

Physik Department
Technische Universität München
Theoretische Physik

Beschreibung, Messung, Vorhersage
und Charakterisierung
des Antwortverhaltens von Neuronen
im auditorischen System

Dissertation
von
Jan-Moritz P. Franosch

Februar 2007

Physik Department
Technische Universität München
Theoretische Physik

Beschreibung, Messung, Vorhersage und Charakterisierung des Antwortverhaltens von Neuronen im auditorischen System

Jan-Moritz P. Fransch

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Matthias Rief

Prüfer der Dissertation: 1. Univ.-Prof. Dr. J. Leo van Hemmen

2. Univ.-Prof. Dr. Roland Netz

Die Dissertation wurde am 15.2.2007 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 20.9.2007 angenommen.

Inhaltsverzeichnis

1	Einleitung	21
1.1	Allgemeine Fragestellung	21
1.2	Ansätze zur Messung des Antwortverhaltens	29
1.3	Mögliche Experimente	34
1.4	Beschreibender und modellbasierter Ansatz	41
1.5	Neuronenmodelle	42
1.6	Vorarbeiten	48
1.6.1	Reverse Correlation	48
1.6.2	Reverse Correlation mit Pieptönen	49
1.6.3	Wiener-Entwicklung	50
1.6.4	Ripple-Spektren	56
1.6.5	Methode von Theunissen	60
2	Psychoakustik	63
2.1	Psychophysik	63
2.2	Lautstärke	65
2.2.1	Schalldruckpegel und Schallintensitätspegel	65
2.2.2	Lautstärkepegel	66
2.2.3	Lautheit	67
2.3	Tonhöhe	67
2.3.1	Verhältnistonhöhe	68
2.3.2	Tonheit	71
2.4	Verdeckung	73
2.5	Funktionsschemata	74
2.5.1	Anregung und Erregung	74
2.5.2	Schwellenfunktionsschema für langsame Schalländerungen	76
2.5.3	Funktionsschema der Lautheit	77
3	Modelle für das periphere auditorische System	79
3.1	Cochlea	79

3.1.1	Erregung	80
3.1.2	Lineare Schwinger	80
3.1.3	Lineare Filter vierter Ordnung	82
3.2	Innere Haarzellen	90
3.2.1	„State-Partition“-Modell	90
3.2.2	Meddis-Modell	92
3.3	Gesamtmodell der auditorischen Vorverarbeitung	96
4	Optimales rezeptives Feld in einer Dimension	99
4.1	Optimale lineare Näherung	99
4.2	Bayes'scher Schätzer	101
4.3	Regularisierungsmethode	103
4.4	Minimale Varianz	106
4.5	Optimale lineare Schätzung	107
4.6	Bayes'scher Schätzer unter erweiterten Verteilungsannahmen	110
4.7	Optimale Vorhersage	113
4.8	Zusammenfassung	115
5	Rezeptives Feld in mehreren Dimensionen	121
5.1	Optimale lineare Schätzung	121
5.2	Regularisierungsmethode	124
5.3	Bayes'scher Schätzer	125
5.4	Zusammenfassung	127
5.5	Spikende Neurone	127
6	Lineare rezeptive Felder in einer Dimension	131
6.1	Das Verfahren	131
6.2	Tiefpassrauschen als Eingabe	131
6.2.1	Einfluss der oberen Grenzfrequenz	132
6.2.2	Einfluss des Regularisierungsparameters	134
6.2.3	Einfluss der Ordnung der Regularisierung	137
6.2.4	Einfluss der Messdauer	139
6.2.5	Einfluss der Form des rezeptiven Feldes	141
6.2.6	Einfluss des Rauschens in der Messung	148
6.3	Hochpassrauschen als Eingabe	159
6.3.1	Einfluss der unteren Grenzfrequenz	159
6.3.2	Einfluss des Rauschens in der Messung	161
6.4	Tiefpassrauschen plus Hochpassrauschen als Eingabe	164
6.5	Cochleogramm mit Tönen als Eingabe	167
6.6	Spikende Neuronen	171

7	Regularisierung in mehreren Dimensionen	175
8	Statistische Lerntheorie	177
8.1	Einführung	177
8.2	Beispiele	179
8.2.1	Lernen von „groß“ und „klein“	179
8.2.2	Statistische numerische Integration	180
8.3	Lernende Maschinen	181
8.3.1	Klassifikatoren	181
8.3.2	Modellfreies Lernen	183
8.4	Abschätzung des Generalisierungsfehlers	184
8.5	Das Perzeptron	186
8.6	Optimale Hyperebenen	188
8.7	Support-Vektor Maschinen	191
8.8	Numerische Behandlung	194
8.9	Beispiele	196
8.10	Effiziente Algorithmen für Support-Vektor-Maschinen	204
8.10.1	Optimierungsprobleme	204
8.10.2	Das Optimierungsproblem für Support-Vektor-Maschinen	206
8.10.3	Ein effizienter Algorithmus zur Lösung des Optimierungsproblems für Support-Vektor-Maschinen	208
8.11	Kolmogorov-Komplexität	213
9	Rezeptive Felder mit Support-Vektor-Maschinen	219
9.1	Eindimensionale rezeptive Felder	219
9.2	Mehrdimensionale rezeptive Felder	219
10	Diskussion und Ausblick	223
11	Zusammenfassung	227
12	Summary	229
A	Kovarianz und Faltung	231
A.1	Kovarianz	231
A.2	Faltung	233
B	Fouriertransformation	237
B.1	Fouriertransformation und Rücktransformation	238
B.2	Fouriertransformation und Faltung	240
B.3	Fouriertransformation und Kovarianz	240

B.4	Fouriertransformation der Ableitung	240
B.5	Fouriertransformation und lineare Systeme	241
B.6	Fouriertransformation und Delta-Funktion	241
	Literaturverzeichnis	245

Abbildungsverzeichnis

1.1	Reaktion eines Neurons auf einen Stimulus	22
1.2	Aufbau zur Messung des Antwortverhaltens eines Neurons . .	24
1.3	Sinusoide Stimuli mit <i>gleichem</i> Peri-Stimulus-Zeit-Histogramm, jedoch <i>unterschiedlichem</i> Antwortverhalten des Neurons	27
1.4	Zwei unterschiedliche Neuronen mit dem gleichen Stimulus und gleichem Peri-Stimulus-Zeit-Histogramm, jedoch unterschiedlichem Antwortverhalten	28
1.5	Antwortverhalten des Neurons als linearer Filter	30
1.6	Korrelation des Cochleogramms mit Gauß'schem weißem Rauschen	33
1.7	Minimierung der mittleren quadratischen Abweichung zwischen Messung und Schätzung	35
1.8	Neuronales Netz zur Vorhersage von Aktionspotentialen	36
1.9	Postsynaptisches Potential und Refraktärpotential	45
2.1	Breite von 1 mel in Hz	69
2.2	Frequenz als Funktion der Verhältnistönhöhe	70
2.3	Bestimmung der Frequenzgruppenbreite	72
2.4	Erregung durch einen reinen Ton	75
3.1	Impulsantwort der linearen cochleären Filter vierter Ordnung .	83
3.2	Vergleich cochleärer Filter	84
3.3	Vergleich cochleärer Filter als Pegel	84
3.4	Vergleich der cochleären Erregung	85
3.5	Vergleich der cochleären Erregung auf einer Barkskala	85
3.6	Cochleogramm von Tiefpassrauschen	88
3.7	Cochleogramm von einem Reinton	89
3.8	„State-Partition“-Modell	91
3.9	Meddis-Modell	93
3.10	Neurotransmitterkonzentration	95
3.11	Gesamtmodell der auditorischen Vorverarbeitung	96

4.1	Zusammenhang zwischen den Methoden zur Rekonstruktion rezeptiver Felder	119
6.1	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit verschiedenen oberen Grenzfrequenzen	133
6.2	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit verschiedenen Regularisierungsparametern	135
6.3	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit verschiedenen Regularisierungsparametern	136
6.4	Rekonstruierte rezeptive Felder Regularisierungsmethode mit verschiedenen Regularisierungsordnungen	137
6.5	Rekonstruierte rezeptive Felder Regularisierungsmethode mit verschiedenen Regularisierungsordnungen und Regularisierungsparametern	138
6.6	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode bei verschiedenen Messdauern	139
6.7	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode bei verschiedenen Messdauern und verschiedenen Regularisierungsparametern	140
6.8	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode	142
6.9	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode mit verschiedenen Regularisierungsparametern	143
6.10	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode mit verschiedenen Regularisierungsparametern	144
6.11	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode mit verschiedenen Regularisierungsparametern	145
6.12	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode mit verschiedenen Regularisierungsparametern	146
6.13	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode mit verschiedenen Regularisierungsparametern	147
6.14	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen	149
6.15	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen mit verschiedenen Regularisierungsparametern	150
6.16	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen	151
6.17	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen	152

6.18	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen mit verschiedenen Regularisierungsordnungen	153
6.19	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen mit verschiedenen Regularisierungsordnungen	154
6.20	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen mit verschiedenen Regularisierungsordnungen	155
6.21	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen mit verschiedenen Messdauern	156
6.22	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen mit verschiedenen Regularisierungsordnungen	157
6.23	Rekonstruierte sinusoidale rezeptive Felder nach der Regularisierungsmethode unter Einfluss von Rauschen mit verschiedenen Regularisierungsordnungen	158
6.24	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Hochpassrauschen als Eingabe	160
6.25	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Hochpassrauschen als Eingabe und Rauschen in der Messung	161
6.26	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Hochpassrauschen als Eingabe und Rauschen in der Messung	162
6.27	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Hochpassrauschen als Eingabe	163
6.28	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Tiefpass- plus Hochpassrauschen als Eingabe	165
6.29	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Tiefpass- plus Hochpassrauschen als Eingabe	166
6.30	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Tönen als Eingabe	168
6.31	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Tönen als Eingabe und Rauschen in der Messung	169
6.32	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit Tönen als Eingabe und Rauschen in der Messung	170
6.33	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit spikenden Neuronen	172

6.34	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit spikenden Neuronen	173
6.35	Rekonstruierte rezeptive Felder nach der Regularisierungsmethode mit spikenden Neuronen	174
7.1	Rekonstruierte multidimensionale rezeptive Felder nach der Regularisierungsmethode mit spikenden Neuronen	176
8.1	Optimale Hyperebene und Supportvektoren	190
8.2	Support-Vektor-Maschine zur Beschreibung eines rektifizierenden Neurons	198
8.3	Support-Vektor-Maschine zur Beschreibung eines rektifizierenden Neurons mit Fehler	199
8.4	Support-Vektor-Maschine zur Beschreibung eines rektifizierenden Neurons mit Fehler	199
8.5	Support-Vektor-Maschine zur Beschreibung eines rektifizierenden Neurons mit Fehler	199
8.6	Support-Vektor-Maschine zur Beschreibung eines rektifizierenden Neurons mit Fehler	199
8.7	Support-Vektor-Maschine zur Rekonstruktion eines rezeptiven Feldes	200
8.8	Support-Vektor-Maschine zur Rekonstruktion eines rezeptiven Feldes	201
8.9	Support-Vektor-Maschine zur Rekonstruktion eines rezeptiven Feldes	202
9.1	Rekonstruierte rezeptive Felder nach der Support-Vektor-Methode	220
9.2	Rekonstruierte rezeptive Felder nach der Support-Vektor-Methode	221

Tabellenverzeichnis

2.1	Umrechnung zwischen Tonheit, Tonunterschiedsschritten, Tonhöhe und Zahl der Haarzellen	71
3.1	Parameter einer Nervenfasernach dem Meddis-Modell	94
4.1	Rekonstruktion rezeptiver Felder nach unterschiedlichen Me- thoden	116
5.1	Rekonstruktion rezeptiver Felder in mehreren Dimensionen nach unterschiedlichen Methoden	128

Die Welt ist alles, was der Fall ist.

Ludwig Wittgenstein¹

Vorwort

Im Gegensatz zum visuellen System (Buser und Imbert 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 und Fastl 1999, 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. Im Rahmen dieser Arbeit sollten nun Methoden entwickelt und getestet werden, die zur Beantwortung derartiger Fragen geeignet sein könnten. Was ist unter einer „Messung“ an einem Neuron zu verstehen? Man legt einen Stimulus an und misst die Antwort (d. h. das Membranpotential oder Aktionspotentiale) des Neurons oder eine aus der Antwort abgeleitete Größe, beispielsweise die Feuerrate. Nun wäre die Messung wertlos, würde sie nicht eine gewisse *Vorhersage* des Antwortverhaltens des Neurons erlauben. Wenigstens muss man fordern, dass eine Messung reproduzierbar ist, d. h. eine Messung mit dem gleichen Stimulus sollte die gleiche Antwort des Neurons ergeben. Dies ist jedoch bei den meisten Neuronen nicht der Fall, da die Antwort von Stichprobe zu Stichprobe trotz gleichem Stimulus variiert. Man ist also i. d. R. auf aus mehreren Antworten abgeleitete statistische Größen beschränkt, beispielsweise das Poststimulus-Zeit-Histogramm (PSTH). Das

¹Wittgenstein (1966).

PSTH gibt an, mit welcher Wahrscheinlichkeitsdichte das Neuron zur Zeit t nach Einsetzen des Stimulus feuert. Nun ist es relativ uninteressant, das PSTH eines Neurons für nur einen einzigen Stimulus aufzunehmen. Man wird daher das PSTH für eine Reihe von Stimuli (z. B. alle reinen Töne der Dauer 100 ms mit Frequenzen zwischen 10 und 10000 Hz in Abständen von 100 Hz) ermitteln. „Vorhersage der neuronalen Antwort“ bedeutet nun, dass sich aus den gemessenen PSTHs das PSTH für einen noch nicht gemessenen Stimulus (etwa 1050 Hz) wenigstens ungefähr vorhersagen lässt.

Allgemein stellt sich bei der Vorhersage der Antwort eines Neurons folgendes Problem: Was ist das PSTH des Neurons bezüglich eines Stimulus, der noch nicht gemessen wurde? Gelingt die Beantwortung dieser Frage, so verfügen wir offenbar über eine relativ gute Beschreibung des Antwortverhaltens dieses Neurons. Diese Beschreibung wird eine Reihe von Parametern enthalten, die bei der Messung ermittelt wurden. Was ist nun eine möglichst allgemeine Beschreibung des Antwortverhaltens eines Neurons? Für Zellen im CGL ist es weitgehend ausreichend, eine *lineare* Beschreibung der Neuronen zu wählen, d. h. das PSTH ergibt sich durch Faltung des „rezeptiven Felds“ des Neurons mit dem Stimulus (Cai et al. 1997). Es muss also nur das rezeptive Feld des Neurons gemessen werden, um das PSTH des Neurons auf unbekannte Stimuli vorherzusagen. Im Allgemeinen wird eine lineare Beschreibung von Neuronen jedoch nicht ausreichen. Beispielsweise hat jedes Neuron eine maximale Feuerrate, d. h. die Wahrscheinlichkeitsdichten im PSTH können nicht beliebig anwachsen. Lineare Modelle sagen jedoch ein beliebiges Anwachsen der Feuerwahrscheinlichkeit mit Erhöhung der Intensität des Stimulus voraus. Zur Beschreibung des Antwortverhaltens der Neuronen sollten also auch nichtlineare Ansätze gewählt werden. Nun ist es jedoch selbst mit einem linearen Ansatz i. d. R. möglich, das *gemessene* Verhalten *exakt* zu reproduzieren, da das rezeptive Feld (im Visuellen eine Funktion mit zwei Raumkoordinaten und einer Zeitkoordinate) beliebig viele Parameter enthält, nämlich einen Funktionswert zu jeder Zeit an jedem Raumpunkt. Eine solche „Messung“ des rezeptiven Feldes hätte jedoch keinerlei *Vorhersagekraft*, d. h. bei dem Versuch einer Vorhersage würden weitgehend zufällige Werte resultieren². Man muss sich also überlegen, wie man eine möglichst exakte Beschreibung des Antwortverhaltens eines Neurons erhält, die es auch gestattet, Vorhersagen zu machen und nicht nur die Messungen reproduziert. Auch eine Abschätzung der Qualität der Vorhersage („Wie groß ist der

²Dies ist das unter „overfitting“ bekannte Phänomen: Eine Funktion mit sehr vielen Parametern wird an eine Messung mit zu wenig Messpunkten angepasst. Die resultierende Funktion wird zwar durch die Messpunkte laufen, nicht aber der „wirklichen“ Kurve entsprechen. Cai et al. (1997) umgeht dieses Problem durch Anwendung der „reverse correlation“-Technik, die jedoch von vorneherein auf lineare Beschreibungen beschränkt ist.

durchschnittliche Vorhersagefehler?“) wäre wünschenswert³.

Verfügt man über eine allgemeine Beschreibung des Antwortverhaltens eines Neurons, so will man die darin vorkommenden freien Parameter messen. Hierzu sind geeignete Stimuli zu generieren, so dass die Vorhersagequalität möglichst gut bzw. die Messzeiten möglichst klein werden.

Eine Beschreibung des Neurons in dem Sinn, dass die Antwort des Neurons auf unbekannte Stimuli vorhergesagt werden kann, reicht jedoch u.U. noch nicht aus. Es sollen ja Fragen beantwortet werden wie „Was macht dieses Neuron?“, „Was detektiert dieses Neuron?“ oder sogar „Wozu dient dieses Neuron?“. Hierzu müssen Konzepte entwickelt werden, um das beschriebene Antwortverhalten des Neurons mit einfachen Worten zu charakterisieren. Beispielsweise kann der optimale Stimulus ermittelt werden, d. h. der Stimulus, auf den das Neuron mit der höchsten durchschnittlichen Feuerrate reagiert oder die Reaktion des Neurons auf spezielle Stimuli wie im auditorischen z. B. reine Töne verschiedener Frequenz und Rauschen verschiedener Bandbreiten und Lagen⁴. Auch kann man versuchen, die Parameter in der Beschreibung des Neurons zu interpretieren⁵.

Ziel der vorliegenden Arbeit ist es, aufgrund einer geeigneten Messung mit möglichst kurzer Messzeit die Antwort (d. h. das PSTH bzw. Membranpotential) eines Neurons im auditorischen System⁶ auf noch nicht gemessene Stimuli möglichst gut vorherzusagen. Darauf aufbauend kann dann versucht werden, das gemessene Antwortverhalten des Neurons anschaulicher zu charakterisieren als dies durch die reine mathematische Beschreibung möglich

³Eine Möglichkeit wäre, die Anzahl der Parameter zu beschränken. Dies würde jedoch die Messgenauigkeit stark herabsetzen, zumindest in dem einfachen Fall, dass die betrachteten Neuronen tatsächlich weitgehend linear sind. Andere Möglichkeiten mit Abschätzung des Vorhersagefehlers finden sich z. B. in Vapnik (1998).

⁴Letztere können auch direkt im Experiment gemessen werden. Um dagegen den optimalen Stimulus zu ermitteln, benötigte man jedoch sehr lange Messzeiten.

⁵Dies fällt bei einem linearen Modell mit rezeptiven Feldern leicht, bei anderen Modellen nur, wenn sie so angelegt sind, dass sie auch leicht interpretierbare Parameter enthalten.

⁶Eine Übertragung auf andere sensorische Systeme sollte möglich sein. Mein primäres Interesse gilt jedoch dem auditorischen, da ich hier schon Erfahrung aufgrund der Diplomarbeit Franosch (1998), Franosch und van Hemmen (1999), Fastl et al. (2001), Franosch et al. (2003a) habe und über das auditorische System noch nicht so viel bekannt ist wie über das visuelle. Außerdem müssen vergleichbare Techniken aus dem visuellen zuerst geschickt auf das auditorische übertragen werden: Die Cochlea führt eine Kurzzeit-Spektralanalyse des angelegten Schalls durch. Dieses Kurzzeitspektrum ist die Eingabe der Neuronen in den ersten Verarbeitungsstufen. Das bedeutet, rezeptive Felder von auditorischen Neuronen haben nicht nur eine zeitliche, sondern auch eine spektrale Dimension. Verwendet man also ein lineares rezeptives-Feld-Modell für auditorische Neuronen, so sollte dieses durch ein geeignetes Cochleamodell ergänzt werden, wenn man möglichst einfache Ergebnisse erzielen möchte.

ist.

Im Einzelnen verfolgt diese Arbeit folgende Ziele:

- Sie soll eine kritische Übersicht über Verfahren zur Messung des Antwortverhaltens von Neuronen im auditorischen System geben und deren Grenzen aufzeigen.
- Entwicklung von verbesserten, wohlbegründeten Verfahren zur Messung des Antwortverhaltens von Neuronen im auditorischen System.
- Der für die Analyseverfahren notwendige Bedarf an Rechenleistung sollte nur so hoch sein, dass sie in der Praxis eingesetzt werden können. Eine Auswertung in Echtzeit wird dabei als optimal, eine Auswertung „über Nacht“ in zehnfacher Echtzeit als vertretbar angesehen.
- Die zur Charakterisierung eines Neurons notwendige Messzeit sollte idealerweise zehn Minuten, jedoch höchstens etwa eine Stunde, nicht überschreiten.
- Testen der Verfahren anhand von künstlichen, simulierten Neuronen und Aufzeigen der Grenzen und systematischen wie zufälligen Fehler.

Als Grundlage zur Erstellung eigener Programme in C++ unter Linux bzw. Unix empfehle ich die Lektüre von Breyman (1999), Stallman und Goyal (1994), Stroustrup (1998), Loukides und Oram (1997), Stallman et al. (1995) und Stallman und Support (1995) in etwa dieser Reihenfolge. Hilfreich waren auch Glass und Schuchert (1996), Josuttis (1997), Chan (1997) und Meyers (1999a,b, 2001). Eine Einführung zur Erstellung objektorientierten Codes findet gibt Gamma et al. (1997).

Dieses Dokument wurde mit $\LaTeX 2_{\epsilon}$, $\TeX 3.14159$ und $\text{BIB}\TeX 0.99c$ erstellt (Knuth 1993, Kopka 1994, 1995, 1997, Downes 2002). Die Bilder wurden mit Hilfe von Xfig 3.2 (Sato und Smith 2000), Xmgrace 5.1.2 (Grace Team 2000) und OpenDX 4.1.3 (IBM 2001) erstellt. Als Editor diente XEmacs (Stallman und Goyal 1994, Stallman et al. 1995), als Literaturdatenbank Xmbase-Grok. Unter anderem wurden folgende Programme zur Programmierung der Simulationen unter dem Betriebssystem Suse Linux 7.1 bis 9.2 verwendet: GNU C++-Compiler GCC 2.95.2 und GCC 3.x (Stallman 2001), GNU Debugger GDB 5.0 (Stallman et al. 2001), der GNU Profiler gprof (Fenlason 2006), der Parsergenerator GNU Bison 1.28 (Donnelly und Stallman 2002), GNU Make (Stallman und McGrath 2000), GNU Automake 1.4 (MacKenzie und Tromey 2002) und Autoconf 2.13 (MacKenzie und Elliston 2001). Als IDE diente KDevelop 2.0 (Nolden 2001). Programmiersprachen waren C++ (C++ Standard Committee 1997, Stroustrup 1998), Perl (Wall et al. 1998) und Bash

(Free Software Foundation 2002). Die numerische Bibliothek „GNU Scientific Software Library“ wurde verwendet (Pierce 1996). Berechnungen wurden verteilt auf einem Computer-Cluster mit Hilfe der Sun Grid Engine durchgeführt (Sun Microsystems 2002a,b,c).

Bei analytischen Berechnungen und zur Darstellung einiger Graphen half Maple V, Release 4 (Char 1991b,a, 1993). Zur Verwaltung der Simulationsergebnisse dienten die SQL-Datenbanken MySQL (MySQL AB 2001) und PostgreSQL (Lockhart 2000).

Beim Schreiben hilfreich waren Steenrod et al. (1973), Krantz (1997) und Beutelspacher (1999).

«*L'homme pense; donc je suis* »,
dit l'univers.¹

Paul Valéry

1

Einleitung

Hier werden die theoretischen und experimentellen Grundlagen und Vorarbeiten zur vorliegenden Arbeit erläutert.

1.1 Allgemeine Fragestellung

In diesem Abschnitt werden die allgemeinen Fragestellungen dieser Arbeit vorgestellt. Dies sind die

- Beschreibung,
- Messung,
- Vorhersage und
- Charakterisierung

des Antwortverhaltens von Neuronen. Im Folgenden sollen diese Begriffe näher erläutert werden.

Ein Neuron reagiert auf einen Stimulus, dargestellt in Abbildung 1.1, mit einer Folge von Aktionspotentialen („Spiketrain“).

Was kann man nun als das „Antwortverhalten“ des Neurons auf den betrachteten Stimulus bezeichnen? Offenbar nicht den Spiketrain, denn der fällt ja

¹„Der Mensch denkt, also bin ich“, sagte das Universum.

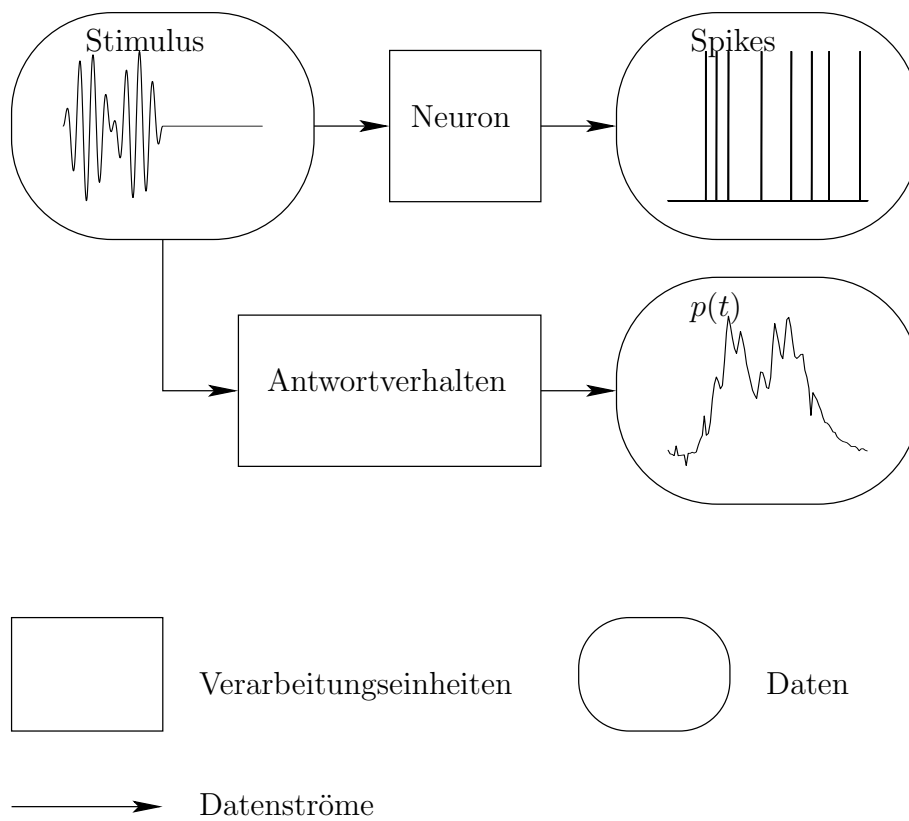


Abbildung 1.1: Reaktion eines Neurons auf einen Stimulus mit Aktionspotentialen. Kästen bedeuten Verarbeitungseinheiten, Kästen mit abgerundeten Ecken Daten und Pfeile Datenströme. Das Antwortverhalten des Neurons ist eine Verarbeitungseinheit, die aus dem Stimulus die Wahrscheinlichkeitsdichte $p(t)$ berechnet, dass das Neuron zum Zeitpunkt t ein Aktionspotential erzeugt.

mit jeder Wiederholung des Stimulus verschieden aus. Das Antwortverhalten muss allerdings auf jeden Fall die Wahrscheinlichkeitsdichte, dass das Neuron zur Zeit t nach dem Einsetzen des Stimulus feuert, beinhalten. Diese Wahrscheinlichkeitsdichte beschreibt das Antwortverhalten des Neurons auf diesen Stimulus allerdings nicht vollständig, denn sie berücksichtigt nicht die Abhängigkeit *zwischen* Aktionspotentialen, z. B. durch das Refraktärverhalten des Neurons (vgl. Abschnitt 1.5). Außerdem wird die Abhängigkeit der Feuerzeitpunkte *verschiedener* Neuronen untereinander nicht berücksichtigt, wenn man nur ein einzelnes Neuron betrachtet.²

Wenn die Wahrscheinlichkeitsdichte $p(t)$ für das Auftreten eines Aktionspotentials zum Zeitpunkt t nach Einsetzen des Stimulus das Antwortverhalten eines Neurons *auf diesen Stimulus* weitgehend vollständig beschreibt, was beschreibt dann das Antwortverhalten des *Neurons*? Offenbar die Wahrscheinlichkeitsdichte für das Auftreten eines Aktionspotentials *in Abhängigkeit* vom Stimulus. Wenn wir also zu jedem *beliebigen* Stimulus die Wahrscheinlichkeitsdichten für die Erzeugung eines Aktionspotentials durch das Neuron angeben können, dann haben wir das Antwortverhalten des Neurons.³

Wie kann man nun das Antwortverhalten eines Neurons grundsätzlich messen? Abbildung 1.2 gibt einen Überblick.

Das Antwortverhalten eines Neurons auf *einen* Stimulus kann man durch sein Peri-Stimulus-Zeit-Histogramm (PSTH) schätzen⁴. Um das Antwortver-

²Diese Abhängigkeiten beobachtet man, wenn Korrelationen zwischen den Aktionspotentialen einzelner Neuronen festgestellt werden, wie z. B. bei den Ganglionzellen in der Retina des Salamanders (Brivanlou et al. 1998). Natürlich wird ein Neuron, das von einem anderen eine exzitatorische Eingabe erhält, bevorzugt dann ein Aktionspotential erzeugen, wenn dieses andere Neuron gerade ein Aktionspotential erzeugt hat. Diese Abhängigkeiten von anderen Neuronen werden in einer Beschreibung des Antwortverhaltens, das nur den gegebenen *Stimulus* berücksichtigt und nicht auch die Aktivität anderer Neuronen, nicht behandelt.

³Wenn wir uns allerdings beispielsweise auf auditorische Stimuli beschränken, dann haben wir die Wechselwirkungen verschiedener Sinnesmodalitäten untereinander nicht berücksichtigt. Im optischen Tektum der Schleiereule gibt es z. B. Neurone, die sowohl auf visuelle als auch auf auditorische Stimuli reagieren (Knudsen und Brainard 1995, Miller und Knudsen 1999). Vergleichbare Neurone findet man im optischen Tektum des Kralenfrosches, welche sowohl auf optische Reize als auch auf Wasserwellen reagieren (Lowe 1987). Es sind auch Neuronen denkbar, die nur bei gleichzeitiger auditorischer *und* visueller Stimulation verstärkt Aktionspotentiale erzeugen. Möglicherweise hat man derartige Neurone bisher nicht gefunden, da Experimente gewöhnlich nur mit einer Sinnesmodalität gleichzeitig durchgeführt werden. Selbst wenn die Antwort des Neurons unabhängig von der Stimulation anderer Sinnesmodalitäten ist, kann sie noch vom inneren Zustand des Lebewesens abhängen, z. B. von der Aufmerksamkeit des Lebewesens (Kastner et al. 1998) oder davon, was das Lebewesen gerade denkt.

⁴Um *ein* Peri-Stimulus-Zeit-Histogramm (PSTH) auf einen Stimulus zu ermitteln, wiederholt man den Stimulus N -mal, man „spielt“ dem Neuron also N -mal einen identischen

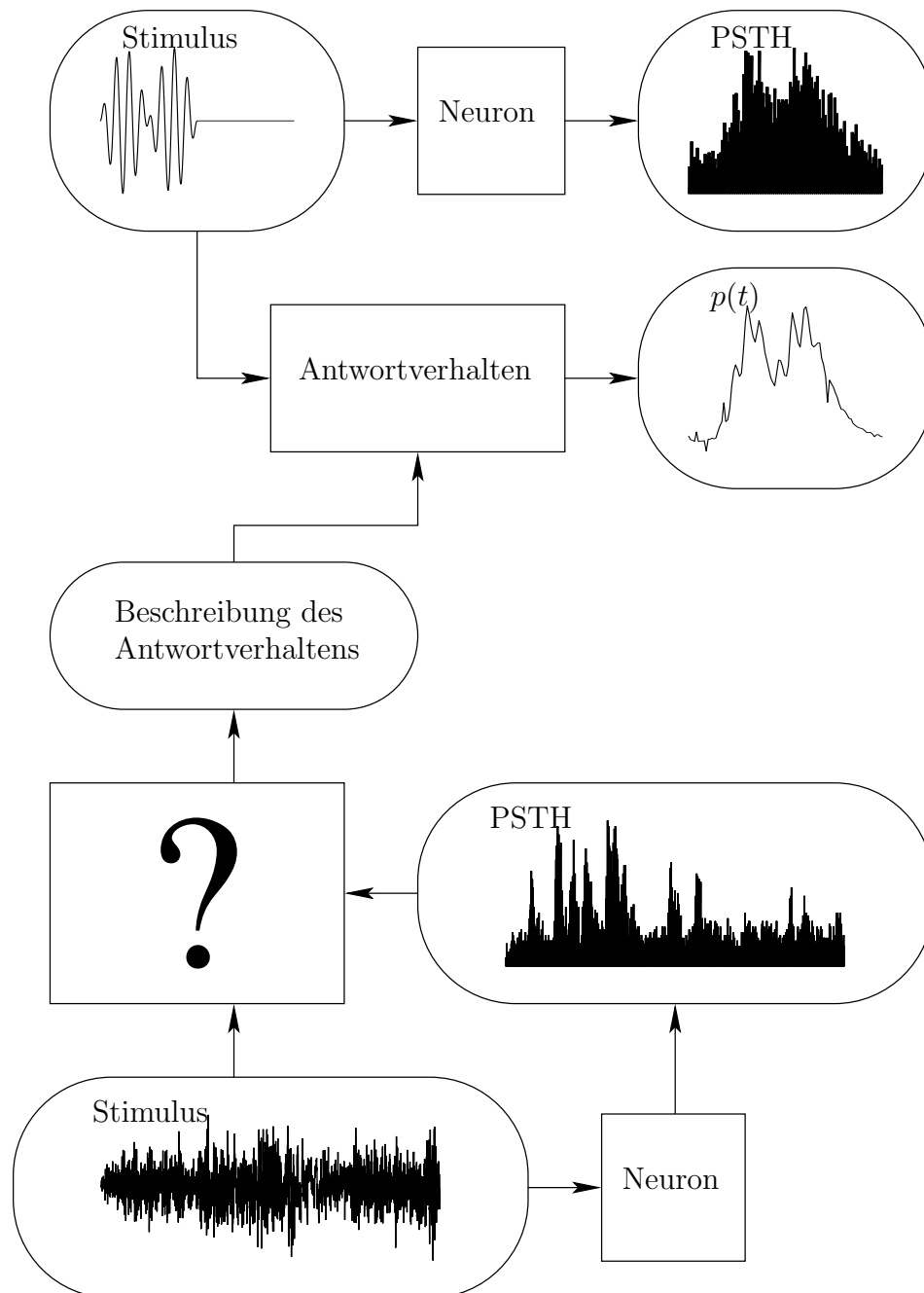


Abbildung 1.2: Grundsätzlicher Aufbau zur Messung des Antwortverhaltens eines Neurons. Das unten Zeile dargestellte Experiment besteht darin, einen (u.U. sehr langen) Stimulus zu wählen, und das Peri-Stimulus-Spike-Histogramm (PSTH) des Neurons elektrophysiologisch zu messen. Durch Anwendung eines mit „?“ bezeichneten, noch unbekanntes Verfahrens, wird eine Beschreibung des Antwortverhaltens des Neurons berechnet. Hieraus ergeben sich die Wahrscheinlichkeitsdichten $p(t)$ zur Erzeugung eines Aktionspotentials bei einem *anderen* Stimulus (oben links). Symbole wie in Abbildung 1.1.

halten eines Neurons auf *beliebige* Stimuli zu messen, kann man ebenfalls nur folgendes Experiment machen: Man legt einen Stimulus an und misst die Aktionspotentiale des Neurons.⁵ Die „Messung“ des Antwortverhaltens ist die

Schall vor. Seien t_i die Zeitpunkte, zu denen das Neuron, gemessen seit Beginn des jeweiligen Schalls, ein Aktionspotential erzeugt (nicht des Gesamtschalls, der aus einer Wiederholung von einzelnen identischen Schallen besteht, sondern seit Beginn eines einzelnen eben dieser identischen Schalle). Das Tupel (n_1, \dots) mit $n_j = |\{t_i | (j-1)\Delta t \leq t_i \leq j\Delta t\}|$ heißt Peri-Stimulus-Spike-Histogramm mit „Binbreite“ Δt . Hierbei bezeichnen die „|“ die Mächtigkeit einer Menge, also hier die Anzahl Aktionspotentiale in einem Bin. Natürlich ist das PSTH eine Zufallsgröße, d. h. man erhält bei Wiederholung des Experiments ein anderes PSTH. Ist die Zahl N der Wiederholungen jedoch ausreichend und die Binbreite Δt klein genug, so ist $n_j/(N\Delta t)$ eine gute Schätzung für die Wahrscheinlichkeitsdichte $p(t)$, dass das Neuron im Zeitbereich $[t, t + dt]$ ein Aktionspotential erzeugt, mit $t \in [(j-1)\Delta t, j\Delta t]$. Dies gilt jedoch nur für den Fall, dass die Erzeugung eines Aktionspotentials zum Zeitpunkt t nach dem Beginn einer Schallpräsentation unabhängig ist von der Erzeugung von Aktionspotentialen während einer anderen Präsentation des identischen Schalls. Dies ist nicht immer der Fall. Neurone können an einen sich immer wiederholenden Schall „adaptieren“, d. h. sie erzeugen von Mal zu Mal weniger Aktionspotentiale. Dies versucht man bei der Messung von PSTHs dadurch auszugleichen, dass man zufallsgesteuert eine Auswahl aus all den Schallen präsentiert, deren PSTH man messen will, so dass das Neuron nicht auf einen bestimmten Schall adaptieren kann (Nieder und Klump 1999). Dann berücksichtigt man bei der Bestimmung eines PSTH eines Schalls natürlich nur die Aktionspotentiale, die während dieses bestimmten Schalls aufgetreten sind. Dieses Verfahren kann ein eventuell auftretendes komplizierteres Adaptionsverhalten jedoch grundsätzlich nicht ganz vermeiden. So können Neurone im auditorischen System ihr Antwortverhalten durchaus langfristig ändern (man bezeichnet dies auch als „lernen“), und zwar *in Abhängigkeit* von den vorhergehenden Schallpräsentationen (Ohl und Scheich 1997, Ohl et al. 2001). Das Problem der Adaption ist wohl grundsätzlich nicht gänzlich behandelbar, da allgemeine Adaption zu Lernen äquivalent ist und sich durch Lernen die Eigenschaften der Neurone langfristig ändern, was dazu führt, dass die Stichproben aus einer nicht konstanten Wahrscheinlichkeitsverteilung gezogen werden (vgl. Abschnitt 8.3.2). Deshalb wird eventuelle Adaption in dieser Arbeit nicht speziell berücksichtigt. Zum Poisson-Neuron vgl. Kempster et al. (1998b), allgemein zur Ereignisanalyse Blossfeld et al. (1986), Blossfeld und Rohwer (1995).

⁵Der Stimulus kann natürlich beliebig lang sein und natürlich können innerhalb des Gesamtstimulus einzelne Stimuli wiederholt werden, z. B. auch um zu beurteilen, wie genau einzelne Folgen von Aktionspotentialen von Mal zu Mal reproduzierbar sind. Dies kann sehr genau sein, z. B. auf etwa 10 ms im Vorderhirn des Staren (Nieder und Klump 2001) und auf 1–10 ms bei Ganglionzellen in der Retina des Salamanders in Folgen von Aktionspotentialen, bei denen die einzelnen Aktionspotentiale etwa 100 ms voneinander getrennt sind (Berry et al. 1997, Berry und Meister 1998) sowie bei natürlichen Stimuli in einem visuellen Neuron der Fliege (de Ruyter van Steveninck et al. 1997). Szucs et al. (2004) stellt fest, dass Spiketrains bei gleichem Input-Strom reproduzierbar sind. J. et al. (2001) stellt ein Modell für retinale Ganglionzellen von Salamander, Katze und Kaninchen vor, das in der Lage ist, einzelne Spikes vorauszusagen. Die Modellierung des Antwortverhaltens dieser Neurone als Poisson-Prozess genügt dann *nicht*, es ist vielmehr auch das Refraktärverhalten der Neurone zu berücksichtigen (Berry und Meister 1998). Die durch den *Zeitpunkt* eines Feuer-Ereignisses übertragene Information ist auch viel größer als die

Folgerung des Antwortverhaltens aus dem Experiment. d. h. aus angelegtem Stimulus und gemessenen Aktionspotentialen wird das Antwortverhalten bei *beliebigem* Stimulus ermittelt. Die Messung sollte also eine *Verallgemeinerung* von Bekanntem (bei gegebenem Stimulus gemessene Aktionspotentiale) auf Unbekanntes (Aktionspotentiale oder Wahrscheinlichkeitsdichte derselben bei beliebigem Stimulus) ermöglichen.

Eine prinzipielle Einschränkung hierbei ist, dass die Vorhersage des Antwortverhaltens eines Neurons bei beliebigem Stimulus nur mit gewisser Wahrscheinlichkeit möglich ist. d. h. alles was man verlangen kann, ist, dass die Wahrscheinlichkeit, dass die Abweichung zwischen tatsächlichem und vorausgesagtem Verhalten groß ist, gering ist (vgl. hierzu Abschnitt 8.3.2).

Außerdem genügt es nicht unbedingt, nur das Peri-Stimulus-Zeit-Histogramm eines Neurons bei einem bestimmten Stimulus anzugeben. Es könnte z. B. sein, dass das Neuron auf einen Stimulus mit einem Aktionspotential nahe des Maximums und mit einem Aktionspotential nahe des Minimums antwortet, wie in Abbildung 1.3 oben. Das gleiche Neuron antwortet auf einen anderen Stimulus mit zwei Aktionspotentialen nahe des Maximums in der Hälfte aller Fälle wie in Abbildung 1.3 unten. Dann kann das Peri-Stimulus-Zeit-Histogramm in beiden Fällen *gleich* sein, das Neuron reagiert aber trotzdem offenbar *unterschiedlich* auf die beiden Stimuli. Dies liegt daran, dass man eigentlich die Wahrscheinlichkeit für einen *bestimmten* Spike-train angeben müsste, gegeben den jeweiligen Stimulus. Dies ist aber ein sehr schwieriges Problem *und* man würde dazu vermutlich riesige Datenmengen benötigen. Deshalb gibt es hierzu auch keinerlei Ansätze.

Abbildung 1.4 zeigt ein Beispiel, bei dem zwei *unterschiedliche* Neurone bei *gleichem* Stimulus das gleiche Peri-Stimulus-Zeit-Histogramm haben, jedoch dennoch unterschiedliches Antwortverhalten zeigen. Ist die Phase, bei der ein Aktionspotential erfolgt, zu Beginn des Stimulus auch noch zufällig, dann ist das Peri-Stimulus-Zeit-Histogramm der beiden Neuronen nicht nur gleich, sondern auch konstant. Dieses Ergebnis würde suggerieren, dass der Spike-train garnicht mit dem Stimulus zusammenhängt. In Wirklichkeit hängt die Antwort des Neurons aber sehr wohl vom Stimulus ab.

Die Messung aus Abbildung 1.2 soll also eine Beschreibung des Antwortverhaltens liefern. Die Frage ist: Wie sollte eine solche Messung aussehen? Was zunächst geklärt werden muss, ist jedoch, wie das Antwortverhalten eines

durch die *Zahl* der Aktionspotentiale übertragene (Berry et al. 1997), so dass Methoden benötigt werden, die einzelne Aktionspotentiale vorhersagen. Tatsächlich wird Länge des Stimulus bei einem betäubten Tier i. d. R. auf einige Stunden eingeschränkt, z. B. 2-3 Stunden bei Faulstich und Kössl (2000). Bei wachen Tieren, denen man chronisch Ableit-elektroden implantiert hat, kann man Tage lang Aktionspotentiale vom gleichen Neuron oder zumindest der gleichen Gruppe von Neuronen messen (Nieder und Klump 1999).

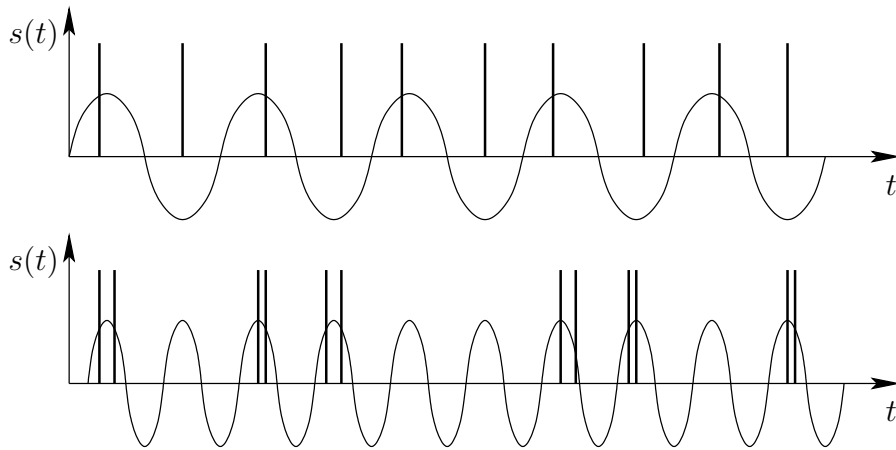


Abbildung 1.3: Sinusoide Stimuli s mit *gleichen* Peri-Stimulus-Zeit-Histogrammen, jedoch offenbar *unterschiedlichem* Antwortverhalten des Neurons. Senkrechte Striche bedeuten Aktionspotentiale. Der obere Stimulus veranlasst das Neuron, regelmäßig ungefähr im Maximum und Minimum des Stimulus ein Aktionspotential zu erzeugen. Beim unteren Stimulus mit der doppelten Frequenz tritt nur im Durchschnitt bei der Hälfte aller Maxima Aktivität auf, jedoch dann jeweils zwei Aktionspotentiale.

Neurons auf einen *beliebigen* Stimulus überhaupt beschrieben werden kann. Die Verarbeitungseinheit „Antwortverhalten“ in Abbildung 1.2 kann als „Modell“ des Neurons aufgefasst werden. Welches sind die Anforderungen an ein gutes Modell?

- Es soll Vorhersagen ermöglichen.
- Es soll wirkliche Verhältnisse widerspiegeln. d. h. es sollte über den tatsächlichen Aufbau des untersuchten neuronalen Systems Auskunft geben.
- Es soll eine anschauliche Beschreibung des Systems bieten, so dass Eigenschaften des Systems leicht aus den Modellparametern gefolgert werden können.
- Die Messung der Modellparameter sollte wenig Messzeit beanspruchen und auch sonst nicht zu aufwändig sein.
- Die zur Berechnung der Modellparameter aus der Messung beanspruchte Rechenzeit sollte akzeptabel sein.

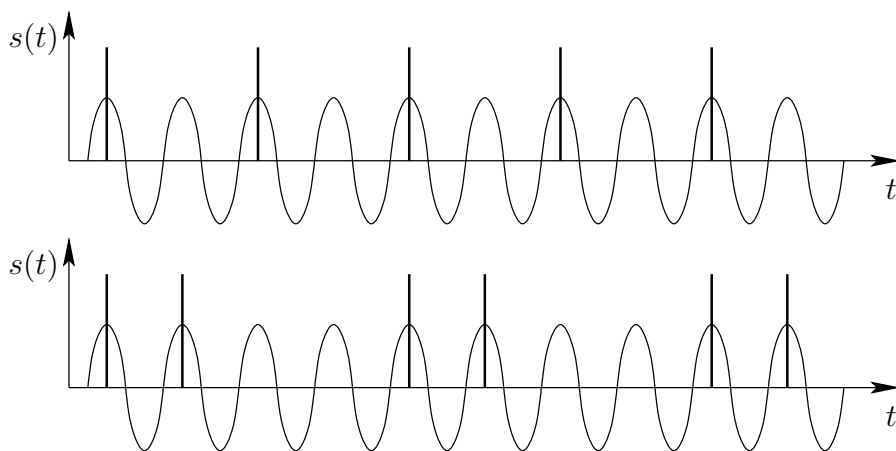


Abbildung 1.4: Zwei unterschiedliche Neurone mit dem gleichen Stimulus s und gleiche Peri-Stimulus-Zeit-Histogrammen, jedoch unterschiedlichem Antwortverhalten. Das obere Neuron erzeugt regelmäßig an jedem zweiten Maximum des Stimulus ein Aktionspotential. Das untere Neuron erzeugt regelmäßig an jedem ersten und zweiten Maximum ein Aktionspotential, nicht jedoch an den dritten und vierten Maxima. Mit welchem „Takt“ die Neurone bei Einsetzen des Stimulus beginnen, sei zufällig. Das Peri-Stimulus-Zeit-Histogramm beider Neurone ist dann gleich, das Antwortverhalten jedoch offenbar verschieden.

„Vorhersagen ermöglichen“ bedeutet, Auskunft über das Antwortverhalten nicht nur bei den *gemessenen* Stimuli, sondern auch bei *neuen* Stimuli geben.⁶ Ein Modell, das zwar das Antwortverhalten beschreibt, dessen Modellparameter jedoch nicht auf Parameter des untersuchten Systems abbildbar sind, ist nicht mehr als eben eine Beschreibung. Es trägt nichts zum Verständnis des funktionalen Aufbaus des Systems bei. Wenn das Modell keine anschauliche Beschreibung des untersuchten neuronalen Systems bietet, sondern eine sehr komplizierte, z. B. wenn die Beschreibung ein Backpropagation-Netzwerk (Rojas 1996, Zell 2000) ist, dann fällt es schwer, aus dem Modell Folgerungen über die Funktionsweise des Systems zu ziehen, also nicht nur zu verstehen *wie* das System funktioniert, sondern auch *warum*. Falls die Messung der Modellparameter zu lange dauert oder aus sonstigen Gründen in der Praxis nicht durchführbar ist, ist das Modell wertlos. Auch wenn die physikalische Messung selbst nur wenig Zeit beansprucht, jedoch die Auswertung der Messung zu lange dauert, ist das Modell nicht praktikabel. Ideal wäre, wenn die Auswertung nicht länger dauert als die Messung selbst, d. h. in Echtzeit stattfinden kann, denn dann kann der Experimentator sofort auf die Ergebnisse reagieren.

1.2 Ansätze zur Messung des Antwortverhaltens

Am einfachsten wäre es, das Neuron wie in Abbildung 1.5 als linearen Filter zu beschreiben. Der Stimulus wird also mit einer Impulsantwort gefaltet. Diese kann man auch als „rezeptives Feld“ bezeichnen.⁷

Ein Vorteil dieser Beschreibung als linearer Filter ist, dass die Impulsantwort einfach durch eine Korrelation zwischen Stimulus und gemessener Antwort y ermittelt werden kann, falls der Stimulus Gauß'sches weißes Rauschen ist, vgl. Abschnitt 4.1. Man benötigt also wenig Rechenzeit für die Auswertung

⁶Ein Modell, das nur über die Messung selbst Auskunft gibt, kann bestenfalls als sehr eingeschränkt bezeichnet werden. Ein Modell, das nichteinmal dies leistet, ist rein zufällig und somit wertlos.

⁷Falls $r(t)$ in der Umgebung von $t = 0$ sehr groß ist und ansonsten sehr klein, bedeutet das, dass der Stimulus durch das Neuron unverfälscht übertragen wird. Falls $r(t)$ in der Umgebung von $t = 100$ ms sehr groß ist und ansonsten sehr klein, dann wird der Stimulus durch das Neuron um 100 ms verzögert. Falls $r(t) = -1$ für $-1 \text{ ms} \leq t < 0$, $r(t) = 1$ für $0 \leq t \leq 1 \text{ ms}$ und $r(t) = 0$ sonst, dann wird durch das Neuron im Wesentlichen die Ableitung des Stimulus übertragen, d. h. das Neuron reagiert mit einer „on“-Antwort nur auf den Anfang eines konstanten Stimulus $x(t) = 1$ für $t \geq 0$ und $x(t) = 0$ sonst. Im *Nucleus cochlearis* des auditorischen Systems und im auditorischen Cortex gibt es Neurone, die nur auf den Beginn eines Reintons mit Aktionspotentialen reagieren (Pickles 1991, Yost 1994).

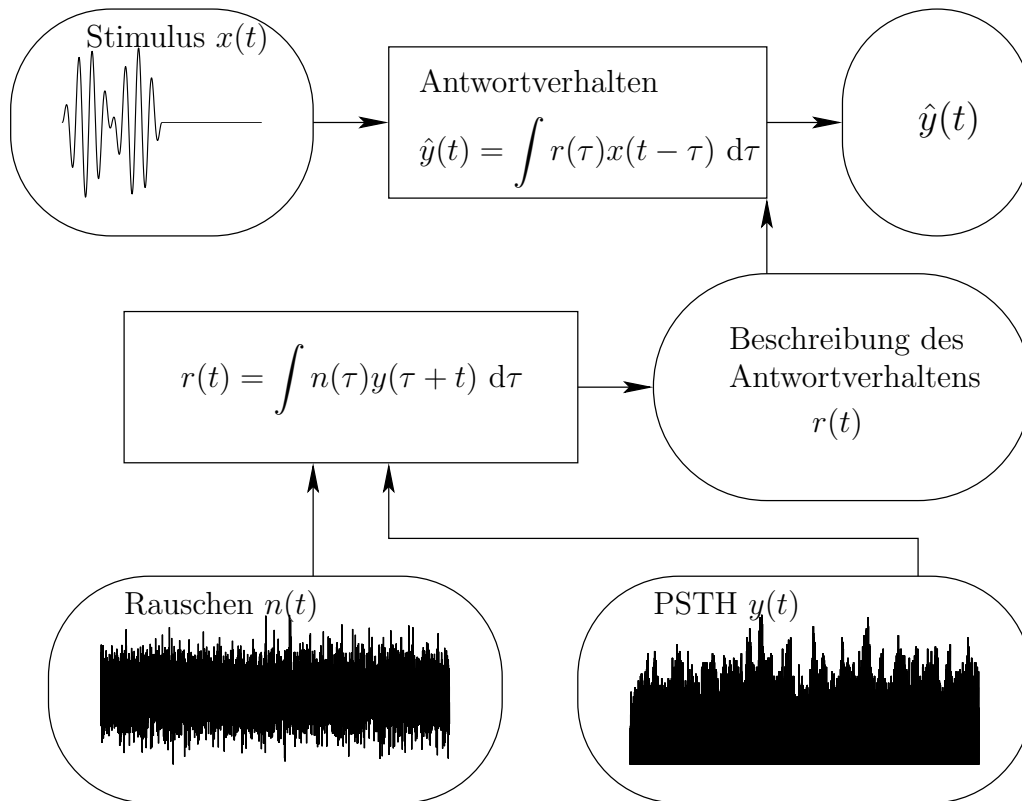


Abbildung 1.5: Das Antwortverhalten des Neurons wird als linearer Filter mit Impulsantwort r beschrieben. Die geschätzte neuronale Antwort \hat{y} ergibt sich durch Faltung von r mit dem Stimulus x . Mit Gauß'schem weißem Rauschen n als Stimulus für die unten abgebildete Messung der neuronalen Antwort y kann die Impulsantwort r durch die angegebene Korrelation zwischen Rauschen n und gemessener Antwort y geschätzt werden.

der Messung. Allerdings sind z. B. auditorische Neurone im Allgemeinen nicht linear. Warum *können* Neurone des auditorischen Systems nicht als lineare Elemente beschrieben werden?

- Ein lineares Neuron würde in seiner Feuerrate einem harmonischen Stimulus mit der gleichen Frequenz folgen. Dies ist aber im Hörnerven nur bei niedrigen Frequenzen der Fall (Popper und Fay 1992) und in höheren Arealen des auditorischen Systems unter Umständen überhaupt nicht.
- Ein Neuron kann nicht beliebig schnell feuern, es sättigt bei Stimuli hoher Intensität. Im auditorischen System müssen Stimuli mit Schalldruckpegeln von 0 bis etwa 100 dB betrachtet werden⁸, also 10 Größenordnungen Unterschied in der Intensität und 5 Größenordnungen in der Amplitude. Sättigungseffekte spielen daher eine Rolle.
- Bei einem linearen Ansatz treten unweigerlich negative Feuerwahrscheinlichkeiten auf, die nicht zu interpretieren sind.⁹

Solch ein einfacher linearer Ansatz ist zwar, mit einiger Übung, anschaulich und beansprucht wenig Rechenzeit. Zumindest für das auditorische System muss er aber aus den angegebenen Gründen verworfen werden.¹⁰

Ein wesentlicher Grund dafür, warum auditorische Neuronen nicht linear im Schall sind, ist die nichtlineare Vorverarbeitung in der Cochlea¹¹. In der Cochlea wird der Schall, d.i. die am Trommelfell ankommende zeitabhängige Luftdruckschwankung, nach Frequenzbereichen gefiltert, vgl. Abschnitt 3.1.

⁸Der Schalldruckpegel, der gerade noch wahrgenommen werden kann, beträgt etwa 0 dB, ist allerdings frequenzabhängig (Zwicker und Fastl 1999). Bei Schalldruckpegeln von ca. 130 dB liegt die Schmerzgrenze (Zenner 1994).

⁹Ist allerdings die spontane Feuerrate eines Neurons hoch, treten bei Stimuli geringer Intensität keine negativen Feuerraten auf. Es *kann* also sein, dass sich Neurone ausreichend gut als lineare Filter beschreiben lassen, so z. B. in Teilen des vestibulären Systems (Wilden et al. 2002).

¹⁰Bei der Beschreibung von Neuronen des visuellen Systems im Thalamus hatte man allerdings mit dem linearen Ansatz aus Abbildung 1.5 gute Erfolge (Cai et al. 1997). Teile der menschlichen visuellen Wahrnehmung lassen sich ebenfalls gut durch eine lineare Systemtheorie beschreiben (Hauske 1994). Allerdings sind der Kapazität von linearen Filtern als Feature-Dektoren theoretische Grenzen gesetzt (Zetzsche und Barth 1990). Beim auditorischen System des Grasfrosches scheitert die lineare Beschreibung, wie zu erwarten (Aertsen et al. 1980).

¹¹Nichtlinearitäten noch höherer Ordnung, sogenannte Zweiton-Interaktionen, beobachtet man allerdings auch schon auf der Ebene des *Nucleus cochlearis dorsalis* (Nelken und Young 1997, Nelken et al. 1997).

Die inneren Haarzellen feuern um so schneller, je mehr Energie in den entsprechenden Frequenzfilter gelangt. Die Operation der Berechnung der Energie aus der Auslenkung (Bildung des Betragsquadrats) ist nicht linear.

Die zeitabhängige Feuerwahrscheinlichkeit der inneren Haarzellen bezeichnen wir als „Cochleogramm“. Ein besserer Ansatz ist es nun, ein auditorisches Neuron als linear im Cochleogramm anzunehmen. Die Antwort ist dann eine Faltung des Cochleogramms mit einem „spekro-temporalen“ rezeptiven Feld, wie in Abbildung 1.6.

Dieser Ansatz ist auch anschaulich. Am rezeptiven Feld kann man z. B. sofort die Isointensitätskurve des Neurons ablesen.¹² Allerdings lässt sich das rezeptive Feld nicht mehr einfach durch eine Korrelation wie in Abbildung 1.6 angeben, bestimmen, da weißes Gauß'sches Rauschen als Stimulus kein weißes Gauß'sches Rauschen im Cochleogramm erzeugt. Vielmehr wird durch Gauß'sches weißes Rauschen als Stimulus eine eher konstantes Cochleogramm erzeugt, da im Mittel jeder Frequenzkanal zu jeder Zeit gleich stark angeregt wird. Tatsächlich ist es unmöglich, auf der Cochlea Gauß'sches weißes Rauschen zu erzeugen.¹³ Selbst wenn dies möglich wäre oder wenn man vernachlässigt, dass das Cochleogramm kein Gauß'sches weißes Rauschen ist und sich mit näherungsweise Gauß'schem weißem Rauschen auf der Cochlea zufrieden gibt, sind lange Messzeiten für diesen Ansatz notwendig, da zufällige Korrelationen im Eingabesignal das Messergebnis verfälschen, und sich erst „herausmitteln“ müssen.¹⁴

¹²Die „Isointensitätskurve“ eines auditorischen Neurons ist die Feuerrate, aufgetragen über der Frequenz eines stimulierenden Reintons (Yost 1994). Die Feuerrate bei der Frequenz, die dem Frequenzkanal i des Cochleogramms entspricht, ist in dem in Abbildung 1.6 angegebenen Modell proportional zu $\int r_i(t) dt$. Alternativ wird auch oft die „Tuningkurve“ („Isoratenkurve“) gemessen. Hier wird die Schallintensität, die zu einer bestimmten Feuerrate führt, über der Frequenz des stimulierenden Reintons aufgetragen (Yost 1994).

¹³ Will man in einem einzigen Frequenzkanal auf der Basilarmembran Gauß'sches weißes Rauschen n erzeugen, und ist h die Impulsantwort dieses Frequenzkanals, so muss man einen Schall s wählen mit $n(t) = \int h(\tau)s(t - \tau) d\tau$. Fouriertransformation ergibt $1 = |H(\omega)S(\omega)|$, somit liegt die Fouriertransformierte S des Schalls bis auf die Phase fest. Die Forderung, Rauschen in einem einzigen Frequenzkanal der Basilarmembran zu erzeugen, legt also bereits den zu verwendenden Schall praktisch fest. Man könnte einwenden, dass für die Fouriertransformierte H der Impulsantwort h gilt $H(\omega) \approx 0$ für ω , die weit entfernt von der Bestfrequenz ω_0 der betrachteten Stelle auf der Basilarmembran sind und daher die Gleichung $1 = |H(\omega)S(\omega)|$ nicht eindeutig nach $|S(\omega)|$ auflösbar ist. Dennoch legt die Gleichung $|S(\omega)|$ in der Nachbarschaft von ω_0 fest, so dass benachbarte Frequenzkanäle niemals unabhängige Aktivität zeigen können, was bei Gauß'schem weißem Rauschen auf der Basilarmembran gefordert wäre. Will man in einem Frequenzkanal x_i des Cochleogramms Gauß'sches weißes Rauschen erzeugen, so hat man die Schwierigkeit, dass $x_i(t) \geq 0$ ist, jedoch Gauß'sches weißes Rauschen auch negative Auslenkungen enthalten müsste.

¹⁴Richard et al. (1995) spricht noch davon, dass die für die Auswertung der Messung

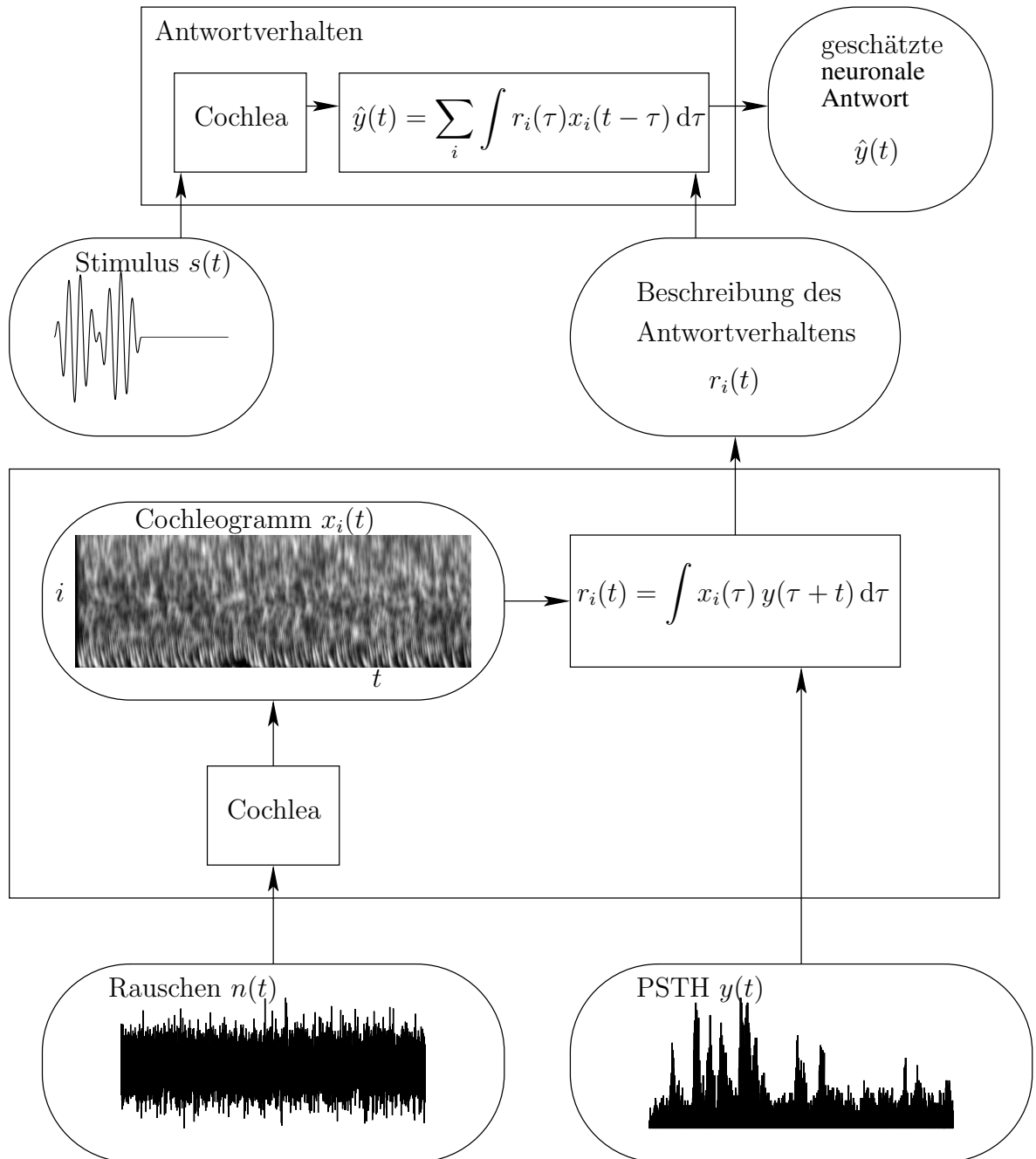


Abbildung 1.6: Das Antwortverhalten des Neurons wird als linear im Cochleogramm x_i angenommen. Das rezeptive Feld r_i wird (fehlerhaft!) durch Korrelation des Cochleogramms aus Gauß'schem weißem Rauschen n und der gemessenen Antwort y berechnet.

Einen Ausweg stellen Optimierungsverfahren dar, wie in Abbildung 1.7 dargestellt. Dieser naive Ansatz scheitert jedoch, da das rezeptive Feld r_i sehr viele Parameter enthält¹⁵ und nur wenige unabhängige Messungen vorliegen. Auch eine Beschreibung des Neurons in Form eines neuronalen Netzes ist denkbar, wie in Abbildung 1.8. Diese hat jedoch den Nachteil, dass das abgebildete „Modell“ sehr wenig mit der Wirklichkeit zu tun hat, da *ein* Neuron durch ein ganzes neuronales Netz aus vielen Neuronen simuliert wird und die nichtlineare Vorverarbeitung in der Cochlea komplett fehlt. Dadurch wären auch die Synapsengewichte sehr schwer zu interpretieren.

1.3 Mögliche Experimente

Um die Funktionsweise des Gehirns von Menschen und Tieren zu verstehen, können anatomische und physiologische Experimente sowie Verhaltensexperimente durchgeführt werden.

Bei *anatomischen Experimenten* beobachtet man, i. d. R. durch geeignete Färbung und Mikroskopie, welche Neuronen mit welchen anderen über Axone verbunden sind. Größe und Form der Neurone sowie ihrer Dendritenbäume können ebenfalls beobachtet werden. Die Art der Verbindungen lässt Rückschlüsse auf die Funktion und Funktionsweise der Neuronen zu. Diese Schlüsse können zunächst recht einfach sein. Da z. B. der Hörnerv direkt in den *Nucleus cochlearis* mündet (Yost 1994), kann man davon ausgehen, dass der *Nucleus cochlearis* im Hörsystem eine Rolle spielt. Durch anatomische Untersuchungen ist es auch möglich, Hinweise auf die genauere Funktionsweise eines neuronalen Subsystems zu erhalten. Beispielsweise beobachtet man im *Nucleus laminaris* der Schleiereule eine geordnete Struktur, die an das Jeffreys-Modell (Shepherd 1993) zur Richtungslokalisierung erinnert und dieses auch tatsächlich realisiert (Kempster et al. 2001, Leibold et al. 2001).¹⁶ Jedoch haben anatomische Experimente ihre Grenzen. Es ist anatomisch i. d. R. nicht möglich, Aussagen über die Stärke der synaptischen

benötigte Rechenzeit der limitierende Faktor sei, nicht die Messzeit selbst, obwohl diese auch sehr lang war (80 Stunden für ein Neuron). Hat der Experimentator vollständige Kontrolle über den Stimulus, lässt sich auch die Technik der „m-sequences“ anwenden (Bernardete und Victor 1994). Hier hat man jedoch gerade aufgrund vorgeschalteter Cochlea keine Kontrolle über den Stimulus, also das Cochleogramm.

¹⁵ Geht man von 20 Frequenzkanälen aus, einer zeitlichen Auflösung von 1 ms und einer maximalen Dauer des rezeptiven Feldes von 100 ms, d. h. $r(t) = 0$ für $t > 100$ ms, dann sind 2000 Parameter $r_i(t)$ zu bestimmen. Für 2000 unabhängige Messungen sind daher schon $2000 \cdot 100$ ms = 200 s erforderlich. Man benötigt jedoch mehr unabhängige Messungen, falls das Messergebnis verrauscht ist. Außerdem wäre ein lineares Gleichungssystem mit 2000 Unbekannten zu lösen, das noch dazu schlecht konditioniert ist.

¹⁶Neuroanatomische Untersuchungen am Krallenfrosch *Xenopus* zeigen eine dem

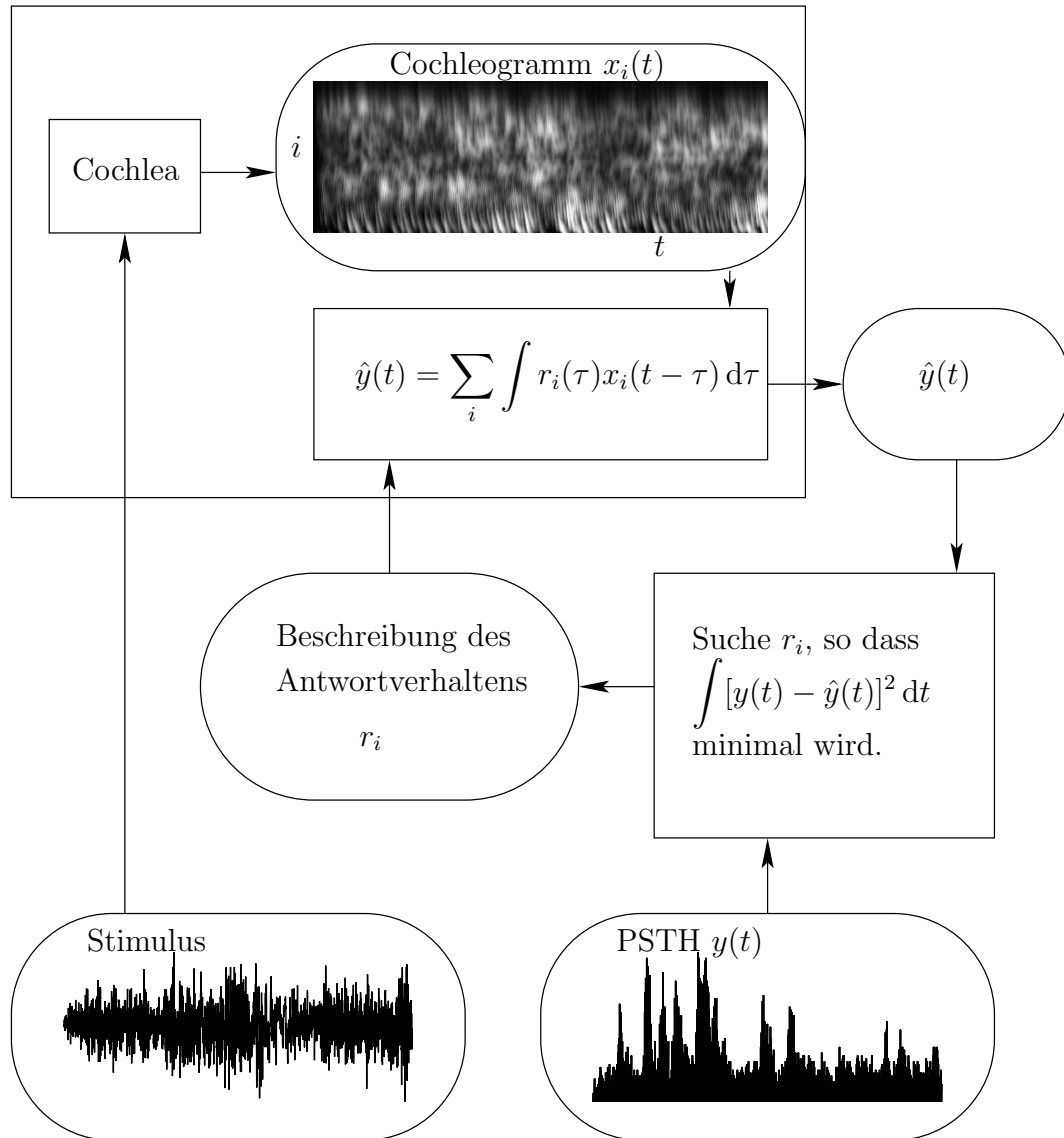


Abbildung 1.7: Das Neuron wird als linear im Cochleogramm x_i angenommen, wie in Abbildung 1.6. Das receptive Feld r_i wird so gewählt, dass die mittlere quadratische Abweichung zwischen der Messung y der neuronalen Antwort und deren Schätzung \hat{y} aus dem rezeptiven Feld minimal wird. Der Stimulus ist in diesem Fall Rauschen, das spektral und temporal korreliert ist, wodurch sich entsprechende Korrelationen im Cochleogramm ergeben.

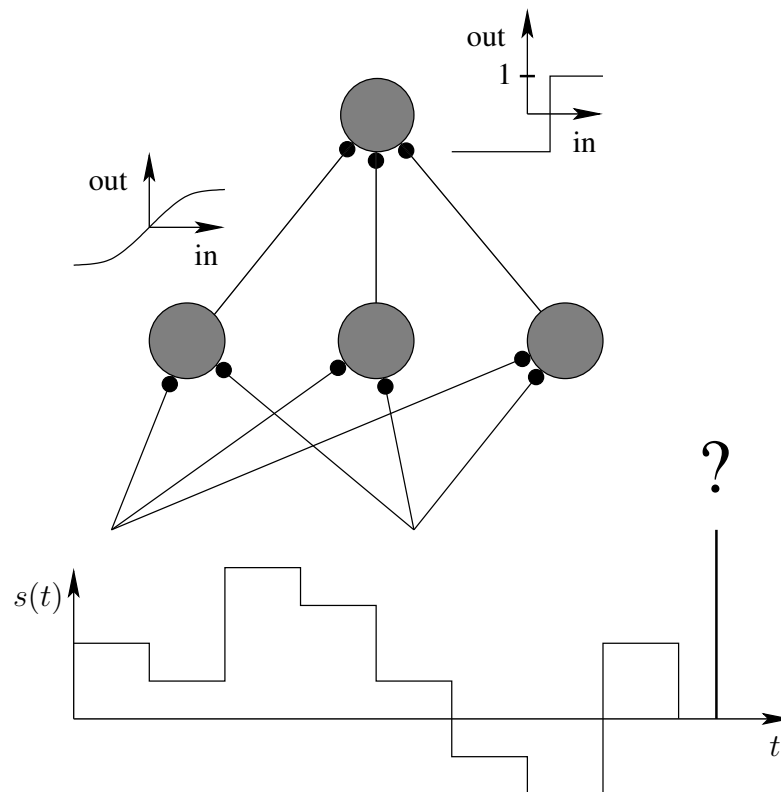


Abbildung 1.8: Neuronales Netz zur Vorhersage von Aktionspotentialen. Graue Kreise bedeuten Neurone, kleine schwarze Kreise Synapsen und gerade Striche Axone. Der zeitabhängige Stimulus s dient als Eingabe in ein künstliches neuronales Netz. Die Eingabe wird in einer Zwischenschicht von „Neuronen“ verarbeitet. Die Ausgabe dieser Zwischenschicht dient als Eingabe eines Perzeptrons (Rojas 1996), das mit seiner Ausgabe entscheidet, ob auf den Stimulus ein Aktionspotential folgt oder nicht. Die Ein-Ausgabe-Funktionen der Neurone sind schematisch dargestellt.

Verbindungen zwischen Neuronen zu machen.¹⁷ Leider ist es hauptsächlich diese *Stärke* der Verbindungen, die für die korrekte Funktion eines Neurons verantwortlich sind.¹⁸

Physiologische Experimente sollen Aufschluss über die *Funktion* eines neuronalen Subsystems geben. Bei einem elektrophysiologischen Experiment wird beispielsweise das Potential nahe einer Membran des Zellkörpers eines Neurons gemessen, wodurch die Zeitpunkte bestimmt werden können, zu denen das Neuron ein Aktionspotential erzeugt.¹⁹ I.d.R. wird dem untersuchten Tier gleichzeitig mit der Ableitung ein Stimulus präsentiert. Dies kann beispielsweise ein visueller, auditorischer oder vestibulärer Stimulus sein, etwa sich bewegende Balken, verschiedene Töne oder Drehungen und Beschleunigungen des Kopfes. Ist der Stimulus geeignet, verstärkt Aktionspotentiale hervorzurufen²⁰, so liegt der Schluss nahe, dass das untersuchte Neuron dazu beiträgt, die präsentierte Art von Stimulus zu „verarbeiten“. Ist das Neuron relativ scharf auf einen bestimmten Stimulus abgestimmt²¹, so wird oft vermutet,

Nucleus laminaris der Schleiereule ähnliche Struktur im Seitenlinienkern in der Medulla. Hier verzweigen sich die von den Seitenlinienorganen des Frosches kommenden Nervenfasern sehr stark (Lowe und Russell 1982, Altman und Dawes 1983, Will et al. 1985a,b). Der Seitenlinienkern könnte somit die von vielen verschiedenen Seitenlinienorganen kommenden Aktionspotentiale zusammen auswerten. Da die Seitenlinienorgane die Partikelgeschwindigkeit des Wassers an der Hautoberfläche des Frosches messen, könnte der Frosch durch Auswertung der neuronalen Antworten im Seitenlinienkern in der Lage sein, die Richtung, aus der eine Wasserwelle kommt, zu bestimmen (Sobotka 2002, Franosch et al. 2003b, 2005a,b). Diese Information nutzt der Frosch bei seinem Beutefangverhalten aus (Elepfandt 1996).

¹⁷Es gibt Ausnahmen. Beispielsweise enthält der *Nucleus magnocellularis* im auditorischen System der Schleiereule Axone, die das Zielneuron mit sehr vielen Synapsen kelchförmig umschließen (Carr und Boudreau 1991). Diese Verbindung ist auch tatsächlich besonders effektiv und dient zur Weitergabe exakter Zeitinformation zur Richtungslokalisierung (Kempner 1997).

¹⁸Die Stärke der Verbindungen kann durch Lernen verändert werden. Man vermutet hierin den Mechanismus für das Gedächtnis (Kandel et al. 1996).

¹⁹Mit extrazellulären Elektroden leitet man oft die Aktionspotentiale einer *Gruppe* von Neuronen ab. Diese Aktionspotentiale haben jedoch unterschiedliche Höhe und Form, je nach dem von welchem Neuron der Gruppe sie stammen. Mit „Spikesortern“ lassen sich die Aktionspotentiale den jeweiligen Neuronen zuordnen (Lewicki 1994).

²⁰Neuronen haben auch spontane Aktivität, d. h. sie erzeugen Aktionspotentiale auch ohne äußeren Stimulus. Diese Spontanaktivität kommt durch Eingaben von anderen Neuronen zustande. Echt zufällige Aktionspotentiale treten i. d. R. nicht auf, da Neuronen Aktionspotentiale bei definierter Eingabe sehr präzise und reproduzierbar erzeugen (de Ruyter van Steveninck et al. 1997, Ariav et al. 2003, Uzzell und Chichilnisky 2004, Szucs et al. 2004).

²¹d. h. ändert man die Stimulusparameter nur geringfügig, so erzeugt das Neuron wesentlich weniger Aktionspotentiale. Variiert man beispielsweise die Frequenz eines präsentierten Reintons und das Neuron erzeugt im Wesentlichen nur bei einer ganz bestimmten

dass das Neuron diesen Stimulus „detektiert“. Z.B. im *Nucleus cochlearis ventralis* des auditorischen Systems gibt es Neuronen, die auf bestimmte Frequenzen abgestimmt sind. Die Tatsache, dass benachbarte Neuronen auf ähnliche Frequenzen abgestimmt sind (Pickles 1991, Popper und Fay 1992), legt die Vermutung nahe, dass der *Nucleus cochlearis ventralis* tatsächlich dazu dient, festzustellen, welche Frequenzen im Schall enthalten sind.²² Da es sehr viele unterscheidbare Schalle gibt²³, erscheint es allerdings bei komplexeren Neuronen als sehr schwierig, unverhältnismäßig zeitaufwändig oder gar unmöglich, die Schalle bzw. Schallparameter zu bestimmen, auf welche die Neurone vorzugsweise reagieren. Mit Multielektroden-Ableitungen²⁴ las-

Frequenz Aktionspotentiale, so sagt man, das Neuron sei scharf auf eine „beste“ Frequenz abgestimmt. Natürlich kann es auch sein, dass das Neuron „in Wirklichkeit“ auf eine bestimmte Zeitdifferenz zwischen aufeinanderfolgenden Schalldruckmaxima reagiert, also etwa auf eine geeignete Pulsfolge genauso intensiv antwortet. Auch könnte es sein, dass die neuronale Antwort bei bleichzeitiger Präsentation zweier Reintöne unterschiedlicher Frequenz wieder verschwindet oder dass das Neuron noch intensiver antwortet, wenn man dem Tier arteigene Laute präsentiert, welche die beste Frequenz als Komponente enthalten. An diesem Beispiel zeigt sich schon, dass man sich widersprechende Deutungen eines Experiments nicht vermeiden kann, wenn man nur *einen* Stimulusparameter variiert.

²²Zweifel kommt auf, wenn man bedenkt, dass die Frequenzinformation auch im Hörnerv schon enthalten ist und dass man ähnliche Frequenzkarten auch in höheren Arealen des auditorischen Systems findet, die nachweislich nicht mehr ausschließlich der Weitergabe von Frequenzinformation dienen, wie der *Nucleus laminaris* der Schleiereule, der zur Richtungslokalisierung der Schallquelle dient (Kempster et al. 2001).

²³Die Schallpegelauflösung des menschlichen Gehörs ist etwa 1 dB (Zwicker und Fastl 1999). Bei einer Hörschwelle von ca. 0 dB und einer maximalen, gerade noch nicht unangenehmen Schalldruckamplitude von ca. 100 dB (Zenner 1994) ergibt dies etwa 100 unterscheidbare Lautstärken. Das menschliche Gehör verfügt über etwa 24 weitgehend unabhängige Frequenzkanäle (Zwicker und Fastl 1999). Somit ergeben sich etwa $100^{24} = 10^{48}$ unterscheidbare stationäre Schalle. Mit einer Zeitauflösung von etwa 100 ms (Zwicker und Fastl 1999) kann der Mensch potentiell $10^{48 \cdot 10} = 10^{480}$ verschiedene Schalle von 1 s Dauer unterscheiden. Effekte wie Maskierung und Nachverdeckung (Yost 1994, Zwicker und Fastl 1999) können die Zahl der unterscheidbaren Schalle reduzieren, sind aber in obiger Rechnung zum Teil schon berücksichtigt. Der übertragene Informationsfluss wäre somit maximal etwa 1,6 kBit/s. Zum Vergleich: Der Informationsgehalt der deutschen Sprache beträgt etwa 2 Bit pro Zeichen Singh (2002), was bei ca. 30 Zeichen pro Sekunde einem Informationsfluss von 60 Bit/s entspricht.

Der innere foveale Bereich des Auges hat eine räumliche Auflösung von etwa einer Bogenminute und ist ca. 1.5×1.5 Grad groß (Hauske 1994). Bei zwei unterscheidbaren Graustufen (tatsächlich sind natürlich mehr Graustufen unterscheidbar, jedoch nicht bei dieser hohen Auflösung (Hauske 1994)) ergeben sich $2^{(1.5 \cdot 60)^2} \approx 10^{2438}$ unterscheidbare stationäre Graustufenbilder. Mit einer Zeitauflösung von etwa 100 ms erhält man ungefähr 10^{24380} unterscheidbare Filme von 1 s Dauer.

²⁴Multielektroden-Ableitungen sind elektrophysiologische Experimente mit mehreren Elektroden gleichzeitig, z. B. einer Anordnung von 7×7 Elektroden in einem Gitter von $2 \times 2 \text{ mm}^2$ auf dem Cortex eines Affen (deCharms et al. 1999).

sen sich die Aktionspotentiale mehrerer Neuronen gleichzeitig beobachten. Hierbei ist es dann nicht mehr möglich, iterativ nach einem Stimulus zu suchen, der jedes einzelne Neuron maximal erregt, denn dann wäre der Zeitvorteil, den man durch die Multielektroden-Ableitungen erhält, wieder verloren. Natürlich kann man die Frequenz eines Reintons systematisch variieren und erhält so für jedes Neuron diejenige Frequenz, bei der es mit maximaler Rate feuert, man tastet so jedoch nur einen relativ kleinen Stimulusraum ab. Die Situation verschärft sich, wenn die abgeleiteten Neuronen stark unterschiedliche Antwortigenschaften haben, somit Stimuli unterschiedlicher Art, z. B. *sowohl* Reintöne verschiedener Frequenz *als auch* Rauschen unterschiedlicher Bandbreite, getestet werden müssen. Es wäre also wünschenswert, könnte man möglichst „generische“ Stimuli, z. B. aus der natürlichen Schallumgebung eines Tieres, verwenden, die jedes abgeleitete Neuron mehr oder weniger erregen, und aus den Antworten der Neurone ein Modell ihrer Antwortigenschaften berechnen. Dieser Weg wird hier vorgeschlagen. Verfahren, welche es erlauben derartige „generischen“ Stimuli zu verwenden, werden untersucht. Eine spezielle Form physiologischer Experimente sind die sog. „bildgebenden“ Verfahren der Positronenemissionstomographie (PET)²⁵, der funktionellen Kernspinresonanz („functional magnetic resonance imaging“ FMRI)²⁶, des

²⁵Hierbei wird Wasser mit dem Sauerstoffisotop ^{15}O in den Blutkreislauf injiziert. Das Wasser sammelt sich im Gehirn, und zwar proportional zur Stärke der Durchblutung der jeweiligen Stelle. Das Sauerstoffisotop ^{15}O zerfällt unter Emission eines Positrons, das sofort mit einem Elektron annihiliert und zwei Photonen in entgegengesetzten Richtungen aussendet. Ringförmig um den Kopf der Versuchsperson angeordnete Detektoren zählen nur solche Photonen, die genau gleichzeitig ankommen, denn genau diese stammen aus der Positronenemission. Es ist möglich, während etwa einer Minute Daten zu sammeln. Aus diesen lässt sich dann auf die Konzentration von ^{15}O in verschiedenen Regionen des Gehirns, somit auf die Durchblutung und somit wiederum auf die Aktivität schließen (Posner und Raichle 1996).

²⁶Bei der bildgebenden magnetischen Kernspinresonanz (MRI) wird die Konzentration von Protonen sowie deren Relaxationszeiten im magnetischen Feld gemessen (longitudinale Relaxationszeit T_1 und transversale Relaxationszeit T_2). Dies geschieht mit einer Ortsauflösung von etwa 1 mm. Die transversale Relaxationszeit fällt mit der lokalen Inhomogenität des magnetischen Felds ab. Sauerstofftragendes Hämoglobin ist diamagnetisch, während Desoxi-Hämoglobin paramagnetisch ist. Je größer der Anteil an Desoxi-Hämoglobin an einer Stelle, desto inhomogener ist das lokale Magnetfeld. Je höher die Aktivität einer Hirnregion, desto *geringer* der Anteil an Desoxi-Hämoglobin. Beim FMRI wird die T_2 -Relaxationszeit mit Stimulus mit derjenigen ohne Stimulus (oder mit einem anderen Stimulus) verglichen, um die durch den Stimulus verursachten Aktivitätsunterschiede zu messen (Robb 1995, Andrä 1998).

„optical imaging“²⁷ und der spannungssensitiven Farbstoffe²⁸. Sie ermöglichen eine gleichzeitige Messung der Aktivität von vielen verschiedenen Neuronen. Dies ist allerdings nur mit begrenzter räumlicher und zeitlicher Auflösung möglich, d. h. man erhält ein Signal, das nur die räumlich und zeitlich über viele Neuronen *gemittelte* Aktivität wiedergibt²⁹. Die Signale bildgebender Verfahren können mit den hier vorgestellten Methoden behandelt werden, indem ein „Punkt“ des Bildes über den Zeitverlauf als Aktivität eines Neurons aufgefasst wird.

Verhaltensexperimente können entscheidende Hinweise darauf liefern, auf welche Art von Stimuli Neuronen reagieren *könnten*. Beispielsweise ist bekannt, dass der Bereich der vom Menschen gehörten Reintöne etwa von 16 Hz bis 16 kHz reicht (Zwicker und Fastl 1999). Die Vermutung liegt also nahe, dass es keine Neurone gibt, die auf Reintöne außerhalb des angegebenen Bereichs reagieren.³⁰ Daher lässt man Reintöne außerhalb des Hörbereichs der jeweiligen Spezies bei neurophysiologischen Experimenten weg (Nieder und Klump 1999). Aus Verhaltensexperimenten kann man nicht nur entnehmen, auf welche Stimuli Neurone vermutlich *nicht* reagieren, sondern auch, auf welche Stimuli sie vermutlich *besonders* ansprechen. Beispielsweise ist aus Be-

²⁷Beim „optical imaging of intrinsic signals“ wird der freigelegte Cortex mit rotem Licht beleuchtet. Sauerstoffarmes Hämoglobin absorbiert das Licht stärker und erscheint somit dunkler als sauerstoffreiches Hämoglobin. Die auftretenden Helligkeitsunterschiede werden als Aktivitätsunterschiede bei der Wahrnehmung verschiedener Stimuli gedeutet (Bonhoeffer und Grinvald 1996).

²⁸Ein Farbstoff wird auf den Cortex aufgebracht. Dieser ändert sein Absorptionsverhalten von Licht in Abhängigkeit des Membranpotentials der Neuronen. Einen Überblick gibt Shoham et al. (1999).

²⁹Mit den spannungssensitiven Farbstoffen können auch Einzelneuronen in zeitlicher Auflösung von Millisekunden beobachtet werden. Dies allerdings nur *in vitro* oder allenfalls an optisch zugänglichen Teilen des Gehirns, also im Wesentlichen der Cortex-Oberfläche.

³⁰Weiß man, dass die Unhörbarkeit von Reintönen außerhalb des angegebenen Frequenzbereichs ihre Ursache darin hat, dass die Basilarmembran von diesen Reintönen nicht mehr zu Schwingungen angeregt wird, derartige Schalle also garnicht an das zentrale Nervensystem weitergegeben werden, so weiß man *sicher*, dass es keine Neuronen geben kann, die auf diese Schalle reagieren. Bei anderen Schallen ist dies jedoch nicht ohne Weiteres klar. Beispielsweise werden Lücken von etwa 5 ms Dauer in weißem Rauschen nicht gehört (Musiek et al. 2005). Hieraus lässt sich nicht schließen, dass es keine Neurone gibt, die in ihrer Aktivität die Lücke repräsentieren. Es könnte vielmehr sein, dass Neurone zwar die Lücke repräsentieren, diese Information jedoch nicht zur Steuerung des Verhaltens der Versuchsperson oder des Versuchstiers verwendet wird. Umgekehrt lässt sich daraus, dass man kein Einzelneuron findet, das einen Schall hinreichend von einem anderen trennt, nicht schließen, dass die jeweiligen Schalle für das Versuchstier nicht zu unterscheiden sind, denn es könnten *viele* Neurone an der Repräsentation des Schalls beteiligt sein. Was daher die Diskriminierbarkeit zweier Schalle durch ein Einzelneuron (Nieder und Klump 2001) für eine Bedeutung haben soll, bleibt rätselhaft.

obachtungen in freier Natur und aus Verhaltensexperimenten bekannt, dass die Schleiereule ein ausgezeichnetes Richtungshören besitzt. Tatsächlich gibt es im *Nucleus laminaris* der Schleiereule auf verschiedene Schallrichtungen spezialisierte Neurone (Leibold et al. 2001). Es ist also sinnvoll, binaurale Schalle mit ausgeprägter Schallrichtung³¹ zu verwenden, wenn Neurone im *Nucleus laminaris* der Schleiereule untersucht werden. Wenn man weiß, dass der Krallenfrosch *Xenopus* sich in die Richtung des Ursprungs einer Wasserwelle dreht (Elepfandt 1996), so liegt es nahe, nach Neuronen zu suchen, die spezifisch auf Wasserwellen aus bestimmten Richtungen antworten. Diese hat man im *Tectum opticum* des Krallenfrosches auch gefunden (Zittlau et al. 1986). Diese Überlegungen führen dazu, möglichst *natürliche* Stimuli zur Reizung von Neuronen zu verwenden, auch wenn deren mögliche Funktion im Vorhinein nicht bekannt ist. Man kann sich also von Vorneherein auf eine gewisse Auswahl von einigermaßen natürlichen Schallen beschränken.³² Methoden, welche es nicht erlauben, die Stimuli relativ frei zu wählen, sind also solchen Methoden, die im Prinzip die Auswertung der neuronalen Antworten mit beliebigen Stimuli erlauben, weit unterlegen. Diese Arbeit beschränkt sich deshalb auf Methoden, die ein möglichst breites Spektrum an Stimuli erlauben.

1.4 Beschreibender und modellbasierter Ansatz

Um das Verhalten von Neuronen zu verstehen, gibt es grundsätzlich zwei Ansätze, nämlich den *beschreibenden* und den *modellbasierten* Ansatz. Der beschreibende Ansatz versucht, das Antwortverhalten von Neuronen zu beschreiben, d. h. er klärt die Frage, bei welchen Stimuli das Neuron wann innerhalb des Stimulus ein Aktionspotential erzeugt. Z.B. könnte ein Neuron nur auf Reintöne der Frequenz 100 Hz reagieren und dann nur am Schalldruckmaximum ein Aktionspotential erzeugen. Dies wäre eine vollständige Beschreibung des Antwortverhaltens.

³¹Siehe Blauert (1999) zur Frage, welche Schalle beim Menschen zu einer mehr oder weniger ausgeprägten Richtungswahrnehmung führen.

³²Z.B. ist es u.U. nicht sinnvoll, weißes Rauschen mit gleicher Intensität unabhängig auf beide Ohren zu geben und zu erwarten, dass Neurone hierauf reagieren, denn natürliche Schalle gehen immer von einer Schallquelle aus, somit sind die Schalle auf den beiden Ohren korreliert. Außerdem haben die meisten natürlichen Schalle spektrale Eigenschaften, bestehen zum Beispiel aus Klängen, d.i. Töne mit den entsprechenden Oberschwingungen (Zwicker und Fastl 1999). Weißes Rauschen benötigt man jedoch für die „Reverse Correlation“ Methode (vgl. Abschnitt 1.6.1), was die Nützlichkeit derselben aus den angesprochenen Gründen stark einschränkt.

Der modellbasierte Ansatz versucht nicht nur zu beschreiben *wie* ein Neuron reagiert, sondern auch *warum*. Z.B. könnte das oben beschriebene Neuron sich so verhalten, weil es von einer Hörnervenfaser Eingaben bekommt, welche auf die Frequenz 100 Hz abgestimmt ist, und seine Schwelle so eingestellt ist, dass es nur am Amplitudenmaximum feuert. Ist man nur an einer Beschreibung des Neurons interessiert, dann lässt sich im Prinzip jedes Verfahren einsetzen, das eine Vorhersage des Antwortverhaltens erlaubt. Beispielsweise könnte ein Neuron im auditorischen System durch ein Backpropagation-Netzwerk (Rojas 1996, Zell 2000) beschrieben werden, welches als Eingabe die gerade vergangenen Schalldrücke zu Zeitpunkten im Abstand von $1/44.1$ ms erhält³³, vgl. Abbildung 1.8. Aus zwei Gründen würde man hieraus jedoch wenig über das auditorische System von Menschen und Tieren lernen. Erstens kommen Schalldrücke im auditorischen System niemals als Eingabe eines Neurons vor, denn bereits in der Cochlea wird eine Zerlegung des Schalls nach Frequenzbereichen durchgeführt. d. h. das auditorische System *muss* mit der Eingabe aus dem Hörnerv weiterarbeiten. Das Backpropagation-Netzwerk würde jedoch Schalldrücke verarbeiten. Das „echte“ auditorische System muss also ganz anders funktionieren als das konstruierte Backpropagation-Netzwerk. Zweitens lässt sich anhand der Parameter eines Backpropagation-Netzwerks oft nur schwer verstehen, *warum* es funktioniert. Man hätte also nur ein Problem, nämlich wie ein bestimmtes Neuron im auditorischen System funktioniert, durch ein anderes ersetzt, nämlich wie ein konstruiertes Backpropagation-Netzwerk funktioniert. Der modellbasierte Ansatz hat daher gegenüber dem rein beschreibenden den Vorteil, dass man anhand eines Modells, das so oder so ähnlich wie das echte System funktioniert, das untersuchte System wirklich *verstehen* kann. Deshalb wird in dieser Arbeit hauptsächlich der modellbasierte Ansatz verfolgt. Darüber hinaus kann man beim modellbasierten Ansatz bekannte Eigenschaften von Neuronen ausnutzen, um die Effizienz der eingesetzten Verfahren zu erhöhen. Diese Eigenschaften von (Modell-) Neuronen werden im nächsten Abschnitt besprochen.

1.5 Neuronenmodelle

Neuronen werden hier als „Spike-Response-Neuronen“ (Gerstner 1992, Gerstner und van Hemmen 1994) behandelt.³⁴ Seien t_{ij}^{syn} , $i = 1 \dots N$ und

³³„Digital Audio CDs“ haben eine Abtastrate von 44.1 kHz.

³⁴In Gerstner (1992) werden die Eigenschaften des Spike-Response-Modells mit dem Neuronenmodell von Hodgkin und Huxley 1952 (Bower und Beeman 1994) verglichen. Das Spike-Response-Modell stellt eine gute Näherung dieses aus vier gekoppelten nicht linearen Differentialgleichungen bestehenden Modells dar (Kistler et al. 1997). Es eignet

$j = 1 \dots N_i$, die Zeitpunkte, zu denen an Synapse Nummer i ein Aktionspotential ankommt. Sei weiter $\varepsilon_i : \mathbb{R} \rightarrow \mathbb{R}$ das postsynaptische Potential³⁵, welches durch ein Aktionspotential an Synapse i im betrachteten Neuron erzeugt wird. Das synaptische Potential $h_{\text{syn}}(t)$ eines Neurons zum Zeitpunkt t ist dann

$$h_{\text{syn}}(t) = \sum_{i=1}^N \sum_{j=1}^{N_i} \varepsilon_i(t - t_{ij}^{\text{syn}}). \quad (1.1)$$

Um zu berücksichtigen, dass ein Neuron nicht sogleich wieder ein Aktionspotential erzeugt, wenn es gerade gefeuert hat³⁶, wird ein Refraktärpotential η eingeführt.³⁷ Sind t_f die Zeitpunkte, zu denen das Neuron ein Aktionspotential erzeugt hat, dann ergibt sich das effektive Membranpotential h des Neurons als

$$h(t) = h_{\text{syn}}(t) + \sum_f \eta(t - t_f). \quad (1.2)$$

Da der Betrag des Refraktärpotentials η schnell abfällt, bestimmt im Wesentlichen der *letzte* Zeitpunkt, zu dem das Neuron ein Aktionspotential erzeugt hat, wie groß das gesamte Refraktärpotential $\sum_f \eta(t - t_f)$ ist.

Falls ϑ die „Schwelle“ des Neurons ist, erzeugt das Neuron genau dann ein

sich jedoch wegen seiner einfachen Berechenbarkeit auch zur Simulation größerer Netze. Im Spike-Response-Modell nicht enthalten ist eine Synapsen-Dynamik. Synapsen können ihre Effektivität in Abhängigkeit ihrer „Beanspruchung“ verändern. Obwohl zuweilen als ausschlaggebend betrachtet (Tsodyks et al. 1998, Fuhrmann et al. 2002), wird die Dynamik von Synapsen hier nicht berücksichtigt. Vielmehr wird davon ausgegangen, dass an der Synapse ankommende Aktionspotentiale immer die gleiche Wirkung beim postsynaptischen Neuron hervorrufen. Es ist auch möglich, dass Synapsen stochastisch arbeiten, also z. B. bei einem ankommenden Aktionspotential mehr oder weniger Neurotransmitter ausschütten, was hier ebenfalls vernachlässigt wird.

³⁵Das postsynaptische Potential (PSP) kann exzitatorisch (EPSP), also $\varepsilon > 0$, oder inhibitorisch (IPSP), also $\varepsilon < 0$, sein.

³⁶Dies bezeichnet man als „Refraktärverhalten“. Man unterscheidet die *absolute* Refraktärzeit nach einem Aktionspotential, während der das Neuron unabhängig von jeder vernünftigen Eingabe kein weiteres Aktionspotential mehr erzeugen kann, und die *relative* Refraktärzeit, während der das Neuron erst bei größerem synaptischen Potential h_{syn} als gewöhnlich ein Aktionspotential erzeugt (Kandel et al. 1996).

³⁷Gewöhnlich ist $\eta < 0$, es gibt jedoch auch Fälle, z. B. bei „burstenden“ Neuronen, bei denen es Zeitpunkte t mit $\eta(t) > 0$ gibt, so dass ein Neuron eine gewisse Zeit nach einem Aktionspotential leichter wieder zum Feuern gebracht werden kann (Gerstner 1992).

Aktionspotential zum Zeitpunkt t , wenn³⁸

$$h(t) > \vartheta. \quad (1.3)$$

Als postsynaptisches Potential ε_i kann in Simulationen eine sog. „alpha-Funktion“³⁹ verwendet werden (Gerstner 1992), siehe Abbildung 1.9.

$$\varepsilon_i(t) = \begin{cases} 0 & \text{falls } t \leq 0 \\ J_i t e^{-t/\tau_s} & \text{sonst} \end{cases} \quad (1.4)$$

Hierbei bezeichnet τ_s die Membranzeitkonstante und J_i die „Stärke“ von Synapse i . Als Refraktärpotential kann nach Gerstner (1992) die Funktion

$$\eta(t) = \begin{cases} 0 & \text{falls } t \leq 0 \\ -\infty & \text{falls } 0 < t \leq \gamma_{\text{ref}} \\ -\frac{\tau_{\text{ref}}}{t - \gamma_{\text{ref}}} & \text{falls } t > \gamma_{\text{ref}} \end{cases} \quad (1.5)$$

gewählt werden, siehe Abbildung 1.9. Hierbei bezeichnet γ_{ref} die absolute Refraktärzeit und τ_{ref} die Zeitkonstante des Refraktärpotentials.

Das in (1.4) und (1.5) angegebene postsynaptische Potential ε_i bzw. Refraktärpotential η werden in dieser Arbeit jedoch *nicht* weiter verwendet. Vielmehr werden Neuronen mit *beliebigen* postsynaptischen Potentialen und Refraktärpotentialen angenommen. Es sollen ja gerade Messvorschriften entwickelt werden, um die Verbindungen zwischen Neuronen zu messen. Da man jedoch nur die potentiellen Eingabe-Aktionspotentiale eines Neurons und dessen Ausgabe-Aktionspotentiale messen kann⁴⁰, kann man nicht herausfinden,

³⁸Es handelt sich also um ein deterministisch feuerndes Neuron. Nach Gerstner (1992) 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, $P(h(t)) = e^{\beta(h(t) - \vartheta)} dt$ gilt. Die Größe β^{-1} bestimmt dabei die Stärke des Rauschens und heißt „formale Temperatur“ des Systems.

³⁹Sie ergibt sich als Lösung der Differentialgleichung $\dot{h}(t) = -\frac{1}{\tau} h(t) + j(t)$ mit $j(t) = J_i e^{-t/\tau_{\text{syn}}}$ als Inhomogenität im Spezialfall $\tau = \tau_{\text{syn}}$ (Bower und Beeman 1994). Diese Differentialgleichung beschreibt einen Kondensator mit Leckstrom, in den ein exponentiell abfallender Strom j injiziert wird. Eine Synapse injiziert in ein Neuron nach einem ankommenden Aktionspotential ebenfalls einen mit der Zeit abfallenden Strom und ein Neuron verhält sich bei wenig vom Ruhepotential entfernten Membranpotentialen wie ein Kondensator mit Leckstrom. Siehe Tuckwell (1988) für genaue Ausführungen. Warum allerdings die Zeitkonstante τ des Neuronen-Kondensators und diejenige des synaptischen Stroms τ_{syn} gleich sein sollen, ist für den Autor nicht nachvollziehbar.

⁴⁰In einem Experiment *in vitro* wäre es allerdings schon denkbar, in einem Eingangs-Axon eines Neurons künstlich ein Aktionspotential hervorzurufen und dann das postsynaptische Potential des Neurons zu messen. In einem lebenden Tier dürfte dies jedoch derzeit unmöglich oder äußerst schwierig und zeitaufwändig sein. Wir beschränken uns hier auf Experimente, die am lebenden Tier routinemäßig durchgeführt werden können.

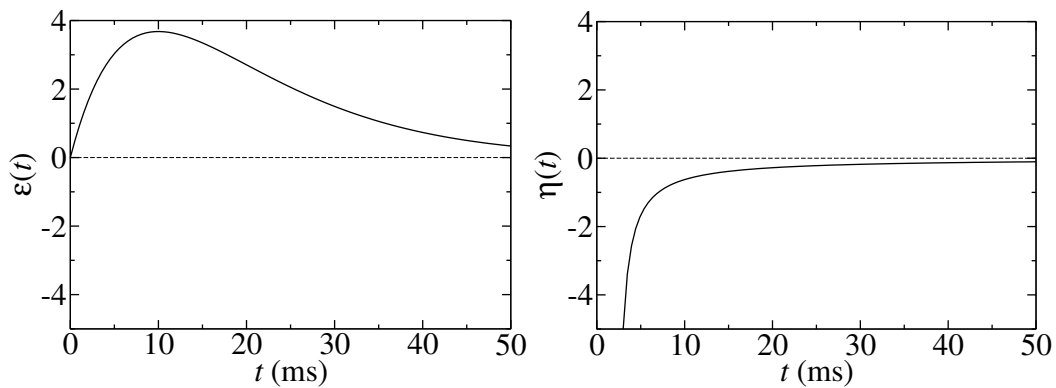


Abbildung 1.9:

Links: Postsynaptisches Potential ε aus (1.4). Die Membranzeitkonstante τ_s beträgt 10 ms, die synaptische Stärke $J = 1$. Das postsynaptische Potential steigt zunächst linear an, erreicht bei $t = \tau_s$ sein Maximum und fällt dann exponentiell ab. Es ist $\varepsilon \geq 0$, das abgebildete postsynaptische Potential ist also exzitatorisch.

Rechts: Refraktärpotential η aus (1.5). Die absolute Refraktärzeit γ_{ref} beträgt 2 ms und die Zeitkonstante der relativen Refraktärzeit $\tau_{\text{ref}} = 5$ ms. Das Refraktärpotential ist $-\infty$ für $t \leq \gamma_{\text{ref}}$ und erreicht bei etwa 20 ms fast wieder den Wert null.

ob ein Neuron mit einem anderen über nur eine oder über mehrere Synapsen verbunden ist, denn ein besonders hohes exzitatorisches postsynaptisches Potential könnte ja von einer starken oder von zwei schwächeren Synapsen stammen.⁴¹ Es wird sich herausstellen, dass das, was man messen kann, das „rezeptive Feld“ eines Neurons ist. Um diesen Begriff zu entwickeln, müssen wir einen Zusammenhang zwischen Eingabe- und Ausgabe-Aktionspotentialen aufstellen. Dazu schreiben wir mit den Zeiten t_{nk} , zu denen das Neuron n ein Aktionspotential erzeugt, die Ausgabe-Aktionspotentiale s_n eines Neurons n als Summe von δ -Pulsen.⁴²

$$s_n(t) = \sum_k \delta(t - t_{nk}) \quad (1.6)$$

Seien Δt_{ni} die Zeiten, die ein vom Eingabe-Neuron n erzeugtes Aktionspotential braucht, um zur Synapse i des Ausgabe-Neurons zu gelangen, dann können wir $h_{\text{syn}}(t)$ aus (1.1) schreiben als

$$\begin{aligned} h_{\text{syn}}(t) &= \sum_{i=1}^N \sum_{j=1}^{N_i} \varepsilon_i(t - t_{ij}^{\text{syn}}) = \\ &= \sum_{i=1}^N \sum_{n,k} \varepsilon_i(t - \Delta t_{ni} - t_{nk}) \end{aligned}$$

Mit Hilfe der Funktion

$$r_n(t) = \sum_{i=1}^N \varepsilon_i(t - \Delta t_{ni}) \quad (1.7)$$

und dem „Spiketrain“ $s_n(t)$ aus (1.6) ergibt sich

$$h_{\text{syn}}(t) = \sum_n (r_n \star s_n)(t).$$

Hierbei bezeichnet der Stern „ \star “ die Faltung zweier Funktionen, siehe Abschnitt A.2. Wir bezeichnen das Tupel (r_1, \dots) aus (1.7) als das „rezeptive Feld“ des Neurons.

⁴¹Wenn man freilich nach einem Aktionspotential eines Eingabeneurons ein postsynaptisches Potential mit zwei klar getrennten Maxima sehen kann, dann deutet dies auf mindestens zwei Synapsen hin.

⁴²Die verallgemeinerte Funktion δ ist die Dirac'sche δ -Funktion mit der Eigenschaft $\int_{-\infty}^{\infty} \delta(\tau)x(t - \tau) d\tau = x(t)$, vgl. Abschnitt B.6.

Sind t_j die Zeitpunkte, zu denen unser Neuron feuert, und schreiben wir den Spiketrain dieses Neurons als

$$s(t) = \sum_j \delta(t - t_j),$$

so lässt sich das effektive Membranpotential h aus (1.2) ausdrücken durch

$$h(t) = \sum_n (r_n \star s_n)(t) + (\eta \star s)(t). \quad (1.8)$$

Kennen wir das rezeptive Feld (r_1, \dots) und das Refraktärpotential η eines Neurons, so können wir aus den Eingabe-Spiketrains s_n und aus dem bisher berechneten Spiketrain s des Neurons selbst nach (1.8) das effektive Membranpotential h des Neurons berechnen. Aus (1.3) ergeben sich dann, bei bekannter Schwelle ϑ , die Zeitpunkte, zu denen das Neuron Aktionspotentiale erzeugt. Die Kenntnis des rezeptiven Feldes ist also zusammen mit dem Refraktärpotential und der Schwelle hinreichend, um das Antwortverhalten eines Neurons zu charakterisieren. Sind also die rezeptiven Felder r_n^1 und r_n^2 von zwei Neuronen gleich, d. h. gilt

$$\begin{aligned} r_n^1(t) &= r_n^2(t) \\ \sum_{i=1}^{N^1} \varepsilon_i^1(t - \Delta t_{ni}^1) &= \sum_{i=1}^{N^2} \varepsilon_i^2(t - \Delta t_{ni}^2) \end{aligned}$$

nach (1.7), und ist auch $\eta^1 = \eta^2$ sowie $\vartheta^1 = \vartheta^2$, so ist auch das Antwortverhalten der beiden Neurone gleich, d. h. die Neurone erzeugen bei gleicher Eingabe Aktionspotentiale zu gleichen Zeitpunkten. d. h. Neurone mit gleichen rezeptiven Feldern aber unterschiedlichen postsynaptischen Potentialen ε_i und unterschiedlichen Verzögerungszeiten Δt_{ni} können durch die oben definierte Messung der Aktionspotentiale der Eingabe-Neuronen und der Ausgabe-Aktionspotentiale nicht unterschieden werden.⁴³ Unser Ziel kann es also nur sein, Messverfahren für das rezeptive Feld (r_1, \dots) eines Neurons zu entwickeln und eventuell dessen Refraktärpotential η und Schwelle ϑ , nicht jedoch für einzelne postsynaptische Potentiale ε_i (Form und absolute Stärke) und Verzögerungszeiten Δt_{nj} . Allerdings reicht das rezeptive Feld eines Neurons in oben definiertem Sinne sowohl aus, um sein Antwortverhalten zu

⁴³Zu einer Unterscheidung wären dann neben elektrophysiologischen Experimenten am lebenden Tier noch zusätzliche Experimente notwendig, etwa anatomische Untersuchungen oder elektrophysiologische Experimente *in vitro*, in denen durch Reizung von Eingabe-Axonen theoretisch die Wirkung unterschiedlicher Synapsen getrennt werden könnte. Aufgrund des unverhältnismäßigen Aufwands für derartige Experimente werden sie hier nicht weiter berücksichtigt.

beschreiben, als auch um es zu *erklären*, d. h. also ein funktionales Modell des Neurons zu entwickeln, das von weiteren (Modell-) Neuronen Eingaben erhält und sich so verhält wie das echte Neuron.⁴⁴

1.6 Vorarbeiten

1.6.1 Reverse Correlation

„Reverse Correlation“ ist eine Methode, die Impulsantwort eines zeitinvarianten linearen Systems (vgl. Abschnitt A.2) herauszufinden. Dazu wird die Systemantwort y bei Gauß'schem weißem Rauschen σn mit Varianz σ^2 als Eingabe gemessen. Es ist dann nach (A.5)

$$y = r \star \sigma n. \quad (1.9)$$

Eine Schätzung \hat{r} für die Impulsantwort r des Systems erhält man, indem man die Eingabe σn mit der Ausgabe y korreliert:

$$\hat{r} = \frac{1}{\sigma} \text{cov}(n, y)$$

Für den Erwartungswert von $\langle \hat{r} \rangle$ von $\hat{r}(t)$ gilt unter Verwendung des gemischten Assoziativgesetzes (A.9)

$$\begin{aligned} \langle \hat{r} \rangle &= \langle \text{cov}(n, r \star n) \rangle = \\ &= \langle \text{cov}(n, n \star r) \rangle = \\ &= \langle \text{cov}(n, n) \star r \rangle = \\ &= \langle \text{cov}(n, n) \rangle \star r = \\ &= \delta \star r = \\ &= r. \end{aligned}$$

⁴⁴Wie wir später sehen werden, können auch potentielle Eingabe-Neuronen n identifiziert werden, die tatsächlich *nicht* mit dem untersuchten Neuron verbunden sind, für die also $r_n = 0$ gilt. So lassen sich eventuell anatomische Untersuchungen ersetzen oder umgekehrt anregen, wenn man eine bisher unbekannte Verbindung feststellt. Weiter könnte man anhand des rezeptiven Feldes eventuell die wahrscheinliche *Funktion* eines Neurons erschließen. Ist beispielsweise $(r_1, r_2) = (t \mapsto \delta(t), t \mapsto \delta(t - \Delta t))$, erzeugen also Aktionspotentiale der Eingabeneuronen 1 und 2 peak-förmige postsynaptische Potentiale, wobei dasjenige von Eingabeneuron 2 um Δt verzögert ankommt, könnte man vermuten, dass das Neuron mit dem rezeptiven Feld (r_1, r_2) als Koinzidenzdetektor dient, da es bei geeigneter Schwelle nur dann ein Aktionspotential erzeugt, wenn Eingabeneuron 1 um Δt später ein Aktionspotential erzeugt als Eingabeneuron 2. Für die Bedingungen, unter denen ein Neuron als Koinzidenzdetektor fungieren kann und wie dies im auditorischen System vermutlich zur Schalllokalisation beiträgt, siehe Kempster et al. (1998a,b) und Kempster (1997).

Der Schätzer \hat{r} ist also erwartungstreu. Er ist unter den Voraussetzungen von Abschnitt 4.4 auch derjenige mit der kleinsten Varianz. In der Praxis ist die Eingabe allerdings niemals Gauß'sches weißes Rauschen, denn dies wäre unendlich lang. Die Endlichkeit der Eingabe bewirkt, dass der Schätzwert \hat{r} zwar erwartungstreu ist, allerdings nicht mehr die kleinste Varianz hat. Ein weiterer Nachteil der Reverse-Correlation-Methode ist, dass die Annahme eines linearen Systems gemacht werden muss, andernfalls handelt es sich bei \hat{r} nurmehr um den optimalen *linearen* Schätzer, dies allerdings auch nur bei unendlicher Länge der Eingabe, vgl. Abschnitt 4.1. Der größte Nachteil ist allerdings, dass Gauß'sches weißes Rauschen als Eingabe verwendet werden muss. Verwendet man im auditorischen System Gauß'sches weißes Rauschen als Schall, so ist wegen der nichtlinearen Vorverarbeitung in der Cochlea die lineare Annahme 1.9 keinesfalls gerechtfertigt. Es ist unmöglich, Gauß'sches weißes Rauschen auf der Cochlea zu erzeugen, vgl. Fußnote 13 auf Seite 32. Eine einfache Reverse-Correlation-Analyse scheidet somit aus den genannten Gründen für das auditorische System praktisch aus, vgl. auch Aertsen und I. (1981a).

1.6.2 Reverse Correlation mit Pieptönen

Um die Reverse-Correlation-Methode in einfacher Weise doch für das auditorische System verwenden zu können, macht deCharms et al. (1998) einen Ansatz mit zufälligen Pieptönen: Töne der Dauer 20 ms (mit cosinusförmigen Rampen der Dauer 5 ms) mit zufälligen Frequenzen werden zu zufälligen Zeitpunkten gespielt, insgesamt 10 min lang. Dabei können auch mehrere Töne zur gleichen Zeit vorhanden sein. Seien ω_i die (diskreten) Frequenzen der Töne. Ein künstliches „Kurzzeitspektrum“ $\Xi(\omega_i, t)$ wird definiert als

$$\Xi(\omega_i, t) = \text{Zahl der zum Zeitpunkt } t \text{ aktiven Töne der Frequenz } \omega_i .$$

Seien t_j die Zeitpunkte, zu denen das untersuchte Neuron ein Aktionspotential erzeugt hat. Dann wird der Ausgabe-Spiketrain definiert als

$$y(t) = \sum_j \delta(t - t_j) .$$

In deCharms et al. (1998) wird das spektro-temporale rezeptive Feld $r(\omega_i, t)$ gemessen durch

$$r(\omega_i, t) = \int_{-\infty}^{\infty} \Xi(\omega_i, \tau) y(\tau + t) dt ,$$

also durch eine Korrelation des künstlichen Kurzzeitspektrums $\Xi(\omega_i, t)$ mit dem Ausgabe-Spiketrain y . Die Schwierigkeiten, die dadurch entstehen, dass

$\Xi(\omega_i, t)$ kein weißes Rauschen ist, sondern allenfalls Poisson-Rauschen bei Poisson-verteilten Anfangszeitpunkten der Töne, wird in deCharms et al. (1998) nicht diskutiert. Eine Bedingung für ein „vernünftiges“ Ergebnis $r(\omega_i, t)$ ist wohl, dass die Töne hinreichend kurz und zufällig sind. Dies schränkt die möglichen Stimuli stark ein. Ein Einfluss des gewählten Stimulus, insbesondere der Dauer der Pieptöne, auf das Ergebnis $r(\omega_i, t)$ ist selbst bei linearen Neuronen im Sinne von

$$y(t) = \sum_i \int_{-\infty}^{\infty} r(\omega_i, \tau) \Xi(\omega_i, t - \tau) d\tau$$

gegeben.

1.6.3 Wiener-Entwicklung

Die Wiener-Entwicklung ist eine nichtlineare Erweiterung der Reverse-Correlation-Methode.

Die Volterra-Entwicklung eines nichtlinearen Systems mit Antwort y und Eingabe x lautet (Unbehauen 1993)

$$\begin{aligned} y(t) = & v_0 + \int_{-\infty}^{\infty} v_1(\tau) x(t - \tau) d\tau + \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_2(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) d\tau_1 d\tau_2 + \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_3(\tau_1, \tau_2, \tau_3) x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) d\tau_1 d\tau_2 d\tau_3 + \\ & + \dots \end{aligned} \quad (1.10)$$

Die i -stelligen Funktionen v_i werden als „Volterra-Kerne“ i -ter Ordnung bezeichnet. Im Prinzip können die Volterra-Kerne durch Messung der Antwort des Systems auf Impuls-Kombinationen gemessen werden (van Hemmen et al. 2000). Die Volterra-Entwicklung selbst ist allerdings nur für Eingaben x mit kleiner Auslenkung geeignet, da sie praktisch eine Taylor-Entwicklung um den Nullpunkt darstellt, während die Wiener-Entwicklung für Signale jeden Pegels geeignet ist (Eggermont 1993). Bei Gauß'schem weißem Rauschen als Eingabe können die Volterra-Kerne nicht auf einfache Weise ermittelt werden, die Wiener-Kerne w_i ergeben sich in diesem Fall jedoch durch eine einfache Kovarianz zwischen Eingabe σn und Ausgabe y . Hierzu werden die

„Wiener’schen G-Funktionale“ G_i eingeführt. Ist die Eingabe des Systems Gauß’sches weißes Rauschen σn mit Varianz σ^2 , dann lauten sie bis zur dritten Ordnung (Schetzen 1980)

$$\begin{aligned}
 G_0(w_0, x)(t) &:= w_0 \\
 G_1(w_1, x)(t) &:= \int_{-\infty}^{\infty} w_1(\tau) x(t - \tau) d\tau \\
 G_2(w_2, x)(t) &:= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w_2(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) d\tau_1 d\tau_2 - \sigma^2 \int_{-\infty}^{\infty} w_2(\tau, \tau) d\tau \\
 G_3(w_3, x)(t) &:= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w_3(\tau_1, \tau_2, \tau_3) x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) d\tau_1 d\tau_2 d\tau_3 \\
 &\quad - 3\sigma^2 \int_{-\infty}^{\infty} x(t - \tau) w_3(\tau, \tau, \tau) d\tau
 \end{aligned}$$

Die Wiener-Entwicklung eines Systems mit Ausgabe y und Eingabe x lautet

$$y = \sum_{i=0}^{\infty} G_i(w_i, x). \quad (1.11)$$

Die i -stelligen Funktionen w_i heißen „Wiener-Kerne“. Sie werden so bestimmt, dass bei Gauß’schem weißem Rauschen σn mit Varianz σ^2 als Eingabe der Erwartungswert der mittleren quadratischen Abweichung zwischen der „Wiener-Approximation der Ordnung N “

$$\hat{y}_N = \sum_{i=0}^N G_i(w_i, \sigma n). \quad (1.12)$$

und der Systemantwort y für jedes $N = 0, \dots, \infty$ minimal ist. Mit weiteren Forderungen an die Symmetrie der Wiener-Kerne w_i sind diese eindeutig bestimmt (Schetzen 1980). Die Wiener’schen G-Funktionale haben den Vorteil, dass sie orthogonal bezüglich Gauß’schem weißem Rauschen σn mit Varianz σ^2 als Eingabe sind, d. h.

$$\langle \text{cov}[G_i(w_i, \sigma n), G_j(w_j, \sigma n)] \rangle = 0$$

für $i \neq j$. Dies hat zur Folge, dass sich der mittlere quadratische Fehler zwischen der Systemausgabe y aus (1.11) und der Wiener-Approximation \hat{y}_N

aus (1.12) leicht berechnen lässt, vorausgesetzt die Eingabe ist Gauß'sches weißes Rauschen mit Varianz σ^2 :

$$\begin{aligned} \left\langle \int_{-\infty}^{\infty} [y(t) - \hat{y}(t)]^2 dt \right\rangle &= \left\langle \int_{-\infty}^{\infty} \left[\sum_{i=0}^{\infty} G_i(w_i, \sigma n) - \sum_{i=0}^N G_i(w_i, \sigma n) \right]^2 dt \right\rangle = \\ &= \left\langle \int_{-\infty}^{\infty} \left[\sum_{i=N+1}^{\infty} G_i(w_i, \sigma n) \right]^2 dt \right\rangle = \\ &= \left\langle \int_{-\infty}^{\infty} \sum_{i=N+1}^{\infty} [G_i(w_i, \sigma n)]^2 dt \right\rangle \end{aligned}$$

Jede zusätzliche Ordnung macht den Erwartungswert des Fehlers also kleiner, was auch so sein muss, da die Wiener-Kerne w_1, \dots, w_N ja gerade so gewählt sind, dass \hat{y} aus (1.12) eine optimale Näherung ist. Zudem hat die Wiener-Entwicklung den Vorteil, dass sich die Wiener-Kerne sehr einfach berechnen lassen. Falls die Eingabe x Gauß'sches weißes Rauschen mit Varianz σ^2 ist, gilt nämlich (Schetzen 1980)

$$\begin{aligned} w_0 &= \langle y(t) \rangle \\ w_1(\tau) &= \frac{1}{\sigma^2} \langle y(t) x(t - \tau) \rangle \\ w_2(\tau_1, \tau_2) &= \frac{1}{2\sigma^4} \langle y(t) x(t - \tau_1) x(t - \tau_2) \rangle \\ w_3(\tau_1, \tau_2, \tau_3) &= \frac{1}{3\sigma^6} \langle y(t) x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) \rangle, \end{aligned}$$

falls die τ_i paarweise verschieden sind. Schätzungen \hat{w}_i für die Wiener-Kerne w_i erhält man, indem die Erwartungswerte durch die entsprechenden Inte-

grale ersetzt werden, also

$$\begin{aligned}\hat{w}_0 &= \frac{1}{T} \int_0^T y(t) dt \\ \hat{w}_1(\tau) &= \frac{1}{\sigma^2} \frac{1}{T} \int_0^T y(t) x(t - \tau) dt \\ \hat{w}_2(\tau_1, \tau_2) &= \frac{1}{2\sigma^4} \frac{1}{T} \int_0^T y(t) x(t - \tau_1) x(t - \tau_2) dt \\ \hat{w}_3(\tau_1, \tau_2, \tau_3) &= \frac{1}{3\sigma^6} \frac{1}{T} \int_0^T y(t) x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) dt.\end{aligned}$$

Die Einschränkung an die τ_i kann umgangen werden, indem zur Schätzung von w_1 die Ausgabe y durch $y - G_0(w_0, x)$, zur Schätzung von w_2 durch $y - G_0(w_0, x) - G_1(w_1, x)$ u.s.w. ersetzt wird.

Die Volterra-Kerne v_i aus (1.10) lassen sich aus den Wiener-Kernen w_i berechnen durch (Aertsen und I. 1981b)

$$\begin{aligned}v_n(t_1, \dots, t_n) &= \sum_{i=0}^{\infty} (-1)^i \frac{(n+2i)!}{n! i!} \left(\frac{\sigma^2}{2}\right)^i \\ &\quad \times \int \dots \int w_{n+2i}(t_1, \dots, t_n, \tau_1, \tau_1, \dots, \tau_i, \tau_i).\end{aligned}\tag{1.13}$$

Für ein System mit Wiener-Kernen höchstens dritter Ordnung, d. h. $w_i = 0$ für $i > 3$, gilt daher

$$\begin{aligned}v_1(t) &= w_1(t) - \frac{3}{2}\sigma^2 \int w_3(t, \tau, \tau) d\tau \\ v_2 &= w_2 \\ v_3 &= w_3\end{aligned}$$

Leider werden i. A. *alle* Wiener-Kerne zur Berechnung der ersten Volterra-Kerne v_1 und v_2 benötigt.

Sei $g : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto g(t)$ eine „Fenster“-Funktion (Gewichtsfunktion), z. B. $g(t) = \exp(-t/\tau_0)$ für $t \geq 0$ und 0 sonst. Eine zeitabhängige Frequenzfunktion $X : \mathbb{R}^2 \rightarrow \mathbb{C}; (\omega, t) \mapsto X(\omega, t)$ des Signals x wird definiert als (Terhardt 1998)

$$X(\omega, t) := \int_{-\infty}^{\infty} x(\tau) g(t - \tau) e^{-i\omega\tau} d\tau\tag{1.14}$$

Ein zeitabhängiges Power-Spektrum $\Xi : \mathbb{R}^2 \rightarrow \mathbb{R}; (\omega, t) \mapsto \Xi(\omega, t)$ kann dann definiert werden als⁴⁵

$$\Xi(\omega, t) = |X(\omega, t)|^2. \quad (1.15)$$

Bei einem System, das linear im zeitabhängigen Power-Spektrum Ξ ist, für dessen Systemantwort y auf das Signal x also

$$y(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(\omega, \tau) \Xi(\omega, t - \tau) d\tau d\omega \quad (1.16)$$

gilt, sind alle Volterra-Kerne $v_i = 0$ bis auf den Volterra-Kern zweiter Ordnung v_2 , denn wegen (1.15) und (1.14) gilt

$$\begin{aligned} \Xi(\omega, t) &= |X(\omega, t)|^2 = \\ &= X(\omega, t) X^*(\omega, t) = \\ &= \int_{-\infty}^{\infty} x(\tau_1) g(t - \tau_1) e^{-i\omega\tau_1} d\tau_1 \int_{-\infty}^{\infty} x(\tau_2) g(t - \tau_2) e^{i\omega\tau_2} d\tau_2 = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau_1) x(\tau_2) g(t - \tau_1) g(t - \tau_2) e^{i\omega(\tau_2 - \tau_1)} d\tau_1 d\tau_2 \end{aligned}$$

Somit ergibt sich mit $\tau'_1 = t - \tau_1$ und $\tau'_2 = t - \tau_2$ für die Systemantwort y

⁴⁵Definiert man die zeitabhängige Autokorrelationsfunktion c_{xx} des Signals x als

$$\text{cov}_{xx}(\tau, t) = \int_{-\infty}^{\infty} x(\tau') g(t - \tau') x(\tau' + \tau) g(t - \tau' - \tau) d\tau',$$

so gilt für die zeitabhängige Frequenzfunktion $X(\omega, t)$ eine zum Wiener-Khinchin-Theorem (B.8) analoge Beziehung (Terhardt 1998)

$$\mathcal{F}\text{cov}_{xx}(\omega, t) = |X(\omega, t)|^2.$$

nach (1.16)

$$\begin{aligned}
y(t) &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty x(\tau_1) x(\tau_2) g(t - \tau - \tau_1) g(t - \tau - \tau_2) \\
&\quad \times e^{i\omega(\tau_2 - \tau_1)} r(\omega, \tau) d\tau_1 d\tau_2 d\tau d\omega = \\
&= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty x(t - \tau'_1) x(t - \tau'_2) g(\tau'_1 - \tau) g(\tau'_2 - \tau) \\
&\quad \times e^{i\omega(\tau'_1 - \tau'_2)} r(\omega, \tau) d\tau'_1 d\tau'_2 d\tau d\omega = \\
&= \int_{-\infty}^\infty \int_{-\infty}^\infty x(t - \tau'_1) x(t - \tau'_2) \int_0^\infty \int_{-\infty}^\infty g(\tau'_1 - \tau) g(\tau'_2 - \tau) \\
&\quad \times e^{i\omega(\tau'_1 - \tau'_2)} r(\omega, \tau) d\tau d\omega d\tau'_1 d\tau'_2 = \\
&= \int_{-\infty}^\infty \int_{-\infty}^\infty x(t - \tau'_1) x(t - \tau'_2) v_2(\tau'_1, \tau'_2) d\tau'_1 d\tau'_2
\end{aligned}$$

mit

$$v_2(\tau_1, \tau_2) = \int_0^\infty \int_{-\infty}^\infty g(\tau_1 - \tau) g(\tau_2 - \tau) r(\omega, \tau) e^{i\omega(\tau_1 - \tau_2)} d\tau d\omega. \quad (1.17)$$

Gleichung (1.17) ist eine lineare Beziehung zwischen dem rezeptiven Feld r und dem Volterra-Kern zweiter Ordnung v_2 (Aertsen und I. 1981b). Im Prinzip können also aus der Systemantwort y mit Gauß'schem weißem Rauschen als Eingabe x die Wiener-Kerne w_i geschätzt werden, woraus sich nach (1.13) Schätzungen für die Volterra-Kerne ergeben. Aus diesen ließe sich durch Umkehrung von (1.17) ein rezeptives Feld r berechnen. Oder man ersetzt in (1.17) den Volterra-Kern zweiter Ordnung v_2 durch den Wiener-Kern w_2 und erhält ein rezeptives Feld r , das die geschätzte Systemantwort

$$\hat{y}(t) = \int_{-\infty}^\infty w_1(\tau) x(t - \tau) + \int_0^\infty \int_{-\infty}^\infty \Xi(\omega, t - \tau) r(\omega, \tau) d\tau d\omega$$

optimal an die echte Systemantwort y im Sinne des quadratischen Fehlers $\langle \int (\hat{y} - y)^2 \rangle$ anpasst, falls die Eingabe x des Systems Gauß'sches weißes Rauschen mit Varianz σ^2 ist.

Die Methode, Neuronen mit Hilfe der Wiener-Entwicklung zu charakterisieren hat den Nachteil, dass die Eingabe des Systems auf Gauß'sches weißes

Rauschen beschränkt ist. „Natürliche“ Schalle wie z. B. Kommunikationslaute zur Untersuchung des auditorischen Systems sind nicht möglich. Für das visuelle System können natürliche Bilder oder Filme einer natürlichen Umgebung nicht als Stimuli verwendet werden⁴⁶. Nun ist zwar die Erzeugung von Rauschen für das auditorische und visuelle System kein Problem. Jedoch lassen beispielsweise die zur Stimulation für das vestibuläre System konstruierten Drehstühle u.U. nur sinusoidale Stimuli zu, und diese auch nur in einer Richtung Wilden et al. (2002). Rauschen auf dem vestibuläre Bogengangsystem zu erzeugen würde ein in alle Richtungen beliebig drehbares Gestell und bei neurophysiologischen Experimenten eine entsprechende Verkabelung erfordern. Auf die angedeutete Art erzeugtes Rauschen wäre dann jedoch immer noch nicht unabhängig vom Stimulus auf das Odolithenorgan. Hierzu wären dann noch beliebige *lineare* Beschleunigungen in alle Richtungen erforderlich.

1.6.4 Ripple-Spektren

Einige Nachteile der Wiener-Entwicklung werden für das auditorische System durch die Verwendung von „Ripple-Spektren“ vermieden Shamma et al. (1996a,b), Klein et al. (2000), Depireux et al. (2001), Klein et al. (2006). Zum einen sind Ripple-Spektren „natürlichere“ Schalle als Rauschen, zum anderen kann man erwarten, dass das auditorische System in den Ripple-Spektren linearer ist als im zeitlichen Verlauf des Schalldrucks, da die Cochlea schon ein Kurzzeitspektrum erzeugt und jedes Kurzzeitspektrum als lineare Überlagerung von Ripple-Spektren dargestellt werden kann. Im Folgenden wird der Zweck von Ripple-Spektren erläutert.

Das Power-Spektrum $\Xi(\omega)$ eines Signals x

$$\Xi(\omega) = |X(\omega)|^2$$

genügt nicht, um das Eingangssignal in das Gehör vollständig zu charakterisieren, denn das Power-Spektrum berücksichtigt die Phasenbeziehungen zwischen den harmonischen Komponenten des Signals überhaupt nicht. Zwar werden Phasenbeziehungen i. d. R. nicht gehört⁴⁷, beispielsweise wird die Überlagerung von Tönen der Frequenzen 100 Hz und 200 Hz unabhängig von

⁴⁶In Einhauser et al. (2003) und Betsch et al. (2004) werden Videofilme vom Kopf einer Katze aus aufgenommen, natürliche visuelle Stimuli stehen also grundsätzlich zur Verfügung.

⁴⁷Dass Phasenbeziehungen nicht gehört werden bezeichnet man als Ohm'sches Gesetz der Akustik, benannt nach Georg Simon Ohm (1789–1845) (von Campenhausen 1981).

der Phasenbeziehung zwischen den überlagerten Tönen wahrgenommen⁴⁸. Dagegen tritt bei einer Überlagerung eines Reintons der Frequenz 100 Hz mit einem der Frequenz 101 Hz eine Schwebung auf, man hört also einen Ton, dessen Lautstärke periodisch schwankt (Terhardt 1998). Ob die Lautstärke gerade ein Maximum oder ein Minimum hat, hängt von der Phasenbeziehung zwischen den Reintönen ab. Das Power-Spektrum genügt also nicht, um daraus auf den Höreindruck eines Schalls zu schließen. Nehmen wir hingegen das Kurzzeitspektrum eines amplitudenmodulierten Reintons, oder, was das gleiche ist, das Kurzzeitspektrum der Überlagerung zweier Reintöne mit fast der gleichen Frequenz, dann gibt das Kurzzeitspektrum $\Xi(\omega, t)$ aus Gleichung (1.15) den Höreindruck insofern richtig wieder, als dass $\Xi(2\pi 100 \text{ Hz}, t)$ ungefähr eine Harmonische mit der Schwebungsfrequenz ist und $\Xi(\omega, t) \approx 0$ für $\omega \not\approx 2\pi 100 \text{ Hz}$ ist. Das Kurzzeitspektrum gibt also den Schall in etwa „gehör richtig“ wieder, jedenfalls bei geeignetem Zeitfenster g in Gleichung (1.14).

Die folgende Diskussion hat nur Sinn, wenn die betrachteten auditorischen Neuronen ungefähr linear auf das Kurzzeitspektrum reagieren, deren Peri-Stimulus-Zeit-Histogramm y also ein lineares Funktional des Kurzzeitspektrums ist. Der Einfachheit halber betrachten wir zunächst stationäre Schalle, also solche, deren Kurzzeitspektren $\Xi(\omega, t)$ von der Zeit t unabhängig ist, für die also $\Xi(\omega, t) = \Xi(\omega)$ gilt. Anstelle der Frequenz ω verwenden Shamma et al. (1996a,b) und die Position $p(\omega)$ auf der Cochlea, die durch die Frequenz ω maximal angeregt wird, was in etwa einer logarithmischen Frequenzskala entspricht. Eine logarithmische Frequenzskala ist „gehör richtig“ als eine lineare: Angenommen, man kann zwei Frequenzen gerade noch unterscheiden. Dann hängt die Frequenz*differenz* der beiden Töne von den Tonhöhen ab, das Frequenz*verhältnis* bleibt aber etwa konstant (Zwicker und Fastl 1999). Es gibt noch einen weiteren Grund, warum die logarithmische Frequenzskala gehör richtig ist: Die weiter unten beschriebenen Ripple-Spektren sollten im ganzen Frequenzbereich gleich laut sein, d. h. jeder gleich breite Ausschnitt sollte den gleichen Lautstärkeindruck hervorrufen. Würde man eine lineare Frequenzskala verwenden, so wären die Linien des Spektrums bei höheren Frequenzen zu dicht, d. h. höhere Frequenzen würden lauter erscheinen. Bei der logarithmischen Frequenzskala erscheint jeder Abschnitt des Spektrums ungefähr gleich laut (Zwicker und Fastl 1999).

⁴⁸Dies gilt jedenfalls für monaurales Hören. Beim binauralen Hören wird eine Phasenbeziehung zwischen Reintönen an beiden Ohren mit der *gleichen* Frequenz in eine horizontale Richtungswahrnehmung umgesetzt (Blauert 1999).

Die Antwort eines linearen Neurons auf ein stationäres Kurzzeitspektrum ist

$$y = \int_0^{\infty} r(p) \Xi(p) dp.$$

Hier ist r das „klassische“ (nicht spektro-temporale) rezeptive Feld, das einfach angibt, wie stark das Neuron auf welche Frequenz reagiert. Falls das Neuron linear ist, ist dessen Antwort dann die Summe der Antworten $r(p)$ auf die einzelnen Frequenzen bzw. Orte p auf der Cochlea.

Ein stationäres Ripple-Spektrum wird definiert durch sein Power-Spektrum

$$\Xi_R(p) = \Re [e^{-i(\Omega p + \Phi)}].$$

Hierbei heißt Ω Ortsfrequenz und Φ Ortsphase des Ripple-Spektrums. Die Antwort y eines linearen Neurons auf dieses Ripple-Spektrum ist dann mit $R(\Omega) := \int_0^{\infty} r(p) e^{-i\Omega p} dp$

$$\begin{aligned} y &= \int_0^{\infty} r(p) \Re [e^{-i(\Omega p + \Phi)}] dp = \\ &= \Re [R(\Omega) e^{-i\Phi}] = \\ &= |R(\Omega)| \cos\{\arg[R(\Omega)] - \Phi\}. \end{aligned}$$

Aus der von der Ortsphase Φ abhängigen Antwort $y(\Phi)$ des Neurons lässt sich somit $R(\Omega)$ berechnen⁴⁹ und hieraus durch Fourierrücktransformation das rezeptive Feld $r(p)$ selbst. Hierzu muss natürlich die Antwort y des Neurons bei allen Ortsfrequenzen Ω gemessen werden.

Die (zeitabhängige) Reaktion y eines Neurons auf ein zeitabhängiges Kurzzeitspektrum $\Xi(p, t)$ ist

$$y(t) = \int_0^{\infty} \int_{-\infty}^{\infty} r(p, \tau) \Xi(p, t - \tau) d\tau dp$$

Hierbei heißt $r(p, t)$ das spektro-temporale rezeptive Feld des Neurons. Um die Reaktion eines Neurons auf zeitabhängige Kurzzeitspektren $\Xi(p, t)$ vorherzusagen, benötigt man auch zeitabhängige Test-Stimuli. Hierzu sind nach

⁴⁹Es genügt, bei zwei Phasen, z. B. Φ und $\Phi + \pi$ zu messen, um das komplexe $R(\Omega)$ herauszufinden, die Autoren in Shamma et al. (1996a) und Depireux et al. (2001) messen jedoch bei vielen Phasen.

Shamma et al. (1996a) und Depireux et al. (2001) zeitabhängige Ripple-Spektren geeignet. Ein zeitabhängiges Ripple-Spektrum $\Xi_R(p, t)$ ist definiert als

$$\Xi_R(p, t) = \Re [e^{-i(\Omega p + \omega t)}] .$$

Hierbei heißt Ω Ortsfrequenz und ω Modulationsfrequenz. Die Antwort $y(t)$ eines linearen Neurons auf ein zeitabhängiges Ripple-Spektrum $\Xi_R(p, t)$ ist mit $T(\Omega, \omega) = \int r(p, t) e^{-i(\Omega p + \omega t)} dp dt$

$$\begin{aligned} y(t) &= \int_0^\infty \int_{-\infty}^\infty r(p, \tau) \Xi(p, t - \tau) d\tau dp \\ &= \int_0^\infty \int_{-\infty}^\infty r(p, \tau) \Re [e^{-i(\Omega p + \omega t)}] d\tau dp \\ &= \int_0^\infty r(p) \Re [e^{-i(\Omega p + \Phi)}] dp = \\ &= \Re [T(\Omega, \omega) e^{-i\omega t}] = \\ &= |T(\Omega, \omega)| \cos\{\arg[T(\Omega, \omega)] - \omega t\} \end{aligned}$$

Die Antwort $y(t)$ ist also eine Harmonische mit bestimmter Amplitude und Phase. Aus Amplitude und Phase der zeitabhängigen Antwort $y(t)$ lässt sich wiederum die Fouriertransformierte $T(\Omega, \omega)$ des raumzeitlichen rezeptiven Feldes $r(p, t)$ ablesen. Um das raumzeitliche rezeptive Feld $r(p, t)$ zu berechnen, muss man also die Antwort auf jedes beliebige zeitabhängige Ripple-Spektrum kennen, und zwar in Amplitude und Phase. Dann erhält man das raumzeitliche rezeptive Feld $r(p, t)$ durch eine zweidimensionale Fourierrücktransformation.

Die Schwierigkeit bei dem beschriebenen Ansatz mit Ripple-Spektren besteht darin, dass die betrachteten Neurone lediglich linear sein müssen. Darüber hinaus muss die neuronale Antwort bei sehr vielen Orts- und Modulationsfrequenzen gemessen werden, was zeitaufwändig ist⁵⁰. Jede dieser Mes-

⁵⁰Die Analyse wird allerdings wesentlich vereinfacht, wenn das spektro-temporale rezeptive Feld separierbar ist. „Separierbar“ heißt, das spektro-temporale rezeptive Feld setzt sich multiplikativ aus einem spektralen und einem temporalen Anteil zusammen:

$$r(p, t) = r_s(p) r_t(t) .$$

Es ist dann

$$T(\Omega, \omega) = R_s(\Omega) R_t(\omega) .$$

Nun muss, um $r(p, t)$ zu bestimmen, $T(\Omega, \omega)$ nur bei einer festen Ortsfrequenz Ω und ver-

sungen liefert relativ wenig Information, nämlich nur Amplitude und Phase, obwohl zahlreiche $y(t)$ -Werte gemessen werden. Dies liegt daran, dass man schon Linearität voraussetzt, woraus folgt, dass die Antwort $y(t)$ auf ein zeitabhängiges Ripple-Spektrum $\Xi_R(p, t)$ harmonisch ist. Das Ergebnis $T(\Omega, \omega)$ jeder Einzelmessung mit dem Wertepaar (Ω, ω) ist also jeweils nur als beste Annäherung der echten Antwort $y(t)$ durch eine lineare Näherung zu verstehen. Hieraus folgt jedoch leider noch nicht, dass das spektro-temporale rezeptive Feld $r(p, t)$ in irgendeinem Sinn die *beste* lineare Näherung der Antwort des Neurons in Bezug auf Ripple-Spektren als Stimuli darstellt. Über die Antwort auf „natürliche“ Kurzzeitspektren, welche sich aus vielen Ripple-Spektren linear zusammensetzen, kann daher noch weniger ausgesagt werden. Die Verwendung anderer Schalle als Ripple-Spektren zur Messung, z. B. Schalle aus der natürlichen Umgebung oder Kommunikationslaute, ist nicht möglich.

1.6.5 Methode von Theunissen

Theunissen et al. (2000) ist der erste, der methodisch die Einschränkung auf bestimmte Klassen von Stimuli überwindet. Tatsächlich sind die spektro-temporalen rezeptiven Felder, die mit „natürlichen“ Schallen gemessen werden, von denen verschieden, die mit zufälligen Pieptönen gemessen werden. Darüber hinaus werden auditorische Neuronen auf höheren Ebenen nicht stark durch einfache Stimuli wie weißes Rauschen angeregt, denn solche Neuronen seien eher auf natürliche Stimuli spezialisiert (Theunissen et al. 2000). Die gemessene neuronale Antwort sei y , die modellierte \hat{y} , wobei

$$\hat{y} = \sum_i \int_{-\infty}^{\infty} \hat{r}_i(\tau) x_i(t - \tau) d\tau.$$

Hierbei ist x_i der Logarithmus der Ausgabe einer Filterbank im Frequenzkanal i und \hat{r}_i das geschätzte spektro-temporale rezeptive Feld. Theunissen et al. (2000) minimiert

$$L = \int_{-\infty}^{\infty} (y - \hat{y})^2(t) dt, \quad (1.18)$$

schiedenen Modulationsfrequenzen ω gemessen werden, sowie bei einer festen Modulationsfrequenz ω und verschiedenen Ortsfrequenzen Ω . Dies reduziert den Aufwand gegenüber dem Fall mit nicht separierbarem rezeptivem Feld erheblich.

was dem Fall $\sigma = 0$ in Abschnitt 5.2 entspricht, und erhält das Gleichungssystem

$$\sum_j X_i^*(\omega)X_j(\omega)\hat{R}_j(\omega) = X_i^*(\omega)Y(\omega), \quad (1.19)$$

ein lineares Gleichungssystem für $R_j(\omega)$ für jede Frequenz ω . Großbuchstaben X_i, \hat{R} und Y bedeuten die Fouriertransformierten der Funktionen x_i, \hat{r} bzw. y . Theunissen löst das Gleichungssystem (1.19) mit der SVD-Methode (singular value decomposition). Dabei werden die Eigenvektoren der Matrix $(X_i^*(\omega)X_j(\omega))_{ij}$ mit betragsmäßig zu kleinen Eigenwerten vernachlässigt (Press et al. 1995). „Zu klein“ heißt, das Verhältnis zwischen größtem und betrachtetem Eigenwert übersteigt einen kritischen Wert. Dieser kritische Wert wurde so bestimmt, dass die in Marmarelis (1978) definierte durchschnittliche „coherence“ zwischen y und \hat{y} bei *neuen* Stimuli aus dem gleichen Ensemble maximal wurde. Warum nicht auch hier das Kriterium „minimiere L “ aus Gleichung (1.18) verwendet wurde, ist dem Autor der vorliegenden Arbeit schleierhaft. Die Methode von Theunissen et al. (2000) erscheint daher etwas ad-hoc.

*Wem ein Floh ins Ohr gesetzt
wird, der muss stochern.*

Wilhelm Busch

2

Psychoakustik

In diesem Kapitel sollen einige psychoakustische Grundlagen vorgestellt werden, die für das Verständnis der auditorischen Modelle in Kapitel 3 notwendig sind. Sie stammen größtenteils aus Zwicker (1982) und Zwicker und Fastl (1999).

2.1 Psychophysik

Die Psychophysik¹ interessiert sich für die Zusammenhänge zwischen der physikalischen Beschaffenheit des Reizes und den entsprechenden Wahrnehmungen. 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.

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 (von Campenhausen 1981).

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

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

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 Verhältniswerte: 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.

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) 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“ wahr. Hier steht also offenbar noch eine zusätzliche Qualität oder Modalität der Empfindung zur Verfügung, die sich von der Empfindung des Hörens wesentlich unterscheidet (von Campenhausen 1981, Moore 1997). „Sehen“ und „Hören“ sind also verschiedene Modalitäten. Modalitäten unterteilt man wiederum in „Qualitäten“. So sind „Helligkeit“, „Tonhöhe“ und „Lautstärke“ Qualitäten.

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. Die Relation „höher“ ist offenbar nicht transitiv, wie man an der Illusion einer unendlichen Folge immer aufsteigender Tonhöhen sieht (Deutsch 1995, Cave und Risset 2005).

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

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

2.2 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 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.

2.2.1 Schalldruckpegel und Schallintensitätspegel

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

Die Schalldruckamplitude \hat{p} eines Sinustons ist die Amplitude des Schalldrucks dieses Sinustons.

Der Schalldruckpegel L eines Sinustons ist definiert als

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

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 u. U. nicht eindeutig und wesentlich schwieriger zu schätzen als der Median.

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 ist die akustische Energie, die eine Fläche pro Zeiteinheit durchdringt. Schalldruckpegel L und Schallintensität I eines Sinustons hängen nach

$$L = 10 \log(I/I_0) \text{ dB} \quad (2.1)$$

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

Die Schallintensität pro Frequenzbereich heißt Schallintensitätsdichte, das entsprechende logarithmische Maß ist der Schallintensitätsdichteppegel.

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

2.2.2 Lautstärkepegel

Der Lautstärkepegel 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.

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

⁷Die „Einheit“ Bel, Kurzzeichen B, ist nach Alexander Bell (1847–1922) benannt (Breuer 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 (von Campenhausen 1981, Breuer 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 Abschnitt 2.2.3 und Abschnitt 2.5.3), 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 Abschnitt 2.2.2.)

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 und Pfeifer 1994). Eine bessere Messvorschrift für die (subjektive) Lautheit beliebiger Schalle ist in Abschnitt 2.2.3 bzw. in Abschnitt 2.5.3 beschrieben.

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 Ruhehörschwelle 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.

2.2.3 Lautheit

Die Lautheit 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 daher 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

$$N \approx \frac{1}{16} \left(\frac{I}{I_0} \right)^{0,3} \text{ sone.} \quad (2.2)$$

2.3 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

¹¹d. h. bei einem wirksamen Querschnitt des Ohres von $0,5 \text{ cm}^2$ wird nach Gleichung (2.1) 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 (von Campenhauen 1981).

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“. Einen Schall, der aus einer Überlagerung von reinen Tönen besteht, bezeichnet man als „Klang“. Wenn die Frequenzen alle ganzzahlige Vielfache einer Grundfrequenz sind, bezeichnet man diesen Klang als „harmonisch“.

Nicht nur reine Töne rufen eine Tonhöhenempfindung hervor, sondern auch z. B. 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: Die Verhältnistönhöhe misst man in mel¹⁴, wobei die Skala der hörbaren Tonhöhen von 0 bis 2400 mel reicht. Die Tonheit 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.

2.3.1 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

¹³Ein Glockenton wird mit einer Tonhöhe gehört, der keine Frequenz im Spektrum des Glockentons 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“ (von 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).

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 Ausgangs-

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

$$\frac{df(H_v)}{dH_v} \approx \begin{cases} 1 \text{ Hz/mel} & \text{für } 0 < f \leq 500 \text{ Hz} \\ 0,002f/\text{mel} & \text{für } f > 500 \text{ Hz} \end{cases} \quad (2.3)$$

$$f(H_v) \approx \begin{cases} H_v \text{ Hz/mel} & \text{für } 0 < f \leq 500 \text{ Hz} \\ 500 e^{0,002(H_v/\text{mel}-500)} \text{ Hz} & \text{für } f > 500 \text{ Hz} \end{cases} \quad (2.4)$$

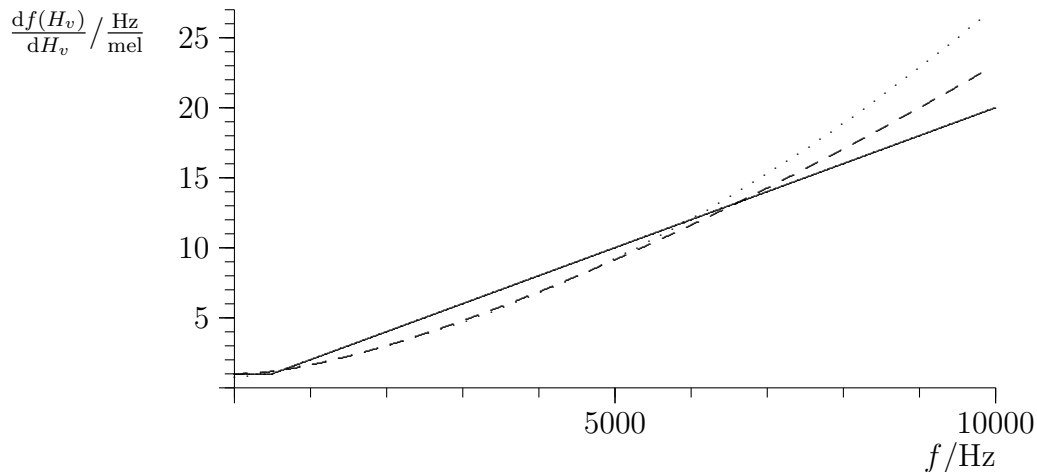


Abbildung 2.1: Breite von 1 mel in Hz bei verschiedenen Frequenzen nach Gleichung (2.3). Gestrichelt: Näherung nach Zwicker, gepunktet: Näherung nach Traummüller (siehe Fußnote 21 auf Seite 71).

Es stellt sich heraus, dass der Ort der maximalen Auslenkung der Basilarmembran in der Cochlea, vom Helicotrema aus gemessen (siehe Kapitel 3.1), proportional zur Verhältnistönhöhe ist. D. h. 0,7 mel entsprechen $9 \mu\text{m}$ oder einer inneren Haarzelle auf der Basilarmembran (Zwicker 1982).

stonhöhe 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 in Bezug zu 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.

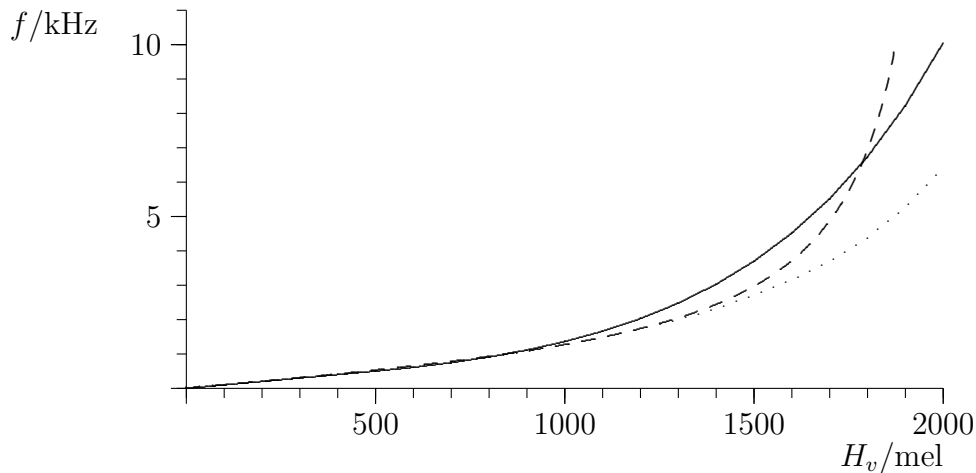


Abbildung 2.2: Frequenz f als Funktion der Verhältnistonhöhe H_v nach Gleichung (2.4). Gestrichelte und gepunktete Linien wie Abbildung 2.1.

Die gerade noch wahrnehmbare Tonhöhendifferenz oder Tonunterschiedsschwelle¹⁹ 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 andauern, 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 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 größer als 500 Hz entspricht dies einer Tonhöhendifferenz von etwa 30 mel.

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.

2.3.2 Tonheit

Die Tonheit eines reinen Tons ist

$$z \approx 100 H_v \text{ Bark/mel}, \quad (2.5)$$

wobei H_v die Verhältnistonhöhe bedeutet.

$\Delta z/\text{Bark}$	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

Tabelle 2.1: 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 Δl entlang der Basilarmembran.

Der Frequenzabstand, der 1 Bark entspricht, ist nach Gleichungen (2.4) und (2.5) für Frequenzen kleiner 500 Hz ungefähr 100 Hz und für höhere Frequenzen ungefähr $0,2f$.²¹ Er heißt auch „Frequenzgruppenbreite“. Alle Frequenzen, die innerhalb von einem Bark liegen, heißen auch „Frequenzgruppe“. Die Tonheit bzw. die Frequenzgruppe haben ihre eigene psychoakustische Berechtigung neben der Verhältnistonhöhe, da sie das geeignete Maß für einige

²¹In Zwicker (1982) und Zwicker und Fastl (1999) werden genauere Näherungsformeln als Gleichungen (2.3) und (2.4) angegeben:

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

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 (Sottek 1993):

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

psychoakustische Phänomene darstellen. Im Folgenden sollen 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 einwirkt.

Eine zweite Methode zur Bestimmung der Frequenzgruppenbreite arbeitet analog zur ersten (siehe Abbildung 2.3). 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).

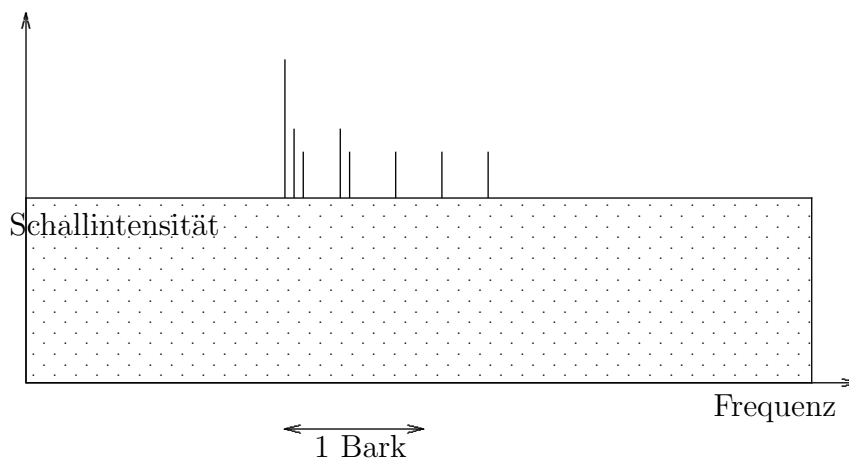


Abbildung 2.3: Bestimmung der Frequenzgruppenbreite. Die Schallintensität des lautesten reinen Tons genügt gerade, um ihn trotz des Rauschens hörbar zu machen. Die beiden mittellauten Töne addieren sich zur selben Schallintensität. Bei drei oder mehr Tönen nimmt die Lautstärke eines Tons nicht mehr ab, 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.

2.4 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“.²²

Die „Mithörschwelle“ ist der Schalldruckpegel eines reinen Tons, den dieser haben muss, damit er trotz des Störschalls gerade noch wahrgenommen wird. Als Störschall kann z. B. Schmalbandrauschen verwendet werden. Wichtig sind die Mithörschwellen für frequenzgruppenbreites Schmalbandrauschen. Je näher sich die Frequenz des Testtons an der Mittenfrequenz des Schmalbandrauschens 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. 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“ und „Nachverdeckung“: 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. Die Existenz von Vorverdeckung erscheint zunächst paradox, da ein Schall, wie laut auch immer, nicht mehr beeinflusst

²²Das Phänomen der Verdeckung ist monaural, d. h. werden Test- und Störschall verschiedenen Ohren dargeboten, so ist keine Verdeckung mehr nachweisbar (von Campenhäusen 1981).

sen kann, was *vor* Beginn des Störschalls geschah. Jedoch dauert sowohl die Registrierung des Testtons in der Cochlea als auch die nachfolgende Verarbeitung im Gehirn etwas, und diese beeinflusst der Störschall.

2.5 Funktionsschemata

Psychoakustische Funktionsschemata oder Funktionsmodelle 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.

2.5.1 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. Für die Anregung I_G einer Frequenz f gilt

$$I_G(f) := \int_{f-\frac{1}{2}\Delta f_G(f)}^{f+\frac{1}{2}\Delta f_G(f)} \frac{dI(f)}{df} df, \quad (2.6)$$

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.

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

Ein frequenzgruppenbreites Schmalbandrauschen hat auf der Tonheitsskala eine dreieckförmige Anregung, ein reiner Ton hat eine rechteckige Anregung. Zur Berechnung der Anregung kann man nach Gleichung (2.6) 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 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 Abschnitt 2.4 entspricht, d. h.

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

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 Sotttek (1993). Darüberhinaus wird die nichtlineare Auffächerung der Mithörschwellen vernachlässigt.

Das entsprechende logarithmische Maß heißt „Erregungspegel“.

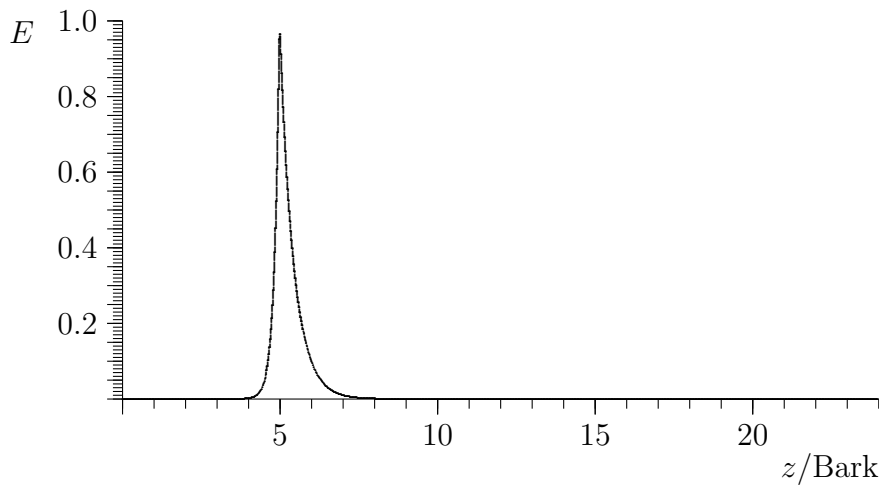


Abbildung 2.4: Erregung $E(z) = M(z - 5 \text{ Bark})$ des Gehörs durch einen reinen Ton mit der Tonheit 5 Bark.

Man beachte, dass die in Abbildung 2.4 gezeigte Erregung in Übereinstimmung mit Abschnitt 2.4 zu tiefen Frequenzen hin schneller (etwa 27 dB/Bark) als zu hohen abfällt (etwa 10 dB/Bark).

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 und Fastl 1999).

2.5.2 Schwellenfunktionsschema für langsame Schalländerungen

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

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

$$\Delta z = \frac{1}{27} \text{ Bark} = 3,7 \text{ mel}.$$

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.

2.5.3 Funktionsschema der Lautheit

Zwei Töne mit großem Frequenzabstand addieren sich in der Lautheit. Zwei Töne der gleichen Frequenz und Phase addieren sich natürlich in der Schallintensität und führen zu einer Änderung der Lautheit entsprechend Gleichung (2.2). 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'

$$N = \int_0^{24 \text{ Bark}} N' dz,$$

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

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

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

Man beachte, dass sich für eine gerade noch unhörbare Erregung $E = E_{RHS}$ ergibt $N' = 0$. Der Vorfaktor 0,08 ist so gewählt, dass die Lautheit eines Sinustons mit dem Lautstärkepegel 40 phon gerade $N = 1$ sone ist, wie in Abschnitt 2.2.3 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 Breuer (1987) verbreitet.

All models are wrong but some are useful.

George Box

3

Modelle für das periphere auditorische System

Hier sollen einige alternative Modelle für Komponenten des auditorischen Systems vorgestellt werden. In Abschnitt 3.3 wird das in dieser Arbeit verwendete Modell für das periphere auditorische System beschrieben.

3.1 Cochlea

Die Cochlea korrekt zu modellieren ist sehr aufwändig. Vermutlich gelingt dies letztlich nur mit dreidimensionalen Modellen der beteiligten Membranen, evtl. ist sogar eine detaillierte hydrodynamische Modellierung der Flüssigkeit in der Cochlea notwendig (Keidel 1975, Allen 1976, Holmes und Rubenfeld 1981). Auch sind die aktiven Prozesse, verursacht durch die äußeren Haarzellen, zu berücksichtigen (Mammano und Nobili 1993, Kössl 1997, Nobili et al. 1998). Stellt man geringere Ansprüche an die Exaktheit der Modellierung, so genügt es, die Cochlea als linearen Filter aufzufassen, d. h. die Auslenkung einer Stelle an der Basilarmembran ist ein lineares Funktional des auf das Trommelfell einwirkenden Schalls. Im Folgenden werden einige lineare Cochleamodelle vorgestellt. Gehörmodelle sind in AIM (Patterson et al. 1995) und Lutear (O'Mard et al. 1997) als Software implementiert.

3.1.1 Erregung

Es scheint zunächst naheliegend, die Cochlea einfach psychoakustisch anhand der in Abschnitt 2.5.1 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 auf der Basilarmembran 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. Wie groß sollten außerdem diese Zeitabschnitte sein? Welches Zeitfenster (cosinusförmig, gaußförmig, etc.) sollte verwendet werden?
- Nach obiger Vorgehensweise 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. Dies wäre relativ rechenaufwändig.
- 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 cochleare Eingabesignal jedoch einfach ist, z. B. nur Tiefpassrauschen, Reintöne o. ä. enthält, 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.

3.1.2 Lineare Schwinger

Die einzelnen Orte der Basilarmembran werden als unabhängige² gedämpfte lineare Schwinger modelliert. Die Gleichung für die Auslenkung $y(x, t)$ an

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

²Die einzelnen quer über die Basilarmembran verlaufenden Fasern sind zwar durch Gewebe verbunden, jedoch hat dies eine geringere Festigkeit als die Fasern selbst. Desweiteren kann man davon ausgehen, dass zwei Orte auf der Basilarmembran, die genügend weit voneinander entfernt sind, weitgehend unabhängig schwingen. Da die Schallgeschwindigkeit in der Flüssigkeit der Cochlea viel größer als die Fortpflanzungsgeschwindigkeit einer durch die Kopplung der Fasern aneinander eventuell über die Basilarmembran lau-

einer Stelle x der Basilarmembran ist dann

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

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 den Kraftstoß $F(t)/m = \delta(t)$ folgende Auslenkung, ist

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

Hierbei bedeutet $\text{H}(t)$ die Heaviside'sche Sprungfunktion.

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

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

wobei die Amplitude

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

und die Phasenverschiebung

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

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

Wählt man $\beta = \frac{1}{2\text{ms}} = 500 \text{ Hz}$,³ so ergibt sich als Halbwertsbreite⁴ des Filters für $\beta \ll \omega_0$ ungefähr $\beta/2\pi \approx 80 \text{ Hz}$. Die Halbwertsbreite der cochleären Filter

fenden Welle ist, erscheint auch die Annahme als realistisch, dass alle Orte zugleich die gleiche anregende Kraft erfahren.

³Man 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. Schalle wirken je nach Intensität etwa 10-100 ms nach, d. h. z. B. kurze Reintöne oder Pulse, die 10-100 ms nach weißem Rauschen gespielt werden, werden nicht oder mit geringerer Lautstärke gehört (Zwicker und Fastl 1999). Deshalb sollten die Anstiegszeiten der verwendeten Filter relativ kurz sein.

⁴Der Abstand zur Resonanzfrequenz $f_R = 2\pi\omega_R$, bei dem das Amplitudenquadrat $A^2(\omega)$ auf die Hälfte des Wertes bei der Resonanzfrequenz abgefallen ist.

beträgt jedoch nur etwa 10–30 Hz,⁵ d. h. die Frequenzauflösung der so wie oben beschrieben modellierten Cochlea wäre zu schlecht.

3.1.3 Lineare Filter vierter Ordnung

Die Cochlea wird nach Sottek (1993) durch lineare Filter vierter Ordnung⁶ mit der Impulsantwort

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

beschrieben (siehe Abbildung 3.1).

Die Antwort auf einen Cosinus ist im stationären Fall

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

wobei die Amplitude

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

⁵Nach Abschnitt 2.5.1 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 ca. 30 Hz führt.

⁶Vierter Ordnung deshalb, da die Impulsantwort durch eine Differenzialgleichung vierter Ordnung beschrieben werden kann. Es handelt sich um einen sogenannten *minimalphasigen* Bandpassfilter vierter Ordnung, d. h. die Phase $\varphi(\omega)$ der Antwort $y(t) = A(\omega) \cos[\omega t - \varphi(\omega)]$ auf das Eingangssignal $x(t) = \cos(\omega t)$ ist minimal in dem Sinne, dass für die Phase $\varphi_0(\omega)$ jedes Filters, der den Absolutbetrag $A(\omega)$ der Übertragungsfunktion realisiert, $\varphi(\omega) < \varphi_0(\omega)$ für alle ω gilt. Zu jeder vorgegebenen Amplituden-Übertragungsfunktion $A(\omega)$ eines kausalen, stabilen, zeitinvarianten, linearen Filters, existiert ein derartiger minimalphasiger Filter und dieser ist eindeutig (Unbehauen 1993). Die Wahl eines minimalphasigen Filters für das auditorische System ist plausibel, denn der minimalphasige Filter mit Amplitude $A(\omega)$ hat die geringste Verzögerung unter allen Filtern mit Amplitude $A(\omega)$. „Will“ das auditorische System also eine bestimmte Amplitude $A(\omega)$ und damit eine bestimmte Trennschärfe für verschiedene Frequenzen realisieren, so ist der minimalphasige Filter die beste Wahl, denn dieser realisiert $A(\omega)$ mit der geringsten Verzögerung. Der Betrag der Übertragungsfunktion dieser Filter stimmt im Durchlassbereich gut mit den von R. D. Patterson 1982 angegebenen ROEX-Filtern 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.

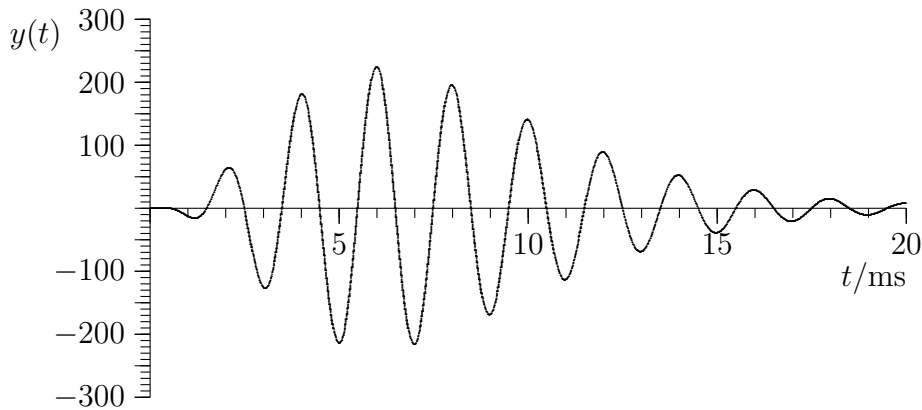


Abbildung 3.1: Impulsantwort nach Gleichung (3.1) der Stelle der Cochlea mit $f_0 = \omega_0/2\pi = 500$ Hz, wobei $\tau = 2$ ms.

beträgt⁷. Die Halbwertsbreite dieses Filters ist ungefähr $0,07/\tau$. Mit $\tau = 2$ ms ergibt sich eine Halbwertsbreite von 35 Hz. Das Gehör hat zwar eine noch bessere Leistung, dieser Wert ist jedoch akzeptabel.

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. Abschnitt 2.5.1). Dies wird durch die Beziehung

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

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

Abbildung 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 sind⁸.

Um die Auslenkung der Basilarmembran zu berechnen, müsste man nun das Eingangssignal der Cochlea, d. h. den Schall, mit der Impulsantwort falten.

⁷Die Fouriertransformierte von Gleichung (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 $\tau\omega_0 \gg 1$. Aus $\tilde{y}(\omega) = \frac{1}{[1+i\tau(\omega-\omega_0)]^4}$ ergibt sich die angegebene Amplitude.

⁸Die cochleäre Erregung ist allerdings auch im Modell der linearen Filter vierter Ordnung *etwas* asymmetrisch. Dies kommt daher, dass die Bandbreite der Filter wegen Gleichung (3.2) mit steigender Bestfrequenz ω_0 zunimmt (Sottek 1993). Die Asymmetrie ist allerdings nicht so stark wie in der Realität, wie Abbildung 3.4 zeigt. Hier irrt Sottek (1993).

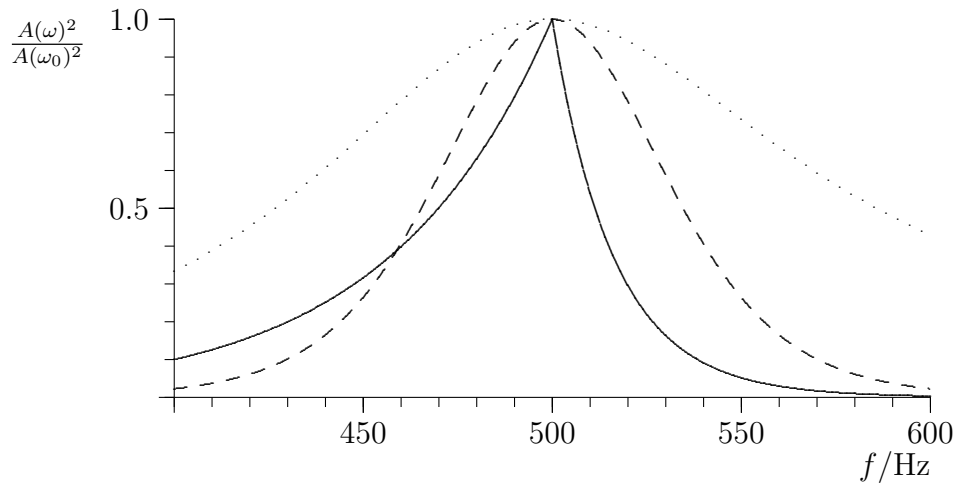


Abbildung 3.2: Verlauf des cochleären Filters mit der Bestfrequenz $f_0 = 500$ Hz in verschiedenen Cochleamodellen. Aufgetragen ist das auf die Bestfrequenz bezogene Amplitudenquadrat $A^2(\omega)/A^2(\omega_0)$ in Abhängigkeit von der Frequenz $f = 2\pi\omega$.

Durchgezogen: Erregungsmodell nach Abschnitt 2.5.1.

Gestrichelt: Lineare Filter vierter Ordnung.

Punktiert: Lineare Schwinger nach Abschnitt 3.1.2.

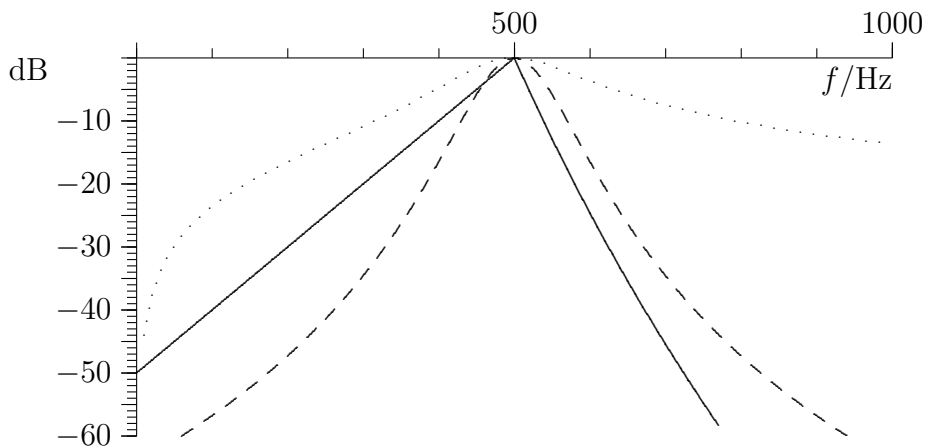


Abbildung 3.3: Wie Abbildung 3.2, nur als Pegel.

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 $t + \Delta t$.

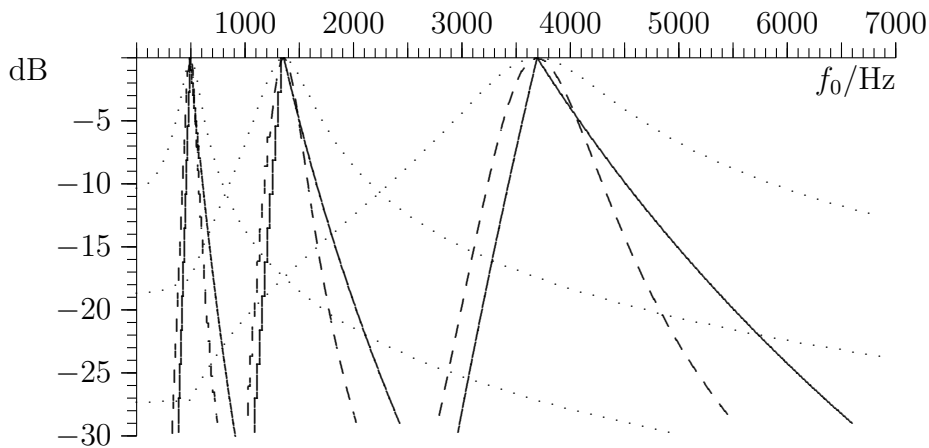


Abbildung 3.4: Verlauf der cochleären Erregung in verschiedenen Cochleamodellen. Aufgetragen ist der Erregungspegel $10 \log[A_{\omega_0}^2(\omega)/A_{\omega_0}^2(\omega_0)]$ der cochleären Filter bei Stimulation mit einem reinen Ton der Tonheit 5, 10 bzw. 15 Bark in Abhängigkeit der Bestfrequenz $f_0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

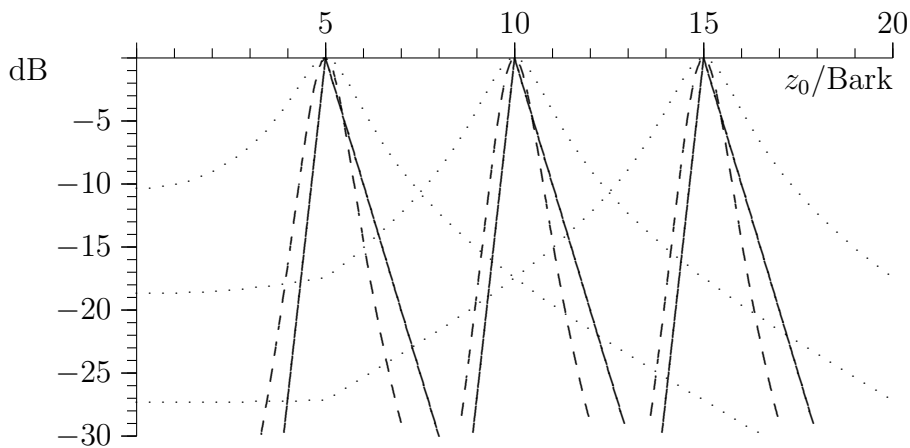


Abbildung 3.5: Wie Abbildung 3.4, nur ist die Bestfrequenz z_0 in der Tonheit-Skala aufgetragen. Die cochleären Erregungsmuster erscheinen somit gleich breit.

Dieses Konzept führt zu folgendem effizienten Algorithmus⁹:

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

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.

⁹Mit $\bar{y}(t) := \frac{1}{3\tau^4} t^3 e^{ct}$, 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

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

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.

Beispiele für die Reaktion einer cochleären Filterbank im Modell der linearen Filter vierter Ordnung zeigen Abbildung 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. Gleichung (3.1) und (3.2). d. h. Filter mit hoher Bestfrequenz relaxieren nach Erregung schneller als solche mit niedrigerer Bestfrequenz.

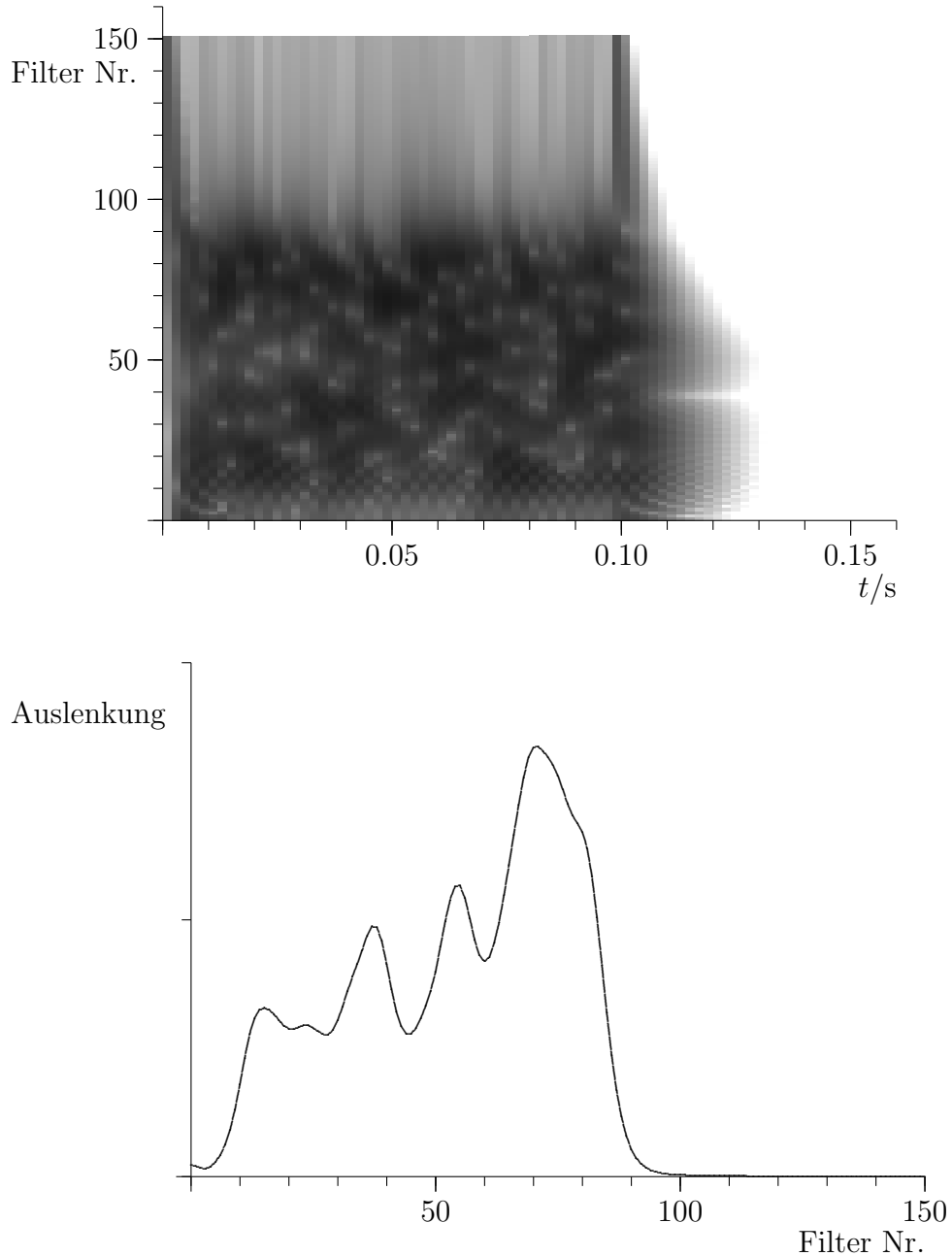


Abbildung 3.6: Stimulus war Tiefpassrauschen von 100 bis 1000 Hz. Die obere Grafik zeigt den Logarithmus der jeweils über 20 ms gemittelten quadratischen Auslenkung der Cochlea im Modell linearer Filter vierter Ordnung. Nach rechts ist die Zeit in s aufgetragen, nach oben die Nummer der cochleären Filter. 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.

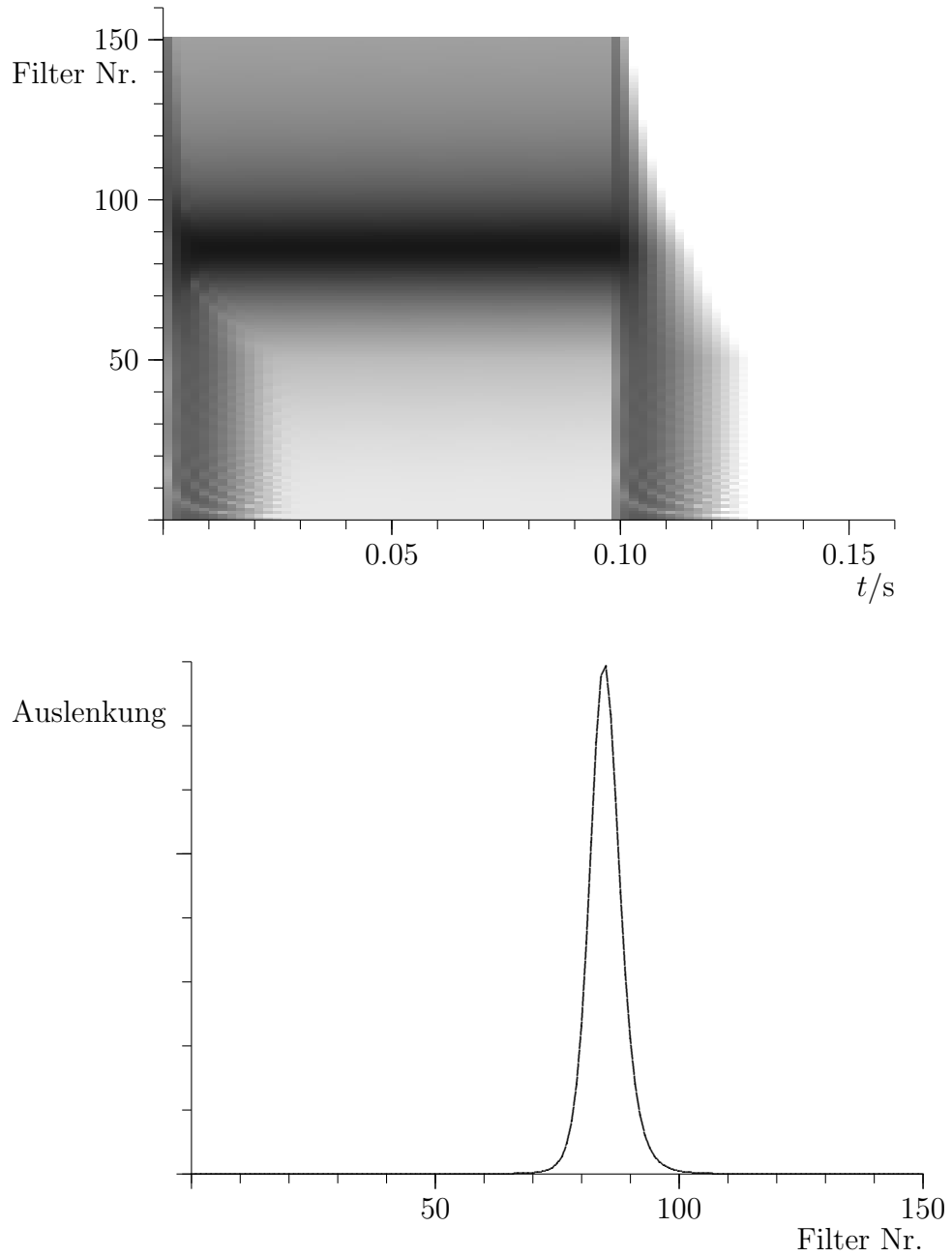


Abbildung 3.7: Wie Abbildung 3.6, nur war der Stimulus diesmal ein Reinton mit Frequenz 1000 Hz.

3.2 Innere Haarzellen

3.2.1 „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 gibt, die Neurotransmitter ausschütten können (siehe Abbildung 3.8).

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

$$\begin{aligned}C_a &= xc_a \\ C_{na} &= (1 - x)c_{na}.\end{aligned}\tag{3.3}$$

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.

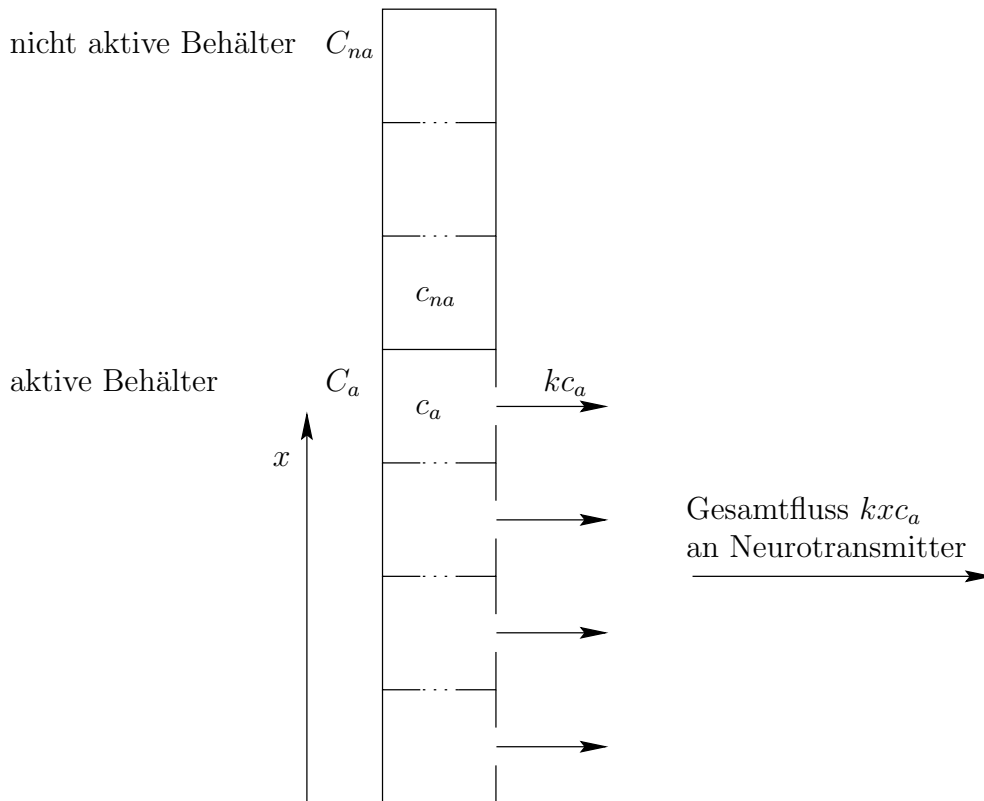


Abbildung 3.8: Das „State-Partition“-Modell für innere Haarzellen. Kästen bedeuten Behälter mit Neurotransmitter. Die nach rechts offenen Behälter sind aktiv und schütten Neurotransmitter aus. Die von den Behältern ausgehenden Pfeile bedeuten Neurotransmitterströme. Im eigentlichen Modell sind alle Behälter voneinander getrennt, d. h. die gepunkteten Sperren sind geschlossen. In der im Text beschriebenen Näherung denkt man sich die gepunkteten Sperren jedoch als offen. Die Größen 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} .

Um die Modellierung 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 Abbildung 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 Gleichung (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“.

3.2.2 Meddis-Modell

In Sottek (1993) wird ein Modell für innere Haarzellen vorgestellt. Es beruht auf Meddis (1986, 1988).

Siehe Abbildung 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 Nervenfasern diffundieren. Die Konzentration $c(t)$ des Transmitters im synaptischen Spalt bestimmt die Wahrscheinlichkeit für das Auftreten eines Aktionspotentials im postsynaptischen Nerv. Die Permeabilität $k(t)$

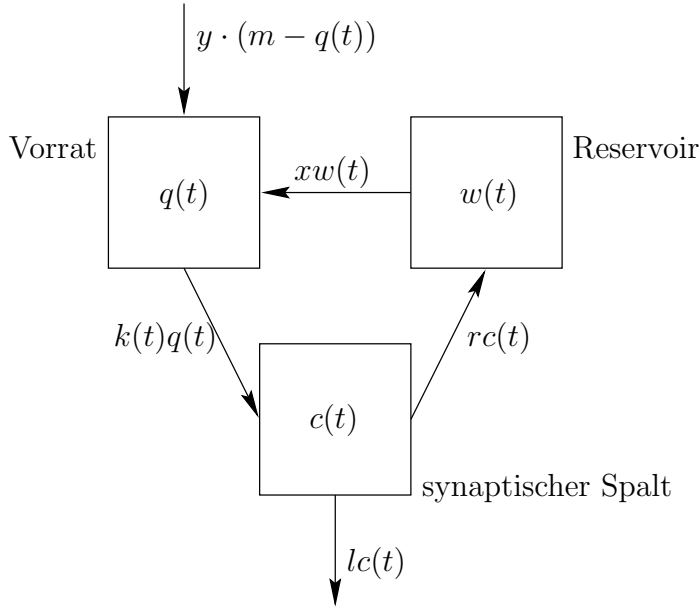


Abbildung 3.9: Das Meddis-Modell für innere Haarzellen. Kästen bedeuten Behälter mit Neurotransmitter, Pfeile Neurotransmitterströme.

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 Tabelle 3.1.

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

Tabelle 3.1: Parameter einer Nervenfasern 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)¹⁰

$$P = hc(t).$$

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

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

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

¹⁰ 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.

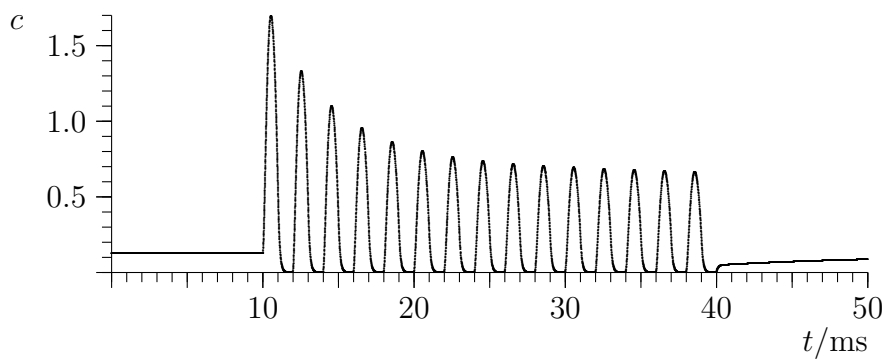


Abbildung 3.10: Neurotransmitterkonzentration $c(t)$ nach dem Meddis-Modell bei einem von 10–40 ms dauernden reinen Ton der Frequenz $f = 500$ Hz mit Amplitude 50 als Stimulus $s(t)$.

3.3 Gesamtmodell der auditorischen Vorverarbeitung

Im Gesamtmodell der auditorischen Vorverarbeitung wurde auf eine genaue Modellierung der inneren Haarzellen nach Abschnitt 3.2 zu Gunsten größerer Einfachheit und Nachvollziehbarkeit verzichtet. Stattdessen wurde als Eingabe ein zeitabhängiges Spektrum verwendet, ähnlich wie in Gleichung 1.15, allerdings psychoakustisch und physiologisch besser begründet. Abbildung 3.11 fasst die verwendete auditorische Vorverarbeitung zusammen.

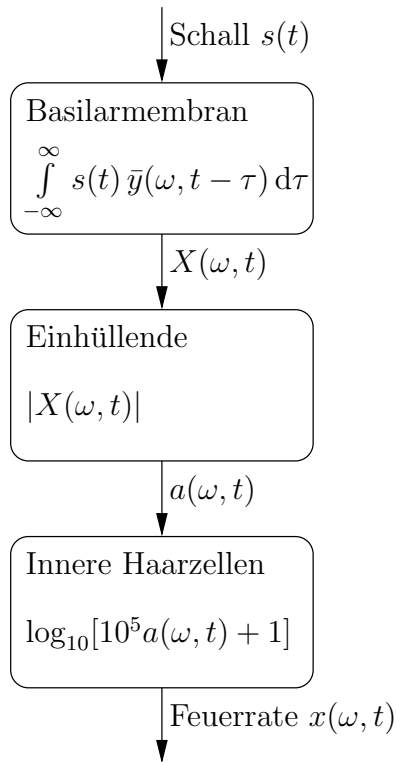


Abbildung 3.11: Gesamtmodell der auditorischen Vorverarbeitung. Pfeile bedeuten Datenströme, Kästen bedeuten Verarbeitungseinheiten. Der Schall gelangt durch das Trommelfell als Druck an jede Position (beschrieben durch die Bestfrequenz ω) an der Basilarmembran. Die Auslenkung $\Re X(\omega, t)$ an der Stelle ω ergibt sich durch Faltung mit \bar{y} aus Gleichung (3.3). Die Einhüllende $a(\omega, t)$ ergibt sich als $|X(\omega, t)|$ aus der „komplexen Auslenkung“ $X(\omega, t)$. Die Feuerrate $x(\omega, t)$ der inneren Haarzellen hängt logarithmisch von der Einhüllenden ab.

Um eine psychoakustisch und physiologisch korrekte Vorverarbeitung zu

erhalten, ist die erste Stufe ein linearer Filter vierter Ordnung nach Abschnitt 3.1.3. Wie in Abschnitt 1.6.4 ausgeführt, wird die Phase der Auslenkung der Basilarmembran bei monauralem Hören nicht wahrgenommen. Daher genügt es, die Einhüllende der Auslenkung der Basilarmembran zu berechnen. Dies ist möglich, indem die Impulsantwort in Gleichung (3.1) durch ihren Imaginärteil ergänzt wird¹¹:

$$\bar{y}(\omega, t) := \frac{1}{3\tau^4} t^3 e^{(-1/\tau + i\omega)t} H(t)$$

Sei s der Schall, dann berechnet sich die Einhüllende a folgendermaßen:

$$\begin{aligned} X(\omega, t) &:= \int_{-\infty}^{\infty} s(\tau) \bar{y}(\omega, t - \tau) d\tau \\ a(\omega, t) &:= |X(\omega, t)| \end{aligned}$$

Die Größe $X(\omega, t)$ entspricht derjenigen in Gleichung (1.14) mit dem verallgemeinerten Zeitfenster

$$g(t) = \frac{1}{3\tau^4} t^3 e^{-t/\tau}.$$

Man beachte, dass im Ggs. zu Gleichung (1.14) das verallgemeinerte Zeitfenster von τ und damit nach Gleichung (3.2) von ω *abhängt*, so dass die auf Gleichung (1.14) folgenden Ausführungen hier nicht mehr gelten.

Da die Feuerrate der inneren Haarzellen etwa mit dem Logarithmus des Schalldrucks zunimmt (Yost 1994), wird diese wesentliche Nichtlinearität in der Vorverarbeitung berücksichtigt, indem

$$x_i(t) := x(\omega_i, t) := \log_{10} [10^5 a(\omega_i, t) + 1]$$

als vorverarbeiteter Schall verwendet wird. Hierbei sind die ω_i logarithmisch verteilte Frequenzen, was ungefähr der gehörrichtigen Bark-Skala entspricht (siehe Abschnitt 2.3.2). Die Konstante 10^5 ist dabei so gewählt, dass bei einem Schallintensitätspegel von weniger als 0 dB, der etwa der Ruheshwelle (Zwicker 1982) entspricht, $x(\omega_i, t) \approx 0$ ist, denn eine Amplitude $a(\omega_i, t) = 10^{-5}$ entspricht einem Schallintensitätspegel von 0 dB, falls eine Amplitude von $a(\omega_i, t) = 1$ einem Schallintensitätspegel von 100 dB entspricht, der durch einen Reinton $s(t) = \cos(\omega_i t)$ mit Amplitude 1 erreicht wird und hier das Ende der schmerzfrei wahrzunehmenden Lautstärkenskala darstellen soll.

¹¹Das gleiche Verfahren verwendet Raab (2001), um Schalle gehörrichtig statistisch zu analysieren.

*Premature optimization is the root
of all evil.*

Donald Knuth

4

Optimales rezeptives Feld in einer Dimension

In diesem Kapitel werden verschiedene Verfahren zur Rekonstruktion des rezeptiven Feldes bei verrauschten Messdaten aus Optimierungsprinzipien abgeleitet und untereinander verglichen.

4.1 Optimale lineare Näherung

Das Neuron wird auf diejenige Art und Weise linear genähert, dass der quadratische Fehler zwischen dem Output des linearen Neurons bei gegebenem Input und den Messdaten minimal ist.

Der Output eines Neurons sei $y(t)$, der Input $x(t)$. Gesucht ist das rezeptive Feld $\hat{r}(t)$ mit

$$\hat{y} = \hat{r} \star x \tag{4.1}$$

für das der quadratische Fehler

$$L = \int_{-\infty}^{\infty} [y(t) - \hat{y}(t)]^2 dt \tag{4.2}$$

zwischen approximiertem Output \hat{y} und gemessenem Output y minimal ist.

Gleichungen (4.2) und (4.1) eingesetzt ergibt

$$\begin{aligned} L &= \int [y(t) - (\hat{r} \star x)(t)]^2 dt = \\ &= \int \left[y(t) - \int \hat{r}(\tau)x(t - \tau) d\tau \right]^2 dt \end{aligned}$$

Notwendige Bedingung für ein Minimum ist

$$\frac{\delta L}{\delta \hat{r}(\tau)} = 0,$$

woraus sich

$$\int 2 [y(t) - (\hat{r} \star x)(t)] x(t - \tau) dt = 0$$

ergibt.

Aufgrund der Definition der Kovarianz (A.1) bedeutet dies

$$\begin{aligned} \text{cov}(y - \hat{r} \star x, x)(-\tau) &= 0 \\ \text{cov}(x, y - \hat{r} \star x)(\tau) &= 0 \\ \text{cov}(x, y)(\tau) - \text{cov}(x, x \star \hat{r})(\tau) &= 0 \\ \text{cov}(x, y)(\tau) - [\text{cov}(x, x) \star \hat{r}](\tau) &= 0 \end{aligned}$$

unter Verwendung der Antisymmetrie (A.3) und Distributivität (A.4) der Kovarianz und des gemischten Assoziativgesetzes zwischen Kovarianz und Faltung (A.9).

Fouriertransformation nach der Variablen τ ergibt wegen (B.7), (B.6) und (B.8)

$$\begin{aligned} X^*(\omega)Y(\omega) - \mathcal{F}\text{cov}(x, x)(\omega)\hat{R}(\omega) &= 0 \\ X^*(\omega)Y(\omega) - |X(\omega)|^2\hat{R}(\omega) &= 0, \end{aligned}$$

wobei die Großbuchstaben X , Y und R die Fouriertransformierten $\mathcal{F}x$, $\mathcal{F}y$ bzw. $\mathcal{F}r$ der Funktionen x , y und r bezeichnen.

Falls der Input x alle Frequenzkomponenten enthält, also $|X(\omega)|^2 > 0$ für alle Frequenzen ω , ergibt sich für die Fouriertransformierte \hat{R} des rezeptiven Feldes r

$$\hat{R} = \frac{X^*Y}{|X|^2}.$$

Falls der Input x alle Frequenzkomponenten mit Intensität Eins enthält, also $|X(\omega)|^2 = 1$ für alle Frequenzen ω gilt, ergibt sich wegen (B.7) nach Fourier-Rücktransformation

$$\hat{r} = \text{cov}(x, y), \tag{4.3}$$

also einfach die Kovarianz zwischen Input und Output.

Dies ist beispielsweise für weißes Rauschen oder einen Delta-Puls als Input der Fall. Für weißes Rauschen erhält man die „reverse correlation“-Methode (Cai et al. 1997).

Für einen Delta-Puls $x(t) = \delta(t)$ als Input ergibt sich

$$\hat{r}(t) = \int \delta(\tau)y(\tau + t) d\tau = y(t),$$

d. h. die Impulsantwort des (i. d. R. nichtlinearen) Neurons ist dasjenige lineare rezeptive Feld, das den im quadratischen Mittel optimalen Output liefert. Dies ist nicht erstaunlich, da für $x = \delta$ und $\hat{r} = y$ wegen (4.1) $\hat{y} = y$, somit $L = 0$ (4.2) gilt.

4.2 Bayes'scher Schätzer

Anwendung der Bayes'schen Methode soll in diesem Fall heißen: Berechne das wahrscheinlichste lineare rezeptive Feld unter der Annahme, dass das Neuron linear ist, dass der gemessene neuronale Output durch Gauß'sches Rauschen verfälscht ist und das rezeptive Feld a-priori wie Gauß'sches weißes Rauschen verteilt ist. Das Rauschen soll dabei unabhängig von jeder der anderen betrachteten Größen sein.

Das „Neuron ist linear“ bedeutet, es gibt ein (lineares) rezeptives Feld r , so dass für den Erwartungswert $E(y)$ des Outputs y , gegeben der Input x , gilt

$$E(y) = r \star x.$$

Der neuronale Output ist durch Gauß'sches weißes Rauschen $\sigma_n n$ mit Varianz σ_n^2 verfälscht, d. h.

$$y = r \star x + \sigma_n n. \quad (4.4)$$

Gesucht ist dasjenige rezeptive Feld \hat{r} , das die Wahrscheinlichkeit $p(r|y)$, dass das rezeptive Feld r vorliegt, gegeben die Messung y , maximiert. Nach der Bayes'schen Regel ist

$$p(r|y) = \frac{p(y|r)p(r)}{p(y)}.$$

Es genügt also, die Größe

$$L = \ln[p(y|r)p(r)] \quad (4.5)$$

bezüglich des rezeptiven Feldes r zu maximieren, da der Nenner $p(y)$ unabhängig von r ist.

Wegen (4.4) gilt

$$p(y|r) \sim \exp \left[-\frac{\int_{-\infty}^{\infty} n(t)^2 dt}{2} \right] \quad (4.6)$$

mit $n = (y - r \star x)/\sigma_n$ wegen (4.4).

Da das rezeptive Feld r a-priori wie weißes Rauschen mit Varianz σ_r verteilt sein soll, gilt

$$p(r) \sim \exp \left[-\frac{\int_{-\infty}^{\infty} r(t)^2 dt}{2\sigma_r^2} \right]. \quad (4.7)$$

Die Wahrscheinlichkeiten aus (4.6) und (4.7) in (4.5) eingesetzt ergibt

$$L \sim -\frac{1}{2} \left\{ \frac{\int_{-\infty}^{\infty} [y(t) - (r \star x)(t)]^2 dt}{\sigma_n^2} + \frac{\int_{-\infty}^{\infty} r(t)^2 dt}{\sigma_r^2} \right\}. \quad (4.8)$$

Notwendige Bedingung für ein Maximum ist

$$\begin{aligned} \frac{\delta L}{\delta r(\tau)} \Big|_{r=\hat{r}} &= 0 \\ \frac{\int [y(t) - (r \star x)(t)]x(t-\tau)}{\sigma_n^2} - \frac{r(\tau)}{\sigma_r^2} &= 0. \end{aligned}$$

Mit $\sigma := \sigma_n/\sigma_r$ und den Eigenschaften von Kovarianz und Faltung aus Anhang A ergibt sich

$$\text{cov}(x, y) - \text{cov}(x, x) \star r - \sigma r = 0.$$

Fouriertransformation ergibt wegen (B.7), (B.6) und (B.8)

$$X^*Y - |X|^2 R - \sigma^2 R = 0.$$

Falls $\sigma > 0$ erhält man die eindeutige Lösung¹

$$R = \frac{X^*Y}{|X|^2 + \sigma^2}. \quad (4.9)$$

Der Fall $\sigma_r \rightarrow \infty$ bedeutet wegen (4.8), dass keine a-priori Information über das rezeptive Feld r vorliegt, da dann der zweite, zu $\ln p(r)$ proportionale, Term wegfällt. Da dann $\sigma = 0$ ist, erhält man im Fall keiner a-priori Information über das rezeptive Feld r die gleiche Lösung wie mit der optimalen linearen Näherung aus Abschnitt 4.1.

¹Diese Lösung entspricht einem Wiener-Filter, vgl. Press et al. (1995).

Bei sehr starkem Rauschen $\sigma_n \rightarrow \infty$ und somit $\sigma \rightarrow \infty$ erhält man für $\sigma^2 r$ die „reverse correlation“ Formel (4.3) aus Abschnitt 4.1.

$$\lim_{\sigma \rightarrow \infty} \sigma^2 r = \text{cov}(x, y).$$

Die „reverse correlation“-Methode ist also auch bei Input x , der nicht weißes Rauschen ist, im Sinne der Maximum Likelihood optimal, falls die Messdaten y stark (mit Gauß'schem weißem Rauschen) verrauscht sind. Man erhält in diesem Fall aufgrund des Vorfaktors σ^2 allerdings nur die *Form* des rezeptiven Feldes, nicht seine richtige Größe im Sinne des Absolutbetrags.

4.3 Regularisierungsmethode

Die Regularisierungsmethode ist eine Verallgemeinerung des Bayes'schen Schätzers aus Abschnitt 4.2. Statt die L aus (4.8) zu maximieren, wird nun

$$L = \int_{-\infty}^{\infty} [y(t) - (r \star x)(t)]^2 dt + \sigma^2 \int_{-\infty}^{\infty} (g \star r)(t)^2 dt \quad (4.10)$$

minimiert, d. h. der quadratische Fehler zwischen Messung und Approximation plus ein mit σ^2 gewichteter sogenannter „Strafterm“ $\int (g \star r)(t)^2$, die L^2 -Norm eines auf das rezeptive Feld angewandten linearen zeitinvarianten Operators $g \star$. Eine übliche Wahl ist $g = \delta^{(d)}$, die d -te Ableitung der Delta-Funktion. Man erhält dann als Strafterm

$$\begin{aligned} \int_{-\infty}^{\infty} (g \star r)(t)^2 &= \int_{-\infty}^{\infty} (\delta^{(d)} \star r)(t) = \\ &= \int_{-\infty}^{\infty} r^{(d)}(t)^2, \end{aligned}$$

also die L^2 -Norm der d -ten Ableitung. Ein derartiger Strafterm bewirkt, dass sowohl der quadratische Fehler $\int [y(t) - (r \star x)(t)]^2 dt$ zwischen Messung y und Approximation $r \star x$ klein, als auch das rezeptive Feld r möglichst „glatt“ wird im Sinne der d -ten Ableitung. Falls $n = 0$ wird das rezeptive Feld im quadratischen Mittel möglichst klein gewählt. Der Parameter σ^2 heißt Regularisierungsparameter. Er gewichtet den Strafterm und stellt somit die Art des Kompromisses zwischen möglichst genauer Reproduktion der Messung

und möglichst glattem bzw. kleinem rezeptivem Feld dar. Ist der Regularisierungsparameter $\sigma^2 = 0$, so fällt der Strafterm weg und die Regularisierungsmethode geht in die Methode der optimalen linearen Näherung aus Abschnitt 4.1 über bzw. in die Maximum-Likelihood-Methode aus Abschnitt 4.2 ohne a-priori Information über das rezeptive Feld. Ist $n = 0$, so ergibt sich der Bayes'sche Schätzer mit a-priori Information und $\sigma_n/\sigma_x = \sigma$, wie man aus dem Vergleich der Extremierungsprobleme (4.8) und (4.10) sieht. Ist der Regularisierungsparameter σ^2 zu groß ($\sigma^2 \rightarrow \infty$), so spielt die eigentliche Messung y kaum mehr eine Rolle, da in erster Linie der Strafterm minimiert wird, d. h. $r^{(d)}(t) = 0$ für alle t . Ist $n = 0$, bedeutet dies $r = 0$, d. h. das rezeptive Feld verschwindet. Im Fall $n \geq 1$ wird das rezeptive Feld r zum Polynom $n - 1$ -ten Grades

$$r(t) = \sum_{i=0}^{n-1} a_i t^i,$$

wobei die a_i noch an die Messung angepasst werden, so dass $\int [y(t) - (r \star x)(t)]^2 dt$ minimal wird. Ein großer Regularisierungsparameter σ^2 bewirkt also eine Reduktion der Anzahl der effektiven Parameter von unendlich (r enthält unendlich viele Zeiten) auf n . Somit kann man sich eine bessere Generalisierung erhoffen, d. h. die Abweichungen zwischen *neuen* Messungen y (mit evtl. neuer Eingabe x) und Vorhersagen $r \star x$ sollten mit zunehmendem Regularisierungsparameter σ^2 zunächst kleiner werden. Da bei größer werdendem Regularisierungsparameter σ^2 der Raum der möglichen Lösungen für das rezeptive Feld r jedoch zunehmend auf Polynome eingeschränkt wird, ist bei zu großem Regularisierungsparameter i. d. R. wieder eine Zunahme des Generalisierungsfehlers zu erwarten. Eine Möglichkeit, einen „geeigneten“ Regularisierungsparameter zu wählen ist, aus den Messdaten zwei Teile zu bilden. Der erste Teil dient dazu, das rezeptive Feld r_σ in Abhängigkeit des Regularisierungsparameters σ^2 zu berechnen, der zweite Teil y_2 dient zur Kontrolle des Generalisierungsfehlers $\int [y_2(t) - (r_\sigma \star x_2)(t)]^2 dt$. Es wird dann der Regularisierungsparameter σ^2 so gewählt, dass der Generalisierungsfehler zufriedenstellend klein ist. Bleibt der Generalisierungsfehler dabei unabhängig von σ^2 zu groß, dann ist das (lineare) Modell keine geeignete Beschreibung des Neurons.

Trotz obiger suggestiver Argumente bleibt der Ansatz der Regularisierungsmethode, einen Ausdruck der Form L aus (4.10) zu minimieren, ad-hoc. Not-

wendige Bedingung für ein Minimum ist

$$\begin{aligned} \left. \frac{\delta L}{\delta r(\tau)} \right|_{r=\hat{r}} &= 0 \\ -2 \int [y(t) - (\hat{r} \star x)(t)] x(t - \tau) dt + 2\sigma^2 \int (g \star \hat{r})(t) g(t - \tau) &= 0 \\ -\text{cov}(x, y) + \text{cov}(x, x) \star \hat{r} + \sigma^2 \text{cov}(g, g) \star \hat{r} &= 0. \end{aligned}$$

Fouriertransformation ergibt

$$X^*Y + |X|^2 \hat{R} + \sigma^2 |G|^2 \hat{R} = 0,$$

wobei Großbuchstaben wieder die Fouriertransformierten der entsprechenden Funktionen bezeichnen. Falls $|X(\omega)| + |G(\omega)| > 0$ für alle ω , ist diese Gleichung eindeutig nach \hat{R} auflösbar.

$$\hat{R} = \frac{X^*Y}{|X|^2 + \sigma^2 |G|^2} \quad (4.11)$$

Für $n = 0$ bzw. $g = \delta$ ist $G = 1$, somit ergibt sich in diesem Fall, wie schon oben gezeigt, die gleiche Lösung (4.9) wie beim Bayes'schen Schätzer aus Abschnitt 4.2.

Mit der Wahl $g = \delta^{(d)}$ erhält man $|G(\omega)| = \omega^n$ (vgl. Abschnitt B.6), also

$$\hat{R} = \frac{X^*Y}{|X|^2 + \sigma^2 \omega^{2d}}$$

Hohe Frequenzen in $\mathcal{F}\text{cov}(x, y) = X^*Y$ werden somit tendenziell zunehmend durch den Term $\sigma^2 |G|^2 = \sigma^2 \omega^{2d}$ gedämpft.

Ist $|G(\omega)| > 0$ für alle Frequenzen ω , so enthält r Frequenzen, welche die Eingabe x nicht enthält, ebenfalls nicht, denn $R(\omega) = 0$ für alle ω mit $X(\omega) = 0$, da dann der Zähler in (4.11) $X(\omega)^*Y(\omega) = 0$ ist. Dies liegt daran, dass Frequenzen, die in der Eingabe x nicht enthalten sind, bei einem linearen System natürlich auch in der Ausgabe y nicht enthalten sein können. Zumindest kann eine Frequenz ω , die nicht in der Eingabe x ist ($X(\omega) = 0$), keine Information über $R(\omega)$ enthalten, da die Ursache dieser Frequenz in der Ausgabe y ($Y(\omega) \neq 0$) nicht das rezeptive Feld r , sondern allenfalls Rauschen in den Messdaten sein kann. Man sollte also bei einer Messung dafür sorgen, dass x alle Frequenzen enthält. Ansonsten erhält man zwar ein rezeptives Feld r , das gute Vorhersagen bei gleicher Frequenzzusammensetzung des Stimulus x macht, das jedoch manche „Teile“, d. h. Frequenzen, nicht enthalten kann, unabhängig vom „echten“ rezeptiven Feld.

Bei $g = \delta$ ist $|G(\omega)| > 0$ für alle Frequenzen ω erfüllt. Falls $g = \delta^{(d)}$ mit $n \geq 1$ ist allerdings $|G(\omega)| \geq 0$ mit $G(\omega) = 0$ genau für $\omega = 0$. Zumindest die Konstante sollte die Eingabe x also enthalten.

Die Regularisierungsmethode hat gegenüber dem Bayes'schen Schätzer aus Abschnitt 4.2 den Vorteil, dass ein weiterer Parameter g neben σ zur Verfügung steht, um den Generalisierungsfehler zu kontrollieren. Mit der Wahl $g = \delta^{(d)}$ wird dieser Parameter $n \in \mathbb{N}_0^+$ diskret, er bringt also wenige zusätzliche Komplikationen².

4.4 Minimale Varianz

Gewählt wird das rezeptive Feld \hat{r} , das die Varianz

$$L = \mathbb{E} \left\{ \int_{-\infty}^{\infty} [r(t) - \hat{r}(t)]^2 dt \right\} \tag{4.12}$$

zwischen „echtem“ rezeptivem Feld r und der Schätzung \hat{r} minimiert. Voraussetzung ist, dass die Messung y mit weißem Gauß'schen Rauschen n mit Varianz σ_n^2 verfälscht ist, d. h.

$$y = r \star x + \sigma_n n,$$

und das rezeptive Feld r selbst a-priori wie Gauß'sches weißes Rauschen mit Varianz σ_r^2 verteilt ist und die gleichen Unabhängigkeitsannahmen wie beim Bayes'schen Schätzer aus Abschnitt 4.2 gelten.

Dieser Ansatz ist deshalb vielversprechend, weil er tatsächlich eine Rekonstruktion des rezeptiven Feldes versucht, wobei der Erwartungswert des quadratischen Fehlers, gemittelt über mehrere Messungen (x, y) , minimal sein soll.

Nun ist jedoch die Lösung mit minimaler Varianz unter den getroffenen Annahmen auch die wahrscheinlichste Lösung. Da r und n Gauß'sch verteilt sind, ist es auch y , denn Linearkombinationen Gauß'scher Zufallsgrößen sind wieder Gauß'sche Zufallsgrößen. Da somit die gemeinsame Verteilung (y, r) von y und r Gauß'sch ist, ist es auch die Projektion $p(r|y)$. Somit ist dasjenige \hat{r} , das $p(r|y)$ wie in Abschnitt 4.2 maximiert, zugleich dasjenige, das L aus (4.12) minimiert. Die Methode der minimalen Varianz ist also mit der Bayes'schen Methode äquivalent. Es gilt also auch hier (4.9)

$$\hat{R} = \frac{X^*Y}{|X|^2 + \sigma^2} \tag{4.13}$$

²Man könnte daran denken, in der Formel (4.11) für $g = \delta^{(d)}$, $\hat{R} = \frac{X^*Y}{|X|^2 + \sigma^2 \omega^{2d}}$, $n \in \mathbb{R}_0^+$ zu wählen. Zu gebrochenzahligen Ableitungen siehe Weisstein (1999).

mit $\sigma = \sigma_n / \sigma_r$.

4.5 Optimale lineare Schätzung

Gleichungen (4.9) und (4.13) lassen sich auch schreiben als

$$\hat{r} = s \star y \quad (4.14)$$

mit

$$S = \frac{X^*}{|X|^2 + \sigma^2}.$$

Die Schätzung \hat{r} des rezeptiven Feldes ist also ein lineares zeitinvariantes Funktional der Messung y . Somit stellt sich die Frage, welche Schätzung \hat{r} wir erhalten, wenn wir die Annahmen der Gaußverteilung fallen lassen und dafür nur fordern, dass das geschätzte rezeptive Feld \hat{r} ein lineares zeitinvariantes Funktional der Messung y sein soll. Dass das Funktional zeitinvariant sein muss ist klar, denn eine um die Zeit Δt verzögerte Messung $y'(t) = y(t - \Delta t)$ sollte auch ein entsprechend um Δt verzögertes rezeptives Feld $\hat{r}'(t) = \hat{r}(t - \Delta t)$ zur Folge haben, d. h. $\hat{r}(x, y') = \hat{r}'$, falls $\hat{r}(x, y) = \hat{r}$. Dass das Funktional linear sein soll ist eine echte Einschränkung. Sie ist sinnvoll, wenn eine effiziente Methode gesucht ist, das rezeptive Feld zu schätzen.

Setzen wir also für die Schätzung \hat{r} ein lineares zeitinvariantes Funktional $s \star$ der Messung y nach (4.14) an. „Beste“ Schätzung sei wieder diejenige mit minimaler Varianz. Zu minimieren ist also

$$L = \mathbb{E} \left\{ \int_{-\infty}^{\infty} [r(t) - \hat{r}(t)]^2 dt \right\}.$$

Gleichung (4.14) eingesetzt ergibt

$$L = \mathbb{E} \left\{ \int_{-\infty}^{\infty} [r(t) - (s \star y)(t)]^2 dt \right\}. \quad (4.15)$$

Für die Ausgabe y des „echten“ Neurons soll gelten

$$y = r \star x + n, \quad (4.16)$$

wobei diesmal im Ggs. zu (4.4) des Bayes'schen Schätzers in Abschnitt 4.2 die Messung y durch additives, jedoch nicht notwendigerweise gaußverteiltes

Rauschen n mit Mittelwert $E(n) = 0$ verfälscht wird. Die Messung y aus (4.16) in (4.15) eingesetzt ergibt

$$L = E \left\{ \int_{-\infty}^{\infty} [r(t) - (s \star (r \star x + n))(t)]^2 dt \right\}.$$

Notwendige Bedingung für ein Minimum von L ist

$$\begin{aligned} \frac{\delta L}{\delta s(\tau)} &= 0 \\ E \left\{ \int -2[r(t) - (s \star (r \star x + n))(t)](r \star x + n)(t - \tau) dt \right\} &= 0 \\ E \{ \text{cov}(r \star x, r) - \text{cov}(r \star x, s \star r \star x) - \text{cov}(n, s \star n) \} &= 0. \end{aligned}$$

Hierbei wurde die Unabhängigkeit des Rauschens n und seine Mittelwertfreiheit $E(n) = 0$ verwendet.

Mit Hilfe der Gesetzmäßigkeiten (A.9), (A.10) und (A.11) aus Abschnitt A erhält man

$$\begin{aligned} E \{ \text{cov}[x, \text{cov}(r, r)] - \text{cov}(r \star x, r \star x) \star s - \text{cov}(n, n) \star s \} &= 0 \\ E \{ \text{cov}[x, \text{cov}(r, r)] - s \star [\text{cov}(r, r) \star \text{cov}(x, x)] - s \star \text{cov}(n, n) \} &= 0. \end{aligned}$$

Da der Erwartungswert $E(\cdot)$ linear ist und die einzigen zufallsbehafteten Größen nach Voraussetzung r und n sind, gilt

$$\text{cov}[x, E\text{cov}(r, r)] - s \star [E\text{cov}(r, r) \star \text{cov}(x, x)] - s \star E\text{cov}(n, n) = 0.$$

Fouriertransformation ergibt

$$X^*E(|R|^2) - SE(|R|^2)|X|^2 - SE(|N|^2) = 0.$$

Hinreichende Bedingung für die eindeutige Auflösbarkeit dieser Gleichung nach S ist $E(|N(\omega)|^2) \neq 0$ für alle Frequenzen ω , d. h. das Rauschen n kann beliebige Frequenzen enthalten. Man erhält³

$$S = \frac{X^*E(|R|^2)}{E(|R|^2)|X|^2 + E(|N|^2)}$$

³Das durch seine Übertragungsfunktion S beschriebene Filter ist bekannt als ein Bildwiederherstellungsfiler (Cattermole 1988). Wir wollen hier jedoch nicht Bilder wieder herstellen, sondern rezeptive Felder.

oder mit (4.14) und unter der Voraussetzung $E(|R(\omega)|^2) \neq 0$ für alle ω , d. h. auch das rezeptive Feld r kann beliebige Frequenzen enthalten,

$$\hat{R} = \frac{X^*Y}{|X|^2 + E(|N|^2)/E(|R|^2)}. \quad (4.17)$$

Somit wird bei diesem Ansatz σ^2 aus (4.9) und (4.13) bzw. $\sigma^2|G|^2$ aus (4.11) durch den allgemeineren Ausdruck $E(|N|^2)/E(|R|^2)$ ersetzt.

Ist n und r verteilt wie Gauß'sches weißes Rauschen mit Varianz σ_n^2 bzw. σ_r^2 , dann ist $E(|N|^2)/E(|R|^2) = \sigma_n^2/\sigma_r^2$ und es ergeben sich die Gleichungen (4.9) und (4.13) des Bayes'schen Schätzers aus Abschnitt 4.2 bzw. der minimalen Varianz aus Abschnitt 4.4. Dies wissen wir bereits, da unter den Voraussetzungen der beiden Abschnitte sich die lineare Beziehung (4.14) schon als optimal erwiesen hat.

Nehmen wir nun an, n sei weiterhin Gauß'sches weißes Rauschen mit Varianz σ_n^2 , jedoch die d -te Ableitung $r^{(d)}$ von r sei verteilt wie Gauß'sches weißes Rauschen mit Varianz σ_r^2 .⁴ Es soll nun der Erwartungswert $E(|R|^2)$ der Fouriertransformierten der Autokovarianz $\text{cov}(r, r)$ von r berechnet werden. Ausgehend von

$$E\text{cov}(r^{(d)}, r^{(d)}) = \sigma_r^2 \delta,$$

da $r^{(d)}$ nach Voraussetzung verteilt ist wie weißes Rauschen, vgl. (A.2), erhält man wegen (B.13) $r^{(d)} = \delta^{(d)} \star r$ und der Autokovarianz von Faltungen (A.11)

$$\begin{aligned} E\text{cov}(\delta^{(d)} \star r, \delta^{(d)} \star r) &= \sigma_r^2 \delta \\ E\text{cov}(r, r) \text{cov}(\delta^{(d)}, \delta^{(d)}) &= \sigma_r^2 \delta. \end{aligned}$$

Durch Fouriertransformation ergibt sich wegen (B.6) und (B.6)

$$\begin{aligned} E(|R|^2)|(-i\omega)^n|^2 &= \sigma_r^2 \\ E(|R|^2) &= \frac{\sigma_r^2}{\omega^{2d}}. \end{aligned}$$

Dieses Ergebnis für $E(|R|^2)$ und $E(|N|^2) = \sigma_n^2$ in (4.17) eingesetzt erhalten wir

$$\hat{R} = \frac{X^*Y}{|X|^2 + \sigma_n^2 \omega^{2d}},$$

also das gleiche Ergebnis (4.11) wie bei der Regularisierungsmethode aus Abschnitt 4.3 bei der Wahl von $g = \delta^{(d)}$. Dieses Ergebnis ermöglicht eine Interpretation des Strafterms, der bei der Regularisierungsmethode ad-hoc

⁴Ist die erste Ableitung r' von r Gauß'sches weißes Rauschen, so ist r eine Brown'sche Bewegung (Lütkepohl 2005).

eingeführt wird: Der Strafterm kontrolliert die a-priori Verteilung von r . Jedoch welche Verteilung vom rezeptiven Feld r sollte man a-priori annehmen, um den Erwartungswert der Autokovarianz $E[\text{cov}(r, r)]$ zu berechnen? Um $E[\text{cov}(r, r)]$ zu schätzen, müsste man schon eine Reihe von rezeptiven Feldern gemessen haben. Dies kann man nach (4.17) jedoch erst, wenn man $E[\text{cov}(r, r)]$ kennt. Es handelt sich hierbei also um das klassische Huhn-und-Ei-Problem⁵.

4.6 Bayes'scher Schätzer unter erweiterten Verteilungsannahmen

Wir gehen wie beim Bayes'schen Schätzer mit Gauß'schem weißem Rauschen in Abschnitt 4.2 davon aus, dass das „echte“ Neuron linear ist, für seinen Output y bei Input x also wieder

$$E(y) = r \star x.$$

Weiter sei der Output y mit additivem, zeitinvariantem, Gauß'schen (nicht notwendigerweise weißem) Rauschen n verfälscht, also

$$y = r \star x + n. \tag{4.18}$$

Weiter sei r a-priori verteilt wie zeitinvariantes Gauß'sches (nicht notwendigerweise weißes) Rauschen. Wie bei der Bayes'schen Methode mit weißem Rauschen in Abschnitt 4.4 maximieren wir die Größe

$$L = \ln[p(y|r)p(r)] \tag{4.19}$$

⁵Was war zuerst da? Die Henne oder das Ei?

Ein iteratives oder evolutionäres Verfahren könnte, wie so oft, die Lösung sein. Eine andere mögliche Lösung findet man in (Zink 1966, 1. Mose 1, 20-25).

bezüglich des rezeptiven Feldes r . Da n zeitinvariantes Gauß'sches Rauschen⁶ ist, gibt es eine Funktion c_n , so dass⁷

$$p(n) \sim \exp \left[- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(t) c_n(t-t') n(t') dt dt' \right].$$

Wegen (4.18) gilt $p(y|r) = p(n)$. Da das rezeptive Feld r a-priori wie zeitinvariantes Gauß'sches Rauschen verteilt sein soll, gibt es eine Funktion c_r , so dass

$$p(r) \sim \exp \left[- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(t) c_r(t-t') r(t') dt dt' \right].$$

⁶d. h. das Rauschen ist

- *stationär*, also für alle $t_1 < t_2 < \dots < t_d$ und $\Delta t > 0$ sind die d -dimensionalen Zufallsvektoren $(n(t_1), n(t_2), \dots, n(t_d))$ und $(n(t_1 + \Delta t), n(t_2 + \Delta t), \dots, n(t_d + \Delta t))$ identisch verteilt, gemeinsame Verteilungen der Prozessgröße sind also zeitinvariant und
- *Gauß'sch*, also für alle $t_1 < t_2 < \dots < t_d$ ist der n -dimensionale Zufallsvektor $(n(t_1), n(t_2), \dots, n(t_d))$ multivariat normalverteilt.

Für einen Ornstein-Uhlenbeck-Prozess (Gauß-Markov-Prozess) würde man noch zusätzlich fordern, dass die zukünftige Entwicklung des Prozesses nicht von der Vergangenheit abhängt (Finch 2004).

⁷Der d -dimensionale Zufallsvektor $\mathbf{n} := (n(t_1), \dots, n(t_d))$ ist nach Voraussetzung multivariat normalverteilt, es gibt also eine Funktion c_n , so dass für die Wahrscheinlichkeitsdichte gilt

$$p(\mathbf{n}) \sim \exp \left(- \sum_{i,j=1}^d n(t_i) c_n(t_i, t_j) n(t_j) \right)$$

Da wegen der Voraussetzung der Zeitinvarianz die d -dimensionalen Zufallsvektoren $\mathbf{n} = (n(t_1), n(t_2), \dots, n(t_d))$ und $(n(t_1 + \Delta t), n(t_2 + \Delta t), \dots, n(t_d + \Delta t))$ identisch verteilt sind, muss gelten

$$c_n(t_i, t_j) = c_n(t_i + \Delta t, t_j + \Delta t),$$

somit hängt c_n nur von der Zeitdifferenz ab, also $c_n(t_i, t_j) = c_n(t_i - t_j)$. Der Grenzübergang für unendlich viele Zeitschritte ergibt

$$p(n) \sim \exp \left[- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(t) c_n(t-t') n(t') dt dt' \right].$$

Wir denken uns dabei jede uns interessierende Realisierung n des Rauschens als zeitlich beschränkt, also $n = 0$ außerhalb von etwa $[0, T]$, so dass $p(n) \neq 0$. Diese Einschränkung können wir machen, da wir dies in der Realität immer mit endlich langen Messdauern zu tun haben.

Setzen wir die berechneten Wahrscheinlichkeiten $p(y|r)$ und $p(r)$ in (4.19) ein, so erhalten wir

$$L \sim - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(t)c_n(t-t')n(t') dt dt' - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(t)c_r(t-t')r(t') dt dt' .$$

Notwendige Bedingung für ein Minimum an der Stelle \hat{r} ist

$$\left. \frac{\delta L}{\delta r(\tau)} \right|_{r=\hat{r}} = 0$$

$$-2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left. \frac{\delta n}{\delta r(\tau)} \right|_{r=\hat{r}} c_n(t-t')n(t') dt dt' - 2 \int_{-\infty}^{\infty} c_r(\tau-t')\hat{r}(t') dt' = 0 .$$

Wegen (4.18) gilt

$$n = y - r \star x ,$$

somit

$$\frac{\delta n}{\delta r(\tau)} = -x(t-\tau) .$$

Eingesetzt in die notwendige Bedingung für ein Minimum ergibt dies

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} -x(t-\tau)c_n(t-t')(y-\hat{r}\star x)(t') dt dt' + \int_{-\infty}^{\infty} c_r(\tau-t')\hat{r}(t') dt' = 0$$

$$\int_{-\infty}^{\infty} -x(t-\tau)(c_n \star y - c_n \star \hat{r} \star x)(t) dt + c_r \star \hat{r} = 0$$

$$-\text{cov}(x, c_n \star y) + \text{cov}(x, c_n \star \hat{r} \star x) + c_r \star \hat{r} = 0 .$$

Fouriertransformation ergibt

$$\begin{aligned} -X^*C_nY + X^*C_n\hat{R}X + C_r\hat{R} &= 0 \\ -C_nX^*Y + C_n|X|^2\hat{R} + C_r\hat{R} &= 0 \end{aligned}$$

Eine hinreichende Bedingung für die eindeutige Auflösbarkeit dieser Gleichung nach \hat{R} ist $C_r(\omega) \neq 0$ für alle Frequenzen ω .

$$\hat{R} = \frac{C_nX^*Y}{C_n|X|^2 + C_r}$$

Mit $C_n = 1/E(|N|^2)$ und $C_r = 1/E(|R|^2)$ ergibt sich, falls $E(|N(\omega)|^2) \neq 0$ für alle Frequenzen ω ,

$$\hat{R} = \frac{X^*Y}{|X|^2 + E(|N|^2)/E(|R|^2)} .$$

Dieses Ergebnis ist identisch mit (4.17) aus Abschnitt 4.5. Der lineare Ansatz (4.14) ist also im Sinne des Bayes'schen Schätzers optimal, auch unter erweiterten a-priori Verteilungsannahmen.

4.7 Optimale Vorhersage

Das optimale rezeptive Feld sei nun dasjenige, das den wahren neuronalen Output y , gegeben den Input x , im Mittel am besten *vorhersagt*, das also

$$E \{ [y(t) - \hat{y}(t)]^2 \} \quad (4.20)$$

minimiert.

Der approximierte Output sei angesetzt mit

$$\hat{y} = \hat{r} \star x, \quad (4.21)$$

wobei r das rezeptive Feld ist.

Das „echte Neuron“ sei wieder linear, d. h. es gibt ein rezeptives Feld r mit

$$E(y) = r \star x.$$

Die Messdaten seien mit Gauß'schem Rauschen $\sigma_n n$ mit Varianz σ_n^2 verwechselt, d. h.

$$y = r \star x + \sigma_n n \quad (4.22)$$

Um die mittlere quadratische Abweichung aus (4.20) zu berechnen, benötigen wir die Wahrscheinlichkeit $p(y|y_0)$, ein bestimmtes Messergebnis y zu erzielen, gegeben die Messung y_0 . Es ist

$$p(y|y_0) = \int p(y|r, y_0)p(r, y_0) dr = \int p(y|r, y_0)p(y_0|r)p(r) dr. \quad (4.23)$$

Die a-priori Wahrscheinlichkeit $p(r)$ für ein bestimmtes rezeptives Feld sei wieder verteilt wie Gauß'sches Rauschen mit Varianz σ_r^2 , also

$$p(r) \sim \exp \left[-\frac{\int_{-\infty}^{\infty} r(t)^2 dt}{2\sigma_r^2} \right].$$

Die Wahrscheinlichkeit $p(y_0|r)$, ein bestimmtes Messergebnis y_0 zu erzielen, gegeben das „echte“ rezeptive Feld r , ist wegen (4.22)

$$p(y_0|r) \sim \exp \left[-\frac{\int_{-\infty}^{\infty} (y_0 - r \star x)(t)^2 dt}{2\sigma_n^2} \right].$$

Die Wahrscheinlichkeit $p(y|r, y_0) = p(y|r)$, für eine bestimmte Messung gegeben das rezeptive Feld r ist, ebenfalls wegen (4.22),

$$p(y|r) \sim \exp \left[-\frac{\int_{-\infty}^{\infty} (y - r \star x)(t)^2 dt}{2\sigma_n^2} \right].$$

Da die Wahrscheinlichkeit $p(y|y_0)$ aus analogen Gründen wie in Abschnitt 4.4 Gauß'sch ist, genügt es, \hat{r} aus (4.21) so zu wählen, dass \hat{y} die wahrscheinlichste Messung ist, also $p(y|y_0)$ aus (4.23) maximiert. Dann wird die mittlere quadratische Abweichung zwischen geschätzter Messung \hat{y} und weiteren, zukünftigen, Messungen y minimiert. Da $p(y|r, y_0)p(y_0|r)p(r)$ eine multivariat Gauß'sch in (y, r) ist, genügt es wegen (4.23)

$$L = \ln[p(y|r, y_0)p(y_0|r)p(r)]$$

gemeinsam nach y und r zu maximieren. Die notwendigen Bedingungen für ein Maximum an der Stelle (y_{\max}, r_{\max}) sind

$$\left. \frac{\delta L}{\delta y(\tau)} \right|_{y=y_{\max}} = 0 \tag{4.24}$$

$$\left. \frac{\delta L}{\delta r(\tau)} \right|_{r=r_{\max}} = 0. \tag{4.25}$$

Mit

$$L = -\frac{\int_{-\infty}^{\infty} r(t)^2 dt}{2\sigma_r^2} - \frac{\int_{-\infty}^{\infty} (y_0 - r \star x)(t)^2 dt}{2\sigma_n^2} - \frac{\int_{-\infty}^{\infty} (y - r \star x)(t)^2 dt}{2\sigma_n^2},$$

bis auf eine von y und r unabhängige additive Konstante, ergeben sich aus den Maximalitätsbedingungen (4.24) und (4.25)

$$\begin{aligned} (y_{\max} - r_{\max} \star x)(\tau) &= 0 \\ -\sigma r_{\max}(\tau) + \int (y_0 - r_{\max} \star x)(t)x(t - \tau) dt + \\ + \int (y_{\max} - r_{\max} \star x)(t)x(t - \tau) dt &= 0 \end{aligned}$$

mit $\sigma := \sigma_n/\sigma_r$, somit

$$y_{\max} = r_{\max} \star x \tag{4.26}$$

$$\begin{aligned} -\sigma^2 r_{\max} + \text{COV}(x, y_0 - r_{\max} \star x) + \\ + \text{COV}(x, y_{\max} - r_{\max} \star x) &= 0. \end{aligned} \tag{4.27}$$

Gleichung (4.26) in (4.27) eingesetzt ergibt

$$-\sigma^2 r_{\max} + \text{cov}(x, y_0 - r_{\max} \star x) = 0.$$

Durch Fouriertransformation erhält man

$$-\sigma^2 R_{\max} + X^* Y_0 - X^* R_{\max} X = 0.$$

Falls $\sigma > 0$ ist die eindeutige Lösung

$$R_{\max} = \frac{X^* Y_0}{|X|^2 + \sigma^2}.$$

Wählt man also $\hat{y} = y_{\max}$ mit y_{\max} aus (4.26) mit oben angegebenem R_{\max} , so wird der erwartete quadratische Fehler aus (4.20) minimiert. Wegen (4.21) ist dann, jedenfalls falls x alle Frequenzen enthält, $\hat{r} = r_{\max}$, somit

$$\hat{R} = \frac{X^* Y_0}{|X|^2 + \sigma^2}.$$

Man erhält also wieder die bekannte Gleichung (4.9) bzw. (4.13).

4.8 Zusammenfassung

Aus den Überlegungen aus den vorangegangenen Abschnitten ergibt sich die Tabelle 4.1. Die allgemeinste Formel für die Fouriertransformierte \hat{R} des geschätzten rezeptiven Feldes \hat{r} ergibt sich aus dem optimalen linearen Schätzer von Abschnitt 4.5 und der Bayes'schen Methode unter erweiterten Voraussetzungen aus Abschnitt 4.6. Bezeichnen wir mit Y_0 die Fouriertransformierte der Messung y_0 der neuronalen Antwort, mit X die Fouriertransformierte der Eingabe x , mit $\langle |N|^2 \rangle$ das zu erwartende Betragsquadrat der Fouriertransformierten N des die Messung y verfälschenden additiven Rauschens n und mit $\langle |R|^2 \rangle$ das zu erwartende Betragsquadrat der Fouriertransformierten R des rezeptiven Feldes, dann lautet sie

$$\hat{R} = \frac{X^* Y_0}{|X|^2 + \frac{\langle |N|^2 \rangle}{\langle |R|^2 \rangle}}.$$

Für starkes Rauschen, also $\langle |N|^2 \rangle \rightarrow \infty$, geht die Formel über in die Kovarianz

$$\hat{r} \sim \text{cov}(x, y_0)$$

zwischen Eingabe x und Messung y_0 . Dies ist wieder die „reverse covelation“ Methode.

Methoden	\hat{r}	Voraussetzungen	\hat{R}
Lineare Näherung	$\operatorname{argmin}_r \ y_0 - r \star x\ ^2$		$\frac{X^*Y_0}{ X ^2}$
Bayes	$\operatorname{argmax}_r p(r y_0)$	$y = r \star x + \sigma_n n$ $n \sim N(1)$ $r \sim N(\sigma_r^2)$	$\frac{X^*Y_0}{ X ^2 + \sigma^2}$
Regularisierung	$\operatorname{argmin}_r (\ y_0 - r \star x\ ^2 + \sigma^2 \ g \star r\ ^2)$		$\frac{X^*Y_0}{ X ^2 + \sigma^2 G ^2}$
Regularisierung 2	$\operatorname{argmin}_r (\ y_0 - r \star x\ ^2 + \sigma^2 \ r^{(d)}\ ^2)$		$\frac{X^*Y_0}{ X ^2 + \sigma^2 \omega^{2d}}$
Minimale Varianz	$\operatorname{argmin}_{\hat{r}} \langle \ r - \hat{r}\ ^2 \rangle$	$y = r \star x + \sigma_n n$ $n \sim N(\sigma_n^2)$ $r \sim N(\sigma_r^2)$	$\frac{X^*Y_0}{ X ^2 + \sigma^2}$
Lineare Schätzung	$y_0 \star \operatorname{argmin}_s \langle \ r - s \star y_0\ ^2 \rangle$	$y = r \star x + n$	$\frac{X^*Y_0}{ X ^2 + \frac{\langle N ^2 \rangle}{\langle R ^2 \rangle}}$
Bayes 2	$\operatorname{argmax}_r p(r y_0)$	$y = r \star x + n$ n, r zeitinv. und Gauß'sch	$\frac{X^*Y_0}{ X ^2 + \frac{\langle N ^2 \rangle}{\langle R ^2 \rangle}}$
Optimale Vorhersage	$\hat{r} \star x = \operatorname{argmin}_{\hat{y}} \langle \ y - \hat{y}\ ^2 \rangle$	$y = r \star x + \sigma_n n$ $n \sim N(\sigma_n^2)$ $r \sim N(\sigma_r^2)$ $ X > 0$	$\frac{X^*Y_0}{ X ^2 + \sigma^2}$
Reverse Correlation	$\operatorname{cov}(x, y_0)$		X^*Y_0

Tabelle 4.1: Zusammenfassung der Ergebnisse aus Abschnitt 4.

Die Zeichen bedeuten die Eingabe x , die neuronale Ausgabe y , die Messung y_0 der neuronalen Ausgabe, das rezeptive Feld r und deren d -te Ableitung $r^{(d)}$. Die $\|\cdot\|$ bedeuten die L^2 -Norm, also $\|x\|^2 = \int_{-\infty}^{\infty} x(t)^2 dt$. Die Schreibweise $n \sim N(\sigma_n^2)$ bedeutet, das Rauschen n ist verteilt wie Gauß'sches weißes Rauschen mit Mittelwert 0 und Varianz σ_n^2 . Spitze Klammern $\langle \cdot \rangle$ stehen für den Erwartungswert. Die Größe σ ist bei den Regularisierungsmethoden frei wählbar, ansonsten ist sie definiert als $\sigma = \sigma_n / \sigma_r$.

Zu den in der Tabelle angegebenen Voraussetzungen sei zusätzlich immer angenommen, dass die Nenner in \hat{R} für keine Frequenz null werden. Lineare Näherung und Regularisierung kommen ohne weitere Voraussetzungen aus. Für die optimale lineare Schätzung genügt die Voraussetzung, dass die Messung mit *additivem* Rauschen verfälscht ist. Alle anderen Ansätze erfordern die Linearität $\langle y \rangle = r \star x$ des betrachteten Neurons sowie Gauß'sche Zufallsgrößen n und r .

Die Regularisierungsmethode, die lineare Schätzung und Bayes in der erweiterten Fassung ergeben die allgemeinsten Formeln für die Fouriertransformierte \hat{R} des geschätzten rezeptiven Feldes \hat{r} .

Wählt man für die Regularisierungsmethode den Regularisierungsparameter $\sigma = 1$ und ein geeignetes g mit $|G|^2 = \langle |N|^2 \rangle / \langle |R|^2 \rangle$, so kann man die entsprechende Klasse von linearen bzw. Bayes'schen Schätzern nachbilden. Eine geeignete Wahl ist die d -te Ableitung der δ -Funktion $g = \delta^{(d)}$. Dann ergibt sich nach Abschnitt 4.3

$$\hat{R} = \frac{X^*Y}{|X|^2 + \sigma^2\omega^{2d}}.$$

Es steht also neben dem Regularisierungsparameter σ noch ein weiterer diskreter Parameter n zur Verfügung, um a-priori-Information zu verwenden. Diese entspricht der Annahme, dass erst die d -te Ableitung von r verteilt ist wie Gauß'sches weißes Rauschen, dass also das rezeptive Feld r relativ „glatt“ ist. Hätte man weitergehende Statistiken zur Verfügung, was nicht der Fall ist, könnte man mit Kenntnis von $\langle |R|^2 \rangle$ bessere Schätzungen erzielen.

Ist man nicht von der „Glattheit“ des rezeptiven Feldes überzeugt, wähle man $n = 0$. In diesem Fall bedeutet $\sigma > 0$ dass man erwartet, dass das rezeptive Feld mit großer Wahrscheinlichkeit an vielen Stellen Werte nahe Null annimmt. Diese Annahme ist allein dadurch gerechtfertigt, dass lange zurückliegende Schalle das Neuron nicht mehr beeinflussen, dass also $\lim_{t \rightarrow \infty} r(t) = 0$ gilt. Ein Algorithmus nach (4.8) wird sich unter gleich guten Lösungen ein möglichst „kurzes“ rezeptives Feld \hat{r} suchen, d. h. eines, wofür $r(t) = 0$ für möglichst große Zeitabschnitte gilt. „Gleich gute“ Lösungen r sind dabei solche, die minimale Abweichungen $\|y - r \star x\|$ erzielen.

Es wäre wünschenswert, wenn beliebige a-priori Annahmen über r einfließen könnten. Beispiele für solche Annahmen wären „ r hat nur m Maxima“ oder „ $r(t) = 0$ für $t > t_0$ “⁸. Solche Annahmen führen jedoch i. d. R. zu Algorithmen mit unakzeptabel langer Laufzeit, da nichtlineare Optimierungsaufgaben gelöst werden müssten. Die Support-Vektor-Methode aus Kapitel 9 kommt ohne a-priori Annahmen über Verteilungen aus und erzielt dennoch einen geringen Generalisierungsfehler

$$\left\langle \int_{-\infty}^{\infty} |y(x, t) - \hat{y}(y_0, x_0, x, t)| dt \right\rangle.$$

Hierbei ist $y(x, t)$ die neuronale Antwort bei Eingabe x zu Zeit t und $\hat{y}(y_0, x_0, x, t)$ die aus der Messung y_0 mit Eingabe x_0 geschätzte Vorhersage für Eingabe x zur Zeit t . Dabei verwendet die Support-Vektor-Methode ebenfalls zwei Parameter, den Regularisierungsparameter C und den Grad d des Polynoms, vgl. Abschnitt 8.7.

⁸Das ist eine Beschränkung des Trägers von r , also der Menge $\{t|r(t) \neq 0\}$.

Abbildung 4.1 zeigt, unter welchen Bedingungen die Methoden aus Tabelle 4.1 ineinander übergehen. Regularisierung ist die allgemeinste Methode. Da sie jedoch ein ad-hoc-Ansatz ist, erfährt sie ihre Rechtfertigung erst durch ihre Äquivalenz mit den anderen Methoden wie Bayes oder minimale Varianz. Reverse Correlation lässt sich zwar ebenfalls als ad-hoc-Ansatz immer anwenden, ist jedoch nur unter sehr einschränkenden Voraussetzungen, insbesondere $|X| = 1$, korrekt. Die Voraussetzung $|X| = 1$ bedeutet, jede Frequenzkomponente ist im Eingangssignal x mit Gewicht 1 vorhanden. Dies ist beispielsweise erfüllt für einen Impuls $x = \delta$ oder Gauß'sches weißes Rauschen als Eingabe.

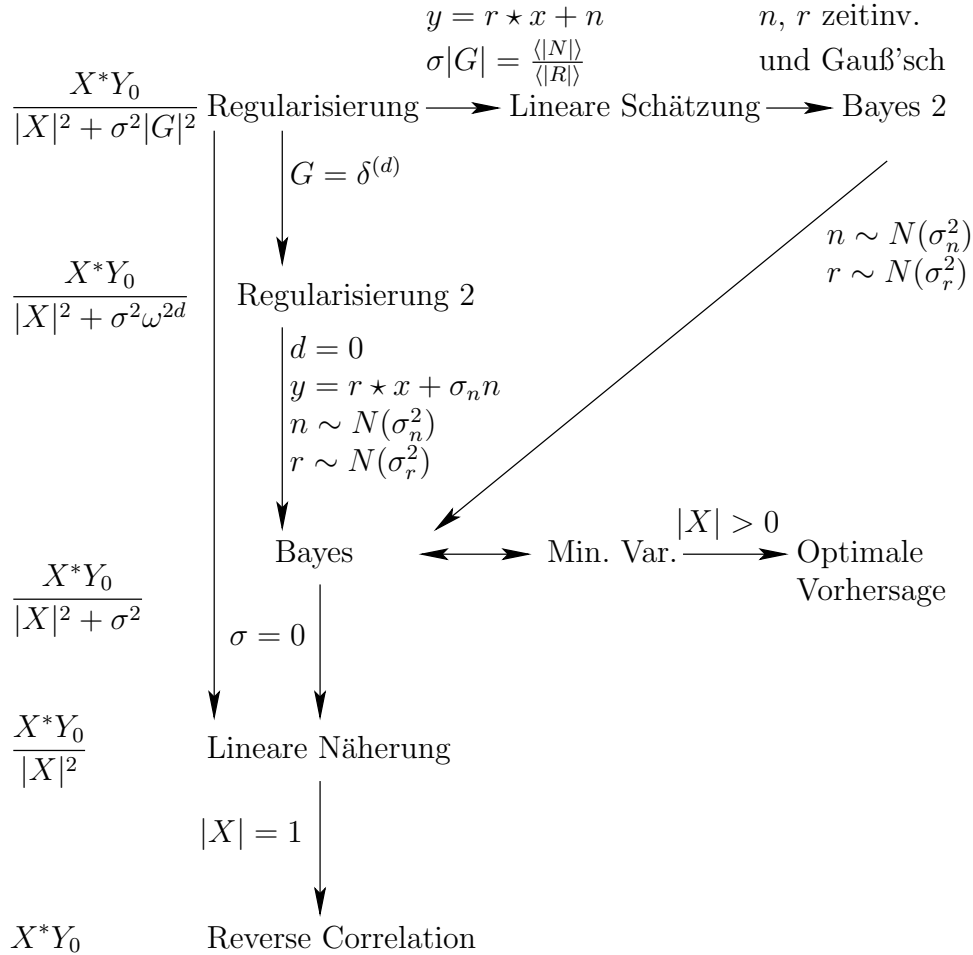


Abbildung 4.1: Zusammenhang zwischen den Methoden aus Tabelle 4.1. Die Pfeile $A \xrightarrow{V} B$ bedeuten, dass Methode A unter den zusätzlichen Voraussetzungen V in Methode B übergeht, d. h. Methode B ist anwendbar und Anwendung von Methode A und Methode B führen zum gleichen Ergebnis. Zu den anderen Zeichen siehe Tabelle 4.1. Methoden mit äquivalenten Verfahren, d. h. äquivalenter Formel auf der linken Seite, stehen auf gleicher Höhe.

Insbesondere kann man durch Wahl einer geeigneten Funktion G immer eine Regularisierungsmethode finden, die jeder der anderen Methoden entspricht. (Die Wahl $G = 1$ und $\sigma \rightarrow \infty$ ergibt die Reverse-Correlation-Methode bis auf einen konstanten (sehr kleinen) Faktor.) Regularisierung ist somit die allgemeinste Methode, Reverse Correlation die speziellste.

*No theory should fit all the facts,
because some of the facts are
wrong.*

Fritz Zwicky

5

Optimales rezeptives Feld in mehreren Dimensionen

Der approximierte Output \hat{y} des Neurons sei nun

$$\hat{y}(t) = \sum_i r_i \star x \quad (5.1)$$

Die Funktionen $r_i(t)$ heißen spektro-temporales rezeptives Feld.

5.1 Optimale lineare Schätzung

Da die optimale lineare Schätzung aus Abschnitt 4.5 die allgemeinste Rekonstruktionsformel (4.17) für das rezeptive Feld liefert, minimieren wir im mehrdimensionalen Fall sogleich

$$L = \mathbb{E} \left\{ \sum_i \int_{-\infty}^{\infty} [r_i(t) - \hat{r}_i(t)]^2 dt \right\}. \quad (5.2)$$

mit dem linearen Ansatz

$$\hat{r}_i = s_i \star y. \quad (5.3)$$

Gleichung (5.3) in (5.2) eingesetzt ergibt

$$L = \sum_i \mathbb{E} \left\{ \int_{-\infty}^{\infty} [r_i(t) - (s_i \star y)(t)]^2 dt \right\}. \quad (5.4)$$

Für die Ausgabe y des „echten“ Neurons soll wieder gelten

$$y = \sum_j r_j \star x_j + n, \quad (5.5)$$

wobei n von r_j unabhängiges, mittelwertfreies Rauschen (nicht notwendigerweise Gauß'sch) ist. Gleichung (5.5) in (5.4) eingesetzt ergibt

$$L = \sum_i \mathbb{E} \left\{ \int_{-\infty}^{\infty} \left\{ r_i(t) - \left[s_i \star \left(\sum_j r_j \star x_j + n \right) \right] (t) \right\}^2 dt \right\}.$$

Notwendige Bedingung für ein Minimum von L bezüglich der Funktionen s_i ist

$$\begin{aligned} \frac{\delta L}{\delta s_i(\tau)} &= 0 \\ \mathbb{E} \left\{ \int -2 \left\{ r_i(t) - \left[s_i \star \left(\sum_j r_j \star x_j + n \right) \right] (t) \right\} \right. \\ &\quad \left. \times \left(\sum_k r_k \star x_k + n \right) (t - \tau) dt \right\} = 0 \\ \mathbb{E} \left\{ \text{cov} \left(\sum_k r_k \star x_k, r_i \right) - \text{cov} \left(\sum_k r_k \star x_k, s_i \star \sum_j r_j \star x_j \right) \right. \\ &\quad \left. - \text{cov}(n, s_i \star n) \right\} = 0. \end{aligned}$$

Hierbei wurde die Unabhängigkeit des Rauschens n und seine Mittelwertfreiheit $E(n) = 0$ verwendet. Die Unterausdrücke in dieser Gleichung sollen nun separat behandelt werden.

Wegen der Rechtsdistributivität der Kovarianz (A.1) und dem gemischten Assoziativgesetz zwischen Kovarianz und Faltung (A.10) gilt

$$\text{cov} \left(\sum_k r_k \star x_k, r_i \right) = \sum_k \text{cov} (r_k \star x_k, r_i) = \sum_k \text{cov} [x_k, \text{cov}(r_k, r_i)].$$

Mit den Distributivgesetzen (A.1) und (A.4), dem gemischten Assoziativgesetz (A.9) sowie der Kovarianz von Faltungen (A.2) erhält man

$$\begin{aligned}
\text{cov} \left(\sum_k r_k \star x_k, s_i \star \sum_j r_j \star x_j \right) &= \\
&= \text{cov} \left(\sum_k r_k \star x_k, \sum_j r_j \star x_j \right) \star s_i = \\
&= s_i \star \sum_{j,k} \text{cov}(r_k \star x_k, r_j \star x_j) \\
&= s_i \star \sum_{j,k} \text{cov}(x_k, x_j) \star \text{cov}(r_k, r_j).
\end{aligned}$$

Mit dem gemischten Assoziativgesetz (A.9) ergibt sich

$$\text{cov}(n, s_i \star n) = s_i \star \text{cov}(n, n).$$

Alle berechneten Unterausdrücke eingesetzt ergibt

$$\begin{aligned}
\mathbb{E} \left\{ \sum_k \text{cov}[x_k, \text{cov}(r_k, r_i)] - s_i \star \sum_{j,k} \text{cov}(x_k, x_j) \star \text{cov}(r_k, r_j) \right. \\
\left. - s_i \star \text{cov}(n, n) \right\} = 0.
\end{aligned}$$

Die einzigen zufallsbehafteten Größen sind das Rauschen n und r_i . Deshalb ergibt sich aufgrund der Linearität des Erwartungswerts

$$\sum_k \text{cov}[x_k, \text{E} \text{cov}(r_k, r_i)] - s_i \star \sum_{j,k} \text{cov}(x_k, x_j) \star \text{E} \text{cov}(r_k, r_j) - s_i \star \text{E} \text{cov}(n, n) = 0.$$

Durch Fouriertransformation erhält man

$$\sum_k X_k^* \text{E}(R_k^* R_i) - S_i \sum_{j,k} X_j^* X_k \text{E}(R_j^* R_k) - S_i \text{E}(|N|^2) = 0.$$

Eine hinreichende Bedingung für die Auflösbarkeit dieser Gleichung nach S_j ist schon, dass das Rauschen n alle Frequenzen enthalten kann, also $N(\omega) \neq 0$ für alle Frequenzen ω . Es gilt

$$S_i = \frac{\sum_k \text{E}(R_k^* R_i) X_k^*}{\sum_{j,k} X_j^* X_k \text{E}(R_j^* R_k) + \text{E}(|N|^2)}.$$

Aufgrund des linearen Ansatzes (5.3) erhalten wir für die Fouriertransformierte \hat{R}_i des geschätzten rezeptiven Feldes \hat{r}_i

$$\hat{R}_i = \frac{\sum_k E(R_k^* R_i) X_k^* Y}{\sum_{j,k} X_j^* X_k E(R_j^* R_k) + E(|N|^2)}. \quad (5.6)$$

Falls die einzelnen Kanäle i des rezeptiven Feldes r_i unabhängig voneinander sind mit verschwindendem Erwartungswert $E(r_i) = 0$, dann gilt $E[\text{cov}(r_j, r_k)] = 0$, somit $E(R_j^* R_k) = 0$, für $j \neq k$. In diesem Fall wird aus (5.6)

$$\hat{R}_i = \frac{E(|R_i|^2) X_i^* Y}{\sum_j |X_j|^2 E(|R_j|^2) + E(|N|^2)}.$$

Ist weiter $E[\text{cov}(r_i, r_i)]$ für alle Kanäle i gleich, also $E(|R_i|^2) =: E(|R|^2)$, ergibt sich

$$\hat{R}_i = \frac{X_i^* Y}{\sum_j |X_j|^2 + E(|N|^2)/E(|R|^2)}. \quad (5.7)$$

Dies ist das gleiche Ergebnis wie für den eindimensionalen Fall (4.17) in Abschnitt 4.5, nur muss das dort eindimensionale Signal x hier durch den jeweiligen Kanal x_i ersetzt werden, und es befindet sich die Summe $\sum_j |X_j|^2$ der quadratischen Beträge der spektralen Komponenten X_j aller Kanäle j im Nenner. Das rezeptive Feld r_i eines Kanals kann also unter den getroffenen Voraussetzungen weitgehend unabhängig vom Inhalt x_j anderer Kanäle $j \neq i$ geschätzt werden.

Beschränkt man sich auf den Fall $E(|N(\omega)|^2)/E(|R(\omega)|^2) = \sigma^2 \omega^{2d}$, wie es in dieser Arbeit in den numerischen Simulationen der Fall ist, erhält man

$$\hat{R}_i = \frac{X_i^* Y}{\sum_j |X_j|^2 + \sigma^2 \omega^{2d}}. \quad (5.8)$$

Als frei wählbare Parameter stehen also nurmehr σ^2 und $n \in \mathbb{N}$ zur Verfügung.

5.2 Regularisierungsmethode

Die Regularisierungsmethode im mehrdimensionalen Fall minimiert

$$L = \int_{-\infty}^{\infty} (y - \sum_i r_i \star x_i)(t)^2 dt + \sigma^2 \sum_i \int_{-\infty}^{\infty} (g_i \star r_i)(t)^2 dt. \quad (5.9)$$

Notwendige Bedingung für ein Minimum an der Stelle \hat{r}_i ist

$$\begin{aligned} \left. \frac{\delta L}{\delta r_i(\tau)} \right|_{r_i=\hat{r}_i} &= 0 \\ \int_{-\infty}^{\infty} -2(y - \sum_j \hat{r}_j \star x_j)(t)x_i(t - \tau) + 2\sigma^2 \int_{-\infty}^{\infty} (g_i \star \hat{r}_i)(t)g_i(t - \tau) &= 0 \\ -\text{cov}(x_i, y - \sum_j \hat{r}_j \star x_j) + \sigma^2 \text{cov}(g_i, g_i \star \hat{r}_i) &= 0. \end{aligned}$$

Fouriertransformation ergibt

$$\begin{aligned} -X_i^* y + X_i^* \sum_j \hat{R}_j X_j + \sigma^2 G_i^* G_i \hat{R}_i &= 0 \\ \sum_j (X_i^* X_j + \sigma^2 |G_i|^2 \delta_{ij}) \hat{R}_j &= X_i^* Y. \end{aligned} \quad (5.10)$$

Dies ist ein lineares Gleichungssystem für \hat{R}_j . Es hat die Lösung

$$\hat{R}_j = \frac{X_j^* Y}{\sum_i |X_i|^2 + \sigma^2 |G_j|^2}, \quad (5.11)$$

wie man durch Einsetzen der Lösung in das Gleichungssystem leicht verifiziert. Falls $\sigma > 0$ ist die Lösung ist eindeutig, da die Matrix $(A^H A + I)$ invertierbar ist¹.

Die Lösung (5.11) entspricht der Lösung (5.7), wenn man $\sigma^2 |G_j| = \langle |N|^2 \rangle / \langle |R|^2 \rangle$ wählt.

Einen Überblick über das verwandte Problem der Bildrekonstruktion mit Regularisierung gibt Demoment (1989).

5.3 Bayes'scher Schätzer

Wir gehen vor wie beim Bayes'schen Schätzer in Abschnitt 4.6, nur in mehreren Dimensionen. Der Output y des Neurons bei Input x_i sei also nun

$$E(y) = \sum_i r_i \star x_i.$$

¹Es ist $\lim_{\sigma \rightarrow 0} (A^H A + \sigma I)^{-1} = A^\dagger$. Die Matrix A^\dagger heißt Pseudoinverse von A (Duda et al. 2000). Falls A injektiv ist, entspricht A^\dagger der Moore-Penrose-Inversen (Cucker und Smale 2002).

Weiter sei der Output y mit additivem, zeitinvariantem, Gauß'schen (nicht notwendigerweise weißem) Rauschen n verfälscht, also

$$y = \sum_i r_i \star x_i + n. \quad (5.12)$$

Weiter sei r_i a-priori verteilt wie zeitinvariantes Gauß'sches (nicht notwendigerweise weißes) Rauschen. Analog zu Abschnitt 4.6 maximieren wir die Größe

$$L = \ln[p(y|(r_1, \dots, r_d))p(r_1, \dots, r_d)] \quad (5.13)$$

bezüglich des rezeptiven Feldes r_i , $i = 1, \dots, d$. Es ist wieder

$$p(n) \sim \exp \left[- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(t) c_n(t-t') n(t') dt dt' \right].$$

Wegen (5.12) gilt $p(y|r) = p(n)$. Da die rezeptiven Teil-Felder r_i a-priori wie zeitinvariantes Gauß'sches Rauschen verteilt sein sollen, gibt es eine Funktion c_{r_i} , so dass

$$p(r_i) \sim \exp \left[- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r_i(t) c_{r_i}(t-t') r_i(t') dt dt' \right].$$

und $p(r_1, \dots, r_d) = \prod_i p(r_i)$. Setzen wir die berechneten Wahrscheinlichkeiten $p(y|(r_1, \dots, r_d))$ und $p(r_i)$ in (5.13) ein, so erhalten wir

$$L \sim - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(t) c_n(t-t') n(t') dt dt' - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_j r_j(t) c_{r_j}(t-t') r_j(t') dt dt'.$$

Notwendige Bedingung für ein Minimum an der Stelle \hat{r}_i ist

$$\begin{aligned} \frac{\delta L}{\delta r_i(\tau)} \Big|_{r_i=\hat{r}_i} &= 0 \\ -2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\delta n}{\delta r_i(\tau)} \Big|_{r_i=\hat{r}_i} c_n(t-t') n(t') dt dt' - 2 \int_{-\infty}^{\infty} c_{r_i}(\tau-t') \hat{r}_i(t') dt' &= 0. \end{aligned}$$

Wegen (5.12) gilt

$$n = y - \sum_j r_j \star x_j,$$

somit

$$\frac{\delta n}{\delta r_i(\tau)} = -x_i(t - \tau).$$

Eingesetzt in die notwendige Bedingung für ein Minimum ergibt dies

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} -x_i(t - \tau) c_n(t - t') (y - \sum_j \hat{r}_j \star x_j)(t') dt dt' + \\ \int_{-\infty}^{\infty} c_{ri}(\tau - t') \hat{r}_i(t') dt' = 0 \\ \int_{-\infty}^{\infty} -x_i(t - \tau) (c_n \star y - c_n \star \sum_j \hat{r}_j \star x_j)(t) dt + c_{ri} \star \hat{r}_i = 0 \\ -\text{cov}(x_i, c_n \star y) + \sum_j \text{cov}(x_i, c_n \star \hat{r}_j \star x_j) + c_{ri} \star \hat{r}_i = 0. \end{aligned}$$

Fouriertransformation ergibt

$$\begin{aligned} -X_i^* C_n Y + \sum_j X_i^* C_n \hat{R}_j X_j + C_{ri} \hat{R}_i &= 0 \\ -C_n X_i^* Y + C_n X_i^* \sum_j \hat{R}_j X_j + C_{ri} \hat{R}_i &= 0 \\ \sum_j \left(X_i^* X_j + \frac{C_{ri}}{C_n} \delta_{ij} \right) \hat{R}_j &= X_i^* Y. \end{aligned}$$

Analog zum Gleichungssystem (5.10) ergibt sich als Lösung

$$\hat{R}_j = \frac{X_j^* Y}{\sum_i |X_i|^2 + \frac{C_{ri}}{C_n}}, \quad (5.14)$$

Dies ist ein Spezialfall der Lösung (5.11) mit $G_j := \frac{C_{ri}}{C_n}$.

5.4 Zusammenfassung

Zusammenfassend ergibt sich im multidimensionalen Fall die Tabelle 5.1.

5.5 Spikende Neurone

Sind die in diesem Kapitel berechneten optimalen rezeptiven Felder auch im Fall spikender Neurone optimal? Wir betrachten zunächst Poisson-Neurone,

Methode	\hat{r}	Voraussetzungen	\hat{R}_i
Lin. Schätzung	$y_0 \star \underset{s_i}{\operatorname{argmin}} \langle \ r_i - s_i \star y_0\ ^2 \rangle$	$y = \sum_j r_j \star x_j + n$	$\frac{X_i^* Y_0}{\sum_j X_j ^2 + \frac{\langle N ^2 \rangle}{\langle R ^2 \rangle}}$
Regularisierung	$\underset{r}{\operatorname{argmin}} \left(\ y_0 - \sum_j r_j \star x_j\ ^2 + \sigma^2 \sum_j \ g_j \star r_j\ ^2 \right)$		$\frac{X_i^* Y_0}{\sum_j X_j ^2 + \sigma^2 G_i ^2}$
Bayes	$\underset{(r_1, \dots, r_d)}{\operatorname{argmax}} p((r_1, \dots, r_d) y_0)$	$y = \sum_j r_j \star x_j + n$ n, r_j zeitinv. und Gauß'sch	$\frac{X_i^* Y_0}{\sum_j X_j ^2 + \frac{\langle N ^2 \rangle}{\langle R ^2 \rangle}}$
Rev. Corr.	$\operatorname{cov}(x_i, y_0)$		$X_i^* Y_0$

Tabelle 5.1: Zusammenfassung der Ergebnisse aus Abschnitt 5.

Die Zeichen bedeuten die Eingabe x_i , die neuronale Ausgabe y , die Messung y_0 der neuronalen Ausgabe und das rezeptive Feld r_i . Die sonstige Bezeichnungen und Voraussetzungen sind analog zu Tabelle 4.1.

bei denen Spike generierende Prozess ein Poissonprozess ist mit Feuerwahrscheinlichkeitsdichte $\lambda(t)$. Sei weiterhin λ die Summe von Faltungen von Inputs x_i mit einem rezeptiven Feldern r_i , also

$$\lambda = \sum_i r_i \star x_i.$$

Im Falle von Spike-Response-Neuronen (ohne Refraktärpotential) ist λ das Membranpotential, wenn $x_i(t) = \sum_{t_f} \delta(t - t_f)$ die Input-Spikes, α das postsynaptische Potential, J_i^Δ die Stärke einer um Δ verzögerten Synapse von Input i und

$$r_i = \sum_{\Delta} \int \alpha(\tau) \star x_i(t - \Delta - \tau). \quad (5.15)$$

Spike-Response-Neuronen, die Poisson-Neuronen sind, lassen sich also durch das rezeptive Feld (5.15) beschreiben.

Falls $\langle y \rangle = \lambda$ gilt z. B. Gleichung (5.10) im Mittel im Fall von Poisson-Neuronen exakt. Also ist das nach Gleichung (5.11) berechnete geschätzte rezeptive Feld r_i , wenn man als neuronale Antwort y einfach den Spiketrain einsetzt, im Mittel das gleiche wie bei linearen (nicht spikenden) Neuronen, nur der Messfehler wird größer. Man kann jedoch nur dann $\langle y \rangle = \lambda$ setzen, wenn $y > 0$ ist, denn negative Wahrscheinlichkeitsdichten machen keinen Sinn. Die Voraussetzung $y > 0$ ist zumindest dann erfüllt, wenn $r_i > 0$, also wenn keine Inhibition auftritt oder wenn die Spontanfeurrate hoch genug ist.

Keep it simple: as simple as possible, but not simpler.

Albert Einstein

6

Lineare rezeptive Felder mit Regularisierung in einer Dimension

In diesem Kapitel wird zunächst das angewandte Regularisierungs-Verfahren zur Messung von rezeptiven Feldern dargestellt. Es folgen Ergebnisse mit verschiedenen Stimuli als Eingabe und verschiedenen Neuronenmodellen.

6.1 Das Verfahren

Das Verfahren zur Rekonstruktion der rezeptiven Felder in diesem Kapitel ist die Regularisierungsmethode nach Gleichung (4.3) aus Abschnitt 4.3.

6.2 Tiefpassrauschen als Eingabe

Zunächst betrachten wir Tiefpassrauschen als Eingabe in ein Neuron, d. h. das Membranpotential eines Neurons wird mit Tiefpassrauschen „gefüttert“. Tiefpassrauschen enthält alle Frequenzen bis zu einem Maximum, der „Grenzfrequenz“ (Unbehauen 1993). Dieser Input ist künstlich. Jedoch ist er insofern natürlich, als dass Aktionspotentiale, die am Soma ankommen, nach dem Spike-Response-Modell (Abschnitt 1.5) mit einem postsynaptischen Potential gefaltet werden. Das postsynaptische Potential enthält zwar alle Frequenzen, jedoch fällt die Intensität mit steigender Frequenz schnell ab,

denn der Betrag der Fouriertransformierten vom postsynaptischen Potential ε aus (1.4) ist

$$|\mathcal{F}\varepsilon(\omega)| = \frac{1}{\frac{1}{\tau_s^2} + \omega^2}.$$

Auch kann man sich vorstellen, dass ein Neuron von sehr vielen Synapsen¹ kontaktiert wird, wodurch, falls diese in etwa zufällig feuern, eine Art Zufallsprozess als Membranpotential erzeugt wird, der durch Tiefpassrauschen modelliert werden kann.

6.2.1 Einfluss der oberen Grenzfrequenz

Welchen Einfluss hat die obere Grenzfrequenz? Um beurteilen zu können, ob rezeptive Felder mit Tiefpassrauschen als Eingabe akkurat rekonstruiert werden können, sollte testweise ein rezeptives Feld rekonstruiert werden, das auch alle Frequenzen *enthält*. Hier wurde

$$r(t) = \begin{cases} 1 & \text{falls } t_0 \leq t \leq t_0 + \tau \\ 0 & \text{sonst} \end{cases}$$

mit $t_0 = 25$ ms und $\tau = 25$ ms gewählt. Der Betrag der Fouriertransformierten ist

$$|\mathcal{F}r(\omega)| = \frac{\sqrt{2[1 - \cos(\tau\omega)]}}{\omega}.$$

Die Intensität fällt mit steigender Frequenz also nur mit $1/\omega$ ab, es sind also genügend hohe Frequenzen im rezeptiven Feld vorhanden, die rekonstruiert werden müssen. Das Ergebnis zeigt Abbildung 6.1. Je höher die Grenzfrequenz der Eingabe, desto ähnlicher wird das rekonstruierte rezeptive Feld dem Original.

¹Im Kleinhirn von bis zu 100000 (Kandel et al. 1996).

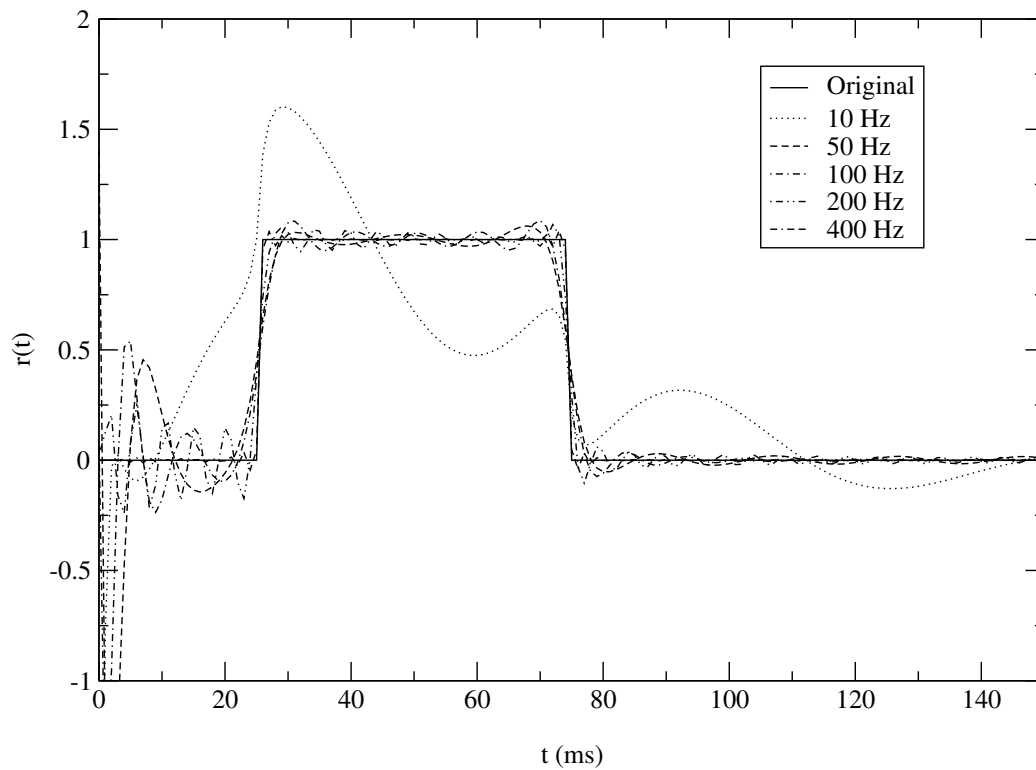


Abbildung 6.1: Rekonstruierte receptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 1$ und Ordnung $d = 0$. Das „Neuron“ ist linear, d. h. die neuronale Ausgabe ist $y = r \star x$. Die Eingabe x ist Tiefpassrauschen der Dauer 100 s mit oberen Grenzfrequenzen 10, 50, 100, 200 bzw. 400 Hz. Das „echte“ receptive Feld r (durchgezogene Linie) ist rechteckig mit $r(t) = 1$ für $t \in [25 \text{ ms}, 75 \text{ ms}]$ und $r(t) = 0$ sonst.

Für eine Frequenz von 10 Hz zeigt das rekonstruierte receptive Feld \hat{r} nur eine entfernte Ähnlichkeit mit dem Original r . Dies ist zu erwarten, denn eine Frequenz von 10 Hz enthält nur eine halbe Schwingung im Bereich mit $r(t) = 1$. Je größer die Grenzfrequenz der Eingabe x , desto mehr nähert sich das geschätzte receptive Feld dem Original an. Bei der Grenzfrequenz 400 Hz ergibt sich eine perfekte Rekonstruktion.

6.2.2 Einfluss des Regularisierungsparameters

Nun untersuchen wir den Einfluss des Regularisierungsparameters σ aus Gleichung (4.3). Wir verwenden wieder das gleiche „rechteckige“ rezeptive Feld aus dem vorigen Abschnitt und eine obere Grenzfrequenz von 200 Hz, da mit dieser Grenzfrequenz das rezeptive Feld schon sehr gut rekonstruiert werden konnte. Ergebnisse zeigt Abbildung 6.2. Man sieht, dass der Regularisierungsparameter einen großen Einfluss auf die Rekonstruktion hat. Insbesondere wenn er zu klein ist, ergeben sich in der Rekonstruktion störende „Schwinger“.

Bei einer Grenzfrequenz von 100 Hz nehmen die Ungenauigkeiten mit abnehmendem Regularisierungsparameter σ zu, wie Abbildung 6.3 zeigt.

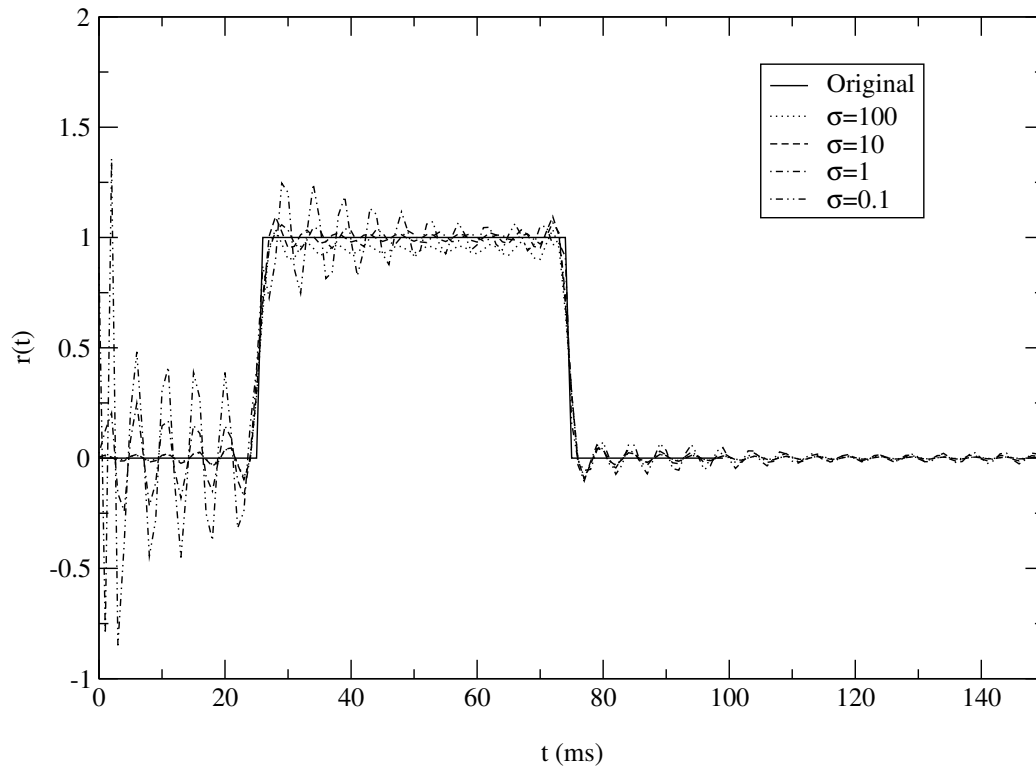


Abbildung 6.2: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit variierenden Regularisierungsparametern $\sigma \in \{100, 10, 1, 0.1\}$ und Ordnung $d = 0$, wie in Abbildung 6.1. Grenzfrequenz der Eingabe ist 200 Hz.

Ein großer Regularisierungsparameter $\sigma = 100$ führt zu einer geringfügigen Unterschätzung der Stärke des rezeptiven Feldes im Bereich $[25 \text{ ms}, 75 \text{ ms}]$. Bei zu geringen Regularisierungsparametern, z. B. $\sigma = 0.1$, werden die störenden „Schwinger“ immer größer, sowohl im Bereich $r(t) = 1$, als auch im Bereich der Nulllinie $r(t) = 0$.

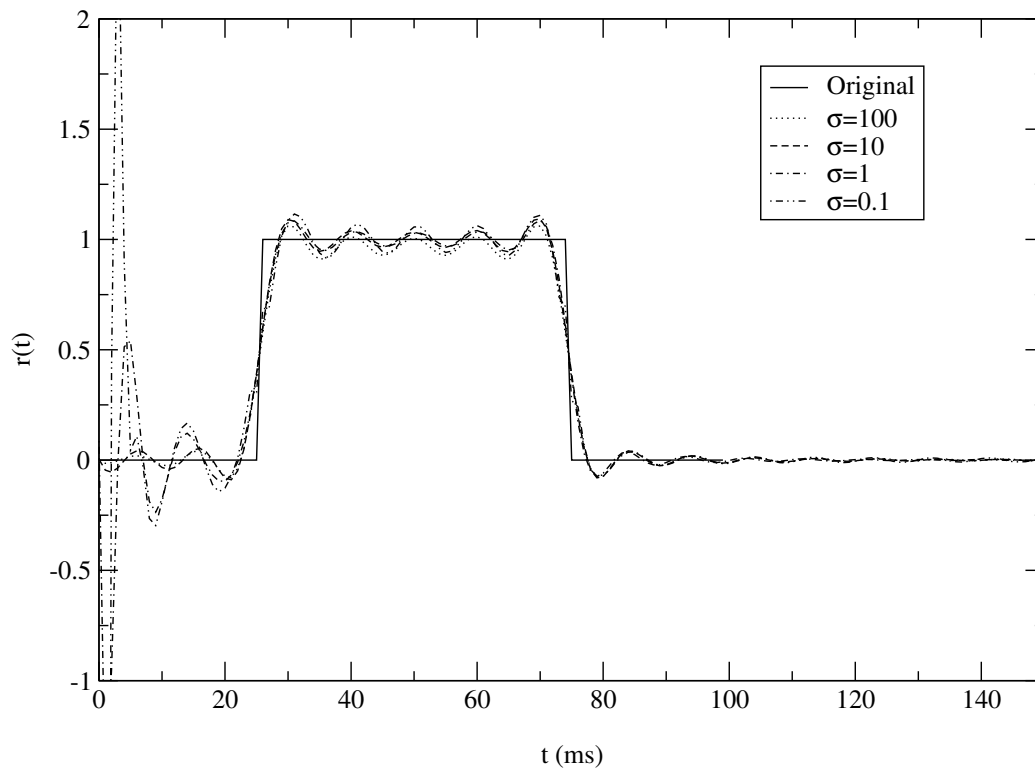


Abbildung 6.3: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit variierenden Regularisierungsparametern wie in Abbildung 6.2. Grenzfrequenz der Eingabe ist nun 100 Hz. In diesem Fall nehmen die Ungenauigkeiten mit abnehmendem Regularisierungsparameter σ zu.

6.2.3 Einfluss der Ordnung der Regularisierung

Gleichung (4.3) enthält außer dem reellen Regularisierungsparameter σ noch einen weiteren ganzzahligen Parameter d , die Ordnung der Regularisierung. Die Wahl der Ordnung erscheint relativ willkürlich. Insofern ist es beruhigend, dass die Ordnung laut Abbildungen 6.4 und 6.5 nur einen geringen Einfluss auf die rekonstruierten rezeptiven Felder hat.

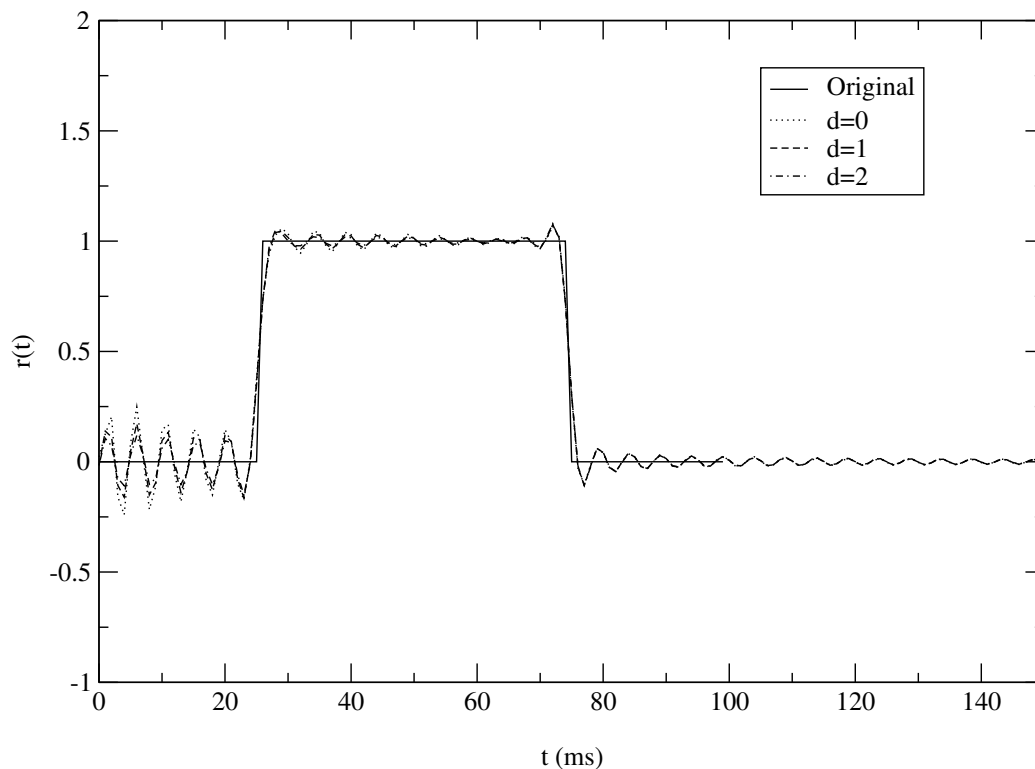


Abbildung 6.4: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 1$ und verschiedenen Ordnungen d , ansonsten wie in Abbildung 6.1. Grenzfrequenz der Eingabe ist 200 Hz. Höhere Ordnungen d wirken leicht „glättend“, haben jedoch insgesamt einen geringen Einfluss.

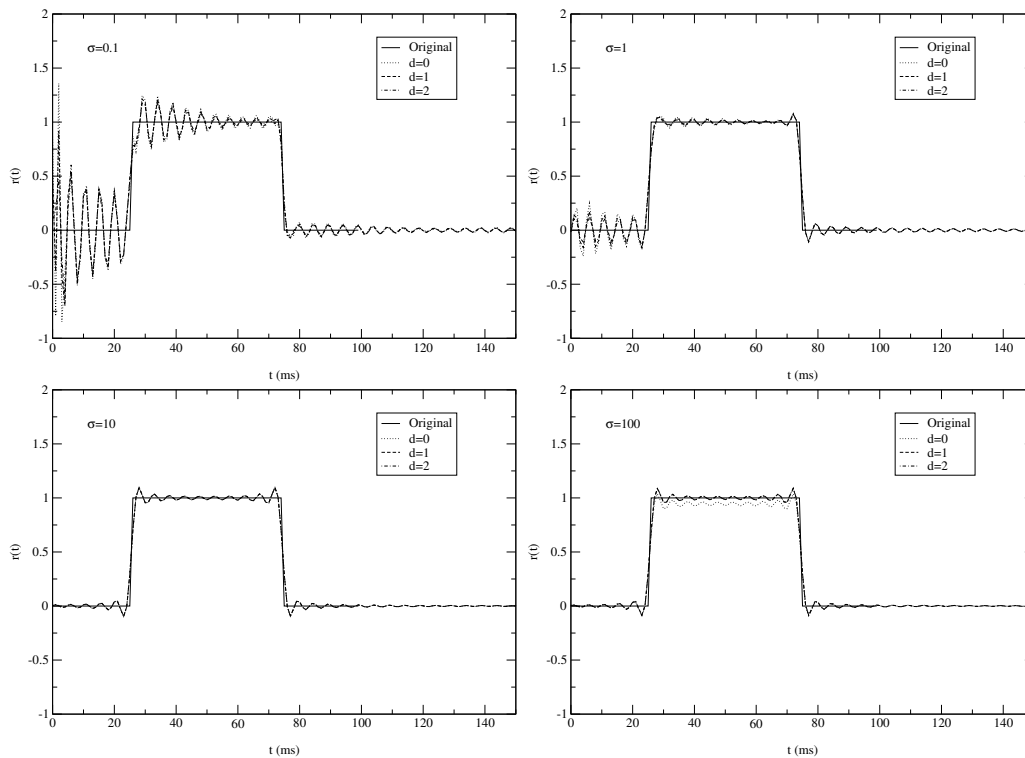


Abbildung 6.5: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und verschiedenen Ordnungen d , wie in Abbildung 6.4.

Bei hohem Regularisierungsparameter $\sigma = 100$ führt eine Ordnung $d > 0$ dazu, dass die Stärke des rezeptiven Feldes im Bereich $r(t) = 1$ nicht unterschätzt wird.

6.2.4 Einfluss der Messdauer

In der Realität steht pro Neuron nur eine begrenzte Messdauer zur Verfügung. Die Messdauer wird zum einen dadurch, dass die Elektrode im Gewebe u.U. nur begrenzte Zeit stabil bleibt, das Versuchstier nur eine begrenzte Zeit unter Narkose am Leber erhalten werden kann und der Experimentator pro Tier möglichst nicht nur ein einziges, sondern viele Neurone untersuchen möchte. Selbst bei fest implantierten Elektroden, bei denen über mehrere Tage oder länger vom gleichen Neuron oder der gleichen Neuronengruppe abgeleitet werden kann, ist die Versuchsdauer durch einen experimentellen Zeitaufwand begrenzt, der noch vernünftig erscheint. Wie Abbildungen 6.6 und 6.7 zeigen, ergeben zu kurze Messdauern natürlich große zufällige Fehler, während sich das Ergebnis bei längerer Messdauer zu noch längeren Messdauern hin nicht mehr wesentlich ändert.

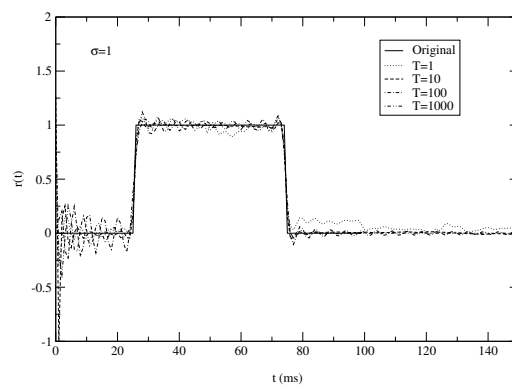


Abbildung 6.6: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 1$ und verschiedenen Messdauern T , ansonsten wie in Abbildung 6.1. Grenzfrequenz der Eingabe ist 200 Hz. Die kurze Messdauer von 1 s führt zu kleineren Ungenauigkeiten. Messdauern von > 10 s verbessern die Schätzung \hat{r} nicht.

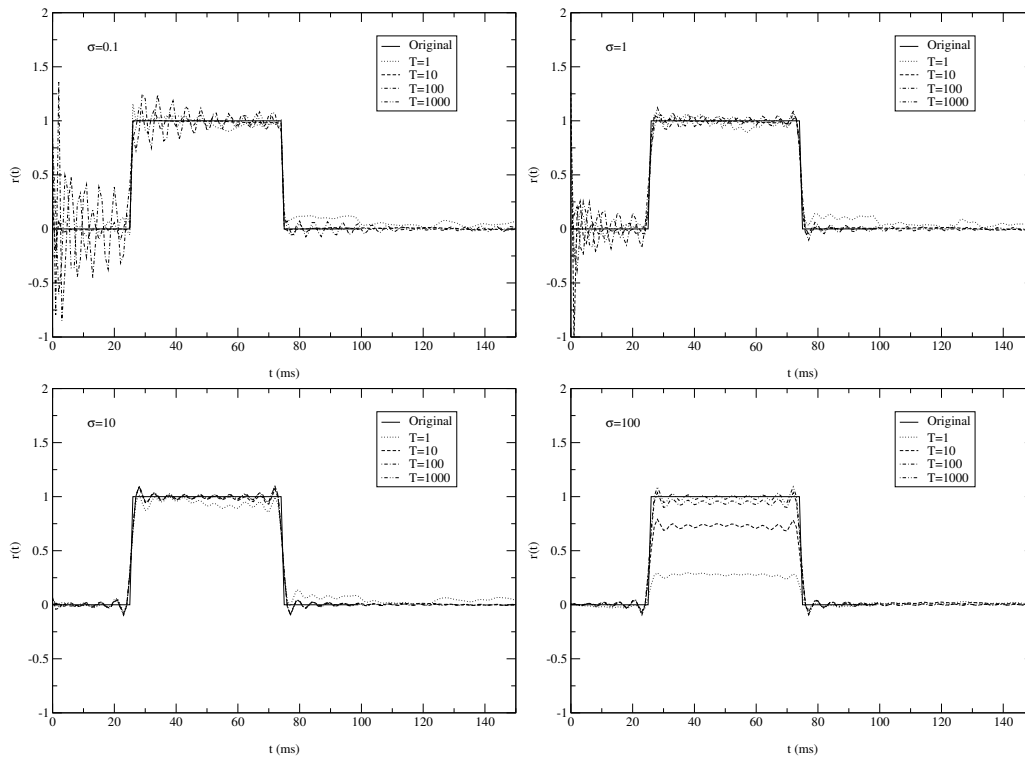


Abbildung 6.7: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ , Ordnung $d = 1$ und verschiedenen Messdauern T , wie in Abbildung 6.6.

Bei kleinen Regularisierungsparametern verbessern Messdauern $T > 10$ s die Schätzungen \hat{r} nicht. Bei dem großen Regularisierungsparameter $\sigma = 100$ wird die Stärke des rezeptiven Feldes allerdings bei Messdauern $T < 100$ s unterschätzt.

6.2.5 Einfluss der Form des rezeptiven Feldes

Anhand der Abbildungen in diesem Abschnitt wird untersucht, inwieweit die Rekonstruktion des rezeptiven Feldes von der speziell gewählten „rechteckigen“ Form abhängig war. Dazu wählen wir eine Form, deren Frequenzbereich im Fourierraum durch einen Parameter variiert werden kann:

$$r(t) = \begin{cases} \sin[2\pi f_{\text{RF}}(t - t_0)] & \text{falls } t_0 \leq t \leq t_0 + \tau \\ 0 & \text{sonst} \end{cases}$$

mit $\tau = 25 \text{ ms}$ und verschiedener Frequenz f_{RF} . Mit $\omega_{\text{RF}} := 2\pi f_{\text{RF}}$ ist der Betrag der Fouriertransformierten dieser Funktion

$$|\mathcal{F}r(\omega)| = \left| \frac{\exp\left(-\frac{i\tau\omega}{2}\right) \left[\omega_{\text{RF}} \cos\left(\frac{\omega_{\text{RF}}\tau}{2}\right) + i\omega \sin\left(\frac{\omega_{\text{RF}}\tau}{2}\right)\right]}{\omega^2 - \omega_{\text{RF}}^2} + \frac{\exp\left(\frac{i\tau\omega}{2}\right) \left[-\omega_{\text{RF}} \cos\left(\frac{\omega_{\text{RF}}\tau}{2}\right) + i\omega \sin\left(\frac{\omega_{\text{RF}}\tau}{2}\right)\right]}{\omega^2 - \omega_{\text{RF}}^2} \right|$$

und hat Intensitäts-Peaks bei $\omega \approx \pm 2\pi f_{\text{RF}}$ und außerhalb von diesen Peaks relativ geringe Intensität. Somit lässt sich der Frequenzgehalt durch Wahl von des Parameters f_{RF} steuern. Es zeigt sich, dass die Wahl des Regularisierungsparameters σ und der Ordnung d relativ unkritisch sind, solange f_{RF} sehr viel kleiner als die Grenzfrequenz des verwendeten Tiefpassrauschens ist. Kommt f_{RF} in die Nähe der oberen Grenzfrequenz des Eingabe-Rauschens, dann werden die Fehler zwischen rekonstruierten rezeptiven Feld und Original mitunter sehr groß (Abbildungen 6.12 und 6.13), wobei dann auch die Wahl der Parameter σ und d eine größere Rolle spielt.

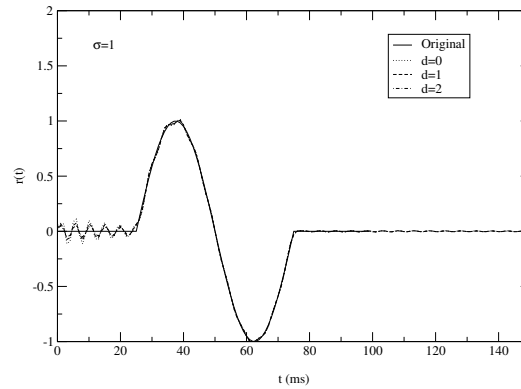


Abbildung 6.8: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 1$ und verschiedenen Ordnungen d , wie in Abbildung 6.4. Die obere Grenzfrequenz der Eingabe ist 200 Hz. Das rezeptive Feld ist nun $r(t) = \sin[2\pi f_{RF}(t - 0.025 \text{ s})]$ $f_{RF} = 20 \text{ Hz}$ im Bereich $[25 \text{ ms}, 75 \text{ ms}]$ und $r(t) = 0$ sonst.

Das rezeptive Feld wird gut rekonstruiert, die Ordnung d der Regularisierung hat geringen Einfluss. Im Bereich um 0 zeigen sich die größten Abweichungen zwischen Schätzung und Original.

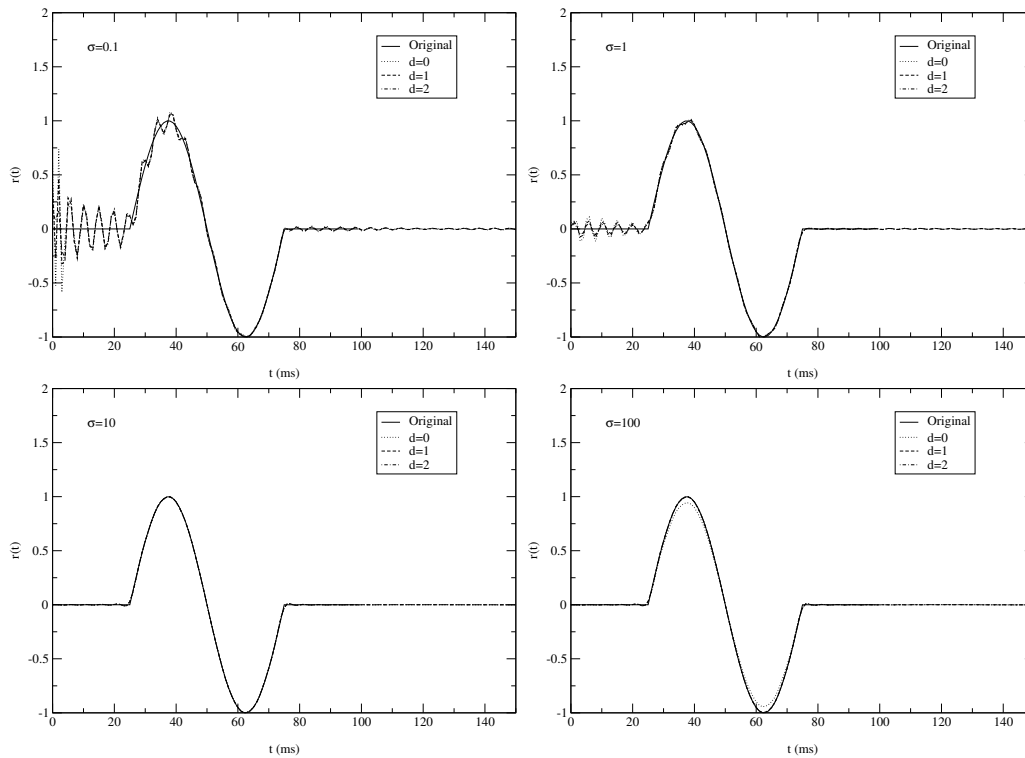


Abbildung 6.9: Rekonstruierte receptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und verschiedenen Ordnungen d , wie in Abbildung 6.8.

Die Abweichungen im Bereich um 0 nehmen mit größer werdendem Regularisierungsparameter σ ab. Bei großem Regularisierungsparameter $\sigma = 100$ wird das Maximum und das Minimum des receptiven Feldes betragsmäßig unterschätzt.

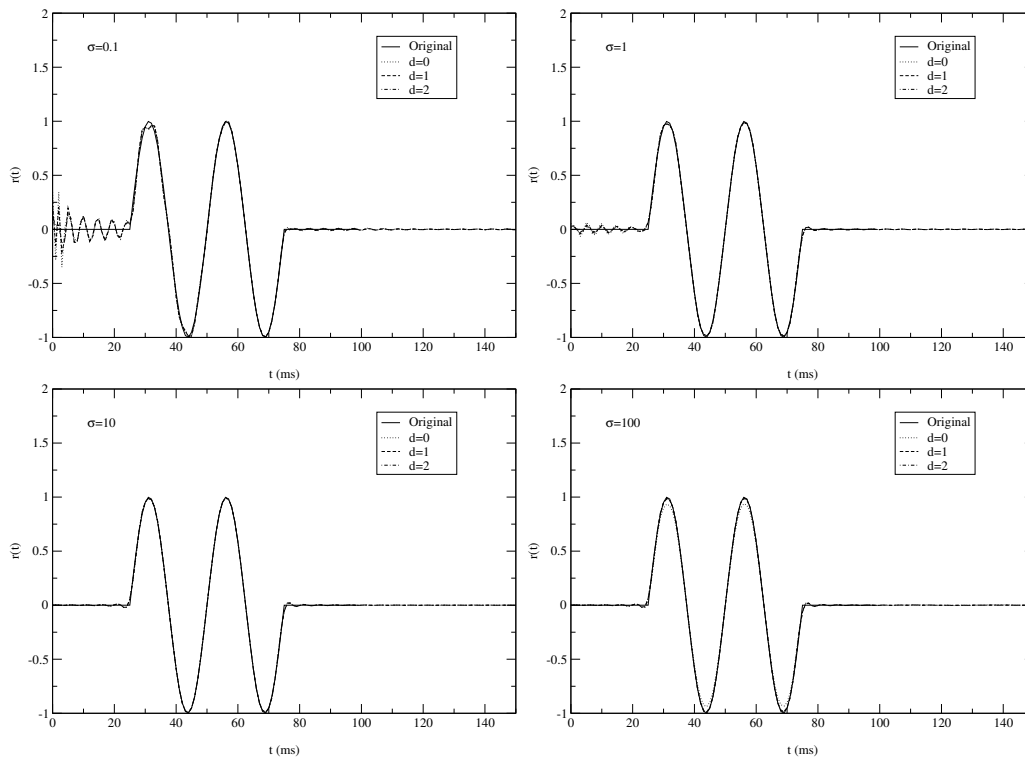


Abbildung 6.10: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und verschiedenen Ordnungen d , wie in Abbildung 6.9 mit $f_{\text{RF}} = 40$ Hz.

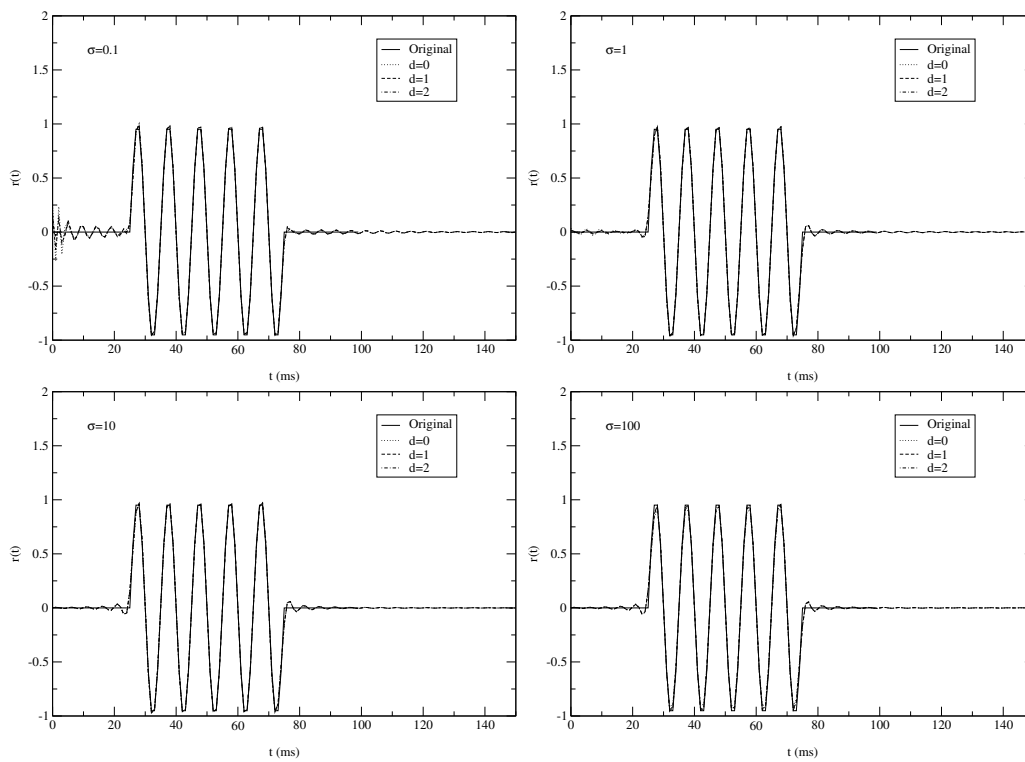


Abbildung 6.11: Wie Abbildung 6.9 mit $f_{RF} = 100$ Hz.

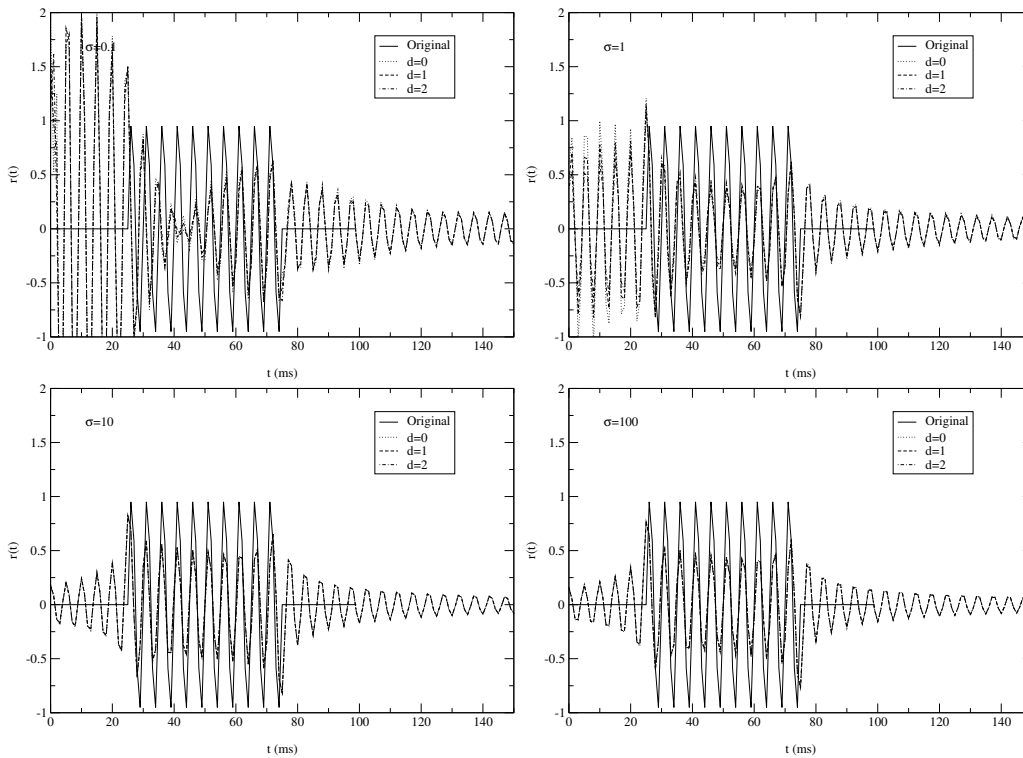


Abbildung 6.12: Wie Abbildung 6.9 mit $f_{\text{RF}} = 200$ Hz.

Starke Abweichungen zwischen rekonstruiertem rezeptivem Feld \hat{r} und Original r zeigen sich insbesondere bei kleinen Regularisierungsparametern $\sigma \in \{0.1, 1\}$. Die Nulllinien im Bereich $r(t) = 0$ werden nurmehr ungenau wiedergegeben. Die Rekonstruktionen mit verschiedenen Ordnungen d decken sich fast.

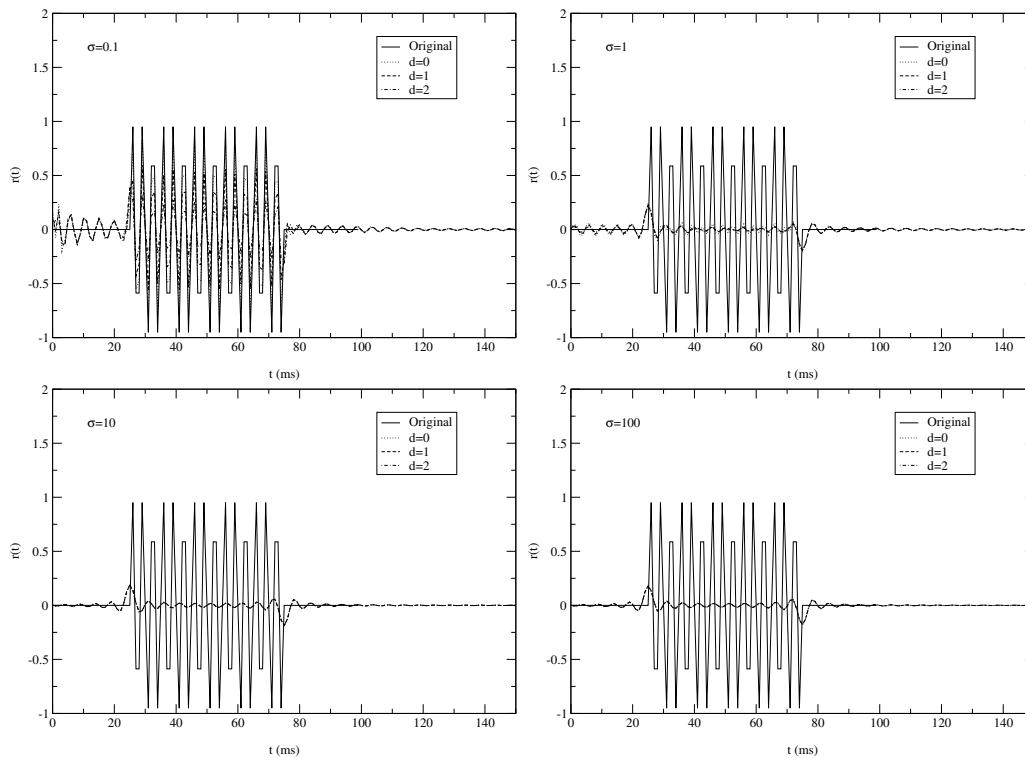


Abbildung 6.13: Wie Abbildung 6.9 mit $f_{\text{RF}} = 300$ Hz. Die Rekonstruktionen \hat{r} sind praktisch null, außer bei kleinem Regularisierungsparameter $\sigma = 0.1$.

6.2.6 Einfluss des Rauschens in der Messung

In diesem Abschnitt wird untersucht, welchen Einfluss unvermeidliches Rauschen auf die Rekonstruktion des rezeptiven Feldes hat. Das „Neuron“ ist wieder linear, nun aber mit weißem Rauschen behaftet, die neuronale Ausgabe ist also

$$y = r \star x + n,$$

wobei n Gauß'sches weißes Rauschen mit Varianz 1 ist. Der Regularisierungsparameter hat hier kritischen Einfluss auf die Messung. Ist er zu klein (Abbildung 6.14), hat das Messergebnis „rauschartigen“ Charakter und das rezeptive Feld kann nicht rekonstruiert werden. Mit größer werdendem Regularisierungsparameter σ erhält man zunächst immer bessere Rekonstruktionen (Abbildung 6.15), bis bei zu großen σ der Betrag des rezeptiven Feldes unterschätzt wird. Dies ist insofern verständlich, da ja bei der Regularisierungsmethode L aus Gleichung (4.10) minimiert wird und das Minimum für die Regularisierungsordnung $d = 0$ bei $\sigma \rightarrow \infty$ gegen $r(t) = 0$ geht.

Bei stärker werdendem Rauschen gelingt die Rekonstruktion nur mehr mit größeren Regularisierungsparametern σ (Abbildung 6.16). Bei zu starkem Rauschen allerdings ist das rezeptive Feld verständlicherweise nicht mehr rekonstruierbar, unabhängig von der Wahl von σ (Abbildung 6.17).

Welchen Einfluss hat nun die Ordnung der Regularisierung auf die Rekonstruktion? Störend war bei Regularisierungsordnung $d = 0$ die betragsmäßige Unterschätzung des rezeptiven Feldes bei großen Regularisierungsparametern σ . Dies lässt sich durch Wahl von $d \geq 1$ vermeiden, wie Abbildungen 6.18 und 6.19 zeigen. Verständlich ist dies, da bei der Minimierung von L aus Gleichung (4.10) und $d \geq 1$ nur zeitliche *Änderungen* des rezeptiven Feldes $r(t)$ „bestraft“ werden und nicht der Absolutbetrag $|r(t)|$ bzw. $r(t)^2$. Daher erscheint eine Wahl von $d \geq 1$, beispielsweise $d = 1$, günstiger als $d = 0$.

Den Einfluss der Messdauer auf eine Messung mit Rauschen zeigt Abbildung 6.21. Je länger die Messdauer, desto genauer wird natürlich die Rekonstruktion. Man sieht, dass selbst in diesem einfachen Fall (eindimensionales rezeptives Feld und lineares Neuron) bei realistischem Rauschen Messdauern von 100 bis 1000 Sekunden zu erwarten sind, um vernünftige Ergebnisse zu erzielen.

Abbildungen 6.22 und 6.23 untersuchen den Einfluss der Ordnung der Regularisierung auf die Rekonstruktion von rezeptiven Feldern mit der sinusoiden Form. Auch hier stellt man fest, dass bei großem Regularisierungsparameter σ der Betrag des rezeptiven Feldes $r(t)$ bei Regularisierungsordnung $d = 0$ unterschätzt wird, was durch Wahl von $d \geq 1$ vermieden wird.

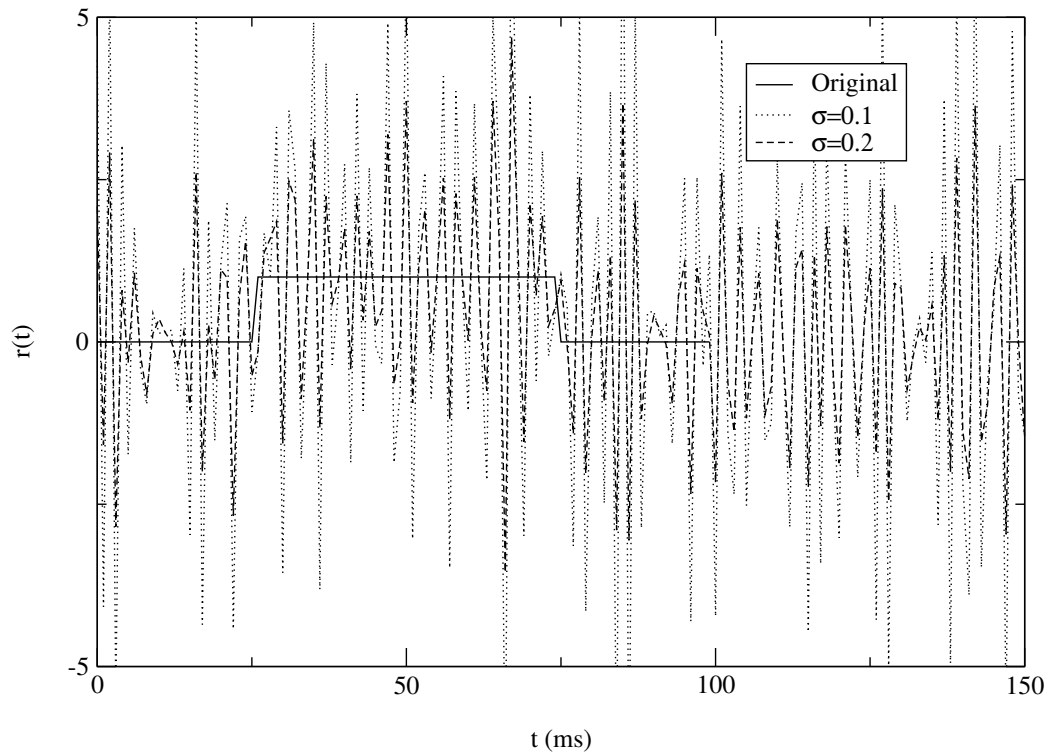


Abbildung 6.14: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und Ordnung $d = 0$. Das „Neuron“ ist linear mit weißem Rauschen, d. h. die neuronale Ausgabe ist $y = r \star x + n$, wobei n Gauß'sches weißes Rauschen mit Varianz 1 ist. Die Eingabe x ist Tiefpassrauschen der Dauer 100 s mit oberer Grenzfrequenz 200 Hz. Das „echte“ rezeptive Feld r (durchgezogene Linie) ist rechteckig, $r(t) = 1$ für $t \in [25 \text{ ms}, 75 \text{ ms}]$, $r(t) = 0$ sonst.

Bei zu kleinen Regularisierungsparametern σ wie in dieser Abbildung hat die Schätzung \hat{r} keine Ähnlichkeit mit dem rezeptiven Feld r . Die Schätzung hat rauschartigen Charakter.

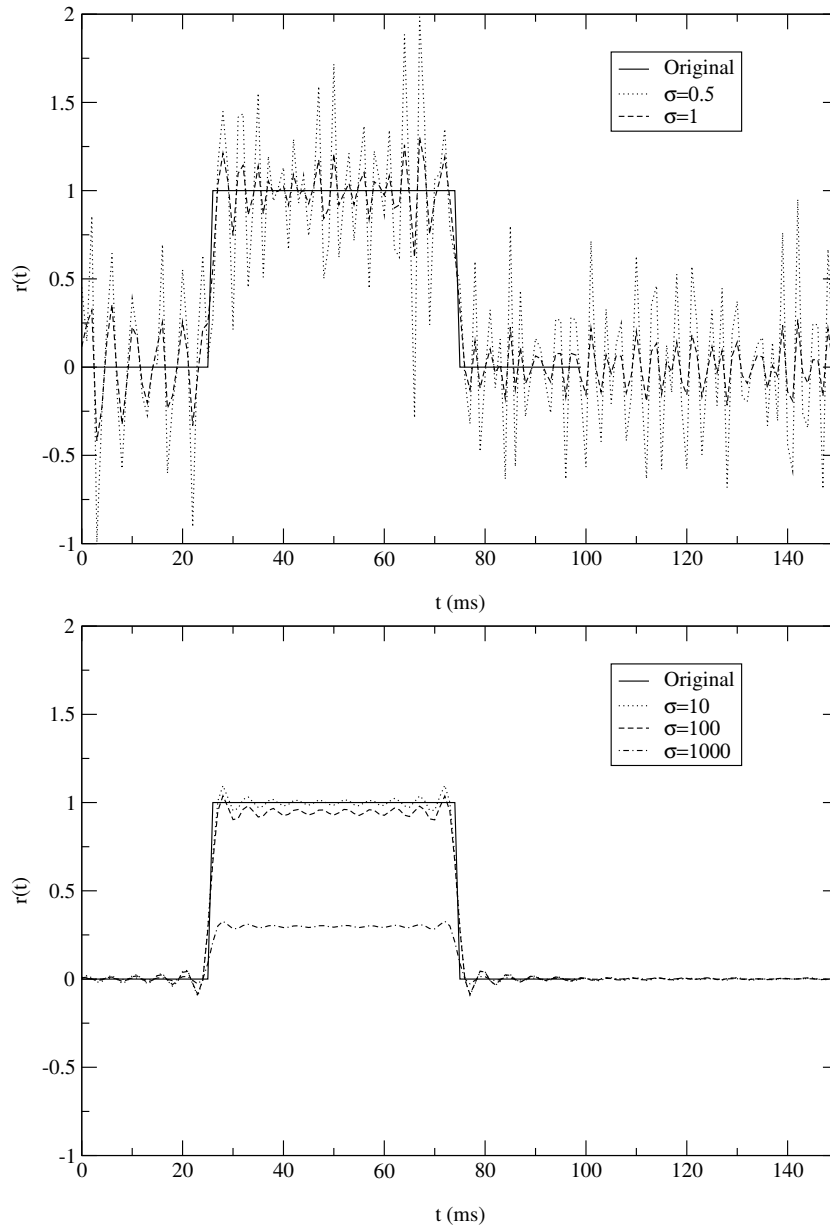


Abbildung 6.15: Wie Abbildung 6.14.

Je größer der Regularisierungsparameter σ , desto besser ist zunächst die Rekonstruktion \hat{r} des rezeptiven Feldes r . Bei großen Regularisierungsparametern ($\sigma \geq 100$) wird zwar die rechteckige Form des rezeptiven Feldes noch richtig rekonstruiert, allerdings wird die absolute Stärke im Bereich $r(t) = 1$ unterschätzt.

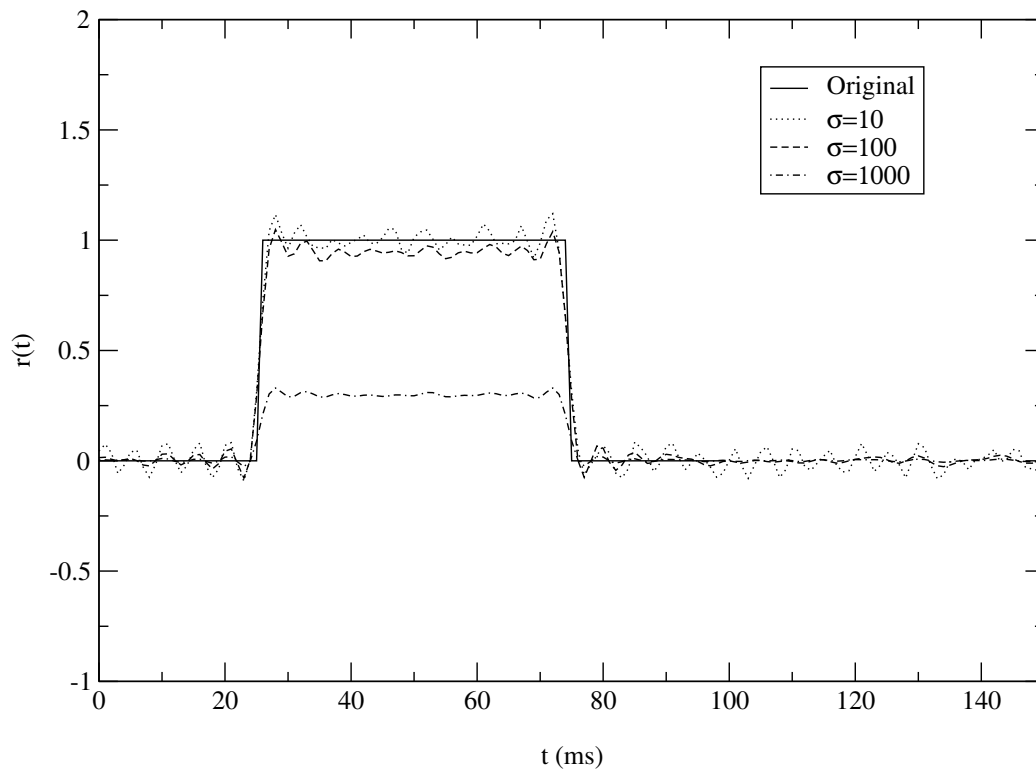


Abbildung 6.16: Wie Abbildung 6.14, nur ist hier die neuronale Antwort $y = r \star x + \sigma_n n$ mit $\sigma_n = 10$.

Mit entsprechend großem Regularisierungsparameter, etwa $\sigma \in \{10, 100\}$, wird das rezeptive Feld noch gut rekonstruiert, obwohl das die Messung verfälschende Rauschen $\sigma_n n$ nun 10-fach stärker ist als in den Abbildungen 6.14 und 6.15.

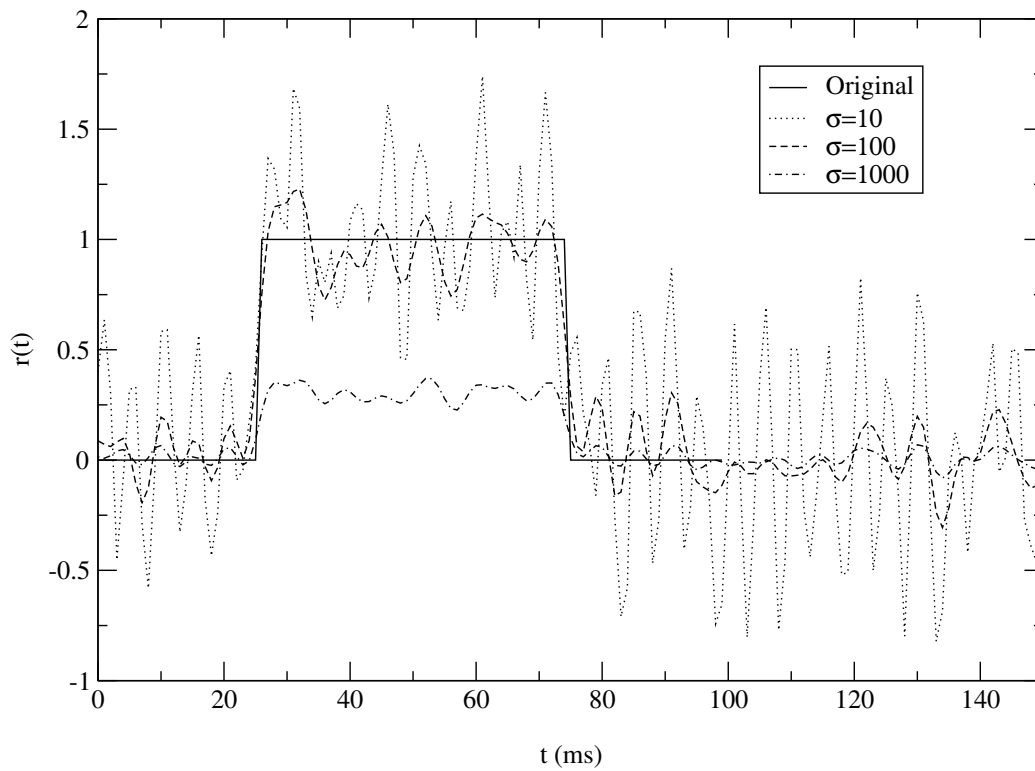


Abbildung 6.17: Wie Abbildung 6.16, mit $\sigma_n = 100$.

Es gibt bei so starkem Rauschen wie hier keinen Regularisierungsparameter mehr, der eine akzeptable Rekonstruktion des rezeptiven Feldes erlauben würde. Die Rekonstruktion mit $\sigma = 100$ gibt den rechteckigen Verlauf des rezeptiven Feldes noch einigermaßen wieder.

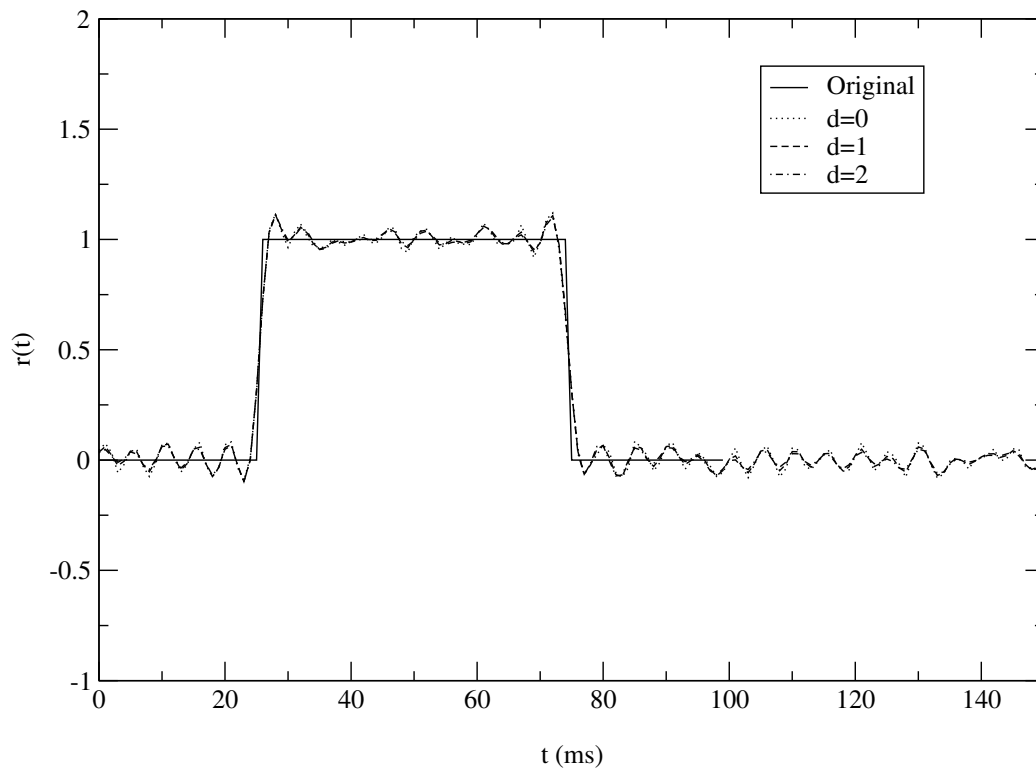


Abbildung 6.18: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 10$ und verschiedenen Ordnungen d , ansonsten wie Abbildung 6.16 mit $\sigma_n = 10$. Verschiedene Ordnungen der Regularisierung führen zu praktisch keiner Veränderung der rekonstruierten rezeptiven Felder \hat{r} .

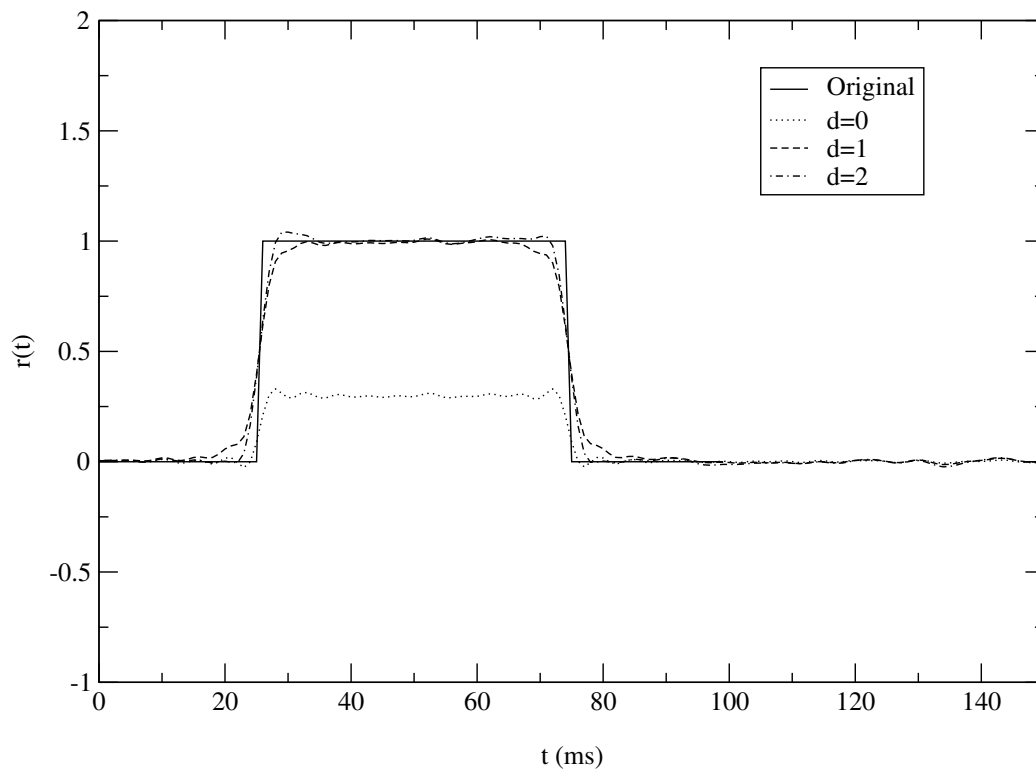


Abbildung 6.19: Wie Abbildung 6.18 mit $\sigma = 1000$.
Ordnungen der Regularisierung $d \geq 1$ führen dazu, dass die Stärke des rezeptiven Feldes im Bereich $r(t) = 1$ nicht unterschätzt wird.

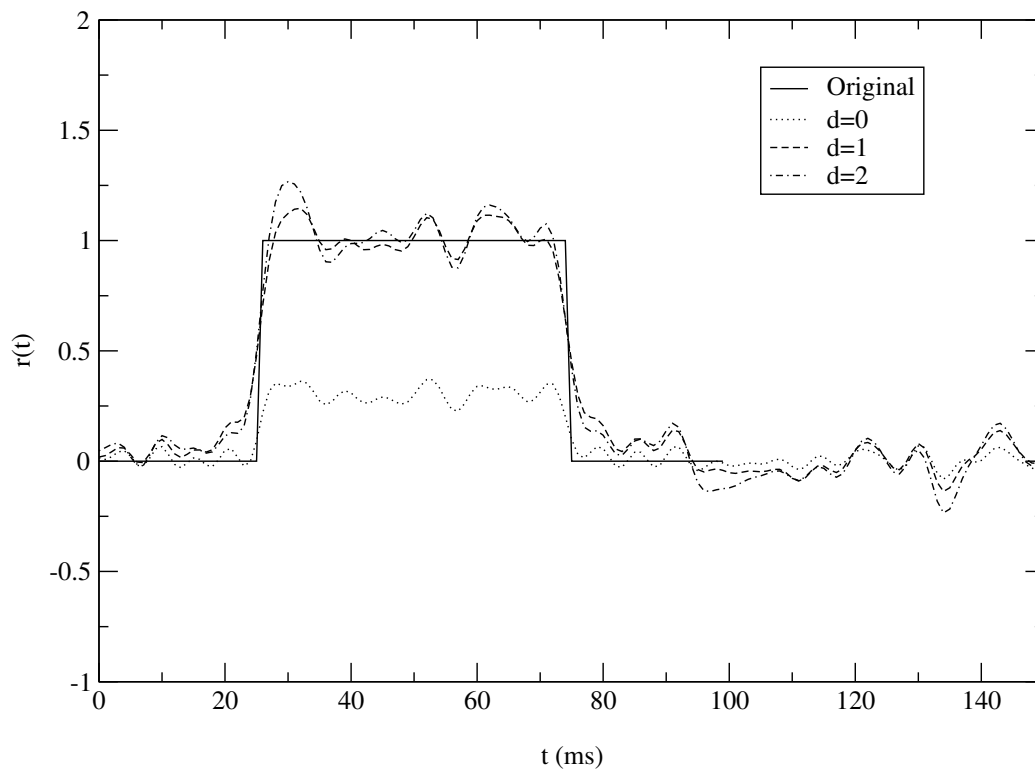


Abbildung 6.20: Wie Abbildung 6.18 mit $\sigma_n = 100$ und $\sigma = 1000$. Ordnungen der Regularisierung $d \geq 1$ führen wie in Abbildung 6.19 dazu, dass die Stärke des rezeptiven Feldes im Bereich $r(t) = 1$ nicht unterschätzt wird. Die Qualität der Wiedergabe der Form des rezeptiven Feldes ändert sich nicht sichtbar.

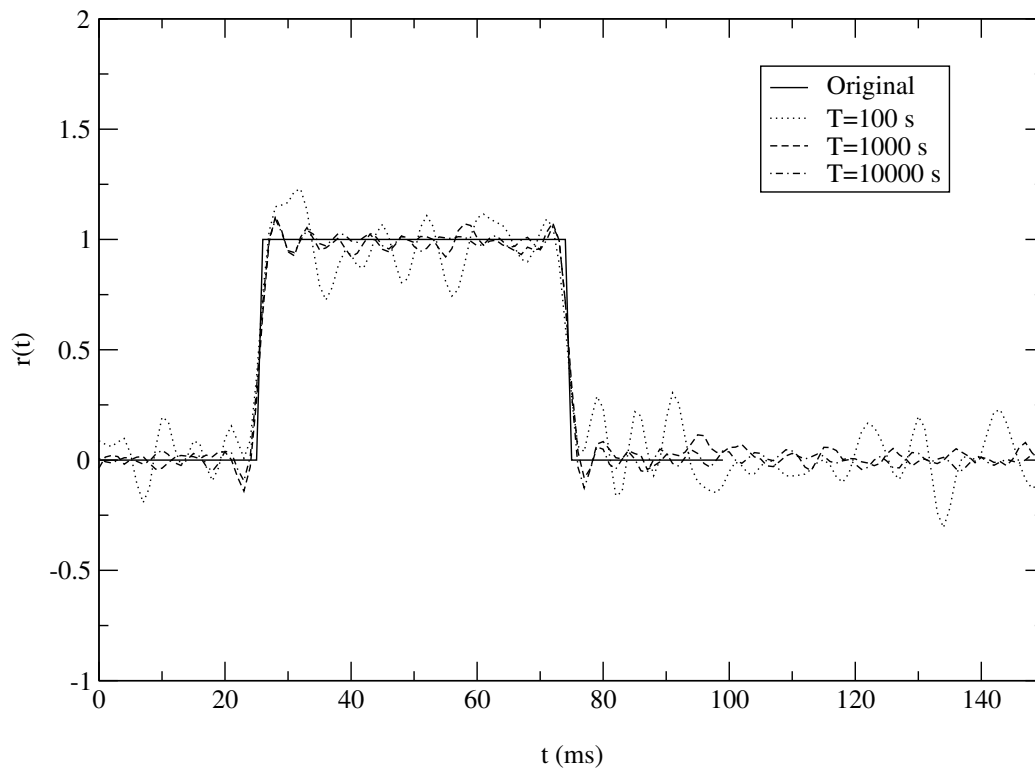


Abbildung 6.21: Wie Abbildung 6.17, also $\sigma_n = 100$, mit $\sigma = 100$ und den angegebenen Messdauern T .

Die Rekonstruktion \hat{r} bei der Messdauer $T = 1000$ s ist deutlich besser als jene bei $T = 100$ s. Bei einer weiteren Erhöhung der Messdauer ist keine weitere Verbesserung der Rekonstruktion erkennbar.

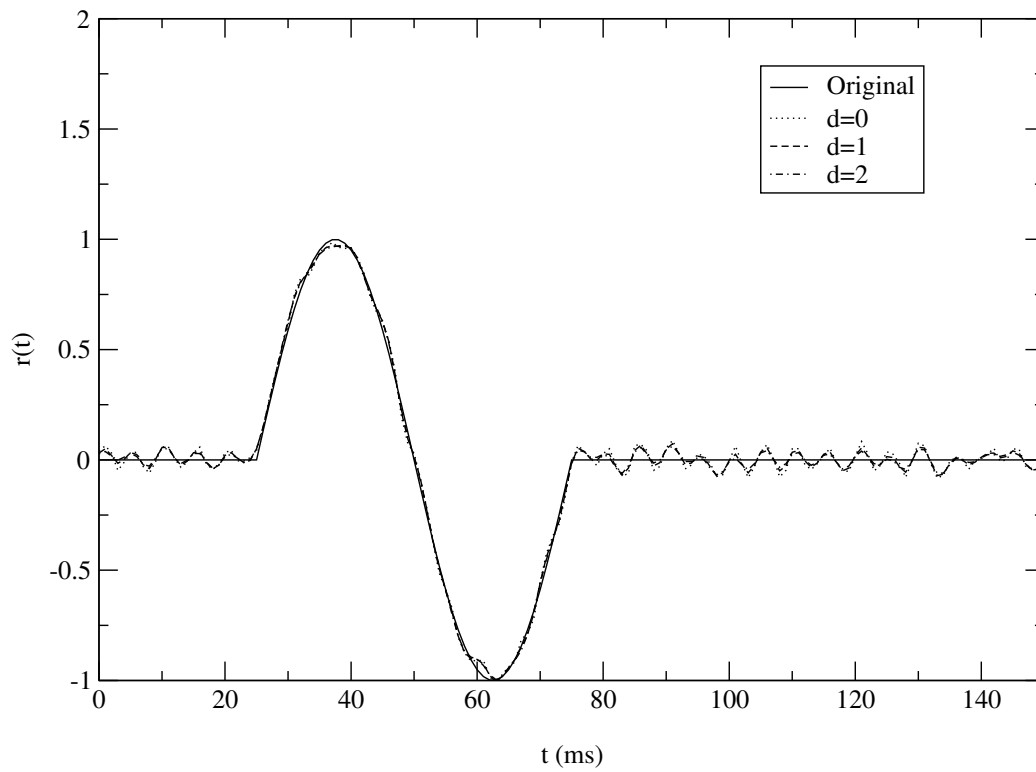


Abbildung 6.22: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 10$ und verschiedenen Ordnungen d , wobei das rezeptive Feld nun $r(t) = \sin[2\pi f_{\text{RF}}(t - 0.025 \text{ s})]$ ist. Ansonsten wie Abbildung 6.16 mit $\sigma_n = 10$. Die Ordnung d der Regularisierung hat hier, wie in Abbildung 6.16, geringen Einfluss auf die Rekonstruktion.

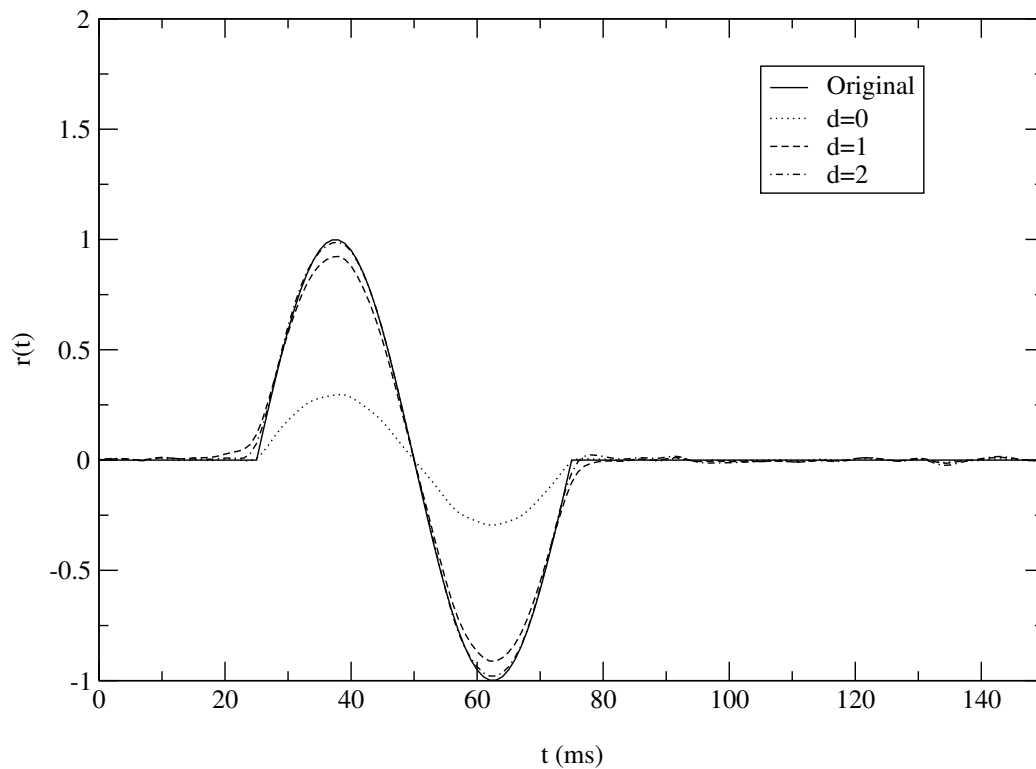


Abbildung 6.23: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 1000$ und verschiedenen Ordnungen d , wie in Abbildung 6.22.

Ein Regularisierungsparameter $d > 0$ verhindert, wie in Abbildung 6.19, dass die betragsmäßige Stärke des rezeptiven Feldes r unterschätzt wird.

6.3 Hochpassrauschen als Eingabe

In diesem Abschnitt wird untersucht, wie die Rekonstruktionen des rezeptiven Feldes aussehen, wenn nur Hochpassrauschen als Eingabe verwendet wird. Dies ist etwas akademisch, da man schon erwarten kann, dass Hochpassrauschen als Eingabe eben nur den hochfrequenten Anteil des rezeptiven Feldes rekonstruieren kann, was denn auch bestätigt wird.

6.3.1 Einfluss der unteren Grenzfrequenz

Man kann erwarten sich dass ein hauptsächlich tieffrequentes rezeptives Feld wie in Gleichung 6.2.1 durch Hochpassrauschen als Eingabe nur dann rekonstruieren lässt, wenn die untere Grenzfrequenz des Hochpassrauschens tief genug ist. Dies bestätigt Abbildung 6.24.

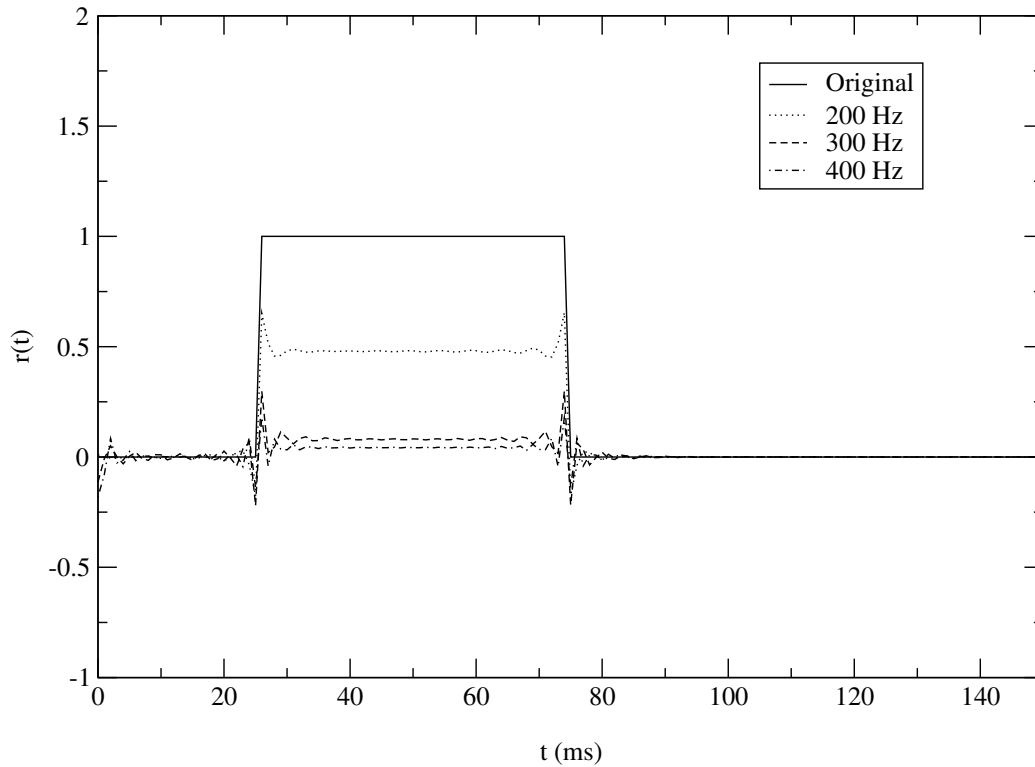


Abbildung 6.24: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 1$ und Ordnung $d = 0$. Das „Neuron“ ist linear, d. h. die neuronale Ausgabe ist $y = r \star x$. Die Eingabe x ist Hochpassrauschen der Dauer 100 s mit unteren Grenzfrequenzen 200, 300 und 400 Hz. Das „echte“ rezeptive Feld r (durchgezogene Linie) ist rechteckig, $r(t) = 1$ für $t \in [25 \text{ ms}, 75 \text{ ms}]$, $r(t) = 0$ sonst. Bei der unteren Grenzfrequenz 200 Hz ergibt sich eine perfekte Rekonstruktion, bei 300 und 400 Hz erscheint nur der entsprechende Hochpassanteil des rezeptiven Feldes.

6.3.2 Einfluss des Rauschens in der Messung

Welchen Einfluss hat Rauschen auf die Messung mit Hochpassrauschen als Eingabe? Im Wesentlichen hängt das Ergebnis stark von der Wahl des Regularisierungsparameters σ und der unteren Grenzfrequenz ab, wie Abbildungen 6.25 und 6.26 zeigen. Bei stärkerem Rauschen ist aufgrund der fehlenden tiefen Frequenzen in der Eingabe eine Rekonstruktion des rezeptiven Feldes nicht mehr möglich (Abbildung 6.27).

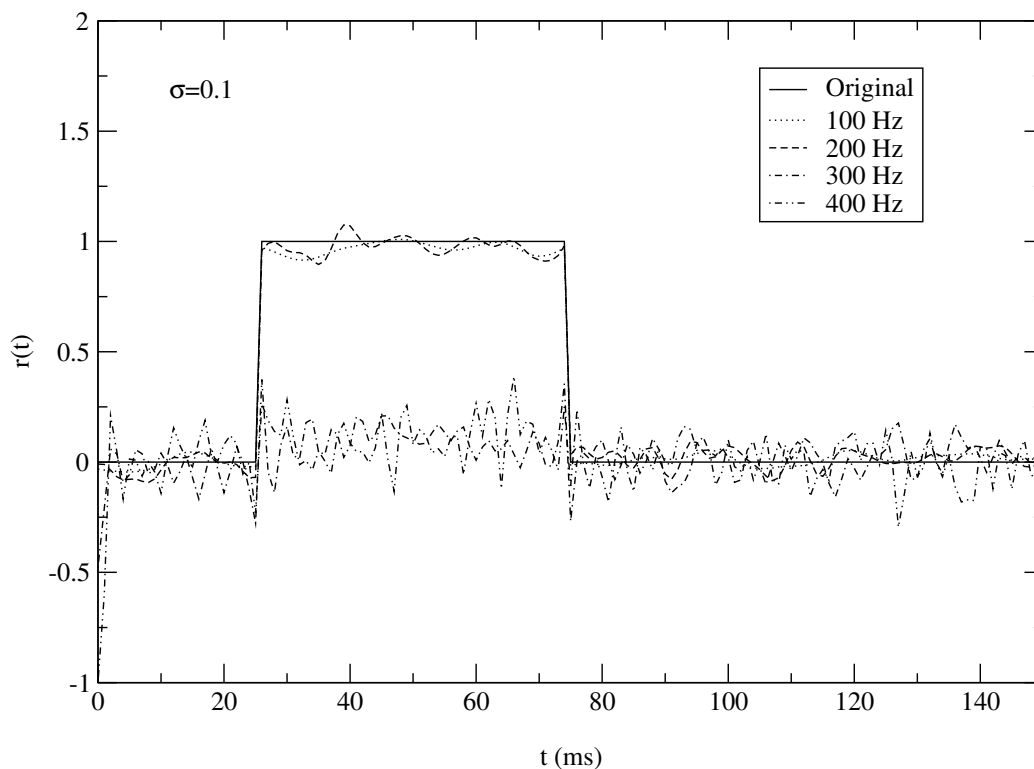


Abbildung 6.25: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und Ordnung $d = 0$. Das „Neuron“ ist linear mit Gauß'schem weißem Rauschen, d. h. die neuronale Ausgabe ist $y = r \star x + \sigma_n n$. Hier ist $\sigma_n = 0.1$. Die Eingabe x ist Hochpassrauschen der Dauer 100s mit unteren Grenzfrequenzen 200, 300 und 400 Hz. Das „echte“ rezeptive Feld r (durchgezogene Linie) ist rechteckig, $r(t) = 1$ für $t \in [25 \text{ ms}, 75 \text{ ms}]$, $r(t) = 0$ sonst.

Bei den niedrigeren unteren Grenzfrequenzen 100 und 200 Hz und entsprechendem Regularisierungsparameter $\sigma = 0.1$ ist die Rekonstruktion gut.

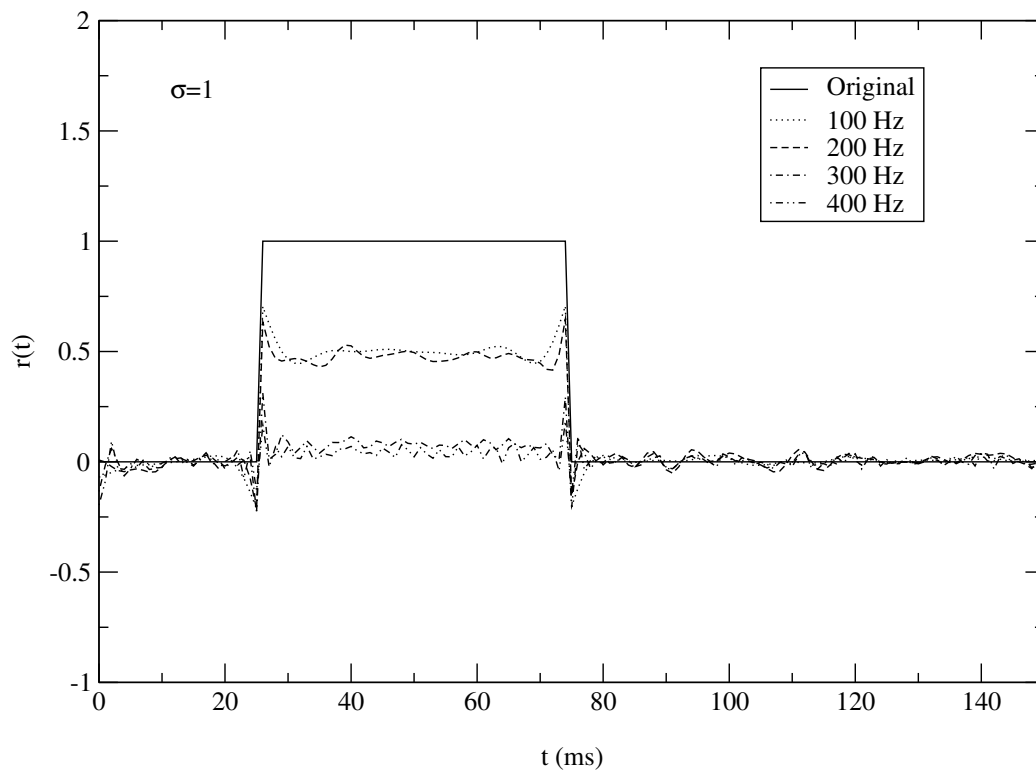


Abbildung 6.26: Wie Abbildung 6.25. Beim größeren Regularisierungsparameter $\sigma = 1$ wird der Hochpassanteil als „Spitzen“ bei 25 und 75 ms rekonstruiert und die Stärke des rezeptiven Feldes im Bereich $r(t) = 1$ wird unterschätzt.

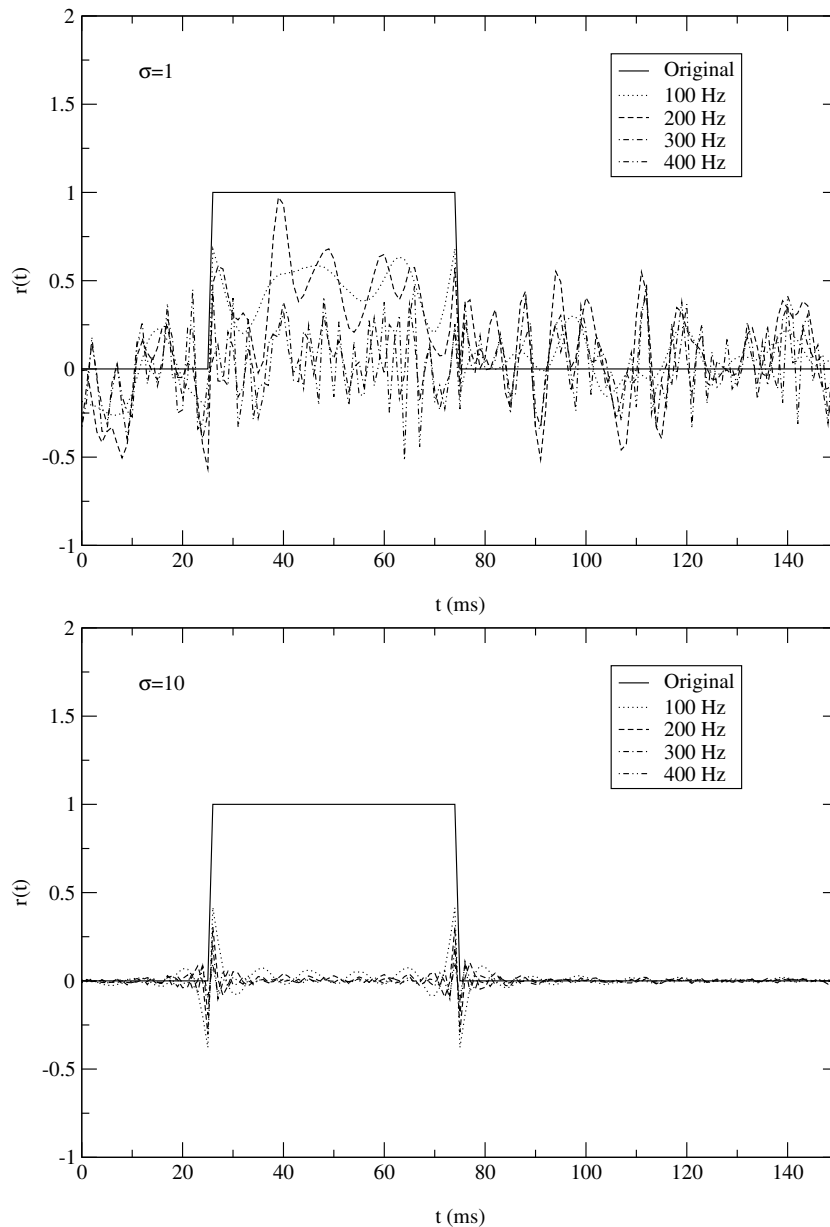


Abbildung 6.27: Rekonstruierte receptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und Ordnung $d = 0$, wie in Abbildung 6.25 mit $\sigma_n = 1$. Eine auch nur annähernde Rekonstruktion ist bei der hier vorhandenen Stärke des Rauschens in der Messung nicht mehr möglich.

6.4 Tiefpassrauschen plus Hochpassrauschen als Eingabe

Wir haben in den vorherigen Abschnitten gesehen, dass sich rezeptive Felder mit hauptsächlich tieffrequentem Anteil durch Tiefpassrauschen als Eingabe gut rekonstruieren lassen, mit Hochpassrauschen dagegen nur mäßig oder gar nicht. Was passiert mit Tiefpassrauschen plus Hochpassrauschen als Eingabe, wobei der Tiefpassanteil eine höhere Intensität hat als der Hochpassanteil? Abbildungen 6.28 und 6.29 zeigen, dass sich durch den zusätzlichen Hochpassanteil bei geeigneter Wahl des Regularisierungsparameters σ fast perfekte Rekonstruktionen ergeben.

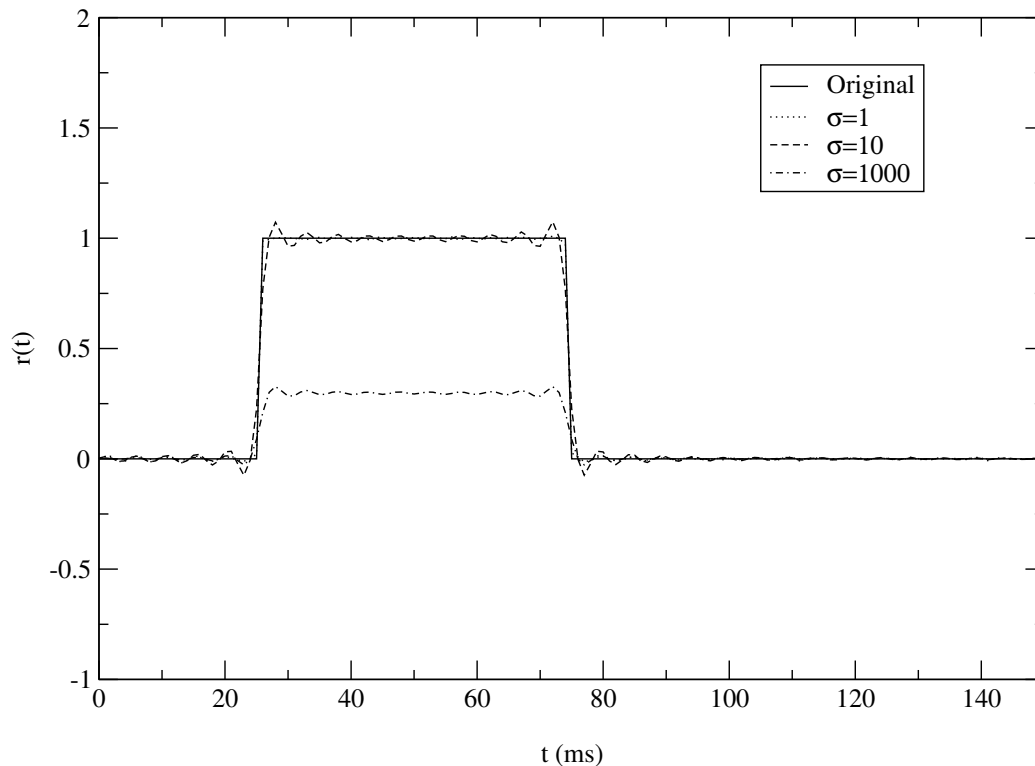


Abbildung 6.28: Rekonstruierte receptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern und Ordnung $d = 0$. Das „Neuron“ ist linear, d. h. die neuronale Ausgabe ist $y = r * x$. Die Eingabe x ist Tiefpass- plus Hochpassrauschen der Dauer 100 s. Der Tiefpassanteil bis 200 Hz hat eine 100-fach größere Standardabweichung als der Hochpassanteil. Das „echte“ receptive Feld r (durchgezogene Linie) ist rechteckig, $r(t) = 1$ für $t \in [25 \text{ ms}, 75 \text{ ms}]$, $r(t) = 0$ sonst.

Bei geeignetem Regularisierungsparameter $\sigma = 1$ erhält man eine fast perfekte Rekonstruktion. Bei zu großen $\sigma \in \{10, 1000\}$ ergibt sich eine Rekonstruktion ähnlich der mit Tiefpassrauschen allein (vergleiche Abbildung 6.1).

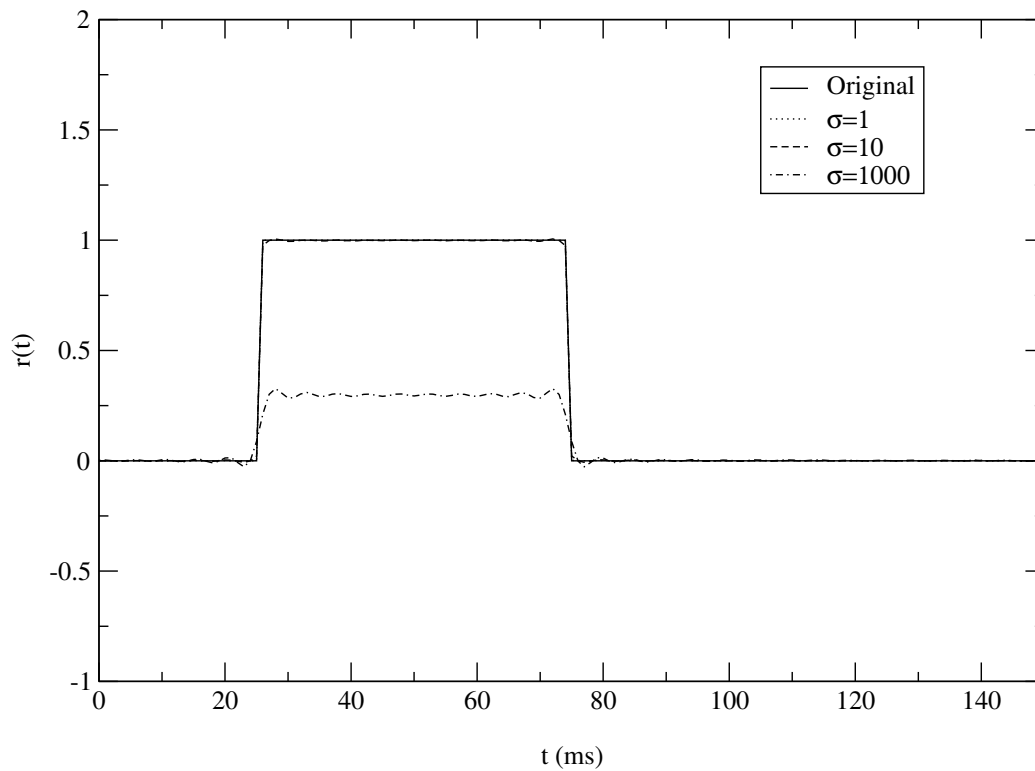


Abbildung 6.29: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern und Ordnung $d = 0$ wie in Abbildung 6.28. Der Tiefpassanteil bis 200 Hz hat nun eine nurmehr 10-fach größere Standardabweichung als der Hochpassanteil. Die Rekonstruktionen sind besser als in Abbildung 6.28 und robuster gegenüber großen Regularisierungsparametern σ .

6.5 Cochleogramm mit Tönen als Eingabe

Bisher waren die Eingaben ziemlich artifiziell, nämlich künstliches Tiefpass- und Hochpassrauschen. Bei realen Messungen kann auf der Basilarmembran niemals Tiefpassrauschen oder dergleichen erzeugt werden. Vielmehr kann nur ein beliebiger Stimulus *auf das Trommelfell* gegeben werden, der dann nach den Vorverarbeitungsschritten aus Abschnitt 3.3 die Basilarmembran anregt. Wir haben gesehen, dass Rauschen oder zumindest ein Stimulus, der viele Frequenzen enthält, auf der Basilarmembran gut geeignet wäre, um rezeptive Felder der nachgeschalteten auditorischen Neuronen zu rekonstruieren. Deshalb werden in diesem Abschnitt zufällige Pieptöne als Stimuli untersucht. Ohne Rauschen (Abbildung 6.30) lässt sich mit diesem Stimulus das rezeptive Feld perfekt rekonstruieren. Mit Rauschen (Abbildungen 6.31 und 6.32) muss der Regularisierungsparameter σ geeignet gewählt werden, um eine akzeptable Rekonstruktion zu erzielen.

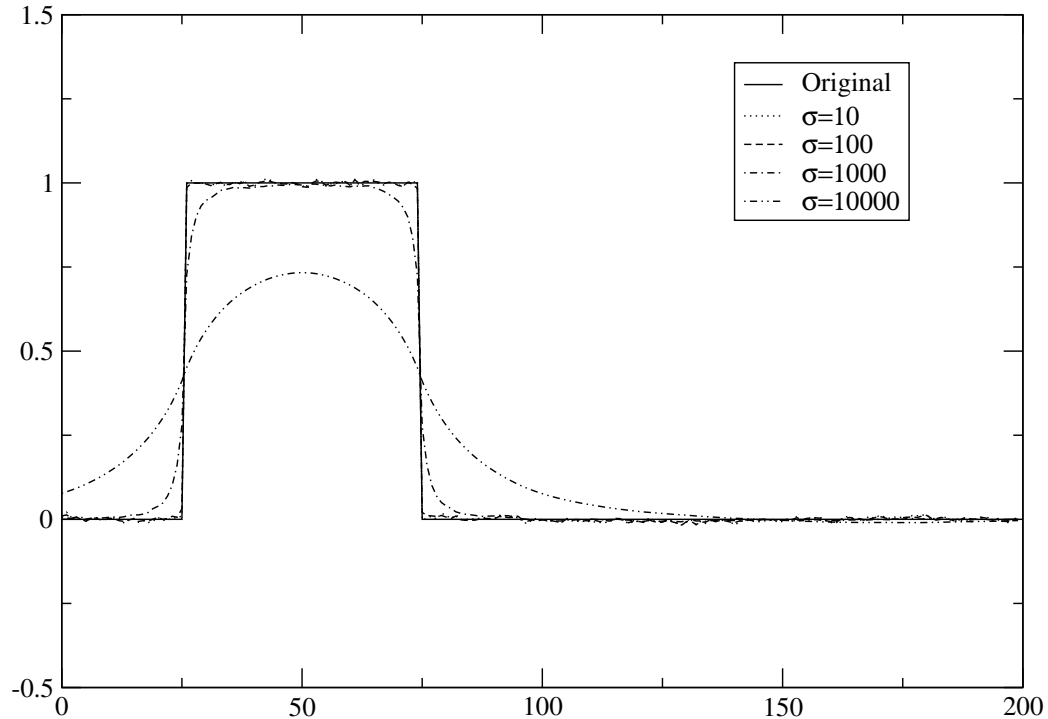


Abbildung 6.30: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern und Ordnung $d = 0$.

Eingabe in das als linear angenommene Neuron ist ein Cochleogramm bei der Frequenz 1000 Hz. Die Cochlea wird mit zufälligen Pieptönen mit logarithmischer Frequenzverteilung zwischen 100 und 16000 Hz, Dichte 100 Töne pro Sekunde, mittlerer Dauer 100 ms (Standardabweichung 50 ms) und Hanning-Rampen von 2 ms angeregt. Die Gesamtdauer der Stimulation beträgt 100 s. Das rezeptive Feld wird mit Regularisierungsparameter $\sigma \leq 100$ perfekt rekonstruiert. Bei größeren Regularisierungsparametern wird die Struktur des rezeptiven Felder zunehmend „abgerundet“.

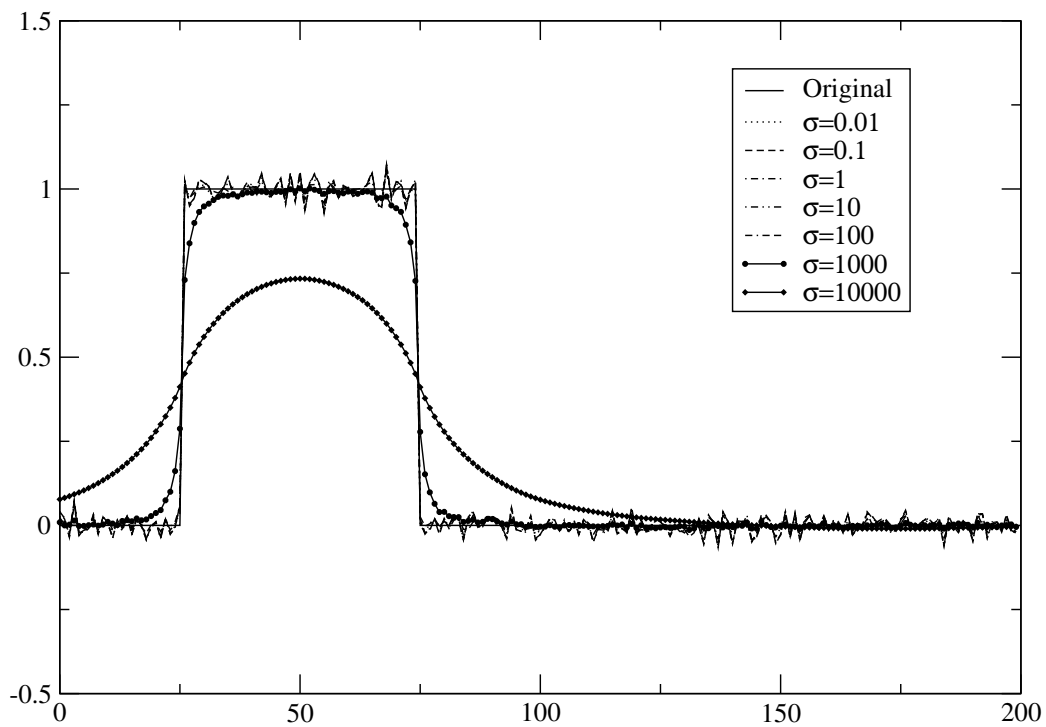


Abbildung 6.31: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern und Ordnung $d = 0$, wie Abbildung 6.30. Das Neuron ist nun mit Rauschen der Standardabweichung $\sigma_n = 10$ behaftet.

Das rezeptive Feld wird mit $\sigma \leq 100$ annähernd perfekt rekonstruiert. Aufgrund des Rauschens in der Messung ist die Rekonstruktion bei kleinen σ ebenfalls etwas verrauscht, größere σ wirken „glättend“.

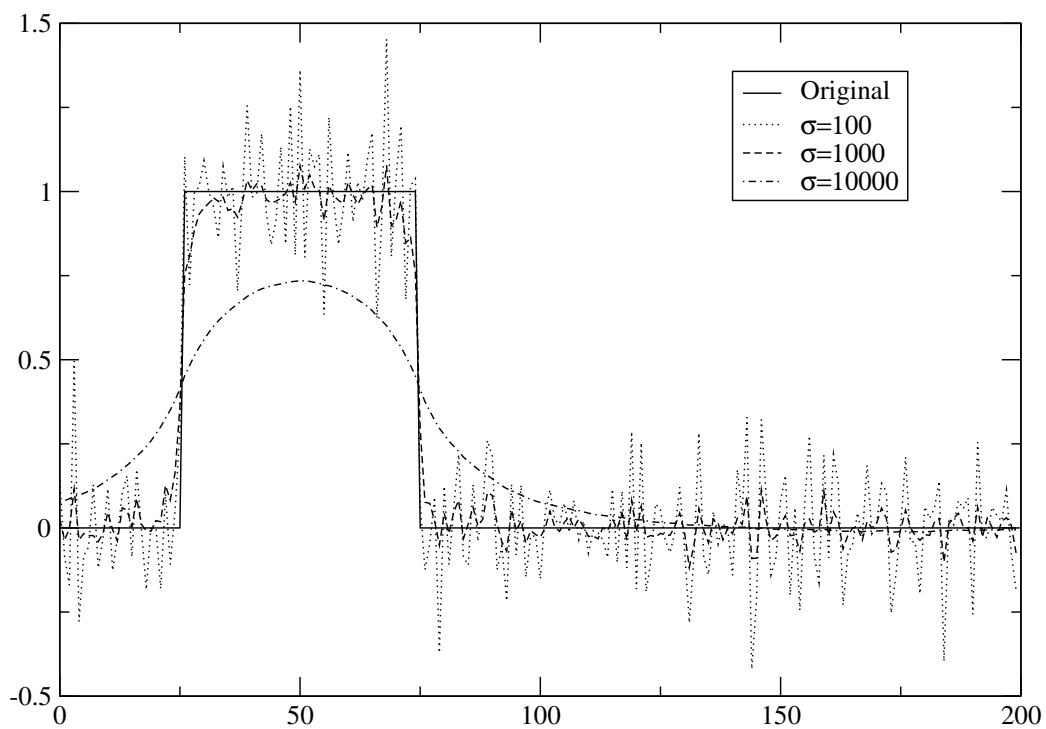


Abbildung 6.32: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern und Ordnung $d = 0$ wie Abbildung 6.31, nur nun mit $\sigma_n = 100$.

Das rezeptive Feld wird nurmehr grob rekonstruiert, eine akzeptable Rekonstruktion ergibt sich allenfalls mit $\sigma = 1000$.

6.6 Spikende Neuronen

Die Neuronen werden nun als nichtlineare Elemente modelliert. Für die neuronale Ausgabe gilt

$$y(t) = H[(r \star x)(t) - \vartheta],$$

wobei H die Heaviside-Funktion mit

$$H(s) = \begin{cases} 0 & \text{falls } s \leq 0 \\ 1 & \text{falls } s > 0 \end{cases}$$

ist. Die Größe ϑ wird als Schwelle bezeichnet. Stimuli sind wieder Cochleogramme wie in Abschnitt 6.5. Nun ist keineswegs klar, dass Optimierungsmethoden wie in den vorangegangenen Abschnitten auch bei solchen nichtlinearen Schwellenelementen vernünftige Ergebnisse erzielen. Dennoch zeigt Abbildungen 6.33, dass sich auch unter dieser Bedingung zumindest noch die rechteckige Form des hier verwendeten rezeptiven Feldes annähernd rekonstruieren lässt. Die absolute Größe des rezeptiven Feldes $r(t)$ wird allerdings bei den großen Regularisierungsparametern σ , die hier notwendig sind, weit unterschätzt. Höhere Ordnungen d der Regularisierung bringen hier leider keine Verbesserung; vgl. Abbildungen 6.34 und 6.35.

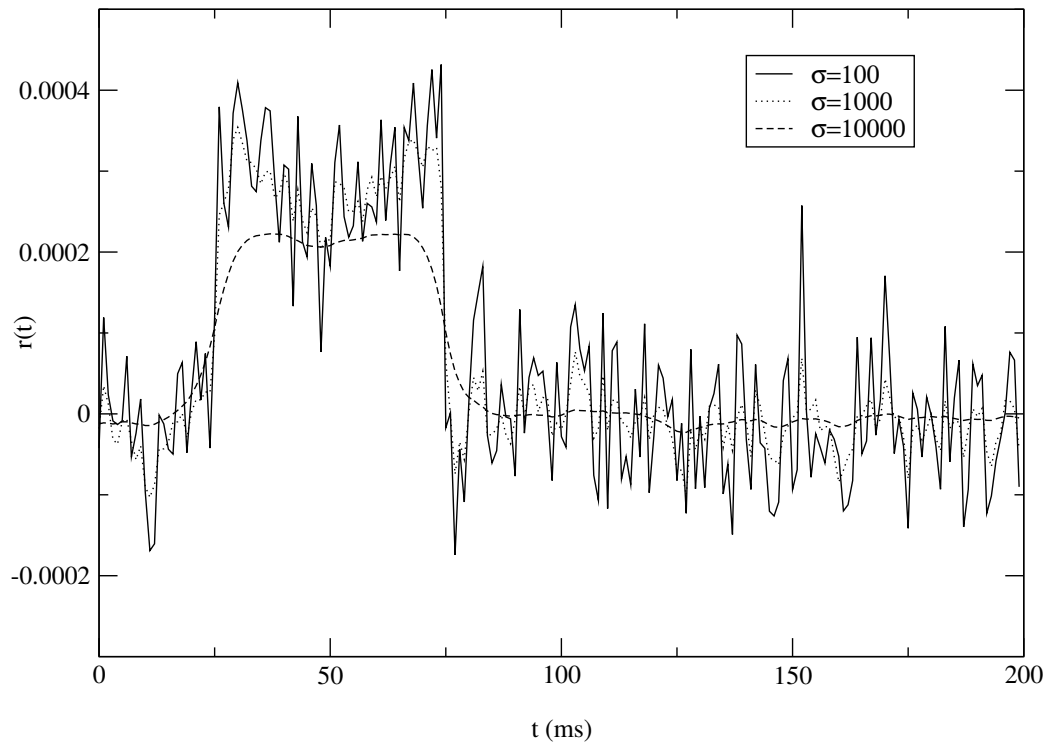


Abbildung 6.33: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und Ordnung $d = 0$.

Eingabe in das als Schwellenelement angenommene Neuron ist ein Cochleogramm bei der Frequenz 1000 Hz. Die Cochlea wird mit zufälligen Pieptönen mit logarithmischer Frequenzverteilung zwischen 100 und 16000 Hz, Dichte 200 Töne pro Sekunde, mittlerer Dauer 10 ms (Standardabweichung 10 ms) und Hanning-Rampen von 2 ms angeregt. Die Gesamtdauer der Stimulation beträgt 100 s.

Zumindest die rechteckige Form des rezeptiven Feldes wird bei geeignetem Regularisierungsparameter $\sigma = 10000$ annähernd wiedergegeben. Das Messergebnis ist allerdings mit starkem Rauschen behaftet.

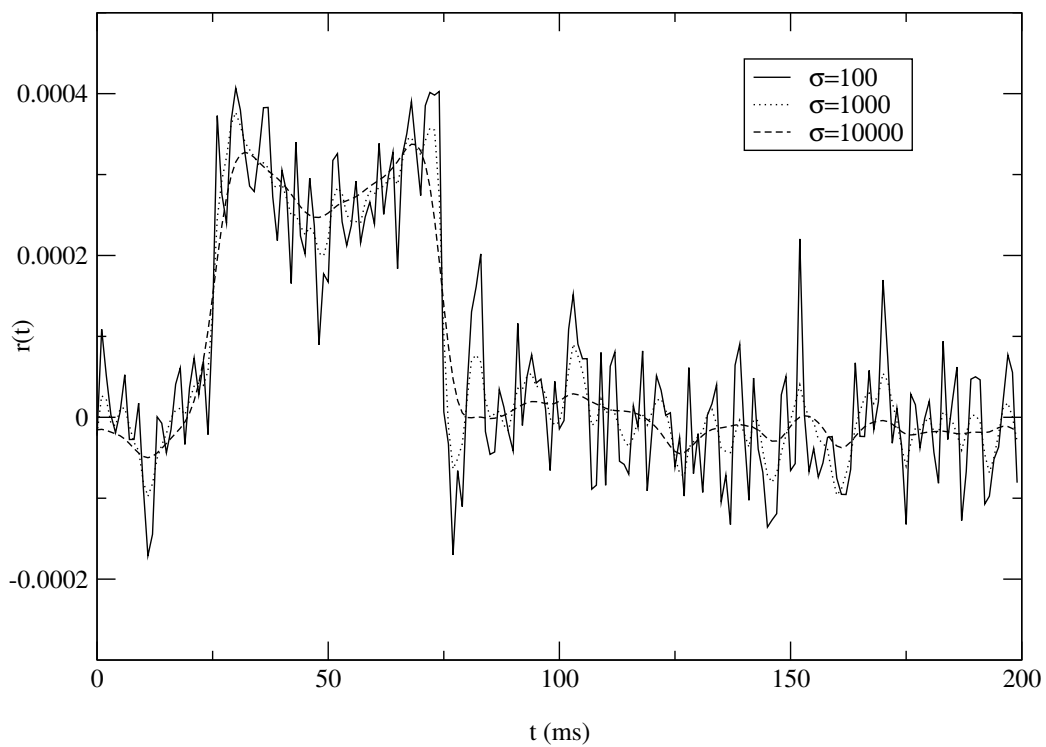


Abbildung 6.34: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und Ordnung $d = 1$, sonst wie Abbildung 6.33.

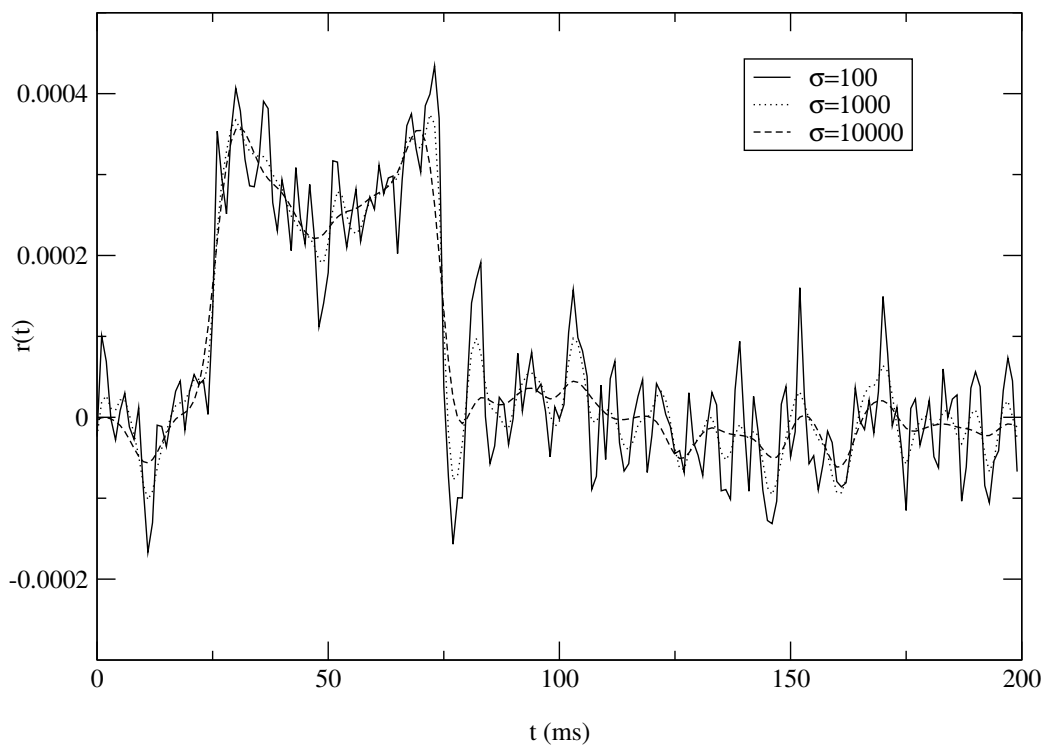


Abbildung 6.35: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit den angegebenen Regularisierungsparametern σ und Ordnung $d = 2$, sonst wie Abbildung 6.33.

A hypothesis or theory is clear, decisive, and positive, but it is believed by no one but the man who created it. Experimental findings, on the other hand, are messy, inexact things, which are believed by everyone except the man who did that work.

Harlow Shapley

7

Lineare rezeptive Felder mit Regularisierung in mehreren Dimensionen

Auditorische Neurone reagieren i. A. nicht nur auf einen einzigen Frequenzkanal, sondern auf mehrere Kanäle in unterschiedlicher Intensität. Hier kann die multidimensionale Erweiterung der Regularisierungsmethode aus Abschnitt 5.2 angewandt werden. Als Eingabe dient wie in Abschnitt 6.5 ein Cochleogramm, das aus zufälligen Pieptönen erzeugt wird. Das Neuronenmodell ist ein Schwellenelement wie in Abschnitt 6.6. Abbildung 7.1 zeigt, dass eine Rekonstruktion des rezeptiven Feldes auch unter diesen Bedingungen möglich ist.

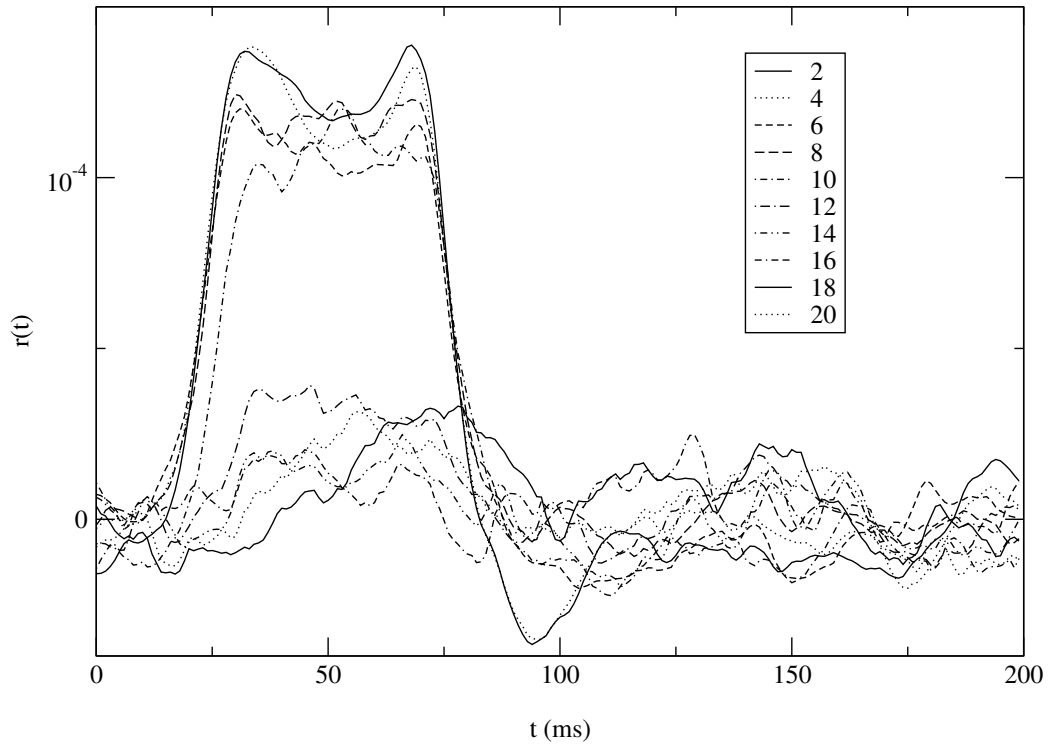


Abbildung 7.1: Rekonstruierte rezeptive Felder \hat{r} nach der Regularisierungsmethode mit Regularisierungsparameter $\sigma = 10000$ und Ordnung $d = 0$. Eingabe in das als Schwellenelement angenommene Neuron ist ein Cochleogramm mit 20 Frequenzkanälen. Die Cochlea wird mit zufälligen Pieptönen mit logarithmischer Frequenzverteilung zwischen 100 und 16000 Hz, Dichte 200 Töne pro Sekunde, mittlerer Dauer 10 ms (Standardabweichung 10 ms) und Hanning-Rampen von 2 ms angeregt. Die Gesamtdauer der Stimulation beträgt 100 s. Das rezeptive Feld ist $r_i(t) = 1$ für $25 \text{ ms} \leq t \leq 75 \text{ ms}$ und Frequenzkanal $i \leq 10$ und $r_i(t) = 0$ sonst. Die Form des rezeptiven Feldes wird gut rekonstruiert.

*In theory there is little difference
between theory and practice. In
practice there is.*

unbekannter Autor

8

Statistische Lerntheorie

In diesem Kapitel werden einige Grundlagen aus der statistischen Lerntheorie vorgestellt. Dabei wird auf zwei konkrete Ansätze näher eingegangen, nämlich Support-Vektor-Maschinen und Induktives Schließen mit Hilfe der Kolmogorov-Komplexität.

8.1 Einführung

Was versteht man unter „statistischer Lerntheorie“? Dazu muss man zuerst definieren, was „lernen“ ist. Lernen ist eine Verhaltensänderung aufgrund von Erfahrung (Zimbardo 1992), d. h. ein Tier oder eine Maschine verhält sich anders, nachdem es oder sie etwas gelernt hat, als vorher. Weiter sollte sich das Verhalten möglichst so ändern, wie es der Lehrer vom Lernenden verlangt, d. h. nach dem Lernprozess sollte der Schüler richtige Antworten anstatt falscher oder zufälliger Antworten geben.¹

Lernen kann auf mindestens zwei Arten stattfinden. Zunächst gibt es Lernen durch Mitteilung. Der Lehrer sagt dem Schüler einfach, was er tun soll oder nach welchem Verfahren er die Antwort erzeugen soll. Damit man dem Schüler etwas mitteilen kann, muss er allerdings zuerst sprechen oder zumindest *verstehen* können. Man kann ihm nicht sagen, wie er sprechen soll. d. h. der Schüler kann letztlich nur anhand von Beispielen oder Versuch und

¹Oder jedenfalls Antworten, welche der Lehrer für richtig *hält*.

Irrtum lernen.² Diese Art von Lernen bezeichnet man auch als „statistisches Lernen“. Die „statistische Lerntheorie“ befasst sich mit statistischem Lernen. Sie lässt sich in drei Probleme unterteilen.

- Wie kann man ein statistisches Lernproblem mathematisch exakt formulieren (möglichst allgemein und möglichst so, dass es auch lösbar ist)?
- Unter welchen Bedingungen kann das statistische Lernproblem gelöst werden?
- Durch welche Algorithmen kann das statistische Lernproblem gelöst werden?

Zum letzten Punkt gehört auch die Frage, wie das statistische Lernproblem möglichst gut gelöst werden kann. Wir werden später noch sehen, was „gut“ bedeutet. Außerdem sollen die Algorithmen das statistische Lernproblem möglichst effizient lösen, also unter Verbrauch von möglichst wenig Rechenzeit.

²Das Lernen nach Versuch und Irrtum entspricht der „operanten Konditionierung“ nach Burrhus Frederick Skinner: Eine „richtige“ Reaktion der Versuchsperson oder des Versuchstieres auf einen Reiz wird belohnt, eine „falsche“ wird bestraft. Falls durch dieses „Training“ eine nachhaltige Verhaltensänderung des Versuchsobjekts bewirkt wird, sagt man, das Versuchsobjekt habe eine Reiz-Reaktions-Beziehung „gelernt“. Bereits Edward L. Thorndike erkannte den Zusammenhang zwischen einer Belohnung und einer Verhaltensänderung und nannte sie „instrumentelle Konditionierung“. Im Gegensatz zur operanten bzw. instrumentellen Konditionierung wird bei der „klassischen“ Konditionierung nach Iwan Pawlow eine Beziehung zwischen *zwei* Reizen gelernt. Gibt man einem Hund beispielsweise immer nach dem Läuten einer Glocke Futter, so löst nach einigen Versuchen das Läuten der Glocke *allein* schon Speichelfluss aus, welcher normalerweise durch Futter ausgelöst wird. Hier lernt das Versuchstier also eine kausale oder zumindest korrelative Beziehung zwischen zwei Reizen, nämlich „Glockenläuten“ und „Futter“. Auch bei der klassischen Konditionierung wird Verhalten geändert, jedoch ist nicht steuerbar, *welches* Verhalten nach dem Läuten der Glocke gezeigt werden soll. Nicht nur ein *bestimmter* Reiz, sondern auch *ähnliche* Reize können das konditionierte Verhalten auslösen. Beispielsweise ein hoher Glockenklang statt ein tiefer. Dies bezeichnet man als „Generalisierung“ (Zimbardo 1992). Bei der Meeresschnecke *Aplysia* führt eine klassische Konditionierung des Kiemenrückziehreflexes zu einer Verstärkung entsprechender Synapsen im neuronalen System (Kandel et al. 1996). Lernen nach Versuch und Irrtum in unserem Sinne entspricht der *operanten* Konditionierung, da das Versuchsobjekt sein Verhalten, d. h. seine Reaktion für „richtig“ und für „falsch“ an vorgegebene Reiz-Reaktions-Beziehungen anpassen soll. Wie wir später sehen werden, ist gerade die Generalisierung das Hauptproblem beim Lernen.

8.2 Beispiele

Statistisches Lernen soll zunächst anhand zweier einfacher Beispiele erläutert werden.

8.2.1 Lernen von „groß“ und „klein“

Der Lehrer zieht zufällig eine Zahl aus dem Intervall $[0, 1]$ und klassifiziert diese Zahlen jeweils als „groß“, wenn sie größer als die Konstante c sind und „klein“, wenn sie kleiner oder gleich c sind.³

Der Schüler will diese Klassifizierungsaufgabe lernen. Er kennt die Vorgehensweise des Lehrers, der Wert von c ist ihm aber unbekannt. Um den Wert von c zu schätzen, bildet er die Hypothese $c = c_0$, wobei c_0 der größte Wert sei, der als „klein“ klassifiziert wurde.

Sei $p(I)$ die Wahrscheinlichkeit, dass der Lehrer eine Zahl aus dem Intervall $I \subseteq [0, 1]$ zieht. Dann ist die Wahrscheinlichkeit, dass der Schüler nach ℓ Beispielen ein neues Beispiel falsch klassifiziert, nur mit Wahrscheinlichkeit $P_\ell \leq (1 - \varepsilon)^\ell$ größer als ε , d. h.

$$P_\ell(p(\cdot|c_0(\ell), c]) \geq \varepsilon) \leq (1 - \varepsilon)^\ell. \quad (8.1)$$

Hierbei muss der Lehrer die Zahlen nicht gleichverteilt aus $[0, 1]$ ziehen, jedoch muss die Wahrscheinlichkeitsverteilung p , mit der die Zahlen gezogen werden, festliegen. Auch die Konstante c muss nicht gleichverteilt aus $[0, 1]$ gezogen worden sein. Eine „Lösung“ des Lernproblems ist ein Algorithmus zur Entscheidung des Schülers zwischen „groß“ und „klein“ bei einer neuen Klassifizierungsaufgabe, gegeben die bereits bekannten Beispiele. Eine bessere Lösung ist eine Lösung, für welche die Wahrscheinlichkeit, dass die Irrtumswahrscheinlichkeit größer ist als ε , im ungünstigsten Fall geringer ist als das angegebene P_ℓ . Angegeben ist nur *eine* mögliche Lösung, nicht notwendigerweise die beste.

Beweis für (8.1): Nehmen wir an, der Schüler habe sich nach ℓ Beispielen auf c_0 festgelegt. Der Schüler weiß $c \geq c_0$. Es gibt normalerweise ein kleinstes c_1 , so dass die Wahrscheinlichkeit $p(\cdot|c_0, c_1] \geq \varepsilon$. Damit sich der Schüler nun mit Wahrscheinlichkeit größer oder gleich ε irrt, muss seine Hypothese c_0 größer oder gleich c_1 sein. d. h. es dürfen in der Lernphase keine Punkte aus $]c_0, c_1[$ gezogen worden sein. Die Wahrscheinlichkeit, dass dies geschehen ist, ist kleiner oder gleich $(1 - \varepsilon)^\ell$, da ja wegen der Wahl von c_1 die Wahrscheinlichkeit $p(\cdot|c_0, c_2] \leq \varepsilon$ für alle $c_2 \leq c_1$. Dies gilt unabhängig von der Wahrscheinlichkeitsverteilung p , mit der der Lehrer die Zahlen zieht. Falls es kein c_1 gibt

³Der Lehrer könnte dem Schüler auch einfach den Wert der Konstanten c mitteilen, dies wäre dann jedoch keine Art von Lernen, welches die statistische Lerntheorie behandelt.

mit $p([c_0, c_1]) \geq \varepsilon$, ist die Irrtumswahrscheinlichkeit $p([c_0(\ell), c]) \leq \varepsilon$, (8.1) gilt also ebenfalls. \square

Das statistische Lernproblem ist also lösbar in dem Sinne, dass nach dem Lernen die Irrtumswahrscheinlichkeit wahrscheinlich klein ist. Dies wird auch als „probably approximately correct“ bezeichnet. Bei ℓ Lernbeispielen gilt für die Wahrscheinlichkeit P_ℓ , dass die Irrtumswahrscheinlichkeit größer als ε ist, sogar $\lim_{\ell \rightarrow \infty} P_\ell = 0$ für $\varepsilon > 0$.

8.2.2 Statistische numerische Integration

Gegeben sei eine reelle Funktion $f : \mathbb{R}^k \rightarrow [0, B]; (x_1, \dots, x_k) \mapsto f(x_1, \dots, x_k)$ von k reellen Variablen, die Werte innerhalb $[0, B]$ annehmen kann. Gesucht ist

$$E(f) := \int_0^1 \dots \int_0^1 f(x_1, \dots, x_k) dx_1 \dots dx_k$$

Werden nun ℓ Punkte $\mathbf{x}_i = (x_1, \dots, x_k)_i$, $i = 1 \dots \ell$, gleichverteilt aus $[0, 1]^k$ gezogen, ergibt sich eine Abschätzung $\hat{E}(f)$ für $E(f)$

$$\hat{E}(f) := \frac{1}{\ell} \sum_{i=1}^{\ell} f(\mathbf{x}_i). \quad (8.2)$$

Die Wahrscheinlichkeit, dass die Abweichung $|E(f) - \hat{E}(f)| \geq \varepsilon$ ist, beträgt höchstens $2e^{-2\ell\varepsilon^2/B^2}$ (Vidyasagar 1997).

Was hat nun dieses zweite Beispiel mit Lernen zu tun? Nun, wir haben etwas über die unbekannte Funktion f gelernt, nämlich ihren Mittelwert. Zu beachten ist, dass die Wahrscheinlichkeit auch hier „doppelt gemoppelt“ ist. Die Wahrscheinlichkeit dafür, dass wir uns stark irren, ist gering. Bei allen statistischen Lernaufgaben ist es *möglich*, dass wir uns *sehr* stark irren. Nur mit einer gewissen *Wahrscheinlichkeit* haben wir recht, und das auch nur ungefähr. Dies liegt daran, dass die Lernbeispiele zufällig ja vollkommen untypisch ausgewählt worden sein könnten. Beispielsweise könnte $f(\mathbf{x}) \in \{0, B\}$ sein und in der Lernphase wurden zufällig nur solche \mathbf{x} ausgewählt, für die $f(\mathbf{x}) = B$ ist. Somit wäre der nach (8.2) geschätzte Mittelwert gleich B . Mit hoher Wahrscheinlichkeit sei jedoch $f(\mathbf{x}) = 0$, so dass tatsächlich $E(f) \approx 0$ ist. Die Wahrscheinlichkeit für einen solch erheblichen Irrtum ist jedoch gering, da *falls* mit hoher Wahrscheinlichkeit $f(\mathbf{x}) = 0$ ist, in der Lernphase auch mit hoher Wahrscheinlichkeit Argumente \mathbf{x} mit $f(\mathbf{x}) = 0$ ausgewählt werden.

Falls f stetig ist und die Ableitung beschränkt, dann kann man f konventionell integrieren, indem man f einfach auf einem rechteckigen Gitter berechnet. Der Fehler ist dann allerdings proportional zu einer Potenz des Abstands der Punkte in einer Dimension (Stoer und Bulirsch 1994, Press et al. 1995). Dieser ist $\ell^{-1/k}$. Der Fehler hängt also von der Dimension k ab und konvergiert bei großem k nur sehr langsam. Verzichtet man auf die absolute Sicherheit, erreicht man wie oben angegeben eine viel schnellere, nämlich exponentielle, Konvergenz. Allerdings nicht Konvergenz des *größtmöglichen* Fehlers, sondern nur Konvergenz der *Wahrscheinlichkeit* gegen null, dass der Fehler größer als ε ist.⁴

8.3 Lernende Maschinen

Die in der Einleitung und den Beispielen angesprochenen Konzepte sollen nun mathematisch exakt formuliert werden.

8.3.1 Klassifikatoren

Es werden ℓ Zufallsvektoren („Punkte“) \mathbf{x}_i , $i = 1 \dots \ell$, mit Wahrscheinlichkeitsverteilung $p : \mathbf{x} \mapsto p(\mathbf{x})$ gezogen. Jeder Punkt wird von einem Lehrer durch einen Klassifikator K als „0“ oder „1“ klassifiziert. Der Klassifikator ist also eine Funktion $K : \mathbb{R}^k \rightarrow \{0, 1\}$; $\mathbf{x} \mapsto K(\mathbf{x})$.

Die lernende Maschine hat nun einen Algorithmus A , der Mengen von ℓ klassifizierten Punkten (Trainingsdaten) auf Klassifikatoren abbildet, etwa

$$A[(\mathbf{x}_1, K(\mathbf{x}_1)), \dots, (\mathbf{x}_\ell, K(\mathbf{x}_\ell))] = K'$$

Die Abweichung $d_p(K, K')$ zwischen K und K' sei die Wahrscheinlichkeit, dass sich K' irrt, d. h. dass ein \mathbf{x} gezogen wird, so dass $K'(\mathbf{x}) \neq K(\mathbf{x})$.

Sei \mathcal{K} die Menge der in Frage kommenden Klassifikatoren und \mathcal{P} die Menge der in Frage kommenden Wahrscheinlichkeitsverteilungen p für die Punkte \mathbf{x} . Betrachte nun die Wahrscheinlichkeit $r(\ell, \varepsilon)$, dass die Abweichung $d_p(K, K')$ größer ist als ε im ungünstigsten Fall:

$$r(\ell, \varepsilon) := \sup_{K \in \mathcal{K}} \sup_{p \in \mathcal{P}} P\{(\mathbf{x}_1, \dots, \mathbf{x}_\ell) | d_p(K, K') > \varepsilon\} \quad (8.3)$$

Mit Hilfe der Wahrscheinlichkeit $r(\ell, \varepsilon)$ für große Abweichungen können wir die folgende Definition formulieren.

⁴Diesen Sachverhalt nutzt man auch beim Monte-Carlo-Verfahren (Press et al. 1995) zur Integration und beim Metropolis-Algorithmus (Press et al. 1995) zur Integration über eine Wahrscheinlichkeitsverteilung, die durch eine Master-Gleichung (Haken 1990) beschrieben werden kann.

Definition 1 (wahrscheinlich ungefähr korrekt)

Der Algorithmus A heißt „wahrscheinlich ungefähr korrekt (PAC=*probably approximately correct*) mit Genauigkeit ε “, falls $\lim_{\ell \rightarrow \infty} r(\ell, \varepsilon) = 0$. Der Algorithmus heißt „wahrscheinlich ungefähr korrekt“, falls $\lim_{\ell \rightarrow \infty} r(\ell, \varepsilon) = 0$ für jedes $\varepsilon > 0$.

Der Klassifikator K heißt „wahrscheinlich ungefähr korrekt lernbar“, falls ein wahrscheinlich ungefähr korrekter Algorithmus existiert.

Das beschriebene Lernproblem ist noch zu speziell, denn es umfasst nur Klassifikationsaufgaben. Dies sieht dann etwa so aus: Der Lehrer oder auch der Schüler zieht zufällige Aufgaben, etwa $1 + 2 = 3$ und $2 + 2 = 5$. Der Schüler sagt dann voraus, ob er einen Haken oder ein „f“ bekommt. Der Schüler sollte jedoch i. d. R. den *Wert* der Lösung vorhersagen. Er bekommt also die Aufgaben $(1, 2)$ und $(2, 2)$ und soll einen Wert zuordnen. Dabei soll er sich dann möglichst selten irren. Eine weitere typische Lernaufgabe ist die Interpolation. Eine Funktion wird an einigen Stellen gemessen. Aufgrund dieser Daten soll dann der Funktionswert an anderen Stellen vorhergesagt werden. Dabei ist es nicht notwendig, dass man sich besonders selten irrt, sondern dass die mittlere Abweichung von Vorhersage und tatsächlichem Wert möglichst gering ist. Die Anforderungen können ganz verschieden sein. Man könnte auch der Meinung sein, die „ $2+2=5$ “ des Schülers sei immerhin *fast* richtig.

Es ist in Abweichung zu (8.3) auch möglich, dass der Schüler gar nicht alle Klassifikatoren \mathcal{K} realisieren kann, die der Lehrer realisieren kann, weil der Schüler einfach zu dumm ist oder weil die Funktionenmenge, die zur Interpolation benötigt wird, die wirkliche, richtige Funktion gar nicht enthält.

Falls sich die Natur natürlich mathematisch oder mit einem Algorithmus beschreiben lässt *und* man nur Ja-Nein-Fragen stellt, beschreibt das Lernen von Klassifikationsaufgaben die empirische Forschung. Experimente sollten ja grundsätzlich zufällig sein, die besten Entdeckungen werden ohnehin durch Zufall gemacht. Die Frage ist nur, welchen Algorithmus zur Lösung des Lernproblems man anwenden soll. Ein Algorithmus, der sicherlich *nicht* wahrscheinlich ungefähr korrekt ist, ist einfach die Versuche zu notieren, die mit „ja“ ausgegangen sind und alle anderen mit „nein“ zu beantworten. Dieser Algorithmus generalisiert allerdings nicht.

Im nächsten Abschnitt soll das statistische Lernproblem noch etwas allgemeiner formuliert werden.

8.3.2 Modellfreies Lernen

Der Lehrer zieht Zufallsvektoren aus \mathbb{R}^k . Diese sollen unabhängig voneinander sein und aus der selben Verteilung mit Wahrscheinlichkeitsdichte $p(\mathbf{x})$ gezogen werden („unabhängige, identisch verteilte Zufallsvektoren“).

Der Lehrer ordnet nun jedem Zufallsvektor \mathbf{x}_i einen Wert y_i zu. Dies geschieht nach der Wahrscheinlichkeitsdichte $p(y|\mathbf{x})$. Der Lehrer kann also auch identischen Zufallsvektoren *verschiedene* Werte zuordnen. Beispielsweise ist Rauschen in den Werten zugelassen, der „Lehrer“ kann also auch ein Messgerät sein.

Die erzeugten Trainingsdaten lauten somit

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell).$$

Die lernende Maschine kann nur eine gewisse Funktionenmenge $f(\alpha) : \mathbb{R}^k \rightarrow \mathbb{R}; \mathbf{x} \mapsto f(\mathbf{x}, \alpha)$, wobei $\alpha \in \Lambda$, realisieren. Die Menge Λ heißt „Parametermenge“.

Die lernende Maschine hat nun einen Algorithmus A , der Mengen von ℓ klassifizierten Punkten, also Paaren (\mathbf{x}_i, y_i) (Trainingsdaten), auf Parameter α abbildet, etwa

$$A((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell)) = \alpha.$$

Gegeben sei eine Funktion $L : \mathbb{R}^{k+1} \rightarrow \mathbb{R}; (\mathbf{x}, y) \mapsto L(y, f(\mathbf{x}, \alpha))$, die ein Maß für die Abweichung der Antwort $f(\mathbf{x}, \alpha)$ der lernenden Maschine von der Antwort y des Lehrers darstellt. Diese Funktion L heißt „Verlustfunktion“ (*loss function*) oder „Diskrepanz“.

Definition 2 (Generalisierungsfehler)

Der Erwartungswert $d_p(\alpha)$ der Verlustfunktion heißt „Generalisierungsfehler“ oder „Risiko“.

$$d_p(\alpha) := \int L(y, f(\mathbf{x}, \alpha)) p(\mathbf{x}, y) d\mathbf{x} dy$$

Im Fall der Klassifikationsaufgabe ist $y, f(\mathbf{x}, \alpha) \in \{0, 1\}$ und

$$L(y, f(\mathbf{x}, \alpha)) = \begin{cases} 0 & \text{falls } y = f(\mathbf{x}, \alpha) \\ 1 & \text{falls } y \neq f(\mathbf{x}, \alpha) \end{cases}$$

Damit ist $d_p(\alpha)$ die Wahrscheinlichkeit eines Irrtums.

Offenbar wird auch das Interpolationsproblem von obiger Definition des statistischen Lernproblems abgedeckt. Es ist nur $L(y, f(\mathbf{x}, \alpha)) = [y - f(\mathbf{x}, \alpha)]^2$ zu wählen.

Auch die statistische Integration wird von dieser Definition erfasst. Der Mittelwert von f ist der Wert mit der geringsten quadratischen Abweichung zwischen diesem Wert und den Funktionswerten von f (Greene 2000). Also nehme $f(\alpha) : \mathbf{x} \mapsto f(\mathbf{x}, \alpha) = \alpha$, d. h. die Funktionen sind einfach die Konstanten, und $L(y, f(x, \alpha)) = (y - \alpha)^2$. Weicht α um ε vom Mittelwert $\langle y \rangle = \int y(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$ ab, so weicht der Generalisierungsfehler $d_p(\alpha)$ von seinem Optimum $\int [y(\mathbf{x}) - \langle y \rangle]^2 p(\mathbf{x}) d\mathbf{x}$ um ε^2 ab.

Das Problem der Konstruktion einer lernenden Maschine ist nun folgendes: Wähle f , Λ und α so, dass der Generalisierungsfehler $d_p(\alpha)$ der lernenden Maschine möglichst klein ist.

Sei d_p^* der kleinstmögliche Generalisierungsfehler.

$$d_p^* := \inf_{\alpha \in \Lambda} d_p(\alpha)$$

Betrachte nun die Wahrscheinlichkeit $r(\ell, \varepsilon)$, dass die Abweichung $d_p(\alpha)$ größer als $d_p^* + \varepsilon$ ist im ungünstigsten Fall.

$$r(\ell, \varepsilon) := \sup_{p \in \mathcal{P}} P\{((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell)) | d_p(A((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell))) > d_p^* + \varepsilon\}$$

Diese gegenüber (8.3) modifizierte Definition von $r(\ell, \varepsilon)$ führt auf eine entsprechend modifizierte Version von „wahrscheinlich ungefähr korrekt“.

Definition 3 (wahrscheinlich ungefähr korrekt)

Der Algorithmus A heißt „wahrscheinlich ungefähr korrekt (PAC=*probably approximately correct*) mit Genauigkeit ε “, falls $\lim_{\ell \rightarrow \infty} r(\ell, \varepsilon) = 0$. Der Algorithmus heißt „wahrscheinlich ungefähr korrekt“, falls $\lim_{\ell \rightarrow \infty} r(\ell, \varepsilon) = 0$ für jedes $\varepsilon > 0$.

Ein Problem heißt „wahrscheinlich ungefähr korrekt lernbar“, falls ein wahrscheinlich ungefähr korrekter Algorithmus für das Problem existiert.

8.4 Abschätzung des Generalisierungsfehlers

Bei der Konstruktion einer lernenden Maschine ist man bestrebt, den Generalisierungsfehler aus Abschnitt 8.3.2 möglichst gering zu halten. Oft, z. B. bei Backpropagation-Netzwerken (Rojas 1996, Zell 2000), wird jedoch nur der *Trainingsfehler* minimiert. Der Trainingsfehler ist der Fehler, den die lernende Maschine auf den Trainingsdaten macht.

Definition 4 (Trainingsfehler)

$$d_{\text{emp}} := \frac{1}{\ell} \sum_{i=1}^{\ell} L(y_i, f(\mathbf{x}_i, \alpha))$$

heißt „Trainingsfehler“ oder „empirisches Risiko“.

Falls der Trainingsfehler schon groß ist, kann man nicht erwarten, dass der Generalisierungsfehler viel kleiner ist. Vielmehr wird der Generalisierungsfehler i. d. R. größer als der Trainingsfehler sein. Man hofft allerdings oft, dass *falls* der Trainingsfehler gering ist, der Generalisierungsfehler ebenfalls gering ist. Dies ist jedoch nicht unbedingt der Fall. Es gibt einfache zweischichtige neuronale Netze mit sigmoiden Aktivierungsfunktionen, welche es erlauben, die synaptischen Gewichte so einzustellen, dass der Trainingsfehler auch bei beliebig komplizierten Trainingsdaten null ist (Vidyasagar 1997). Der Generalisierungsfehler ist dann beliebig groß. Die Funktionen f bzw. die Parametermenge Λ sind also geeignet zu beschränken, so dass der Generalisierungsfehler klein wird. „Geeignet beschränken“ heißt, die sog. VC-Dimension (Vapnik 1998) der lernenden Maschine gering halten.

Definition 5 (VC-Dimension)

Die „VC-Dimension“ (Vapnik-Chervonenkis-Dimension) einer Funktionenmenge $f(\mathbf{x}, \alpha), \alpha \in \Lambda$, ist das maximale h , so dass es $(\mathbf{x}_1, \dots, \mathbf{x}_h)$ gibt, die durch die Funktionenmenge beliebig getrennt werden können.

Das Tupel $(\mathbf{x}_1, \dots, \mathbf{x}_h)$ kann beliebig getrennt werden, wenn zu jedem $(y_1, \dots, y_h) \in \{0, 1\}^h$ ein α existiert mit $f(\mathbf{x}_i, \alpha) = y_i$ für alle $i = 1, \dots, h$.

Falls die VC-Dimension einer lernenden Maschine endlich ist, kann man den Generalisierungsfehler mit Hilfe des folgenden Satzes (Vidyasagar 1997) abschätzen.

Satz 6 (Abschätzung des Generalisierungsfehlers)

Für eine lernende Maschine, die eine Funktionenmenge mit VC-Dimension h realisieren kann, gilt für den Generalisierungsfehler $d_p(\alpha)$ für alle $\alpha \in \Lambda$ mit einer Wahrscheinlichkeit von mindestens $1 - \eta$

$$d_p(\alpha) \leq d_{\text{emp}}(\alpha) + \frac{\mathcal{E}}{2} \left(1 + \sqrt{1 + \frac{4 d_{\text{emp}}(\alpha)}{\mathcal{E}}} \right),$$

wobei

$$\mathcal{E} := 4 \frac{h \cdot [\ln(2\ell/h) + 1] - \ln(\eta/4)}{\ell}.$$

Neuronalen Netzen wird nachgesagt, sie seien lernfähig in dem Sinne, dass sie angeblich gut generalisieren können. Dazu muss man allerdings ihre VC-Dimension so groß einstellen, dass sie das Problem mit geringem Trainingsfehler lernen können, aber auch so klein, dass sie wirklich generalisieren. Fragen, die sich sofort stellen, sind: Lernen Menschen oder Tiere statistisch? Wenn ja, wie wird deren neuronales Netz dazu verwendet? Das Problem besteht darüber hinaus nicht nur darin, die VC-Dimension des Netzes geeignet anzupassen, sondern das Netz muss auch effizient gelernt werden können. Ein effizientes Verfahren unter Verwendung einer lernenden Maschine mit einstellbarer VC-Dimension wird in den folgenden Abschnitten vorgestellt.

8.5 Das Perzeptron

Das Perzeptron wurde von Rosenblatt 1960 vorgeschlagen (Rojas 1996). Ein Neuron bekommt Eingaben in Form von Eingabevektoren $\mathbf{x} \in \mathbb{R}^k$. Die Ausgabe $y \in \{-1, 1\}$ dieses Neurons ist

$$y(\mathbf{x}) = \text{sgn}(\boldsymbol{\omega} \cdot \mathbf{x} - b).$$

Hierbei heißt der Vektor $\boldsymbol{\omega} \in \mathbb{R}^k$ „Gewichtsvektor“ des Neurons und b heißt „Schwelle“ des Neurons.

Das Neuron unterteilt den Eingaberaum \mathbb{R}^k somit durch eine Ebene mit Normalenvektor $\boldsymbol{\omega}$ in einen Bereich mit $y = 1$ und einen Bereich mit $y = -1$. Das Perzeptron kann also einfache Klassifikationsaufgaben lösen, nämlich solche, bei denen die Daten linear separierbar sind. Was macht man jedoch, wenn die Daten nicht linear separierbar sind? Rosenblatt hat hierfür ein neuronales Netz mit einer Zwischenschicht vorgeschlagen (Vapnik 1998): Die Neuronen in einer Zwischenschicht bekommen alle die selbe Eingabe \mathbf{x} . Die Ausgabe dieser Neuronen sei $\mathbf{z} \in \{-1, 1\}^N$, wobei N die Zahl der Neuronen in der

Zwischenschicht ist. Der Vektor \mathbf{z} dient nun als Eingabe des letzten Neurons. Das letzte Neuron soll das Klassifikationsproblem lösen, d. h. sein Gewichtsvektor $\boldsymbol{\omega}$ soll so angepasst werden, dass das Neuron auf Trainingsdaten die „richtige“ Ausgabe erzeugt.

Die Trainingsdaten seien

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell).$$

Hierbei wird die Klassifikation $y_i, i = 1, \dots, \ell$, von einem „Lehrer“ vorgegeben. Das Neuron, das lernen soll, „sieht“ allerdings nur

$$(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_\ell, y_\ell),$$

nämlich seine Eingabe \mathbf{z}_i aus der Zwischenschicht und die vom Lehrer vorgegebene Klassifikation y_i . Rosenblatt schlug folgenden Algorithmus vor, wie das Neuron seine Gewichte anpassen soll:

1. Setze $\boldsymbol{\omega} := \mathbf{0}$ und $i := 1$.
2. Ziehe das nächste Element der Trainingsdaten (\mathbf{z}_i, y_i) . Falls dies vom Neuron korrekt klassifiziert wird, ändere den Gewichtsvektor $\boldsymbol{\omega}$ nicht. Andernfalls setze

$$\boldsymbol{w} := \boldsymbol{\omega} + y_i \mathbf{z}_i.$$

3. Setze $i := i + 1$, falls $i < \ell$, $i := 1$ sonst. Fahre mit 2. fort.

Wann soll das Perzeptron den Lernprozess stoppen? Hierüber gibt folgender Satz (Vapnik 1998) Auskunft.

Satz 7 (Generalisierungsfehler des Perzeptrons)

Falls gilt

1. Es steht eine unendliche Folge von Trainingsdaten (z_i, y_i) zur Verfügung (d. h. $\ell = \infty$).
2. $|z| \leq R$
3. Die Trainingsdaten können mit Rand ϱ separiert werden, d. h. es gibt eine Ebene, welche die Trainingsdaten trennt und der euklidische Abstand von jedem Trainingsvektor x_i zu dieser Ebene ist mindestens ϱ .
4. Das Perzeptron stoppt den Lernprozess, wenn nach der Korrektur Nummer i die nächsten

$$\frac{1 + 2 \ln i - \ln \eta}{-\ln(1 - \varepsilon)}$$

Trainingsdaten richtig erkannt werden.

dann

1. Das Perzeptron stoppt mit

$$i \leq \frac{1 + 4 \ln \frac{R}{\varrho} - \ln \eta R^2}{-\ln(1 - \varepsilon)} \frac{1}{\varrho^2}.$$

2. Der Generalisierungsfehler ist mit Wahrscheinlichkeit größer oder gleich $1 - \eta$ kleiner oder gleich ε .

Das Perzeptron ist nach Satz 7 also jedenfalls dann generalisierungsfähig, wenn die Trainingsvektoren mit Rand $\varrho > 0$ separiert werden können. Ist dies nicht der Fall, dann konvergiert der angegebene Perzeptron-Algorithmus nicht und es gibt auch keine Aussage über den Generalisierungsfehler.

8.6 Optimale Hyperebenen

Das Perzeptron hat VC-Dimension $n + 1$, wobei n gleich der Anzahl der justierbaren Gewichte ist (Vapnik 1998)⁵. In dem in dieser Arbeit behandelten

⁵Betrachten wie eine beliebige Anordnung hintereinander geschalteter Perzeptrene. Sei n die Zahl der justierbaren Parameter im Netz (Schwellen und Gewichte) und bezeichne \lg den Logarithmus zur Basis 2 und e die Basis des natürlichen Logarithmus. Die betrachtete

Problem, Aktionspotentiale von Neuronen vorherzusagen, ist die Dimension des Eingaberaums sehr groß (vgl. Fußnote 15, Abschnitt 1.2). Es ist also wünschenswert, die VC-Dimension der lernenden Maschine beschränken zu können. Dies gelingt mit einer Modifikation des Perzeptrons, nämlich der mit Rand separierenden Hyperebene. Wir bezeichnen mit „ \cdot “ das Skalarprodukt.

Definition 8 (Hyperebene mit Rand M)

Sei $\mathbf{x} \in \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq R\}$ und $y \in \{-1, 1\}$. Die Funktionenmenge

$$f(\mathbf{x}, \boldsymbol{\omega}, b) = \begin{cases} -1 & \text{falls } \boldsymbol{\omega} \cdot \mathbf{x} + b \leq -M \\ 1 & \text{falls } \boldsymbol{\omega} \cdot \mathbf{x} + b \geq M \\ -y & \text{sonst} \end{cases},$$

wobei $(\boldsymbol{\omega}, b) \in \Lambda$, $\Lambda := \{(\boldsymbol{\omega}, b) \mid \boldsymbol{\omega} \in \mathbb{R}^n, \|\boldsymbol{\omega}\| = 1, b \in \mathbb{R}\}$ heißt eine „Hyperebene mit Rand M “.

Die VC-Dimension einer Hyperebene mit Rand M kann viel geringer sein als diejenige einer Hyperebene mit Rand 0, d.i. eines gewöhnlichen Perzeptrons (Vapnik 1998). Bezeichnen wir mit $\lfloor r \rfloor$ die größte natürliche Zahl kleiner oder gleich r (Gauß-Klammer), dann gilt folgender Satz:

Satz 9 (VC-Dimension einer Hyperebene mit Rand M)

Eine Hyperebene mit Rand M hat eine VC-Dimension h , für die gilt

$$h \leq \min(\lfloor R^2/M^2 \rfloor, n) + 1.$$

Um den Generalisierungsfehler aus Theorem 6 einer Hyperebene möglichst zu beschränken, muss der Trainingsfehler so klein wie möglich gemacht werden. Wie definieren daher den Begriff der „optimalen Hyperebene“.

Definition 10 (Optimale Hyperebene)

Eine „optimale Hyperebene“ erhält man durch folgende Minimierungsaufgabe:

Minimiere den Trainingsfehler $d_{\text{emp}}(\boldsymbol{\omega}, b, M)$ einer Hyperebene mit Rand M unter der Nebenbedingung $M \geq M_{\text{min}}$.

Ein Beispiel für eine solche optimale Hyperebene gibt Abbildung 8.1.

Im Zusammenhang mit optimalen Hyperebenen spielen einige der Punkte \mathbf{x}_i eine ausgezeichnete Rolle. Sie werden als „Supportvektoren“ bezeichnet.

Anordnung von Perzeptronen hat dann eine VC-Dimension kleiner oder gleich $n \ln(en)$ (Vidyasagar 1997).

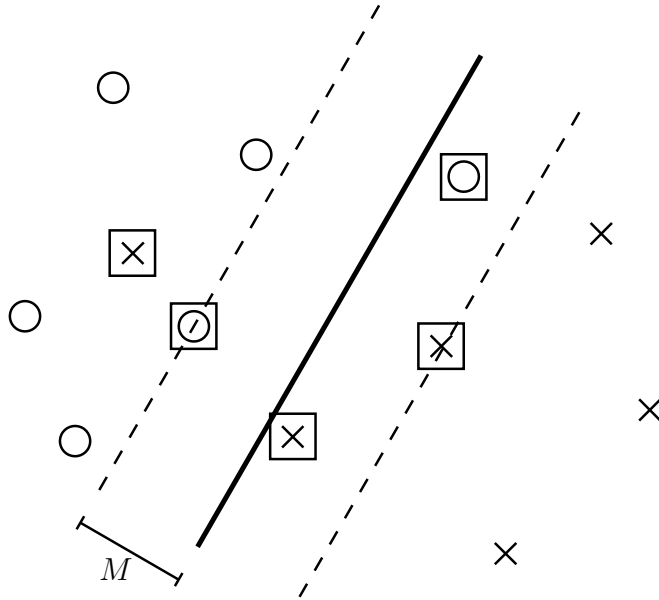


Abbildung 8.1: Kreise und Kreuze werden durch eine optimale Hyperebene mit Rand M getrennt. Supportvektoren sind mit Kästen versehen.

Definition 11 (Support-Vektor)

Die Vektoren \mathbf{x}_i , die durch die separierende optimale Hyperebene mit Rand M falsch oder gerade eben richtig ($|\boldsymbol{\omega} \cdot \mathbf{x}_i + b| = M$) klassifiziert werden, heißen Support-Vektoren.

Beachte hierbei, dass nach Definition 8 der Vektor \mathbf{x}_i nur dann richtig klassifiziert wird, wenn \mathbf{x}_i nicht nur jenseits der Hyperebene liegt, sondern auch noch jenseits des Randes.

Falls die optimale Hyperebene separiert, falls also der Trainingsfehler null ist, dann lässt sich der Generalisierungsfehler durch die Anzahl der Supportvektoren abschätzen (Vapnik 1998). Eine vergleichbare „wunderbare Verbesserung“ des Generalisierungsfehlers ist beim Perzeptron nicht möglich.

Satz 12 (Generalisierungsfehler der optimalen separierenden Hyperebene)

Für den Generalisierungsfehler $d_p(\alpha)$ einer optimalen separierenden Hyperebene gilt

$$\langle d_p(\alpha) \rangle \leq \frac{\langle \text{Zahl der Supportvektoren} \rangle}{\ell - 1}$$

Hierbei bezeichnen die spitzen Klammern $\langle \rangle$ den Erwartungswert über alle Trainingsmengen der Länge ℓ mit der Verteilungsfunktion $p(\mathbf{x}, y)$.

Damit unsere Maschine gut generalisiert hoffen wir, dass die Zahl der Supportvektoren sehr viel kleiner als die Zahl der Trainingsbeispiele ℓ ist.

8.7 Support-Vektor Maschinen

Eine Einführung in Support-Vektor Maschinen findet man in Vapnik (1998) und Kecman (2001).

Optimale Hyperebenen sind nur *lineare* lernende Maschinen in dem Sinn, dass die Punkte nur durch eine lineare Funktion getrennt werden können. Um dieses Problem zu umgehen, wird folgender Trick angewandt: Bilde den Raum der Eingabevektoren \mathbf{x} auf einen höherdimensionalen Merkmalsraum

$$\Psi(\mathbf{x}) := (\psi_1(\mathbf{x}), \dots, \psi_N(\mathbf{x}))$$

ab. Konstruiere in diesem Raum eine optimale Hyperebene. Sie hat die Form

$$f(\Psi(\mathbf{x})) = \text{sgn}(\boldsymbol{\omega} \cdot \Psi(\mathbf{x}) - b) .$$

Die optimale Hyperebene in diesem höherdimensionalen Merkmalsraum hat die im folgenden Satz beschriebene spezielle Form.

Satz 13 (Support-Vektor Maschine)

Für optimale separierende Hyperebenen gilt

$$\boldsymbol{\omega} = \sum_{\text{Support-Vektoren}} \alpha_i \Psi(\mathbf{x}_i),$$

somit

$$f(\Psi(\mathbf{x})) = \text{sgn} \left(\sum_{\text{Support-Vektoren}} \alpha_i \Psi(\mathbf{x}_i) \cdot \Psi(\mathbf{x}) - b \right).$$

Eine lernende Maschine mit dieser Klassifikationsfunktion heißt „Support-Vektor-Maschine“.

Die VC-Dimension h der optimalen Hyperebene im Merkmalsraum Ψ wird nach Satz 9 abgeschätzt durch $h \approx R^2/M^2$. Dabei ist der Rand $M = \boldsymbol{\omega} \cdot \Psi(\mathbf{x}) - b$ (\mathbf{x} ist ein Support-Vektor, der richtig als $y = 1$ klassifiziert wird) und der Radius

$$R^2 = \min_{\mathbf{a}} \max_{i=1, \dots, \ell} (\Psi(\mathbf{x}_i) - \Psi(\mathbf{a}))^2.$$

Zur Berechnung von $f(\Psi(\mathbf{x}))$, der α_i , des Rands M und des Radius R genügt es, die Skalarprodukte $K(\mathbf{x}_i, \mathbf{x}_j) := \Psi(\mathbf{x}_i) \cdot \Psi(\mathbf{x}_j)$, $i, j = 1, \dots, \ell$, zu kennen. Also lässt sich $K(\mathbf{x}_i, \mathbf{x}_j)$ beliebig festlegen, sofern ein Merkmalsraum existiert, in dem sich K als Skalarprodukt darstellen lässt. Diese Bedingung wird im folgenden Satz formuliert.

Satz 14 (Mercers Bedingung)

$K(\mathbf{u}, \mathbf{v})$ sei symmetrisch. Dann existieren Funktionen ψ_k , so dass

$$K(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{\infty} \psi_k(\mathbf{u}) \psi_k(\mathbf{v})$$

genau dann wenn

$$\iint K(\mathbf{u}, \mathbf{v}) g(\mathbf{u}) g(\mathbf{v}) d\mathbf{u} d\mathbf{v} > 0$$

für alle $g \neq 0$ mit $\int g^2(\mathbf{u}) d\mathbf{u} < \infty$.

Beispiele für Support-Vektor-Maschinen sind:

- Polynomiale lernende Maschine mit Grad d :

$$K(\mathbf{u}, \mathbf{v}) := (\mathbf{u} \cdot \mathbf{v} + 1)^d$$

- Zweischichtiges neuronales Netz:

$$K(\mathbf{u}, \mathbf{v}) := S(\sigma \cdot \mathbf{u} \cdot \mathbf{v} + c)$$

Hierbei ist S eine sigmoide Funktion mit geeigneten Konstanten σ und c .

Die Klassifikationsfunktion lautet

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{\text{Support-Vektoren}} \alpha_i S(\sigma \cdot \mathbf{x}_i \cdot \mathbf{x} + c) - b \right).$$

Die VC-Dimension von Support-Vektor-Maschinen ist nicht notwendigerweise klein, ja sie kann sogar, in Abhängigkeit des Kerns, unendlich sein. Falls der Raum der Eingabe-Vektoren \mathbb{R}^n ist, dann hat eine polynomiale Support-Vektor-Maschine mit Grad d die VC-Dimension (Burgess 1998)

$$h = ((n + d - 1)d), \quad (8.4)$$

was bei großem n mit wachsendem d sehr groß werden kann. Ist der zulässige Bereich der Eingaben-Vektoren beschränkt, dann schränkt sich die VC-Dimension ein, in Abhängigkeit des Parameters C , siehe oben. Die sich daraus ergebenden Abschätzungen für den Generalisierungsfehler sind dann allerdings i. d. R. um Größenordnungen größer als der tatsächlich auftretende Generalisierungsfehler. Der Parameter C , bei dem die Abschätzung für den Generalisierungsfehler minimal ist, scheint allerdings zumindest prädiktiv in dem Sinne prädiktiv zu sein, dass hier auch der wahre Generalisierungsfehler minimal ist (Burgess 1998).

8.8 Numerische Behandlung

Eine optimale Hyperebene mit Rand $M = 0$ nach Definition 10 zu finden ist ein NP-hartes⁶ Problem bezüglich der Dimension des Eingaberaums.⁷ Man ist also aus Effizienzgründen gezwungen, von der optimalen Hyperebene abzuweichen und einen etwas größeren Trainingsfehler zu akzeptieren. Man minimiert also *nicht* den Trainingsfehler d_{emp} aus Definition 4, wobei die Klassifikationsfunktion f eine Hyperebene mit Rand aus Definition 8 und $L(y, f(\mathbf{x}, \boldsymbol{\omega}, b)) = 1$ falls $y \neq f(\mathbf{x}, \boldsymbol{\omega}, b)$ und 0 sonst ist. Sondern man

⁶ Jedes Problem, das sich von Hand, mit Hilfe eines Rechners oder sonstiger informationsverarbeitender Anlagen berechnen lässt, lässt sich auch mit einer Turing-Maschine berechnen. Dies ist der Inhalt der Church-Turing-Hypothese. Eine Turingmaschine ist ein mathematisch exakt definierter, einfacher Rechenautomat mit unendlich großem Speicher. Turing-Maschinen führen diskrete Rechenschritte aus. Ein Problem hat die „Zeitkomplexität“ $O(f(n))$, falls für die Anzahl $N(n)$ der Rechenschritte, die zur Lösung eines Problems der „Länge“ n im ungünstigsten Fall notwendig sind, gilt: Es gibt von n unabhängige Konstanten C_1 und C_2 mit $C_1 f(n) \leq N(n) \leq C_2 f(n)$ für alle n . Ein Problem hat „polynomiale Zeitkomplexität bezüglich n “, falls es die Zeitkomplexität $O(1 + n + n^2 + \dots + n^p)$ hat. Man bezeichnet Probleme mit höherer als polynomialer Zeitkomplexität als „nicht effizient lösbar“, da diese zur Lösung eine mit der Länge des Problems exponentiell anwachsende Zahl von Rechenschritten benötigen (Hopcroft und Ullman 1992). Die Menge von Problemen mit polynomialer Zeitkomplexität bezeichnet man als P . Eine „nicht deterministische“ Turingmaschine kann sich in zwei Turingmaschinen aufspalten, die dann parallel arbeiten. Diese können sich wieder aufspalten u.s.w. Die Menge von Problemen, die durch nicht deterministische Turingmaschinen mit polynomialer Zeitkomplexität lösbar sind, bezeichnet man als NP. Probleme, auf die sich *jedes* Problem in NP in polynomialer Zeit zurückführen lassen, heißen „NP-hart“. Ein Problem lässt sich „in polynomialer Zeit“ auf ein anderes „zurückführen“, genau dann wenn es eine Turingmaschine mit polynomialer Zeitkomplexität gibt, die die Eingabedaten des Problems in Eingabedaten für das andere Problem verwandelt und es eine Turingmaschine mit polynomialer Zeitkomplexität gibt, die die Lösung des anderen Problems in eine Lösung des Problems verwandelt. NP-harte Probleme, die selbst in NP sind, bezeichnet man als „NP-vollständig“. Es ist nicht geklärt, ob $P = \text{NP}$, man geht jedoch davon aus, dass $P \subset \text{NP}$. Stimmt dies, dann benötigen NP-harte Probleme Rechenzeit, die exponentiell mit der Länge des Problems anwächst (Hopcroft und Ullman 1992). Auf die Lösung des $P = \text{NP}$ -Problems ist ein Preis von 1 Mio. \$ ausgesetzt (Stewart 2002). Weitere interessante ungelöste mathematische Probleme hat Smale (1998) zusammengestellt.

⁷ Folgendes Problem ist NP-hart in k : Gegeben ist eine Punktmenge aus $\{0, 1\}^k$, eine Klassifikation, die jedem dieser Punkte einen Wert aus $\{0, 1\}$ zuordnet und eine Zahl $d \geq 0$. Bestimme, ob es ein Perzeptron gibt, das höchstens d Punkte falsch klassifiziert (Vidyasagar 1997). Falls man ein Perzeptron gefunden hat, das d Punkte aus \mathbb{R}^k falsch klassifiziert, ist es also ein NP-hartes Problem zu zeigen, dass dieses Perzeptron die optimale Lösung darstellt, es also kein Perzeptron gibt, das höchstens $d - 1$ Punkte falsch klassifiziert. Das Finden eines optimalen Perzeptrons stellt also erst recht ein NP-hartes Problem dar.

minimiert die Näherung

$$\hat{d}_{\text{emp}} := \frac{1}{\ell} \sum_{i=1}^{\ell} L(y_i, f(\mathbf{x}_i, \boldsymbol{\omega}, b))$$

für den Trainingsfehler d_{emp} , wobei

$$f(\mathbf{x}, \boldsymbol{\omega}, b) = K(\boldsymbol{\omega}, \mathbf{x}) - b$$

und

$$L(y, f) := \begin{cases} |1 - yf| & \text{falls } yf < 1 \\ 0 & \text{sonst.} \end{cases}$$

Mit $|\cdot|^\sigma$ anstatt $|\cdot|$ und hinreichend kleinem $\sigma > 0$ wäre diese Näherung exakt der Trainingsfehler. Der Exponent $\sigma = 1$ ist der kleinste Parameter, der numerisch noch einfach zu behandeln ist (Vapnik 1998).⁸

Die Minimierung von \hat{d}_{emp} ist allerdings unter der Nebenbedingung

$$\boldsymbol{\omega}^2 \leq \frac{1}{M^2}$$

durchzuführen, so dass die durch $f(\mathbf{x}, \boldsymbol{\omega}, b)$ beschriebene Hyperebene einen Rand mit Breite M hat.

Das Minimum $\boldsymbol{\omega}_0$ von \hat{d}_{emp} hat die Form

$$\boldsymbol{\omega}_0 = \sum_{i=1}^{\ell} \alpha_i y_i \mathbf{x}_i.$$

Die Koeffizienten α_i , $i = 1, \dots, \ell$, lassen sich durch die Maximierung von

$$W(\boldsymbol{\alpha}) := \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$$

unter den Nebenbedingungen

$$0 \leq \alpha_i \leq C \quad \text{für } i = 1, \dots, \ell$$

und

$$\sum_{i=1}^{\ell} \alpha_i y_i = 0$$

⁸Für $\sigma \rightarrow 0$ ergibt sich das ein Problem ähnlich dem Auffinden des optimalen Perzeptrons, das NP-hart ist, vgl. Fußnote 7.

berechnen. Je größer der „Kontrollparameter“ C , desto kleiner der Rand M . Leider lässt sich nicht zu jedem M auf effiziente Weise ein C berechnen, so dass der resultierende Rand M ist. Dafür lässt sich die Maximierung von $W(\alpha)$ effizient durchführen, siehe Abschnitt 8.10.

Es stellt sich heraus, dass sehr viele $\alpha_i = 0$ sind. Darüber hinaus sind es im separierenden Fall gerade die Support-Vektoren aus Definition 11, für die $\alpha_i \neq 0$. Wir erweitern die Definition der Support-Vektoren auf den nicht separierenden Fall.

Definition 15 (Support-Vektoren)

Alle Vektoren \mathbf{x}_i mit $\alpha_i \neq 0$ heißen Support-Vektoren.

Sowohl im separierenden als auch im nicht separierenden Fall gilt der folgende Satz.

Satz 16 (Lage von Support-Vektoren)

Support-Vektoren liegen auf dem Rand ($0 < \alpha_i < C$), oder werden falsch klassifiziert ($\alpha_i = C$), liegen also jenseits des der richtigen Seite zugewandten Randes, somit nicht unbedingt auf der falschen Seite der Hyperebene.

8.9 Beispiele

Einige einfache Beispiele sollen die Fähigkeiten von Support-Vektor-Maschinen illustrieren. In den Abbildungen 8.2 bis 8.6 ist die Ausgabe eines sehr einfachen „Neurons“ dargestellt. Das Neuron hat zwei Eingaben x_1 (Rechtswertachse) und x_2 (Hochwertachse). Dessen Membranpotential ist

$$h(x_1, x_2) = \text{rect}(x_1) + \text{rect}(x_2), \quad (8.5)$$

wobei „rect“ die „rektifizierende“ Funktion mit

$$\text{rect}(x) := \begin{cases} x & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

ist. Das Neuron erzeugt die Ausgabe $y = 1$ (Aktionspotential) falls $h(x_1, x_2) > 1$ und die Ausgabe $y = -1$ (kein Aktionspotential) sonst. Man kann sich vorstellen, dass das Neuron gleich starke exzitatorische Eingaben von zwei Nervenfasern bekommt. Die Nervenfasern zeigen Feuerraten $\text{rect}(x_1)$

bzw. $\text{rect}(x_2)$. Die Feuerraten sind bei positivem Stimulus, also $x_1, x_2 > 0$, proportional zum Stimulus und bei negativem Stimulus null.⁹

In Abbildung 8.2 sieht man, dass eine polynomiale Support-Vektor-Maschine dritten Grades mit kleinem Rand die nicht lineare Funktion $h(x_1, x_2)$ aus (8.5) recht genau nachbildet. Die durchgezogene Linie ist die trennende Hyperebene der Support-Vektor-Maschine, in diesem Fall also die Nullstellen eines Polynoms dritten Grades in x_1 und x_2 , und stimmt weitgehend mit der gestrichelten Linie überein, die die exakte Trennlinie zwischen den Bereichen mit und ohne Aktionspotentiale darstellt. In Abbildung 8.3 ist der Rand breiter gewählt und die Übereinstimmung der Support-Vektor-Maschine mit der gestrichelten Linie weniger gut. Der Generalisierungsfehler der Support-Vektor-Maschine wäre proportional zu der zwischen gestrichelter und durchgezogener Linie eingeschlossenen Fläche auf dem Quadrat $[-0.5, 1.5]^2$, da die Eingabevektoren (x_1, x_2) gleichverteilt aus $[-0.5, 1.5]^2$ gezogen wurden.

Abbildungen 8.4 bis 8.6 zeigen Support-Vektor-Maschinen fünften Grades. Zusätzlich ist ein „Fehler“ eingebaut: Das Neuron hat in Abweichung von (8.5) ein Aktionspotential bei der Eingabe $(0, 0.25)$ erzeugt. Die Support-Vektor-Maschine aus Abbildung 8.4 kringelt das fehlerhafte Aktionspotential ein, wodurch zwar der Trainingsfehler kleiner wird (der Eingabevektor $(0, 0.25)$ wird richtig klassifiziert), der Generalisierungsfehler wird jedoch größer, wie man an der Größe des von der gestrichelten und der durchgezogenen Kurve eingeschlossenen Bereichs plus dem eingekreisten Bereich um $(0, 0.25)$ herum sieht.

Wird der Rand größer, wie in Abbildung 8.5, nimmt der Trainingsfehler zu, da nun die Eingabe $(0, 0.25)$ von der Support-Vektor-Maschine als „kein Aktionspotential“ klassifiziert wird, der Generalisierungsfehler nimmt jedoch ab. In Abbildung 8.6 ist der Rand der Support-Vektor-Maschine noch größer, was wieder zu einer Zunahme sowohl des Trainingsfehlers als auch des Generalisierungsfehlers führen kann.

Abbildungen 8.7 bis 8.9 zeigen Neuronen mit Membranpotential

$$h(x_1, \dots, x_k) = \sum_{i=1}^k r_i x_i.$$

Sie erzeugen ein Aktionspotential (Ausgabe $y = 1$), falls $h(x_1, \dots, x_k) > \vartheta$, andernfalls erzeugen sie kein Aktionspotential (Ausgabe $y = -1$). Die Eingabe x_1, \dots, x_k wurde gleichverteilt aus $[-1, 1]^k$ gezogen. Die Schwelle wurde

⁹Dies ist z. B. der Fall bei inneren Haarzellen, deren Nervenfasern nur Aktionspotentiale erzeugen, wenn die Stereocilien in positive Richtung ausgelenkt werden (Yost 1994) oder bei den Nerven von den Seitenlinienorganen des Krallenfrosches *Xenopus*, die nur erhöhte Feuerrate zeigen, wenn das entsprechende Sinnesorgan durch die Partikelgeschwindigkeit des Wassers in die „richtige“ Richtung ausgelenkt wird (Russell 1976).

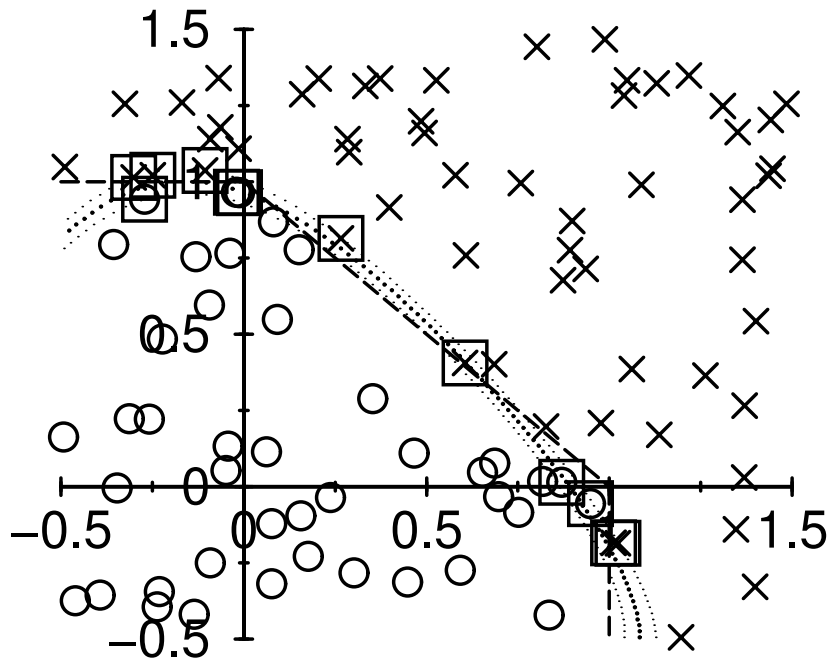


Abbildung 8.2:

Aktionspotentiale (Kreuze) und keine Aktionspotentiale (Kreise) des Neurons mit dem Membranpotential $h(x_1, x_2)$ aus (8.5). Die Stimuli x_1 (Rechtswertachse) und x_2 (Hochwertachse) wurden gleichverteilt aus $[-0.5, 1.5]$ gezogen. Die gestrichelte Linie trennt den Bereich, in dem das Neuron Aktionspotentiale erzeugt (rechts oben) von dem Bereich, indem es keine Aktionspotentiale erzeugt (links unten). Die durchgezogene Linie ist die trennende Hyperebene einer Support-Vektor-Maschine mit $K(\mathbf{u}, \mathbf{v}) = (\mathbf{u} \cdot \mathbf{v} + 1)^3$ und schmalen Rand (Kontrollparameter C groß). Die trennende Hyperebene (durchgezogene Linie) der Support-Vektor-Maschine stimmt gut mit der exakten Trennlinie (gestrichelte Linie) der Bereiche mit und ohne Aktionspotentiale überein. Die Support-Vektoren (umkastelt) legen auf dem Rand oder innerhalb des Randes (gepunktet) und machen nur einen kleinen Bruchteil der Eingabe-Vektoren (Kreuze und Kreise) aus.

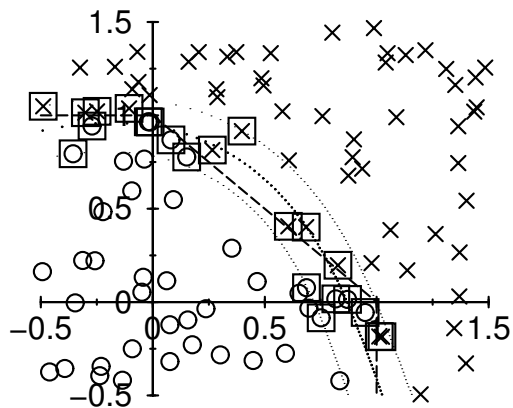


Abbildung 8.3:
 $K(\mathbf{u}, \mathbf{v}) = (\mathbf{u} \cdot \mathbf{v} + 1)^3$,
 breiter Rand.

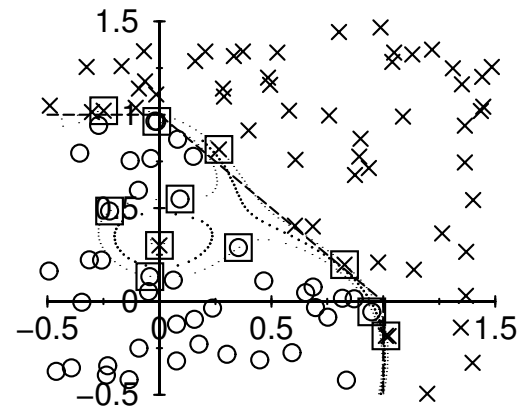


Abbildung 8.4:
 $K(\mathbf{u}, \mathbf{v}) = (\mathbf{u} \cdot \mathbf{v} + 1)^5$,
 unbeschränkter Rand.

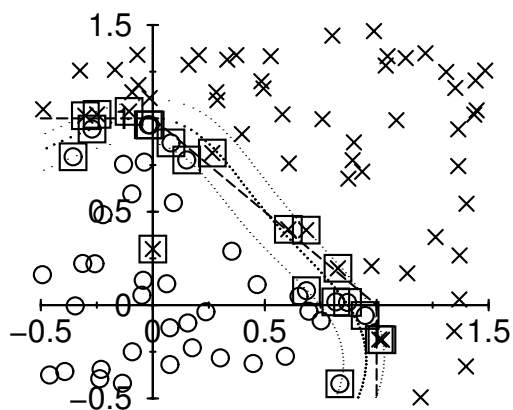


Abbildung 8.5:
 $K(\mathbf{u}, \mathbf{v}) = (\mathbf{u} \cdot \mathbf{v} + 1)^5$,
 breiter Rand.

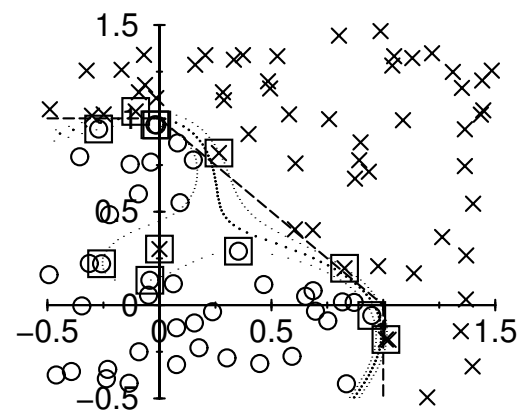


Abbildung 8.6:
 $K(\mathbf{u}, \mathbf{v}) = (\mathbf{u} \cdot \mathbf{v} + 1)^5$,
 schmaler Rand.

so angepasst, dass die Neurone in der Hälfte aller Fälle ein Aktionspotential erzeugt haben.

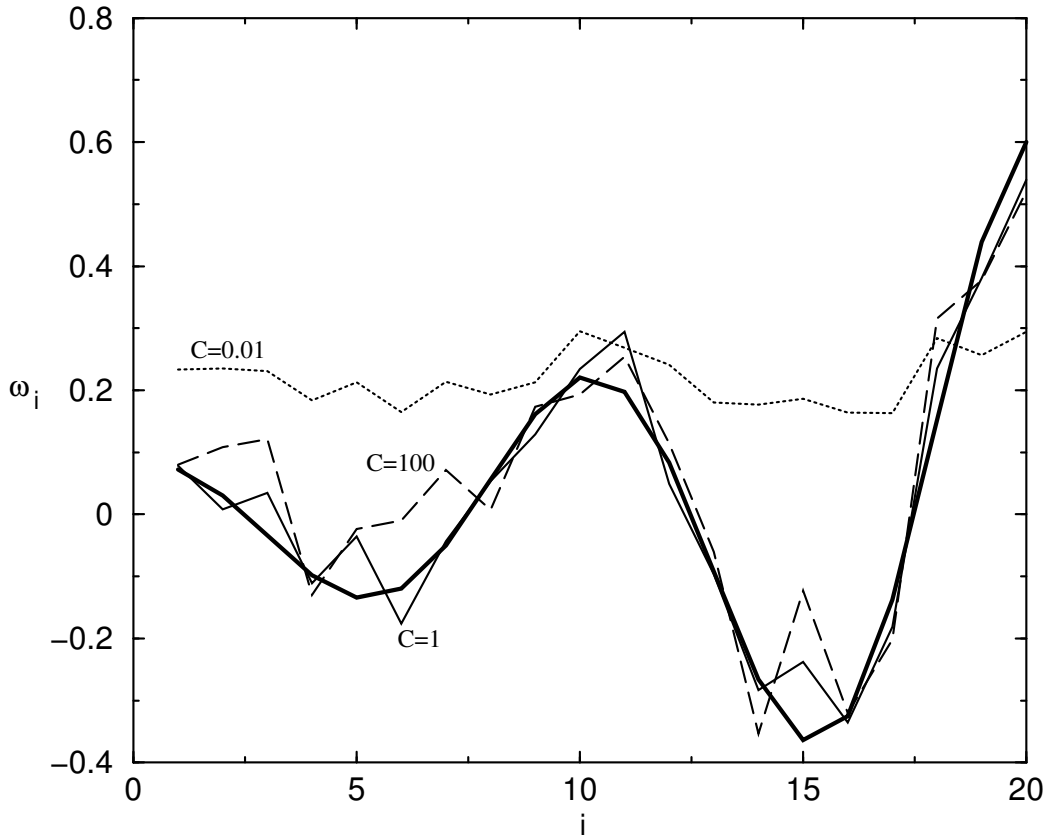


Abbildung 8.7: Gewichte ω_i einer linearen Support-Vektor-Maschine mit $K(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} + 1$ bei verschiedenen Kontrollparametern C und Anzahl der Trainingseingaben $\ell = 100$. Die Tabelle zeigt den numerisch geschätzten Generalisierungsfehler in Abhängigkeit des Kontrollparameters C . Mit größer werdendem Kontrollparameter C nimmt der Generalisierungsfehler zunächst ab, dann allerdings wieder zu. Entsprechend wird die Übereinstimmung der Gewichte ω_i der trennenden Hyperebene der Support-Vektor-Maschine mit den Gewichten r_i (dicke durchgezogene Linie) des Neurons zunächst besser, dann wieder schlechter.

C	0.01	0.1	1	10	100
d_{est}	0.33	0.12	0.03	0.09	0.11

In Abbildung 8.7 wurden $\ell = 100$ Ein-Ausgabe-Paare (\mathbf{x}_i, y_i) , $i = 1, \dots, \ell$, durch eine lineare Support-Vektor-Maschine ausgewertet. Die Wahl des Kontrollparameters $C = 1$ erzeugt den kleinsten Generalisierungsfehler d_{est} . Mit

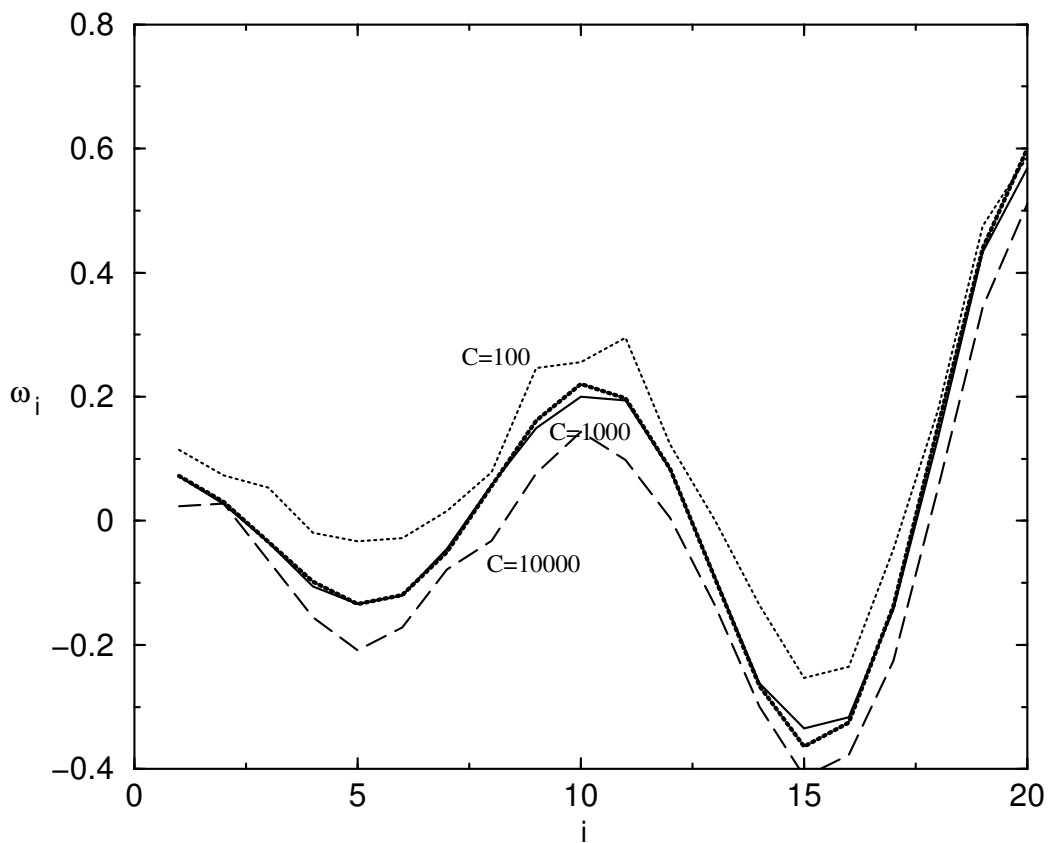


Abbildung 8.8: Wie Abbildung 8.7 mit $\ell = 1000$. Der Generalisierungsfehler ist geringer als in Abbildung 8.7 mit $\ell = 100$ und die Übereinstimmung der Gewichte ω_i der trennenden Hyperebene der Support-Vektor-Maschine mit den Gewichten r_i (dicke durchgezogene Linie) des Neurons besser.

C	100	1000	10000
d_{est}	0.04	0.02	0.05

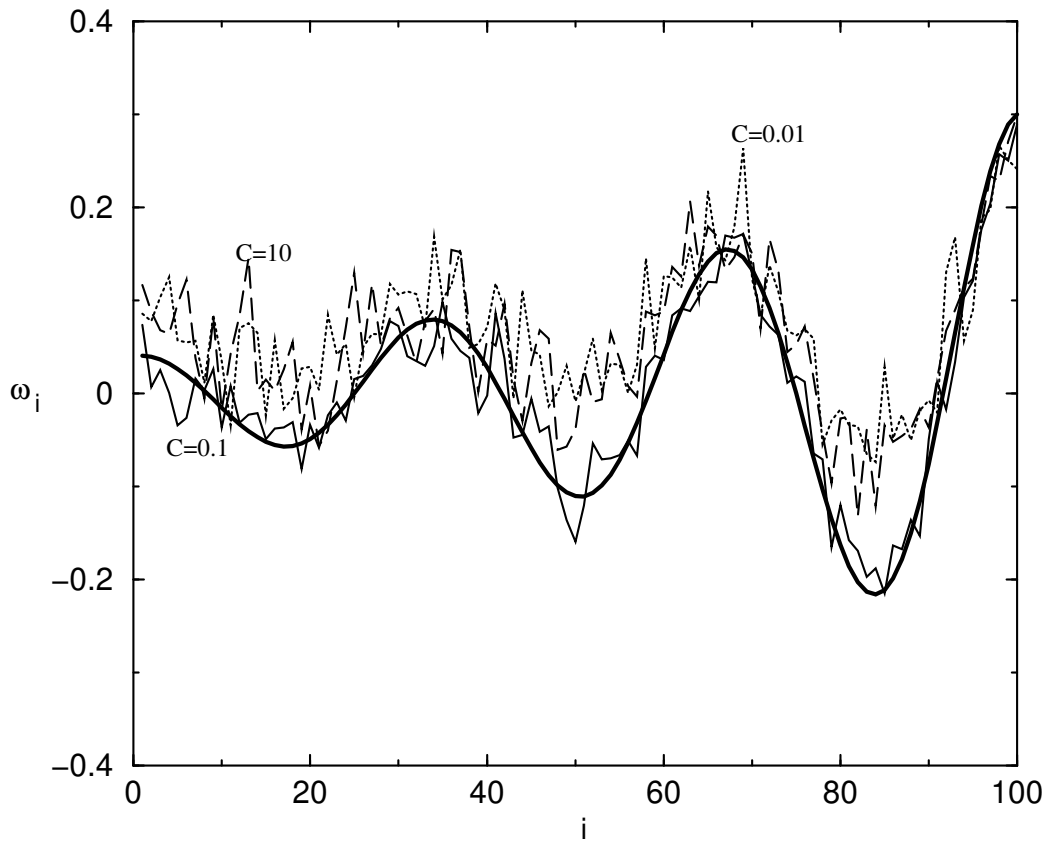


Abbildung 8.9: Wie Abbildung 8.8, nur ist die Dimension des Eingaberaums $k = 100$ anstatt $k = 20$. Die Generalisierungsfehler werden größer und die Übereinstimmung der Gewichte ω_i der Support-Vektor-Maschine mit den Gewichten r_i des Neurons (dicke durchgezogene Linie) wird generell schlechter.

C	0.01	0.1	1	10
d_{est}	0.17	0.09	0.10	0.15

dieser Wahl des Kontrollparameters stimmen auch die Gewichte ω_i der Hyperebene der Support-Vektor-Maschine mit den Gewichten r_i des Neurons gut überein.

Abbildung 8.7 entstand durch Berechnung einer Support-Vektor-Maschine mit einer Trainingsmenge aus $\ell = 1000$ Ein-Ausgabe-Paaren. Der Generalisierungsfehler (0.02) bei $C = 1000$ ist geringer als der Generalisierungsfehler (0.03) für $\ell = 100$ und $C = 1$ in Abbildung 8.7.

In Abbildung 8.9 ist die Dimension des Eingaberaums $k = 100$ anstatt $k = 20$ in den Abbildungen 8.7 und 8.7. Der Generalisierungsfehler wird mit zunehmender Dimension des Eingaberaums größer (0.09 für $C = 0.1$).

8.10 Effiziente Algorithmen für Support-Vektor-Maschinen

Nach einer allgemeinen Einführung in Optimierungsprobleme wird in diesem Abschnitt ein effizienter Algorithmus entwickelt, um das bei Support-Vektor-Maschinen auftretende Optimierungsproblem zu lösen.

8.10.1 Optimierungsprobleme

Ein Optimierungsproblem darin, ein globales Minimum¹⁰ einer Funktion $f : X \subset \mathbb{R}^m \rightarrow \mathbb{R}$ zu finden. Oft sind bei der Optimierung auch Nebenbedingungen einzuhalten, z. B. gegeben die Funktionen $c_i : X \rightarrow \mathbb{R}$, $i = 1, \dots, n$:

$$\begin{aligned} &\text{Minimiere } f(\mathbf{x}), \mathbf{x} \in X \\ &\text{unter den Nebenbedingungen } c_i(\mathbf{x}) \leq 0 \text{ für alle } i \in \{1, \dots, n\} \end{aligned} \quad (8.6)$$

Im Allgemeinen gibt es mehrere lokale Minima, von denen wiederum eins oder mehrere globale Minima sein können. Ein Satz, der Eindeutigkeit des Minimums gewährleistet, ist folgender (Schölkopf und Smola 2002):

Satz 17 (Konvexe Minimierung unter Nebenbedingungen)

Seien $f, c_1, \dots, c_n : \mathbb{R}^m \rightarrow \mathbb{R}$ streng konvexe Funktionen auf der konvexen Menge X , dann hat das Optimierungsproblem (8.6) genau eine Lösung.

Welche (notwendigen und hinreichenden) Bedingungen müssen nun erfüllt sein, damit ein Punkt $\hat{\mathbf{x}} \in \mathbb{R}^m$ ein Minimum ist? Auskunft hierüber gibt folgender Satz (Schölkopf und Smola 2002):

¹⁰Falls das Maximum gefunden werden soll, minimiere die Funktion $-f$.

Satz 18 (Kuhn-Tucker Sattelpunktbedingung)

Seien $f, c_1, \dots, c_n : \mathbb{R}^m \rightarrow \mathbb{R}$ beliebige Funktionen und die „Lagrange-Funktion“

$$L(\mathbf{x}, \boldsymbol{\alpha}) := f(\mathbf{x}) + \sum_{i=1}^n \alpha_i c_i(\mathbf{x}).$$

Falls $\mathbf{x} \in \mathbb{R}^m$ und $\hat{\alpha}_i \geq 0$ für alle $i \in \{1, \dots, n\}$, so dass für alle $\mathbf{x} \in \mathbb{R}^m$ und alle $\boldsymbol{\alpha} \in [0, \infty]^n$ gilt

$$L(\hat{\mathbf{x}}, \boldsymbol{\alpha}) \leq L(\hat{\mathbf{x}}, \hat{\boldsymbol{\alpha}}) \leq L(\mathbf{x}, \hat{\boldsymbol{\alpha}}) \quad (\text{Sattelpunkt}),$$

dann ist $\hat{\mathbf{x}}$ eine Lösung des Optimierungsproblems (8.6).

Gleichheitsnebenbedingungen $e_i(\mathbf{x}) = 0$ können behandelt werden, indem man sie in die Ungleichungsnebenbedingung $e_i(\mathbf{x}) \leq 0$ und $e_i(\mathbf{x}) \geq 0$ aufspaltet.

Welche Bedingungen sind nicht nur hinreichend, sondern auch notwendig für ein Minimum (Schölkopf und Smola 2002)?

Satz 19 (Notwendige Kuhn-Tucker Sattelpunktbedingung)

Gegeben die Annahmen von Satz 18 und unter den weiteren Annahmen, dass f und c_i konvex auf der konvexen Menge $X \subset \mathbb{R}^m$ sind, $\{\mathbf{x} | c_i(\mathbf{x}) \leq 0\} \subset X$ und dass es ein \mathbf{x} gibt mit $c_i(\mathbf{x}) < 0$, dann ist das Sattelpunkt-Kriterium aus Satz 18 notwendig für ein Minimum von f .

Für differenzierbare Funktionen lässt sich das Sattelpunktkriterium folgendermaßen formulieren:

Satz 20 (Kuhn-Tucker Sattelpunktbedingung für differenzierbare Funktionen)

Seien f und c_i aus Satz 18 konvexe, differenzierbare Funktionen. Falls $\mathbf{x} \in \mathbb{R}^m$ und $\hat{\alpha}_i \geq 0$ für alle $i \in \{1, \dots, n\}$, so dass gilt

$$\text{grad}_{\mathbf{x}} L(\hat{\mathbf{x}}, \hat{\boldsymbol{\alpha}}) = \text{grad}_{\mathbf{x}} f(\hat{\mathbf{x}}) + \sum_{i=1}^n \hat{\alpha}_i \text{grad}_{\mathbf{x}} c_i(\hat{\mathbf{x}}) = 0$$

(Sattelpunkt in $\hat{\mathbf{x}}$)

$$\partial_{\alpha_i} L(\hat{\mathbf{x}}, \hat{\boldsymbol{\alpha}}) = c_i(\hat{\mathbf{x}}) \leq 0 \quad (\text{Sattelpunkt in } \hat{\boldsymbol{\alpha}})$$

$$\sum_{i=1}^n \hat{\alpha}_i c_i(\hat{\mathbf{x}}) = 0,$$

dann ist $\hat{\mathbf{x}}$ eine Lösung des Optimierungsproblems (8.6).

8.10.2 Das Optimierungsproblem für Support-Vektor-Maschinen

In dem Fall, dass die Trainingsdaten durch eine Hyperebene separierbar sind, ist diejenige Hyperebene mit dem breitesten trennenden Rand gesucht. Das Optimierungsproblem, das bei Support-Vektor-Maschinen zu lösen ist, lautet also im separierbaren Fall:

$$\text{Minimiere } \tau(\mathbf{w}) = \frac{1}{2} \mathbf{w}^2$$

unter den Nebenbedingungen $y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1$ für alle $i \in \{1, \dots, m\}$.

Mit Satz 20 erhalten wir für die Lagrange-Funktion

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w}^2 - \sum_{i=1}^m \alpha_i [y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1]$$

und somit als Bedingungen für ein Minimum

$$\frac{\partial}{\partial b} L(\mathbf{w}, b, \boldsymbol{\alpha}) = \sum_{i=1}^m \alpha_i y_i = 0$$

$$\text{grad}_{\mathbf{w}} L(\mathbf{w}, b, \boldsymbol{\alpha}) = \mathbf{w} - \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i = 0$$

Die Funktion τ und die Nebenbedingungen erfüllen auch die Konvexitätsbedingungen aus Satz 20. Nach Satz 18 muss die Lagrange-Funktion L bezüglich

α maximiert werden. Setzen wir obige Bedingungen in die Lagrange-Funktion L ein, so erhalten wir folgendes alternatives Optimierungsproblem:

$$\begin{aligned} &\text{Maximiere } W(\mathbf{w}) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ &\text{unter den Nebenbedingungen } \alpha_i \geq 0 \text{ für alle } i \in \{1, \dots, m\} \\ &\text{und } \sum_{i=1}^m \alpha_i y_i = 0. \end{aligned}$$

Um nichtlineare Entscheidungsfunktionen einzuführen, denken wir uns den Eingaberaum \mathbb{R}^n abgebildet durch eine Funktion Ψ auf einen anderen, i.d.R. höherdimensionalen, Eingaberaum. Wir denken uns $\Psi(\mathbf{x}_i)$ als neue Eingabevektoren und brauchen in den Optimierungsproblemen somit nur \mathbf{x}_i durch die Bilder $\Psi(\mathbf{x}_i)$ zu ersetzen. Um $\Psi(\mathbf{x}_i)$ nicht jedesmal berechnen zu müssen, führt man wie oben eine positiv definite Funktion K ein, so dass

$$\Psi(\mathbf{x}) \cdot \Psi(\mathbf{x}_i) = K(\mathbf{x}, \mathbf{x}_i)$$

Damit lautet das Optimierungsproblem

$$\begin{aligned} &\text{Maximiere } W(\mathbf{w}) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &\text{unter den Nebenbedingungen } \alpha_i \geq 0 \text{ für alle } i \in \{1, \dots, m\} \\ &\text{und } \sum_{i=1}^m \alpha_i y_i = 0. \end{aligned}$$

Falls dann K positiv definit ist, so ist $(y_i y_j K(\mathbf{x}_i, \mathbf{x}_j))_{i,j}$ eine positiv definite Matrix (Schölkopf und Smola 2002), so dass das Optimierungsproblem wiederum die Konvexitätsbedingungen erfüllt.

Bisher haben die angegebenen Optimierungsprobleme nur dann eine Lösung, wenn die Trainingsdaten \mathbf{x}_i bzw. $\Psi \mathbf{x}_i$ durch eine Hyperebene separierbar sind. Im nichtseparierbaren Fall könnte man daran denken, eine Hyperebene zu berechnen die die Zahl der Trainingsfehler minimiert. Dieses Problem ist allerdings NP-hart. Auch die Näherungslösung „Finde eine Hyperebene die höchstens um einen konstanten Faktor schlechter ist als die optimale.“ ist NP-hart. Werden allerdings Eingabevektoren, die innerhalb eines Randes liegen, ignoriert, wird das Problem polynomiell (Schölkopf und Smola 2002). Alternativ verwendet Vapnik (1998) um Misklassifikationen zu erlauben sogenannte Schlupfvariablen

$$\xi_i \geq 0 \text{ für alle } i \in \{1, \dots, m\}$$

und weicht die Nebenbedingungen folgendermaßen auf:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \xi_i \text{ für alle } i \in \{1, \dots, m\}$$

Natürlich müssen die ξ_i irgendwie nach oben beschränkt werden. Deshalb wird in der zu minimierenden Funktion ein Strafterm eingeführt:

$$\begin{aligned} \text{Minimiere } \tau(\mathbf{w}, \boldsymbol{\xi}) &= \frac{1}{2} \mathbf{w}^2 + C \sum_{i=1}^m \xi_i \\ \text{unter den Nebenbedingungen } &\xi_i \geq 0 \\ \text{und } y_i(\mathbf{x}_i \cdot \mathbf{w} + b) &\geq 1 - \xi_i \text{ für alle } i \in \{1, \dots, m\}. \end{aligned}$$

Analog zu oben kann folgendes alternatives Optimierungsproblem abgeleitet werden (Schölkopf und Smola 2002):

$$\begin{aligned} \text{Minimiere } W(\boldsymbol{\alpha}) &= \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^m \alpha_i \quad (8.7) \\ \text{unter den Nebenbedingungen } &0 \leq \alpha_i \leq C \text{ für alle } i \in \{1, \dots, m\} \\ \text{und } \sum_{i=1}^m \alpha_i y_i &= 0. \end{aligned}$$

8.10.3 Ein effizienter Algorithmus zur Lösung des Optimierungsproblems für Support-Vektor-Maschinen

Wir lösen das Minimierungsproblem (8.7). Sei $\mathbf{g} := \text{grad } W(\boldsymbol{\alpha})$ der Gradient der zu minimierenden Funktion W an einem Punkt $\boldsymbol{\alpha}$. Sind \mathbf{e}_i Einheitsvektoren, dann ist nach der Kuhn-Tucker Sattelpunktbedingung 20 notwendige und hinreichende Bedingung für das Minimum: Es gibt λ und λ_i , so dass

$$\begin{aligned} \mathbf{g} &= \lambda \mathbf{y} + \sum_i \lambda_i \mathbf{e}_i \quad (8.8) \\ \lambda_i &= 0 \quad \text{für } 0 < \alpha_i < C \\ \lambda_i &\geq 0 \quad \text{für } \alpha_i = 0 \\ \lambda_i &\leq 0 \quad \text{für } \alpha_i = C. \end{aligned}$$

Wir bezeichnen dabei Ungleichungsnebenbedingungen mit $\alpha_i = 0$ und $\lambda_i > 0$ oder $\alpha_i = C$ und $\lambda_i < 0$, also solche, für die $\lambda_i \neq 0$, als „aktiv“, die anderen Ungleichungsnebenbedingungen als „inaktiv“. Die Idee des Algorithmus ist nun folgende: Minimiere W im Unterraum aller α_i mit $0 < \alpha_i < C$ sowie aller α_i , die zu inaktiven Ungleichungsnebenbedingungen gehören. Während dieser Minimierungprozedur können u.U. Ungleichungsnebenbedingungen, welche vorher inaktiv waren, aktiviert werden. Es können aber auch Ungleichungsnebenbedingungen, welche vorher aktiv waren, also im Unterraum nicht vorhanden und somit während der Minimierung auch nicht überprüft werden, inaktiv werden. Falls dies passiert, setzte den Algorithmus in

8.10. Effiziente Algorithmen für Support-Vektor-Maschinen 209

einem entsprechenden neuen Unterraum fort. Ansonsten beende den Algorithmus und das Minimum ist gefunden.

Wann werden nun Ungleichungsnebenbedingungen, die vorher aktiv waren, inaktiv? Der Algorithmus führt Buch über die aktiven Ungleichungsnebenbedingungen in einem Vektor \mathbf{e} mit

$$e_i = \begin{cases} 0 & \text{falls die Ungleichungsnebenbedingung } i \text{ aktiv ist} \\ y_i & \text{sonst} \end{cases}$$

Gegeben seien die – hypothetischen – aktiven Ungleichungsnebenbedingungen \mathbf{e} . Wir berechnen zuerst λ in (8.8). Es ist

$$\mathbf{e} \cdot \mathbf{g} = \lambda \mathbf{y} \cdot \mathbf{e},$$

da $e_i = 0$ für die aktiven Ungleichungsnebenbedingungen und $\lambda_i = 0$ für die inaktiven Ungleichungsnebenbedingungen. Mit $\mathbf{y} \cdot \mathbf{e} = \mathbf{e}^2$ erhalten wir

$$\lambda = \frac{\mathbf{e} \cdot \mathbf{g}}{\mathbf{e}^2}.$$

Da \mathbf{e}_i Einheitsvektoren sind, gilt für

$$\mathbf{s} := \lambda \mathbf{y} - \mathbf{g}$$

$s_i = -\lambda_i$, falls i eine aktive Ungleichungsnebenbedingung ist, also falls $e_i = 0$. Damit die Bedingungen (8.8) für ein Minimum erfüllt sind, muss also für alle i mit $e_i = 0$ gelten:

$$\begin{aligned} \alpha_i &= 0 \text{ und } s_i \leq 0 \text{ oder} \\ \alpha_i &= C \text{ und } s_i \geq 0. \end{aligned}$$

Falls eine der Bedingungen für den aktuellen „Kandidaten“ für ein Minimum α nicht erfüllt ist, wird mindestens eine Ungleichungsnebenbedingung, die vorher aktiv war, inaktiv und die Suche nach einem Minimum wird fortgesetzt.

Da zu Anfang der Suche sehr viele Bedingungen für ein Minimum nicht erfüllt sind, somit sehr viele Ungleichungsnebenbedingungen inaktiv sind, jedoch gegen Ende der Suche sehr viele α_i gleich 0 oder C sind, mit entsprechend aktiven Ungleichungsnebenbedingungen, führt nachfolgender Pseudocode die Minimierung nicht auf dem vollständigen Unterraum inaktiver Ungleichungsnebenbedingungen aus, sondern zunächst nur auf einer Teilmenge. Die Zahl der aktuell gefundenen „ungebundenen“ Supportvektoren, also derjenigen mit $0 < \alpha_i < C$, dient dabei als Schätzung für die Größe der zu verwendenden Teilmenge.

Die Optimierung im Unterraum wird mit dem konjugierten Gradientenverfahren durchgeführt. Die Unterfunktion `conjugate_gradient` berechnet den konjugierten Gradienten (Press et al. 1995). Dies hat den Vorteil, dass bei einem quadratischen Problem, wie dem vorliegenden, das Minimum *exakt* in n Schritten berechnet wird, wobei n die Dimension des Unterraums ist. Da sich allerdings der betrachtete Unterraum durch Aktivierung von Ungleichungsnebenbedingungen während der Minimumssuche ändern kann, sind die Voraussetzungen zur Anwendung des konjugierten Gradienten nicht exakt erfüllt. Deshalb fällt der Algorithmus auf einen herkömmlichen Gradienten zurück, falls die Abstiegsbedingung nicht mehr erfüllt ist, falls also die „Suchrichtung“ für ein Minimum keinen spitzen Winkel mehr mit dem negativen Gradienten der Funktion einschließt. In diesem Fall werden weitere n Schritte zur Berechnung des Minimums verwendet. Es kann nicht garantiert werden, dass mit diesem Verfahren das Minimum in n Schritten gefunden wird, falls Ungleichungsnebenbedingungen aktiviert werden und sich dadurch der Unterraum ändert. Auch numerische Ungenauigkeiten können auftreten. Deshalb hat nachfolgende Implementierung einen `high_accuracy`-Modus, bei dem statt der Zahl der Iterationen tatsächlich die Bedingungen für ein Minimum geprüft werden.

Der Algorithmus wurde in C++ implementiert und entspricht folgendem Pseudocode:

$\mathbf{e} := \mathbf{0}$

$\boldsymbol{\alpha} := \mathbf{0}$

$\mathbf{g} := \text{grad } W(\boldsymbol{\alpha})$

$N_{\text{SV}} := |\{\alpha_i \neq 0\}|$

$N_{\text{SVunbound}} := |\{\alpha_i \neq 0 \text{ und } \alpha_i \neq C\}|$

Falls $\mathbf{e} = \mathbf{0}$: $\mathbf{s} := -\mathbf{g}$, sonst: $\mathbf{s} := (\mathbf{g} \cdot \mathbf{e} / \mathbf{e}^2) \mathbf{y} - \mathbf{g}$.

Für alle i , für die $e_i = 0$ und (($\alpha_i = 0$ und $s_i > 0$) oder ($\alpha_i = C$ und $s_i < 0$)) setze $e_i := y_i$. Falls \mathbf{e} im vorigen Schritt geändert wurde, haben wir eine Ungleichungsnebenbedingung deaktiviert. Falls wir keine Ungleichungsnebenbedingung deaktiviert haben und `high_accuracy = true`, dann beende den Algorithmus. Falls wir keine Ungleichungsnebenbedingung deaktiviert haben, setze `high_accuracy := true`, sonst `high_accuracy := false`.

Sei *unsatisfied_Kuhn_Tucker_conditions_p* ein Index-Vektor mit den Indizes i für die gilt $e_i < 0$, sortiert nach absteigenden $|s_i|$.

Sei *unsatisfied_Kuhn_Tucker_conditions_n* ein Index-Vektor mit den Indizes i für die gilt $e_i > 0$, sortiert nach absteigenden $|s_i|$.

Falls $N_{\text{SVunbound}} = 0$, setze $n_{\min} := n_{\min} + 1$.

Setze alle $e(\text{unsatisfied_Kuhn_Tucker_conditions_}p_i) := 0$, beginnend mit $i = N_{\text{SVunbound}} + n_{\min}$.

Setze alle $e(\text{unsatisfied_Kuhn_Tucker_conditions_}n_i) := 0$, beginnend mit $i = N_{\text{SVunbound}} + n_{\min}$.

Setze $e_i := y_i$ für alle i mit $0 \leq \alpha_i \leq C$.

Sei $m := |\{i | e_i \neq 0\}|$ die Dimension eines reduzierten Optimierungsproblems.

Setze $\mathbf{e}^{\text{red}}, \boldsymbol{\alpha}^{\text{red}}, \mathbf{y}^{\text{red}}$ und $\mathbf{a}^{\text{red}} \in \mathbb{R}^m$ auf die jeweiligen Elemente i von $\mathbf{e}, \boldsymbol{\alpha}, \mathbf{y}$ und \mathbf{a} mit $e_i \neq 0$. Setze entsprechend $Q^{\text{red}} \in \mathbb{R}^{m \times m}$. Sei $i^* \in \{1, \dots, m\}$ und i der entsprechende Index mit $e_i \neq 0$, dann setze $a_{i^*}^{\text{red}} := a_i + \sum_{j, \alpha_j = C, e_j = 0} CQ_{ij}$.

Löse nun das reduzierte Problem der Minimierung von

$$W^{\text{red}} : \boldsymbol{\alpha} \rightarrow W^{\text{red}}(\boldsymbol{\alpha}) := \frac{1}{2} \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} + \mathbf{a} \cdot \boldsymbol{\alpha} \quad (8.9)$$

durch ein modifiziertes konjugiertes Gradientenverfahren:

while ($m > 0$ oder `high_accuracy`) **do**

$m := m - 1$

$\mathbf{g}^{\text{red}} := -\text{grad } W(\boldsymbol{\alpha}^{\text{red}})$

$\mathbf{s}^{\text{red}} := (\mathbf{g}^{\text{red}} \cdot \mathbf{e}^{\text{red}} / |\mathbf{e}^{\text{red}}|^2) \mathbf{y}^{\text{red}} - \mathbf{g}^{\text{red}}$

Für alle i mit $e_i = 0$ setze $s_i^{\text{red}} := 0$.

$\mathbf{c}^{\text{red}} := \text{conjugate_gradient}(-\mathbf{s}^{\text{red}})$

$\mathbf{s}^{\text{red}} := (\mathbf{c}^{\text{red}} \cdot \mathbf{e}^{\text{red}} / |\mathbf{e}^{\text{red}}|^2) \mathbf{y}^{\text{red}} - \mathbf{c}^{\text{red}}$

Für alle i mit $e_i = 0$ setze $s_i^{\text{red}} := 0$.

Falls die Abstiegsbedingung nicht erfüllt ist, also falls $\mathbf{s}^{\text{red}} \cdot \mathbf{g}^{\text{red}} > 0$, dann gehe zurück zum Gradientenverfahren: Setze den konjugierten Gradienten zurück, $m := |\{i | e_i \neq 0\}|$, $\mathbf{s}^{\text{red}} := (\mathbf{g}^{\text{red}} \cdot \mathbf{e}^{\text{red}} / |\mathbf{e}^{\text{red}}|^2) \mathbf{y}^{\text{red}} - \mathbf{g}^{\text{red}}$ und für alle i mit $e_i^{\text{red}} = 0$ setze $s_i^{\text{red}} := 0$.

Das Minimum $\alpha_{\min}^{\text{red}}$ in Richtung \mathbf{s} erhalten wir durch quadratische Interpolation von $W^{\text{red}}(\alpha)$. Setze $\alpha^{\text{red}} := \alpha_{\min}^{\text{red}}$.

Falls das neue α^{red} eine Ungleichungsnebenbedingung verletzt, es also nun einen Index i gibt mit $\alpha_i^{\text{red}} < 0$ oder $\alpha_i^{\text{red}} > C$, dann korrigiere α^{red} in Richtung $-\mathbf{s}$ so weit, bis keine Ungleichungsnebenbedingung mehr verletzt ist. Falls eine Ungleichungsnebenbedingung verletzt war, setze `high_accuracy := false`. Der Index, der für die letzte, und damit größte, Korrektur verantwortlich ist, sei i_{corr} .

Falls $e_{i_{\text{corr}}}^{\text{red}} \neq 0$, setze $e_{i_{\text{corr}}}^{\text{red}} := 0$ und $m := |\{i | e_i^{\text{red}} \neq 0\}|$.

Aus Gründen der numerischen Stabilität, korrigiere α^{red} , so dass die Gleichheitsnebenbedingung erfüllt ist und setze $\alpha_i \approx 0$ auf 0 und $\alpha_i \approx C$ auf C .

Falls die Änderung von W^{red} kleiner ist als eine vorgegebene Genauigkeit `ACCURACY` oder kein Fortschritt mehr erzielbar ist wegen $\mathbf{g}^{\text{red}} \cdot \mathbf{s}^{\text{red}} \leq \text{ACCURACY}$, dann brich ab, wenn diese Bedingung 5 mal hintereinander erfüllt war.

end

Aktualisiere α entsprechend α^{red} und \mathbf{e} entsprechend \mathbf{e}^{red} .

Berechne $\mathbf{g} := \text{grad } W(\alpha)$ neu.

Bei der Berechnung von $W^{\text{red}}(\alpha^{\text{red}})$ müssen hochdimensionale Skalarprodukte berechnet werden. Dies beansprucht einen Großteil der Laufzeit des Algorithmus. Deshalb enthält das im Rahmen dieser Arbeit entwickelte Programm einen Cache, der zu den entsprechenden Index-Paaren (i, j) das zugehörige Skalarprodukt $\mathbf{x}_i \cdot \mathbf{x}_j$ speichert und gleich den gespeicherten Wert zurückgibt, falls ein Index-Paar wiederholt aufgerufen wird.

Der Algorithmus ist effizient, wenn die Zahl der Support-Vektoren im Vergleich zur Zahl der Eingabe-Vektoren klein ist, was auch i. d. R. der Fall ist. Der Algorithmus vermeidet das Lösen von Gleichungssystemen einer Dimension, die größer ist als die Zahl der Support-Vektoren. Falls dann allerdings gute Kandidaten für Support-Vektoren bekannt sind, löst der Algorithmus das Optimierungsproblem gewissermaßen in einem Schritt mit dem konjugierten Gradientenverfahren, dessen Laufzeit kubisch in der Zahl der Support-Vektoren ist. Der „Sequential Minimal Optimization“ Algorithmus (Platt 1999) vermeidet das Lösen großer linearer Gleichungssysteme gänzlich, stellte sich in Experimenten allerdings als unzulänglich für vorliegendes Problem heraus. Einen ähnlichen Ansatz wie der vorgestellte Algorithmus verfolgt „SVM light“ (Joachims 1999).

8.11 Induktives Schließen mit Hilfe der Kolmogorov-Komplexität

„Induktives Schließen“ bedeutet hier, anhand von Einzelereignissen eine Wahrscheinlichkeitsverteilung für das Eintreten eines bestimmten Ereignisses anzugeben. Vom Besonderen (den Beobachtungen) wird also auf etwas Allgemeineres (eine Wahrscheinlichkeitsverteilung) geschlossen. Es existieren i. d. R. mehrere allgemeine Theorien, die die Beobachtungen erklären. Die Schwierigkeit besteht darin, eine dieser Theorien auszuwählen. Man bevorzugt dabei oft „einfachere“ Theorien gegenüber „komplexeren“, wendet also das Ockham'sche Prinzip an: „Von mehreren Theorien, die die gleichen Sachverhalte erklären, ist die einfachste allen anderen vorzuziehen.“ Die Aufgabe dieses Abschnitts ist es, den Begriff der „Einfachheit“ oder seines Gegenteils, der „Komplexität“, geeignet zu präzisieren. Hierzu wird der Begriff der „Kolmogorov-Komplexität“ eingeführt, nämlich grob gesagt die Länge des kürzesten Programms als Eingabe einer universellen Turing-Maschine¹¹ in Bits, das die Theorie beschreibt.

Der Einfachheit halber drücken wir die Aufgabe des „Lernens“ in Form der Vorhersage des nächsten Elements einer endlichen Folge von Elementen eines Basis-Alphabets \mathcal{B} aus. Wir beschreiben unser gesamtes bisheriges Wissen durch ein Element aus \mathcal{B}^* , der Menge aller endlichen Folgen aus Elementen von \mathcal{B} . Beispielsweise codieren wir ein Experiment durch eine endliche Folge aus Elementen von \mathcal{B} und hängen das Ergebnis in Form einer endlichen Folge aus Elementen von \mathcal{B} an, woran wir die Beschreibung des nächsten Experiments anhängen u.s.w. Eine Theorie ist dann ein Wahrscheinlichkeits-

¹¹Eine Turing-Maschine ist eine Zustandsmaschine mit einem unendlich langen Eingabeband, einem unendlich langen Ausgabeband, einem Lesekopf und einem Schreibkopf. Abhängig vom Zeichen auf dem Eingabeband unter dem Lesekopf, dem Zeichen auf dem Ausgabeband unter dem Schreibkopf und dem aktuellen Zustand der Maschine wechselt die Maschine in einen neuen Zustand. Bei jedem Zustandswechsel geschieht, abhängig von Start- und Zielzustand, folgendes:

- Die Maschine bewegt den Lesekopf um ein Zeichen vor.
- Die Maschine bewegt den Schreibkopf um ein Zeichen vor oder zurück.
- Die Maschine schreibt ein Zeichen auf das Ausgabeband.

Man vermutet, dass es keine Maschine gibt, welche mehr Funktionen berechnen kann als eine Turing-Maschine (Hypothese von Church) (Hopcroft und Ullman 1992). Jedes in einer beliebigen Programmiersprache geschriebene Programm lässt sich also in eine Turing-Maschine übersetzen. Eine universelle Turing-Maschine erwartet als Eingabe zunächst die Codierung einer beliebigen Turingmaschine und simuliert diese dann mit der auf die Codierung folgenden Eingabe. Eine universelle Turingmaschine entspricht also einem Compiler, der ein beliebiges Programm übersetzen und ausführen kann.

maß $\mu : \mathcal{P}(\mathcal{B}^\infty) \rightarrow \mathbb{R}$ über dem Ereignisraum \mathcal{B}^∞ aller unendlicher Folgen aus Elementen von \mathcal{B} . Die Menge aller Mengen $\mathcal{P}(\mathcal{B}^\infty)$ aus Elementen von \mathcal{B}^∞ (Potenzmenge) ist der Ergebnisraum. Wir definieren

$$\mu(x) := \mu(\{\omega \in \mathcal{B}^\infty \mid \omega \text{ beginnt mit } x\})$$

als die Wahrscheinlichkeit, dass eine Folge aus \mathcal{B}^∞ mit der Anfangssequenz x beginnt. Kennen wir μ , dann können wir dasjenige Element $a \in \mathcal{B}$ bestimmen, dass die größte Auftretenswahrscheinlichkeit $\mu(a|x)$ nach einer Anfangsfolge x hat. Nach der Bayes'schen Regel gilt nämlich

$$\mu(a|x) = \frac{\mu(xa)}{\mu(x)}.$$

Wir definieren allgemein $\mu(y|x)$ als die Wahrscheinlichkeit, dass die endliche Anfangsfolge $x \in \mathcal{B}^*$ mit der endlichen Folge $y \in \mathcal{B}^*$ fortgesetzt wird durch

$$\mu(y|x) = \frac{\mu(xy)}{\mu(x)}.$$

Die „korrekte“ Theorie μ ist meistens unbekannt. Wir suchen daher eine *universelle Theorie* ϱ , die wir für μ einsetzen können, so dass wir gute Vorhersagen $\varrho(y|x) \approx \mu(y|x)$ erhalten¹². Eine solche universelle Theorie existiert i. A. nicht (Li und Vitanyi 1997). Wir beschränken uns daher im Folgenden auf $\mathcal{B} = \{0, 1\}$ ¹³ und auf „rekursive“ Wahrscheinlichkeitsmaße μ .

Das Wahrscheinlichkeitsmaß μ heißt „rekursiv“ genau dann wenn die Funktion μ rekursiv ist. Eine Funktion mit Wertebereich \mathcal{B}^* heißt „rekursiv“ („berechenbar“), wenn eine Turingmaschine existiert, die für alle Eingaben im Definitionsbereich der Funktion hält und als Ausgabe das Funktionsergebnis zurückgibt (Hopcroft und Ullman 1992). Zum Begriff der Turingmaschine siehe Fußnote 6 auf Seite 194. Wir führen die rekursive Funktion $g : D \times \mathbb{N} \rightarrow \mathbb{Z} \times \mathbb{N}; (x, k) \mapsto g(x, k)$ ein. Als rekursive Funktion sollte

¹²Ein typischer Wert für x mit $\mathcal{B} = \{a-z; A-Z; 0-9; (;); \cdot; \cdot; \cdot; \cdot\}$ ist „Ich habe eine Kugel vom schiefen Turm von Pisa aus verschiedenen Höhen fallen gelassen, und dabei folgende Wertepaare (Höhe, Zeit) aus Höhe in Metern und Fallzeit in Sekunden, gerundet auf eine Stelle nach dem Komma, gemessen: (6.9,1.2), (13.8,1.7), (20.6,2.2), (27.5,2.5), (34.4,“ Die Theorie μ gibt nun an, mit welcher Wahrscheinlichkeit die gemessene Fallzeit bei der Höhe 34.4 m gleich y ist. Dies könnte die „Theorie des freien Falls“ sein. Eine universelle Theorie induziert aus den gegebenen Informationen in x zunächst eine Theorie des freien Falls und wendet sie an, um $\mu(y|x)$ zu bestimmen. Wir geben unten eine solche universelle Theorie an.

¹³Diese Einschränkung ist äquivalent zur Beschränkung von \mathcal{B} auf eine *endliche* Menge von „Symbolen“.

sie Sequenzen aus Elementen von \mathcal{B} erzeugen, wir fassen ihre Ausgabe jedoch als die codierte Form eines Zahlenpaars (p, q) auf, das wir als rationale Zahl p/q interpretieren. Eine Funktion f mit Definitionsbereich D und reellem Wertebereich heißt „rekursiv“, wenn es eine rekursive Funktion g wie oben beschrieben gibt, so dass $|f(x) - g(x, k)| < 1/k$ für alle $x \in D$. Eine Funktion f mit reellem Wertebereich heißt „rekursiv aufzählbar“, wenn eine rekursive, in k monoton steigende Funktion $(x, k) \mapsto g(x, k)$ existiert, so dass $f(x) = \lim_{k \rightarrow \infty} g(x, k)$.

Satz 21 (Induktionsprinzip von Solomonoff)

Es gibt ein („universelles“) Wahrscheinlichkeitsmaß $\mathbf{M} : \mathcal{B}^* \rightarrow \mathbb{R}$, so dass folgendes gilt:

Sei $\mu : \mathcal{B}^* \rightarrow \mathbb{R}$ ein rekursives Wahrscheinlichkeitsmaß und $\mathcal{B} = \{0, 1\}$.

Sei S_n der μ -Erwartungswert des quadratischen Fehlers zwischen μ -Wahrscheinlichkeit und \mathbf{M} -Wahrscheinlichkeit, dass das Ergebnis bei der n -ten Vorhersage 0 ist, also

$$S_n = \sum_{\{x \in \mathcal{B}^* | l(x) = n-1\}} \mu(x) [\mathbf{M}(0|x) - \mu(0|x)]^2,$$

wobei $l(x)$ die Länge von $x \in \mathcal{B}^*$ bezeichnet.

Dann konvergiert die Reihe $\sum_{n=1}^{\infty} S_n$.

Weiter gilt

$$\lim_{l(x) \rightarrow \infty} \frac{\mathbf{M}(y|x)}{\mu(y|x)} = 1$$

mit μ -Wahrscheinlichkeit 1.

Der Erwartungswert des quadratischen Vorhersagefehlers konvergiert daher schneller gegen null als $1/n$, wobei n die Zahl der nach μ verteilten Bits sind, die zur Vorhersage des nächsten Bits verwendet werden. Wird also eine universelle Wahrscheinlichkeitsverteilung \mathbf{M} als a-priori Verteilung in der Bayes'schen Regel verwendet, so konvergieren die vorhergesagten (bedingten) Wahrscheinlichkeiten sehr schnell gegen die wahren Wahrscheinlichkeiten. Eine schnellere Konvergenz kann man nicht erwarten, denn schätzt man allein die Auftretenswahrscheinlichkeiten von 0 und 1 aus den bisher beobachteten Häufigkeiten unter der Voraussetzung, dass die Einzelereignisse unabhängig von der Vergangenheit auftreten, so nimmt die Varianz des Schätzfehlers ebenfalls nur mit $1/n$ ab.

Das Problem der Wahl einer geeigneten a-priori Verteilung in der Bayes'schen Regel ist mit Satz 21 beweisbar gelöst. Eine mögliche Wahl von $\mathbf{M}(x)$ ist die Wahrscheinlichkeit, dass eine „monotone“ universelle Turingmaschine bei zufälliger, durch gerechten Münzwurf ausgewürfelte Eingabe aus $\{0, 1\}^\infty$ eine Ausgabe erzeugt, die mit x beginnt. Eine „monotone“ Turingmaschine ist eine Turingmaschine mit einem nur lesbaren, nur in eine Richtung beweglichen Eingabeband, mit einem nur einmal beschreibbaren, ebenfalls nur in eine Richtung beweglichen Ausgabeband und einem Arbeitsband. Berechnet daher eine monotone Turingmaschine die Funktion ψ und sind $x, y \in \mathcal{B}^*$ Eingaben, so ist die Ausgabe-Sequenz $\psi(x) \in \mathcal{B}^* \cup \mathcal{B}^\infty$ ein Präfix von $\psi(xy) \in \mathcal{B}^* \cup \mathcal{B}^\infty$, d. h. es gibt ein z aus $\mathcal{B}^* \cup \mathcal{B}^\infty$ mit $\psi(x)z = \psi(xy)$. Eine „universelle“ Turingmaschine erhält als Eingabe zunächst die *Codierung* einer Turingmaschine und simuliert dann dieselbe mit der auf diese Codierung folgenden Eingabe. Eine universelle monotone Turingmaschine simuliert nur (und alle) monotonen Turingmaschinen. Wir verwenden eine dieser universellen monotonen Turingmaschinen als „universelle monotone Referenzmaschine“. Diese berechne die Funktion U . Die Wahrscheinlichkeit, dass die universelle monotone Referenzmaschine bei zufälliger Eingabe aus $\{0, 1\}^\infty$ eine Ausgabe mit Präfix x berechnet, ist

$$\lambda_U(x) = \sum_{y \in Y} 2^{-l(y)},$$

wobei die Menge Y alle Eingaben $y \in \{0, 1\}^*$ enthält, so dass x Präfix von $U(y)$ ist und es kein y' mit $l(y') < l(y)$ gibt mit x Präfix von $U(y')$. Die Menge Y enthält daher keine Elemente, die Präfix eines anderen Elements aus Y sind. Wie oben schon erwähnt können wir in Satz 21 das universelle Wahrscheinlichkeitsmaß $\mathbf{M} = \lambda_U$ setzen.

Nun gibt es leider kein \mathbf{M} für Satz 21, das rekursiv geschweige denn rekursiv aufzählbar wäre. Es gibt allerdings ein solches rekursiv aufzählbares *Halbmaß*¹⁴ \mathbf{M} . Die bedingten „Wahrscheinlichkeiten“ $\mathbf{M}(y|x)$ bleiben definiert als

$$\mathbf{M}(y|x) = \frac{\mathbf{M}(xy)}{\mathbf{M}(x)}.$$

¹⁴Das Symbol ε steht für die leere Zeichenkette „“. Eine Funktion $\mu : \mathcal{B}^* \rightarrow \mathbb{R}$ heißt „Halbmaß“ in unserem Sinne, wenn für alle $x \in \mathcal{B}^*$ gilt

$$\begin{aligned} \mu(\varepsilon) &\leq 1 \\ \mu(x) &\geq \sum_{a \in \mathcal{B}} \mu(xa). \end{aligned}$$

Ist μ ein Wahrscheinlichkeitsmaß, so gelten Gleichheitszeichen.

Was hat nun die universelle Wahrscheinlichkeitsverteilung \mathbf{M} mit Komplexität zu tun? Dazu führen wir das Komplexitätsmaß der „monotonen Komplexität“ $Km : \mathcal{B}^* \rightarrow \mathbb{N}$ ein, definiert als

$$Km(x) = \min\{l(p) \mid x \text{ ist Präfix von } U(p)\},$$

d. h. $Km(x)$ ist die Länge eines kürzesten Programms p für die universelle monotone Referenzmaschine, das x am Anfang der Ausgabe erzeugt. Die monotone Komplexität Km ist monoton in dem Sinn, dass $Km(x) \leq Km(xy)$ für alle $x, y \in \mathcal{B}^*$, was direkt aus der Definition folgt.

Satz 22 (Vorhersage durch Datenkompression)

Sei μ ein rekursives Wahrscheinlichkeitsmaß. Die Folge $\omega \in \mathcal{B}^\infty$ mit $\mathcal{B} = \{0, 1\}$ sei nach der Wahrscheinlichkeitsverteilung μ zufällig gewählt. Dann gilt mit Wahrscheinlichkeit 1: Ist $l(x)$ groß genug, dann ist

$$\operatorname{argmax}_y \mu(y|x) = \operatorname{argmin}_y [Km(xy) - Km(x)].$$

d. h. nach einer Vorgeschichte x , die lang genug ist, kommt dasjenige y mit der größten Wahrscheinlichkeit $\mu(y|x)$, das, gegeben x , am besten komprimiert werden kann, in dem Sinne dass $Km(xy) - Km(x)$ minimal ist. Hat man ein kürzestes Programm für x gefunden und erzeugt dieses nach der Ausgabe von x die weitere Ausgabe y , so gilt $Km(xy) - Km(x) = 0$ und das Programm sagt y als einen wahrscheinlichsten Wert voraus. Somit ist Satz 22 eine exakte Formulierung des „minimum description length“-Prinzips (MDL) von J. J. Rissanen, C. S. Wallace und anderen (Rissanen 2005), das besagt: „Gegeben seien Daten und eine Menge von Theorien, die die Daten beschreiben. Die beste Theorie ist diejenige, die die Summe aus der Länge der Theorie in Bits und der Beschreibung der Daten in Bits unter Verwendung dieser Theorie minimiert.“ Nun ist besagte Summe nichts anderes als die Länge des kürzesten Programms für eine universelle Turingmaschine, das die Daten reproduziert, denn dieses Programm enthält als „Kopf“ zunächst die Codierung einer Turingmaschine (d. i. die „Theorie“ im Sinne von MDL) und danach das eigentliche Programm (d. i. die „Beschreibung der Daten unter Verwendung der Theorie“ im Sinne von MDL). Verwendet man anstatt der universellen Turingmaschine eine *monotone* Turingmaschine, somit Km als „Länge des kürzesten Programms“, dann ist MDL im Sinne von Satz 22 beweisbar richtig. Leider ist Km nicht rekursiv. Praktische Anwendungen von MDL sind daher zur Zeit nur Näherungen.

Lassen sich die Sinneseindrücke eines Lebewesens also als binäre Zeichenfolge beschreiben und ist die Wahrscheinlichkeitsverteilung μ dieser Zeichenfolgen rekursiv (berechenbar), dann gibt es eine Funktion, welche diese Wahrscheinlichkeitsverteilung „lernt“ im Sinne der Sätze 21 und 22. Analoges gilt für physikalische Theorien. Leider sind die oben angegebenen Funktionen \mathbf{M} und Km selbst nicht rekursiv (berechenbar) (Vidyasagar 1997), wodurch dieser Ansatz für alle praktischen Zwecke ausscheidet.

Man kann die „Intelligenz“ eines Lebewesens oder einer Maschine als deren Lern-Fähigkeit im obigen Sinn definieren. Ein Lebewesen ist demnach intelligent, wenn es sich möglichst schnell an möglichst viele im Sinne der Kolmogorov-Komplexität einfache Umgebungen anpasst, in diesen Umgebungen also eine hohe durchschnittliche Belohnung erzielt (Legg und Hutter 2006).

*With ten parameters, you could
fit an elephant. With eleven, you
could wag its tail too!*

Douglas Buttrey, Lord Kelvin,
Richard Feynman

9

Rekonstruktion rezeptiver Felder mit Support-Vektor-Maschinen

In diesem Kapitel werden rezeptive Felder mit Hilfe der Support-Vektor-Maschinen aus Abschnitt 8.7 mit dem in Abschnitt 8.10.3 entwickelten Algorithmus rekonstruiert.

9.1 Eindimensionale rezeptive Felder

Das zu rekonstruierende rezeptive Feld ist hier wieder das „rechteckige“ Feld aus Abschnitt 6.2.1. Stimulus sind die zufälligen Pieptöne aus Abschnitt 6.5. Das Neuronenmodell ist das Schwellenelement aus Abschnitt 6.6. Wie Abbildung 9.1 zeigt, lässt sich die Form des rezeptiven Feldes mit einem weiten Bereich von Kontrollparametern C gut rekonstruieren. Die eingesetzte Support-Vektor-Maschine ist hier linear.

9.2 Mehrdimensionale rezeptive Felder

Abbildung 9.2 zeigt, dass sich auch mehrdimensionale rezeptive Felder mit Hilfe von Support-Vektor-Maschinen rekonstruieren lassen. Stimulus und Neuronenmodell sind gewählt wie in Abschnitt 9.1.

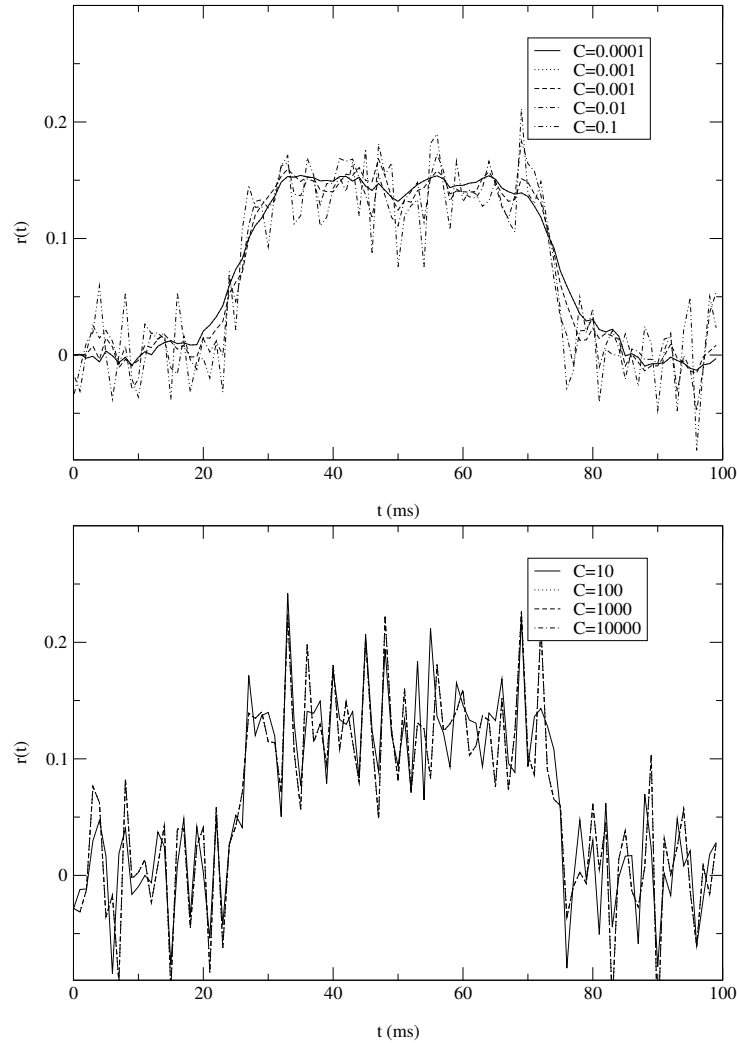


Abbildung 9.1: Rekonstruierte rezeptive Felder \hat{r} nach der Support-Vektor-Methode mit den angegebenen Kontrollparametern C .

Eingabe in das als Schwellenelement angenommene Neuron ist ein Cochleogramm bei der Frequenz 1000 Hz. Die Cochlea wird mit zufälligen Pieptönen mit logarithmischer Frequenzverteilung zwischen 100 und 16000 Hz, Dichte 200 Töne pro Sekunde, mittlerer Dauer 10 ms (Standardabweichung 10 ms) und Hanning-Rampen von 2 ms angeregt. Die Gesamtdauer der Stimulation beträgt 100 s.

Die rechteckige Form des rezeptiven Feldes wird in einem weiten Bereich von Kontrollparametern richtig rekonstruiert. Das „Rauchen“ in der Rekonstruktion wird umso stärker, desto größer der Kontrollparameter ist.

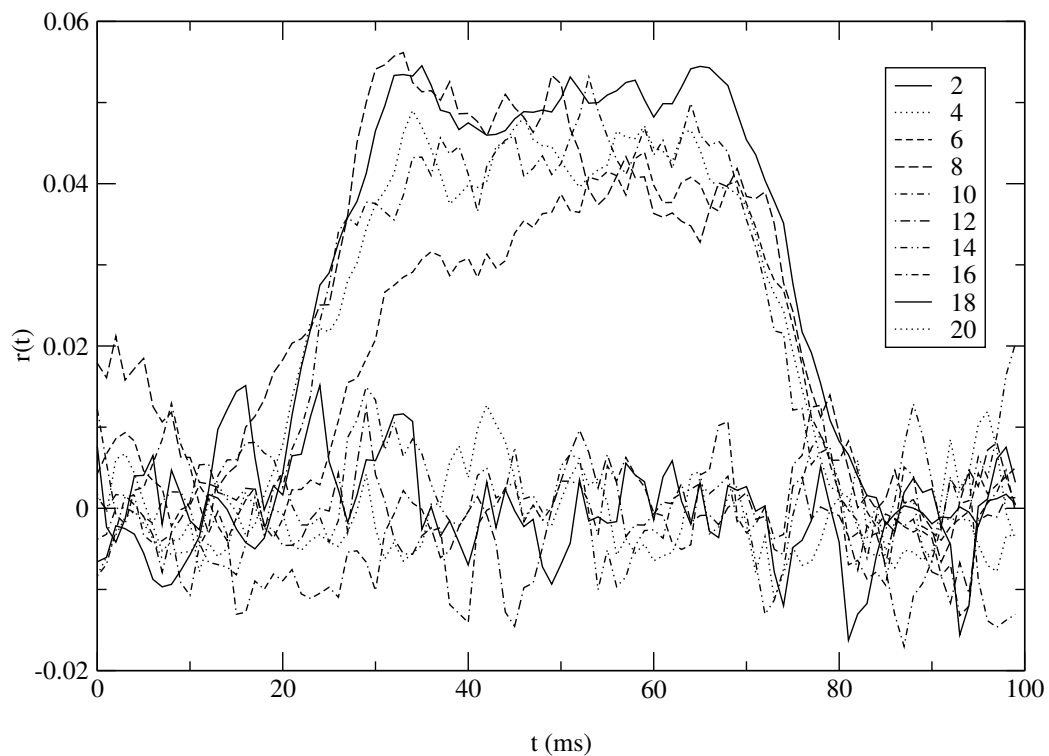


Abbildung 9.2: Rekonstruierte rezeptive Felder \hat{r} nach der Support-Vektor-Methode mit Kontrollparameter $C = 0.001$.

Wie Abbildung 9.1, nur ist das rezeptive Feld diesmal mehrdimensional wie in Abbildung 7.1.

Die Form des rezeptiven Feldes wird näherungsweise durch die Rekonstruktion wiedergegeben.

*Wovon man nicht sprechen kann,
darüber muss man schweigen.*

Ludwig Wittgenstein¹

10

Diskussion und Ausblick

Wir haben gesehen, dass sich rezeptive Felder von spikenden Neuronen mit der Regularisierungsmethode (Abschnitte 5.2 und 7) und mit Hilfe von Support-Vektor-Maschinen (Abschnitt 9) rekonstruieren lassen. Die Maximum-Likelihood-Methode versagt bei Rauschen in der Messung, denn die Maximum-Likelihood-Methode entspricht der Regularisierungsmethode mit $\sigma \rightarrow 0$ aus Gleichung (5.9), vgl. Abschnitt 6.2.6. Die oft ad-hoc angewandte „Reverse-Correlation“-Methode (Cai et al. 1997, Klein et al. 2000, 2006) ist von vorneherein mit systematischen Fehlern behaftet, da hier von weißem Rauschen als Stimulus ausgegangen wird, somit die Frequenzverteilung echter Stimuli in keiner Weise berücksichtigt wird.

Diese Arbeit beschränkt sich auf künstliche, also simulierte, Neuronen, deren Antwortverhalten natürlich genau bekannt ist, so dass sich rekonstruierte und wirkliche rezeptive Felder auch miteinander vergleichen lassen. Dabei haben sich die Regularisierungsmethode und die Support-Vektor-Methode als relativ zuverlässig erwiesen, wenn auch diese noch freie Parameter (im Fall der Regularisierungsmethode σ und d und im Fall der Support-Vektor-Methode C) enthalten, welche das Messergebnis beeinflussen. Die wahren Antwortigenschaften von echten Neuronen wird man durch eine Messung mit statistischem Charakter, die nur bekannte Input-Output-Relationen verwendet, wohl nie genau berechnen können, denn bei der Vielzahl von synaptischen Verbindungen und potentiellen Verzögerungszeiten wären sehr lange Mes-

¹Wittgenstein (1966).

szeiten notwendig, ganz abgesehen davon, dass man den Input nicht genau kennt. Bislang ist es jedoch nicht möglich, einzelne Synapsenstärken in vivo auf irgendeine Art und Weise zu messen, geschweige denn die Stärken aller Synapsen eines relevanten Teils des Gehirns. Man wird also vorerst auf die Messung von Input-Output-Relationen beschränkt bleiben. Deshalb sollten rezeptive Felder zunächst mit geeignet erscheinenden Parametern gemessen werden, dann jedoch sollte die Messung überprüft werden, indem Modellneurone mit den gemessenen oder ähnlichen rezeptiven Feldern der gleichen „Messung“ in einer Simulation ausgesetzt werden, wodurch die zu erwartenden Messfehler abgeschätzt werden können. Eine andere Möglichkeit, Messfehler abzuschätzen, wäre Bootstrap- oder Jackknife-Methoden anzuwenden, bei denen bei der Auswertung der Messung unterschiedliche Teile der Messdaten weggelassen werden (Duda et al. 2000). Die Abweichungen der resultierenden Statistiken² sind dann eine Näherung für die statistischen Fehler. Diese letztere Methode erfordert jedoch eine multiple Auswertung, was derzeit nur mit schnellen Methoden wie der Regularisierungsmethode möglich wäre. Die im Rahmen dieser Arbeit untersuchten Methoden und die entwickelte Software sind geeignet, beide Methoden praktisch durchzuführen. Das hier vorgeschlagene Vorgehen wäre nun, nach einer geeigneten Vorverarbeitung des Stimulus, wie sie im peripheren Teil des untersuchten Systems vermutlich stattfindet, die Antworteigenschaften der gemessenen Neurone mit Hilfe deren spektro- oder spatio-temporalen rezeptiven Felder anhand der in dieser Arbeit vorgestellten Verfahren zu messen. Hierdurch erhält man eine anschauliche Beschreibung der Neurone im an den peripheren Bereich anschließenden Kern und brauchbare Modelle für diese Neurone. Natürlich wird es nicht möglich sein, alle Neurone oder auch nur einen größeren Bruchteil in einem Kern nachzumessen. Man wird also anhand der gemessenen rezeptiven Felder gewisse Systematiken feststellen müssen, so dass sich diese Felder mit wenigen Parametern (z. B. „Breite“ und „Verzögerungszeit“) beschreiben lassen. Dann wird man Hypothesen aufstellen, welchen Parameterbereich die Neurone im entsprechenden Kern abdecken, d. i. welche *neuronalen Karte* die Neurone bilden. Somit lässt sich dann der Input *der nächst höheren Ebene* durch ein Modell beschreiben. Diesen nun leidlich bekannten Input kann man verwenden, um eine weitere Input-Output-Relation zu messen, nun einen Kern höher. Das eben skizzierte Verfahren geht davon aus, dass sich Neurone hinreichend gut als lineare Schwellenelemente beschreiben lassen, was der Fall sein dürfte, vgl. Abschnitt 1.5. Es geht jedoch auch

²Eine „Statistik“ ist ein sich aus der Messung einer Zufallsgröße ergebender Zahlenwert, oder, anders ausgedrückt, eine Funktion, die die Ausprägungen einer Zufallsgröße in die reellen Zahlen abbildet (Rüger 1996).

davon aus, dass es keine allzu starke Rückkopplung zwischen den untersuchten Kerngebieten gibt. In der Realität ist die Rückkopplung jedoch u. U. sogar stärker als die Vorwärts-Kopplung. Wie eine solche Rückkopplung im Allgemeinen zu behandeln ist, stellt ein noch ungelöstes Problem dar. Man könnte iterativ vorgehen, indem zunächst so getan wird, als sei keine Rückkopplung vorhanden, rezeptive Felder gemessen werden und dann anhand bekannter Eigenschaften der Modell-Neurone mögliche Rückkopplungen berechnet werden. Inwieweit ein solches Verfahren konvergiert, ist jedoch noch zu untersuchen.

Reale Neurone besitzen ein Refraktärverhalten (Abschnitt 1.5), das den Spike-Generierungs-Prozess beeinflusst. Das Refraktärpotential ist Teil des Neuronenmodells. Es wäre zu überlegen, wie man auch das Refraktärverhalten des untersuchten Neurons messen könnte, um das Neuronenmodell in dieser Hinsicht genauer zu machen. Hierzu könnte beispielsweise weiterer „Dummy-“ Input $x_0(t)$ eingeführt werden, der mit $x_0(t) = 1$, anzeigt, zu welcher Zeit t in der Vergangenheit das betrachtete Neuron Spikes erzeugt hat. Das rekonstruierte rezeptive Feld $\hat{r}_i(t)$ würde dann im Kanal 0 als $\hat{r}_0 t$ das Refraktärpotential enthalten. Dabei ist jedoch zu beachten, dass nun erzwungen werden muss, dass $r_0(t) = 0$ für $t < 0$, damit das Neuronenmodell kausal bleibt, also keine Information über *zukünftige* Spikes enthält. Bei den in den Abschnitten 4 und 5 entwickelten Verfahren ist dies nicht gewährleistet, es könnten also theoretisch auch rezeptive Felder geschätzt werden, die Informationen aus *zukünftigen* Stimuli verwenden, d. h. es tritt $\hat{r}_i(t) \neq 0$ für $t < 0$ auf. Dabei ist zu berücksichtigen, dass die eingesetzte FFT (Press et al. 1995) grundsätzlich alle diskretisierten Funktionen als zyklisch betrachtet, d. h. die Eingabe und die Ausgabe wiederholen sich zyklisch, was eine genaue Unterscheidung von Vergangenheit und Zukunft unmöglich macht, da sich die Vergangenheit in der Zukunft im nächsten Zyklus exakt wiederholt. Es gibt weitere hier nicht untersuchte statistische Verfahren, um Klassifikatoren zu schätzen. Während das hier eingesetzte Verfahren der Minimalen Quadrate für Poisson-Neuronen konsistent ist (vgl. Abschnitt 5.5), ist für Schwellenelemente oder reale Neuronen eventuell ein Logit- oder Probit-Schätzer³ zur Schätzung der *Feuerwahrscheinlichkeit* besser an das Problem

³Sei $y \in \{0,1\}$ und \hat{y} eine vorherzusagende Wahrscheinlichkeit für $y = 1$ und x_i , $i = 1, \dots, n$, Variablen, von denen \hat{y} abhängt. Sei $\Phi : \mathbb{R} \rightarrow [0,1]$ eine streng monoton steigende Funktion mit $\lim_{x \rightarrow -\infty} \Phi(x) = 0$ und $\lim_{x \rightarrow \infty} \Phi(x) = 1$. Dann wird für \hat{y} der Ansatz

$$\hat{y}(\mathbf{x}) = \Phi \left(r_0 + \sum_{i=1}^n r_i x_i \right)$$

gewählt. Die direkte Anwendung der Maximum-Likelihood-Methode und der Ansatz, die durchschnittliche quadratische Abweichung $\langle [\hat{y}(\mathbf{x}) - y]^2 \rangle$ der Messdaten zu mi-

angepasst. Inwieweit solch ein Schätzer mit Hilfe von FFT effizient zu implementieren ist und wie hier eine Regularisierung durchgeführt werden könnte, wäre zu untersuchen.

nimieren, führen auf ein, u. U. sehr großes, nichtlineares Gleichungssystem für die Unbekannten r_i . Deshalb die Idee, obige Gleichung nach $r_0 + \sum_{i=1}^n r_i x_i$ aufzulösen und $\langle [r_0 + \sum_{i=1}^n r_i x_i - \Phi^{-1}(y)]^2 \rangle$ zu minimieren, was auf ein *lineares* Gleichungssystem in den Unbekannten r_i führt. Beim Logit-Schätzer ist $\Phi(x) = \frac{1}{1+e^{-x}}$, beim Probit-Schätzer ist $\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ die Verteilungsfunktion der Standard-Normalverteilung. Wird obiges lineares Gleichungssystem entsprechend gewichtet iterativ gelöst, dann sind die geschätzten r_i Maximum-Likelihood-Schätzwerte (Greene 2000).

11

Zusammenfassung

Ein Neuron reagiert auf einen Stimulus mit einer Folge von Aktionspotentialen, die mit jeder Wiederholung des Stimulus verschieden ausfallen. Die Wahrscheinlichkeitsdichte $p(t)$ für das Auftreten eines Aktionspotentials zum Zeitpunkt t nach Einsetzen des Stimulus beschreibt das Antwortverhalten eines Neurons *auf diesen Stimulus* unter gewissen Voraussetzungen weitgehend vollständig. Das Antwortverhalten des Neurons ist dann offenbar die Wahrscheinlichkeitsdichte für das Auftreten eines Aktionspotentials *in Abhängigkeit* vom Stimulus.

Die „Messung“ des Antwortverhaltens ist die *Folgerung* des Antwortverhaltens aus dem Experiment, d. h. aus angelegtem Stimulus und gemessenen Aktionspotentialen wird das Antwortverhalten bei *beliebigem* Stimulus ermittelt.

Am einfachsten wäre es, das Neuron als linearen Filter zu beschreiben. Die Impulsantwort des Filters kann man auch als „rezeptives Feld“ bezeichnen. Solch ein Ansatz muss jedoch für das auditorische System verworfen werden, denn die Vorverarbeitung des Schalls in der Cochlea ist nichtlinear, denn hier wird der Schall nach Frequenzbereichen gefiltert. Die inneren Haarzellen feuern um so schneller, je mehr Energie in den entsprechenden Frequenzfilter gelangt. Die Operation der Berechnung der Energie aus der Auslenkung (Bildung des Betragsquadrats) ist nicht linear.

Die zeitabhängige Feuerwahrscheinlichkeit der inneren Haarzellen ist das „Cochleogramm“. Ein besserer Ansatz ist es nun, ein auditorisches Neuron als linear im Cochleogramm anzunehmen. Die Antwort ist dann eine Faltung

des Cochleogramms mit einem „spekro-temporalen“ rezeptiven Feld.

Ein bisheriger Ansatz, um ein lineares rezeptives Feld zu messen, ist die Messung der Korrelation mit Gauß'schem weißem Rauschen. Es ist allerdings nicht möglich, Gauß'sches weißes Rauschen auf die Cochlea zu bekommen. Einen Ausweg stellen Optimierungsverfahren dar. Ein naiver Ansatz scheitert jedoch, da das rezeptive Feld sehr viele Parameter enthält.

Psychoakustische Grundlagen sind für das Verständnis jedweder auditorischen Modelle notwendig. Die Psychoakustik als Teilgebiet der Psychophysik beschäftigt sich also mit dem Zusammenhang zwischen den physikalischen Eigenschaften des Schalls und den Wahrnehmungen.

Für das auditorische System gibt es verschiedene Modelle, welche zum Verständnis der neuronalen Antworten wichtig sind. Die Cochlea korrekt zu modellieren ist sehr aufwändig, es genügt jedoch die Cochlea als linearen Filter aufzufassen, d. h. die Auslenkung einer Stelle an der Basilarmembran ist ein lineares Funktional des auf das Trommelfell einwirkenden Schalls.

Verschiedene Verfahren zur Rekonstruktion des rezeptiven Feldes bei verauschten Messdaten werden aus Optimierungsprinzipien abgeleitet und untereinander verglichen. Das Neuron wird zunächst auf diejenige Art und Weise linear genähert, dass der quadratische Fehler zwischen dem Output des linearen Neurons bei gegebenem Input und den Messdaten minimal ist. Eine Alternative ist die Anwendung der Bayes'schen Methode: Berechne das wahrscheinlichste lineare rezeptive Feld unter der Annahme, dass der gemessene neuronale Output durch Gauß'sches Rauschen verfälscht ist und das rezeptive Feld a-priori wie Gauß'sches weißes Rauschen verteilt ist. Die Regularisierungsmethode ist eine Verallgemeinerung.

Ein ganz anderer Ansatz zur Vorhersage von Verhalten kommt aus der statistischen Lerntheorie. Zwei konkrete Ansätze werden vorgestellt, nämlich Support-Vektor-Maschinen und induktives Schließen mit Hilfe der Kolmogorov-Komplexität. „Lernen“ ist eine Verhaltensänderung aufgrund von Erfahrung. Bei der Konstruktion einer lernenden Maschine ist man bestrebt, den Generalisierungsfehler möglichst gering zu halten. Eine Erweiterung einer einfachen generalisierenden lernenden Maschine, des Perzeptrons, sind optimale Hyperebenen, die zur Konstruktion von Support-Vektor-Maschinen führen. Es wird ein effizienter Algorithmus entwickelt, um das bei Support-Vektor-Maschinen auftretende Optimierungsproblem zu lösen. Induktives Schließen mit Hilfe der Kolmogorov-Komplexität ist ein interessantes mathematisches Konstrukt, dessen praktische Verwertung jedoch an bewiesenermaßen exponentieller Laufzeitkomplexität scheitert. Support-Vektor-Maschinen allerdings lassen sich mit Hilfe des entwickelten Algorithmus verwenden, um rezeptive Felder von Neuronen zu rekonstruieren. Dies ist auch für mehrdimensionale rezeptive Felder möglich.

12

Summary

To a stimulus, a neuron responds with a series of action potentials which is different after each repetition of the stimulus. The probability density $p(t)$ for the occurrence of an action potential at time t after stimulus onset describes the response properties of the neuron *to this stimulus* more or less completely, under some conditions. The response property of the *neuron* are then the probability for the occurrence of an action potential *in dependence of* the stimulus.

A “measurement” of the response property is the *deduction* of the response property from an experiment, i.e. after measuring the neuronal response that occurs after presentation of several stimuli, one generalizes to *arbitrary* stimuli.

It would be simplest to describe a neuron as a linear filter. The impulse response of such a filter would then be the “receptive field” of the neuron. However, such an approach must be rejected for the auditory system as the sound preprocessing in the cochlea is nonlinear because the cochlea filters the sound according to the different frequencies. The inner hair cells are firing as fast as the more energy is contained in the respective frequency band. The operation of calculating the energy from amplitude (squaring) is nonlinear. The time dependent probability of spiking of the inner hair cells is the “cochleogram”. It is far better to assume that auditory neurons are linear in the cochleogram. Then, the response is a convolution of the cochleogram with the “spectro-temporal” receptive field.

To measure the receptive field one can correlate the neuronal response with

Gaussian white noise. Unfortunately it is not possible to generate Gaussian white noise at a position in the cochlea. A resort would be optimization methods. However, a naive approach fails because the receptive field contains too many parameters.

A psychoacoustic basis is essential for understanding of models of the auditory system. Psychoacoustics is the part of psychophysics that is concerned with the relation of physical properties of sound with the relating subjective perceptions.

Different models have been developed for the auditory system which are essential for understanding neuronal responses. To model the cochlea accurately is very elaborate, but here it is sufficient to model it as a linear filter, i. e. the deflection at one position is a linear functional of the sound arriving at the eardrum.

Several methods to reconstruct a receptive field from noisy measurements are derived as optimization problems and compared. First of all, the neuron is approximated linearly such that the quadratic error between the output of the linear neuron given the input and the measurements is minimal. An alternative is Bayes' method that calculates the most probable linear receptive field under the condition that the measured neuronal output contains Gaussian white noise and the receptive field has an a-priori probability distribution that can be described by Gaussian noise. Regularization methods are a generalization of the other concepts.

A completely different approach to predict responses comes from statistical learning theory. Two concrete attempts are presented: support vector machines and inductive reasoning with the help of Kolmogorov complexity. "Learning" is a change of behavior due to experience. When constructing a learning machine one should minimize the generalization error. The perceptron is a simple generalizing learning machine. A generalization are optimal hyperplanes that lead to the concept of support vector machines. An efficient algorithm is developed to solve the optimization problem that arises. Inductive reasoning with the help of Kolmogorov complexity is an interesting mathematical construct whose applicability fails due to provably exponential runtime complexity of the necessary algorithms. Support vector machines however can be used with the developed algorithm to reconstruct receptive fields of neurons. This is also possible for multidimensional receptive fields.

Never underestimate the joy people derive from hearing something they already know.

Enrico Fermi

A

Kovarianz und Faltung

In diesem Kapitel sollen einige Eigenschaften der Kovarianzfunktion und der Faltung aufgeführt werden, welche in Berechnungen immer wieder benötigt werden.

A.1 Kovarianz

Die (Kreuz-) „Kovarianzfunktion“ zwischen zwei reellen Funktionen $x : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto x(t)$ und $y : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto y(t)$ ist definiert als (Schlittgen und Streitberg 1999)

$$\text{cov}(x, y)(t) = \int_{-\infty}^{\infty} x(\tau) y(\tau + t) d\tau. \quad (\text{A.1})$$

Bezeichne $\|x\| := \int_{-\infty}^{\infty} x^2(t) dt$ die L^2 -Norm der Funktion x . Die mit dem Produkt aus den Normen der Funktionen normierte Kovarianzfunktion

$$\text{corr}(x, y)(t) := \frac{\text{cov}(x, y)(t)}{\|x\| \|y\|}, \quad (\text{A.2})$$

heißt (Kreuz-) „Korrelationsfunktion“.

Die Funktion $\text{corr}(x, y)(0)$ an der Stelle null ist ein Maß für die Ähnlichkeit der Funktionen x und y , denn die sie ist nach oben und unten beschränkt

$$0 \leq \text{corr}(x, y)(0) \leq 1,$$

und nimmt den maximalen Wert an, d. h.

$$\text{corr}(x, y)(0) = 1,$$

genau dann wenn die Differenz der beiden Funktionen in der L^2 -Norm verschwindet, also $\|x - y\| = \int_{-\infty}^{\infty} [x(t) - y(t)]^2 dt = 0$ ist.

Die Funktion

$$\text{cov}(x, x) : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto \text{cov}(x, x)(t)$$

heißt „Autokovarianzfunktion“ der Funktion x .

Die Autokovarianzfunktion von x an der Stelle null ist die L^2 -Norm von x .

$$\text{cov}(x, x)(0) = \int_{-\infty}^{\infty} x^2(t) dt$$

Die Kovarianz ist antisymmetrisch

$$\text{cov}(y, x)(t) = \text{cov}(x, y)(-t) \quad (\text{A.3})$$

wegen

$$\begin{aligned} \text{cov}(y, x)(t) &= \int_{-\infty}^{\infty} y(\tau) x(\tau + t) d\tau = \\ &= \int_{-\infty}^{\infty} x(\tau + t) y(\tau) d\tau = \\ &= \int_{-\infty}^{\infty} x(\tau') y(\tau' - t) d\tau' = \\ &= \text{cov}(x, y)(-t) \end{aligned}$$

mit $\tau' := \tau + t$.

Die Kovarianz ist distributiv von links

$$\text{cov}(x, y + z) = \text{cov}(x, y) + \text{cov}(x, z) \quad (\text{A.4})$$

und von rechts

$$\text{cov}(x + y, z) = \text{cov}(x, z) + \text{cov}(y, z).$$

Es gilt das gemischte Assoziativgesetz

$$c \text{cov}(x, y) = \text{cov}(cx, y) = \text{cov}(x, cy)$$

für eine Konstante c .

A.2 Faltung

Die Faltung zwischen zwei reellen Funktionen $x : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto x(t)$ und $y : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto y(t)$ ist definiert als

$$(x \star y)(t) = \int_{-\infty}^{\infty} x(\tau) y(t - \tau) d\tau.$$

Die Faltung spielt eine große Rolle, da ein lineares, zeitinvariantes System mit Systemantwort y auf die Eingabe x beschrieben werden kann als

$$y = r \star x, \tag{A.5}$$

wobei r Impulsantwort des Systems heißt (Unbehauen 1993). Neuronen sind i. A. zwar nicht linear, jedoch kann zumindest das Membranpotential als näherungsweise linear im Eingabe-Spiketrain aufgefasst werden. Im Spike-Response-Neuronenmodell (Gerstner und van Hemmen 1994) gilt diese Linearität exakt, siehe Gleichung (1.8). Falls sich die synaptischen Gewichte des Neurons nicht ändern, also kein Lernen vorliegt, ist das Neuron ein zeitinvariantes System. Zeitinvariant bedeutet, die Antwort auf einen (genügend langen) Eingang-Spiketrain ist immer gleich, unabhängig von der Vorgeschichte. Falls das Neuron eine Zufallskomponente beinhaltet, ist zumindest das PSTH zeitinvariant.

Ist die Eingabe x des Systems ein Puls, also $x = \delta$, so ergibt sich aus (A.5) für die Systemantwort

$$y(t) = (r \star \delta)(t) = \int_{-\infty}^{\infty} r(\tau) \delta(t - \tau) d\tau = r(t)$$

die Impulsantwort r .

Die Faltung ist kommutativ

$$x \star y = y \star x \tag{A.6}$$

wegen

$$\begin{aligned}
 (x \star y)(t) &= \int_{-\infty}^{\infty} x(\tau) y(t - \tau) d\tau = \\
 &= \int_{-\infty}^{\infty} y(t - \tau) x(\tau) d\tau = \\
 &= \int_{\infty}^{-\infty} -y(\tau') x(t - \tau') d\tau' = \\
 &= \int_{-\infty}^{\infty} y(\tau') x(t - \tau') d\tau' = \\
 &= (y \star x)(t)
 \end{aligned}$$

mit $\tau' := t - \tau$.

Die Faltung ist assoziativ

$$x \star (y \star z) = (x \star y) \star z \quad (\text{A.7})$$

wegen

$$\begin{aligned}
 [x \star (y \star z)](t) &= \int_{-\infty}^{\infty} x(\tau) (y \star z)(t - \tau) d\tau = \\
 &= \int_{-\infty}^{\infty} x(\tau) \int_{-\infty}^{\infty} y(\tau') z(t - \tau - \tau') d\tau' d\tau = \\
 &= \int_{-\infty}^{\infty} x(\tau) \int_{-\infty}^{\infty} y(\tau'' - \tau) z(t - \tau'') d\tau'' d\tau = \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau) y(\tau'' - \tau) z(t - \tau'') d\tau d\tau'' = \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau) y(\tau'' - \tau) d\tau z(t - \tau'') d\tau'' = \\
 &= [(x \star y) \star z](t)
 \end{aligned}$$

mit $\tau'' = \tau' + \tau$.

Es gilt das Distributivgesetz

$$x \star (y + z) = x \star y + x \star z$$

und das gemischte Assoziativgesetz

$$c(x \star y) = (cx) \star y$$

für eine Konstante c .

Sei die Funktion

$$x^- : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto x^-(t) := x(-t)$$

die an der Hochwertachse gespiegelte Funktion x . Kovarianz und Faltung hängen zusammen über

$$\text{cov}(x, y) = x^- \star y \tag{A.8}$$

Der Zusammenhang (A.8) ergibt sich mit der Antisymmetrie der Kovarianz (A.3) und der Kommutativität der Faltung (A.6) aus

$$\begin{aligned} \text{cov}(x, y)(t) &= \text{cov}(y, x)(-t) = \\ &= \int_{-\infty}^{\infty} y(\tau) x(\tau - t) \, d\tau = \\ &= \int_{-\infty}^{\infty} y(\tau) x^-(t - \tau) \, d\tau = \\ &= (y \star x^-)(t) = \\ &= (x^- \star y)(t) \end{aligned}$$

Es gilt weiter

$$(x \star y)^- = x^- \star y^-,$$

denn

$$\begin{aligned} (x \star y)^-(t) &= (x \star y)(-t) = \\ &= \int_{-\infty}^{\infty} x(\tau) y(-t - \tau) \, d\tau = \\ &= \int_{-\infty}^{\infty} x^-(-\tau) y^-(t + \tau) \, d\tau = \\ &= - \int_{\infty}^{-\infty} x^-(\tau) y^-(t - \tau) \, d\tau = \\ &= (x^- \star y^-)(t). \end{aligned}$$

Außerdem gilt zwischen Kovarianz und Faltung das gemischte Assoziativgesetz

$$\text{cov}(x, y \star z) = \text{cov}(x, y) \star z \quad (\text{A.9})$$

und

$$\text{cov}(x \star y, z) = \text{cov}[x, \text{cov}(y, z)]. \quad (\text{A.10})$$

Das erste Gesetz (A.9) gilt aufgrund des Zusammenhangs zwischen Kovarianz und Faltung (A.8)

$$\text{cov}(x, y \star z) = x^- \star (y \star z),$$

woraus sich aufgrund der Assoziativität der Faltung (A.7)

$$\begin{aligned} \text{cov}(x, y \star z) &= (x^- \star y) \star z = \\ &= \text{cov}(x, y) \star z \end{aligned}$$

ergibt. Das zweite Gesetz (A.10) erhält man mit Hilfe der selben Zusammenhänge aus

$$\begin{aligned} \text{cov}(x \star y, z) &= (x \star y)^- \star z = \\ &= (x^- \star y^-) \star z = \\ &= x^- \star (y^- \star z) = \\ &= \text{cov}[x, \text{cov}(y, z)]. \end{aligned}$$

Insbesondere ist die Autokovarianz einer Faltung die Faltung der Autokovarianzen

$$\text{cov}(x \star y, x \star y) = \text{cov}(x, x) \star \text{cov}(y, y), \quad (\text{A.11})$$

denn

$$\begin{aligned} \text{cov}(x \star y, x \star y) &= (x^- \star y^-) \star (x \star y) = \\ &= (x^- \star x) \star (y^- \star y) = \\ &= \text{cov}(x, x) \star \text{cov}(y, y). \end{aligned}$$

Die Kovarianz von zwei Faltungen ist ebenfalls die Faltung entsprechender Kovarianzen

$$\text{cov}(x \star y, u \star v) = \text{cov}(x, u) \star \text{cov}(y, v),$$

denn

$$\begin{aligned} \text{cov}(x \star y, u \star v) &= (x^- \star y^-) \star (u \star v) = \\ &= (x^- \star u) \star (y^- \star v) = \\ &= \text{cov}(x, u) \star \text{cov}(y, v). \end{aligned}$$

Die Differentialgleichungen beschreiben die ganz allgemeinen Bedingungen für die Ausbreitung von Wärme und führen die physikalischen Fragestellungen auf reine Mathematik zurück, und genau das sollte eine gute Theorie leisten.

Jean Baptiste Joseph Fourier

B

Fouriertransformation

Die Fouriertransformierte einer reellen Funktion kann aufgefasst werden als die Entwicklung dieser Funktion nach den harmonischen Funktionen $\sin(\omega)$ und $\cos(\omega)$. Da die harmonischen Funktionen Eigenfunktionen linearer zeitinvarianter Systeme sind, ist die Antwort eines linearen zeitinvarianten Systems auf eine Harmonische wieder eine Harmonische mit i. d. R. verschiedener Amplitude und Phase (Unbehauen 1993). Kennt man also die Antwort eines linearen zeitinvarianten Systems auf die Harmonischen, so kann man aus der Fouriertransformierten der Eingabe die Fouriertransformierte der Systemantwort konstruieren.

Fouriertransformation, Faltung und Kovarianz hängen auf einfache Weise zusammen, so dass sich Gleichungen zwischen Faltungen und Kovarianzen oft als rationale Funktionen der Fouriertransformierten der enthaltenen Funktionen schreiben lassen. Dies macht es möglich, solche Gleichungen effizient zu lösen, denn die Fouriertransformation sowie die entsprechende Rücktransformation lassen sich in ihrer digitalen Form („digital fourier transform“ DFT) mit Hilfe von FFT-Algorithmen („fast fourier transform“) effizient numerisch ausführen (Press et al. 1995, Frigo und Johnson 1999). Auch die effiziente Berechnung von Faltungen ist mit Hilfe der Fouriertransformation möglich (Myers 1990, Nussbaumer 1990). Dieser Anhang fasst die hierzu benötigten Eigenschaften der Fouriertransformation zusammen.

B.1 Fouriertransformation und Rücktransformation

Die Fouriertransformierte $\mathcal{F}x$ einer Funktion $x : \mathbb{R} \rightarrow \mathbb{C}; t \mapsto x(t)$ von den reellen in die komplexen Zahlen ist definiert als (Unbehauen 1993)

$$\mathcal{F}x(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt. \quad (\text{B.1})$$

Die Rücktransformation einer Funktion $X : \mathbb{R} \rightarrow \mathbb{C}; \omega \mapsto X(\omega)$ ist definiert als¹

$$\mathcal{F}^{-1}X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega. \quad (\text{B.2})$$

¹Nach Weisstein (1999) wird die Fouriertransformation auch als

$$\mathcal{F}x(\omega) = \sqrt{\frac{|b|}{(2\pi)^{1-a}}} \int_{-\infty}^{\infty} x(t) e^{ib\omega t} dt$$

und die entsprechende Rücktransformation als

$$\mathcal{F}^{-1}X(t) = \sqrt{\frac{|b|}{(2\pi)^{1+a}}} \int_{-\infty}^{\infty} X(\omega) e^{-ib\omega t} d\omega$$

mit beliebigen reellen Konstanten a und b definiert. Die hier verwendete Wahl lautet $a = 1$ und $b = -1$. Es gilt folgende Tabelle (Char 1991a, Wolfram 1993, Weisstein 1999):

Fachgebiet	a	b
moderne Physik	0	1
Mathematik und Systemtheorie	1	-1
Wahrscheinlichkeitstheorie	1	1
klassische Physik	-1	1
Signalverarbeitung	0	-2π
Mathematica	0	1
Maple	1	-1

Für stetige², absolut integrierbare³ Funktionen $x : \mathbb{R} \rightarrow \mathbb{C}$ gelten die Sätze (Pinkus und Zafrany 1997)

$$\mathcal{F}^{-1}(\mathcal{F}x) = x \tag{B.3}$$

und

$$\mathcal{F}(\mathcal{F}^{-1}x) = x. \tag{B.4}$$

Da neuronale Antworten und rezeptive Felder reell sind, benötigen wir meist nur die Fouriertransformierten reeller Funktionen. Bezeichnen wir mit „*“ das komplex Konjugierte, dann gilt für reelle Funktionen $y : \mathbb{R} \rightarrow \mathbb{R}$

$$\mathcal{F}y(\omega) = \mathcal{F}y(-\omega)^*.$$

Sei $y^- : t \mapsto y^-(t) := y(-t)$ wieder die an der Hochwertachse gespiegelte Funktion y . Es gilt

$$\mathcal{F}(y^-)(\omega) = \mathcal{F}y^* \tag{B.5}$$

wegen

$$\begin{aligned} \mathcal{F}(y^-)(\omega) &= \int_{-\infty}^{\infty} y^-(t) e^{-i\omega t} dt = \\ &= \int_{-\infty}^{\infty} y(-t) e^{-i\omega t} dt = \\ &= - \int_{\infty}^{-\infty} y(t) e^{i\omega t} dt = \\ &= \left[\int_{-\infty}^{\infty} y(t) e^{-i\omega t} dt \right]^* = \\ &= \mathcal{F}y(\omega)^*. \end{aligned}$$

²Falls die Funktion x nur stückweise stetig ist, gilt (Pinkus und Zafrany 1997)

$$\mathcal{F}^{-1}(\mathcal{F}x)(t) = \begin{cases} x(t) & \text{falls } x \text{ stetig an der Stelle } t \text{ ist} \\ \frac{1}{2} [\lim_{t' \rightarrow t-} x(t') + \lim_{t' \rightarrow t+} x(t')] & \text{sonst.} \end{cases}$$

Für möglichst allgemeine Bedingungen, unter denen die Fourier-Rücktransformation die inverse Operation ist, also (B.3) gilt, siehe Titchmarsh (1948). Verwendet man verallgemeinerte Funktionen (vgl. auch Abschnitt B.6), so verschwinden diese Schwierigkeiten und es gilt (B.3) (Lighthill 1966).

³Eine Funktion $x : \mathbb{R} \rightarrow \mathbb{C}$ heißt „absolut integrierbar“, falls das Integral $\int_{-\infty}^{\infty} |x(t)| dt$ existiert.

B.2 Fouriertransformation und Faltung

Für stückweise stetige, absolut integrierbare Funktionen $x : \mathbb{R} \rightarrow \mathbb{C}$ und $y : \mathbb{R} \rightarrow \mathbb{C}$ gilt der Faltungssatz (Pinkus und Zafrany 1997)

$$\mathcal{F}(x \star y) = \mathcal{F}x \mathcal{F}y, \quad (\text{B.6})$$

d. h. die Fouriertransformierte der Faltung zweier Funktionen ist das Produkt der Fouriertransformierten der Funktionen.

B.3 Fouriertransformation und Kovarianz

Die Fouriertransformierte $\mathcal{F}\text{cov}(x, y)$ der Kovarianzfunktion $\text{cov}(x, y)$ zweier reeller Funktionen $x : \mathbb{R} \rightarrow \mathbb{R}$ und $y : \mathbb{R} \rightarrow \mathbb{R}$ ist (Unbehauen 1993)

$$\mathcal{F}\text{cov}(x, y) = (\mathcal{F}x)^* \mathcal{F}y, \quad (\text{B.7})$$

denn wegen dem Zusammenhang zwischen Kovarianz und Faltung (A.8), dem Faltungssatz (B.6) und der Fouriertransformierten der gespiegelten Funktion (B.5) gilt

$$\begin{aligned} \mathcal{F}\text{cov}(x, y)(\omega) &= \mathcal{F}(x^- \star y)(\omega) = \\ &= \mathcal{F}(x^-)(\omega) \mathcal{F}y(\omega) = \\ &= \mathcal{F}x(\omega)^* \mathcal{F}y(\omega). \end{aligned}$$

Insbesondere ist die Fouriertransformierte $\mathcal{F}\text{cov}(x, x)$ der Autokovarianzfunktion $\text{cov}(x, x)$ wegen (B.7) (Wiener-Khinchin-Theorem)

$$\mathcal{F}\text{cov}(x, x) = |\mathcal{F}x|^2. \quad (\text{B.8})$$

B.4 Fouriertransformation der Ableitung

Die Fouriertransformierte $\mathcal{F}x^{(n)}$ der n -ten Ableitung $x^{(n)}$ einer Funktion x ist

$$\mathcal{F}x^{(n)}(\omega) = (i\omega)^n \int \mathcal{F}x(\omega). \quad (\text{B.9})$$

Für $n = 0$ ist (B.9) offenbar richtig. Für $n \geq 1$ leitet man die Gleichung

$$x(t) = \frac{1}{2\pi} \int \mathcal{F}x(\omega) e^{i\omega t} d\omega,$$

die sich aus der Fourier-Rücktransformation (B.2) und (B.3) ergibt, n -mal nach der Variablen t ab und erhält

$$x^{(n)}(t) = \frac{1}{2\pi} (i\omega)^n \mathcal{F}x(\omega) e^{i\omega t} d\omega.$$

Fouriertransformation dieser Gleichung ergibt wegen (B.4) das gewünschte Ergebnis (B.9).

B.5 Fouriertransformation und lineare Systeme

Lineare Systeme mit Systemantwort y und Eingabe x lassen sich nach (A.5) durch die Faltung

$$y = r \star x \tag{B.10}$$

beschreiben, wobei r „Impulsantwort“ des Systems heißt. Wir bezeichnen mit Großbuchstaben Y , R und X hier, wie auch im Folgenden, die Fouriertransformierten $\mathcal{F}x$, $\mathcal{F}r$ bzw. $\mathcal{F}x$ der entsprechenden Funktionen x , r bzw. y . Fouriertransformation von (B.10) ergibt nach dem Faltungssatz (B.6)

$$Y(\omega) = R(\omega) X(\omega), \tag{B.11}$$

Die Fouriertransformierte R der Impulsantwort r heißt „Übertragungsfunktion“ des linearen Systems (Unbehauen 1993). Kennt man die Übertragungsfunktion, so lässt sich die Systemantwort nach (B.11) mit Hilfe von Fouriertransformation (B.1) der Eingabe und Rücktransformation (B.2,B.3) des Produkts aus derselben mit der Fouriertransformierten der Übertragungsfunktion numerisch effizient⁴ berechnen.

B.6 Fouriertransformation und Delta-Funktion

In der linearen Systemtheorie benötigt man oft dasjenige lineare System, das die Eingabe x unverändert lässt, d. h. für dessen Ausgabe y gilt $y = x$. Die Impulsantwort dieses Systems sei δ , d. h. nach (B.10) gilt

$$x = \delta \star x. \tag{B.12}$$

⁴Eine naive Implementation der Fouriertransformation benötigte bei Berechnung von n Zeitschritten n Faltungen, jede Faltung wiederum n Rechenschritte, also insgesamt n^2 Schritte. Der FFT-Algorithmus (*fast fourier transform*) kommt mit einer Zeitkomplexität $n \log(n)$ aus (Press et al. 1995).

Es gibt aber keine Funktion $\delta : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto \delta(t)$, welche Gleichung (B.12) für beliebige Funktionen x erfüllt. Daher verallgemeinert man den Funktionsbegriff auf „verallgemeinerte Funktionen“ oder „Distributionen“ (Lighthill 1966, Unbehauen 1993). Wir unterscheiden hier nicht und sprechen einfach von δ -Funktion, obwohl es sich um eine verallgemeinerte Funktion handelt.⁵

⁵ Für eine mathematisch exakte Behandlung verallgemeinerter Funktionen sei auf Lighthill (1966) verwiesen. Eine verallgemeinerte Funktion f ist hier die Menge $\{f_n\}$ aller Folgen f_n , $n \in \mathbb{N}$, von „Grundfunktionen“, so dass

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(t) g(t) dt$$

für alle Folgen f_n aus $\{f_n\}$ und alle Grundfunktionen g existiert und unabhängig vom Element f_n das gleiche Ergebnis liefert. Diese Folgen heißen zu f_n äquivalent. Grundfunktionen $g(t)$ müssen für $|t| \rightarrow \infty$ sehr schnell gegen null abfallen. Für eine genaue Definition siehe Lighthill (1966). Da Messungen und Eingaben für reale Systeme i. d. R. endlich sind, d. h. $g(t) = 0$ für hinreichend großes $|t|$, entstehen aus der Einschränkung auf Grundfunktionen i. d. R. keine Probleme.

Das Integral über ein Produkt einer verallgemeinerten Funktion f mit einer Grundfunktion g ist definiert als

$$\int_{-\infty}^{\infty} f(t) g(t) dt = \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(t) g(t) dt.$$

Das Integral ist wohldefiniert, da nach der Definition der verallgemeinerten Funktion $f = \{f_n\}$ das Ergebnis unabhängig vom Repräsentanten f_n das gleiche ist.

Die zu $t \mapsto \delta_n(t) := \sqrt{n/\pi} e^{-nt^2}$ äquivalenten Folgen definieren eine verallgemeinerte Funktion δ mit

$$\int_{-\infty}^{\infty} \delta(t) g(t) dt = g(0)$$

für alle Grundfunktionen g .

Unter Notationsmissbrauch wird $\delta(at - b)$ mit reellen Konstanten a und b als diejenige verallgemeinerte Funktion definiert, für welche die Folge $t \mapsto \delta_n(at - b)$ ein Repräsentant ist. Somit bedeutet die symbolische Schreibweise

$$\int_{-\infty}^{\infty} \delta(at - b) g(t) dt := \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \delta_n(at - b) g(t) dt = \frac{1}{|a|} g\left(\frac{b}{a}\right).$$

Auf ähnliche Weise können die Summe von verallgemeinerten Funktionen, das Produkt einer Funktion mit einer verallgemeinerten Funktion, die Ableitung sowie die Fouriertransformierte einer verallgemeinerten Funktion definiert werden. Das Produkt zweier verallgemeinerter Funktionen ist nach Lighthill (1966) nicht befriedigend definierbar. Einen Überblick über verschiedene Verfahren, Distributionen zu definieren und deren Vor- und Nachteile gibt Temple (1953).

Wenn der Faltungssatz (B.6) mit verallgemeinerten Funktionen weiter gelten soll, erhält man durch Fouriertransformation aus (B.12)

$$X(\omega) = \mathcal{F}\delta(\omega) X(\omega).$$

Wählt man eine Funktion x mit $X(\omega) \neq 0$ für alle Frequenzen ω , sieht man, dass

$$\mathcal{F}\delta = 1$$

gelten muss. Fouriertransformationen verallgemeinerter Funktionen sind dazu entsprechend zu definieren (Lighthill 1966).

Auch das System ist linear, das die n -te Ableitung der Eingabe x als Ausgabe y zurückliefert, also

$$y = x^{(n)},$$

denn die Ableitung ist ein lineares Funktional. Also sollte sich dieses System nach (B.10) durch

$$x^{(n)} = (-1)^n \delta^{(n)} \star x \tag{B.13}$$

mit Impulsantwort $(-1)^n \delta^{(n)}$ beschreiben lassen. Wiederum gibt es jedoch keine Funktion $\delta^{(n)} : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto \delta^{(n)}(t)$, welche Gleichung (B.13) für beliebige Funktionen x erfüllt. d. h. auch $\delta^{(n)}$ ist eine entsprechend definierte⁶ verallgemeinerte Funktion.

Fouriertransformation von (B.13) ergibt nach dem Faltungssatz (B.6) und der Fouriertransformation der Ableitung (B.9)

$$(i\omega)^n X(\omega) = (-1)^n \mathcal{F}\delta^{(n)}(\omega) X(\omega).$$

Wählt man wieder ein x mit Fouriertransformierter $X(\omega) \neq 0$ für alle Frequenzen ω , so erhält man für die Fouriertransformation $\mathcal{F}\delta^{(n)}$ der n -ten Ableitung der δ -Funktion

$$\mathcal{F}\delta^{(n)}(\omega) = (-i\omega)^n.$$

⁶Nach Lighthill (1966) ist die Ableitung f' einer verallgemeinerten Funktion $f = \{f_k\}$ die Menge aller zu einer Repräsentantenfolge $f'_k, k \in \mathbb{N}$, äquivalenter Folgen. Dabei bezeichnet f'_k die Ableitung der gewöhnlichen Funktion f_k . Die Ableitung einer verallgemeinerten Funktion ist natürlich wieder eine verallgemeinerte Funktion nach Definition, vgl. Fußnote 5 auf Seite 242. Die n -te Ableitung der δ -Funktion ist somit die Menge $\delta^{(n)} = \{\delta_k^{(n)}\}$ der zu $\delta_k^{(n)}, k \in \mathbb{N}$, äquivalenten Folgen.

Literaturverzeichnis

- Aertsen, A. M. und I., J. P. (1981a). A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biol. Cybern.*, **42**(2):145–156.
- Aertsen, A. M. und I., J. P. (1981b). The spectro-temporal receptive field. a functional characteristic of auditory neurons. *Biol Cybern*, **42**(2):133–143.
- Aertsen, A. M., Olders, J. H. und I., J. P. (1980). Spectro-temporal receptive fields of auditory neurons in the grassfrog. iii. analysis of the stimulus-event relation for natural stimuli. *Biol. Cybern.*, **39**(3):195–209.
- Allen, J. B. (1976). Two-dimensional cochlear fluid model: New results. *J. Acoust. Soc. Am.*, **61**(1):110–119.
- Altman, J. S. und Dawes, E. A. (1983). A cobalt study of medullary sensory projections from lateral line nerves, associated cutaneous nerves, and the VIIIth nerve in adult *Xenopus*. *J. Comp. Neurol.*, **213**:310–326.
- Andrä, W. (Hg.) (1998). *Magnetism in medicine*. Wiley-VCH, Berlin.
- Ariav, G., Polsky, A. und Schiller, J. (2003). Submillisecond precision of the input-output transformation function mediated by fast sodium dendritic spikes in basal dendrites of ca1 pyramidal neurons. *J. Neurosci.*, **23**(21):7750–7758.
- Benardete, E. A. und Victor, J. D. (1994). An extension of the m-sequence technique for the analysis of multi-input nonlinear systems. In *Advanced Methods of Physiological Systems Modeling, Volume III* (Hg. V. Z. Maremarelis), S. 87–110. Plenum, New York.
- Berry, M. J. und Meister, M. (1998). Refractoriness and neural precision. *J. Neurosci.*, **18**(6):2200–2211.
- Berry, M. J., Warland, D. K. und Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. U. S. A.*, **94**(10):5411–5416.

- Betsch, B. Y., Einhauser, W., Kording, K. P. und Konig, P. (2004). The world from a cat's perspective – statistics of natural videos. *Biol. Cybern.*, **90**(1):41–50.
- Beutelspacher, A. (1999). *Das ist o.B.d.A. trivial! - Tips und Tricks zur Formulierung mathematischer Gedanken*. Vieweg, Braunschweig.
- Blauert, J. (1999). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge.
- Blossfeld, H.-P., Hamerle, A. und Mayer, K. U. (1986). *Ereignisanalyse - Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften*. Campus Verlag, Frankfurt/Main.
- Blossfeld, H.-P. und Rohwer, G. (1995). *Techniques of Event History Modeling - New Approaches to Causal Analysis*. Lawrence Erlbaum Associates, Mahwah.
- Bonhoeffer, T. und Grinvald, A. (1996). Optical imaging based on intrinsic signals: The methodology. In *Brain Mapping - The Methods* (Hg. A. W. Toga und J. C. Mazziotta), S. 55–97. Academic Press.
- Bower, J. M. und Beeman, D. (1994). *The Book of GENESIS: Exploring Realistic Neural Models with the General Neural Simulation System*. Springer, New York.
- Breuer, H. (1987). *dtv-Atlas zur Physik: Mechanik, Akustik, Thermodynamik*, Band 1. Deutscher Taschenbuch Verlag.
- Breyman, U. (1999). *C++: Eine Einführung*. Hanser.
- Brivanlou, I. H., Warland, D. K. und Meister, M. (1998). Mechanisms of concerted firing among retinal ganglion cells. *Neuron*, **20**(3):527–539.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**(2):121 – 167.
- Buser, P. und Imbert, M. (1992). *Vision*. MIT Press.
- C++ Standard Committee (1997). November 1997 draft C++ standard. Technischer Bericht.
- Cai, D., DeAngelis, G. C. und Freeman, R. D. (1997). Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *Journal of Neurophysiology*, **78**:1045–1061.

- Carr, C. E. und Boudreau, R. E. (1991). Central projections of auditory nerve fibers in the barn owl. *J. Comp. Neurol.*, **314**(2):306–318.
- Cattermole, K. W. (1988). *Signale und Wellen - mathematische Grundlagen und technische Anwendungen Signale und Wellen*. VCH, Weinheim.
- Cave, C. und Risset, J.-C. (2005). Musikalische Illusionen – Vom Ohr an der Nase herumgeführt. *Gehirn & Geist*, **3**:38–44.
- Chan, T. (1997). *UNIX system programming using C++*. Prentice Hall, Upper Saddle River.
- Char, B. W. (Hg.) (1991a). *Maple V Language Reference Manual*. Springer, New York.
- Char, B. W. (Hg.) (1991b). *Maple V Library Reference Manual*. Springer, New York.
- Char, B. W. (Hg.) (1993). *First Leaves: A Tutorial Introduction to Maple V*. Springer, New York.
- Cooke, M. (1993). *Modelling Auditory Processing and Organisation*. Dissertation, Cambridge University Press, Cambridge.
- Cucker, F. und Smale, S. (2002). On the mathematical foundations of learning. *Bull. AMS*, **34**(1):1–49.
- de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R. und Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, **275**(5307):1805–1808.
- deCharms, R. C., Blake, D. T. und Merzenich, M. M. (1998). Optimizing sound features for cortical neurons. *Science*, **280**:1439–1443.
- deCharms, R. C., Blake, D. T. und Merzenich, M. M. (1999). A multielectrode implant device for the cerebral cortex. *J. Neurosci. Methods*, **93**(1):27–35.
- Demoment, G. (1989). Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **37**(12):2024 – 2036.
- Depireux, D. A., Simon, J. Z., Klein, D. J. und Shamma, S. A. (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol.*, **85**(3):1220–1234.

- Deutsch, D. (1995). *Musical Illusions and Paradoxes, Hör-CD*. Philomel.
- Donnelly, C. und Stallman, R. (2002). *Bison: The YACC-compatible Parser Generator*. Free Software Foundation, Boston.
- Downes, M. (2002). Tex and latex2e. *Notices of the AMS*, **42**(11):1384–1391.
- Duda, R. O., Stork, D. G. und Hart, P. E. (2000). *Pattern Classification and Scene Analysis*. Wiley, New York, 2. Auflage.
- Eggermont, J. J. (1993). Wiener and volterra analyses applied to the auditory system. *Hear Res*, **66**(2):177–201.
- Einhauser, W., Kayser, C., Kording, K. P. und König, P. (2003). Learning distinct and complementary feature selectivities from natural colour videos. *Rev. Neurosci.*, **14**(1-2):43–52.
- Elephant, A. (1996). Sensory perception and the lateral line system in the clawed frog, *Xenopus*. In *The Biology of Xenopus* (Hg. R. C. Tinsley und H. R. Kobel), Kapitel 7, S. 97–120. Clarendon Press, Oxford.
- Fastl, H., Patsouras, D., Franosch, J.-M. P. und van Hemmen, J. L. (2001). Zwicker tones for pure tone plus band limited noise. In *Physiological and Psychophysical Bases of Auditory Function, Proc. of the 12th International Symposium on Hearing* (Hg. D. B. et al.), S. 67–74. Shaker, Maastricht.
- Faulstich, M. und Kössl, M. (2000). Evidence for multiple dpoae components based upon group delay of the 2f1-f2 distortion in the gerbil. *Hearing Research*, **140**(1-2):99–110.
- Fenlason, J. (2006). GNU gprof – profiling a program: Where does it spend its time? In *Documentation for binutils 2.17*. GNU Software Foundation.
- Finch, S. (2004). Ornstein-uhlenbeck process.
- Franosch, J.-M. P. (1998). *Der Zwicker-Ton - Ein Phänomen der auditorischen Informationsverarbeitung*. Diplomarbeit, TU München, München.
- Franosch, J.-M. P., Kempster, R. und van Hemmen, J. L. (2003a). Zwicker tone illusion and noise reduction in the auditory system. *Physical Review Letters*, **90**:178103.
- Franosch, J.-M. P., Lingenheil, M. und van Hemmen, J. L. (2005a). How a frog can learn what is where in the dark. *Phys. Rev. Lett.*, **95**(7):078106.

- Franosch, J.-M. P., Sichert, A. B., Suttner, M. D. und van Hemmen, J. L. (2005b). Estimating position and velocity of a submerged moving object by the clawed frog *Xenopus* and by fish — a cybernetic approach. *Biol. Cybern.*, **93**(4):231–238.
- Franosch, J.-M. P., Sobotka, M. C., Elepfandt, A. und van Hemmen, J. L. (2003b). Minimal model of prey localization through the lateral-line system. *Phys. Rev. Lett.*, **91**:1581011–1581014.
- Franosch, J.-M. P. und van Hemmen, J. L. (1999). Lateral inhibition enhances the detectability of a pure tone in the presence of background noise. In *Proceedings of the 1st Göttingen Conference of the German Neuroscience Society 1999* (Hg. N. Elsner und U. Eysel), Band 1, S. 280. Thieme, Stuttgart.
- Free Software Foundation (2002). *The GNU Bash Reference Manual*.
- Frigo, M. und Johnson, S. G. (1999). *FFTW User's Manual*. Massachusetts Institute of Technology.
- Fuhrmann, G., Segev, I., Markram, H. und Tsodyks, M. (2002). Coding of temporal information by activity-dependent synapses. *J. Neurophysiol.*, **87**(1):140–148.
- Gamma, E., Helm, R. und Johnson, R. (1997). *Design Patterns - Elements of Reusable Object-Oriented Software*. Addison-Wesley, Amsterdam.
- Gerstner, W. (1992). *Kodierung und Signalübertragung in neuronalen Systemen*. Dissertation, TU München, München.
- Gerstner, W. und van Hemmen, J. L. (1994). Coding and information processing in neural networks. In *Models of Neural Networks II* (Hg. E. Domany, J. L. van Hemmen und K. Schulten), Kapitel 1, S. 39–47. Springer, New York.
- Glass, G. und Schuchert, B. L. a. (1996). *The STL <Primer>*. Prentice Hall.
- Grace Team (2000). *Grace User's Guide*.
- Greene, W. H. (2000). *Econometric analysis*. Prentice Hall, Upper Saddle River.
- Grimsehl, E. (1987). *Lehrbuch der Physik: Mechanik*, Band 1. Teubner.

- Haken, H. (1990). *Synergetik: Eine Einführung, Nichtgleichgewichts-Phasenübergänge und Selbstorganisation in Physik, Chemie und Biologie*. Springer, Berlin.
- Hauske, G. (1994). *Systemtheorie der visuellen Wahrnehmung*. Teubner, Stuttgart.
- Holmes, M. H. und Rubenfeld, L. A. (Hg.) (1981). *Mathematical Modeling of the Hearing Process: Proceedings of the NSF-CBMS Regional Conference Held in Troy, NY*. Springer.
- Hopcroft, J. E. und Ullman, J. D. (1992). *Einführung in die Automatentheorie, formale Sprachen und Komplexitätstheorie*. Addison-Wesley, Bonn.
- IBM (2001). *IBM Visualization Data Explorer Complete Documentation: Version 3 Release 1 Modification 4*. International Business Machines Corporation.
- J., K., P., R., C., R. R. und Meister, M. (2001). Predicting every spike: a model for the responses of visual neurons. *Neuron*, **30**(3):803–817.
- Joachims, T. (1999). Making large-scale svm learning practical. In *Advances in Kernel Methods - Support Vector Learning* (Hg. B. Schölkopf, C. Burges und A. Smola). MIT Press.
- Josuttis, N. (1997). *Die C++-Standardbibliothek, Eine detaillierte Einführung in die vollständige ANSI/ISO-Bibliothek*. Addison-Wesley, Bonn.
- Kandel, E. R., Schwartz, J. H. und M., J. T. (Hg.) (1996). *Neurowissenschaften: Eine Einführung*. Spektrum Akad. Verl., Heidelberg.
- Kastner, S., De Weerd, P., Desimone, R. und Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. *Science*, **282**:108–111.
- Kecman, V. (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. MIT Press, Cambridge.
- Keidel, W. D. (1975). *Physiologie des Gehörs: Akustische Informationsverarbeitung, Einführung für Ärzte, Biologen, Psychologen und Nachrichtentechniker*. Thieme.

- Kempster, R., Gerstner, W. und van Hemmen, J. L. (1998a). How the threshold of a neuron determines its capacity for coincidence detection. *Biosystems*, **48**(1-3):105–112.
- Kempster, R., Gerstner, W., van Hemmen, J. L. und Wagner, H. (1998b). Extracting oscillations. Neuronal coincidence detection with noisy periodic spike input. *Neural Comput.*, **10**(8):1987–2017.
- Kempster, R., Leibold, C., Wagner, H. und van Hemmen, J. L. (2001). Formation of temporal-feature maps by axonal propagation of synaptic learning. *Proc. Natl. Acad. Sci. USA*, **98**(7):4166–4171.
- Kempster, Richard Kempster, R. (1997). *Hebbsches Lernen zeitlicher Codierung: Theorie der Schallortung im Hörsystem der Schleiereule*. Dissertation, TU München, München.
- Kistler, W. M., Gerstner, W. und van Hemmen, J. L. (1997). Reduction of the hodgkin-huxley equations to a single-variable threshold model. *Neural Computation*, **9**(5):1015–1045.
- Klein, D. J., Depireux, D. A., Simon, J. Z. und Shamma, S. A. (2000). Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J. Comput. Neurosci.*, **9**(1):85–111.
- Klein, D. J., Simon, J. Z., Depireux, D. A. und Shamma, S. A. (2006). Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *J. Comput. Neurosci.*, **20**(2):111–136.
- Knudsen, E. I. und Brainard, M. S. (1995). Creating a unified representation of visual and auditory space in the brain. *Annu. Rev. Neurosci.*, **18**:19–43.
- Knuth, D. E. (1993). *The TeXbook*. Addison-Wesley.
- Kopka, H. (1994). *LaTeX Band 1: Einführung*. Addison-Wesley, Bonn.
- Kopka, H. (1995). *LaTeX Band 2: Ergänzungen – mit einer Einführung in METAFONT*. Addison-Wesley, Bonn.
- Kopka, H. (1997). *LaTeX Band 3: Erweiterungen*. Addison-Wesley, Bonn.
- Krantz, S. G. (1997). *A Primer of Mathematical Writing*. American Mathematical Society.
- Kuchling, H. (1988). *Taschenbuch der Physik*. Harri Deutsch.

- Kössl, M. (1997). Sound emission from cochlear filters and foveae - does the auditory sense organ make sense? *Naturwissenschaften*.
- Legg, S. und Hutter, M. (2006). A formal measure of machine intelligence. ArXiv:cs.AI/0605024 v1 6 May 2006.
- Leibold, C., Kempner, R. und van Hemmen, J. L. (2001). Temporal map formation in the barn owl's brain. *Phys. Rev. Lett.*, **87**(24):248101.
- Lewicki, M. S. (1994). Bayesian modeling and classification of neural signals. *Neural Computation*, **6**:1005–1030.
- Li, M. und Vitanyi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. Springer, New York.
- Lighthill, M. J. (1966). *Einführung in die Theorie der Fourieranalysis und der verallgemeinerten Funktionen*. Hochschultaschenbücher-Verlag, Mannheim.
- Lockhart, T. (2000). *PostgreSQL User's Guide*. PostgreSQL Inc.
- Loukides, M. und Oram, A. (1997). *Programming with GNU software*. O'Reilly, Cambridge.
- Lowe, D. A. (1987). Single-unit study of lateral line cells in the optic tectum of *Xenopus laevis*: Evidence for bimodal lateral line/optic units. *J. Comp. Neurol.*, **257**(3):396–404.
- Lowe, R. und Russell, I. (1982). The central projections of lateral line and cutaneous sensory fibres (VII and X) in *Xenopus laevis*. *Proc. R. Soc. Lond.*, **216**:279–297.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
- MacKenzie, D. und Elliston, B. (2001). *Autoconf: Creating Automatic Configuration Scripts*. Free Software Foundation, Boston.
- MacKenzie, D. und Tromey, T. (2002). *GNU Automake*. Free Software Foundation, Boston.
- Mammano, F. und Nobili, R. (1993). Biophysics of the cochlea: Linear approximation. *J. Acoust. Soc. Am.*, **93**(6):3320–3332.
- Marmarelis, V. Z. (1978). *Analysis of physiological systems, the white-noise approach*. Plenum Pr.

- Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. *Journal of the Acoustical Society of America*, **79**(3):702–711.
- Meddis, R. (1988). Simulation of auditory-neural transduction: Further studies. *Journal of the Acoustical Society of America*, **83**(3):1056–1063.
- Meyers, S. (1999a). *Effektiv C++ programmieren: 50 Wege zur Verbesserung Ihrer Programme und Entwürfe*. Addison-Wesley, Bonn.
- Meyers, S. (1999b). *Mehr effektiv C++ programmieren: 35 neue Wege zur Verbesserung Ihrer Programme und Entwürfe*. Addison-Wesley, Bonn.
- Meyers, S. (2001). *Effective STL: 50 specific ways to improve your use of the Standard Template Library*. Addison-Wesley, Boston.
- Miller, G. L. und Knudsen, E. I. (1999). Early visual experience shapes the representation of auditory space in the forebrain gaze fields of the barn owl. *J. Neurosci.*, **19**(6):2326–2336.
- Moore, B. C. J. (1997). *An introduction to the psychology of hearing*. Acad. Press.
- Musiek, F. E., Shinn, J. B., Jirsa, R., Bamiaou, D. E., Baran, J. A. und Zaida, E. (2005). Gin (gaps-in-noise) test performance in subjects with confirmed central auditory nervous system involvement. *Ear Hear*, **26**(6):608–618.
- Myers, D. G. (1990). *Digital signal processing: Efficient convolution and Fourier transform techniques*. Prentice Hall, New York.
- MySQL AB (2001). *MySQL Reference Manual*. MySQL AB.
- Nelken, I., Kim, P. J. und Young, E. D. (1997). Linear and nonlinear spectral integration in type iv neurons of the dorsal cochlear nucleus. ii. predicting responses with the use of nonlinear models. *Journal of Neurophysiology*, **78**:800–811.
- Nelken, I. und Young, E. D. (1997). Linear and nonlinear spectral integration in type iv neurons of the dorsal cochlear nucleus. i. regions of linear interaction. *Journal of Neurophysiology*, **78**:790–799.
- Nieder, A. und Klump, G. M. (1999). Adjustable frequency selectivity of auditory forebrain neurons recorded in a freely moving songbird via radiotelemetry. *Hear. Res.*, **127**(1-2):41–54.

- Nieder, A. und Klump, G. M. (2001). Signal detection in amplitude-modulated maskers. II. Processing in the songbird's auditory forebrain. *European Journal of Neuroscience*, **13**(5):1033–1044.
- Nobili, R., Mammano, F. und Ashmore, J. (1998). How well do we understand the cochlea? *TINS*, **21**(4):159–167.
- Nolden, R. (2001). *The User Manual to KDevelop*.
- Nussbaumer, H. J. (1990). *Fast Fourier Transform and Convolution Algorithms*. Springer, Berlin.
- Ohl, F. W. und Scheich, H. (1997). Learning-induced dynamic receptive field changes in primary auditory cortex of the unanaesthetized mongolian gerbil. *J. Comp. Physiol. A*, **181**(6):685–96.
- Ohl, F. W., Scheich, H. und Freeman, W. J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. *Nature*, **412**(6848):733–736.
- O'Mard, L. P., Hewitt, M. J. und Meddis, R. (1997). *LUTEar: Core Routines Library, for Flexible, Auditory Computer Simulation Development*. Hearing Research Laboratory, University of Essex.
<http://www.essex.ac.uk/psychology/hearinglab/lutear/home.html>.
- Patterson, R. D., Allerhand, M. H. und Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *Journal of the Acoustical Society of America*, **98**(4):1890–1894.
- Pickles, J. O. (1991). *An introduction to the physiology of hearing*. Acad. Press, London.
- Pierce, R. D. (1996). *The GNU Scientific Software Library*. Department of Mathematics, The Pennsylvania State University, University Park, PA 16802.
- Pinkus, A. und Zafrany, S. (1997). *Fourier Series and Integral Transforms*. Cambridge University Press, Cambridge.
- Platt, J. C. (1999). Fast training of support vector machines using sequential minimal optimization. In *Advances in kernel methods: support vector learning*, S. 185–208. MIT Press, Cambridge.

- Popper, A. N. und Fay, R. R. (Hg.) (1992). *The Mammalian Auditory Pathway: Neurophysiology*. Springer, New York.
- Posner, M. I. und Raichle, M. E. (1996). *Bilder des Geistes*. Spektrum Akad. Verl., Heidelberg.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. und Flannery, B. P. (1995). *Numerical Recipes in C*. Cambridge University Press.
- Profos, P. und Pfeifer, T. (Hg.) (1994). *Handbuch der industriellen Messtechnik*. Oldenbourg.
- Raab, B. (2001). *Diskriminierung von Hörobjekten*. Diplomarbeit, TU München.
- Richard, J. P., Leppelsack, H. J. und Hausberger, M. (1995). A rapid correlation method for the analysis of spectro-temporal receptive fields of auditory neurons. *J Neurosci Methods*, **61**(1-2):99–103.
- Rissanen, J. (2005). An introduction to the MDL principle.
- Robb, R. A. (1995). *Three-dimensional biomedical imaging*. VCH, New York.
- Rojas, R. (1996). *Theorie der neuronalen Netze: Eine systematische Einführung*. Springer, Berlin.
- Russell, I. J. (1976). Amphibian lateral line receptors. In *Frog Neurobiology: A Handbook* (Hg. R. Llinas und W. Precht), S. 513–550. Springer, Berlin.
- Rüger, B. (1996). *Induktive Statistik: Einführung für Wirtschafts- und Sozialwissenschaftler*. Oldenbourg, München.
- Sato, T. und Smith, B. V. (2000). *Xfig User Manual Version 3.2.3c*.
- Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York.
- Schlittgen, R. und Streitberg, B. (1999). *Zeitreihenanalyse*. Oldenbourg, München.
- Schölkopf, B. und Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.
- Shamma, S. A., Depireux, D. A. und Kowalski, N. (1996a). Analysis of dynamic spectra in ferret primary auditory cortex. i. characteristics of single-unit responses to moving ripple spectra. *Journal of Neurophysiology*, **76**(5):3503–3523.

- Shamma, S. A., Depireux, D. A. und Kowalski, N. (1996b). Analysis of dynamic spectra in ferret primary auditory cortex. ii. prediction of unit responses to arbitrary dynamic spectra. *Journal of Neurophysiology*, **76**(5):3524–3534.
- Shepherd, G. M. (1993). *Neurobiologie*. Springer, Berlin.
- Shoham, D., Glaser, D. E., Arieli, A., Tenet, T., Wijnbergen, C., Toledo, Y., Hildesheim, R. und Grinvald, A. (1999). Imaging cortical dynamics at high spacial and temporal resolution with novel blue voltage-sensitive dyes. *Neuron*, **24**:791–802.
- Singh, S. (2002). *Geheime Botschaften, Die Kunst der Verschlüsselung von der Antike bis in die Zeiten des Internet*. dtv.
- Smale, S. (1998). Mathematical problems for the next century. *Mathematical Intelligencer*, **20**(2):7–15.
- Sobotka, M. C. (2002). *Ein mathematisches Modell zur Beutelokalisation des Krallenfrosches Xenopus laevis*. Diplomarbeit, TU München.
- Sottek, R. (1993). *Modelle zur Signalverarbeitung im menschlichen Gehör*. Dissertation, Technische Hochschule Aachen.
- Stallman, R., Pesch, R. und Shebs, S. (2001). *Debugging with GDB*. Free Software Foundation, Boston.
- Stallman, R. M. (2001). *Using and Porting the GNU Compiler Collection*. Free Software Foundation, Boston.
- Stallman, R. M. und Goyal, R. (1994). *Getting Started with XEmacs*.
- Stallman, R. M., Inc., L. und Wing, B. (1995). *XEmacs User's Manual*.
- Stallman, R. M. und McGrath, R. (2000). *GNU Make: A Program for Directing Recompilation*. Free Software Foundation, Boston.
- Stallman, R. M. und Support, C. (1995). *Debugging with GDB*. Free Software Foundation.
- Steenrod, N. E., Halmos, P. R., M., S. M. und Dieudonne, J. R. (1973). *How to write Mathematics*. American Mathematical Society.
- Stewart, I. (2002). Wer wird millionär? *Spektrum der Wissenschaft*, **5**:114–115.

- Stoer, J. und Bulirsch, R. (1994). *Einführung in die numerische Mathematik: Band 1*. Springer, Berlin.
- Stroustrup, B. (1998). *Die C++-Programmiersprache*. Addison-Wesley, Bonn.
- Sun Microsystems (2002a). *Sun ONE Grid Engine Administration and User's Guide*. Sun Microsystems, Santa Clara.
- Sun Microsystems (2002b). *Sun ONE Grid Engine and Sun ONE Grid Engine, Enterprise Edition Reference Manual*. Sun Microsystems, Santa Clara.
- Sun Microsystems (2002c). *Sun ONE Grid Engine, Enterprise Edition Administration and User's Guide*. Sun Microsystems, Santa Clara.
- Szucs, A., Vehovszky, A., Molnar, G., Pinto, R. D. und Abarbanel, H. D. (2004). Reliability and precision of neural spike timing: simulation of spectrally broadband synaptic inputs. *Neuroscience*, **126**(4):1063–1073.
- Temple, G. (1953). Theories and applications of generalized functions. *J. Lond. Math. Soc.*, **28**:134–148.
- Terhardt, E. (1998). *Akustische Kommunikation: Grundlagen mit Hörbeispielen*. Springer, Berlin.
- Theunissen, F. E., Sen, K. und Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *The Journal of Neuroscience*, **20**(6):2315–2331.
- Titchmarsh, E. C. (1948). *Introduction to the Theory of Fourier Integrals*. Oxford University Press, Oxford.
- Tsodyks, M., Pawelzik, K. und Markram, H. (1998). Neural networks with dynamic synapses. *Neural Comput.*, **10**(4):821–835.
- Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology: 1. Linear Cable Theory and Dendritic Structure*. Cambridge Univ. Pr., Cambridge.
- Unbehauen, R. (1993). *Systemtheorie*. Oldenbourg, München.
- Uzzell, V. J. und Chichilnisky, E. J. (2004). Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol.*, **92**(2):780–789.
- van Hemmen, L. J., Kistler, W. M. und Thomas, E. G. F. (2000). Calculation of volterra kernels for solutions of nonlinear differential equations. *SIAM Journal on Applied Mathematics*, **61**(1):1–21.

- Vapnik, V. N. (1998). *The Nature of Statistical Learning Theory*. Springer, New York.
- Vidyasagar, M. (1997). *A Theory of Learning and Generalization: With Applications to Neural Networks and Control Systems*. Springer, London.
- von Campenhausen, C. (1981). *Die Sinne des Menschen/1, 1. Einführung in die Psychophysik der Wahrnehmung*. Thieme, Stuttgart.
- Wall, L., Christiansen, T. und Schwartz, R. L. (1998). *Programmieren mit Perl*. O'Reilly, Köln.
- Weisstein, E. W. (1999). *CRC Concise Encyclopedia of Mathematics*. Chapman & Hall/CRC, Boca Raton.
- Wilden, A., Glasauer, S., Kleine, J. F. und Büttner, U. (2002). Modelling transfer characteristics of vestibular neurons in the fastigial nucleus of the behaving monkey on the basis of canal-otolith interaction. *Neuroreport*, **13**(6):799–804.
- Will, U., Luhede, G. und Görner, P. (1985a). The area octavo-lateralis in *Xenopus laevis*, I. The primary afferent projections. *Cell Tissue Res.*, **239**:147–161.
- Will, U., Luhede, G. und Görner, P. (1985b). The area octavo-lateralis in *Xenopus laevis*, II. Second order projections and cytoarchitecture. *Cell Tissue Res.*, **239**:163–175.
- Wittgenstein, L. (1966). *Tractatus logico-philosophicus: Logisch-philosophische Abhandlung*. edition suhrkamp.
- Wolfram, S. (1993). *Mathematica, A system for Doing Mathematics by Computer*. Addison-Wesley.
- Yost, W. A. (1994). *Fundamentals of Hearing: An Introduction*. Academic Press, San Diego, 3. Auflage.
- Zell, A. (2000). *Simulation neuronaler Netze*. Oldenbourg, München.
- Zenner, H.-P. (1994). *Hören : Physiologie, Biochemie, Zell- und Neurobiologie*. Thieme, Stuttgart.
- Zetsche, C. und Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Res.*, **30**(7):1111–1117.

- Zimbardo, P. G. (1992). *Psychologie*. Springer, Berlin.
- Zink, J. (1966). *Das Alte Testament - ausgewählt, übertragen und in geschichtlicher Folge angeordnet*. Kreuz-Verlag, Stuttgart.
- Zittlau, K. E., Claas, B. und Münz, H. (1986). Directional sensitivity of lateral line units in the clawed toad *Xenops laevis* Daudin. *J. Comp. Physiol. A*, **158**:469–477.
- Zwicker, E. (1982). *Psychoakustik*. Springer, Berlin.
- Zwicker, E. und Fastl, H. (1999). *Psychoacoustics: Facts and Models*. Springer, Berlin.

Danksagung

All denen, die bei der Anfertigung dieser Arbeit geholfen haben, sei herzlichst gedankt. Insbesondere gilt mein Dank Prof. Dr. J. Leo van Hemmen, der diese Arbeit ermöglichte. Ganz herzlich bedanken möchte ich mich bei allen Kollegen. Insbesondere auch bei Christof Schatz, durch den ich neben der interessanten Arbeit bei ASKOS (Büro für Analyse und Simulation komplexer Systeme) auch noch Zeit für die Forschung fand. Mein besonderer Dank gilt meinen Eltern Ursula und Peter Franosch die mir jederzeit mit Rat und Tat beiseite standen, sowie Mariana Claudy für ihre unermüdliche Hilfe.

Vielen Dank für eure Unterstützung!

Während der Arbeit wurde ich vom Graduiertenkolleg „Sensorische Interaktion in biologischen und technischen Systemen“ GRK 267/1-96 unterstützt.