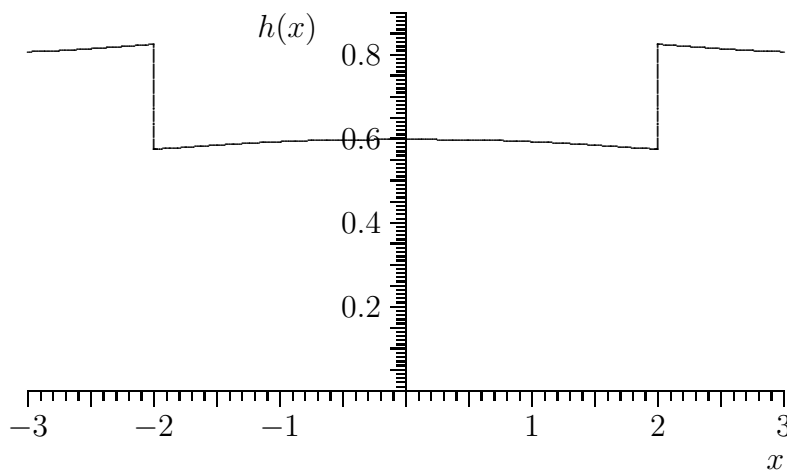


Abb. 4.5: Membranpotential $h(x)$ bei einem Eingangssignal $E(x)$ mit Lücke wie in Abb. 4.4. Stärke der lateralen Inhibition ist $I = 0,1$, die Reichweite $d = 1$ und die Verschiebung $\Delta x = 0$.

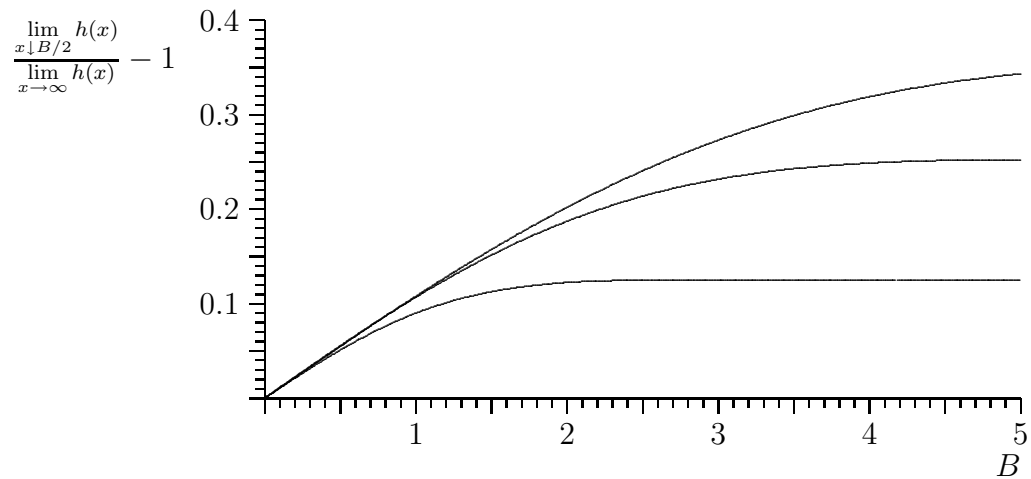


Abb. 4.6: Gezeigt ist die Verbiegung $\frac{\lim_{x \downarrow B/2} h(x)}{\lim_{x \rightarrow \infty} h(x)} - 1$ in Abhängigkeit von der Breite B der Lücke bei verschiedenen Reichweiten $d = 1; 2; 3$ der lateralen Inhibition. Tiefe der Lücke ist $T = 1$, Stärke der Inhibition ist $I = 0,1$.

4.3 Laterale Inhibition in einem Modell mit spikenden Neuronen

Wir haben im vorhergehenden Kapitel gesehen, dass laterale Inhibition Kannten verstärkt und somit Strukturen hervorhebt. Man vermutet daher, dass laterale Inhibition dazu dient, als Vorverarbeitungsstufe Signale besser detektierbar zu machen [Zenner 1994]. Beispielsweise sollte nach einer Schicht von Neuronen mit lateraler Inhibition ein reiner Ton von Rauschen besser getrennt werden können als dies vorher möglich war. D.h. es sollte mit geringerer Fehlerrate festgestellt werden können, ob ein reiner Ton vorhanden war oder nicht. Um abzuschätzen, mit welcher Fehlerrate das Gehör eine solche Erkennungsleistung vollbringen kann, benötigt man stochastische Datenanalyse². Deshalb soll im nächsten Abschnitt kurz hierauf eingegangen werden, bevor das Neuronenmodell zur Simulation spikender Neurone³ und Simulationsergebnisse vorgestellt werden.

4.3.1 Stochastische Datenanalyse

Die stochastische Datenanalyse befasst sich mit Datenvektoren \vec{x} , die aus stochastischen Quellen (Klassen) k gezogen werden. Die Wahrscheinlichkeit, dass aus Quelle k gezogen wird, sei $p(k)$. Die Wahrscheinlichkeit, dass Quelle k den Datenvektor \vec{x} liefert, sei $p(\vec{x}|k)$.

Gesucht sind nun Methoden, die es erlauben, vom Datenvektor \vec{x} auf die Quelle k zurückzuschließen. Außerdem ist die Fehlerrate gesucht, mit der dies möglich ist.

Beispielsweise wird bis zu einer gewissen Vorverarbeitungsstufe im Gehör ein Schall in Feuerzeiten \vec{x} von Neuronen umgewandelt. Als Schall kommen zwei Wörter $k \in 1, 2$ in Frage, die allerdings immer etwas verschieden ausgesprochen werden. Gesucht ist der Fehler, mit dem es möglich ist, anhand der Feuerzeiten der Neuronen in der Vorverarbeitungsstufe zu entscheiden, welches Wort k gesprochen wurde.

Bayes-Klassifikator

Die Wahrscheinlichkeit, dass der Datenvektor \vec{x} aus der Quelle k gezogen wurde, ist

²Eine Alternative wäre ein informationstheoretischer Ansatz wie in [Haft 1998]. Hier wird die Transinformation zwischen externem Stimulus und interner Repräsentation des Stimulus unter Anwesenheit von Rauschen maximiert.

³Man könnte auch daran denken, diese Untersuchungen mit linearen Neuronen wie in Kap. 4.2 durchzuführen. Das Eingabesignal $E(x)$ wäre dann etwa ein Kurzzeitspektrum des angelegten Stimulus. Das Signal $h(x, t)$ der Outputschicht berechnet man nach (4.1). Das Signal $h(x, t)$ sähe dann in jedem Zeitabschnitt von 100 ms wegen der Zufallskomponente im Rauschen verschieden aus. Das Entscheidungskriterium, ob ein zusätzlicher reiner Ton vorlag, basiert darauf, wie groß $h(x_0, t)$ im Zeitraum von 100 ms durchschnittlich war, wobei x_0 die Stelle ist, an der das Eingabesignal $E(x)$ eines reinen Tons maximal ist.

$$p(k|\vec{x}) = \frac{p(k)p(\vec{x}|k)}{p(\vec{x})} = \frac{p(k)p(\vec{x}|k)}{\sum_k p(k)p(\vec{x}|k)}.$$

Wählt man aufgrund der Messung \vec{x} die Klasse $k_{\vec{x}}$, so ist die mittlere Wahrscheinlichkeit, dass richtig zugeordnet wird

$$\int p(\vec{x})p(k_{\vec{x}}|\vec{x})d\vec{x}.$$

Diese ist genau dann maximal, wenn

$$(4.4) \quad p(k_{\vec{x}}|\vec{x}) = \max_k p(k|\vec{x}).$$

Die entsprechende Zuordnung $\vec{x} \mapsto k_{\vec{x}}$ heißt dann auch Bayes-Klassifikator.

Die mittlere Fehlerwahrscheinlichkeit (Fehlerrate) E ist dann nach obigen Gleichungen

$$(4.5) \quad E = 1 - \int p(\vec{x})p(k_{\vec{x}}|\vec{x})d\vec{x} = 1 - \int \max_k (p(k)p(\vec{x}|k))d\vec{x}.$$

Abstandsmaße

Ein Abstandsmaß ist ein Maß für die Trennbarkeit mehrerer Wahrscheinlichkeitsverteilungen. D.h. je größer der Abstand, desto sicherer kann von einer Messung \vec{x} gesagt werden, aus welcher stochastischen Quelle k sie stammt. Abstandsmaße sind also Abschätzungen für die mittlere Fehlerwahrscheinlichkeit bei Verwendung des Bayes-Klassifikators.

Im Folgenden betrachten wir Abstände zwischen zwei Wahrscheinlichkeitsverteilungen $p(\vec{x}|1)$ und $p(\vec{x}|2)$.

Kolmogorov-Abstand Der Kolmogorov-Abstand J_K ist definiert als

$$J_K := \int |p(1)p(\vec{x}|1) - p(2)p(\vec{x}|2)| d\vec{x}.$$

Für die Fehlerrate E gilt

$$E = \frac{1}{2}(1 - J_K).$$

Bhattacharyya-Abstand Der Bhattacharyya-Abstand J_B ist definiert als

$$J_B := -\ln \int \sqrt{p(\vec{x}|1)p(\vec{x}|2)} d\vec{x}.$$

Für die Fehlerrate E gilt

$$E \leq \sqrt{p(1)p(2)} e^{-J_B}.$$

4.3.2 Das verwendete Neuronenmodell und die Verschaltung der Neurone

Wir nehmen an, dass der Hörnerv die Reihe der inneren Haarzellen („Inputschicht“) tonotop in den Nucleus cochlearis („Outputschicht“) abbildet. Dies zeigt Abb. 4.7. Die Verbindungen zwischen der Input- und der Outputschicht stellen also den Hörnerv dar. Weiter wird angenommen, dass sich die Neuronen in der Outputschicht gegenseitig (lateral) inhibieren, wie in Abb. 4.7 angedeutet.

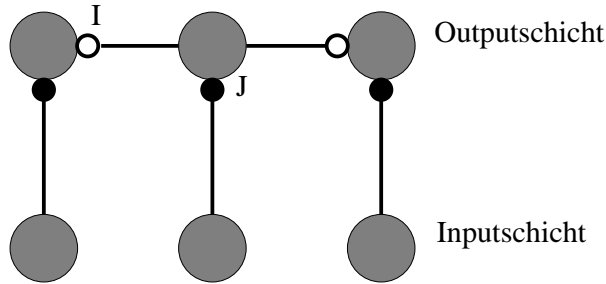


Abb. 4.7: Modell der Verschaltung der Neurone. Graue Kreise bezeichnen Neurone, Linien Axone, ausgefüllte Punkte exzitatorische und offene Punkte inhibitorische Synapsen.

Die Neuronen wurden als „Spike-Response-Neurone“ nach [Gerstner 1993] modelliert⁴.

Das synaptische Potential h_{syn} des Neurons Nr. i in der Outputschicht ist

$$h_{syn}(i, t) = J \sum_{t_f^{in}(i)} \varepsilon(t - t_f^{in}(i)) + \sum_{j \neq i} I(i, j) \sum_{t_f(j)} \varepsilon(t - t_f(j)).$$

Hierbei sind $t_f^{in}(i)$ die Feuerzeiten des Neurons Nr. i in der Inputschicht, $t_f(i)$ die Feuerzeiten des Neurons Nr. i in der Outputschicht und J die Stärke der „feed-forward“-Kopplung.

Die Stärke $I(i, j)$ der lateralen Inhibition zwischen den Neuronen i und j in der Outputschicht wurde als

$$I(i, j) = I e^{-(i-j)^2/d^2}$$

angenommen.

Das Refraktärpotential h_{ref} des Neurons i ist

$$h_{ref}(i, t) = \sum_{t_f(i)} \eta(t - t_f(i)).$$

⁴In [Gerstner 1993] werden die Eigenschaften des Spike-Response-Modells mit dem Neuronenmodell von HODGKIN und HUXLEY 1952 verglichen. Das Spike-Response-Modell stellt eine gute Näherung dieses aus vier gekoppelten nicht linearen Differenzialgleichungen bestehenden Modells dar (siehe [Domany 1994, Gerstner 1993, Kistler 1996, Kistler 1997]). Es eignet sich jedoch wegen seiner leichten Berechenbarkeit auch zur Simulation größerer Netze.

Das Membranpotential ist

$$h(i, t) = h_{syn}(i, t) + h_{ref}(i, t).$$

Das Neuron feuert, wenn⁵

$$h(i, t) > \theta.$$

Für $\varepsilon(t)$ und $\eta(t)$ wurde

$$\varepsilon(t) = \frac{t}{\tau_s} e^{(1-t/\tau_s)} \quad \text{für } t \geq 0$$

bzw.

$$\eta(t) = \begin{cases} -\infty & \text{für } 0 < t \leq \gamma_{ref} \\ -\frac{\tau_{ref}}{t - \gamma_{ref}} & \text{für } t > \gamma_{ref} \end{cases}$$

angesetzt.

d	θ	τ_s	τ_{ref}	γ_{ref}
3	1	10 ms	5 ms	2 ms

Tab. 4.1: Die Tabelle zeigt die in der Simulation in Kap. 4.3.4 verwendeten Konstanten. Die Größen I und J wurden variiert.

4.3.3 Die in der Simulation verwendeten Stimuli

Das Gehör soll nun einen reinen Ton aus weißem Rauschen heraushören, d.h. es soll entscheiden, ob ein reiner Ton vorliegt oder nicht. Aus Kap. 2 wissen wir, dass der Hörnerv ein in Feuerraten codiertes Kurzzeitspektrum liefert. Angenommen, das an den Hörnerv anschließende „Gehör“ wäre ein idealer Klassifikator. Dann könnte es schon anhand des Eingangssignals entscheiden, ob ein reiner Ton vorliegt oder nicht. Da es sich bereits um einen idealen Klassifikator handelt, kann die Fehlerrate durch irgendeine Art der Vorverarbeitung nur verschlechtert werden, jedoch niemals verbessert. D.h. es wäre am besten, wenn der Nucleus cochlearis das Signal des Hörnervs möglichst unverfälscht an das „Gehör“ weitergeben würde. Da das „Gehör“ jedoch kein idealer Klassifikator ist, ist eine Vorverarbeitung sinnvoll.

Um zu entscheiden, ob z. B. laterale Inhibition als Vorverarbeitung sinnvoll ist, müssen wir also Hypothesen darüber machen, welche Art von Klassifikator das Gehör einsetzt, um etwa zwischen einem reinen Ton mit Rauschen und nur Rauschen zu unterscheiden. Die einfachste Annahme ist die folgende: Das Gehör stellt das Neuron fest, das am häufigsten gefeuert hat. Falls die Feuerrate einen gewissen Schwellenwert übersteigt, vermutet das Gehör

⁵Es handelt sich also um ein deterministisch feuerndes Neuron. Nach [Gerstner 1993] können auch stochastisch feuernde Neuronen modelliert werden, indem man annimmt, dass für die Wahrscheinlichkeit $P(h(t))$, dass das Neuron innerhalb des Zeitabschnitts $[t, t + dt]$ feuert, $P(h(t)) = e^{\beta(h(t) - \theta)} dt$ gilt. Die Größe β^{-1} bestimmt dabei die Stärke des Rauschens und heißt „formale Temperatur“ des Systems.

einen reinen Ton an dieser Stelle, andernfalls nimmt es an, dass die Feuerrate allein durch das Hintergrundrauschen erzeugt wurde. Dies muss etwa innerhalb eines Zeitraums von 100 ms geschehen⁶. Noch nicht festgelegt ist die Schwelle, ab der das Gehör sich für die eine oder die andere Möglichkeit entscheidet. Sie wird hier nach Gleichung (4.4) immer so gewählt, dass die Irrtumswahrscheinlichkeit minimal ist⁷.

Die Irrtumswahrscheinlichkeit wurde bei verschiedenen starker Inhibition und Exzitation nach Gleichung (4.5) berechnet. Hierzu benötigt man folgende Wahrscheinlichkeitsverteilungen:

- Falls der Stimulus mit Rauschen und dem reinen Ton vorliegt die Wahrscheinlichkeit, dass das Neuron, das auf den reinen Ton optimal anspricht⁸, genau n mal feuert.
- Falls der Stimulus mit nur Rauschen vorliegt die Wahrscheinlichkeit, dass das Neuron, das gerade am häufigsten gefeuert hat, genau n mal gefeuert hat.

Um die erste Wahrscheinlichkeitsverteilung zu messen, wurde das simulierte Gehör eine zeitlang mit reinen Tönen und Rauschen stimuliert. Die Frequenz der reinen Töne wechselte alle 100 ms zufällig. Um die zweite Wahrscheinlichkeitsverteilung zu messen, wurde das simulierte Gehör eine zeitlang nur mit Rauschen stimuliert.

Der Stimulus „Rauschen“ wurde dabei erzeugt, indem alle Neuronen der Inputschicht poissonverteilt durchschnittlich mit einer Rate von 50 Hz feuerten. Der Stimulus „Rauschen mit reinem Ton“ wurde erzeugt, indem das Neuron der Inputschicht, das am besten auf diesen Ton anspricht, poissonverteilt mit einer durchschnittlichen Rate von 150 Hz zum Feuern gebracht wurde⁹. Alle anderen Neuronen der Inputschicht feuerten wie beim Stimulus „Rauschen“.

4.3.4 Simulationsergebnisse

Zunächst wird die Stärke I der lateralen Inhibition auf Null gesetzt und wir variieren nur die Stärke J der feed-forward-Kopplung. Abb. 4.8 zeigt die Fehlerrate E nach (4.5) in Abhängigkeit von der feed-forward-Kopplung J .

Man erkennt ein relativ breites Minimum der Fehlerrate. Die Ursache ist, dass bei geringer feed-forward-Kopplung J die Feuerraten einfach zu gering sind, um zwischen den beiden Stimuli zuverlässig zu unterscheiden. Bei zu

⁶Ein Schall muss mindestens 100–200 ms dauern, damit seine Lautheit gleich der Lautheit des Dauerschalls gleicher Schallintensität ist [Sottek 1993]. Dies ist ein Hinweis darauf, dass die Integrationszeiten des Gehörs etwa 100–200 ms betragen.

⁷Man kann sich etwa vorstellen, dass das Gehör diese Aufgabe lernen kann und die Schwelle in „Vorversuchen“ bereits auf ihren optimalen Wert eingestellt hat.

⁸Das „Erkennen“ eines reinen Tons, das nur darauf beruht, dass *irgendein* Neuron zufällig sehr häufig gefeuert hat, soll als Fehler gewertet werden.

⁹Die Raten unterscheiden sich stark. Dies wird allerdings dadurch relativiert, dass die Neuronen der Inputschicht ja innerhalb des Auswertungszeitraums von 100 ms nur etwa 5–15-mal feuern.

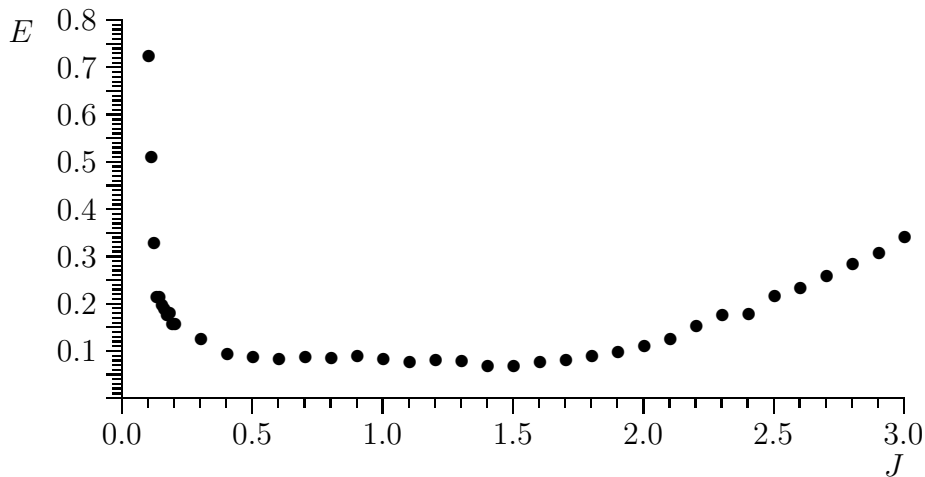


Abb. 4.8: Aufgetragen ist die Fehlerrate E in Abhängigkeit der Stärke J der feed-forward-Kopplung. Die Stärke der lateralen Inhibition ist $I = 0$, die Dauer der Simulation betrug 100 s.

großer feed-forward-Kopplung dagegen werden die Feuerraten so groß, dass sich die Neuronen in der Sättigung befinden, unabhängig davon, ob Rauschen mit zusätzlichem reinen Ton oder nur Rauschen allein vorliegt.

Die entsprechenden Wahrscheinlichkeitsverteilungen der Anzahl der Aktionspotentiale zeigt Abb. 4.9.

In Abb. 4.8 haben wir ein breites Minimum in der Irrtumswahrscheinlichkeit von etwa $J = 0,5 - 1,5$ beobachtet. Es erscheint sinnvoll, nun die Stärke der feed-forward-Kopplung auf dem Wert $J = 1,5$ konstant zu lassen, und die Stärke der lateralen Inhibition zu variieren. Die daraus resultierenden Irrtumswahrscheinlichkeiten zeigt Abb. 4.10, die entsprechenden Wahrscheinlichkeitsverteilungen Abb. 4.11.

Die Abb. 4.12 und 4.13 zeigen noch einmal eine genauere Analyse der Abhängigkeit der Irrtumswahrscheinlichkeit E von der Stärke der feed-forward-Kopplung J und der Stärke der lateralen Inhibition I .

Was ist der Grund dafür, dass die Irrtumswahrscheinlichkeit bei zu starker lateraler Inhibition wieder wächst? Man könnte zunächst vermuten, dass dann die Feuerrate einfach wieder zu gering wird, um zwischen den beiden Stimuli zu unterscheiden. Das Hauptproblem bei zu großer lateraler Inhibition ist allerdings ein anderes. Wir haben in Kap. 4.2.2 schon die „Überschwinger“ gesehen, die zu starke laterale Inhibition verursachen kann. Dasselbe Phänomen tritt auch im Modell mit spikenden Neuronen wieder auf, was Abb. 4.14 zeigt¹⁰. Noch deutlicher wird dies allerdings, indem man die Feuerzeitpunkte im Einzelnen aufträgt, siehe Abb. 4.15. Im Falle geringer lateraler Inhibition feuern die Neuronen schon nicht sehr gleichmäßig. Sie zeigen eher ein burstendes Verhalten, wobei die Bursts allerdings noch relativ kurz sind. Bei sehr starker lateraler Inhibition hingegen werden die Bursts sehr lang (länger als 100 ms). Auch zeigt sich eine deutliche Streifenstruktur im Feuermuster.

¹⁰Man beachte, dass die Neuronen am Rand mit höherer Rate feuern, da sie von einer Seite nicht inhibiert werden. Dies wird bei der Behandlung des Zwicker-Tons in Kap. 6 noch eine bedeutende Rolle spielen.

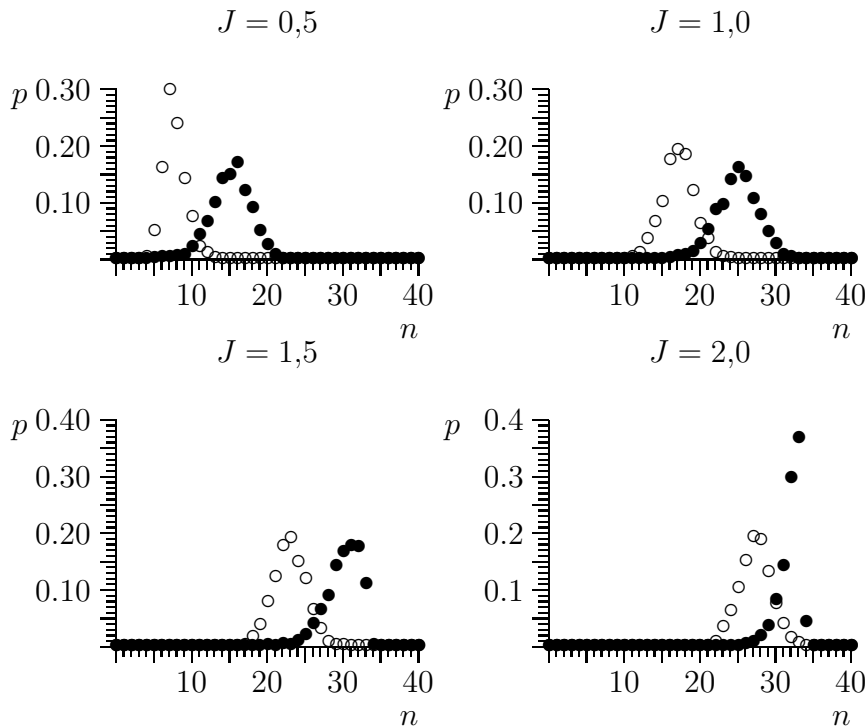


Abb. 4.9: Aufgetragen ist die Wahrscheinlichkeit, dass das Neuron, das am häufigsten gefeuert hat, innerhalb von 100 ms genau n -mal gefeuert hat (nur Rauschen als Stimulus) als „ \circ “ bzw. die Wahrscheinlichkeit, dass das Neuron, das auf den reinen Ton optimal anspricht, genau n -mal gefeuert hat (Rauschen und reiner Ton als Stimulus) als „ \bullet “.

Die Stärke J der feed-forward-Kopplung variiert von Bild zu Bild. Die Stärke der lateralen Inhibition ist $I = 0$, die Dauer der Simulation betrug 100 s.

Die Streifen sind relativ stabil, denn die verstärkt feuernden Neuronen werden ja von den Neuronen in ihrer Nachbarschaft nicht inhibiert, da diese nicht feuern. Die verstärkt feuernden Neuronen selbst inhibieren allerdings die Neuronen in ihrer Nachbarschaft.

4.3.5 Zusammenfassung

In den letzten Abschnitten haben wir gesehen, dass laterale Inhibition die Irrtumswahrscheinlichkeit des Gehörs bei der Detektion von Merkmalen in den angebotenen auditorischen Stimuli herabsetzen kann. Hierzu haben wir ein System spikender Modellneuronen mit lateraler Inhibition studiert, die zwischen zwei Stimuli unterscheiden sollten, nämlich Rauschen mit bzw. ohne zusätzlichen reinen Ton. Dies gelang mit lateraler Inhibition deutlich besser als ohne. Weiter dient laterale Inhibition im auditorischen System vermutlich zur Datenreduktion unter weitgehender Informationserhaltung, was auch für das visuelle System angenommen wird [Hauske 1994]. Beiden Zwecken der lateralen Inhibition ist jedoch gemeinsam, dass eine zu starke laterale Inhibition den Stimulus durch zusätzlich vorgetäuschte Kanten so verfälscht, dass eine Diskrimination bzw. Rekonstruktion nicht mehr möglich ist. Bei unserer Diskriminationsaufgabe erwies sich die genaue Stärke der lateralen Inhibition jedoch als relativ unkritisch.

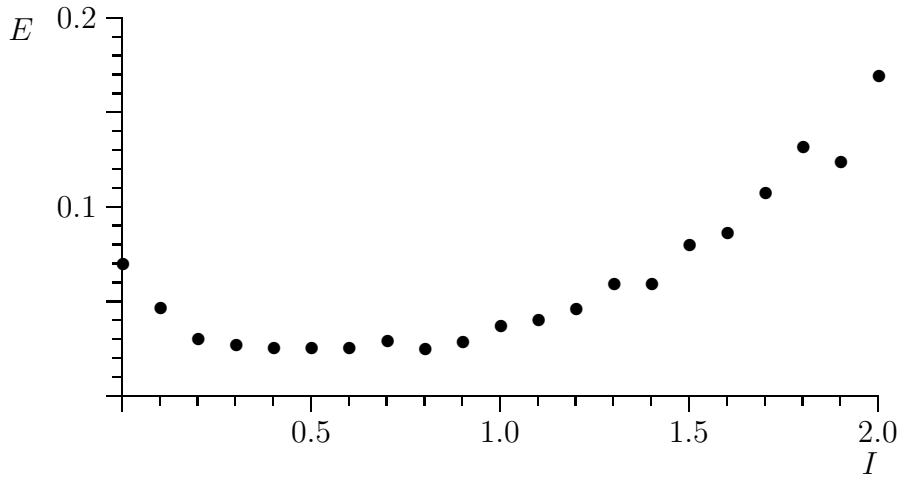


Abb. 4.10: Aufgetragen ist die Fehlerrate E in Abhängigkeit der Stärke I der lateralen Inhibition. Die Stärke der feed-forward-Kopplung ist $J = 1,5$, die Dauer der Simulation betrug 100 s.

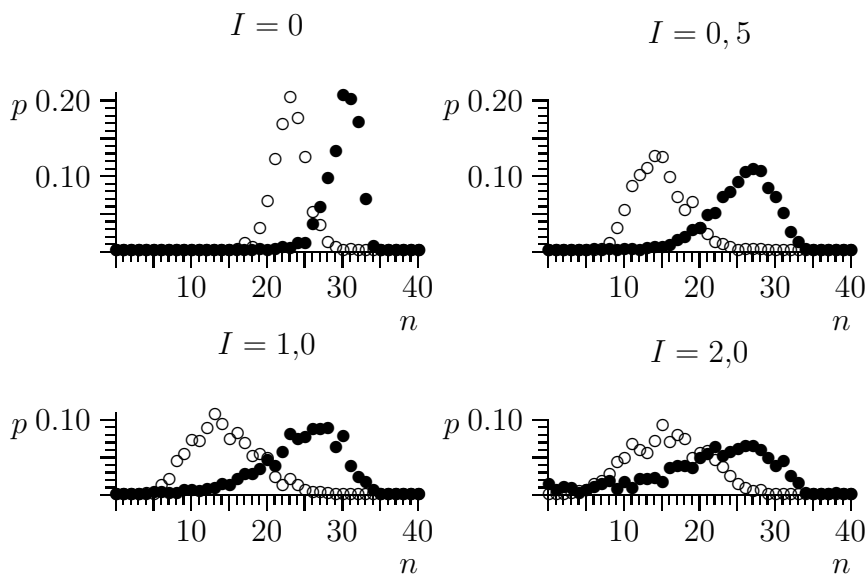


Abb. 4.11: Siehe Abb. 4.9.

Die Stärke I der lateralen Inhibition variiert von Bild zu Bild. Die Stärke der feed-forward-Kopplung $J = 1,5$, die Dauer der Simulation betrug 100 s.

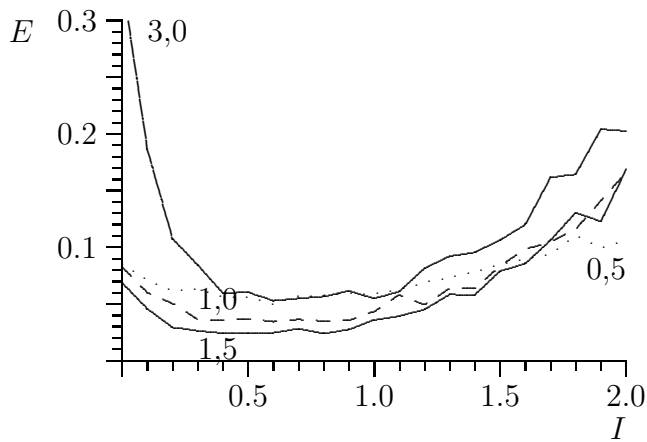


Abb. 4.12: Aufgetragen ist die Irrtumswahrscheinlichkeit E in Abhängigkeit der Stärke der lateralen Inhibition I . Die Kurven gelten für verschiedene feed-forward-Kopplungen J , die jeweils angegeben sind. Die Dauer der Simulation betrug 100 s.

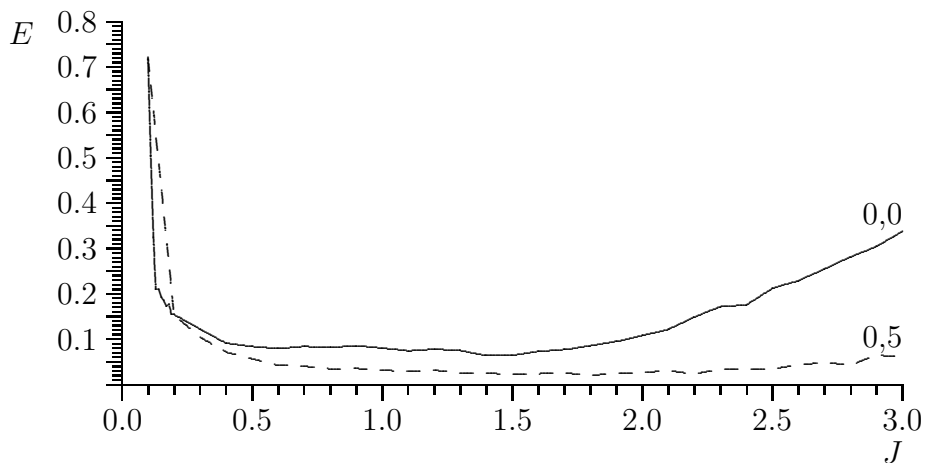


Abb. 4.13: Aufgetragen ist die Irrtumswahrscheinlichkeit E in Abhängigkeit der Stärke der feed-forward-Kopplung J . Die Kurven gelten für verschiedene Stärken I der lateralen Inhibition, die jeweils angegeben sind. Die Dauer der Simulation betrug 100 s.

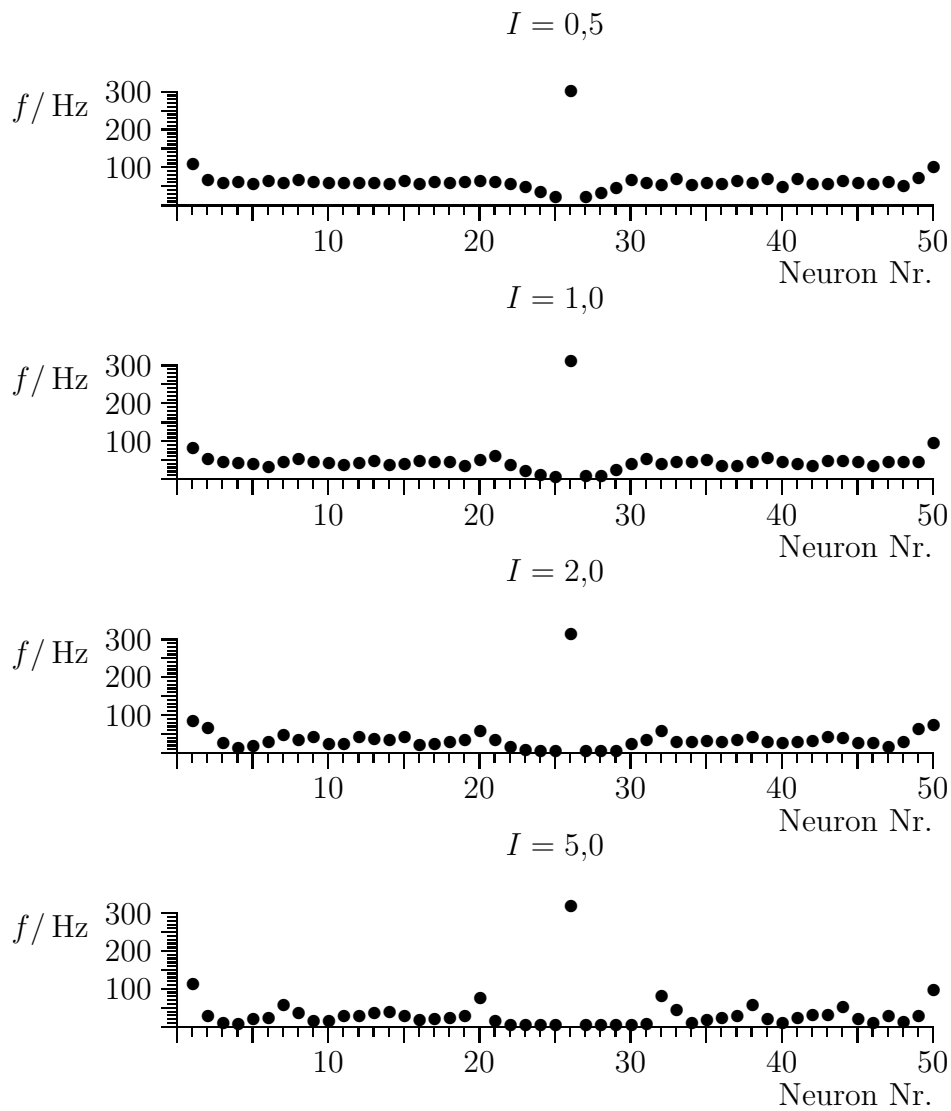


Abb. 4.14: Gezeigt sind die Feuerraten der einzelnen Neuronen der Outputschicht bei einem Stimulus aus Rauschen und einem reinen Ton. Die Stärken I der lateralen Inhibition sind jeweils angegeben. Die Stärke der feed-forward-Kopplung ist $J = 1,5$, die Dauer der Simulation betrug 100 s.

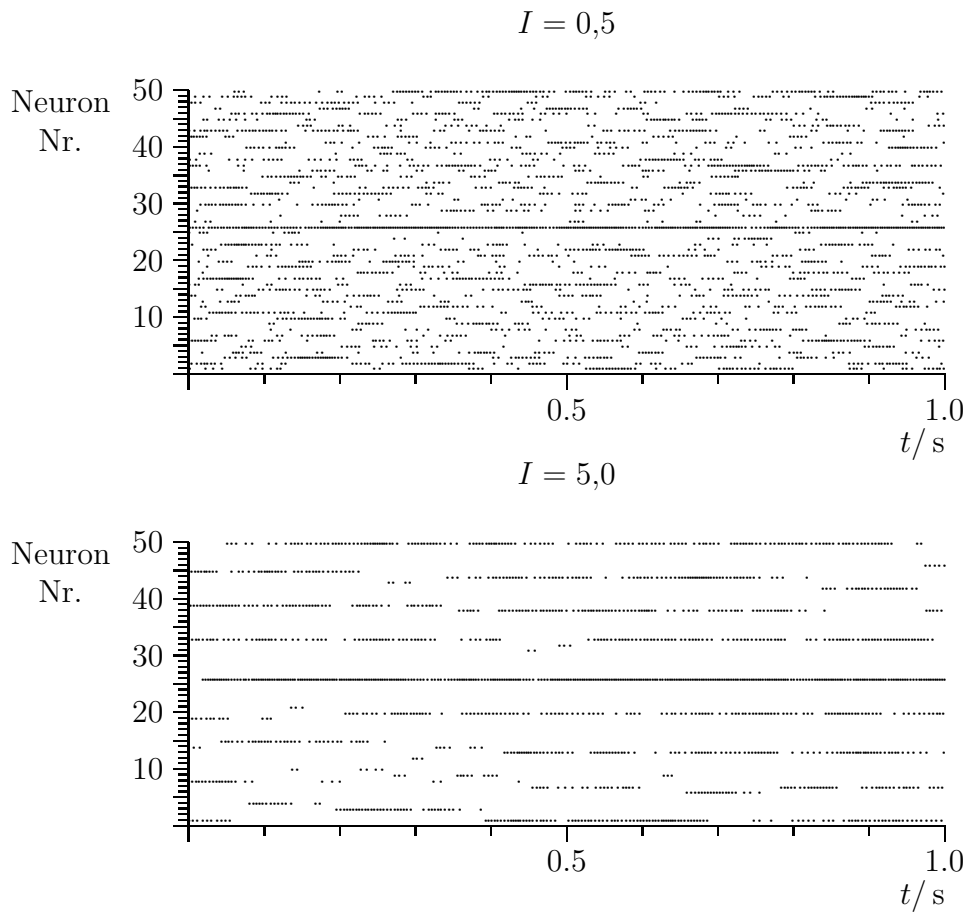


Abb. 4.15: Stimulus ist Rauschen und ein reiner Ton, der Neuron Nr. 26 erregt. Jeder Punkt bedeutet ein Aktionspotential des entsprechenden Neurons zur Zeit t . Die Stärken I der lateralen Inhibition sind jeweils angegeben. Die Stärke der feed-forward-Kopplung ist $J = 1,5$, die Dauer der Simulation betrug 1 s.

Kapitel 5

Zwicker-Ton

Der Zwicker-Ton (*Zwicker-tone*) ist ein akustischer Nacheffekt (*acoustic afterimage*). Er entsteht unmittelbar nach dem Abschalten von Schallen verschiedener Beschaffenheit. Es handelt sich um einen fast reinen Ton, der üblicherweise 1–2, aber auch bis zu 6 Sekunden nach dem Abschalten des erzeugenden Schalls noch zu hören ist. Der Ton ist sehr leise (10–15 phon) und klingt mit der Zeit ab. [Krump 1993, Zwicker 1990]

Zunächst sollen die wesentlichen Eigenschaften einiger Zwicker-Ton Erzeugerschalle vorgestellt werden, bevor im Abschnitt 5.2 darüber berichtet wird, welchen Einfluss diese Eigenschaften auf Tonhöhe und Qualität des Zwicker-Tons haben. Die Abhängigkeit der Qualität (siehe Kap. 5.2) und der Tonhöhe des Zwicker-Tons von vielen verschiedenen Parametern der Zwicker-Ton Erzeugerschalle ist in [Fastl 1986, Fastl 1989, Krump 1993] eingehend untersucht worden.

5.1 Zwicker-Ton Erzeugerschalle

Schalle, die nach dem Abschalten den Zwicker-Ton hervorrufen, heißen Zwicker-Ton Erzeugerschalle. „Klassischer“ Erzeugerschall ist bandsperrrenbegrenztes Rauschen, jedoch auch ein frequenzmodulierter Ton und weißes Rauschen mit einem überlagerten Sinuston kann einen Zwicker-Ton hervorrufen.

5.1.1 Rauschen

Weißes Rauschen (*white noise*) besteht aus einer Überlagerung von Sinustönen aller Frequenzen mit gleicher Amplitude und zufälliger Phase. Alle Phasen zwischen 0 und 2π treten dabei mit gleicher Wahrscheinlichkeit auf. Dies führt zu einer gaußförmigen Amplitudenverteilung, d.h. die Wahrscheinlichkeitsverteilung der Amplitude von weißem Rauschen zu einem bestimmten Zeitpunkt ist eine Gaußverteilung. Außerdem ist die Intensitätsdichte über alle Frequenzen konstant.

Tiefpassrauschen (*low pass noise*) ist weißes Rauschen, das alle Frequenzen bis zu einer bestimmten Grenzfrequenz (*cutoff frequency*) enthält.

Hochpassrauschen (*high pass noise*) ist weißes Rauschen, das alle Frequenzen ab einer bestimmten Grenzfrequenz enthält.

Bandpassrauschen (*bandpass noise*) ist weißes Rauschen, das alle Frequenzen zwischen zwei Grenzfrequenzen enthält.

Bandsperrenbegrenztes Rauschen (*band reject noise*) ist weißes Rauschen, das alle Frequenzen außer diejenigen zwischen zwei Grenzfrequenzen enthält.

Weißes Rauschen erzeugt keinen Zwicker-Ton. Zwicker-Ton Erzeugerschalle sind aber Tiefpassrauschen, bandsperrenbegrenztes Rauschen mit einer hinreichend breiten Frequenzlücke und, mit Einschränkungen, auch Hochpassrauschen¹. Die Tonhöhe des Zwicker-Tons liegt dann stets im Bereich der Frequenzlücke.

Statt echtem weißen Rauschen kann auch ein Linienspektrum mit Sinustönen gleicher Amplitude und zufälliger Phase in konstantem Frequenzabstand verwendet werden. Ist der Frequenzabstand z. B. 1 Hz, so entsteht ein „weißes“ Rauschen mit einer Periode von 1 s. Ein solches Rauschen kann sehr leicht mit einem Rechner erzeugt werden. Läßt man Frequenzen in einem Bereich weg, so kann man bandsperrenbegrenztes Rauschen mit sehr steilen Flanken erzeugen. Die Periodizität kann deutlich gehört werden, das Rauschen unterscheidet sich aber ansonsten nicht von echtem weißen Rauschen. Deshalb wird im Folgenden Rauschen aus einem dichten Linienspektrum und weißes Rauschen nicht unterschieden, es sei denn, die Linienabstände werden größer als 1 Hz.

5.1.2 Andere Erzeugerschalle

Offenbar muss der Zwicker-Ton Erzeugerschall breite Frequenzbereiche abdecken. Er muss aber auch abrupte Sprünge im Frequenzbereich enthalten. Dies wird durch bandsperrenbegrenztes Rauschen erfüllt. Jedoch auch ein Gleitsinus (Frequenzsweep) von niedrigen zu hohen Frequenzen erfüllt diese Bedingung. Folgen die Sweeps rasch genug aufeinander, so kann auch ein Gleitsinus einen Zwicker-Ton hervorrufen.

Weißes Rauschen allein erzeugt keinen Zwicker-Ton, auch kein einzelner reiner Ton. Überlagert man jedoch weißes Rauschen mit einem reinen Ton, so kann ein Zwicker-Ton erzeugt werden. Dieser liegt in der Tonhöhe stets unterhalb des Sinustons. Auch weißes Rauschen kombiniert mit einem Schmalbandrauschen erzeugt einen Zwicker-Ton, dessen Tonhöhe unterhalb der unteren Bandkante des Schmalbandrauschens liegt.

5.2 Psychoakustische Messungen

Der Zwicker-Ton wird als relativ reiner Ton empfunden. Man kann ihm deshalb recht genau eine bestimmte Tonhöhe zuordnen. Weiter hat er eine bestimmte Dauer und eine bestimmte Lautstärke bzw. einen bestimmten

¹Mit Hochpassrauschen werden allerdings nur otoakustische Emissionen hörbar gemacht, siehe Kap. 5.2.2.

Lautstärkeverlauf.

Lautstärke und Dauer werden am meisten dafür verantwortlich sein, wie „gut“ oder wie deutlich eine Versuchsperson den Zwicker-Ton hört. Dies wird als Qualität des Zwicker-Tons im engeren Sinn bezeichnet. Eine Einteilung der Qualität von „nicht gehört“ bis „sehr gut wahrgenommen“ ist nach Tab. 5.1 möglich ([Krump 1993]).

Symbol	Qualität
++	sehr gut
+	gut
o	mittel
–	schlecht
– –	sehr schlecht
×	nicht gehört

Tab. 5.1: Qualität des Zwicker-Tons und symbolische Bezeichnungen nach [Krump 1993].

Die Zwicker-Ton Erzeugerschalle lassen sich durch verschiedene Parameter, wie beispielsweise Breite der Frequenzlücke und Lautstärke sowie Darbietungslänge des Rauschens, beschreiben. Über die Abhängigkeit der Tonhöhe und der Qualität des Zwicker-Tons von diesen Parametern wird im Folgenden berichtet.

5.2.1 Verbreitung

Pauschal kann man sagen, dass der Zwicker-Ton bei relativ optimalen Bedingungen von über 90 % der Versuchspersonen gehört wird [Krump 1993]. Bei Zwicker hörten etwa 10 % der Versuchspersonen den Zwicker-Ton länger als 6 s [Zwicker 1990].

5.2.2 Otoakustische Emissionen

Schalle, die vom Gehör selbst erzeugt werden, heißen otoakustische Emissionen (*otoacoustic emissions*). Sie werden im verschlossenen Gehörgang mit sehr empfindlichen Mikrofonen gemessen. [Zwicker 1990, Krump 1993]

Eigenschaften otoakustischer Emissionen

Otoakustische Emissionen ohne Stimulation des Gehörs heißen spontan (*spontaneous otoacoustic emissions*). Otoakustische Emissionen, die als Antwort auf eine vorherige Stimulation des Gehörs auftreten, heißen stimuliert (*evoked otoacoustic emissions*). Um sie zu messen, muss sich natürlich auch ein kleiner Lautsprecher im Gehörgang befinden. Die stimulierten otoakustischen Emissionen unterteilt man weiter in simultane und verzögerte otoakustische Emissionen (*spontaneous and delayed evoked otoacoustic emissions*). Spontane otoakustische Emissionen treten während gleichzeitiger Stimulation des

Gehörs auf. Verzögerte otoakustische Emissionen misst man nach dem Abschalten eines äußeren Stimulus. Schließlich gibt es otoakustische Emissionen, die auf nichtlineare Verzerrungsprodukte des Gehörs zurückzuführen sind (*distortion product otoacoustic emissions*).

Spontane otoakustische Emissionen treten bei mehr als 50 % der Versuchspersonen auf. Die Intensität der Emissionen ist nicht sehr stabil, sie schwankt von Tag zu Tag so stark, dass einige Emissionen manchmal auftreten, manchmal nicht. Die Frequenz der spontanen otoakustischen Emissionen ist allerdings sehr stabil, d.h. sie treten immer wieder bei den selben Frequenzen auf.

Der minimale Frequenzabstand zwischen benachbarten spontanen sowie verzögerten otoakustischen Emissionen beträgt ungefähr 0,4 Bark. Außerdem können spontane otoakustische Emissionen auch als simultane stimulierte Emissionen gemessen werden. Dies deutet darauf hin, dass diese drei Arten otoakustischer Emissionen die selbe Ursache haben.

Die Feinstruktur der Ruhehörschwelle hat einen engen Zusammenhang mit den otoakustischen Emissionen: An Stellen mit otoakustischen Emissionen treten Minima der Ruhehörschwelle auf, d.h. das Gehör ist für Frequenzen, an denen otoakustische Emissionen auftreten, sensibler. [Zwicker 1990]

Wird ein sehr leiser Sinuston mit einem Schalldruckpegel von 3 bis 6 phon über der Ruhehörschwelle in der Frequenz durchgestimmt, so treten bei den meisten Versuchspersonen an einigen Stellen Schwebungen auf. Diese Schwebungsstellen korrespondieren mit den otoakustischen Emissionen der jeweiligen Person.

Wechselwirkung otoakustischer Emissionen mit dem Zwicker-Ton

Ein Schall, der normalerweise keinen oder nur einen sehr leisen Zwicker-Ton erzeugen würde, kann einen gut hörbaren Nachton an einer otoakustischen Emissionsstelle erzeugen. Es handelt sich hierbei um ein vom Zwicker-Ton verschiedenes Phänomen. Beispielsweise kann ein Hochpassrauschen, das keinen unbeeinflussten Zwicker-Ton verursacht, einen solchen Ton verursachen.

Der Zwicker-Ton selbst verändert die otoakustischen Emissionen nicht: Wird der Versuchsperson ein geeignetes Rauschen mit spektraler Lücke vorgespielt, so hört sie nach dem Abschalten einen Zwicker-Ton in der Frequenzlücke. Dieser kann auch an der Stelle einer otoakustischen Emission auftreten. Die otoakustischen Emissionen bleiben jedoch vor, während (in der Lücke) und nach dem Rauschen gleich. Der Zwicker-Ton regt also weder eine otoakustische Emission an, die vorher nicht vorhanden war, noch verstärkt er die Intensität einer Emission.

Auch läßt sich der Zwicker-Ton nicht durch Schwebungen nachweisen: Präsentiert man zusätzlich zum Zwicker-Ton einen leisen Sinuston, der in der Frequenz angeglichen werden kann, und läßt die Versuchsperson auf Schwebungen achten, so sind keine Schwebungen feststellbar.

Weiter kann zwar ein zusätzlicher Sinuston den Zwicker-Ton maskieren, aber nicht umgekehrt.

Dies alles deutet darauf hin, dass der Zwicker-Ton nicht durch Schwingungen

der Basilarmembran repräsentiert ist.

5.2.3 Bandbreite des Erzeugerschalls

Bei den im Folgenden beschriebenen Experimenten wird bandpassbegrenztes Rauschen als Erzeugerschall verwendet. Die Tonhöhe des Zwicker-Tons befindet sich stets außerhalb des Rauschens, bei Tiefpassrauschen stets oberhalb der oberen Bandgrenze. Die Tonhöhe bleibt während der Messungen stets konstant, deshalb wird nur die Veränderung der Qualität des Zwicker-Tons beschrieben.

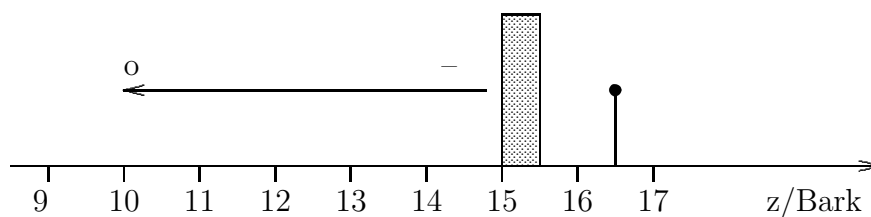


Abb. 5.1: Bandpassbegrenztes Rauschen als Zwicker-Ton Erzeugerschall. Die Höhe des schraffierten Bereichs gibt die Intensitätsdichte des Rauschens bei der jeweiligen Tonheit an. Der Pfeil deutet an, wie der Erzeugerschall bei diesem Experiment verändert wird. Hier wird also die untere Kante des bandpassbegrenzten Rauschens verschoben. Die Symbole über dem Pfeil beschreiben, wie gut der Zwicker-Ton beim jeweiligen Erzeugerschall wahrgenommen wird (siehe Tab. 5.1). Das Symbol \uparrow bezeichnet die Tonhöhe, bei der der Zwicker-Ton wahrgenommen wird.

Als Erzeugerschall werde zunächst ein bandpassbegrenztes Rauschen verwendet (siehe Abb. 5.1). Die obere Bandgrenze sei konstant, die untere wird variiert. Ab etwa 1 Bark Bandbreite kann der Zwicker-Ton bereits „sehr schlecht“ bis „schlecht“ gehört werden. Je größer die Bandbreite, desto besser kann der Zwicker-Ton gehört werden, bis die Qualität bei etwa 5 Bark sättigt, d.h. hier entspricht sie der Qualität bei Tiefpassrauschen.

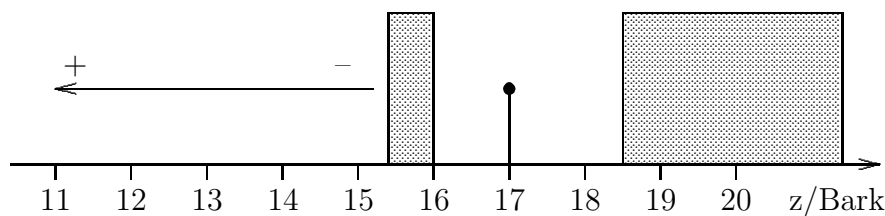


Abb. 5.2: Bandpassbegrenztes Rauschen mit zusätzlichem Hochpassanteil als Zwicker-Ton Erzeugerschall. Die untere Kante des bandpassbegrenzten Rauschens wird verschoben. Bezeichnungen wie in Abb. 5.1.

Nun wird dem bandpassbegrenzten Rauschen ein Hochpassanteil hinzugefügt (siehe Abb. 5.2). Die untere Bandgrenze und damit die Breite des bandbegrenzten Rauschens wird wieder variiert. Mit steigender Breite des Tiefpassbereichs erhöht sich auch hier die Qualität, bis sie bei 5 Bark sättigt. Das Hochpassrauschen erhöht die Qualität des Zwicker-Tons.

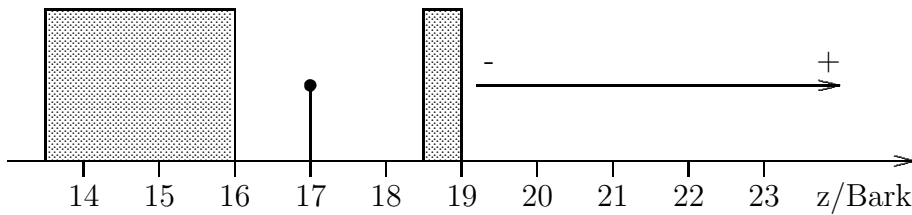


Abb. 5.3: Tiefpassrauschen mit zusätzlichem bandpassbegrenztem Rauschen als Zwicker-Ton Erzeugerschall. Die obere Kante des bandpassbegrenzten Rauschens wird verschoben. Bezeichnungen wie in Abb. 5.1.

Erzeugerschall ist nun Tiefpassrauschen kombiniert mit bandpassbegrenztem Rauschen, dessen obere Grenzfrequenz variiert wird (siehe Abb. 5.3). Je breiter das bandpassbegrenzte Rauschen, desto besser wird der Zwicker-Ton gehört. Ab einer Breite von 5 Bark sättigt die Qualität wieder.

Zusammenfassend kann man sagen, dass der Zwicker-Ton bereits bei einem Schmalbandrauschen der Breite 1 Bark gehört wird. Je breiter das bandpassbegrenzte Rauschen, desto besser wird der Zwicker-Ton gehört. Der Zwicker-Ton verbessert sich zusätzlich, wenn ein weiteres bandpassbegrenzte Rauschen, eine Frequenzlücke lassend, hinzugefügt wird. Auch hier wird die Qualität des Zwicker-Tons um so besser, je breiter das Rauschen wird. Die Qualität verbessert sich bei Breiten des oberen bzw. unteren Anteils von 5 Bark nur mehr unwesentlich.

5.2.4 Bandbreite der spektralen Lücke

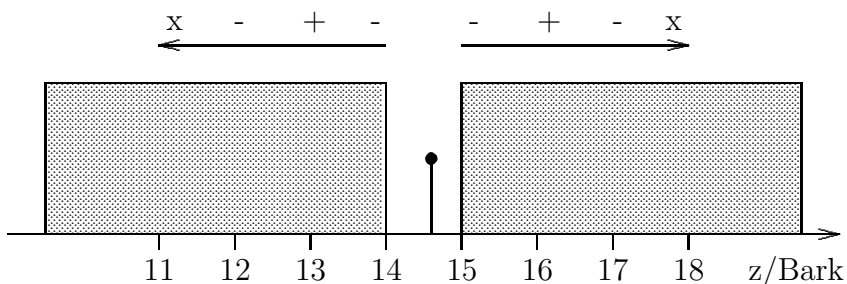


Abb. 5.4: Rauschen mit spektraler Lücke als Zwicker-Ton Erzeugerschall. Bezeichnungen wie in Abb. 5.1.

Als Erzeugerschall wird ein bandsperrbegrenzte Rauschen verwendet, d.h. ein Rauschen mit spektraler Lücke. Die Lückenbreite wird variiert. Optimale Lückenbreiten liegen zwischen 2 und 4,5 Bark. Ab einer Lückenbreite von etwa 5 Bark bewirkt das Hochpassrauschen keine wesentliche Verbesserung der Qualität mehr.

Die Tonhöhe des Zwicker-Tons befindet sich immer etwa 1 Bark oberhalb der oberen Bandgrenze des Tiefpassrauschens. Nur bei Lücken kleiner 2 Bark wird die Tonhöhe mehr zur Mitte der Bandlücke hin verschoben.

5.2.5 Lautstärke des Erzeugerschalls

Als Erzeugerschall wird bandsperrbegrenztetes Rauschen mit einer Frequenzlücke von etwa 2,5 Bark verwendet. Je lauter das Rauschen, desto mehr verschiebt sich die Tonhöhe des Zwicker-Tons zu höheren Frequenzen.

Die Qualität des Zwicker-Tons nimmt mit steigender Lautstärke des Erzeugerschalls zunächst zu, bis sie ein Optimum erreicht, und dann wieder ab.

5.2.6 Tiefe der spektralen Lücke

Als Erzeugerschall wird wieder bandsperrbegrenztetes Rauschen mit einer Frequenzlücke von etwa 2,5 Bark verwendet. Je tiefer die Lücke, desto besser ist der Zwicker-Ton zu hören.

5.2.7 Andere Einflüsse

Viele Schalle, die weißem Rauschen mit spektraler Lücke ähneln, können einen Zwicker-Ton hervorrufen.

Variiert man etwa die Steilheit der Flanke des Tiefpassrauschens, so wird der Zwicker-Ton zu höheren Tonhöhen hin „abgestoßen“. Verschiebt man das Hochpassrauschen zu niedrigeren Frequenzen oder stattet es ebenfalls mit einer Flanke aus, so wird der Zwicker-Ton zu niedrigeren Frequenzen verschoben.

5.3 Berechnung der Tonhöhe

Nach [Krump 1993] berechnet man zunächst die Erregung des Zwicker-Ton Erzeugerschalls nach Kap. 1.5.1.

Die Tonhöhe des Zwicker-Tons befindet sich dann bei der Frequenz, bei der die Erregung die Ruhehörschwelle schneidet (siehe Abb. 5.5). Kommt ein solcher Schnittpunkt aufgrund des Hochpassrauschens nicht zustande, so befindet sich der Zwicker-Ton am Minimum der Erregung. Der Zwicker-Ton ist bei einer Pegeldifferenz zwischen maximaler und minimaler Erregung von 15 dB gerade noch wahrnehmbar.

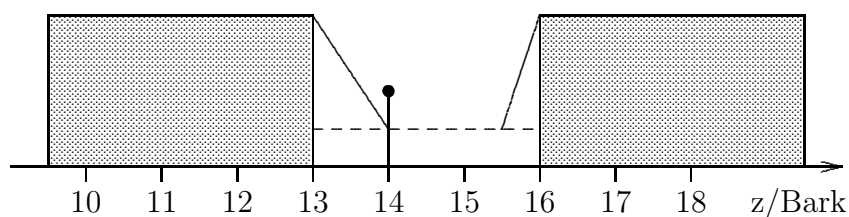


Abb. 5.5: Die wahrgenommene Tonhöhe des Zwicker-Tons befindet sich am Schnittpunkt zwischen der Erregung des Erzeugerschalls (durchgezogen eingezeichnet) und der Ruhehörschwelle (gestrichelt eingezeichnet).

Bei einem Erzeugerschall, der aus weißem Rauschen und einem überlagerten reinen Ton besteht, befindet sich der Zwicker-Ton an der Stelle, an der die Erregung des reinen Tons die Erregung des Rauschens schneidet.

Man beachte, dass es sowohl bei der Frequenzlücke als auch bei dem reinen Ton mit überlagertem Rauschen als Erzeugerschall zwei Schnittpunkte zwischen der durch den Erzeugerschall hervorgerufenen Erregung und der Ruhehörschwelle gibt. Das in [Krump 1993] vorgeschlagene Modell zur Berechnung der Tonhöhe des Zwicker-Tons sagt leider nicht voraus, an welchem der beiden Schnittpunkte Versuchspersonen den Zwicker-Ton wahrnehmen.

5.4 Offene Fragen

Offen bleibt, warum der Zwicker-Ton überhaupt entsteht. Weiter ist nicht geklärt, warum der Zwicker-Ton ein fast reiner Ton ist und nicht ein zum Schall komplementäres Erregungsmuster besitzt. Außerdem kann die Lautstärke und Dauer bzw. Qualität des Zwicker-Tons nicht bestimmt werden. Die Qualität des Zwicker-Tons hängt von Anteilen des Erzeugerschalls ab, die um mehr als 5 Bark von der Tonheit des Zwicker-Tons entfernt sind. Erregungsmuster (Kap. 1.5.1) haben aber eine weit geringere Ausdehnung.

Kapitel 6

Modelle für den Zwicker-Ton

Der Zwicker-Ton entsteht nicht in der Cochlea, ausgenommen im Fall otoakustischer Emissionen (siehe Kap. 5.2.2). Auch im Hörnerv wurde er bisher nicht nachgewiesen [Hoke 1996]. Mit magnetoenzephalographischen Methoden wurde allerdings ein neurophysiologisches Korrelat des Zwicker-Tons im auditorischen Cortex des Menschen festgestellt [Hoke 1996]. Durch elektro-physiologische Ableitungen am auditorischen Cortex des Chinchilla konnten [Tomlinson, Biebel, Langner 1998] feststellen, dass Neurone nach dem Abschalten des Zwicker-Ton Erzeugerschalls eine erhöhte Feuerrate zeigen. Diese liegen im Wesentlichen entlang eines tonotopen Steifens (zur Tonotopie siehe Kap. 2.2.1).

Der erste Ort, an dem der Zwicker-Ton entstehen könnte, ist die erste Verarbeitungsstufe der Signale aus dem Hörnerv, der Nucleus cochlearis (Beschreibung in Kap. 2.2.1). Für eine Entstehung des Zwicker-Tons in einer Verarbeitungsstufe, die nur durch ein Ohr stimuliert wird, spricht nach [Krump 1993, Hoke 1996] auch, dass der Zwicker-Ton ein rein monaurales Phänomen ist. Verarbeitungsstufen oberhalb des Nucleus cochlearis sind bereits binaural (siehe Abb. 2.1 in Kap. 2.2).

Da dies die einfachste Hypothese ist, gehen die im Folgenden vorgestellten Modelle davon aus, dass der Zwicker-Ton bereits im Nucleus cochlearis entsteht. Dies ist, wie wir sehen werden, für das Modell allerdings nicht wesentlich. Wesentlich ist nur, dass in einer zentralen tonotopen Verarbeitungsstufe im Gehör ein Signal ankommt, das in seiner Zusammensetzung den Aktionspotentialen im Hörnerv ähnelt. Der Verlauf der Feuerraten in den ankommenden Nervenfasern, etwa bei Stimulation durch Tiefpassrauschen oder einen reinen Ton, sollte also ungefähr den Feuerraten im Hörnerv entsprechen.

In den folgenden beiden Kapiteln werden zwei Modelle zur Entstehung des Zwicker-Tons vorgestellt, das Habituationsmodell und das Adaptionsmodell. Das Habituationsmodell geht von einem weitgehend passiven Nucleus cochlearis aus, während das Adaptionsmodell kompliziertere Rückkopplungsmechanismen voraussetzt. Die wesentlichen Ideen beider Modelle werden kurz vorgestellt. Es folgen eine genaue Beschreibung der Simulationen, einige Simulationsergebnisse sowie eine Diskussion der Eigenschaften des jeweiligen Modells.

6.1 Habituationsmodell

In diesem Kapitel wird ein Modell zur Erklärung des Zwicker-Tons vorgestellt, das auf einer Ermüdung (Habituation) der Neuronen im Nucleus cochlearis beruht. Nach einer Beschreibung der grundlegenden Ideen zu diesem Modell werden die Simulationen und Simulationsergebnisse detailliert beschrieben. Abschließend werden die Simulationen qualitativ mit der Realität verglichen.

6.1.1 Beschreibung des Habituationsmodells

In Kap. 4.3 haben wir gesehen, dass laterale Inhibition die Detektierbarkeit von Signalen verbessern kann. Laterale Inhibition ist im Nucleus cochlearis nachgewiesen (siehe Kap. 4.1). Wie wir in Kap. 2.2.1 gesehen haben, besteht der Nucleus cochlearis aus Streifen tonotop angeordneter Neurone. Das einfachste Modell für den Nucleus cochlearis ist eine Reihe tonotop angeordneter Neuronen mit lateraler Inhibition (vergleiche auch das Tinnitusmodell in Kap. 3.2). Anhand der Erfahrungen aus Kap. 4.3 wissen wir, wie stark die laterale Inhibition eingestellt werden muss bzw. höchstens eingestellt werden darf, damit die Wahrnehmung von reinen Tönen verbessert wird. Je stärker die laterale Inhibition, umso deutlicher ist nämlich der Effekt der Überhöhung von Kanten (Kap. 4.2.2). Diese Kantenüberhöhung führt letztlich zum Zwicker-Ton, wie wir weiter unten sehen werden.

Hypothese im Habituationsmodell für den Zwicker-Ton ist eine zusätzliche Habituation (Gewöhnung oder Ermüdung) der Neuronen, d.h. bei anhaltendem Reiz und somit anhaltendem Feuern ermüden sie und die Feuerrate sinkt. Bis sich die Neuronen wieder erholt haben, vergeht eine Zeitspanne, die etwa so lang ist wie die Wahrnehmung des Zwicker-Tons.

Das Szenario zur Erklärung des Zwicker-Tons sieht dann etwa so aus (siehe Abb. 6.1): Betrachten wir Tiefpassrauschen als Zwicker-Ton Erzeugerschall. Die Neurone, die auf das Rauschen ansprechen, ermüden. D.h. die spontane Feuerrate dieser Neuronen nach Abschalten des Rauschens wird niedriger sein als die spontane Feuerrate der Neurone, die nicht dem Rauschen ausgesetzt waren. Dies bedeutet wiederum, dass die laterale Inhibition, die von den ermüdeten Neuronen ausgeht, geringer ist als die Inhibition, die von erhaltenen Neuronen ausgeht. Hieraus folgt, dass die Feuerrate der Neurone, die gerade nicht mehr durch das Tiefpassrauschen angeregt wurden, nun höher ist als die Feuerrate der Neuronen im ausgeruhten Gehör. Denn die Neuronen an der Kante des Tiefpassrauschens werden ja von den ermüdeten Neuronen bei tieferen Frequenzen nicht mehr inhibiert. Die Neuronen an der Kante des Tiefpassrauschens haben also nach dem Abschalten des Rauschens eine höhere Feuerrate als ohne jegliche Stimulation durch Schall. Liegt diese etwas höhere Feuerrate nun deutlich oberhalb der Spontanaktivität, so muss man annehmen, dass diese Feuerrate zu einer Wahrnehmung führt. Denn die Hörschwelle sollte so empfindlich wie möglich eingestellt sein, d.h. so empfindlich, dass es gerade nicht zu einer Schallwahrnehmung kommt, wenn kein äußerer Schall vorhanden ist.

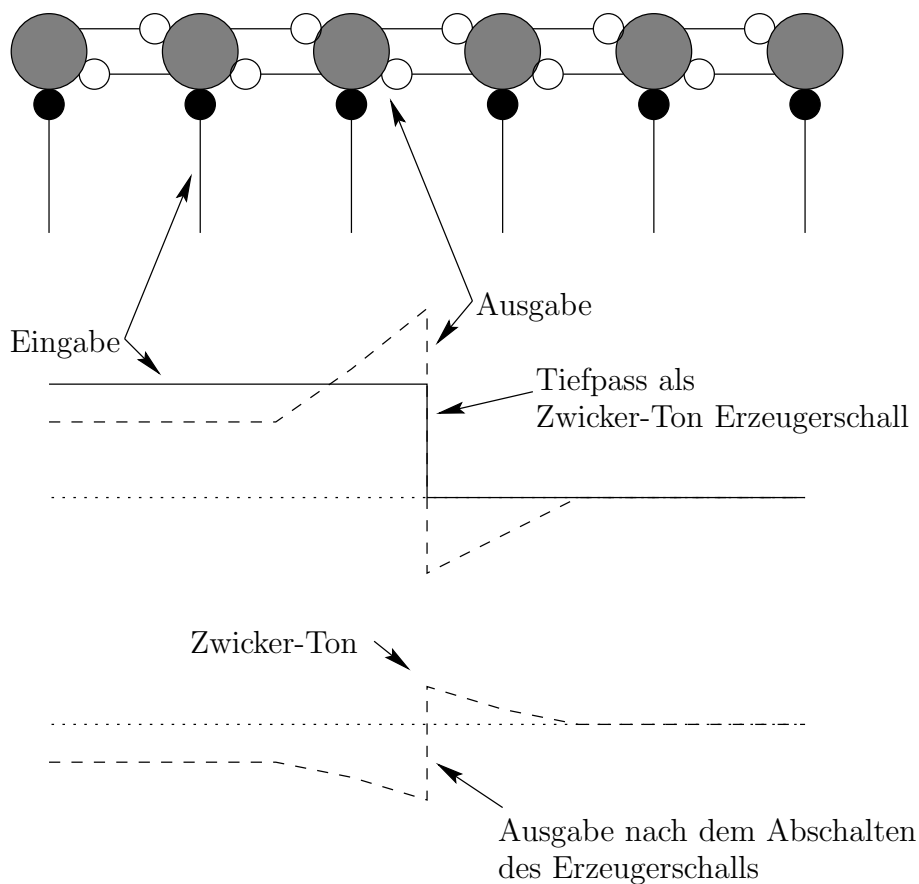


Abb. 6.1: Illustration der Entstehung des Zwicker-Tons im Habitationsmodell. Im oberen Bildteil bedeuten große schraffierte Kreise Neurone, durchgezogene Linien Axone, ausgefüllte kleine Kreise exzitatorische und offene kleine Kreise inhibitorische Synapsen. Die Kurven unterhalb stellen die Eingabe (durchgezogene Linie) und Ausgabe (gestrichelte Linie) der Neuronenschicht vor (oben) und nach (unten) dem Abschalten des Tiefpassrauschens schematisch dar. Die gepunktete Linie bedeutet relativ zur Eingabe (durchgezogen) Stimulus Null und relativ zur Ausgabe (gestrichelt) die spontane Aktivität.

6.1.2 Beschreibung der Simulationen des Habitationsmodells

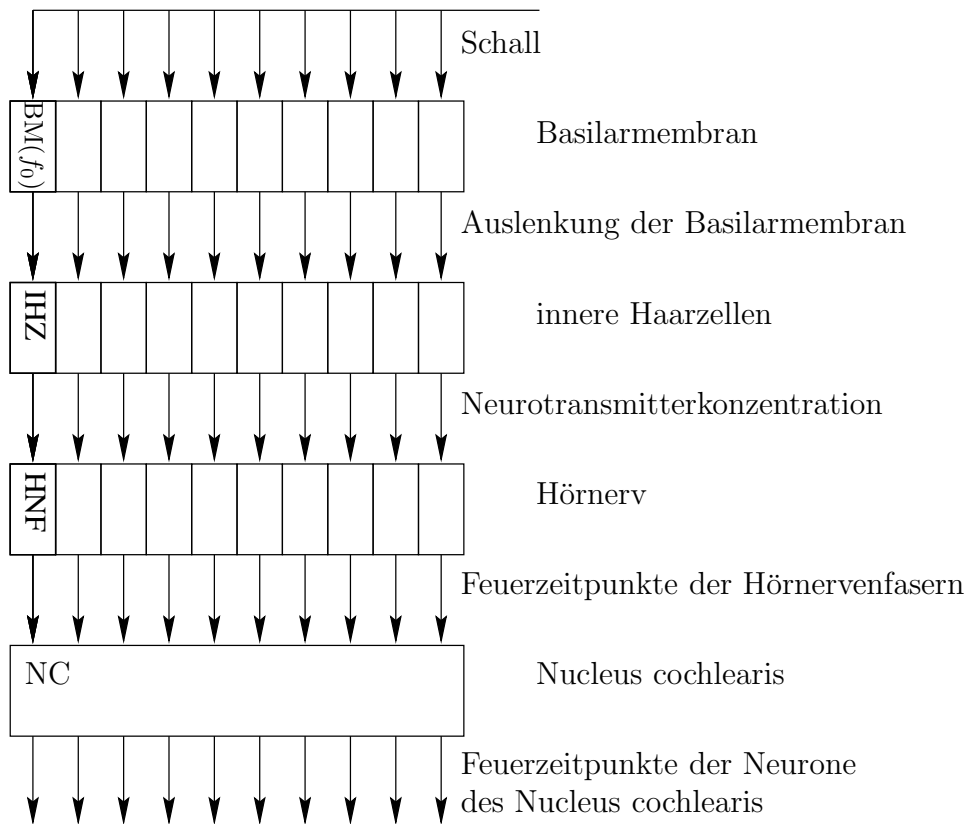


Abb. 6.2: Die Bestandteile des verwendeten Gehörmodells. Kästen bedeuten Verarbeitungseinheiten, Pfeile Datenströme zwischen den Verarbeitungseinheiten.

Die Simulation besteht aus den in Abb. 6.2 gezeigten Modulen. Diese sind die Simulation der Basilar membran, der inneren Haarzellen, des Hörnervs und des Nucleus cochlearis.

Abbildung 6.3 zeigt die Simulation der Basilar membran an der Stelle mit Bestfrequenz f_0 wie in Kap. 3.1.1 beschrieben. Eingabe dieses Moduls ist der Schalldruck $p(t)$, Ausgabe die Auslenkung $y(t)$ der Basilar membran an der Stelle mit Bestfrequenz f_0 . Die wechselnde Auslenkung der Basilar membran an einer Stelle erregt die jeweilige innere Haarzelle.

Abbildung 6.4 zeigt die Simulation einer inneren Haarzelle nach dem Meddis-Modell (Kap. 3.1.2). Eingabe dieses Moduls ist die Auslenkung der Basilar membran am Ort der inneren Haarzelle, Ausgabe ist die Neurotransmitterkonzentration $c(t)$ im synaptischen Spalt, an den der Hörnerv anschließt.

Abbildung 6.5 beschreibt, wie nach Kap. 3.1.3 die Aktionspotentiale in den Hörnervenfasern berechnet werden.

Abbildung 6.6 verdeutlicht die angenommene Verschaltung der Neuronen im Nucleus cochlearis. Jedes Neuron inhibiert über ein inhibitorisches Interneuron seine Nachbarn. Dabei werden allerdings nicht nur die nächsten Nachbarn inhibiert, sondern eine ganze Umgebung (ähnlich wie in Kap. 4.3.2).

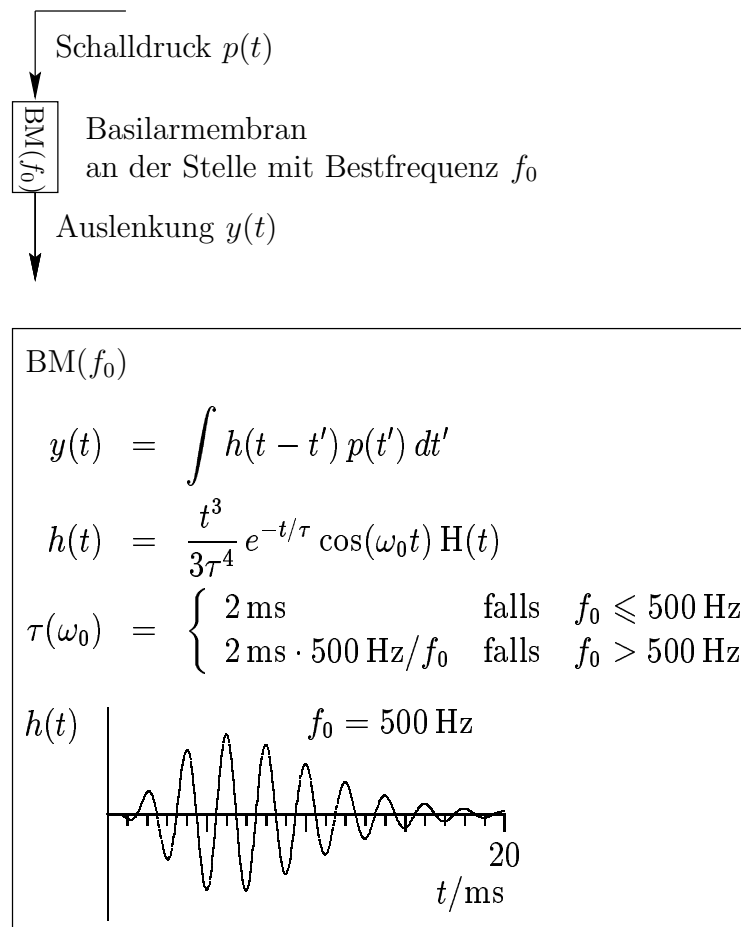
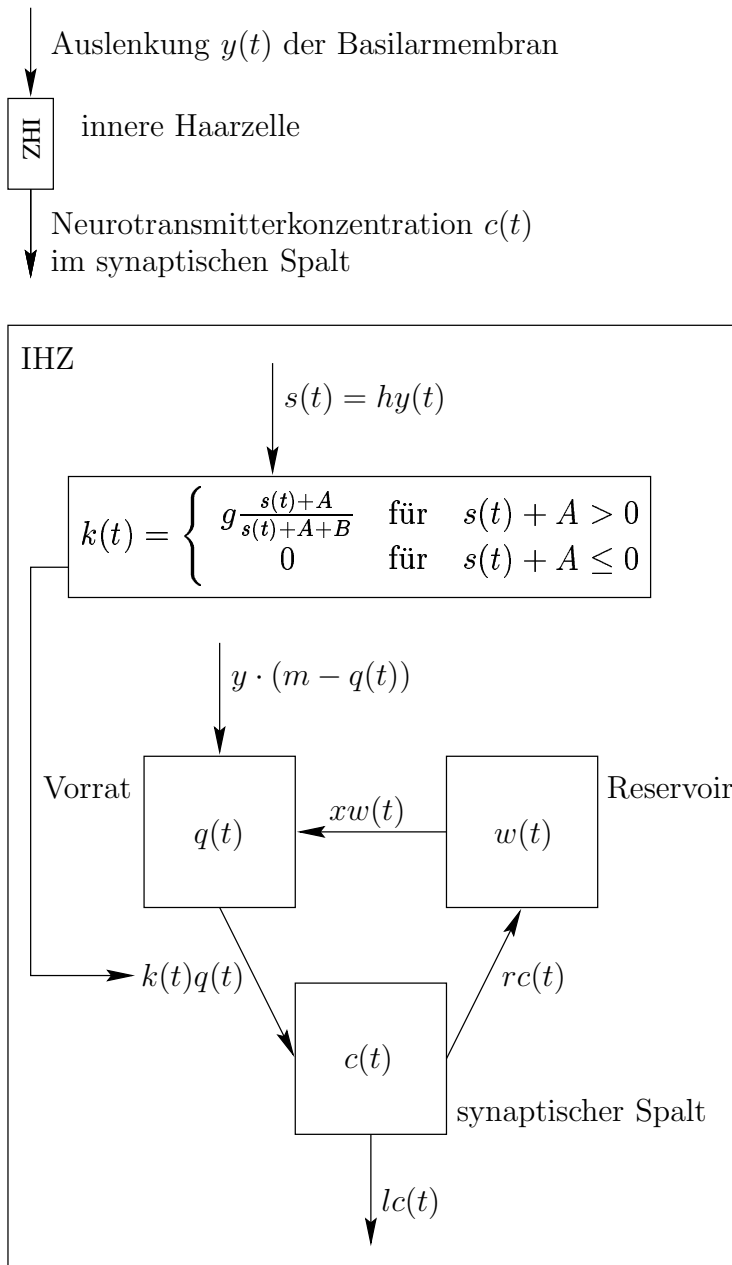


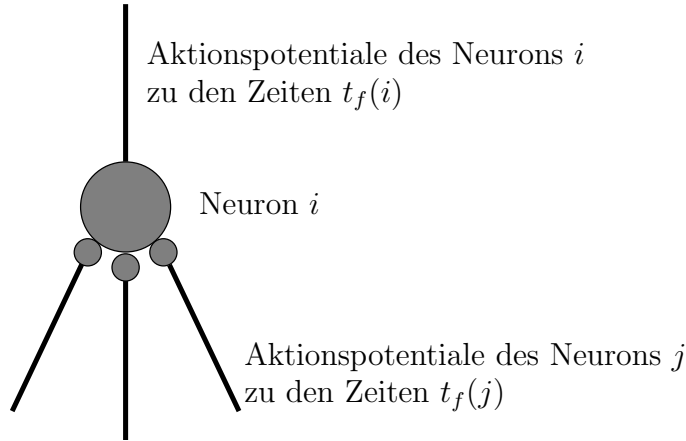
Abb. 6.3: Simulation der Basilarmembran an der Stelle mit Bestfrequenz $f_0 = \omega_0/2\pi$. Signaturen wie in Abb. 6.2.

Abbildung 6.7 fasst nocheinmal das verwendete Spike-Response-Neuronenmodell zusammen (vgl. auch Kap. 4.3.2).

Die Neuronen der Outputschicht sollen jedoch nach den Erläuterungen in Kap. 6.1.1 ermüden. Dies wird durch eine in Abb. 6.8 erläuterte Modifikation des Neuronenmodells aus Abb. 6.7 erreicht: Ein zusätzliches, langsam abklingendes Refraktärpotential sorgt dafür, dass die Feuerrate des Neurons nach andauerndem Feuern sinkt. Das Neuron erholt sich nach Wegfall der starken Erregung nur langsam.

**Abb. 6.4:**

Simulation einer inneren Haarzelle nach dem Meddis-Modell (Kap. 3.1.2).
 Signaturen wie in Abb. 6.2. In der Simulation ist $y(t) \in [-1, 1]$ und $h = 50$.
 Zu den Werten der übrigen Konstanten siehe Tab. 3.1.

Neuron i

$$h_{syn}(i, t) = \sum_j J(i, j) \sum_{t_f(j)} \varepsilon(t - t_f(j))$$

$$h_{ref}(i, t) = \sum_{t_f(i)} \eta(t - t_f(i))$$

$$h(i, t) = h_{syn}(i, t) + h_{ref}(i, t)$$

$$\varepsilon(t) = \frac{t}{\tau_s} e^{1-t/\tau_s} \quad \text{für } t \geq 0$$

$$\eta(t) = \begin{cases} -\infty & \text{für } 0 < t \leq \gamma_{ref} \\ -\frac{\tau_{ref}}{t - \gamma_{ref}} & \text{für } t > \gamma_{ref} \end{cases}$$

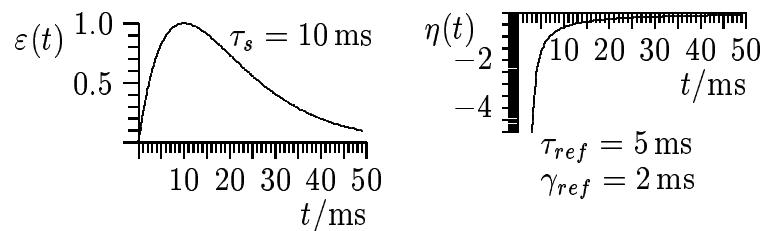
Aktionspotential, falls $h(i, t) > \theta$

Abb. 6.7: Das Neuronenmodell. Graue Punkte bezeichnen (exzitatorische oder inhibitorische) Synapsen. Sonstige Signaturen wie in Abb. 6.2 und 6.6.

ermüdendes Neuron

$$h_{ref}(i, t) = \sum_{t_f(i)} \eta(t - t_f(i)) + \eta_{hab}(t - t_f(i))$$

$$\eta_{hab}(t) = -\gamma_{hab} e^{-t/\tau_{hab}} \quad \text{für } t \geq 0$$

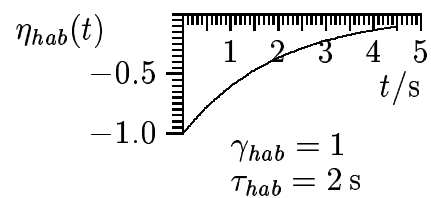


Abb. 6.8: Das Modell für die ermüdenden Neuronen der Outputschicht. Sie besitzen zusätzlich zum kurzzeitigen Refraktärpotential aus Abb. 6.7 ein lang anhaltendes Refraktärpotential. Signaturen wie in Abb. 6.2 und 6.7.

6.1.3 Simulationsergebnisse im Habituationsmodell

In diesem Abschnitt wird die Reaktion des Habituationsmodells auf verschiedene Stimuli beschrieben. Farbbilder mit zeitabhängigen mittleren Feuerraten sind in Anhang A zusammengefasst.

Simuliert wurden bei jedem Stimulus 300 Hörnervenzfasern, die mit 300 Neuronen der Outputschicht verbunden sind. Die Neuronen der Outputschicht sind wiederum mit jeweils einem inhibitorischen Interneuron verbunden, wie in Abb. 6.6 beschrieben. Die Neuronen am Rand des simulierten tonotopen Bereichs werden sehr stark dadurch beeinflusst, dass die laterale Inhibition von jenseits des simulierten Bereichs fehlt. Um den Leser durch diese Randeffekte nicht zu verwirren, werden in den Abbildungen nur die Neuronen 50–250 gezeigt.

Tiefpassrauschen

Stimulus ist nun Tiefpassrauschen mit einer oberen Bandkante von 1300 Hz. Abbildung 6.9 (oben) zeigt die durchschnittliche Feuerrate der Hörnervenzfasern, die die Eingabe zu den Neuronen der Outputschicht bilden.

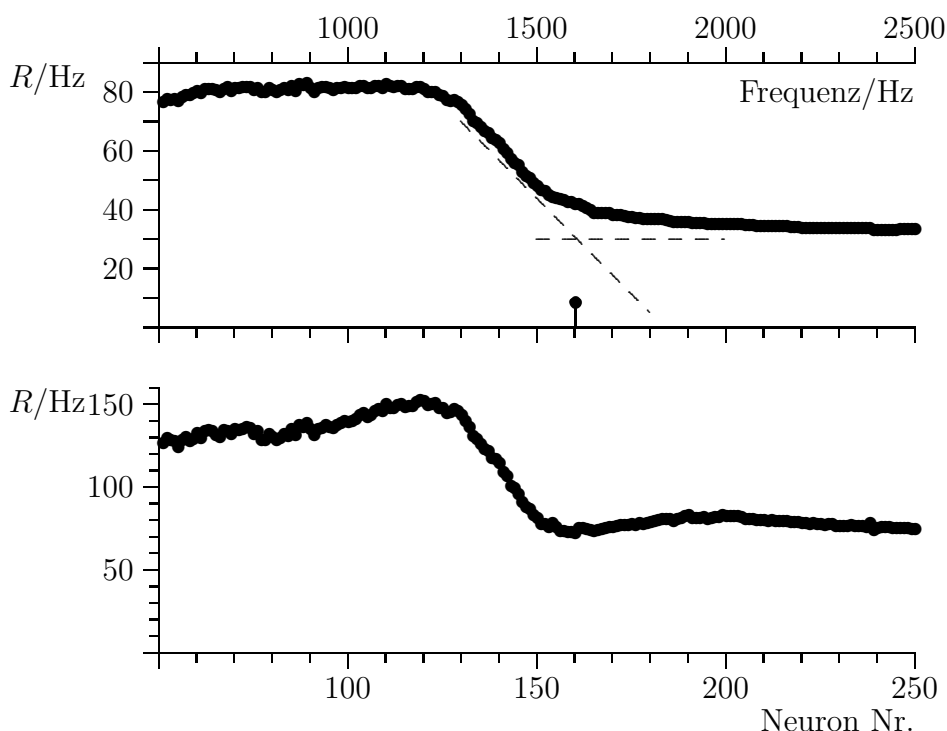


Abb. 6.9: Die Feuerraten R der Hörnervenzfasern (oben) und der Neuronen der Outputschicht (unten) im Habituationsmodell, gemittelt über die Dauer des Stimulus von 2 s. Stimulus ist Tiefpassrauschen mit einer oberen Bandkante von 1300 Hz. Die Frequenzskala bezeichnet die Bestfrequenzen der Neuronen. Der Zwicker-Ton tritt im Habituationsmodell an der mit \uparrow bezeichneten Stelle auf. Die gestrichelten Linien deuten die Berechnung der Tonhöhe des Zwicker-Tons im Mithörschwellenmodell von [Krump 1993] (Kap. 5.3) an, näheres siehe Kap. 6.1.4.

Abbildung A.2 zeigt außerdem den zeitlichen Verlauf der Feuerraten der Neuronen der Outputschicht aus Abb. 6.6 während und nach dem Abschalten des Stimulus. Man erkennt deutlich die erhöhte Aktivität im Bereich etwas oberhalb von 1300 Hz nach dem Abschalten des Stimulus nach 2 s. Grüne Bereiche bedeuten spontane Aktivität der Neurone, orange bis rote Bereiche Aktivität oberhalb der Hörschwelle und blaue Bereiche geringere als spontane Aktivität.

In Abb. 6.9 ist die Wirkung der lateralen Inhibition erkennbar. Die Kante des Tiefpass bei 1300 Hz wird deutlich hervorgehoben.

Erklärung der Simulationsergebnisse: Eingabe ist ideales Tiefpassrauschen mit einer oberen Bandkante von 1300 Hz. Der Tiefpass wird durch die (nicht idealen) cochleären Filter aus Kap. 3.1.1 so beeinflusst, dass die Bandkante flacher wird. D.h. auch Stellen der Basilarmembran mit Bestfrequenzen größer als 1300 Hz werden noch angeregt. Dies führt natürlich nach Kap. 3.1.3 dazu, dass auch Hörnervenfaser mit Bestfrequenzen oberhalb 1300 Hz noch feuern. Dies zeigt der obere Teil von Abb. 6.9. Laterale Inhibition gemäß Abb. 6.6 wirkt nun so wie in Kap. 4.2 näherungsweise berechnet wurde, d.h. Bandkanten werden aufgebogen (vgl. Abb. 4.1). Die Wirkung der lateralen Inhibition sieht man im unteren Teil von Abb. 6.9. Feuerraten von Neuronen mit Bestfrequenzen unterhalb der Bandkante des Tiefpassrauschens werden erhöht, Feuerraten von höherfrequenten Neuronen erniedrigt. Je höher die Feuerrate der Neurone, desto stärker auch die in Abb. 6.8 beschriebene Ermüdung. Diese führt nach dem Abschalten des Stimulus zur relativ geringen Feuerrate der Neuronen bis Nr. 140 in Abb. A.2. Aufgrund der geringen Feuerrate der Neuronen bei niedrigen Frequenzen geht von diesen keine starke laterale Inhibition mehr aus. D.h. die (frischen) Neuronen bei höheren Frequenzen feuern stärker als im unbeeinflussten Gehör. Die Simulation zeigt, dass diese Feuerrate signifikant ist. Dies bedeutet, die erhöhten Feuerraten sind so hoch, dass sie zweifellos (d.h. mit geringer Irrtumswahrscheinlichkeit, siehe Kap. 4.3.1) als solche erkannt werden können, denn solch hohe Feuerraten treten im unbeeinflussten Gehör praktisch nirgends auf. Signifikant erhöhte Feuerraten bedeuten unter der Hypothese einer möglichst empfindlichen Einstellung des Gehörs eine Schallwahrnehmung. Dabei handelt es sich um einen relativ reinen Ton, da die Erhöhung der Feuerrate auf einen sehr engen Frequenzbereich beschränkt ist¹.

Reiner Ton mit Rauschen

Als Stimulus verwenden wir nun einen reinen Ton mit darunterliegendem weißen Rauschen. Der Sinuston hat eine Amplitude von 0,7, das Rauschen besteht aus einzelnen Sinustönen zufälliger Phase im Abstand 1 Hz mit Amplitude 0,007. Die Simulationsergebnisse zeigen Abb. 6.10 und A.3.

Erklärung: Der Zwicker-Ton tritt bei Tonhöhen unterhalb des reinen Tons

¹Beachte, dass selbst ein reiner Ton die Cochlea und den Hörnerv nach Kap. 3.1.1 und 3.1.3 in vielen Frequenzkanälen mehr oder weniger stark anregt. Eine Anregung in einem relativ engen Frequenzbereich führt also zur Wahrnehmung eines reinen Tons und nicht etwa zur Wahrnehmung von Schmalbandrauschen.

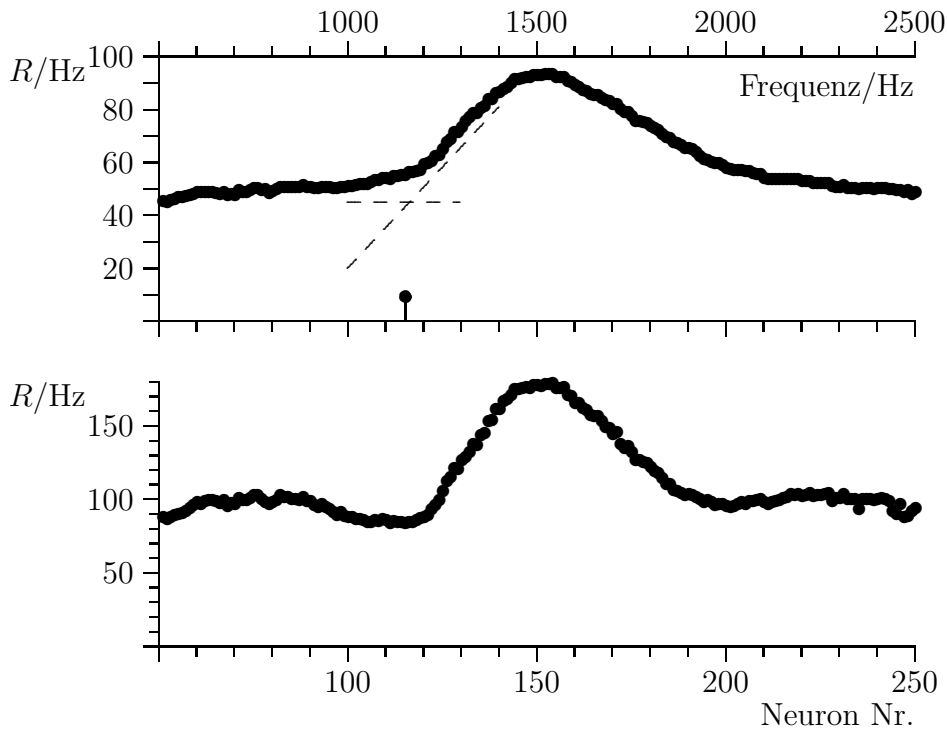


Abb. 6.10: Die Feuerraten R der Hörnervenzfasern (oben) und der Neuronen der Outputschicht (unten) im Habituationsmodell, gemittelt über die Dauer des Stimulus von 2 s. Stimulus war ein Sinuston der Frequenz 1500 Hz und darunterliegendem weißem Rauschen.

auf. Dies stimmt mit den Beobachtungen in Kap. 5.1.2 und Kap. 5.3 überein. Der Grund liegt in der in Kap. 3.1.1 beschriebenen Asymmetrie der cochleären Filter. Die Erregung fällt nach tieferen Tonhöhen wesentlich schneller ab als nach hohen Tonhöhen (siehe auch Kap. 1.5.1). Dies sieht man in Abb. 6.10. Der schnellere Abfall der Erregung nach tieferen Tonhöhen bewirkt auch eine stärkere Inhibition in Richtung tieferer Tonhöhen (siehe Abb. 6.10). Die nach dem Abschalten des Stimulus entstehende Kante ist also bei tieferer Tonhöhe wesentlich steiler als die Kante bei höherer Tonhöhe als der reine Ton. Die steilere Kante wird stärker hervorgehoben als die flachere. Deshalb ist die Aktivität der Neuronen an der tieferfrequenten Kante höher als an der höherfrequenten. Der Zwicker-Ton tritt also nur bei einer Tonhöhe unterhalb der Tonhöhe des Stimulus auf oder ist hier zumindest wesentlich deutlicher.

Reiner Ton

Bei einem reinen Ton als Stimulus ist der Zwicker-Ton im Habituationsmodell in Abweichung zur Realität noch deutlicher als bei einem reinen Ton mit Rauschen. Das zusätzliche Rauschen bewirkt nur Ermüdung im Bereich der Tonhöhe des Zwicker-Tons, schwächt also den Zwicker-Ton im Habituationsmodell eher ab.

Hochpassrauschen

Die Abb. 6.11 und A.4 sind analog zu den bisher erläuterten Abbildungen in diesem Abschnitt. Man beachte jedoch, dass die ansteigende Bandkante des Hochpassrauschens steiler als die abfallende Bandkante des Tiefpassrauschens ist. Da die Kante sehr steil ist, ist auch der Zwicker-Ton recht deutlich, deutlicher als im Fall des Tiefpassrauschens (vgl. Kap. 6.1.3). Dies widerspricht der Realität. Tatsächlich wird bei Hochpassrauschen als Stimulus kein Zwicker-Ton wahrgenommen.

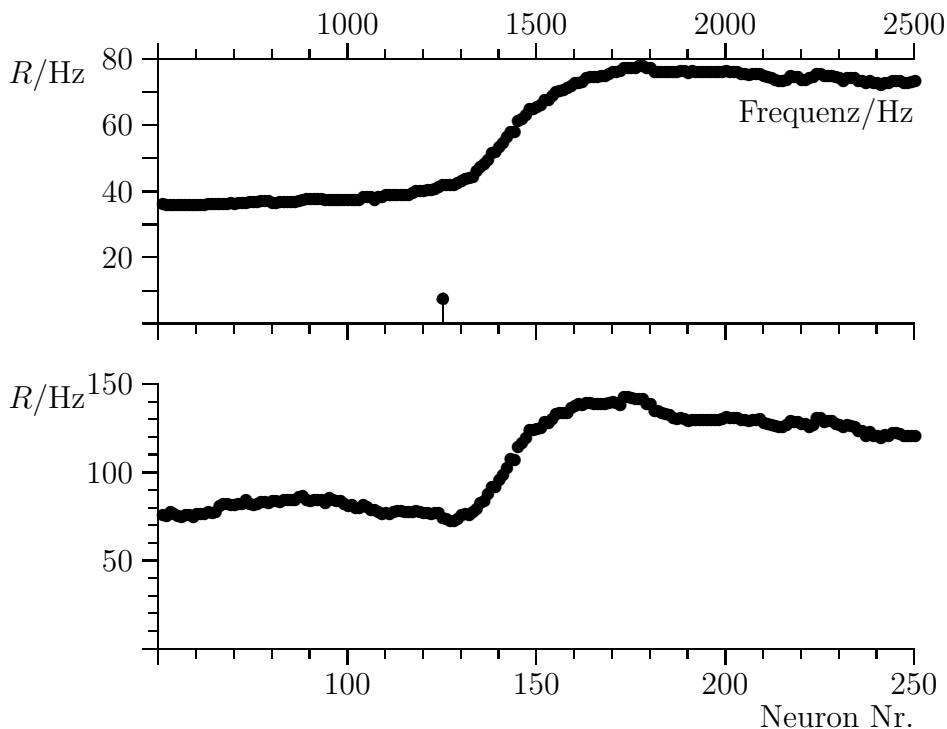


Abb. 6.11: Die Feuerraten R der Hörnervenfasern (oben) und der Neuronen der Outputschicht (unten) im Habituationsmodell, gemittelt über die Dauer des Stimulus von 2 s. Stimulus war Hochpassrauschen mit einer unteren Bandkante von 1500 Hz.

Tiefpass- und Hochpassrauschen mit Rauschen

Tiefpass- und Hochpassrauschen mit zusätzlichem weißem Rauschen erzeugt im Habituationsmodell keinen Zwicker-Ton, da alle Neuronen erregt werden und somit alle Neuronen ermüden. Der Kanteneffekt reicht bei hinreichend intensivem weißem Rauschen nicht mehr aus, um einen Zwicker-Ton oberhalb der Hörschwelle zu erzeugen.

6.1.4 Diskussion der Simulationsergebnisse im Habituationsmodell

Tonhöhe des Zwicker-Tons im Habituationsmodell

Die nach Kap. 5.3 berechnete Tonhöhe des Zwicker-Tons stimmt mit der Tonhöhe des Zwicker-Tons im Habituationsmodell ziemlich genau überein (vgl. die Schnittpunkte der gestrichelten Linien in Abb. 6.9 und 6.10 mit der eingezeichneten Tonhöhe des Zwicker-Tons im Habituationsmodell).

Im Fall des Tiefpassrauschens als Zwicker-Ton Erzeugerschall tritt der Zwicker-Ton nach Kap. 5.2.3 oberhalb der oberen Bandkante des Tiefpassrauschens auf. Dies ist auch im Habituationsmodell der Fall. Weiter tritt der Zwicker-Ton nach Kap. 5.3 am Schnittpunkt der Erregung des Tiefpassrauschens mit der Ruhehörschwelle auf. Dieser Schnittpunkt ist in den Abb. 6.9 und 6.10 als Schnitt zweier gestrichelter Linien eingezeichnet, die die Erregung durch die jeweiligen Bestandteile des Erzeugerschalls beschreiben. Im Habituationsmodell entspricht der Schnittpunkt der Tonhöhe, bei der die Feuerrate der Neuronen etwa auf die spontane Aktivität abgeklungen ist. Dies ist der Ort, an dem nach dem Abschalten des Zwicker-Ton Erzeugerschalls die Kante in den Feuerraten auftritt, die dann durch laterale Inhibition verstärkt wird. D.h. die Tonhöhe des Zwicker-Tons wird vom Modell korrekt wiedergegeben.

Im Fall des einzelnen reinen Tons mit weißem Rauschen als Zwicker-Ton Erzeugerschall tritt der Zwicker-Ton nach Kap. 5.1.2 unterhalb der Tonhöhe des reinen Tons auf. Dies ist auch im Habituationsmodell der Fall. Außerdem gilt auch hier das in Kap. 5.3 beschriebene Modell zur Berechnung der Tonhöhe des Zwicker-Tons. Die Tonhöhe des Zwicker-Tons befindet sich am Schnittpunkt der Erregung des reinen Tons und der Erregung des darunterliegenden weißen Rauschens. Im Modell entspricht dieser Schnittpunkt wieder der Tonhöhe, bei der die Feuerrate der durch den reinen Ton angeregten Neuronen etwa auf die Feuerrate abgeklungen ist, die durch das weiße Rauschen allein verursacht wird. Diese Tonhöhe liegt im Modell etwa an der Stelle, an der der Zwicker-Ton auftritt (vgl. Abb. A.3 und 6.10).

Zwicker-Ton Erzeugerschalle im Habituationsmodell

Tabelle 6.1 vergleicht die Wirkung der in Kap. 5.2 und [Krump 1993] aufgeführten Schalle mit dem Habituationsmodell.

Die entscheidende Abweichung zwischen der Realität und dem Habituationsmodell ist, dass im Habituationsmodell Hochpassrauschen Zwicker-Ton Erzeugerschall ist, in der Realität dagegen nicht. Hochpassrauschen ist im Habituationsmodell sogar besserer Erzeugerschall als Tiefpassrauschen, denn die Bandkante der Erregung des Hochpassrauschens ist wegen der Asymmetrie der cochleären Filter (siehe Kap. 3.1.1) wesentlich steiler als die Bandkante der Erregung des Tiefpassrauschens. Somit wird der Zwicker-Ton-Effekt deutlicher. Auch ein reiner Ton ohne darunterliegendem weißen Rauschen erzeugt, wie ein Hochpass, einen deutlichen Zwicker-Ton in Abweichung zur

Schall	Zwicker-Ton laut psychoakustischer Messung	Zwicker-Ton im Habituationsmodell
Tiefpassrauschen	oberhalb der oberen Bandkante	oberhalb der oberen Bandkante
Hochpassrauschen	kein Zwicker-Ton	unterhalb der unteren Bandkante
Bandsperrerauschen	oberhalb der oberen Bandkante	oberhalb der oberen Bandkante und unterhalb der unteren Bandkante
Hochpassrauschen mit weißem Rauschen	kein Zwicker-Ton	kein Zwicker-Ton
Tiefpassrauschen mit weißem Rauschen	kein Zwicker-Ton	kein Zwicker-Ton
weißes Rauschen	kein Zwicker-Ton	kein Zwicker-Ton
reiner Ton mit weißem Rauschen	unterhalb des reinen Tons	unterhalb des reinen Tons
reiner Ton	kein Zwicker-Ton	unterhalb des reinen Tons

Tab. 6.1: Qualitativer Vergleich des Habituationsmodells mit der Realität.

Realität.

6.1.5 Erweiterungen des Habituationsmodells

In Tab. 6.1 haben wir zwei wesentliche qualitative Unterschiede des Habituationsmodells zur Realität gesehen: Hochpassrauschen sowie ein reiner Ton erzeugen im Habituationsmodell einen Zwicker-Ton, in der Realität dagegen nicht. Das Habituationsmodell muss also durch zusätzliche Annahmen oder Mechanismen erweitert werden.

Betrachten wir zunächst nur Hochpassrauschen und die sich daraus ergebenden Folgerungen. Wäre das Gehör gegenüber Spiegelung von niedrigen nach hohen und hohen nach niedrigen Tonhöhen vollkommen symmetrisch, so müsste natürlich auch Hochpassrauschen einen Zwicker-Ton hervorrufen, denn Hochpassrauschen ist nur das Spiegelbild von Tiefpassrauschen. Das bedeutet, wir müssen nach Asymmetrien im Gehör suchen, um diese Asymmetrie in der Wahrnehmung erklären zu können. Eine deutliche Asymmetrie in der peripheren Verarbeitung haben wir in Kap. 1.5.1 und 3.1.1 erwähnt. Die cochleäre Erregung bei Stimulation durch einen reinen Ton nimmt auf der niederfrequenten Seite schneller zu als auf der höherfrequenten ab. Wir wissen jedoch bereits aus Kap. 6.1.4, dass der Zwicker-Ton durch diesen Effekt eher deutlicher wird. Da die erwähnte Asymmetrie und Folgen hiervon die einzigen sind, denen der Autor beim Studium des Gehörs begegnet ist, wird im Folgenden davon ausgegangen, dass die gesuchte Asymmetrie in der neuronalen Verschaltung zu suchen ist. Hier kommt im Habituationsmodell nur die laterale Inhibition in Frage. Nehmen wir also an, die laterale Inhibi-

tion wirke stärker von tiefen zu hohen Tonhöhen als umgekehrt. Dann würde Tiefpassrauschen nach wie vor einen Zwicker-Ton hervorrufen, nicht jedoch Hochpassrauschen. Denn Hochpassrauschen führt in diesem Fall dazu, dass nach dem Abschalten des Stimulus an der Bandkante die Inhibition von hohen nach niedrigeren Tonhöhen abgeschwächt wird. Diese ist aber bei der angenommenen Asymmetrie ohnehin nicht vorhanden. Die Kante in der Erregung wird also nicht verbogen, somit entsteht kein Zwicker-Ton (vergleiche auch Kap. 4.2.2 zu den Auswirkungen asymmetrischer lateraler Inhibition). Mit asymmetrischer lateraler Inhibition kann also die beobachtete Asymmetrie in den Zwicker-Ton Erzeugerschallen erklärt werden.

Wenden wir uns nun dem einzelnen reinen Ton zu, der keinen Zwicker-Ton erzeugt, und dem einzelnen reinen Ton mit Rauschen, der einen Zwicker-Ton erzeugt. Im Habituationsmodell trägt das zusätzliche Rauschen überhaupt nicht zur Entstehung des Zwicker-Tons bei, im Gegenteil (siehe Kap. 6.1.3). Dennoch ist an allen Zwicker-Ton Erzeugerschallen eine Form von Rauschen beteiligt². Rauschen scheint also eine zentralere Rolle bei der Erzeugung des Zwicker-Tons zu spielen als bisher angenommen. Im Habituationsmodell führte Rauschen nur zu einer Ermüdung der Neuronen. Dies gilt allerdings für einen reinen Ton ebenso, so dass dieser nach dem Habituationsmodell ebenfalls einen Zwicker-Ton erzeugt. Außerdem befände sich der Zwicker-Ton beim einzelnen reinen Ton und der oben eingeführten Asymmetrie rechts vom reinen Ton, in der Realität entsteht der Zwicker-Ton jedoch beim einzelnen reinen Ton mit Rauschen links von diesem. Diesen Widerspruch wird erst das im nächsten Kapitel eingeführte Adaptionsmodell lösen.

²Rauschen lässt sich auch durch schnelle Frequenzmodulation oder Klickfolgen ersetzen. Dies wird in Kap. 7.2 diskutiert.

6.2 Adaptionsmodell

Wir haben gesehen, dass das Habituationsmodell zwar zunächst einfach und einleuchtend aussieht, dass es aber einer genaueren Überprüfung anhand der vielfältigen untersuchten Schalle nicht standhält. Es kann auch nicht auf einfache Weise so erweitert werden, dass mehr Zwicker-Ton-Phänomene erklärt werden können. Deshalb wird in diesem Kapitel ein Ansatz vorgestellt, der nicht auf einer Ermüdung der Neuronen basiert. Vielmehr spielen aktive, rauschunterdrückende Mechanismen eine Rolle. Deshalb heißt das in diesem Kapitel vorgestellte neue Modell „Adaptionsmodell“. Zunächst sollen die wesentlichen Ideen des Adaptionsmodells vorgestellt werden. Es folgen eine genaue Beschreibung der Simulationen, Simulationsergebnisse mit verschiedenen Stimuli sowie eine Diskussion.

6.2.1 Beschreibung des Adaptionsmodells

In Kap. 6.1.5 haben wir gesehen, dass wir vier zentrale Phänomene berücksichtigen müssen: Tiefpassrauschen und reiner Ton mit Rauschen als Zwicker-Ton Erzeugerschalle, sowie Hochpassrauschen und reiner Ton als Schalle, die keinen Zwicker-Ton hervorrufen. Im Habituationsmodell führten natürlich sowohl Rauschen als auch ein reiner Ton zur Ermüdung der Neurone. Hierdurch wurde jeweils ein Zwicker-Ton ausgelöst. Da ein einzelner reiner Ton keinen Zwicker-Ton erzeugt, sollte es hier keine Nacheffekte geben. Ein reiner Ton mit Rauschen erzeugt allerdings einen Zwicker-Ton. Deshalb liegt die Vermutung nahe, dass Rauschen eine besondere Rolle spielt. Betrachten wir also folgendes Modell:

- Rauschen wird durch einen aktiven Mechanismus unterdrückt. Hierzu werden Neuronen der Outputschicht, die Rauschen als Eingabe bekommen, inhibiert.

Diese neue Inhibition ist dabei nicht die laterale Inhibition, sondern eine zusätzliche, von Rauschen detektierenden Neuronen ausgehende Inhibition. Ist der Stimulus in der Nachbarschaft der Bestfrequenz eines Neurons der Outputschicht also Rauschen, so wird dieses Neuron und seine nähere Umgebung inhibiert. Die Reichweite dieser Inhibition wird in Kürze noch genauer spezifiziert. Was sind nun die Konsequenzen dieser Annahme? Bei Tiefpassrauschen führt der postulierte rauschunterdrückende Mechanismus wie im Habituationsmodell zu einer geringeren Feuerrate der stimulierten Neuronen. Damit ein Zwicker-Ton erzeugt wurde, müsste die Ermüdung der Neuronen für die Dauer des Zwicker-Tons anhalten. Daran halten wir fest:

- Die Aktivität der rauschunterdrückenden Neuronen hält auch nach Abschalten des Stimulus noch vorübergehend an.

Wir postulieren also eine gewisse Trägheit des rauschunterdrückenden Mechanismus. Diese führt dazu, dass die Neuronen im Bereich des Tiefpassrauschens nach Abschalten des Stimulus mit geringerer als spontaner Feuerrate

feuern. Das erzeugt zusammen mit lateraler Inhibition einen Zwicker-Ton nach den selben Mechanismen wie im Habituationsmodell (Kap. 6.1.1).

Betrachten wir nun den einzelnen reinen Ton mit darunterliegendem weißen Rauschen. Im Bereich des reinen Tons befindet sich kein Rauschen, also ist der rauschunterdrückende Mechanismus nur oberhalb und *unterhalb* des reinen Tons aktiv. D.h. ein reiner Ton mit Rauschen wirkt wie Rauschen mit einer Frequenzlücke (im Bereich des reinen Tons), die nach den selben Mechanismen wie Tiefpassrauschen einen Zwicker-Ton erzeugt.

Ein einzelner reiner Ton hingegen erzeugt in der Realität keinen Zwicker-Ton. Dies ist auch im bisher beschriebenen Adaptionsmodell der Fall, denn ein reiner Ton enthält kein Rauschen, also wird der rauschunterdrückende Mechanismus nicht aktiv und somit gibt es auch keine lang anhaltenden Nacheffekte.

Damit der Zwicker-Ton beim Hochpassrauschen nicht und beim reinen Ton mit Rauschen nur unterhalb des reinen Tons auftritt, behalten wir auch die schon bei den Erweiterungen des Habituationsmodells diskutierte asymmetrische laterale Inhibition bei.

- Die laterale Inhibition wirkt stärker von tiefen zu hohen Tonhöhen als umgekehrt.

Wie bereits erwähnt werden Neuronen in der „Nachbarschaft“ eines Neurons, das Rauschen als Eingabe erhält, vom rauschunterdrückenden Mechanismus inhibiert. Wie groß sollte diese Nachbarschaft sein? Nehmen wir an, diese Nachbarschaft sei sehr klein. Das würde beim Tiefpassrauschen bedeuten, dass nur Neuronen mit Bestfrequenzen innerhalb des Tiefpassrauschens vom Rauschunterdrückungsmechanismus inhibiert werden. Der Zwicker-Ton entsteht dann an der oberen Bandkante des Tiefpassrauschens. Tatsächlich befindet sich die Tonhöhe jedoch etwa 1 Bark oberhalb der oberen Bandkante des Tiefpassrauschens. Deshalb führen wir noch folgende Modellannahme ein:

- Die Reichweite der Inhibition, die von einem rauschunterdrückenden Neuron ausgeht, ist etwa gleich der Reichweite der Erregung.

Die letzte Annahme führt dazu, dass die Tonhöhe des Zwicker-Tons vom Modell korrekt wiedergegeben wird.

6.2.2 Beschreibung der Simulationen des Adaptionsmodells

Im Adaptionsmodell benötigen wir einen Rauschunterdrückungsmechanismus. Hierzu sind speziell verschaltete Neuronen notwendig, insbesondere solche, die Rauschen detektieren. Die hierzu notwendigen Neuronen sowie deren Verschaltung werden hier vorgestellt.

Die rauschunterdrückenden Neuronen des Adaptionmodells

Wir haben im vorhergehenden Kapitel davon gesprochen, dass im Adaptionmodell Rauschen unterdrückt werden soll. Hierzu muss zunächst geklärt werden, nach welchen Kriterien zwischen „Rauschen“ und „nicht Rauschen“ unterschieden werden soll. Diese Kriterien sollten dann möglichst einfach neuronal implementiert werden können.

Rauschen erzeugt im Hörnerv eine gleichmäßige, d.h. über den Ort konstante, Erregung. D.h. Neuronen der Outputschicht, deren Umgebung im Hörnerv gleichmäßig erregt ist, sollten vom Rauschunterdrückungsmechanismus inhibiert werden, denn in der Umgebung ihrer Bestfrequenz ist der Stimulus Schmalbandrauschen. Variiert dagegen die Erregung in der Umgebung eines Neurons stark, so erhält dieses jedenfalls *kein* (weißes) Rauschen als Eingabe. Variation der Erregung bedeutet entweder eine im Ort ansteigende oder eine abfallende Erregung. Wir benötigen also Neurone, die auf ansteigende bzw. absteigende Flanken in der Erregung reagieren, nicht jedoch auf eine konstante Erregung. Wir nennen sie „Featuredetektoren“, denn sie erkennen gewisse Eigenschaften des Stimulus.

Wir stellen uns also eine tonotope Reihe von Featuredetektoren vor. Wird ein solcher Featuredetektor nun von Hörnerven höherer Bestfrequenz erregt und von Hörnerven niedrigerer Bestfrequenz gehemmt (wie das Neuron Mitte rechts in Abb. 6.12), so reagiert es (bei geeigneter Wahl der synaptischen Gewichte und der Schwelle) nur auf ansteigende Flanken im Stimulus. Denn eine konstante Erregung würde das Neuron gleich stark inhibieren wie erregen, somit würde das Neuron in diesem Fall nicht feuern. Eine abfallende Erregung würde das Neuron gar stärker inhibieren als erregen. Nur bei einer ansteigenden Flanke in der Erregung wird das Neuron stärker (von oberhalb) erregt als (von unterhalb) inhibiert und feuert. Ein Neuron, das nur auf eine abfallende Flanke in der Erregung des Hörnervs reagiert, ist spiegelsymmetrisch aufgebaut (das Neuron Mitte links in Abb. 6.12). Ein reiner Ton ist sicherlich kein Rauschen. Er führt jedoch zu einem fast konstanten Bereich in der Erregung des Hörnervs gerade in der Nachbarschaft des Neurons, dessen Bestfrequenz die Frequenz des reinen Tons ist (vgl. Abb. 6.10 bei 1500 Hz). Deshalb erwies es sich als zweckmäßig, einen weiteren Featuredetektor einzuführen, der auf eine Rechtskrümmung in der Erregung des Hörnervs reagiert. Dieser wird durch ein Neuron verwirklicht, das durch Hörnervenfasern in seiner Nähe erregt und durch weiter entfernte Hörnervenfasern inhibiert wird (das Neuron in der Mitte von Abb. 6.12). Dieser Featuredetektor reagiert dann nur, falls der Hörnerv bei der Bestfrequenz des Featuredetektors stärker feuert als oberhalb und unterhalb davon. Er reagiert also nur auf eine hinreichend starke Rechtskrümmung in der Erregung des Hörnervs, verursacht etwa durch einen reinen Ton.

Wir besitzen nun eine Reihe von Featuredetektoren, die anzeigen, wann an einer Stelle der Outputschicht *kein* Rauschen vorliegt. Die entsprechenden Neuronen der Outputschicht sollen aber nur inhibiert werden, falls sie Rauschen als Eingabe erhalten. Dies wird gelöst, indem das rauschunterdrückende Neuron (oben in Abb. 6.12) von allen Featuredetektoren stark inhibiert wird, so dass es bei Detektion eines Features die Outputschicht nicht inhi-

biert. Das bedeutet, die rauschunterdrückenden Neurone feuern nur, wenn an der entsprechenden Stelle in der Outputschicht kein Feature vorliegt und die entsprechende Stelle der Outputschicht erregt ist (exzitatorische Kopplung von der Outputschicht zum rauschunterdrückenden Neuron in Abb. 6.12). Dies bedeutet wiederum, dass die rauschunterdrückenden Neuronen nur feuern, falls Rauschen in dem Bereich der Outputschicht vorliegt, für den die rauschunterdrückenden Neuronen „zuständig“ sind³.

Gibt es in der Neurophysiologie Hinweise auf solch einen Rauschunterdrückungsmechanismus, wie er oben postuliert und beschrieben wurde? Jedenfalls gibt es im Nucleus cochlearis Neurone, die *nicht* auf Rauschen reagieren (Typ II-Neurone, Kap. 4.1), denn sie haben inhibitorische Seitenbänder wie der Featuredetektor für eine Rechtskrümmung. Darüber hinaus gibt es im Nucleus cochlearis Neuronen mit vielfältig angeordneten exzitatorischen und inhibitorischen Bereichen [Popper 1992]. Diese könnten als weitere Featuredetektoren dienen. Eine Implementation der in Abb. 6.12 angegebenen Verschaltung im Nucleus cochlearis wäre also denkbar, obwohl es nach Wissen des Autors noch keine genauen Vorstellungen der neuronalen Verschaltung des Nucleus cochlearis gibt (vgl. [Voigt, Young 1990, Popper 1992]). Jedoch ist es auch durchaus möglich, und der Autor hält dies für wahrscheinlich, dass sich der beschriebene Mechanismus zur Detektion der Features nicht im Nucleus cochlearis befindet, sondern in höheren Arealen, die auf den Nucleus cochlearis über die absteigende Hörbahn rückprojizieren. Relativ träge Rückprojektionen würden auch die langen Wahrnehmungszeiten des Zwicker-Tons von bis zu 6 Sekunden erklären.

Alle im Adaptionmodell simulierten Neuronen sind Spike-Response-Neuronen wie in Abb. 6.7 beschrieben. Die Neuronen der verschiedenen Schichten haben die Zeitkonstanten und die Schwellen aus Tab. 6.2.

Neuron	τ_s /ms	τ_{ref} /ms	γ_{ref} /ms	θ
Outputschicht	10	5	2	1
Featuredetektor steigend u. fallend	50	5	2	1
Rechtskrümmung	200	5	2	1
Rauschunterdrücker	500	5	2	2
inhibitorisches Interneuron	10	5	2	1

Tab. 6.2: Zeitkonstanten und Schwellen der Neuronen des Adaptionmodells.

Die Neuronen der Outputschicht und die inhibitorischen Interneuronen haben die selben (schnellen) Zeitkonstanten wie im Habituationsmodell, denn sie müssen den Stimulus in möglichst guter Zeitaufösung übertragen. Die Featuredetektoren haben längere Zeitkonstanten, da sie über längere Zeit mitteln müssen, um ein Feature zu „entdecken“, insbesondere für die schwer

³Man könnte noch an weitere Features denken, die bewirken, dass *kein* weißes Rauschen vorliegt, wie beispielsweise eine Linkskrümmung der Erregung im Ort oder eine Variation der Erregung in der Zeit. Diese spielen allerdings bei den meisten Zwicker-Ton Erzeugerschallen keine Rolle, deshalb wird hier der Einfachheit halber davon abgesehen.

zu detektierende Rechtskrümmung. Die Rauschunterdrücker wurden als wesentlicher Teil des Modells als recht träge angenommen, so dass der Zwicker-Ton entsprechend lange andauert.

Die neuronale Verschaltung im Adaptionmodell

In Abb. 6.12 ist schematisch der Aufbau des Nucleus cochlearis im Adaptionmodell abgebildet. Die laterale Inhibition wirkt stärker von tiefen zu hohen Frequenzen als umgekehrt. Die Neuronen oberhalb der Outputschicht dienen der Detektion von Rauschen. Ihre Verbindungen sind schematisch eingezeichnet. Meist steht jedoch eine eingezeichnete Synapse für viele Verbindungen der gleichen Art. Die genaue Struktur der Verbindungen soll in diesem Abschnitt beschrieben werden.

Es werden 300 Hörnervfasern, 300 Neuronen der Outputschicht, 300 inhibitorische Interneurone, 300 rauschunterdrückende Neuronen und 300 Featuredetektoren von jedem Typ simuliert. Jeder Hörnerv ist mit einem Neuron der Outputschicht verbunden (Synapsenstärke 4), dieses wiederum mit einem inhibitorischen Interneuron (Synapsenstärke 1) und mit einem rauschunterdrückenden Neuron (Synapsenstärke 0,00045).

Die Synapsenstärke der lateralen Inhibition von inhibitorischem Interneuron i (d.i. das inhibitorische Interneuron, das seine Eingabe von Neuron i der Outputschicht erhält) zum Neuron j der Outputschicht folgt der Formel $-0,03 e^{-[(i-j+12)/40]^2}$, Reichweite ist also 40 Neuronen, Verschiebung 12 Neuronen. Es gibt keine Selbstinhibition, d.h. keine synaptische Kopplung vom inhibitorischen Interneuron i zum Outputneuron i .

Jedes rauschunterdrückende Neuron i erhält eine inhibitorische Eingabe von Featuredetektor j jedes Typs nach der Formel $-0,0001 e^{-[(i-j)/10]^2}$. Die rauschunterdrückenden Neuronen „hören“ also auf Featuredetektoren in einer Nachbarschaft von etwa 10 Neuronen.

Das rauschunterdrückende Neuron i ist inhibitorisch mit dem Neuron j der Outputschicht verknüpft, falls $|i - j| \leq 30$ (Synapsenstärke $-0,05$).

Der Featuredetektor i ist mit der Hörnervenfaser j mit einer Synapsenstärke nach der Formel

$$I \cdot \left(c_1 e^{-[(j-i+d)/c]^2} + c_2 e^{-[(j-i-d)/c]^2} + c_3 e^{-[(j-i)/c]^2} \right)$$

verknüpft. Für die Konstanten gilt Tab. 6.3.⁴

⁴Beachte, dass die Hörnervfasern, die an den Featuredetektoren enden, sowohl exzitatorische als auch inhibitorische Synapsen besitzen. Biologisch korrekter wären weitere zwischengeschaltete Interneurone, auf die jedoch der Einfachheit halber in der Simulation verzichtet wurde. Weitere Interneurone würden an der Funktion des neuronalen Schaltkreises nichts ändern, bis auf eine geringe Verzögerung inhibitorischer Eingaben. Da der Rauschunterdrückungsmechanismus ohnehin träge ist, spielen diese zusätzlichen Verzögerungen keine Rolle.

Featuredetektor	I	c	d	c_1	c_2	c_3
steigend	0,02	10	5	-1	1	0
fallend	0,02	10	5	1	-1	0
Rechtskrümmung	0,02	10	20	-1	-1	5,8

Tab. 6.3: Konstanten für die Verschaltung der Featuredetektoren des Adaptionmodells.

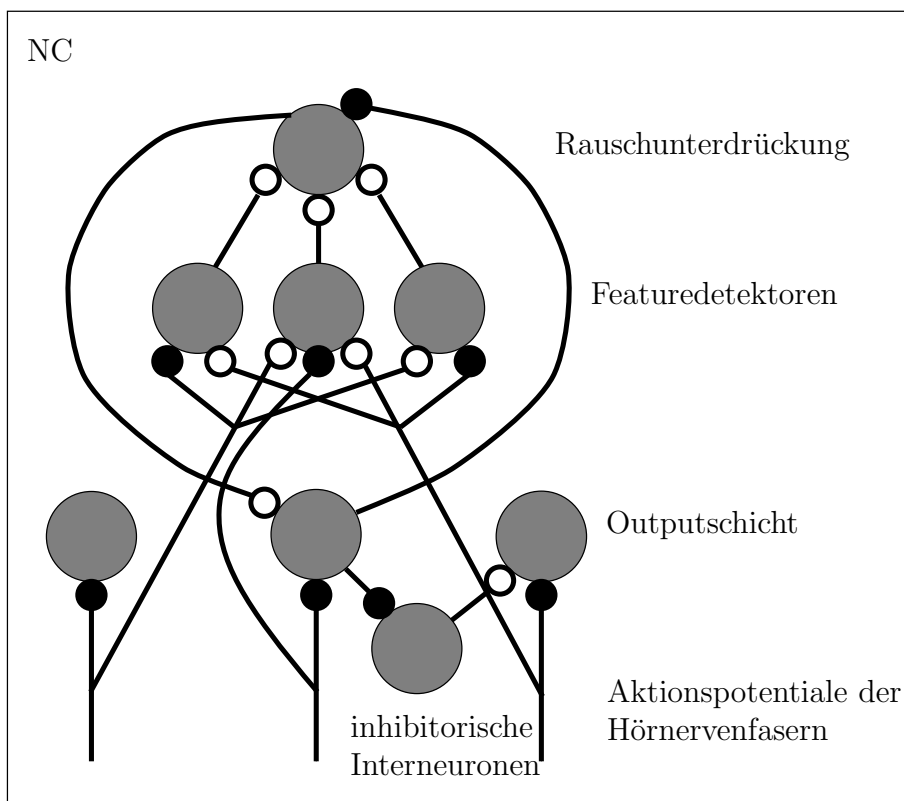
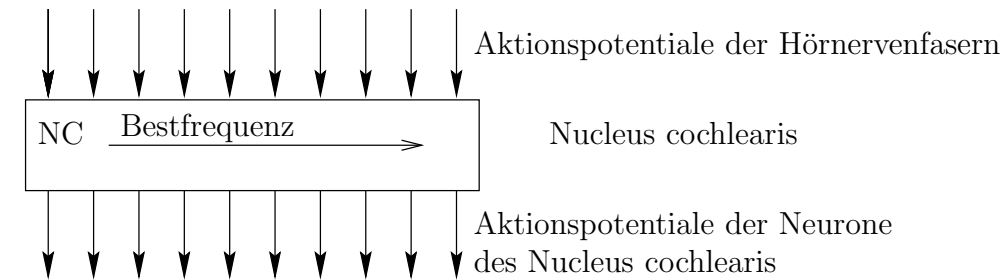


Abb. 6.12: Simulation des Nucleus cochlearis im Adaptionmodell. Signaturen wie in Abb. 6.2 und 6.6. Man beachte wiederum, dass oft nur jeweils eine von vielen Verbindungen dargestellt ist: Die inhibitorischen Interneurone und die rauschunterdrückenden Neuronen inhibieren Neuronen der Outputsicht aus einem ganzen Bereich und die Featuredetektoren bekommen Eingaben von vielen Hörnervenfasern (siehe Text).

6.2.3 Simulationsergebnisse im Adaptionsmodell

In diesem Abschnitt wird die Reaktion des Adaptionsmodells auf verschiedene Stimuli beschrieben. Die Simulationen sollen zeigen, dass der in Abb. 6.12 beschriebene Nucleus cochlearis eine relativ einfache neuronale Implementation des in Kap. 6.2.1 beschriebenen Adaptionsmodells ist. Zeitabhängige mittlere Feuerraten sind wieder in Anhang A farbcodiert abgebildet.

Tiefpassrauschen

Stimulus ist zunächst Tiefpassrauschen mit der Grenzfrequenz 1700 Hz. Die Aktivität des Hörnervs und der Outputschicht zeigt Abb. 6.13.

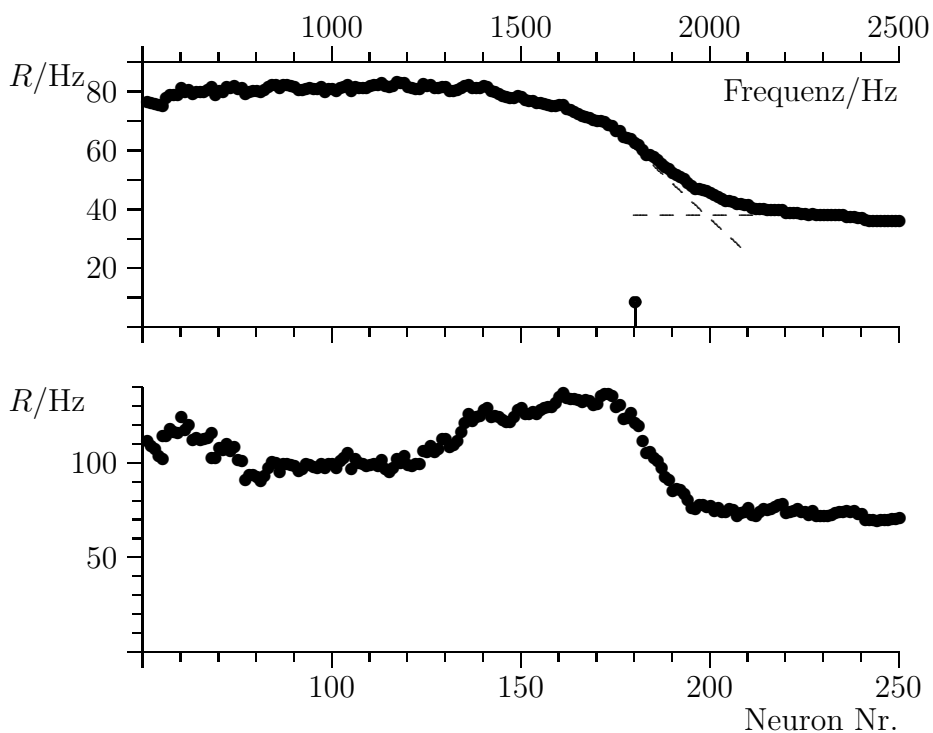


Abb. 6.13: Die Feuerraten R der Hörnervenfasern (oben) und der Neuronen der Outputschicht (unten) im Adaptionsmodell, gemittelt über die Dauer des Stimulus von 2 s. Stimulus war Tiefpassrauschen mit einer oberen Bandkante von 1700 Hz. Die Frequenzskala bezeichnet die Bestfrequenzen der Neuronen. Der Zwicker-Ton tritt im Adaptionsmodell an der mit \uparrow bezeichneten Stelle auf. Die gestrichelten Linien deuten die Berechnung der Tonhöhe des Zwicker-Tons im Mithörschwellenmodell von [Krump 1993] (Kap. 5.3) an, näheres siehe Kap. 6.2.4.

Abb. A.5 zeigt den zeitlichen Verlauf der Feuerraten der Neuronen der Outputschicht im Adaptionsmodell. Das Rauschen wird nach 2 s abgeschaltet. Man erkennt deutlich die gesteigerte Aktivität der Neuronen nach dem Abschalten des Stimulus bei Frequenzen von etwa 1800 Hz. Sie sind im Modell für die Wahrnehmung des Zwicker-Tons verantwortlich. Grüne Bereiche bedeuten wieder spontane Aktivität der Neurone, orange bis rote Bereiche Aktivität oberhalb der Hörschwelle und blaue Bereiche geringere als spon-

tane Aktivität. Blaue Bereiche kommen durch die anhaltende Inhibition der Rauschunterdrücker aus Abb. 6.12 zustande.

Die Erklärung für das Zustandekommen eines Zwicker-Tons ist folgende: Das Rauschen unterhalb 1700 Hz wird durch den Mechanismus in Abb. 6.12 unterdrückt. Dieser Mechanismus ist relativ langsam, er setzt erst nach etwa 0,5 s ein (siehe Abb. A.5). Diese Verzögerung entspricht der Zeitkonstante $\tau_s = 0,5$ s der Rauschunterdrücker aus Tab. 6.2. Die Rauschunterdrücker unterdrücken das Rauschen ziemlich stark, so dass die Kante bei 1700 Hz hervorgehoben wird (vgl. auch Abb. 6.13). Nach dem Abschalten ist der Rauschunterdrückungsmechanismus noch immer aktiv (blaue Bereiche in Abb. A.5). D.h. Neuronen unterhalb 1700 Hz feuern schwächer als im Zustand ohne aktive Rauschunterdrücker. Dadurch können sie Neuronen bei Frequenzen oberhalb 1700 Hz nicht mehr so stark inhibieren (über die inhibitorischen Interneuronen in Abb. 6.12). D.h. Neuronen bei Frequenzen oberhalb 1700 Hz feuern nun stärker als im stationären Zustand ohne Stimulus. Diese erhöhte Feuerrate ist durch das Gehör wahrnehmbar, da sie deutlich oberhalb der spontanen Feuerrate liegt⁵. Im Modell dauert der Zwicker-Ton etwa 0,5 s.

Die Aktivität der Rauschunterdrücker aus Abb. 6.12 zeigt Abb. A.6. Man erkennt das verspätete Einsetzen der Inhibition nach Beginn des Stimulus sowie das Anhalten der Inhibition nach dem Abschalten. Inhibition tritt nur im Bereich des Rauschens auf.

In Abb. A.5 und Abb. 6.13 kann man sehen, wie das Rauschen von den Rauschunterdrückern unterdrückt wird. Die obere Bandkante der Tiefpassgefilterten Eingabe wird dadurch hervorgehoben.

Reiner Ton mit Rauschen

Stimulus ist nun ein reiner Ton mit darunterliegendem weißen Rauschen. Abb. A.7 zeigt die zeitabhängige Feuerrate der Neuronen der Outputschicht. Der obere Teil von Abb. 6.14 zeigt die mittlere Aktivität des Hörnervs und der untere Teil die mittlere Aktivität der Outputschicht während der Dauer des Stimulus von 2 s.

In Abb. A.7 ist nach Abschalten des Stimulus ein deutlicher Zwicker-Ton zu sehen. Dieser kommt auf die selbe Weise wie beim Tiefpassrauschen zustande: Die Neuronen der Rauschunterdrückerschicht inhibieren alle Neuronen der Outputschicht, die Rauschen als Stimulus erhalten. D.h. Inhibition tritt unterhalb und oberhalb des reinen Tons auf, nicht jedoch an der Stelle des reinen Tons (siehe Abb. A.8). Dies wirkt auf die selbe Weise wie ein vorhergegangenes Tiefpassrauschen (siehe Kap. 6.2.3).

⁵Deutlich oberhalb heißt, dass die Feuerrate nicht mit spontaner Feuerrate verwechselt werden kann im Sinne von Kap. 4.3.1. Dies sieht man daran, dass im Normalzustand (Bereiche hoher Frequenzen und nach 2,5 s in Abb. A.5) die über 100 ms gemittelten Feuerraten niemals den grünen Bereich verlassen.

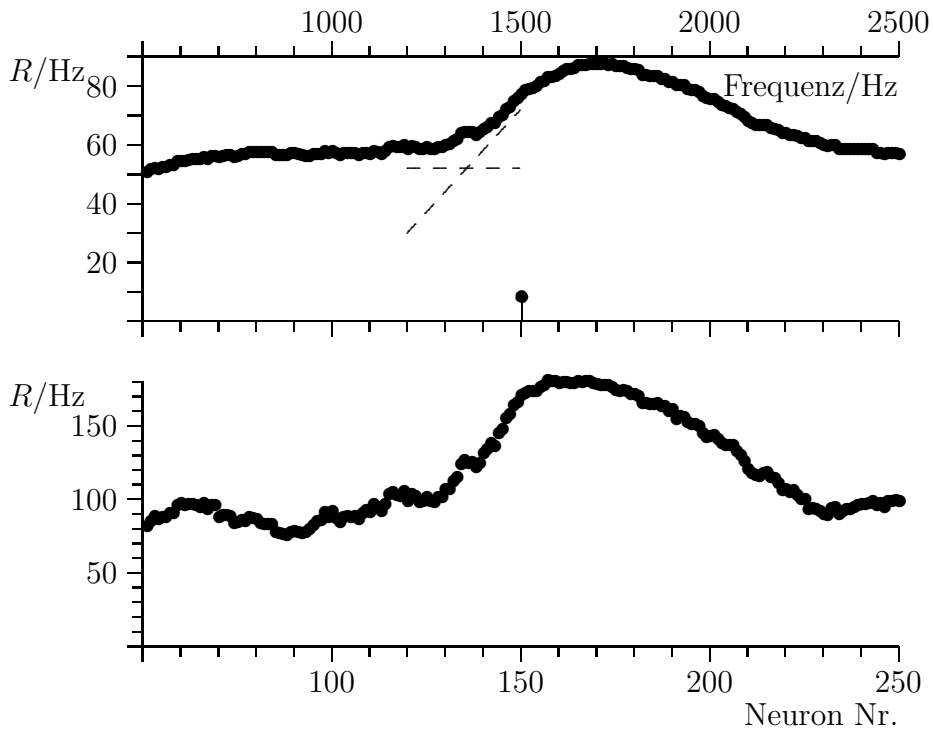


Abb. 6.14: Die Feuerraten R der Hörnervenfasern (oben) und der Neuronen der Outputschicht (unten) im Adaptionsmodell, gemittelt über die Dauer des Stimulus von 2s. Stimulus war ein Sinuston der Frequenz 1700 Hz mit darunterliegendem weißen Rauschen.

Reiner Ton

Ein einzelner reiner Ton erzeugt im Adaptionsmodell keinen Zwicker-Ton, da kein Rauschen im Stimulus vorkommt, somit auch keine anhaltende Inhibition.

Hochpassrauschen

Hochpassrauschen erzeugt im Adaptionsmodell keinen Zwicker-Ton, da die laterale Inhibition hauptsächlich von links nach rechts wirkt. Nach dem Abschalten des Stimulus tritt also keine genügend starke Verbiegung der Kante in der spontanen Aktivität auf (vgl. auch die Diskussion in Kap. 6.1.5 sowie Kap. 4.2.2).

Tiefpass- und Hochpassrauschen mit Rauschen

Tiefpass- und Hochpassrauschen mit zusätzlichem weißen Rauschen erzeugt auch im Adaptionsmodell keinen Zwicker-Ton. Der Grund ist dem im Habituationsmodell ähnlich: Das Tiefpassrauschen sowie das weiße Rauschen erzeugen Inhibition. Die Bereiche mit Inhibition sind also nur durch die schmale abfallende Bandkante getrennt, die ja Inhibition über die entsprechenden Featuredetektoren verhindert. Die schmale Lücke in der Inhibition reicht nicht aus, um einen Zwicker-Ton zu erzeugen. Beachte darüber hin-

aus, dass die von den Neuronen der Rauschunterdrückerschicht ausgehende Inhibition ausgedehnt ist (Modellparameter in Kap. 6.2.2).

6.2.4 Diskussion der Simulationsergebnisse im Adaptionmodell

Tonhöhe des Zwicker-Tons im Adaptionmodell

Die Tonhöhe des Zwicker-Tons im Adaptionmodell ergibt sich folgendermaßen: Rauschen erzeugt Inhibition, die auch nach dem Abschalten des Stimulus noch andauert. Diese erzeugt eine Kante in der Erregung, die nach den oben erläuterten Mechanismen verstärkt wird. Diese Kante befindet sich an der Stelle, an der die Inhibition durch die rauschunterdrückenden Neuronen endet. Damit der Zwicker-Ton nun oberhalb der oberen Bandkante auftritt, mussten wir annehmen, dass die durch ein Neuron der Rauschunterdrückerschicht hervorgerufene Inhibition die gleiche Reichweite wie die Erregung hat. Dann nämlich berechnet sich die Tonhöhe des Zwicker-Tons beim Tiefpassrauschen wie für das Habituationsmodell in Kap. 6.1.4 beschrieben. An die Stelle von durch Erregung hervorgerufener Ermüdung tritt nun durch die Rauschunterdrückerschicht hervorgerufene Inhibition.

Die Tonhöhen der im Adaptionmodell simulierten Zwickertöne und der nach Kap. 5.3 berechneten Tonhöhe (Schnittpunkt der gestichelten Linien in den Abb. 6.13 und 6.14) stimmen zwar nicht exakt überein, zeigen aber die gleiche Tendenz.

Beim einzelnen reinen Ton mit weißem Rauschen sind die Verhältnisse komplizierter: Das Rauschen oberhalb und unterhalb des reinen Tons stimuliert die rauschunterdrückenden Neuronen. Im Bereich des reinen Tons selbst werden die Rauschunterdrücker durch die Featuredetektoren inhibiert. Die rauschunterdrückenden Neuronen sind also nur im Bereich des Rauschens aktiv. Daher ergibt sich nach dem Abschalten dasselbe Bild wie nach einem Bandsperrrauschen: Die anhaltende Inhibition der Neuronen unterhalb des reinen Tons durch die Rauschunterdrückungsschicht führt zu einer Kante in der spontanen Aktivität, die durch laterale Inhibition zum Zwicker-Ton verstärkt wird. Dadurch tritt der Zwicker-Ton oberhalb vom Rauschen unterhalb des reinen Tons, jedoch noch unterhalb des reinen Tons auf.

Zwicker-Ton Erzeugerschalle im Adaptionmodell

Bei allen in Tab. 6.1 aufgeführten Schallen treten beim Adaptionmodell keine qualitativen Abweichungen zwischen Modell und Realität mehr auf.

In den in Kap. 5.2 beschriebenen Experimenten wurde oft die Bandbreite des Rauschens im Zwicker-Ton Erzeugerschall systematisch variiert. Die Qualität des Zwicker-Tons nahm dabei i.d.R. mit zunehmender Bandbreite des Rauschens zu. Dies war bis Bandbreiten von etwa 5 Bark der Fall, darüber blieb die Qualität des Zwicker-Tons konstant. Sowohl das Habituationsmodell als auch das Adaptionmodell beruhen auf einer Verringerung der lateralen Inhibition nach dem Abschalten des Zwicker-Ton Erzeugerschalls.

Je stärker diese Verringerung der lateralen Inhibition ist, desto intensiver ist der Zwicker-Ton in diesen Modellen. Man muss also nach den Experimenten von einer sehr langreichweitigen (Reichweite etwa 5 Bark) lateralen Inhibition ausgehen. Der Bereich, in dem Kanten in der Erregung „verbogen“ werden, ist jedoch etwa so breit wie die Reichweite der lateralen Inhibition (vgl. Kap. 4.2.2). Deshalb würden große Reichweiten der lateralen Inhibition nicht zu einem reinen Ton als Nacheffekt führen, sondern allenfalls zu Schmalbandrauschen. Die Reichweite der lateralen Inhibition beträgt im Adaptionsmodell etwa 1 Bark. Dies steht im Widerspruch zu oben beschriebenem Hinweis auf Reichweiten von etwa 5 Bark. Er ließe sich folgendermaßen lösen: Es besteht eine starke laterale Inhibition geringer Reichweite (etwa 1 Bark) sowie eine schwache laterale Inhibition großer Reichweite (etwa 5 Bark). Wegfall oder Abschwächung der starken lateralen Inhibition würde einen (in der Tonhöhe eng begrenzten) Zwicker-Ton entstehen lassen. Zunehmende Abschwächung der schwachen aber langreichweitigen lateralen Inhibition ließe den Zwicker-Ton intensiver werden. Diese Kombination von starker kurzreichweitiger und schwacher langreichweitiger lateraler Inhibition ist im oben beschriebenen Adaptionsmodell allerdings nicht implementiert, da der Erkenntnisgewinn gering und quantitative Aussagen ohnehin zweifelhaft wären.

In [Krump 1993, Zwicker 1990] wird auch die Lautheit der Zwicker-Ton Erzeugerschalle systematisch variiert, wobei sich herausstellt, dass es für jeden Zwicker-Ton Erzeugerschalle eine optimale Lautstärke gibt, bei der der Zwicker-Ton am deutlichsten zu hören ist. Optimale Lautstärkepegel liegen bei etwa 60 phon, Zwicker-Ton Erzeugerschalle sind also relativ leise. Nach lauten Schallen entsteht kein Zwicker-Ton. Dies kann mehrere Ursachen haben. Zunächst gibt es den Effekt der nichtlinearen Auffächerung (siehe Kap. 1.4), bei dem etwa bei einem Tiefpassrauschen die obere Flanke der Erregung mit zunehmender Lautstärke flacher wird. Die nichtlineare Auffächerung würde den Zwicker-Ton im Habituationsmodell bei großen Lautstärken verhindern, da die Kante in der spontanen Aktivität nach dem Abschalten zu flach wäre. Im Adaptionsmodell könnten die Featuredetektoren nicht mehr so gut zwischen Rauschen und abfallender Kante in der Erregung unterscheiden, so dass die von der Rauschunterdrückerschicht ausgehende Inhibition eher kontinuierlich denn abrupt verlief. Dies würde ebenfalls zu einer weniger scharfen Kante in der spontanen Aktivität nach dem Abschalten führen. Bei der unteren Kante eines reinen Tons ist die nichtlineare Auffächerung weniger stark, so dass beim einzelnen reinen Ton mit Rauschen die Lautstärke des reinen Tons die Wahrnehmbarkeit des Zwicker-Tons kaum beeinflussen dürfte. Ein weiterer stark nichtlinearer Effekt ist die Rekrutierung weiterer Hörnervenfaserhöherer Schwellintensität bei zunehmender Lautstärke des Schalls (Kap. 2.1.4). Diese projizieren möglicherweise auf einen anderen tonotopen Streifen des Nucleus cochlearis mit anderer innerer Verschaltung, bei der kein Zwicker-Ton erzeugt wird. Beide nichtlinearen Effekte werden hier nicht in die Simulationen miteinbezogen.

Kapitel 7

Diskussion und Ausblick

7.1 Rückblick

In Kap. 1 hatten wir grundlegende psychoakustische Begriffe behandelt, um das Phänomen des Zwicker-Tons hierin einordnen zu können. Der psychoakustische Begriff der Erregung stellte sich in Kap. 5 als zentral zur Bestimmung der Tonhöhe des Zwicker-Tons heraus. Um die physiologische Ursache der mit der Erregung zusammenhängenden psychoakustischen Phänomene wenigstens annähernd modellieren zu können, war ein relativ genaues Modell der Cochlea und des Hörnervs notwendig, das in Kap. 3 entwickelt wurde. In Kap. 6 wurden schließlich zwei Modelle vorgestellt, die beschreiben, wie der Zwicker-Ton möglicherweise entstehen könnte.

Als eine zentrale Ursache für die Entstehung des akustischen Nacheffekts war laterale Inhibition postuliert worden. Qualitative Auswirkungen der lateralen Inhibition auf den Erregungszustand der Neuronen in Abhängigkeit von der neuronalen Eingabe wurden in einem einfachen linearen Modell einer Reihe tonotop angeordneter Neurone untersucht (Kap. 4). Wir behandelten auch mögliche Auswirkungen der lateralen Inhibition auf die Qualität der Informationsverarbeitung im Gehör in einem sehr einfachen Modell mit spikenden Neuronen. Hierbei bekamen wir ein Gefühl für eine vernünftige Stärke für laterale Inhibition und „feed-forward“-Kopplungen.

Anhand der qualitativen Auswirkungen lateraler Inhibition konnten zwei Modelle entwickelt werden, die Nacheffekte zeigen. Das Habituationsmodell geht davon aus, dass feuernde Neuronen ermüden, wobei der ermüdete Zustand für die Dauer des Zwicker-Tons anhält. In einigen Fällen werden hierdurch Zwicker-Töne mit richtiger Tonhöhe produziert, in anderen widersprechen sich jedoch die experimentellen Beobachtungen und das Modell. Das Modell wurde daraufhin durch ein verfeinertes, jedoch auch komplizierteres, ersetzt. Das Adaptionsmodell geht davon aus, dass das auditorische System durch Rückkopplungen aktiv in die Informationsverarbeitung in tieferen Stufen eingreift. Genauer soll Rauschen unterdrückt werden, während andere „Features“ ungehindert passieren. Es wurde angenommen, dass die durch Rückkopplung verursachte Inhibition etwa so lange anhält wie der Zwicker-Ton wahrnehmbar ist. Die notwendig gewordenen Featuredetektoren konnten durch eine relativ einfache neuronale Schaltung modelliert werden. Sie sind in

der Realität, sollte das Adaptionmodell zutreffen, jedoch bestimmt weitaus komplizierter aufgebaut.

Das Adaptionmodell stimmt zumindest qualitativ gut mit den Beobachtungen überein. Hieraus quantitative Schlüsse über psychoakustische Größen ziehen zu wollen, erscheint etwas vermessen, da ja nur die erste Stufe der komplizierten auditorischen Informationsverarbeitung (vgl. Kap. 2) modelliert wurde, und diese auch nur in der einfachsten Weise als eine einzige Reihe von 300 tonotop angeordneten Neuronen. Die Dauer des Zwicker-Tons und seine Lautstärke hängen natürlich in beiden Modellen stark von den gewählten Modellparametern ab, beispielsweise von der Zeitkonstante des Rauschunterdrückungsmechanismus (hier etwa 0,5s). Wenn man allerdings einen derartigen Rauschunterdrückungsmechanismus postuliert und laterale Inhibition zur besseren Detektion von Variationen im Stimulus annimmt, dann ist der Zwicker-Ton als Nacheffekt zwar nicht unvermeidlich, jedoch zumindest plausibel. Jedenfalls lohnt es sich für das Gehör offenbar nicht, einen extra Mechanismus zur Unterdrückung dieser Nacheffekte einzubauen, da sie ohnehin schwach sind und bei natürlichen Stimuli praktisch nicht auftreten.

7.2 Weitere psychoakustische Experimente

Das Adaptionmodell kann als Grundlage für weitere psychoakustische Messungen dienen, insbesondere, um weitere Zwicker-Ton Erzeugerschalle zu finden. Beispielsweise erkennt man am Mechanismus der Entstehung des Zwicker-Tons beim einzelnen reinen Ton mit weißem Rauschen in Kap. 6.2.3, dass das Rauschen bei höheren Tonhöhen als der des reinen Tons keine Rolle spielt. D.h. auch ein in Abb. 7.1 skizzierter Schall müsste einen Zwicker-Ton erzeugen.

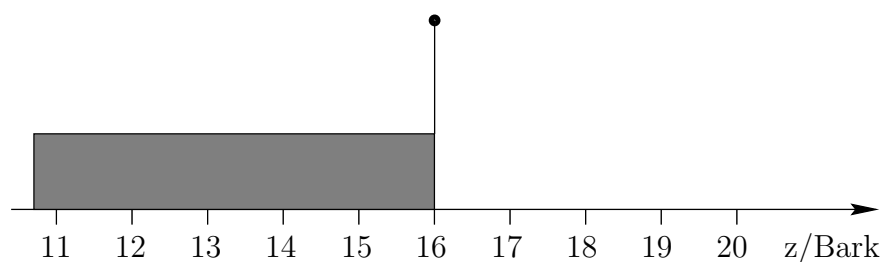


Abb. 7.1: Tiefpassrauschen mit individuellem reinen Ton als Zwicker-Ton Erzeugerschall im Adaptionmodell. Die Höhe des schraffierten Bereichs gibt die Intensitätsdichte des Rauschens bei der jeweiligen Tonhöhe an. Das Symbol \uparrow bezeichnet einen einzelnen reinen Ton.

Die Featuredetektoren reagieren nur auf die Steilheit einer ansteigenden oder abfallenden Flanke. Auf Rauschen mit steil ansteigenden oder abfallenden Flanken müssten sie auch reagieren. Deshalb sollte der einzelne reine Ton auch durch geeignetes steil ansteigendes und abfallendes Schmalbandrauschen ersetzt werden können (Abb. 7.2).

Nach [Krump 1993] kann Rauschen auch durch geeignete Pulsfolgen oder frequenzmodulierte Sinustöne ersetzt werden. Diese sind zwar kein Rau-

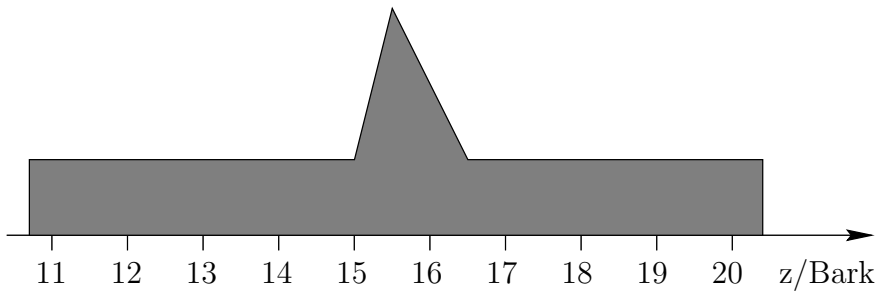


Abb. 7.2: Rauschen mit steil ansteigendem und abfallendem Schmalbandrauschen als Zwicker-Ton Erzeugerschall im Adaptionmodell. Die Höhe des schraffierten Bereichs gibt die Intensitätsdichte des Rauschens bei der jeweiligen Tonheit an.

schen, werden aber von den einfachen Featuredetektoren des Adaptionmodells als solches registriert, da die Featuredetektoren relativ lange Zeitkonstanten (Tab. 6.2) besitzen. Aus der Frequenzmodulationsfrequenz bzw. der Frequenz der Klickfolgen, bei der gerade noch ein Zwicker-Ton entsteht, sollte man die Zeitkonstanten der Featuredetektoren abschätzen können.

Wir haben gesehen, dass im auditorischen System vermutlich ein kompliziertes Wechselspiel von verschiedenen Detektions- und Adaptionsmechanismen stattfindet. Dies wirkt sich nicht in offensichtlicher Weise auf die akustische Wahrnehmung aus, denn sonst wäre unsere akustische Welt voll von Täuschungen. Vielmehr müssen die Adaptionsprozesse möglichst dezent im Hintergrund arbeiten, so dass die Wahrnehmung im Wesentlichen konstant bleibt, sich aber auf unaufdringliche Weise „verbessert“, wie dies im visuellen System, beispielsweise bei der Adaption der Helligkeit über viele Größenordnungen hinweg, auch geschieht. Trotz aller vermutlich eingebauten Maßnahmen zur Unterdrückung störender Effekte, die durch verschiedene Adaptionsmechanismen hervorgerufen werden könnten, gibt es hierdurch hervorgerufene akustische Täuschungen wie den Zwicker-Ton. Daher lohnt sich wohl die Suche nach weiteren, dem Zwicker-Ton verwandten psychoakustischen Phänomenen, d.h. solchen, bei denen charakteristische Zeiten größer als eine Sekunde eine Rolle spielen. Der Zwicker-Ton Erzeugerschall muss etwa 5 Sekunden präsent sein, damit ein optimaler Zwicker-Ton hervorgerufen wird und der Zwicker-Ton kann bis zu 6 Sekunden lang anhalten. Dies ist ein Hinweis auf länger andauernde Adaptionsmechanismen im auditorischen System. Fände man weitere Phänomene auf dieser Zeitskala, so könnte man die zugrundeliegenden Adaptionsmechanismen noch besser charakterisieren. Die in dieser Arbeit vorgestellten Überlegungen könnten hierfür hilfreich sein.

7.3 Weitere neurophysiologische Experimente

Zunächst wäre neurophysiologisch festzustellen, auf welcher Verarbeitungsstufe in der Hörbahn der Zwicker-Ton entsteht. Der Nucleus cochlearis ist hierfür sehr wahrscheinlich (Kap. 6). Dabei ist zu beachten, dass der Zwicker-

Ton vielleicht nicht in jeder Reihe tonotoper Neuronen entsteht, sondern nur in einer oder wenigen, eben denjenigen, die eine Verschaltung ähnlich dem Adaptionsmodell oder eine sonstige spezialisierte Verschaltung aufweisen, die einen Zwicker-Ton hervorruft.

Ist diese Reihe tonotoper Neuronen gefunden, so wäre es zur Diskrimination verschiedener Modelle wichtig, die Antworteigenschaften der Neuronen in dieser Reihe zu untersuchen, insbesondere im Hinblick auf laterale Inhibition, Reaktion auf Rauschen (Adaptionsmodell) und Ermüdungsverhalten (Habituationsmodell). Eine ausgeprägte laterale Inhibition sollte festgestellt werden. Falls ein Modell ähnlich dem Habituationsmodell richtig ist, sollte eine ausgeprägte Ermüdung der Neuronen festgestellt werden. Falls ein Modell wie das Adaptionsmodell zutrifft, sollten die Neuronen zunächst auf Rauschen reagieren, nach Zeiten im Sekundenbereich jedoch sollte ihre Aktivität abnehmen. Nach Abschalten des Rauschens sollten sie ihre unbeeinflusste spontane Feuerrate ebenfalls erst nach Zeiten im Sekundenbereich (Dauer des Zwicker-Tons) wieder erreichen.

Anhang A

Farbtafeln

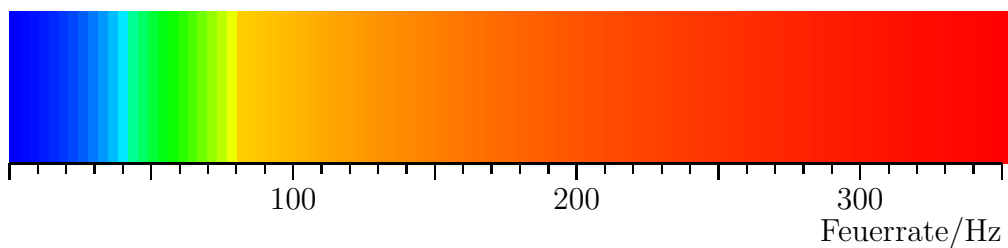


Abb. A.1: Farbcodierung der Feuerraten in den Abbildungen.

A.1 Habituationsmodell

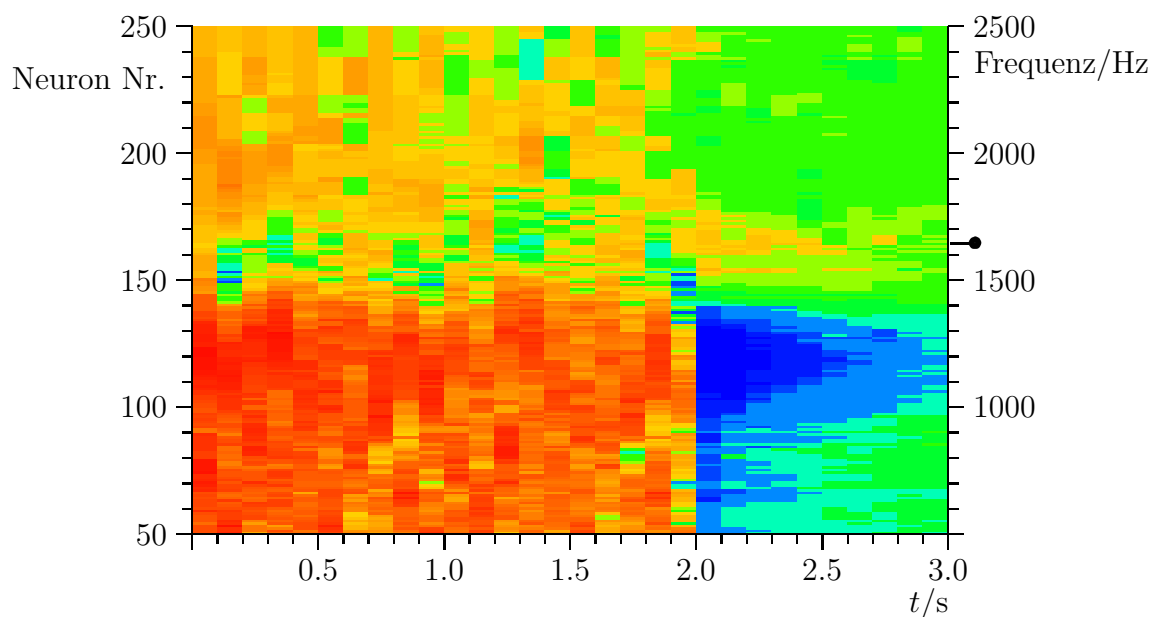


Abb. A.2: Zeitlicher Verlauf der Feuerraten der Neuronen der Outputsicht aus Abb. 6.6 im Habituationsmodell. Nach rechts ist die Zeit, nach oben die Nummer der Neuronen angetragen. Die über 100 ms gemittelte Feuerrate ist farbcodiert aufgetragen. Für die Farbcodes siehe Abb. A.1. Stimulus ist Tiefpassrauschen mit einer oberen Bandkante von 1300 Hz. Die Frequenzskala bezeichnet die Bestfrequenzen der Neuronen. Der Zwicker-Ton tritt im Habituationsmodell an der mit \bullet bezeichneten Stelle auf.

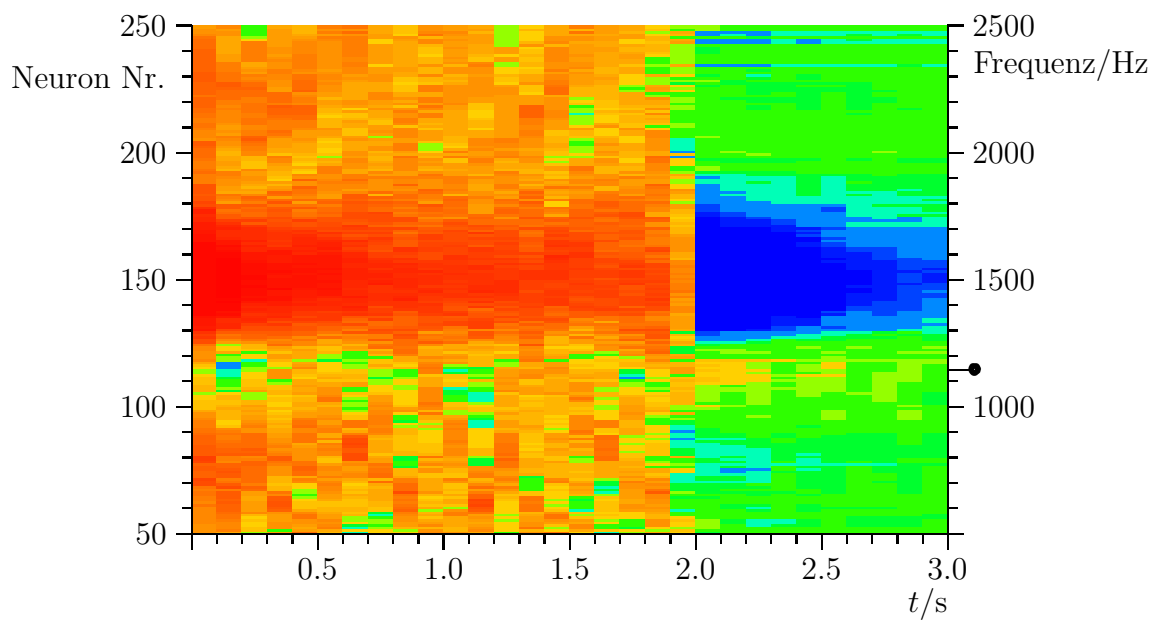


Abb. A.3: Siehe Abb. A.2, nur ist der Stimulus nun ein reiner Ton der Frequenz 1500 Hz mit darunterliegendem weißen Rauschen.

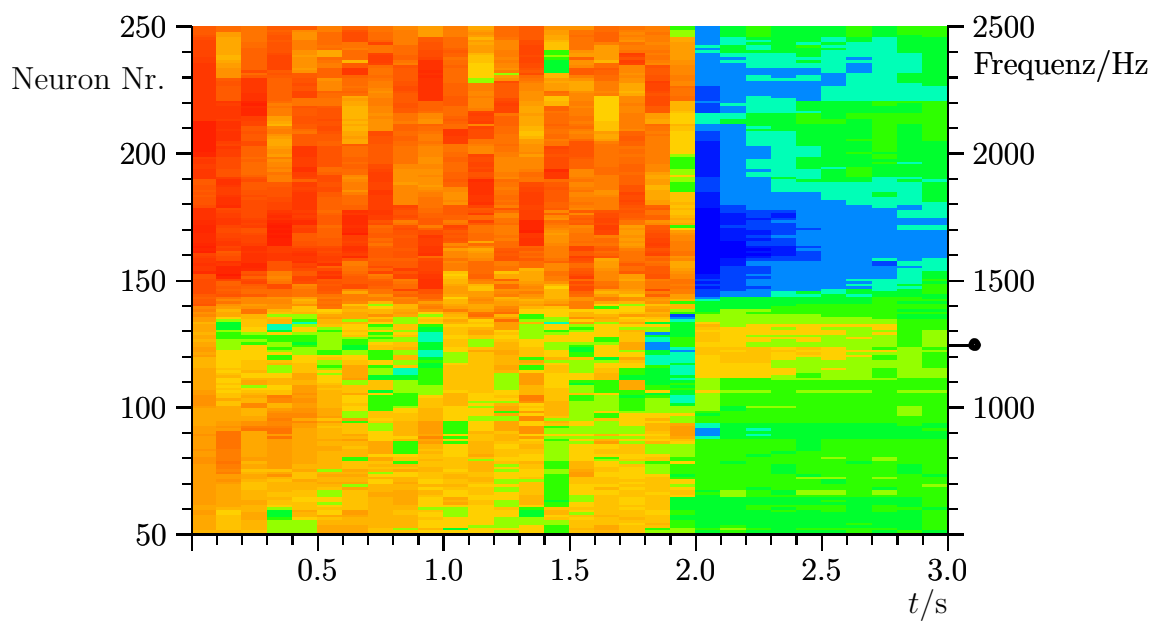


Abb. A.4: Siehe Abb. A.2, nur ist der Stimulus nun ein Hochpass mit Grenzfrequenz 1500 Hz.

A.2 Adaptionmodell

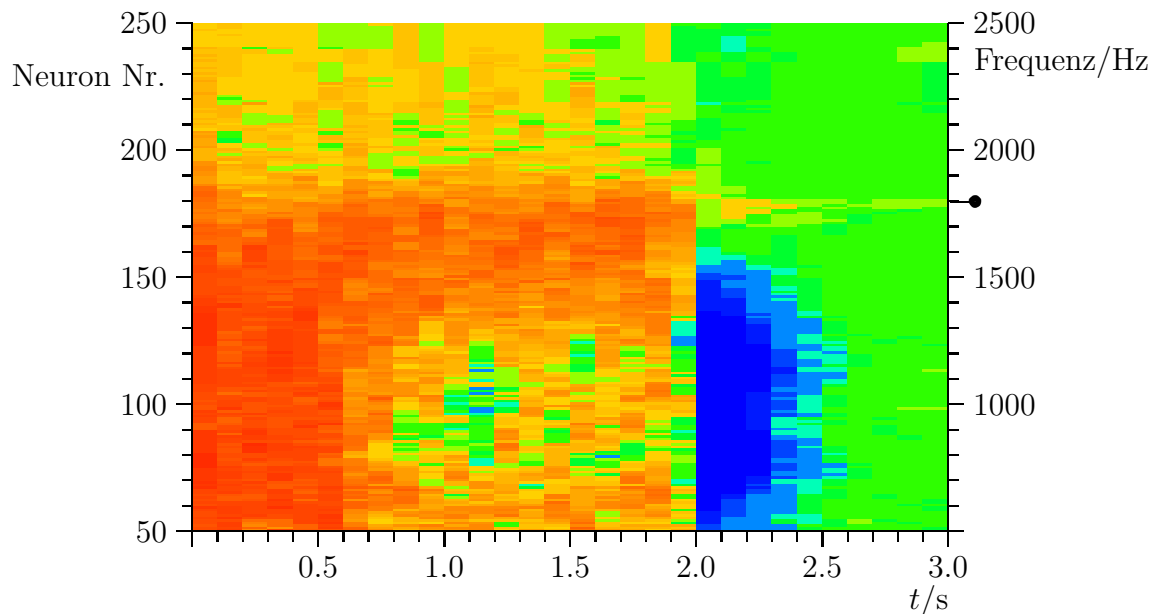


Abb. A.5: Zeitlicher Verlauf der Feuerraten der Neuronen der Outputschicht aus Abb. 6.12 im Adaptionmodell. Nach rechts ist die Zeit, nach oben die Nummer der Neuronen angetragen. Die über 100 ms gemittelte Feuerrate ist farbcodiert aufgetragen. Zu den Farbcodes siehe Abb. A.1. Stimulus ist Tiefpassrauschen mit einer oberen Bandkante von 1700 Hz. Die Frequenzskala bezeichnet die Bestfrequenzen der Neuronen. Der Zwicker-Ton tritt im Adaptionmodell an der mit \bullet bezeichneten Stelle auf.

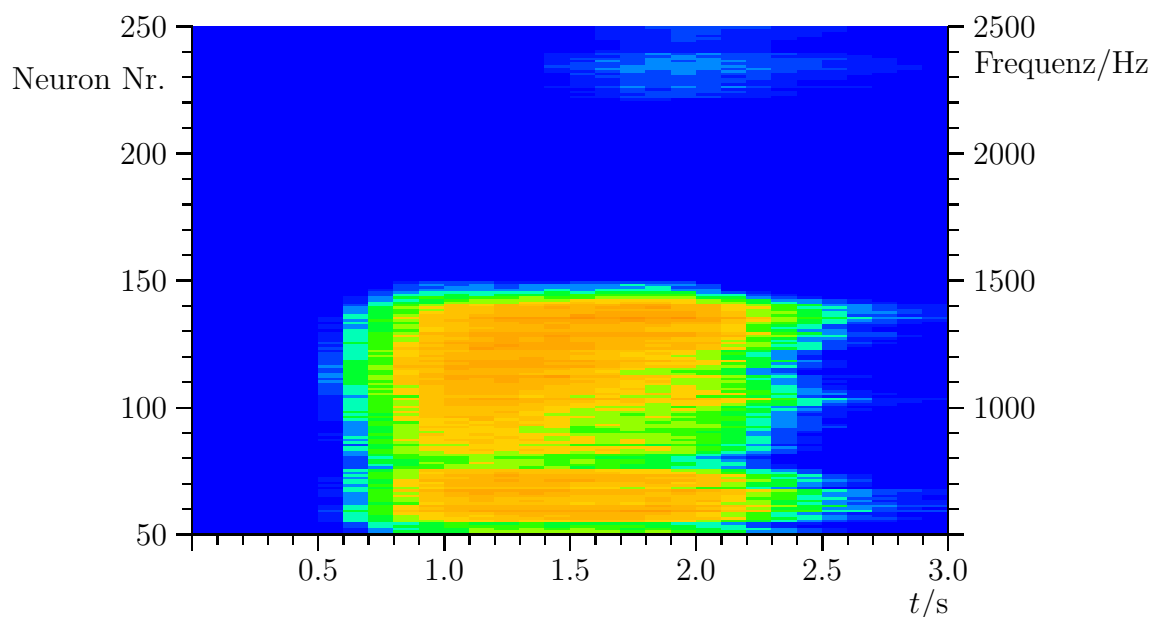


Abb. A.6: Zeitlicher Verlauf der Feuerraten der Neuronen der Rauschunterdrückerschicht im Adaptionmodell. Stimulus und Signaturen wie in Abb. A.5.

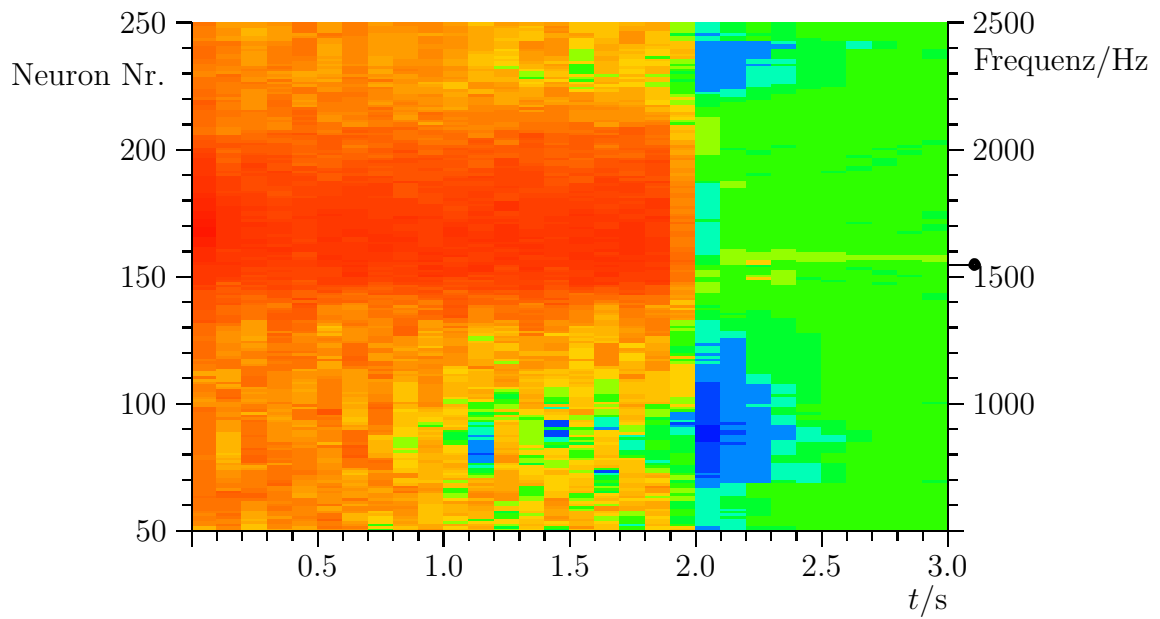


Abb. A.7: Zeitlicher Verlauf der Feuerraten der Neuronen der Outputschicht aus Abb. 6.12 im Adaptionmodell. Signaturen wie in Abb. A.5. Stimulus ist ein reiner Ton der Frequenz 1700 Hz mit darunterliegendem weißen Rauschen.

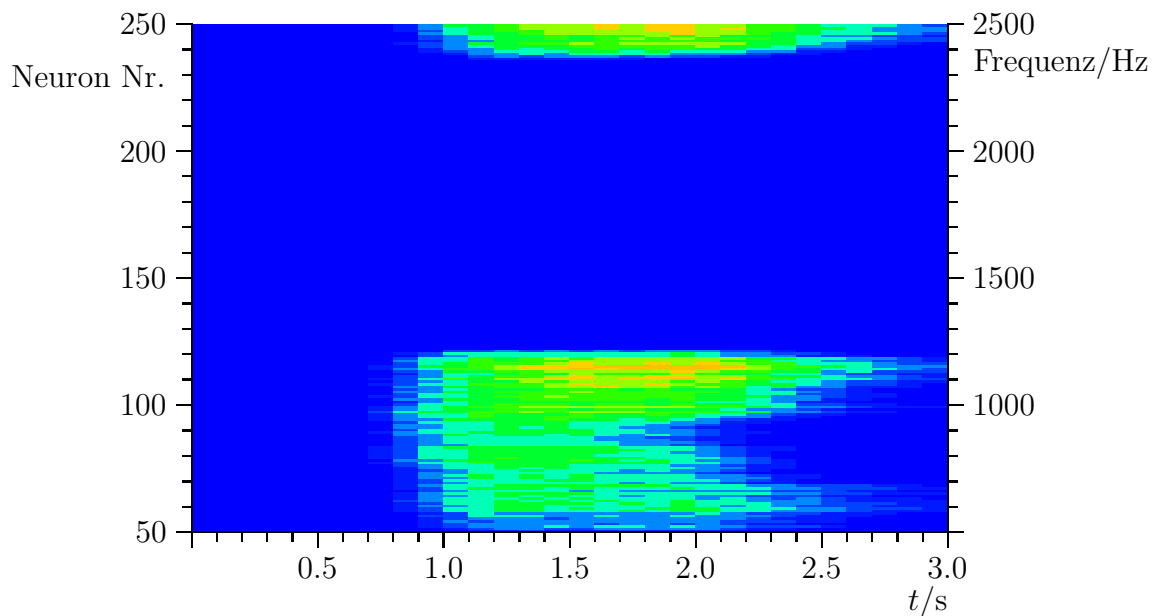


Abb. A.8: Zeitlicher Verlauf der Feuerraten der Neuronen der Rauschunterdrückerschicht im Adaptionmodell. Stimulus und Signaturen wie in Abb. A.7.

Literaturverzeichnis

- [Born I 1995] Günter Born, Referenzhandbuch Dateiformate: Grafik, Text, Datenbanken, Tabellenkalkulation, Addison-Wesley 1995
- [Born II 1995] Günter Born, Noch mehr Dateiformate: neue Dateiformate für Grafik, Text, Tabellenkalkulation und Sound, Addison-Wesley 1995
- [Buser 1992] Pierre Buser, Michel Imbert, Vision, MIT Press 1992
- [C++ 1996] Accredited Standards Committee, Operating under the procedures of the American National Standards Institute (ANSI), Working Paper for Draft Proposed International Standard for Information Systems — Programming Language C++, 1996
- [Campenhausen 1981] Christoph von Campenhausen, Die Sinne des Menschen, Band I: Einführung in die Psychophysik der Wahrnehmung, Thieme 1981
- [Campenhausen II 1981] Christoph von Campenhausen, Die Sinne des Menschen, Band II: Anleitung zu Beobachtungen und Experimenten, Thieme 1981
- [Chassell 1996] Robert J. Chassell, Richard M. Stallman, Texinfo, The GNU Documentation Format, Edition 2.23, for Texinfo Version 3, Free Software Foundation 1996
- [Cooke 1993] Martin Cooke, Modelling Auditory Processing and Organisation, Cambridge University Press 1993
- [dtv-Atlas 1987] dtv-Atlas zur Physik, Band 1, Mechanik, Akustik, Thermodynamik, Optik, Deutscher Taschenbuch Verlag 1987
- [dtv Brockhaus 1986] dtv Brockhaus Lexikon, Deutscher Taschenbuch Verlag 1986
- [Domany 1994] E. Domany, J. Leo van Hemmen, K. Schulten (Hrsg.), Models of Neural Networks II, Temporal Aspects of Coding and Information Processing in Biological Systems, Springer 1994
- [Donnelly 1995] Charles Donnelly, Richard M. Stallman, Bison, The YACC-compatible Parser Generator, Bison Version 1.25, Free Software Foundation 1995

- [Duden 1996] Duden, Rechtschreibung der deutschen Sprache, Dudenverlag 1996
- [Fastl 1986] Hugo Fastl, Auditory after-images produced by complex tones with a spectral gap. In: Proc. 12. ICA Toronto (1986), Vol. I, B 2–5
- [Fastl 1989] Hugo Fastl, Zum Zwicker-Ton bei Linienspektren mit spektralen Lücken. *Acustica* 67, 177–186
- [Frick 1987] Hans Frick, Helmut Leonhardt, Dietrich Starck, Spezielle Anatomie II, Eingeweide — Nervensystem — Systematik der Muskeln und Leitungsbahnen, Thieme 1987
- [Gerken 1996] George M. Gerken, Central Tinnitus and lateral inhibition: an auditory brainstem model, *Hearing Research* 97 (1996) 75–83
- [Gerstner 1993] Wulfram Gerstner, Kodierung und Signalübertragung in Neuronalen Systemen: Assoziative Netzwerke mit stochastisch feuernenden Neuronen, Reihe Physik Bd. 15, Verlag Harri Deutsch 1993
- [Glass 1996] Graham Glass, Brett L. Schuchert, The STL <Primer>, Prentice Hall 1996
- [Grimsehl 1987] Ernst Grimsehl, Lehrbuch der Physik, Band 1: Mechanik, Teubner 1987
- [Haft 1998] M. Haft, J. L. van Hemmen, Theory and implementation of infomax filters for the retina, *Network: Comput. Neural Syst.* 9 (1998) 39–71
- [Hauske 1994] Gert Hauske, Systemtheorie der visuellen Wahrnehmung, Teubner 1994
- [Hawkins 1996] Harold L. Hawkins, Teresa A. McMullen, Arthur N. Popper, Richard R. Fay (Hrsg.), Auditory Computation, Springer 1996
- [Hoke 1996] Ellen S. Hoke, Manfred Hoke, Bernhard Ross, Neurophysiological Correlate of the Auditory After-Image ('Zwicker Tone'), *Audiol Neurootol* 1996;1:161–174
- [Holmes, Lester 1981] Mark H. Holmes, Lester A. Rubenfeld (Editors), Mathematical Modeling of the Hearing Process: Proceedings of the NSF-CBMS Regional Conference Held in Troy, NY, Springer-Verlag 1981
- [Kahle 1979] Werner Kahle, Helmut Leonhardt, Werner Platzer, Taschenatlas der Anatomie für Studium und Praxis, Thieme 1979
- [Kistler 1996] Werner Kistler, Zeitcodierung im Cerebellum, Technische Universität München, Diplomarbeit 1996
- [Kistler 1997] Werner M. Kistler, Wulfram Gerstner, J. Leo van Hemmen, Reduction of the Hodgkin-Huxley Equations to a Single-Variable Threshold Model, *Neural Computation* 9 (1997), 1015–1045

- [Knuth 1993] Donald Ervin Knuth, *The TeXbook*, Addison-Wesley 1993
- [Keidel 1995] Wolf D. Keidel, *Physiologie des Gehörs: Akustische Informationsverarbeitung, Einführung für Ärzte, Biologen, Psychologen und Nachrichtentechniker*, Thieme 1975
- [Kopka I 1994] Helmut Kopka, *LaTeX Band 1: Einführung*, Addison-Wesley 1994
- [Kopka II 1995] Helmut Kopka, *LaTeX Band 2: Ergänzungen — mit einer Einführung in METAFONT*, Addison-Wesley 1995
- [Kopka III 1997] Helmut Kopka, *LaTeX Band 3: Erweiterungen*, Addison-Wesley 1997
- [Krump 1993] Gerhard Krump, *Beschreibung des akustischen Nachtönes mit Hilfe von Mithörschwellenmustern*, München, Techn. Univ. München, Diss., 1993
- [Kuchling 1988] Horst Kuchling, *Taschenbuch der Physik*, Verlag Harri Deutsch 1988
- [Loukides 1997] Mike Loukides, Andy Oram, *Programming with GNU Software*, O'Reilly 1997
- [Møller 1983] Aage, R. Møller, *Auditory Physiology*, Academic Press 1983
- [Maple 1] Bruce W. Char u. a., *First Leaves, A Tutorial Introduction to Maple V*, Springer 1992
- [Maple 2] Bruce W. Char u. a., *Maple V Language Reference Manual*, Springer 1991
- [Maple 3] Bruce W. Char u. a., *Maple V Library Reference Manual*, Springer 1991
- [Meddis 1986] R. Meddis, *Simulation of auditory-neural transduction: further studies*, *J. Acoust. Soc. Am.* 83 1056–1063
- [Moore 1982] Brian C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press 1982
- [Pickles 1982] James O. Pickles, *An Introduction to the Physiology of Hearing*, Academic Press 1982
- [Popper 1992] Arthur N. Popper, Richard R. Fay, *The Mammalian Auditory Pathway: Neurophysiology*, Springer-Verlag 1992
- [Profos 1994] Paul Profos, Tilo Pfeifer (Hrsg.), *Handbuch der industriellen Messtechnik*, Oldenbourg 1994
- [Shepherd 1993] Gordon M. Shepherd, *Neurobiologie*, Springer-Verlag 1993
- [Sottek 1993] Roland Sottek, *Modelle zur Signalverarbeitung im menschlichen Gehör*, Techn. Hochschule Aachen, Diss., 1993

- [Spektrum 4/1997] Spektrum der Wissenschaft, April 1997, Neues zur akustischen Funktion der Ohrmuschel
- [Stallman GCC 1995] Richard M. Stallman, Using and Porting GNU CC, for version 2.7.2.1, Free Software Foundation 1995
- [Stallman GDB 1995] Richard M. Stallman, Cygnus Support, Debugging with GDB, Edition 4.12, for GDB Version 4.16, Free Software Foundation 1995
- [Stallman XEmacs 1994] Richard M. Stallman, Rashmi Goyal, Getting Started with XEmacs, 1994
- [Stallman XEmacs 1995] Richard M. Stallman, Lucid Inc., Ben Wing, XEmacs User's Manual, 1995
- [Stroustrup 1991] Bjarne Stroustrup, The C++ Programming Language, Addison-Wesley 1991
- [Tomlinson, Biebel, Langner 1998] R. W. Ward Tomlinson, Ulrich Biebel, G. Langner, The Hunt for the Phantom's Song: Representation of Zwicker Tones in Neurons of Alert Auditory Cortex, Poster 345 der Göttinger Neurobiologentagung 1998
- [Unbehauen 1997] Rolf Unbehauen, Systemtheorie 1, Oldenbourg 1997
- [Voigt, Young 1980] Herbert F. Voigt, Eric D. Young, Evidence of inhibitory interactions between neurons in dorsal cochlear nucleus, J. Neurophysiol. 44 76–96
- [Voigt, Young 1982] Herbert F. Voigt, Eric D. Young, Response properties of type I and type III units in the dorsal cochlear nucleus, Hear. Res. 6 153–169
- [Voigt, Young 1990] Herbert F. Voigt, Eric D. Young, Cross-correlation analysis of inhibitory interactions in the dorsal cochlear nucleus, J. Neurophysiol. 164 1590–1610
- [Yost 1994] William A. Yost, Fundamentals of Hearing, An Introduction, Third Edition, Academic Press Inc. 1994
- [Young, Brownell 1976] Eric D. Young, W. Brownell, Response to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats, J. Neurophysiol. 39 282–300
- [Zenner 1994] Hans-Peter Zenner, Hören: Physiologie, Biochemie, Zell- und Neurobiologie, Thieme 1994
- [Zwicker 1982] Eberhard Zwicker, Psychoakustik, Springer-Verlag 1982
- [Zwicker 1990] Eberhard Zwicker, H. Fastl, Psychoacoustics: Facts and Models, Springer-Verlag 1990

Dank

Mein Dank an dieser Stelle gilt allen, die zum Gelingen dieser Arbeit beigetragen haben.

Insbesondere Prof. J. Leo van Hemmen für seine guten Vorlesungen, Gespräche und Ideen, ohne die diese Arbeit nicht möglich gewesen wäre.

Ganz herzlich bedanken möchte ich mich bei Richard Kempter und Oliver Wenisch für das Korrekturlesen, bei Richard auch besonders dafür, dass er mir bei einer ersten Orientierung im auditorischen System geholfen und auch sonst die gesamte Arbeit immer hilfreich begleitet hat.

Danke an das gesamte „Team“ Julian Eggert, Armin Bartsch, Richard Kempter, Werner Kistler, Ulrich Hillenbrand, Oliver Wenisch, Claudia Bergmann, Wolfgang Stürzl, Berthold Bäuml und Christian Leibold für viele nicht nur wissenschaftliche Gespräche und die jederzeitige Hilfsbereitschaft.

Große Hilfen waren Georg Klump, Institut für Zoologie der TU München, der mich anfangs mit Büchern und Informationen unterstützte, und Manfred Kössl, Zoologisches Institut der LMU München, der versuchen wird, den Zwicker-Ton neurophysiologisch zu messen. An dieser Stelle auch vielen Dank an alle Teilnehmer und Vortragenden im Seminar der Forschergruppe „Hörobjekte“.

Während des Studiums unterstützten mich die Bayerische Begabtenförderung und die Studienstiftung des Deutschen Volkes, erstere mehr finanziell, letztere mehr ideell, deshalb gilt mein Dank auch diesen beiden Institutionen.

Letztlich noch ein herzliches Dankeschön an meine Eltern, die mir immer mit Rat und Tat zur Seite standen.