

Lehrstuhl für Nachrichtentechnik

Visuelle und intersensorische  
Informationsverarbeitung:  
Signalstatistische Analyse und  
psychophysische Experimente

*Florian Röhrbein*

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing.habil G. Rigoll

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. G. Hauske
2. Univ.-Prof. Dr.rer.nat., Dr.h.c.mult. W. Brauer

Die Dissertation wurde am 21.10.2004 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 15.11.2004 angenommen.

**Visuelle und intersensorische  
Informationsverarbeitung:  
Signalstatistische Analyse und  
psychophysische Experimente**

# Inhalt

1.	Fragestellung	1
2.	Visuelle Verarbeitung in parallelen Kanälen	3
2.1	Einführung	3
2.2	Eigenschaften des visuellen Systems	4
2.2.1	Die frühe visuelle Verarbeitung	4
2.2.2	Ortsfrequenzselektive Zerlegung	7
2.3	Biologische und technische Bildkodierung	13
2.3.1	Übertragbarkeit und Anwendungsrahmen	13
2.3.2	Kanäle mit Additivrauschen	14
2.3.3	Frequenzbereichskodierung	15
2.3.4	Ressourcenzuteilung	18
2.4	Rate-Distortion-Theorie	19
2.4.1	Informationstheorie	19
2.4.2	Herleitung der optimalen Ressourcenaufteilung	22
2.4.3	Vorteil einer mehrkanaligen Repräsentation	24
2.5	Anpassung an die Eingangsstatistik	26
2.5.1	Statistik natürlicher Bilder	26
2.5.2	Das Leistungsdichtespektrum natürlicher Bilder	28
2.5.3	Der 1/f-Abfall	30
2.6	Die Oktavzerlegung	33
2.6.1	Selbstähnlichkeit	33
2.6.2	Lineare vs. selbstähnliche Zerlegung: Herleitung	34
2.6.3	Lineare vs. selbstähnliche Zerlegung: Komplexitätsvergleich	36

2.7	Die Optimalzerlegung	40
2.7.1	Komplexitätsvergleich	42
2.7.2	Resultierende Filterstruktur	44
2.7.3	Vergleich mit biologischer Kanalstruktur	46
2.8	Diskussion	49
3.	Intra- und intersensorische Verarbeitung	54
3.1	Erweiterung des kybernetischen Ansatzes	54
3.2	Modelle der Kanalinteraktion	56
3.2.1	Wahrscheinlichkeitssummutationsmodell	56
3.2.2	Neuronales Summutationsmodell	59
3.2.3	Quick'sches Poolingmodell	60
4.	Interaktion zwischen sensorischen Kanälen	63
4.1	Einführung	63
4.2	Multisensorische Informationsverarbeitung	64
4.2.1	Separate Sinnesmodalitäten	64
4.2.2	Natürliche Reizkomplexe	66
4.2.3	Multisensorische Signal- und Objektstatistik	68
4.2.4	Ausnutzung der polymodalen Korrelationen	69
4.3	Psychophysische Experimente	72
4.3.1	Experiment 1: Akustisch-visuelle Intensität	73
4.3.1.1	Methode	73
4.3.1.2	Ergebnisse	76
4.3.2	Experiment 2: Akustisch-visuelle Frontalbewegung	78
4.3.2.1	Methode	79
4.3.2.2	Ergebnisse	80
4.3.3	Multisensorisches ON-OFF Modell	81

4.3.4	Experiment 3: Akustisch-visuelle Trajektorien	84
4.3.4.1	Methode	85
4.3.4.2	Ergebnisse	89
4.3.5	Experiment 4: Akustisch-visuelles Rezeptivfeld	91
4.3.5.1	Methode	92
4.3.5.2	Ergebnisse	93
4.4	Diskussion	96
5.	Zusammenfassung	101
6.	Literatur	103

# 1. Fragestellung

In der vorliegenden Arbeit untersuche ich den Zusammenhang zwischen grundlegenden Eigenschaften sensorischer Informationsverarbeitung und der Statistik natürlicher Eingangssignale. Mit diesem kybernetischen Ansatz wurden bereits wichtige Merkmale und strukturelle Eigenschaften von Nervenzellen als Ergebnis eines evolutionären Anpassungsprozesses an die Statistik der umgebenden Physik interpretiert. Charakteristisch für neurobiologische Systeme ist dabei sowohl eine Signalaufspaltung in parallele Kanäle, als auch die Integration von Information aus unterschiedlichen Kanälen zu einem gemeinsamen Perzept ("binding").

Divergenz und Konvergenz werden nacheinander betrachtet: Im ersten Teil der Arbeit werden theoretische Überlegungen zur mehrkanaligen Verarbeitung im visuellen Kortex durch ortsfrequenzselektive Neurone angestellt. Zu den bei der Integration entstehenden sensorischen Interaktionen habe ich im zweiten Teil eine Reihe von psychophysischen Experimenten mit akustisch-visuellem Stimulusmaterial durchgeführt.

Für die intrasensorische Verarbeitung war bisher nicht geklärt, ob und wie die spezifische Ortsfrequenzerlegung im visuellen Kortex informationstheoretisch erklärt werden kann und welche Regelmäßigkeiten im Bildsignal dafür herangezogen werden müssen. Ich habe mir deshalb die Frage gestellt, wie im Rahmen der linearen Systemtheorie eine optimale Aufteilung des spezifischen Leistungsdichtespektrums natürlicher Bilder unter biologisch sinnvollen Komplexitätsbeschränkungen zu erzielen ist.

Methodisch können dazu technische Kodiersysteme analysiert werden, die durch ihre Bandpasszerlegung in enger Beziehung zu biologischen Sehsystemen stehen. Verschiedene Partitionierungsstrategien lassen sich dann im Rahmen der Rate-Distortion-Theorie hinsichtlich ihrer Kodierleistung evaluieren. Ein Vergleich zwischen theoretisch ermittelter Bandbreitenverteilung und biologisch dokumentierter Kanalstruktur entscheidet dann über den Erfolg einer Interpretation im Rahmen des gewählten informationstheoretischen Ansatzes.

Im zweiten Teil der Arbeit wird die Fusion von Informationen aus verschiedenen Modalitäten thematisiert und experimentell untersucht. Ausgangspunkt ist die Beobachtung, dass biologische Systeme unter natürlichen Randbedingungen meist mit polysensorischen Stimuli konfrontiert sind, deren modalitätsspezifischen Signale statistisch korreliert sind oder beispielsweise durch Bewegung relativ zu einem Beobachter zu Kovariationen in den Eingangskanälen führen.

Bereits Aristoteles weist indirekt auf diesen Umstand hin. Er wird zwar meist als Urvater einer Aufteilung der Wahrnehmung in verschiedene Sinne zitiert, in seinem Werk *Περὶ Πσυχῆς* (Über die Seele) findet sich neben der bekannten Aufzählung der einzelnen Sinne jedoch noch ein weiterer Sinn: *κοινὴ αἰσθησις*, der "gemeinsame Wahrnehmungssinn". Dieser wird von Aristoteles postuliert, da er feststellt, dass es gewisse "gemeinsame Sinnesqualitäten" gibt. Dazu rechnet er (Buch II, Kap. 7, 418a):

“... κοινὰ δὲ κινήσεις, ἠρεμία, ἀριθμὸς, σχῆμα, μέγεθος. τὰ γὰρ τοιαῦτα οὐδεμίᾳς ἐστὶν ἰδέα, ἀλλὰ κοινὰ πασαισ. καὶ γὰρ ἀφῆ κινήσεις τίς ἐστὶν αἰσθητὴ καὶ ὄψει.”

“... gemeinsame [Objekte] aber sind Bewegung, Ruhe, Zahl, Gestalt, Größe. Solche Objekte sind nämlich keinem einzelnen Sinn eigentümlich, sondern allen gemeinsam; denn eine Bewegung ist sowohl durch den Tastsinn, als auch durch das Gesicht wahrnehmbar.”

In einer Reihe von psychophysischen Experimenten bin ich der Frage nachgegangen, ob und unter welchen Bedingungen diese in multisensorischen Stimuli enthaltene Redundanz die Präzision der Wahrnehmung verbessert. Berichtet wird über Versuche zur Diskriminationsleistung akustisch-visueller Intensitäten und akustisch-visueller Bewegungsinformation, sowie über Versuche mit horizontalen Trajektorien, mit denen spatio-temporale Randbedingungen für eine optimale Integrationsleistung ermittelt wurden.

Die Experimente wurden am Institut für Medizinische Psychologie der Ludwig-Maximilians-Universität München (Experiment 1 und 2) und am MacKay Institute of Communication and Neuroscience an der Keele University in England (Experiment 2 und 3) durchgeführt.

## 2. Visuelle Verarbeitung in parallelen Kanälen

### 2.1 Einführung

Im Zuge des kybernetischen Ansatzes zum Verständnis und zur Modellierung biologischer Wahrnehmungssysteme, haben Attneave, Barlow und andere argumentiert, dass die involvierten neuronalen Strukturen möglicherweise am besten im Sinne einer informationstheoretischen Interpretation verstanden werden können. Dieser Erklärungsansatz nimmt eine optimale Anpassung sensorischer Systeme an die spezifische statistische Struktur der natürlichen Umgebung durch Schaffung effizienter, minimal redundanter Signalrepräsentationen an. Tatsächlich konnte damit bisher für eine ganze Reihe von neuronalen Mechanismen v.a. des visuellen Systems der Nachweis geführt werden, dass sie geeignet sind, durch Dekorrelation die vorhandenen Signalredundanzen auszunutzen.

Zu den untersuchten Eigenschaften gehören die laterale Inhibition der retinalen Ganglienzellen (Laughlin, 1981, Srinivasan et al., 1982, van Hateren, 1992, Barlow, Földiak, 1989, Atick, Redlich, 1990), Weber's Gesetz (Zetsche, Hauske, 1989a, Röhrbein, Zetsche, 2002), die Orientierungsselektivität von Nervenzellen im visuellen Kortex (z.B. Linsker, 1986, Barrow, 1987, Sanger, 1989), der Farbantagonismus (Webster, Mollon, 1991, Zaidi, Shapiro, 1993) sowie Eigenschaften des extra-klassischen rezeptiven Felds (Zetsche, Röhrbein, 2001, Übersicht in Rao et al., 1998).

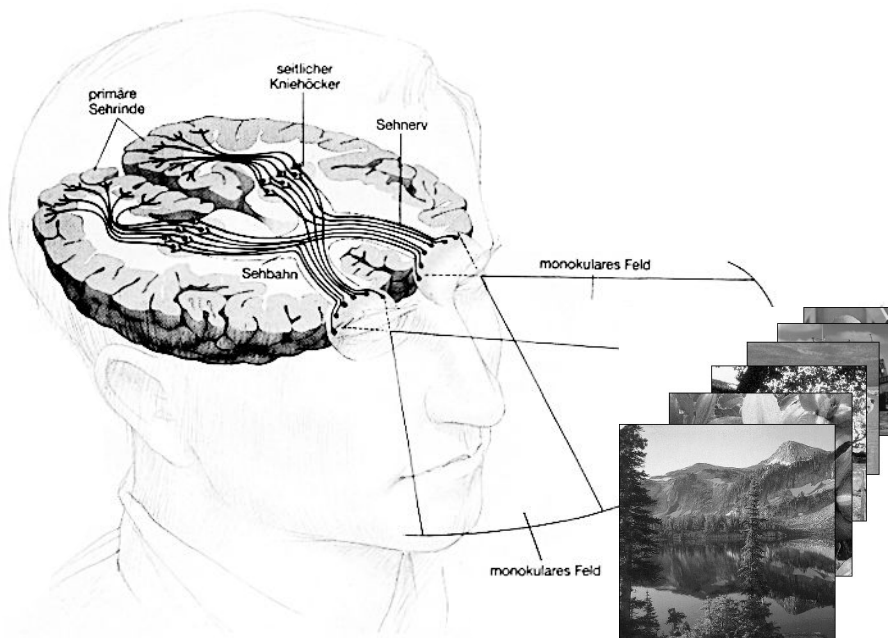
Ungeklärt war bisher, wie eine weitere, grundlegende Eigenschaft kortikaler Neurone informationstheoretisch zu interpretieren ist: die charakteristische Struktur der Bandpasszerlegung innerhalb der Orientierungssäulen des visuellen Kortex. In diesem Kapitel soll nun geprüft werden, ob das Dekorrelationsparadigma auch zur Begründung dieser Eigenschaft herangezogen werden kann. Voraussetzung dafür ist die Klärung des theoretischen Rahmens, die Motivierung der zu optimierenden Größe und die Identifikation der relevanten statistischen Strukturen. Zunächst aber erfolgt zum besseren Verständnis eine knappe Übersicht über die frühen Stufen der visuellen Informationsverarbeitung.



## 2.2 Eigenschaften des visuellen Systems

### 2.2.1 Die frühe visuelle Verarbeitung

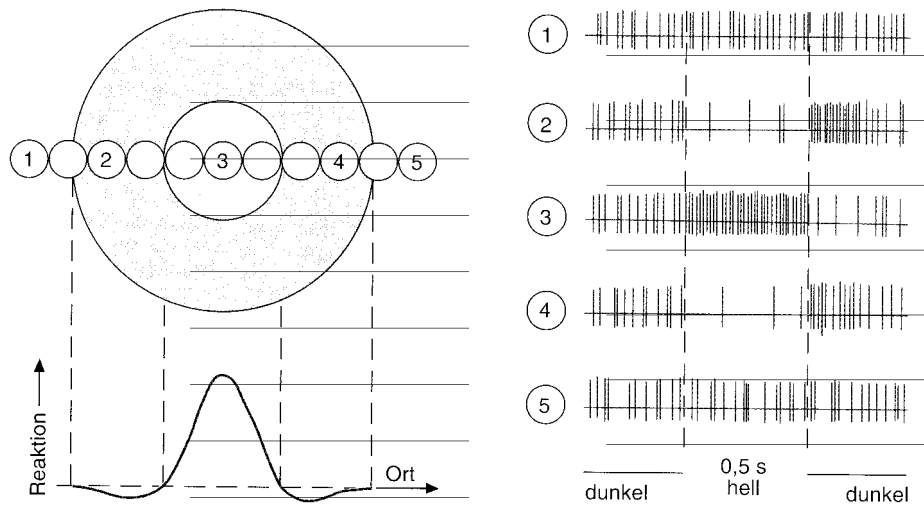
Die Verarbeitung visueller Information erfolgt in einer Hierarchie von Verarbeitungsstufen (Felleman und van Essen, 1991): Das eingehende Lichtsignal wird bereits in der Netzhaut vorverarbeitet, über den lateralen Kniehöcker (LGN) verschaltet und erreicht im primären visuellen Areal V1 den Kortex, in dem es schließlich in ca. 30 extrastriären Arealen einer weiteren Analyse unterzogen wird (Abbildung 1).



*Abbildung 1:* Die Verarbeitung visueller Reize verläuft von der Retina über die seitlichen Kniehöcker zur primären Sehrinde. Von dort aus wird das Signal in einer Reihe weiterer kortikaler Areale einer reiz- und aufgabenspezifischen Analyse unterzogen.

Eine Charakterisierung der beteiligten visuellen Neurone erfolgt dabei durch Aussagen über ihr rezeptives Feld, ein Begriff, der von Sherrington (1906) und Hartline (1938) eingeführt wurde. Das rezept-

tive Feld (classical receptive field - CRF) beschreibt den Bereich des Gesichtsfeldes, für den Antworten bei dieser Nervenzelle ausgelöst werden können. In Abbildung 2 ist eine solche Charakterisierung für Ganglienzellen auf der Retina für unterschiedliche Stimuluspositionen (1-5) und einem zeitlichen dunkel-hell-dunkel Reiz gezeigt.

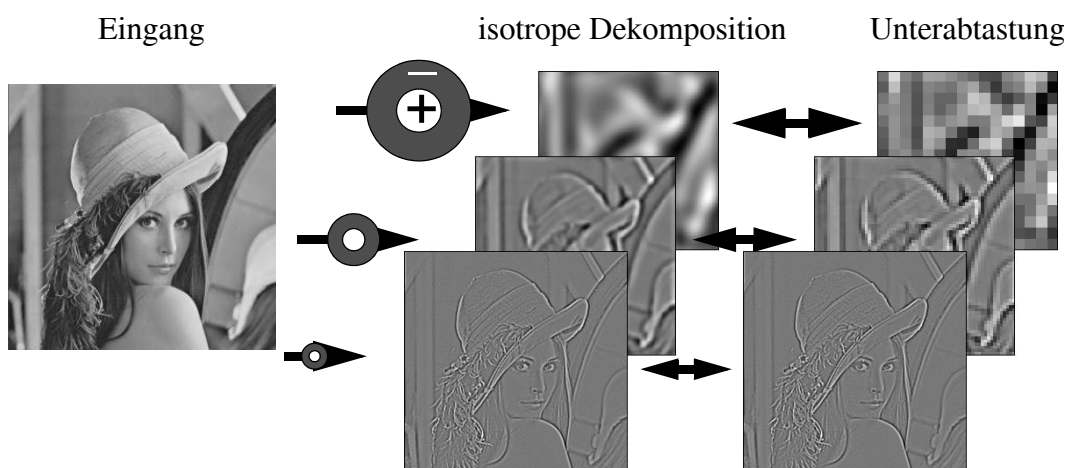


*Abbildung 2:* Die rezeptiven Felder der retinalen Ganglienzellen sind durch eine antagonistische Verschaltung von Zentrum und Umfeld charakterisiert (links). Diese ON/OFF-Struktur führt zu einem Antwortverhalten, das bei inhomogener Beleuchtung den maximalen Wert erreicht (rechts). Abbildungen aus Kandel et al., 1995 nach Kuffler, 1953.

Die Beschreibung der rezeptiven Felder im Ort hat ihre Entsprechung im Fourierbereich, indem die Größe und Form des rezeptiven Feldes die Selektivität der Nervenzelle für einen Bereich von Ortsfrequenzen bestimmt. Retinale Ganglienzellen und Zellen im LGN weisen hier eine Tiefpasscharakteristik auf, mit scharfen oberen Grenzfrequenzen und einem langsamen Sensitivitätsabfall hin zu tiefen Ortsfrequenzen. Das Vorhandensein von unterschiedlich skalierten Versionen desgleichen Typs führt bereits zu einer mehrkanaligen, aber ungerichteten Bildrepräsentation (siehe Abbildung 3).

Die Hierarchie von Verarbeitungsstufen im visuellen System des Menschen kann als Aufeinanderfolge von linearen und auch nichtline-

aren Signaltransformationen betrachtet werden, durch die das Eingangssignal fortlaufend umkodiert wird. Im Zuge dieses Verarbeitungsprozesses steigt einerseits die Selektivität für spezifische, objektgebundene Bildmerkmale, andererseits wird im Zuge einer Objekterkennung über gewisse Stimulusparameter (z.B. Position) generalisiert, etwa mittels einer additiven Verschaltung. Es erfolgt eine Enkodierung der für die Spezies verhaltensrelevanten Eingangsinformation in immer effizientere Repräsentationen, die bestimmte Invarianzeigenschaften aufweisen und bei Stimulation mit natürlichen Bildsignalen eine zunehmend spärlichere Aktivitätsverteilungen der Zellen aufweisen (Übersicht in Reinagel, 2001). Vergleicht man z.B. die Amplitudenhistogramme vor und nach einer isotropen Zerlegung wie in Abbildung 3, stellt man statt einer ursprünglich breiten Verteilung der Grauwerte nun eine deutliche Konzentrierung der neuronalen Aktivierungsniveaus um Null fest.



*Abbildung 3:* Die Ganglienzellen führen durch die Organisation ihres rezeptiven Feldes eine isotrope Teilbandzerlegung auf unterschiedlichen Skalen durch. Das Ausgangssignal kann anschließend aufgrund seiner Bandbegrenzung für eine effiziente Repräsentation passend unterabgetastet werden.

Die Informationsreduktion durch Beseitigung von irrelevanter und redundanter Information kann als Abstraktionsleistung, Mustererkennung oder Klassifikation interpretiert werden. Eine erste Reduktion mit einem Verhältnis zwischen Rezeptoren und Ganglienzellen von etwa

150:1 findet dabei bereits im Auge statt. Schätzungen der einströmenden visuellen Information liegen bei etwa  $5 \cdot 10^7$  bit/s, bewusst aufgenommen und verarbeitet können jedoch nur Informationsflüsse unter 50 bit/s (vgl. Marko, 1982). Es liegen auch Schätzungen für einzelne Neurone vor, Nervenzellen im Areal MT übertragen beispielsweise etwa 5.5 bit/sec (Buracas et al., 1998).

### **2.2.2 Ortsfrequenzselektive Zerlegung**

Auf drei parallelen Pfaden (M, P und K pathway) gelangt die visuelle Information in das primäre Sehareal V1. Als erste visuelle Neurone im Kortex gelten die Simple Cells, die von Hubel und Wiesel zuerst beschrieben wurden (Hubel, Wiesel, 1959). Ihre Ableitungen vom Kortex der Katzen zeigten bereits eine deutliche Selektivität für Orientierung. Da Balken und Kanten als Stimuli benutzt wurden, die selbst ein breites Spektrum enthalten, konnte die Ortsfrequenzspezifität jedoch nicht nachgewiesen werden. Erst in späteren Studien an Katzen (Campbell, Cooper, Enroth-Cugell, 1969, Maffei, Fiorentini, 1973) und auch an Affen (De Valois et al., 1975, Schiller et al., 1976) wurde die hohe Ortsfrequenzselektivität dieser kortikalen Neurone entdeckt.

Auch zahlreiche psychophysische Experimente, die in der Folgezeit unternommen wurden, sprechen für eine Verarbeitung der visuellen Information in getrennten Kanälen, die jeweils für einen bestimmten Bereich von Ortsfrequenzen selektiv sind. Die rezeptiven Felder dieser Zellen im primären visuellen Areal bestehen dabei aus parallel angeordneten exzitatorischen (ON) und inhibitorischen (OFF) Bereichen, die Zellantwort ergibt sich aus linearer Summation innerhalb dieser Bereiche mit anschließender antagonistischer Verschaltung (siehe Abbildung 4). Im Unterschied zu Ganglienzellen besitzen sie recht genaue untere und obere Grenzfrequenzen, bedingt durch eine stärkere Gewichtung der inhibitorischen Felder und eine multioszillatorische Struktur des Rezeptivfeldes.

Die Sensitivität erstreckt sich insgesamt über einen Bereich von etwa acht Oktaven (2 - 8 c/deg im fovealen Bereich) mit der Modulationsübertragungsfunktion als Umhüllende. Erste Messungen zur Bandpasscharakteristik wurden von Schade (1956) und Menzel

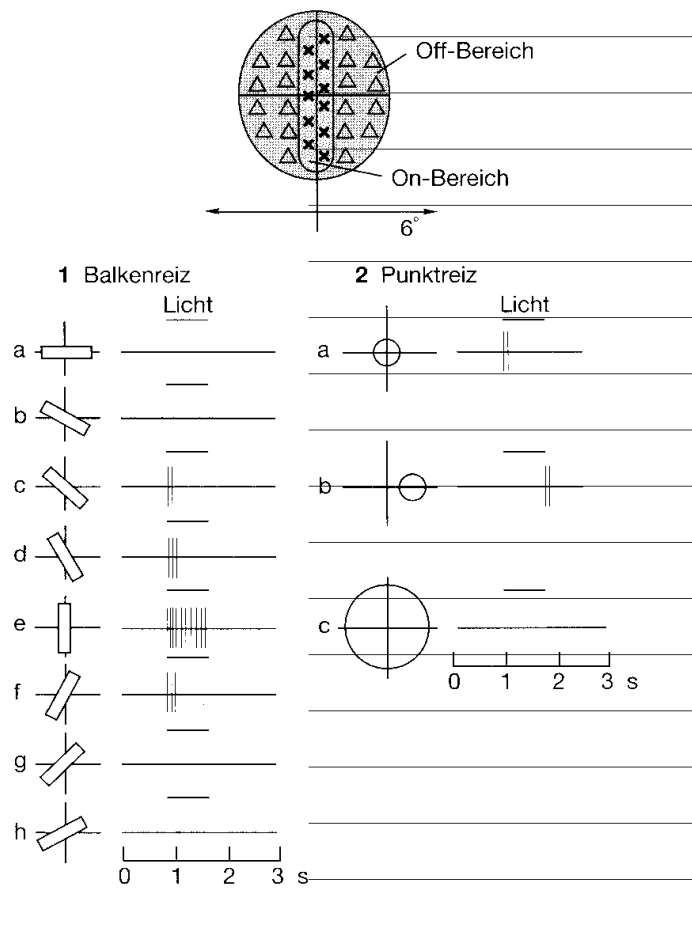


Abbildung 4: Kortikale Simple Cells enthalten elongierte exzitatorische und inhibitorische Bereiche, die ein orientierungsselektives Antwortverhalten auf Balkenreize verursachen. Abbildung aus Kandel et al., 1995 nach Hubel, Wiesel, 1959.

(1959) durchgeführt. Die Ortsfrequenzselektivität der einzelnen Neurone kann quantitativ mit ihrer Bandbreite angegeben werden (full width at half height):

$$\text{Bandbreite} = \frac{\log F_h - \log F_l}{\log 2}$$

Die mittlere Bandbreite von Neuronen im visuellen Kortex beträgt etwa 1.4 Oktaven. Simple Cells und die nichtlinearen Complex Cells haben dabei sehr ähnliche Bestfrequenzen und Bandbreiten, wobei die Mittenfrequenzen von Complex Cells etwas höher liegen (De

Valois et al., 1982) und Simple Cells ein etwas engeres Tuning aufweisen (Movshon et al., 1978, Schiller et al., 1976).

Die Frequenzzerlegung der Simple Cells besitzt eine charakteristische Struktur: Die gemessenen Frequenzbänder haben eine annähernd konstante Bandbreite auf einer logarithmischen Frequenzachse, d.h. ihre linearen Bandbreiten steigen direkt proportional zur Mittenfrequenz. In Abbildung 5 sind die Ergebnisse aus elektrophysiologischen Messungen dargestellt.

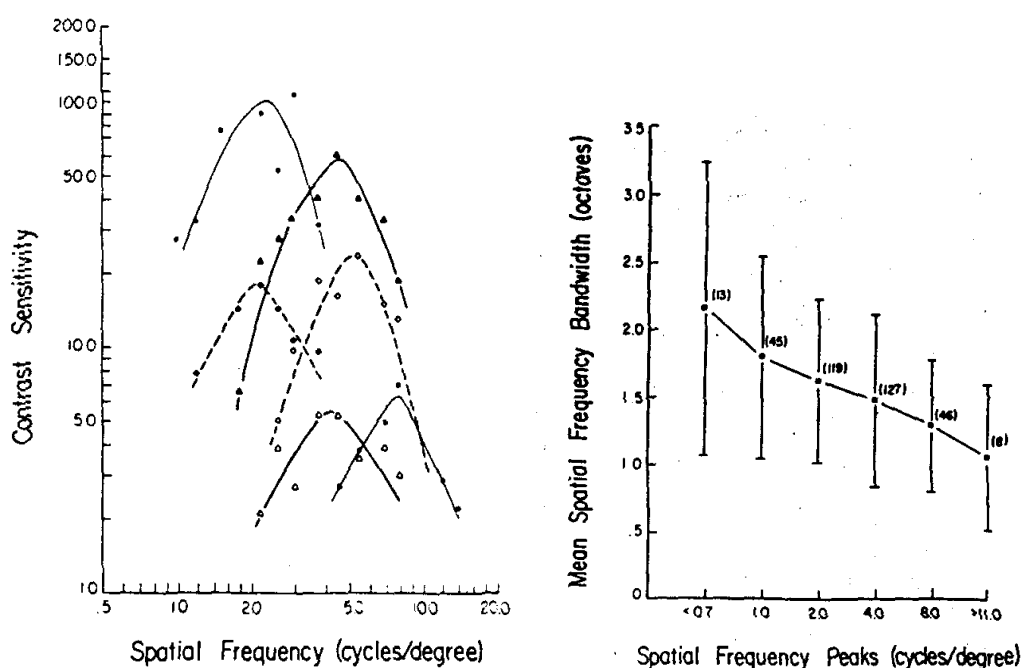


Abbildung 5: Die Teilbandzerlegung durch Simple Cells weist eine spezifische Struktur auf. Sie ist charakterisiert durch annähernd selbstähnliche Kontrastsensitivitätsfunktionen (links), die eine systematische Abweichung in den Oktav-Bandbreiten aufweisen (rechts). Siehe Text für Details. Die Daten stammen von kortikalen Simple Cells aus dem Areal V1 des Affen (De Valois et al., 1982).

Rechts in der Abbildung ist die Verteilung der Bandbreiten der Zellen als Funktion ihrer örtlichen Mittenfrequenz angetragen (Mittelwerte +/- 1 Standardabweichung, in Klammern die Absolutzahl der abgeleiteten Neurone). In dieser Darstellung wird die systematische

Abweichung von einer exakten Selbstähnlichkeit deutlich, denn mit steigender Mittenfrequenz nimmt die Bandbreite auf log-Skala kontinuierlich ab.

Andere elektrophysiologische Studien kommen zu sehr ähnlichen Ergebnissen: Die Arbeit von Foster, Gaska und Pollen (1983) misst am Makaken Bandbreiten von 0.7 bis 2.8 Oktaven, bei einer Untersuchung an Katzen werden Zellen mit Bandbreiten zwischen 0.7 und 3.2 Oktaven (im Mittel 1.3) gefunden (Movshon et al., 1978). Weitere neurophysiologische Evidenzen für den Zusammenhang zwischen Mittenfrequenz und Bandbreite finden sich in Kulikowski, Bishop, 1981 und Tolhurst, Thompson, 1981.

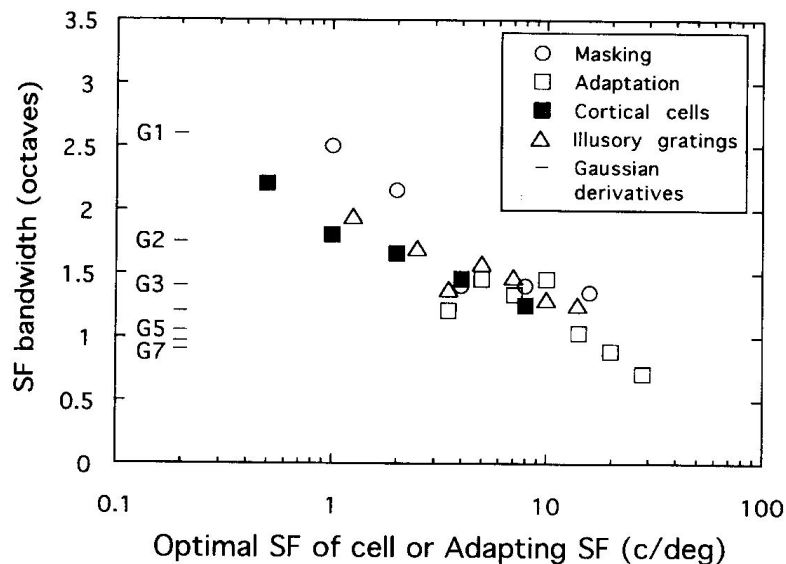


Abbildung 6: Psychophysische Untersuchungen untermauern die neurophysiologische Evidenz gegen eine Invarianz der Bandbreiten. Verschiedene Versuchstechniken belegen konsistent eine kontinuierliche Abnahme der Bandbreiten mit steigender Mittenfrequenz auf einer log-Skala. Abbildung aus Bruce et al., 1996.

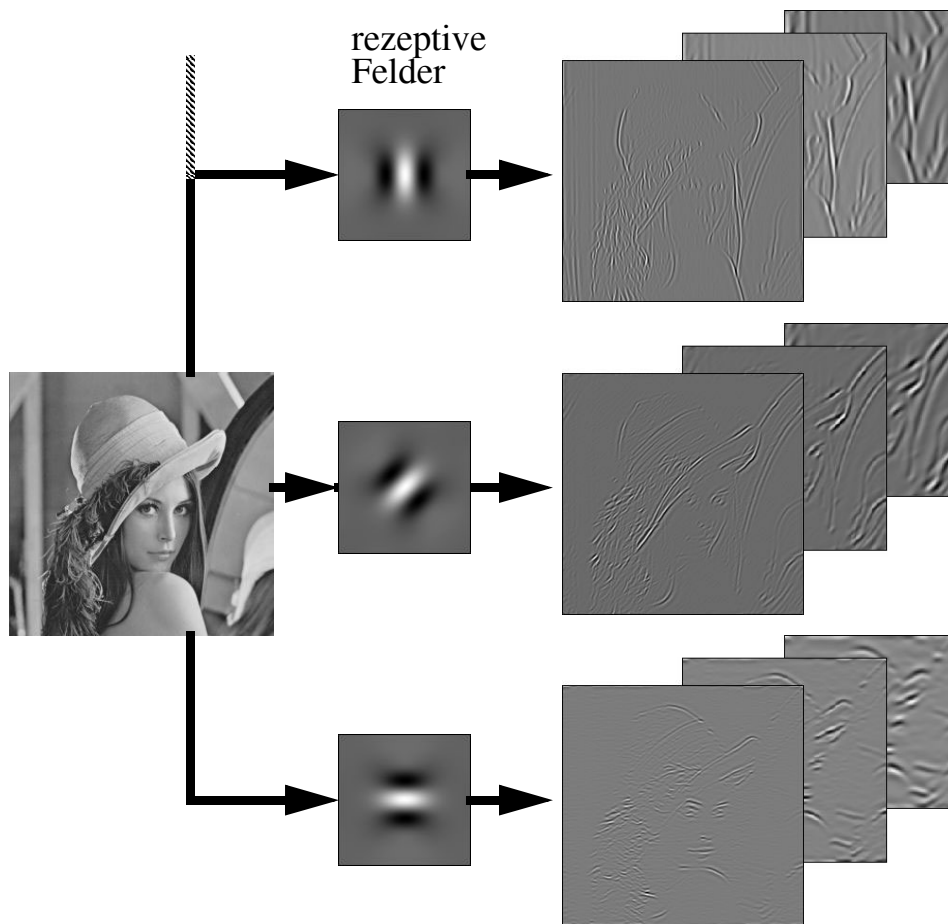
Eine exakte Selbstähnlichkeit würde einer über die Skalen gleich bleibenden, größeninvarianten CRF-Struktur entsprechen. Tatsächlich weisen die rezeptiven Felder der Neurone aber eine

unterschiedliche Anzahl von Oszillationen auf und hochfrequente Neurone besitzen eher zusätzliche "sidelobes". Eine systematische Abnahme der logarithmischen Bandbreiten mit zunehmender Mittenfrequenz zeigen auch zahlreiche psychophysische Experimente. Dabei kamen unterschiedliche Paradigmen zum Einsatz: selektive Adaptation (Pantle und Sekuler, 1968, Blakemore, Campbell, 1969, De Valois, 1977), Maskierung (Stromeyer, Julesz, 1972, Legge, Foley, 1980, Wilson et al., 1983) und unterschwelligen Summation (Sachs et al., 1971, Legge, Foley, 1980, Watson, 1982). Für einen kritischen Vergleich der Methoden siehe Dealy, Tolhurst (1977) und Graham et al. (1978). In Abbildung 6 sind die Ergebnisse psychophysischer Experimente von Blakemore, Campbell, 1969, Stromeyer, Julesz, 1972, Georgeson, 1980, Wilson et al., 1983 und Wilson, Gelb, 1984 berücksichtigt.

Die Entdeckung, dass sich überall im kortikalen Areal V1 Zellen aus dem gesamten Ortsfrequenzspektrum finden, führte zum Mehrkanalkonzept der visuellen Verarbeitung (Campbell, Robson, 1968, Pantle, Sekuler, 1968, Thomas, 1970) und es wurde zunehmend von einer lokal begrenzten Fourieranalyse ("visuelle Frequenzgruppe") durch die kortikalen Simple Cells gesprochen. In Abbildung 7 ist anhand eines in der Nachrichtentechnik häufig verwendeten Testbilds die Bandpasszerlegung durch die Simple Cells illustriert. Diese erfolgt orientierungsselektiv auf unterschiedlichen Skalen und führt dadurch zu der mehrkanaligen Repräsentation. In meiner Arbeit hier bezieht sich die Analyse und informationstheoretische Herleitung auf die Skalenstruktur entlang der Ortsfrequenzachse, weshalb die Orientierungsdimension im Folgenden zu vernachlässigen ist. Es wird daher wird im Eindimensionalen mit einer beliebigen, aber festen Orientierung gerechnet.

Ungeklärt war bisher, warum diese Zellen die beschriebene Ortsfrequenzzerlegung in nur annähernd selbstähnliche Kanäle durchführen, wie diese Zerlegung informationstheoretisch erklärt werden kann und warum sich im Laufe der Evolution gerade diese spezifische Struktur herausgebildet hat.





*Abbildung 7:* Die kortikale Zerlegung des Eingangsbildes durch orientierungsselektive Neurone unterschiedlicher Größe führt zu einer mehrkanaligen Repräsentation. Innerhalb einer kortikalen Orientierungssäule finden sich Bandpassauszüge aus dem gesamten Ortsfrequenzspektrum. Rechts in der Abbildung sind drei solcher Bandpässe für jede Orientierung dargestellt.

## **2.3 Biologische und technische Bildkodierung**

Zur Interpretation der biologischen Frequenzerlegung ist es angebracht, technische Multikanalkonzepte zu betrachten, die aus Gründen der Datenkompression ebenfalls bandpassgefilterte Repräsentationen verwenden.

### **2.3.1 Übertragbarkeit und Anwendungsrahmen**

Der hier angewandte Optimierungsansatz wurde im Kontext von technischen Bildkompressionsverfahren entwickelt. Zunächst ist vielleicht nicht klar, welche Beziehung zwischen der Leistung visueller Neurone und den Prinzipien technischer Kodierverfahren besteht, zumal viele Komponenten in technischen Schemata keine offensichtliche Entsprechung in biologischen Strukturen zu haben scheinen. Wenn wir jedoch die Systeme auf einer abstrakteren Ebene betrachten, wird ihre Ähnlichkeit bezüglich grundlegender Randbedingungen für das Systemdesign ersichtlich. Grundlage dafür ist eine kybernetische Betrachtung des Gesichtssinns als Übertragungskanal.

Um die verschiedenen Variablen, Operationen und Konzepten der Informationstheorie auf biologische Systeme anwenden zu können, müssen zunächst Überlegungen angestellt werden, wie die verschiedenen Aspekte der natürlichen Umgebung und der neuronalen Informationsverarbeitung auf diese Größen bezogen werden sollen. Es ist also die Identifikation von korrespondierenden Einheiten zu leisten. Dabei gibt es eine Reihe von Möglichkeiten, eine solche Abbildung herzustellen, je nach angestrebtem Detaillierungsgrad und gewünschtem Fokus. Im Sinne einer Informationsübertragung über einen verrauschten Kanal kann beispielsweise das retinale Signal als Informationsquelle, der Nervus opticus als störungsbehafteter Kanal und der Kortex als Informationssenke betrachtet werden. Eine andere Herangehensweise würde darin bestehen, die Kanalstörung zu vernachlässigen und stattdessen auf die in der Quelle vorhandenen Redundanzen zu fokussieren. Hierbei werden dann die Stufen der visuellen Verarbeitung als hierarchischer Umkodierungsprozess zur Redundanzminimierung interpretiert.

Die Granularitätsebene kann sowohl weiter verfeinert werden, indem lokale Optimierungen der Informationsverarbeitung zwischen benachbarten neuronalen Schichten betrachtet werden, oder, indem sogar das einzelne Neuron Gegenstand informationstheoretischer Analyse wird (für einen Überblicksartikel hierzu siehe z.B. Borst, Theunissen, 1999). Andererseits kann der Analyserahmen aber auch radikal erweitert werden bis hin zum Informationstransfer zwischen der dreidimensionalen physikalischen Umgebung und der kognitiven Repräsentation im Gehirn (z.B. Young, Purves, 2003). In diesem Fall wird durch die Projektion auf das zweidimensionale Retinabild bereits eine erhebliche, weitere Störgröße in das System eingeführt, weshalb das Ziel einer Optimierungsstrategie dann darin besteht, die intern verfügbare Information über die äußere Umgebung zu maximieren (statt einer Maximierung der Information über das Bild selbst).

Wir wollen hier den üblichen Weg gehen und uns auf das retinale Eingangsbild als Informationsquelle beziehen, da die dort zur Verfügung stehende Information bereits von ausreichender Komplexität ist.

### **2.3.2 Kanäle mit Additivrauschen**

Alle technischen Methoden zur Bilddatenkompression basieren auf zwei fundamentalen Prinzipien: Das eine Prinzip besteht darin, Eigenschaften der Quelle auszunutzen und dadurch Redundanz vom Signal zu entfernen (redundancy reduction), das andere Prinzip nutzt Eigenschaften der Signalsenke, also des visuellen Systems, um überflüssige Information aus dem Signal zu entfernen (irrelevancy reduction).

Der essentielle Punkt für die Übertragbarkeit von technischen Konzepten auf biologische Systeme ist, dass wichtige Ergebnisse der technischen Quell- und Kanalkodierung im gleichen Grundschema beschrieben werden können. Das liegt im Wesentlichen am additiven Rauschmodell der Quantisierung, das besagt, dass Quantisierer in technischen Quellkodierern formal als verrauschte Kanäle behandelt werden können (vgl. Abbildung 8). Im biologischen Kontext können die Verarbeitungsstufe und das Systemrauschen mit der Operation von neuronalen Elementen assoziiert werden, die eine begrenzte Komplexität und eine eingeschränkte Zuverlässigkeit aufweisen. Da Neurone ebenso als verrauschte Kanäle betrachtet werden können,

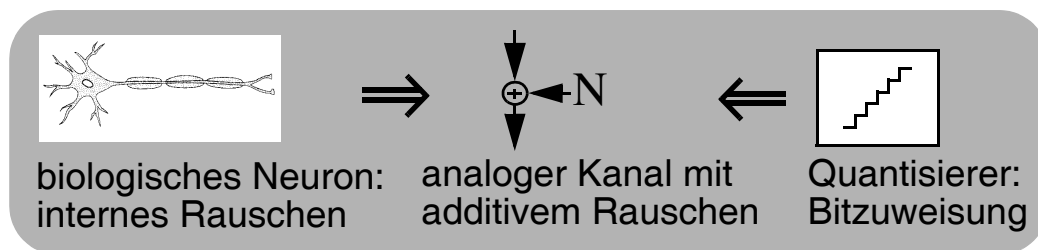


Figure 8: Zusammenhang zwischen Nervenzelle, rauschbehaftetem Analogkanal und technischem Quantisierer. Die grundlegende Äquivalenz zwischen technischem und biologischem Aufwand manifestiert sich in der Bitrate des Quantisierers bzw. in dem Dynamikbereich und in dem internen Rauschniveau des Neurons.

ist diese Äquivalenz auch die Schlüsseleigenschaft, die die Anwendung von gemeinsamen Optimierungsprozeduren für technische Bildkompressionsverfahren und biologischen Sehsystemen rechtfertigt.

### 2.3.3 Frequenzbereichskodierung

Vergleichbar mit der biologischen Zerlegung ist eine auf einer speziellen irregulären Baumstruktur basierende, technische Teilbandzerlegung. Abbildung 9 zeigt das Grundschema, das den meisten technischen Kodierverfahren zu Grunde liegt. Dieses weist mehrere Stufen auf: Die Informationsquelle liefert den Input für das System, in der Zerlegungsstufe wird dieses Signal transformiert und aufgrund von Rauschen und anderen Ressourcenbegrenzungen gestört. Das resultierende Signal kann dann als "interne" Repräsentation des Inputs betrachtet werden. Es wird in der Informationssenke schließlich dekodiert, idealerweise in eine exakte Kopie des Eingangssignals.

In einer biologischen Interpretation kann nun die Stufe der Zuordnung von Systemressourcen mit einer internen Rauschquelle und dem physiologischen Aufwand zur Erreichung einer bestimmten Genauigkeit in Beziehung gebracht werden. In diesem Schema lassen sich sowohl Blocktransformationen, wie KLT/PCA und Cosinustransformation, als auch Verfahren mit Filterbänken, wie die Teilbandkodierung oder der Wavelet-Ansatz darstellen. Bekanntermaßen lassen sich alle orthogonalen Transformationen (DCT, usw.) als Spezialfälle der Teilbandzerlegung interpretieren (Vaidyanathan, 1993).

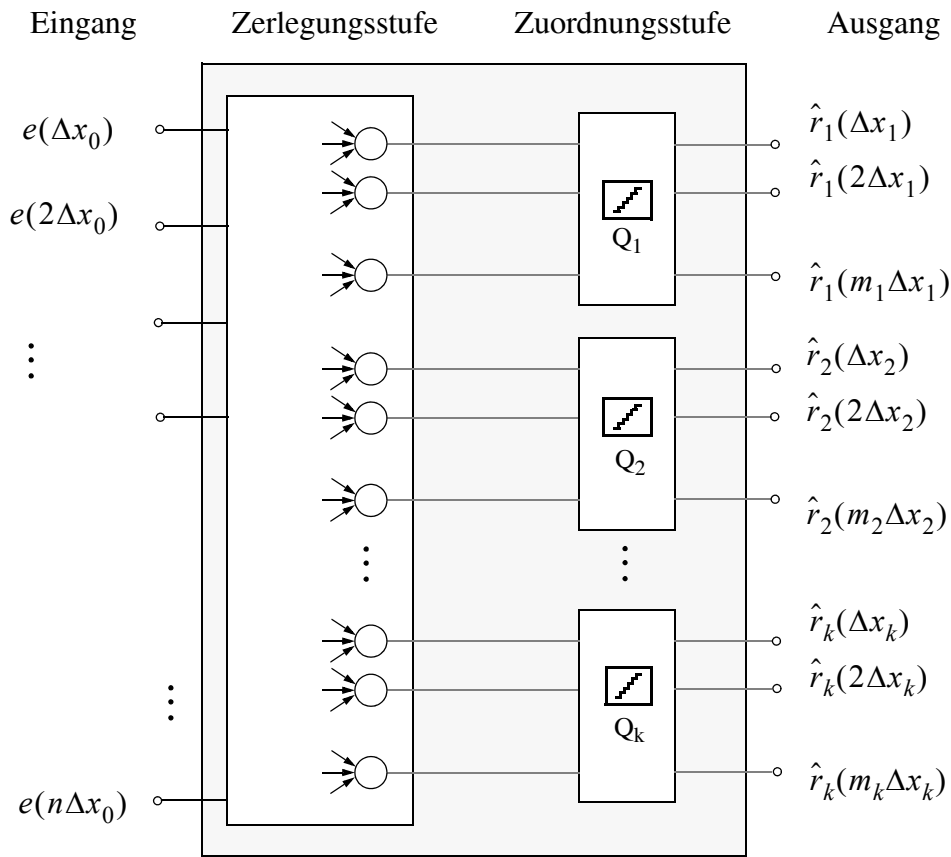
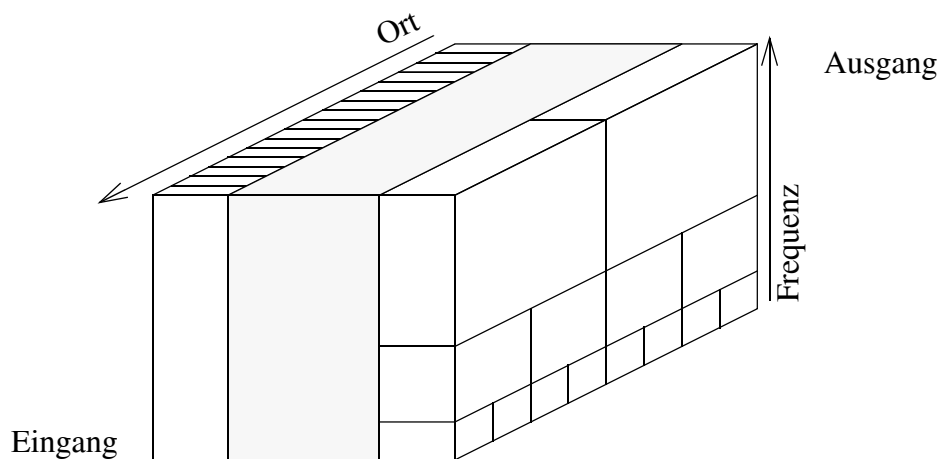


Abbildung 9: Prinzip der Signalkodierung mit Filterbänken oder Blocktransformationen. Das eintreffende Signal mit  $n$  Eingangssamples und Samplingabstand  $\Delta x_0$  wird in Kanäle  $1, \dots, k$  mit Samplingabständen  $\Delta x_l$  aufgeteilt. Dies führt zu insgesamt  $m = \sum_{l=1,k} m_l$  Samples am Ausgang (siehe Text).

Die klassischen linearen Transformationen arbeiten meist mit einer gleichförmigen Bandbreite bei der Spektralanalyse (eine Ausnahme bildet die Haar-Transformation). Wie im vorangegangenen Abschnitt gezeigt, unterscheiden sich biologische Schemata von solchen einfachen technischen Zerlegungsschemata darin, dass sie keine Partitionierung in Frequenzbänder gleicher Breite (auf linearer Ortsfrequenzachse) benutzen, sondern stattdessen auf annähernd selbstähnliche Frequenzbänder zugreifen. In der technischen Notation des Teilbandkodierens (subband coding, SBC; Crochiere et al., 1976, Crochiere, Rabiner, 1983, Woods, O'Neil, 1986) wird die lineare

Zerlegung oft als reguläre Binärbaumzerlegung bezeichnet und die selbstähnliche logarithmische Aufteilung als Oktavbandanalyse, dyadische Zerlegung oder auch constant-q Filterung (Nelson et al., 1972, Younberg, Boll, 1978, Barabell, Crochiere, 1979).

Bei einer solchen ungleichförmigen Frequenzzzerlegung in kaskadierten Teilbandsystemen verfeinert sich die Frequenzauflösung mit jeder Kaskadenstufe um den Faktor 2, während die Länge der gesamten Impulsantwort des Analysefilters in einem einzelnen Frequenzband dadurch zunächst um den Faktor 2 größer wird. Aufgrund ihrer ähnlichen Partitionierungsschemata (Abbildung 10) stehen neben der Oktavbandanalyse auch Multiskalendarstellungen wie z.B. die Laplace-Pyramide (Burt, Adelson, 1983) und die Wavelet-Transformation (Mallat, 1989) in enger Beziehung zu der biologischen Aufteilung durch kortikale Simple Cells.



*Abbildung 10:* Die ortdiskreten Samples am Eingang werden mit “biologienahen” Zerlegungsschemata wie z.B. Wavelets in Ausgangssamples transformiert, die durch Ausnutzung beider Freiheitsgrade (Ort, Frequenz) zu einer lokalisierten, selbstähnlichen Repräsentation führen.

### **2.3.4 Ressourcenzuteilung**

In beiden Systemen, biologischen und technischen, ist ein optimales Verhältnis zwischen begrenzter Verfügbarkeit von Ressourcen und geeigneter Darstellungsqualität zu finden. Während bei technischen Verfahren die Anzahl der Samples und die Anzahl der Quantisierungsstufen die Schlüsselressourcen darstellen, sind biologische Systeme in der Anzahl der Neurone und in dem Signal-Rausch-Verhältnis, das mit diesen erreichbar ist, limitiert. Biologische Kosten nehmen in ähnlicher Art und Weise mit der Dynamik und mit dem Rauschniveau der Ausgabeneurone zu, wie die Bitrate mit der Anzahl der Quantisierungsstufen wächst. Die technische Ausstattung eines Quantisierers mit einer größeren Anzahl von Quantisierungsstufen ist daher äquivalent zu der biologischen Realisierung einer Einheit mit einem geringeren internen Rauschniveau.

Es gibt in der Tat Evidenzen dafür, dass auch biologische Systeme den Freiheitsgrad besitzen, neuronalen Einheiten unterschiedliche relative Rauschniveaus oder unterschiedliche Dynamikbereiche zuzuordnen (Albrecht, Hamilton, 1982, Barlow et al., 1989). Aber auch wenn die Kontrolle des Signal-Rausch-Verhältnisses nicht von einem einzelnen Neuron geleistet werden kann, sondern andere Strategien erforderlich sind, wie etwa eine Populationskodierung durch mehrere Neurone, erscheint das Grundkonzept von einem biologischen Aufwand, das eng mit dem technischen Aufwand verbunden ist, gerechtfertigt. Hierin liegt eine wichtige Einsicht, da es die Möglichkeit eröffnet, Methoden aus der Nachrichten- und Informationstechnik in den Bereich der biologischen Sehsysteme zu transferieren.

## **2.4 Rate-Distortion-Theorie**

Der informationstheoretische Ansatz zur Beschreibung biologischer Sehleistungen stellt keinen homogenen theoretischen Rahmen dar, sondern eher eine Menge von Ansätzen, denen ein Reservoir von Grundideen und Konzepten gemein ist. Die einzelnen Ansätze unterscheiden sich dabei nicht nur in den Details, wie physikalische und biologische Entitäten mit den Konzepten der Informationstheorie in Beziehung gesetzt werden, sondern auch in der Wahl der zu optimierenden informationstheoretischen Variablen.

Bei meiner Analyse hierzu werden Argumente und Herleitungen aus der Rate-Distortion-Theorie bemüht, um die Optimalität der biologischen Kanalstruktur zu zeigen. Das Kriterium des minimalen Rekonstruktionsfehlers ist dabei eine sinnvolle Operationalisierung der Annahme, dass auf frühen Stufen der visuellen Informationsverarbeitung eine informationserhaltende, verlustfreie Umkodierung stattfindet (s.a. Linsker, 1994 und Ruderman, 1994a). Es wird zunächst kurz vorgestellt und in den Kontext der Informationstheorie eingebettet.

### **2.4.1 Informationstheorie**

Ein zentrales Ergebnis der Informationstheorie besagt, dass Quellkodierung und Kanalkodierung getrennt voneinander optimiert werden können ohne dass Einbußen an der Optimalität des Gesamtsystems hingenommen werden müssen. Im Folgenden interessiert nur die (verlustbehaftete) Quellkodierung und für den Kontext biologischer Systeme lässt sich diese Aufteilung auch plausibel interpretieren, denn ein wichtiger Grund für biologische "Kanalfehler" ist in der langfristigen Instabilität oder auch im kompletten Verlust von Neuronen zu sehen. Die passende Redundanz zum Auffangen solcher Fehler, wird mit einer Mehrfachauslegung von Neuronen oder auch mit einer Art Populationskodierung erreicht, bei der die Nervenzellen leicht variierende, aber überlappende Tuningfunktionen besitzen (siehe z.B. Hinton et al., 1986, Snippe, 1996). Bei einer einfachen Verdoppelung von neuronalen Einheiten auf jeder Verarbeitungsstufe würden dagegen schnell kombinatorische Probleme auftauchen.



Eine andere Möglichkeit zur Implementierung eines Fehlerschutzes würde darin bestehen, sehr breitbandige Einheiten zu verwenden. Dieses Vorgehen hätte jedoch den Nachteil, dass die Diskriminierungsfähigkeit des Systems darunter leiden würde. Daher scheint es vernünftig, zuerst im Sinne einer Quellkodierung Tuningfunktionen zum Ausnutzen der Eingangsredundanz zu entwerfen und dann als Kanalkodierung beispielsweise die Abtastrate zu erhöhen oder ein Überlappen zuzulassen, um so das System vor Kanalfehlern zu schützen. Beim Input des Kanalkodierers wird dabei meist davon ausgegangen, dass dieser keine Redundanzen mehr aufweist, da es bereits die Aufgabe des Quellkodierers war, das quantisierte Signal  $\hat{r}(\tilde{x})$  in einen Bitstrom (idealerweise) ohne jegliche Redundanzen zu kodieren. Der Kanalkodierer fügt nun in diesen Bitstrom gezielt Redundanzen ein, die an das Rauschen und an die Beschränkungen des Kanals so angepasst sind, dass der Kanaldekodierer den Input mit vernachlässigbarem Fehler rekonstruieren kann.

Das Quellkodierungstheorem gibt nun zusammen mit der Rate-Distortion-Theorie an, welche Bitrate zur Übertragung eines Signals aufgewendet werden muss und welche Techniken dabei anzuwenden sind. Das Standardwerk hierzu ist Berger (1971), speziell für den Kontext der Bildkodierung siehe aber auch Davisson (1972). Die Rate-Distortion-Funktion stellt für eine gegebene Signalstatistik den funktionalen Zusammenhang zwischen Übertragungsrate  $R$  und entstehender Verzerrung  $D$  her (vgl. Abbildung 11). Im Quellkodierungskontext ist die Quelle in der Regel ein Analogsignal, die Verarbeitung ein einfacher Filter oder eine komplexere Struktur wie z.B. eine Blocktransformation oder eine Multikanalzerlegung. Das Rauschen kommt durch die Quantisierung des Kodiererergebnisses in das System. Das Quellkodierungsproblem besteht nun darin, die statistischen Redundanzen der Quelle so auszunutzen, dass eine hinreichend getreue Rekonstruktion des ursprünglichen Signals möglich ist. Dabei ist mit der minimalen Informationsrate  $R$  auszukommen, die erforderlich ist, um das quantisierte Signal zu kodieren.

Für Quellen mit einem diskreten Alphabet besagt das Theorem über die verlustfreie Kodierung, dass es möglich ist, die Quelle ohne Verzerrung mit einer Rate zu kodieren, die beliebig nahe an der Quellentropie liegt. Die informationstheoretische Grenze einer verlustfreien

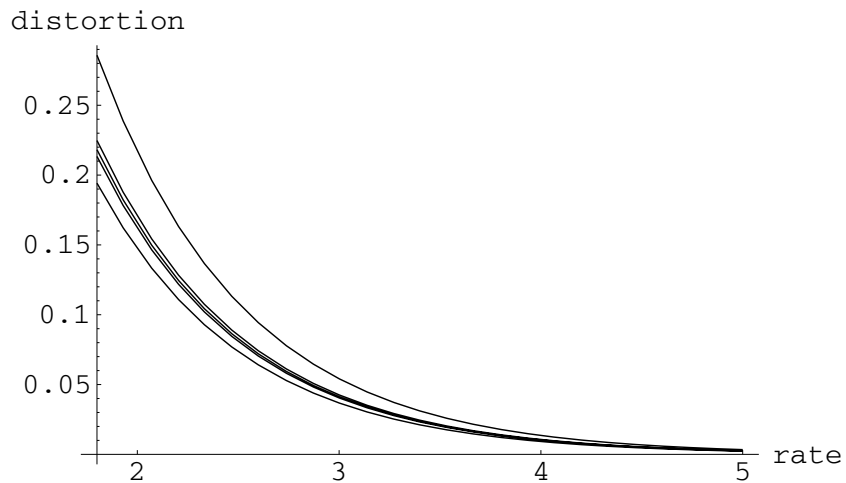


Abbildung 11: Graph mehrerer Rate-Distortion-Funktionen zwischen PCM-Kodierung (oberste Kurve) und absolute Grenze für Teilbandkodierer (untere Kurve). Detail in Abbildung 20. [bit/pixel]

Kodierung ist also durch die Entropie der Quelle bestimmt (Shannon, 1948). Für die wichtige Klasse der Gaußprozesse ist hier sogar eine analytische Lösung für  $R(D)$  bekannt.

Um die Rate-Distortion-Grenze tatsächlich zu erreichen, müssen wir in der Lage sein, die Wahrscheinlichkeitsdichtefunktion des Signals zu transformieren und die Kanalform so zu gestalten, dass der Kanal an das Signal angepasst ist. Das ist mit keinem linearen System möglich, das aus nur einem Kanal besteht, kann aber erreicht werden mit Systemen, die den Eingangsdatenstrom in mehrere Teilkanäle zerlegen, also mit Multikanalzerlegungen.

Eine zentrale Rolle beim Design moderner Kodierungsverfahren spielt der Algorithmus zur Bitzuweisung. Dieser versucht die durch eine Gesamtbitrate beschränkten Bits in einer möglichst effizienten Weise auf die Teilbänder zu verteilen. Der erste Schritt besteht deshalb darin, eine optimale Bitzuweisung für die Teilbandkodierer herzuleiten.

## 2.4.2 Herleitung der optimalen Ressourcenaufteilung

Die Herleitung einer optimalen Zuweisung wird hier für den allgemeinen Fall der beliebigen Teilbandzerlegung gegeben. Sie ist an Jayant, Noll (1984) angelehnt, dort wird jedoch nur der Fall einer linearen Aufteilung in Frequenzbänder gleicher Breite gezeigt. Die folgenden Formeln für Bitzuweisung, Gesamtfehlersignal und Vorteil gegenüber Vollbandkodierung stellen damit eine wichtige Generalisierung dar. Gesucht ist hier also für eine vorgegebene Informationsrate diejenige Strategie, die nicht nur für beliebige Eingangssignale, sondern auch für beliebige Bandaufteilungen zu dem kleinstmöglichen Gesamtfehler führt.

Die Datenübertragungsrate, die auch als geeignetes Maß zur Beschreibung des biologischen Gesamtaufwands interpretiert werden kann, ist gegeben mit  $I = 2WR$ , wobei  $W$  die Bandbreite des Signals ist und

$$R = \sum_{k=1}^M R_k \bar{w}_k$$

die Bitrate als Funktion der gewichteten Einzelkanalraten beschreibt. Die Übertragungsrate  $R$  des Gesamtsystems setzt sich also aus den kanalspezifischen Bitraten  $R_k$  und den kanalspezifischen Samplingraten  $w_k$  zusammen. Die Samplingrate eines Teilbands beträgt  $2 \cdot \Delta w_k$  mit  $\Delta w_k = W \bar{w}_k$  und  $\sum \bar{w}_k = 1$ . Die Gesamtverzerrung

$$\sigma_d^2 = \sum_{k=1}^M \sigma_{d,k}^2$$

setzt sich aus den kanalweisen Verzerrungen der PCM-kodierten Teilbänder zusammen:

$$\sigma_{d,k}^2 = \varepsilon^* 2^{-2R_k} \sigma_{x,k}^2$$

Die Konstante  $\varepsilon^*$  beschreibt dabei den Signalverlauf innerhalb eines Kanals. Diese wird eins bei einer Gleichverteilung und ein

Umschreiben macht deutlich, dass die Rate direkt interpretiert werden kann als das Signal-Rausch-Verhältnis des Kanals:

$$R_k = \frac{1}{2} \text{ld} \left( \frac{\sigma_{x,k}^2}{\sigma_{d,k}^2} \right)$$

Die optimale Bitzuordnung ist nun diejenige, die  $\sigma_d^2$  unter der Randbedingung

$$(I = \text{const}) \Rightarrow \sum_{k=1}^M R_k \bar{w}_k = \text{const}$$

minimiert. Die optimale Bitzuweisung kann jetzt mit der Methode der Lagrange'schen Multiplikatoren gefunden werden, indem gesetzt wird:

$$\left. \frac{\partial}{\partial R_k} \left( \sigma_d^2 - \lambda \left( R - \sum_{k=1}^M R_k \bar{w}_k \right) \right) \right| = 0$$

Die optimale Bitzuweisung ergibt sich dann zu:

$$R_{k, \text{opt}} = R + \frac{1}{2} \text{ld} \left( \frac{\sigma_{x,k}^2}{w_k} \right) - \frac{1}{2} \sum_{l=1}^M \left( \text{ld} \left( \frac{\sigma_{x,l}^2}{w_l} \right)^{\bar{w}_l} \right) \quad (1)$$

Die im Sinne einer optimalen Aufteilung für einen Kanal  $k$  zu wählende Bitzahl hängt hier logarithmisch von der invers gewichteten Energie in diesem Kanal ab. Durch die Vorschrift (1) werden also die Bits proportional zur Signalleistung vergeben, die Bitrate wird hiermit an die Statistik der Quelle angepasst (siehe Abbildung 12).

Die Frage, ob die Ortsfrequenzaufteilung des biologischen Systems im Sinne einer optimalen Ressourcenzuweisung verstanden werden kann, kann jedoch erst geklärt werden, wenn wir Informationen über die Energieverteilung des Eingangssignals haben. Zuvor sollen jedoch die Vorteile einer Teilbandkodierung bei beliebiger Partitionierungsstruktur formal erfasst werden.

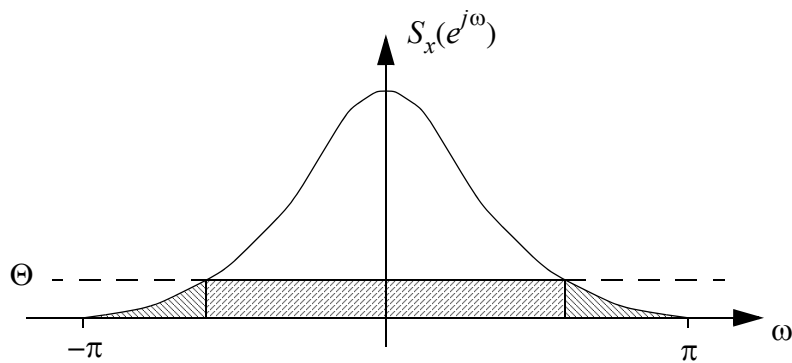


Abbildung 12: Das Rate-Distortion-Optimum für eine korrelierte Gauss-Quelle führt dazu, dass das resultierende Rauschniveau als Ergebnis einer “water-filling” Prozedur interpretiert werden. Die Oberfläche ergibt sich aus  $S_{xx}(e^{j\omega})$ , ein niedriges Rauschniveau erfordert jedoch ein weißes Rauschspektrum.

### 2.4.3 Vorteil einer mehrkanaligen Repräsentation

Der Gesamtfehler des Systems bei optimaler Bitzuweisung ergibt sich durch Einsetzen von (1):

$$\sigma_{d,min}^2 = \varepsilon^* 2 \sum_{k=1}^M \left( \sigma_{x,k}^2 2^{-2R} 2^{ld\left(\frac{\prod_{l=1}^M \left(\frac{\sigma_{x,l}^2}{w_l}\right)^{\bar{w}_l}}{\sigma_{x,k}^2}\right)} \right)$$

Als minimale Verzerrung erhält man:

$$\sigma_{d,min}^2 = \varepsilon^* 2 2^{-2R} \prod_{k=1}^M \left( \frac{\sigma_{x,k}^2}{w_k} \right)^{\bar{w}_k} \quad (2)$$

Der Einzelkanalfehler ist damit proportional zur Bandbreite  $\sigma_{d,k,min}^2 = \bar{w}_k \cdot \sigma_{d,min}^2$ , d.h. die optimale Bitzuweisung führt zu einer konstanten Rauschamplitude.

Der Kodiergewinn gibt an, um welchen Faktor sich die Verzerrung bei gleicher Rate durch Anwendung der Frequenzbereichskodierung gegenüber einer PCM-Vollbandkodierung verringern lässt. Der Vorteil einer linearen Aufteilung in gleichbreite Kanäle ist bekanntlich gerade das Verhältnis zwischen arithmetischem und geometrischem Mittel der Kanalvarianzen. Der Vorteil einer beliebigen Aufteilung gegenüber Vollbandkodierung ist nun das Verhältnis zwischen *gewichtetem* arithmetischem und *gewichtetem* geometrischem Mittel der Einzelkanalvarianzen:

$$G_{SBC} = \frac{\sigma_d^2}{\sigma_{d, min}^2} = \frac{\sigma_x^2}{\prod_{k=1}^M \left( \frac{\sigma_{x,k}^2}{\bar{w}_k} \right)^{\bar{w}_k}} = \frac{\sum_{k=1}^M \left( \frac{\sigma_{x,k}^2}{\bar{w}_k} \cdot \bar{w}_k \right)}{\prod_{k=1}^M \left( \frac{\sigma_{x,k}^2}{\bar{w}_k} \right)^{\bar{w}_k}} \quad (3)$$

Durch Teilbandkodierung wird stets ein Vorteil erzielt, da das gewichtete arithmetische Mittel in aller Regel größer als das gewichtete geometrische Mittel ist. Gleichheit gilt genau dann, wenn die gewichteten Einzelkanalvarianzen identisch sind, denn dies führt gerade dazu, dass für alle Teilbänder gilt:  $R_{k, opt} = R$ . Dieser Gewinn gilt natürlich nur für ideale Filter. Mit physikalisch realisierbaren Filtern ist nur eine suboptimale Lösung erreichbar (siehe z.B. Fischer, 1992).

Der Gewinn ist hier umso größer, je ungleichmäßiger die Verteilung der Energie auf die einzelnen Koeffizienten ist. Das ist z.B. bei einer Energiekonzentration auf wenige Komponenten der Fall. Welcher Gewinn in einer gegebenen Situation tatsächlich erreicht werden kann, ist von mehreren Faktoren abhängig: dem Leistungsdichtespektrum des Eingangssignals, der Kanalanzahl sowie der Art der Zerlegung entlang der Ortsfrequenzachse.

## **2.5 Anpassung an die Eingangsstatik**

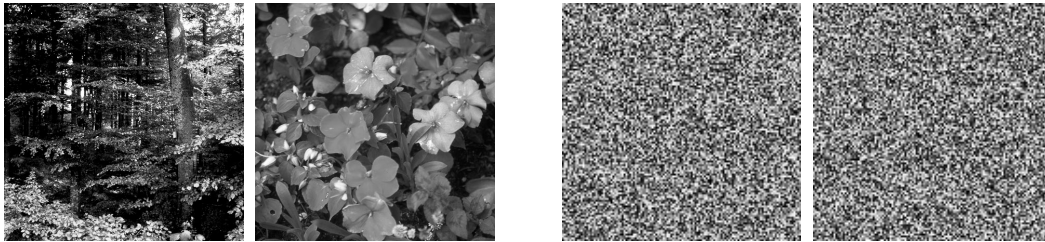
Zur Optimierung technischer Bildkodierverfahren und auch zur Beurteilung biologischer Verarbeitungsstrategien ist es notwendig, Aussagen über die Beschaffenheit des verwendeten Bildmaterials zu machen. Da keine optimale Kodierung einzelner Bilder im Sinne etwa einer inhaltsorientierten Zerlegung angestrebt ist, sondern eine Kodierung, die an die statistischen Regelmäßigkeiten einer großen Menge unterschiedlicher Eingangssignale angepasst ist, sollen hier stationäre, statistische Bildmodelle verwendet werden. Die Bilder, mit denen das visuelle System des Menschen aus evolutionärer Sicht typischerweise konfrontiert ist, konstituieren die Menge der natürlichen Bilder.

### **2.5.1 Statistik natürlicher Bilder**

Gegeben die enorme Variabilität innerhalb einer Kategorie wie die der natürlichen Bilder, erscheint es vielleicht zunächst erstaunlich, wenn nicht sogar aussichtslos, sinnvolle Aussagen über diese Gesamtheit treffen zu wollen. Das dies dennoch möglich ist, liegt an den physikalischen Randbedingungen, die allen diesen Bildern zu Grunde liegt, und zu Phänomenen wie der Kompaktheit von Objekten führt, der Homogenität von Oberflächen und Diskontinuitäten an Objektgrenzen. So beobachtet man in natürlichen Bildern Flächen mit gleichmäßiger Grundhelligkeit, es sind aber auch plötzliche Amplitudensprünge an bestimmten lokalen Positionen vorhanden. Diese charakteristischen Eigenschaften können mit geeigneten statistischen Modellen erfasst werden.

Die Tatsache, dass die Menge der natürlichen Bilder eine im Vergleich zu der Menge aller möglichen Bilder winzige Teilmenge darstellt, erhöht den evolutionären Druck, hier eine Anpassung des visuellen Systems vorzunehmen. Dass diese Anpassung auch tatsächlich stattgefunden hat, soll Abbildung 13 illustrieren (Zetsche, Röhrbein, 2000). Die rechts in der Abbildung gezeigten Zufallsbilder sind gemessen an ihrer Euklidischen Distanz im Signalraum sich sehr viel unähnlicher, als die beiden natürlichen Bilder links. Trotzdem sind es diese, die mühelos unterschieden werden können, während das für

die Zufallsbilder nahezu unmöglich erscheint. Dies zeigt, dass die Anpassung des visuellen Systems zu einer selektiven Unterscheidungsfähigkeit innerhalb einer hochspeziellen Teilmenge aus der Menge aller möglichen Bilder führt.



*Abbildung 13:* Die euklidische Distanz zwischen Zufallsbildern ist größer als die zwischen natürlichen Bildern, obwohl diese sehr viel leichter voneinander zu unterscheiden sind.

Zur Erfassung der charakteristischen Eigenschaften natürlicher Bilder gibt es prinzipiell zwei Möglichkeiten: Zum einen kann man sich der hochdimensionalen Wahrscheinlichkeitsdichtefunktion mit geeigneten Annahmen nähern und diese dann empirisch testen. Alternativ können die statistischen Abhängigkeiten zwischen Bildpunkten mit Hilfe von Korrelationen erfasst werden. Bei diesem Ansatz werden die Wahrscheinlichkeitsdichtefunktionen durch eine Potenzreihe approximiert, deren Koeffizienten die Momente bzw. Kumulanten sind.

Für die Analyse hier stellt sich nun die Frage, welche statistischen Eigenschaften zur Erklärung der biologischen Teilbandkodierung herangezogen werden sollen. Einige Merkmale des visuellen Systems wie die laterale Inhibition konnten bereits mit der Statistik zweiter Ordnung erklärt werden (Laughlin, 1981 und Srinivasan et al., 1982), für andere Eigenschaften kortikaler Neurone, wie die Orientierungsselektivität, wurde dagegen die Bedeutung von Statistiken höherer Ordnung betont, da in diesen wesentliche Informationen über Phasenkopplungen enthalten sind (Zetsche et al., 1993). Die Statistik zweiter Ordnung liefert nur für Gaußsche Signale eine vollständige Beschreibung, während wesentliche Strukturen natürlicher Bilder nur von den Momenten und Kumulanten höherer Ordnung erfasst werden können.



In der Tat gibt es Argumente, die nahe legen, dass auch lokalisierte rezeptive Felder niemals mit einem Ansatz, der nur die Zweipunkt-Statistik benutzt, entstehen können (Olshausen, Field, 1996).

Zunächst ist daher zu klären, ob die in Kapitel 2 beschriebene Zerlegung des retinalen Bildsignals in ungefähr selbstähnliche Bandpässe innerhalb des etablierten Standardrahmens der linearen Systemtheorie erklärt werden kann. Da gemäß Occam's Rasiermesser die Systemkomplexität nie ohne Notwendigkeit erhöht werden sollte, ist zunächst zu überprüfen, ob nicht bereits mit einem linearen Ansatz das beobachtete Phänomen hinreichend gut beschrieben werden kann. Es wird sich deshalb im Folgenden auf die Autokorrelationsfunktion bzw. das Leistungsdichtespektrum natürlicher Bilder beschränkt.

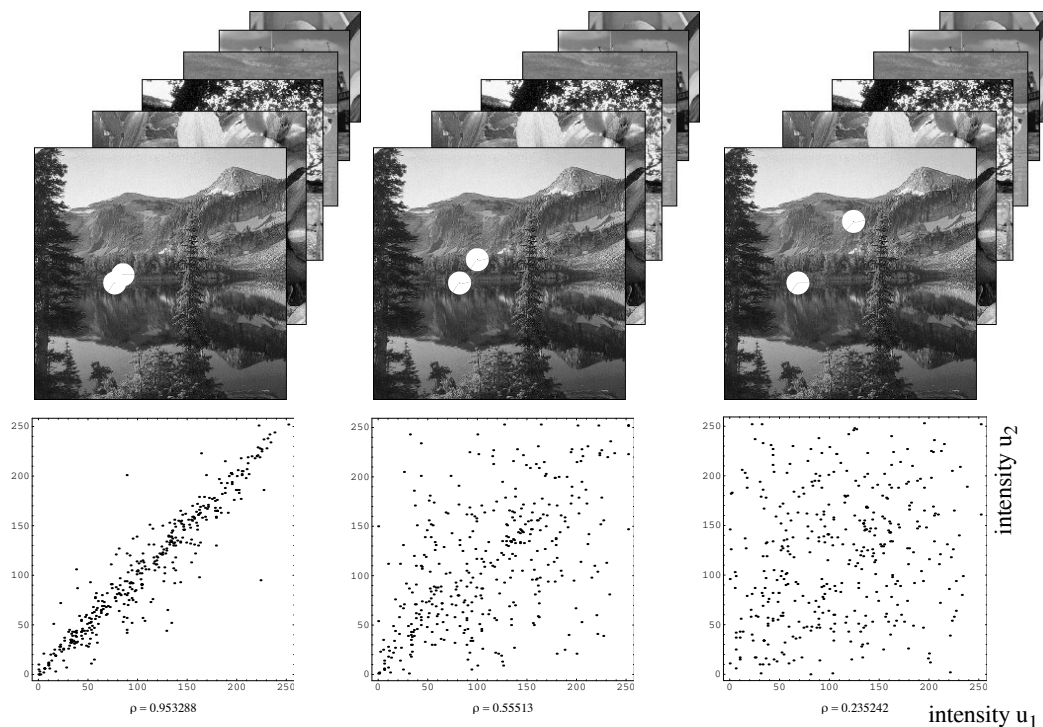
### 2.5.2 Das Leistungsdichtespektrum natürlicher Bilder

Statistiken zweiter Ordnung beschreiben die statistischen Beziehungen zwischen Paaren von Pixeln, indem sie die Momente 2. Ordnung der bivariaten Wahrscheinlichkeitsdichtefunktion dieser Paare berechnen. Mit der Berechnung der Momente für alle möglichen räumlichen Beziehungen erhält man die Autokorrelationsfunktion. Werden die Intensitäten zweier Bildpunkte als Zufallsgrößen  $u_1(x_1, y_1)$  und  $u_2(x_2, y_2)$  mit der Wahrscheinlichkeitsdichtefunktion  $f_{u_1 u_2}(u_1, u_2)$  interpretiert, gilt für die Autokorrelationsfunktion:

$$c^u(\Delta x, \Delta y) = E(u(x, y) \cdot u(x + \Delta x, y + \Delta y))$$

Die Autokorrelationsmatrix  $A = E[(\vec{x})(\vec{x})^{*T}]$  und die Autokovarianzmatrix  $C = E[(\vec{x} - E[\vec{x}])(\vec{x} - E[\vec{x}])^{*T}]$  sind Beschreibungen der Momente 2. Ordnung aller möglichen Paare  $(x_i, x_j)$ , und unterscheiden sich lediglich bezüglich ihres Mittelwertes. Eine Vereinfachung wird durch die Stationaritätsannahme erzielt, also der Annahme, dass die Statistik nicht von den absoluten Pixelpositionen abhängig ist, sondern nur von dem Abstand zwischen ihnen  $\Delta x = x_j - x_i$ . Die Autokorrelationsfunktion  $acf(\Delta x) = E[x_i \cdot x_{i + \Delta x}]$  stellt dann eine vollständige Beschreibung der Variation der Korrelation zweiter Ordnung in Abhängigkeit von der räumlichen Trennung dar. Sie ist damit ein

Maß für die Reichweite der statistischen Bindungen zwischen Bildpaaren (siehe Abbildung 14).



*Abbildung 14:* Korrelation zwischen zwei benachbarten Bildpunkten. Die Korrelation ist für unmittelbar benachbarte Pixel sehr hoch und wird mit zunehmendem Abstand systematisch kleiner. Die Streudiagramme unten geben die Verbundaktivität von Bildpaaren aus natürlichen Bildern an. Die räumliche Trennung zwischen diesen Paaren entspricht jeweils dem Abstand der oben markierten Bildpositionen.

Äquivalent dazu ist als Beschreibung im Frequenzbereich das Leistungsdichtespektrum  $C^U(f_x, f_y)$ , das nach dem Theorem von Wiener-Chintschin mit der Autokorrelationsfunktion  $c''(\Delta x, \Delta y)$  ein Fourierpaar bildet. In vielen Fällen ist es gebräuchlicher, die Beschreibung der Korrelationsstruktur mit Ortsfrequenzkomponenten auszudrücken. Wenn Schwankungen in den spektralen Amplituden auftreten, ist das ein Zeichen für statistische Redundanzen und diese können mit linearen Systemen gezielt ausgenutzt werden. Ist das Spektrum

konstant ("weiß"), kann durch Umkodieren dagegen kein Gewinn erzielt werden.

### 2.5.3 Der 1/f-Abfall

Bereits in den 60er Jahren wurden empirische Bestimmungen der Autokorrelationsfunktionen und Leistungsdichtespektren von Bildern vorgenommen (z.B. O'Neal, 1966, Habibi, Wintz, 1971). Diese pixelbasierten Messungen haben gezeigt, dass massive Korrelationen zwischen den Bildpunkten bestehen (Harrison, 1952), dass diese Korrelationen mit abnehmendem Pixelabstand systematisch zunehmen (Schreiber, 1956) und dass sie sich über 10 Pixel und mehr erstrecken können. So wurde der Abfall von 6dB pro Oktave als ein typisches Charakteristikum von Fernsehsignalen benutzt (Deriugin, 1956, Conner et al., 1972, Abbildung 15). Abweichungen vom  $1/f^2$  Verhalten können im flacheren, tieffrequenten Bereich und im steileren Abfall des Hochfrequenzbereichs festgestellt werden.

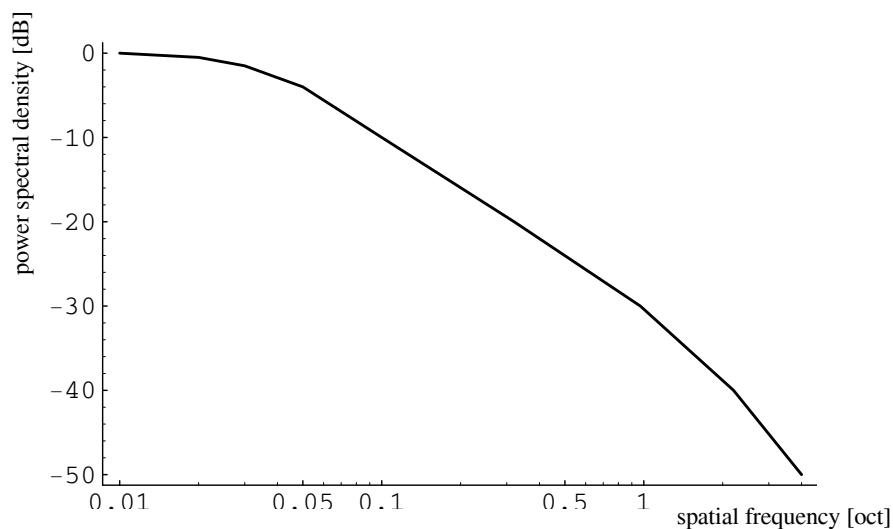


Abbildung 15: Typisches Leistungsdichtespektrum von Videosignalen mit einem Abfall von etwa 6dB/Oktave. Dies entspricht einem  $1/f^2$ -Abfall. Abbildung aus Conner et al., 1972

Weitere Untersuchungen bekräftigten dieses Ergebnis (Field, 1987, Huang, Turcotte, 1990, van Hateren, 1992). In neuerer Zeit wurden dann eine Reihe von weiteren systematischen statistischen Untersuchungen mit großen Bildermengen durchgeführt. Das am bes-

ten etablierte Ergebnis dieser Studien ist, dass der Abfall des Leistungsdichtespektrums mit einer Potenzfunktion beschrieben werden kann:

$$C^U(f_x, f_y) = \frac{c}{(\sqrt{f_x^2 + f_y^2})^\alpha} \quad (4)$$

Typischerweise nimmt der Exponent  $\alpha$  einen Wert von 2 an, womit sich ein Abfall von  $1/f_r^2$  in radialer Richtung ergibt.

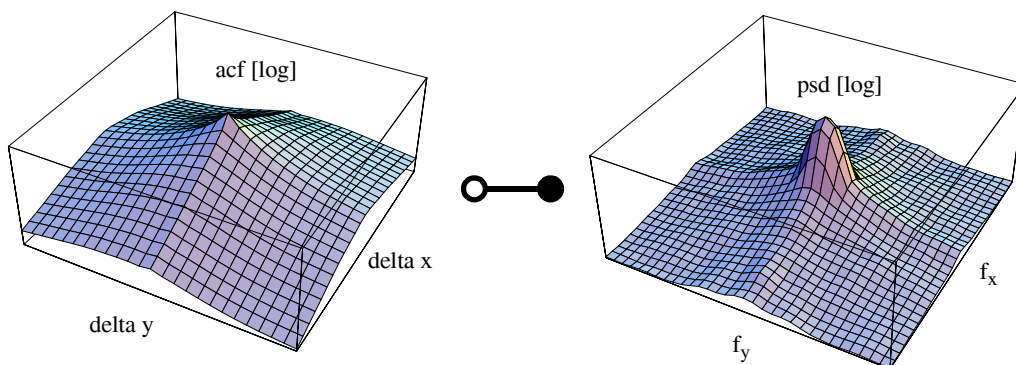


Abbildung 16: Autokorrelationsfunktion und Leistungsdichtespektrum natürlicher Bilder (Bilder aus der van Hateren-Datenbank). Die Statistik zweiter Ordnung zeigt den typischen starken Abfall in Radialrichtung und die ungefähre Konstanz bezüglich der Orientierung.

Eine Autokorrelationsfunktion und das dazugehörige Leistungsdichtespektrum ist als Ergebnis einer Mittelung über eine große Anzahl natürlicher Bilder in Abbildung 16 zu sehen. Auffallend bei diesen beiden äquivalenten Charakterisierungen sind die beträchtliche Variation in radialer Richtung und die (näherungsweise) Unabhängigkeit vom Orientierungswinkel. Im Fourierraum erscheinen die weitreichenden statistischen Abhängigkeiten in Form einer Konzentration spektraler Energie bei tiefen Ortsfrequenzen, während hochfrequente Ereignisse im Sinn von abrupten Luminanzänderungen nur wenig Energie aufweisen.

Die Statistik zweiter Ordnung von natürlichen Bildern kann also gut mit einem "selbstähnlichen"  $1/f^2$  Leistungsdichtespektrum beschrieben werden, während die Bandpasszerlegung in biologischen Sehsystemen durch eine annähernd "selbstähnliche" oktavbreite Strukturierung der Ortsfrequenzkanäle charakterisiert ist. Aufgrund dieser Parallele wurde bisher vermutet, dass möglicherweise eine systematische Beziehung zwischen diesen beiden Eigenschaften besteht. Dabei wurde von einer exakten Selbstähnlichkeit der biologischen Zerlegung ausgegangen, aber auch für diese Näherung hat bisher eine formale Herleitung gefehlt. Dies ist nun Gegenstand des folgenden Kapitels.

## 2.6 Die Oktavzerlegung

### 2.6.1 Selbstähnlichkeit

Allgemein ist das  $1/f$ -Verhalten bei Amplituden bzw. das  $1/f^2$ -Verhalten bei Leistungen ein in der Natur häufig beobachtetes Phänomen: Man findet es z.B. bei der Sonnenaktivität, bei Erdbeben, Waldbränden, Strömungen und Lawinen. Zur Beschreibung solcher Phänomene ist die Theorie der *self-organizing criticality* vorgeschlagen worden (Bak et al., 1987), die einen allgemeinen Mechanismus zur Emergenz skalenfreier Strukturen postuliert.

Bei Bildern mit dieser Eigenschaft wird aufgrund ihres Verhältnisses zur fraktalen Geometrie (Peitgen et al., 1992) oft von Selbstähnlichkeit gesprochen. In der Tat weisen natürliche Bilder über die Skalen hinweg bestimmte Eigenschaften einer Selbstähnlichkeit auf (Ruderman, 1994b). Diese Selbstähnlichkeit ist natürlich nicht vollständig und exakt, sondern partiell und stochastisch. Auch sollte eher von Selbstaffinität gesprochen werden, da zur Bildtransformation affine Abbildungen verwendet werden müssen, die meist nicht längen- und winkeltreu sind.

Der Grund dafür, dass Ausschnitte natürlicher Szenen in der Regel zu ihren isotropen Reskalierungen nicht kongruent sind, liegt an den physikalischen Bedingungen. Z.B. sind topographische Profile durch Tektonik, Erosion und Ablagerungsbedingungen in einem gerichteten Erdschwerefeld gekennzeichnet. Küstenlinien etwa sind nur dann selbstähnlich, wenn sie sich unter Bedingungen gebildet haben, die nicht durch Vorzugsrichtungen bestimmt sind. In der Regel ist ihre Form aber durch Meeresströmungen oder tektonische Strukturen geprägt und sie sind damit (nur) selbstaffin.

Es scheint aber durchaus plausibel, dass die ungefähr selbstähnliche Struktur der biologischen Zerlegung gut auf das in etwa selbstähnliche Leistungsdichtespektrum natürlicher Bilder passt: In Abbildung 17 ist ein  $1/f$  Signal mit einer äquidistanten Aufteilung (oben) und einer oktavbreiten (unten) gezeigt. Die dekorrelierende Wirkung ist dabei unterschiedlich groß: Im ersten Fall ist die spektrale Flachheit in den hohen Frequenzbändern groß, die niedrigen Fre-

quenzbänder enthalten jedoch noch sehr starke Korrelationen. Werden die Grenzfrequenzen dagegen entsprechend der Signalvarianz gewählt, ist die dekorrelierende Wirkung der Signalaufteilung deutlich günstiger auf die Teilbänder verteilt.

Ein formales Argument fehlt zur Klärung der Frage, in welcher Weise damit die statistischen Abhängigkeiten optimal ausgenutzt werden. Es wird allerdings darauf hingewiesen, dass diese Zerlegung als Antwort auf natürliche Bilder zu einer Energiegleichverteilung in den Teilfiltern führt (Field, 1987).

Ich zeige im nächsten Abschnitt mit Rate-Distortion-Argumenten, dass eine strikt selbstähnliche Filteraufteilungen in der Tat zu einer effizienteren Anpassung an das  $1/f_r^2$  Leistungsdichtespektrum ("psd") natürlicher Bilder führt, als eine lineare Aufteilung.

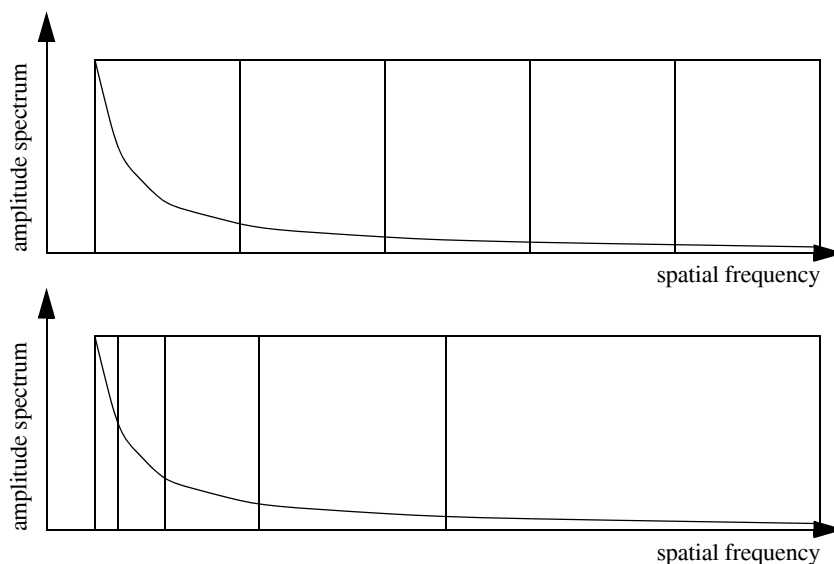


Abbildung 17: Vergleich der dekorrelierenden Wirkung einer konstant linearen Bandbreite (oben) und einer konstant logarithmischen Bandbreite (unten) für ein  $1/f$  Signal.

## 2.6.2 Lineare vs. selbstähnliche Zerlegung: Herleitung

Die vorgestellte Formel zum Rekonstruktionsfehler bei optimaler Ressourcenverteilung (2) gilt für beliebige Zerlegungsschemata mit M

Teilbändern. Von besonderem Interesse ist hier der Vergleich zwischen der einfachen linearen Zerlegung und der "biologischen" logarithmischen Zerlegung. Da nun die Energieverteilung des Eingangssignals bekannt ist und die Störungen in den einzelnen Frequenzbändern nur noch von der Zerlegung abhängen, können die verschiedenen Schemata hinsichtlich ihrer Leistung miteinander verglichen werden.

In beiden Zerlegungen ist die Varianz für ein Teilband  $k$ , das von der Frequenz  $f_{k-1}$  bis  $f_k$  reicht

$$\sigma_{x,k}^2 = \ln(f_k/f_{k-1}) \quad .$$

Um Probleme mit dem Ursprung der Frequenzachse zu vermeiden, nehmen wir eine untere Grenzfrequenz  $f_0$  und einen Skalar  $\beta$  an, so dass insgesamt der Bereich  $[f_0, \beta f_0]$  mit Bandbreite  $W = (2^M - 1)f_0 = (\beta - 1)f_0$  kodiert werden muss.

Bei der linearen Zerlegung besitzen alle Teilbänder dieselbe Breite, so dass  $\bar{w}_k = 1/M$ , und die minimale Verzerrung, die bei der Verarbeitung von  $1/f^2$  Signalen durch eine lineare Zerlegung bei fester Gesamtbitrate, gegebener Anzahl von Kanälen und Gesamtbandbreite  $W$  erreicht werden kann, ist

$$\sigma_{d,min}^2 = \frac{\varepsilon^{*2}}{2^{2R}} \cdot M \cdot \left( \prod_{k=1}^M \ln \frac{M + (\beta - 1)k}{M + (\beta - 1)(k - 1)} \right)^{\frac{1}{M}} \quad (5)$$

Wenn die Teilbänder dagegen selbstähnlich sind, ist das Verhältnis von oberer und unterer Grenzfrequenz für alle Teilbänder konstant. Das führt zu einer konstanten Teilbandenergie.

$$\sigma_{x,k}^2 = (\ln \beta) / M \quad .$$

Da die Gewichte als  $\bar{w}_k = (f_k - f_{k-1}) / W$  definiert sind, betragen sie für die selbstähnlichen Teilbänder



$$\bar{w}_k = \left( \beta^{\frac{k}{M}} - \beta^{\frac{k-1}{M}} \right) / (\beta - 1) \quad .$$

Die minimale Verzerrung für die Verarbeitung von  $1/f^2$  Signalen mit einer selbstähnlichen, logarithmischen Zerlegung und fester Rate ergibt sich dann zu

$$\sigma_{d, min}^2 = \frac{\varepsilon^{*2}}{2^{2R}} \cdot \ln \left( \beta^{\frac{1}{M}} \right) \cdot \frac{\beta - 1}{\beta^{\frac{1}{M}} - 1} \cdot \prod_{k=1}^M \left( \beta^{\frac{1-k}{(\beta-1)M} \left( \beta^{\frac{k}{M}} - \beta^{\frac{k-1}{M}} \right)} \right) \quad (6)$$

Für den biologisch motivierten Fall einer 1-Oktaven Zerlegung kann diese minimale Verzerrung geschrieben werden als:

$$\sigma_{d, min}^2 = \frac{\varepsilon^{*2}}{2^{2R}} \cdot \ln 2 \cdot (2^M - 1) \cdot \prod_{k=1}^M \left( \frac{1}{2^{k-1}} \frac{2^{k-1}}{2^M - 1} \right) \quad (7)$$

### 2.6.3 Lineare vs. selbstähnliche Zerlegung: Komplexitätsvergleich

Um Zerlegungsschemata unter realistischen Bedingungen vergleichen zu können, soll im Folgenden die Komplexität der Teilbandzerlegung berücksichtigt werden. Diese kann als weitere Dimension im "tradeoff space" von Kodierverfahren und Kompressionsschemata betrachtet werden (siehe z.B. Bhaskaran, Konstantinides, 1995).

Für Anwendungen von praktischem Interesse ist eine wichtige Frage, welche Leistung mit einer vernünftigen Anzahl von Kanälen erreicht werden kann. Die Kanalanzahl wird deshalb nun als Komplexitätsmaß benutzt, um einen Vergleich zwischen zunächst linearer

und logarithmischer Zerlegungen durchzuführen. Zu kodieren ist stets ein  $1/f$ -Signal.

Mit (5) und (7) haben wir bereits die Distortion-Rate-Funktionen, von denen wir ableiten können, wie die Verzerrung in Abhängigkeit von Zerlegungstyp und Teilbandzahl variiert. Um die beiden Zerlegungstypen bezüglich ihrer Komplexität vergleichen zu können, ist es nützlich, die Inverse dieser Funktionen zu betrachten. Es ist dann zu fragen, wie viele Teilbänder erforderlich sind, um eine gewünschte Qualität zu erreichen. Die Ergebnisse werden mittels Complexity-Distortion-Funktionen dargestellt. Diese geben jeweils für ein Zerlegungsschema den Zusammenhang zwischen der minimal erforderlichen Kanalanzahl und der tolerierbaren Verzerrung bei festgehaltener Gesamtbitrate an. Die Wahl der Bitrate ist arbiträr, wurde aber zur Veranschaulichung für die Abbildungen auf  $R=4$  bit gesetzt.

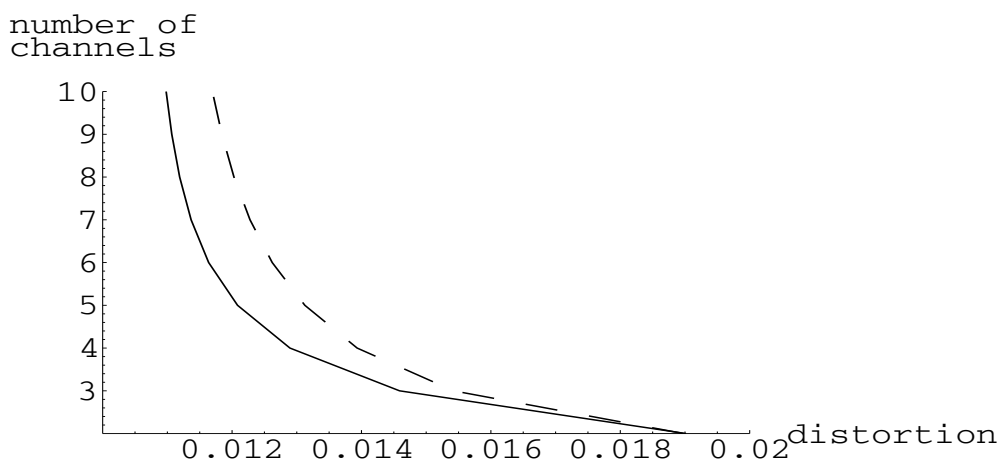


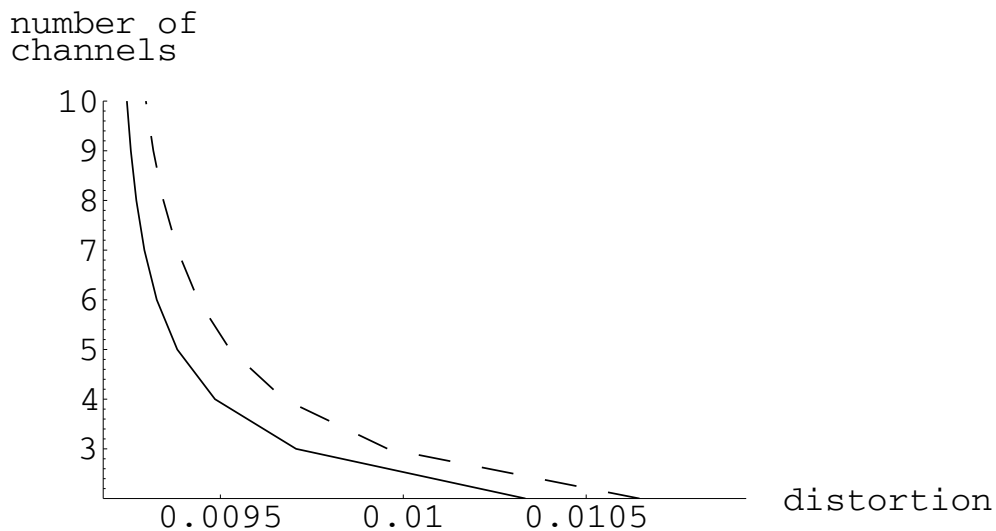
Abbildung 18: Komplexitäts-Verzerrungsfunktionen für eine lineare Zerlegung (gestrichelt) und für eine selbstähnliche, logarithmische Zerlegung (durchgezogen) bei konstanter Bitrate ( $R=4$ ,  $\beta=10^4$ ).

Eine weitere Festlegung muss bezüglich der oberen und unteren Grenzfrequenz gemacht werden. Die Beispielen hier beziehen sich auf zwei Frequenzbereiche: Ein Abschnitt umfasst 5 Oktaven von 0.5

- 16 cyc/deg und kommt damit der Ortsfrequenzempfindlichkeit des visuellen Systems nahe, als anderer Abschnitt wurde zur Kontrastierung der unterschiedlichen Zerlegungsschemata ein großer Bereich gewählt, bei dem die obere Grenzfrequenz den  $10^4$ -fachen Wert der unteren Grenzfrequenz besitzt.

Die hier resultierenden Complexity-Distortion-Funktionen sind für die angegebenen Parameterbereiche in den Abbildungen 18 und 19 gezeigt. Als Ergebnis lässt sich eine deutliche Verbesserung der Kodierleistung von natürlichen Bildern feststellen, wenn statt einer linearen Aufteilung ein selbstähnliches Bandsplitting angewendet wird. So ist ein bestimmter, zu tolerierender Verzerrungsgrad mit erheblich weniger Kanälen realisierbar, wenn eine selbstähnliche Aufteilung gewählt wird.

Für die Berechnung in Abbildung 18 wurde das Signal über einen weiten Bereich kodiert ( $\beta=10.000$ ). Wenn dieser auf ein biologisches Maß reduziert wird, zeigt sich abermals die Überlegenheit der logarithmischen Zerlegung für  $1/f$ -Signale, wie in Abbildung 19 ersichtlich.



*Abbildung 19:* Vergleich zwischen linearer (gestrichelt) und logarithmischer Aufteilung (durchgezogen) bei einem Signal zwischen 0.5 und 16 c/deg und einer Bitrate von 4 bit.

Die Funktionsverläufe zeigen hier einen klaren Vorteil für die logarithmischen Bandpassfilter, da eine gewünschte Qualität mit signifikant weniger Kanälen erreicht werden kann. Der Grund dafür ist darin zu sehen, dass die Bits proportional zur Varianz zugeteilt werden müssen, weshalb bei der linearen Aufteilung die meisten Bits den Tiefpässe zufallen. Diese sind aber stärker korreliert, als bei einer logarithmischen Zerlegung, bei der die spektrale Flachheit für alle Kanäle identisch.

An den Extrempunkten der Complexity-Distortion-Funktion, d.h. für Zerlegungen mit sehr wenig oder sehr vielen Kanälen, verringern sich die aufteilungsspezifischen Unterschiede. Das liegt zum einen daran, dass eine perfekte Kodierung mit jeder Art von Bandaufteilung erreicht werden kann, wenn nur die Kanalanzahl hoch genug gewählt wird und weil sich bei z.B. einer Zweibandzerlegung die Abweichung der logarithmischen Aufteilung von dem tatsächlichen Optimum besonders bemerkbar macht (s.u.).

Unklar ist nun noch, ob die logarithmische Aufteilung bereits die beste aller möglichen Teilbandzerlegungen für das Enkodieren natürlicher Bilder darstellt. In der Tat ist dies nicht der Fall und im nächsten Abschnitt werden Ergebnisse der optimalen Teilbandzerlegung präsentiert.

## 2.7 Die Optimalzerlegung

Eine optimale Lösung hängt stets von den jeweiligen Randbedingungen ab. Gibt es z.B. keine Ressourcenbeschränkung bezüglich der Komplexität der Zerlegung, lautet die optimale Strategie, das Eingangssignal in möglichst viele Teilbänder zu zerlegen, da damit die spektrale Flachheit innerhalb der Teilbänder maximiert und die Korrelationen minimiert werden. Solange die Teilbänder hinreichend schmal sind, spielt bei dieser Betrachtung die genaue Art der Zerlegung überhaupt keine Rolle. Die absolute Rate-Distortion-Grenze kann im Allgemeinen auch nur mit einer unbegrenzten Anzahl von Kanälen erreicht werden. Ausgenommen sind hier Spezialfällen, etwa in denen das Spektrum Frequenzbereiche mit konstanter Energie enthält. Im Allgemeinen gilt aber für den Gewinn:

$$G_{max} = \lim_{M \rightarrow \infty} G_{SBC, M} \quad (8)$$

Ganz anders stellt sich die Situation dar, wenn von einer Ressourcenbeschränkung ausgegangen wird, zum Beispiel hinsichtlich der Anzahl der realisierbaren Kanäle: Eine sinnvolle Restriktion könnte hier etwa darin bestehen, dass die Anzahl der Kanäle deutlich unter der Anzahl der zu kodierenden Samples liegen sollte, dass also die Signale der 11 Millionen Axone des optischen Nerven mit nicht viel mehr als z.B. 10 Kanälen zu enkodieren sind. Diese realistischen Rahmenbedingungen haben dann unausweichlich eine Abweichung von der spektralen Flachheit in jedem Teilband zur Folge.

Die Frage lautet dann, wie die Kanalbandbreiten zu arrangieren sind, um insgesamt eine maximale Dekorrelation des Systems zu erreichen. Bezüglich dieser Frage haben wir bereits zwei alternative Strategien miteinander verglichen: eine Zerlegung mit konstanten absoluten Bandbreiten und eine mit konstanten relativen Bandbreiten. Unter der Annahme einer optimalen Bitallokation auf die einzelnen Teilbänder wird jetzt nach der optimalen Teilbandzerlegung zur Minimierung der Gesamtverzerrung gesucht.

Da dieses Problem komplizierter analytisch zu lösen ist, wurde hier eine numerische Optimierungstechnik verwendet, bei der für eine vor-

gegebene Anzahl von Kanälen die optimalen  $\bar{w}_k$  ermittelt wurden. Die resultierenden Funktionen sind im Folgenden dargestellt. Da die optimale Aufteilung von der zum Einsatz kommenden Bitrate unabhängig ist, kann sie ebenfalls als Rate-Distortion-Funktion dargestellt werden. In Abbildung 20 ist ein Detail aus den Rate-Distortion-Funktionen der linearen, logarithmischen und optimalen Aufteilung für ein Zweikanalsystem gezeigt.

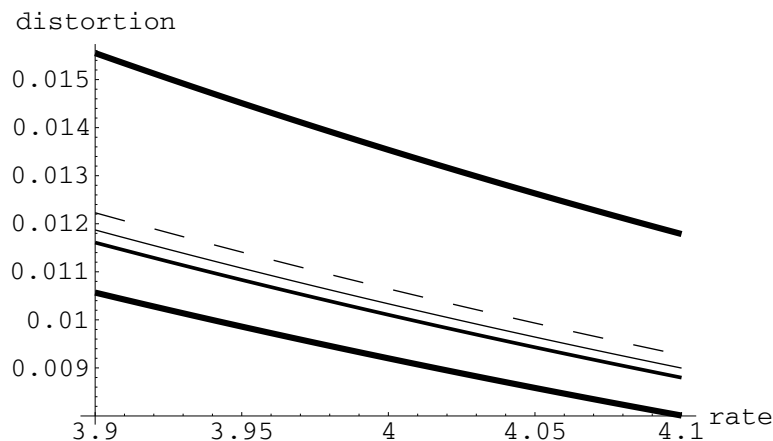


Abbildung 20: R(D) für drei Bandpasszerlegungen von  $1/f$  zwischen Vollbandkodierung (ganz oben) und unterer Grenze mit  $M$  unendlich (ganz unten). Detail aus dem Intervall von 0.5 - 16 cyc/deg bei zwei Kanälen.

Zu erkennen ist nicht nur die verbesserte Leistung bei einem Übergang von einer linearen Aufteilung (gestrichelte Linie) zu einer logarithmischen (durchgezogene Linie), sondern eine weitere, ähnlich große Verbesserung bei Einsatz der ermittelten optimalen Aufteilung (dickere Linie). Graphisch dargestellt sind auch die untere Grenze mit unendlich vielen Kanälen (gültig für alle Aufteilungen) und als obere Begrenzung das Vollbandsystem. Eine logarithmische Aufteilung mindert die Verzerrung um etwa 4% gegenüber der linearen Aufteilung, die Optimalzerlegung sogar um etwa 6%.

## 2.7.1 Komplexitätsvergleich

Das Potential, das sich dadurch zur Einsparung von Ressourcen ergibt, wird bei Bezug auf ein Komplexitätsmaß wie die Kanalanzahl deutlich. Abbildung 21 zeigt die Komplexitäts-Verzerrungs-Funktionen bei optimaler Aufteilung der Frequenzachse im Vergleich zur linearen Partitionierung.

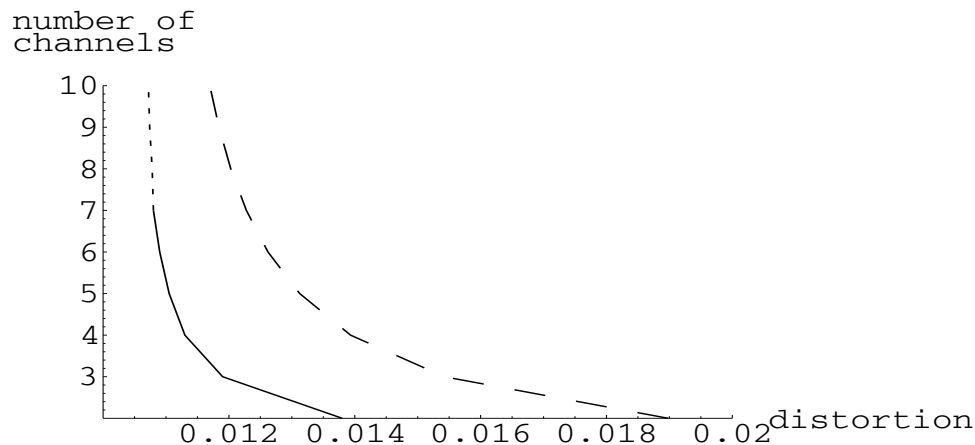


Abbildung 21: Die Komplexitäts-Verzerrungs Funktionen der optimalen Zerlegung (durchgezogene Linie) und die der linearen (gestrichelte Linie). Es ergibt sich ein noch größerer Vorteil als bei der strikt selbstähnlichen Aufteilung. Verwendete Parameter:  $R=4$ ,  $\beta=10^4$ .

Die Optimalzerlegung führt hier zu teilweise gravierenden Einsparung: Wie sich aus dem Graphen ablesen lässt, gibt es Bedingungen, bei denen dieselbe Leistung bei einer Reduktion von 10 auf 3 Kanäle erreicht wird. Die gewählten Parameter entsprechen denen in Abbildung 18. Eine biologisch motivierte Gesamtbandbreite von 5 Oktaven führt zu vergleichbaren Verbesserungen, siehe Abbildung 22.

Zur direkten Gegenüberstellung der linearen, logarithmischen und optimalen Aufteilung sind diese in Abbildung 23 zusammen dargestellt. Hier führt eine logarithmische Zerlegung zu einer Komplexitätsreduktion von etwa 8 auf etwa 6 Kanäle, durch eine optimal an die Quellstatistik angepasste Zerlegung kann sogar noch ein weiterer Bandpass eingespart werden.

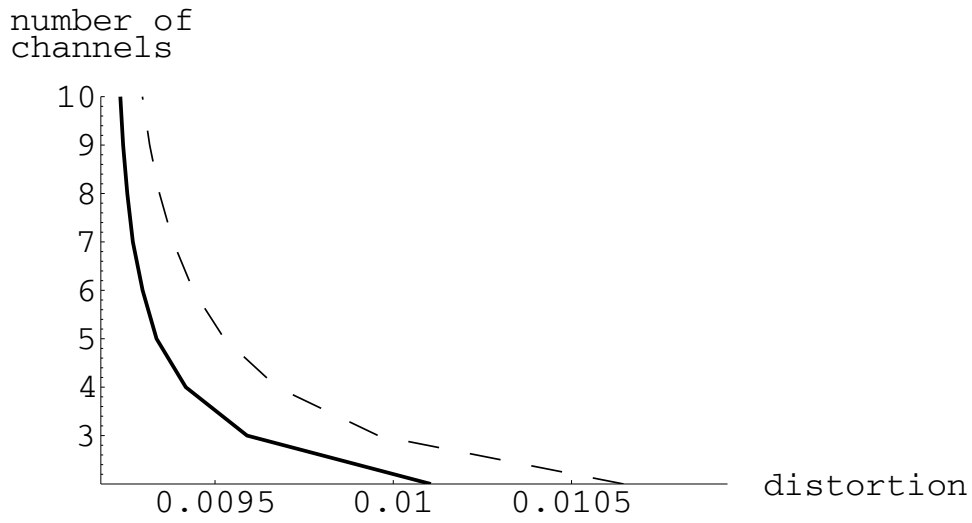


Abbildung 22: Komplexitätsvergleich zwischen linearem Bandsplitting und Optimalaufteilung bei einem kodierten  $1/f$ -Signal zwischen 0.5 und 16 c/deg und einer Bitrate von 4 bit.

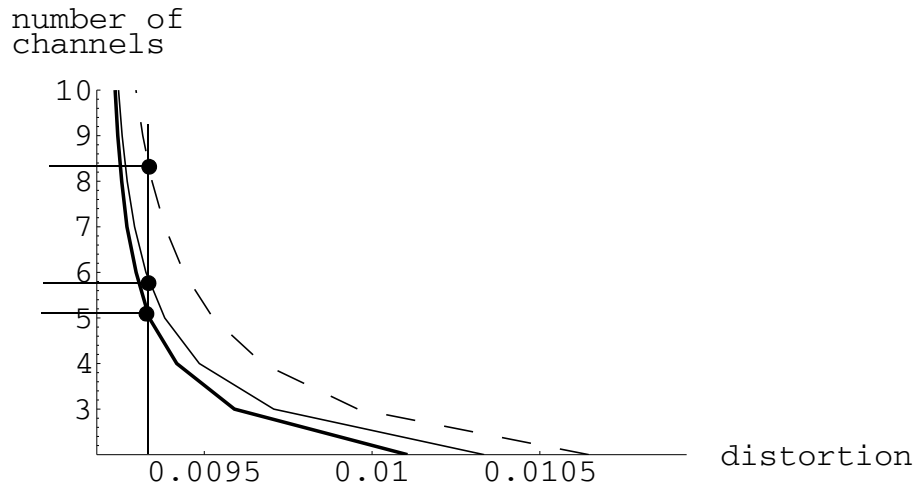


Abbildung 23: Vergleich zwischen linearer (gestrichelt), logarithmischer (durchgezogen) und optimaler (fett) Aufteilung. Hervorgehoben ist die erreichbare Komplexitätsreduktion.



## 2.7.2 Resultierende Filterstruktur

Interessant ist nun die Frage, welche Struktur die ermittelte optimale Zerlegung aufweist, und in welchem Verhältnis diese zu einer strikt selbstähnlichen Aufteilung steht. Die resultierende Filterstruktur ist in Abbildung 24 zusammen mit der linearen und der logarithmischen Aufteilung dargestellt.

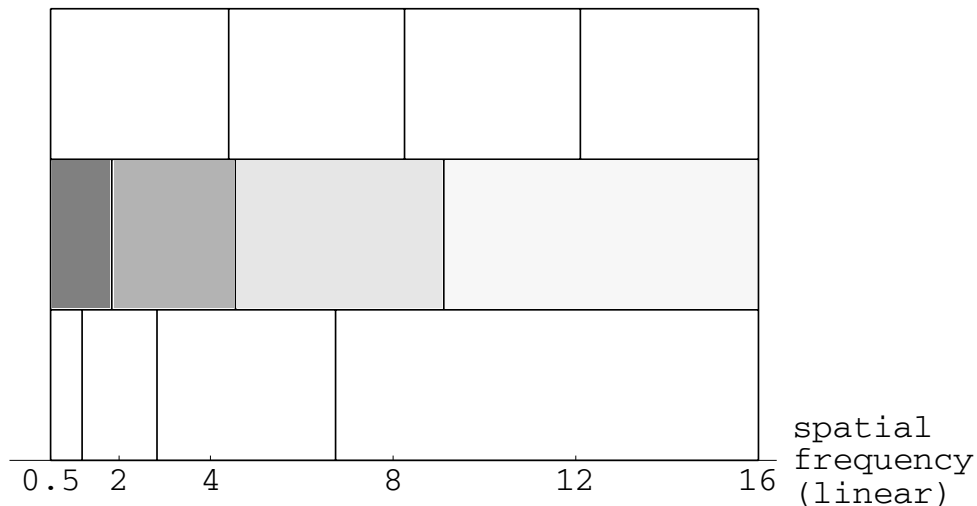


Abbildung 24: Optimalzerlegung (mittig) im Vergleich zu linearer (oben) und selbstähnlicher Aufteilung (unten) bei einem 4-kanaligen Kodierer.

Ersichtlich ist, dass die optimal angepasste Zerlegungsstruktur deutlich von einer linearen Einteilung der Frequenzachse (oben) abweicht, indem niederfrequente Bandpässe schmaler gewählt werden und hochfrequente Bandpässe breiter.

Aber auch von einer oktavbreiten Aufteilung (unten) weicht die gefundene Struktur ab. Dies wird auf der logarithmischen Skala in Abbildung 25, bei der eine strikt selbstähnliche Aufteilung zu gleich breiten Teilbändern führt, besonders deutlich. Im Vergleich dazu sind nun bei der optimalen Aufteilung die log-Bandbreiten für niedrige Mittenfrequenzen größer und für hohe Mittenfrequenzen kleiner. Bei der für  $1/f$ -Signale optimalen Zerlegung werden also die tieffrequenten Komponenten im Signalbereich, die hochfrequenten Komponenten

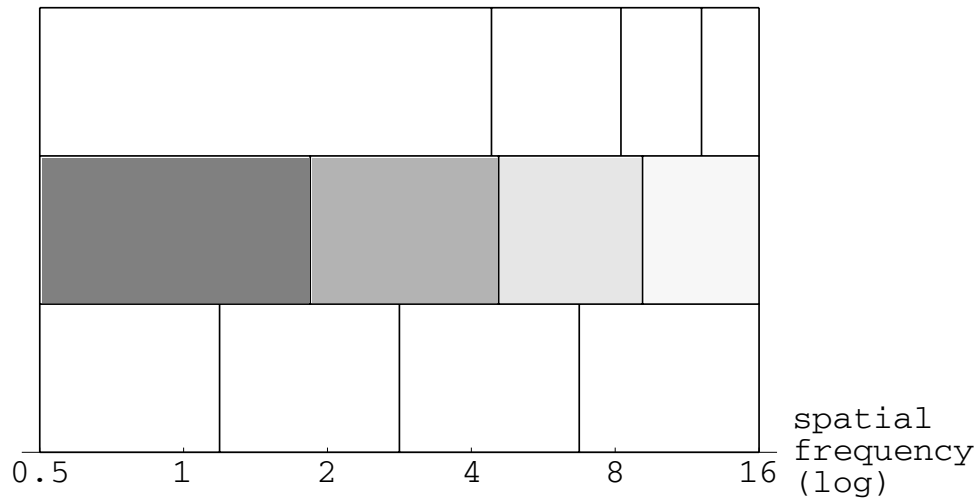


Abbildung 25: Vergleich der Partitionsschemata auf logarithmischer Achse. Daten wie in Abbildung 24.

aber im Frequenzbereich mit einer genaueren Auflösung analysiert, als es eine exakt oktavbreite Partitionierung bewirken würde.

Weitere Analysen bestätigen diese Tendenz der Abweichung von der oktavbreiten Zerlegung: Die Bandzerlegungen in Abbildung 26 beziehen sich auf ein Zweikanalsystem, auf dessen Rate-Distortion-Funktionen bereits eingegangen wurde (siehe Abbildung 20). Links beträgt die Gesamtbandbreite 2 Oktaven, rechts über 13 Oktaven mit unteren und oberen Grenzfrequenzen von 2 und 8 bzw. 1 und  $10^8$  (in beliebigen Einheiten). Abbildung 27 illustriert die Struktur der optimalen Aufteilung für 2 und zusätzlich für 3, 5 und 7 Kanäle. Deutlich wird bei diesen logarithmischen Darstellungen die konsistente Abweichung von der oktavbreiten Zerlegung.

Insgesamt lässt sich feststellen, dass eine strikt selbstähnliche logarithmische Zerlegung weiter modifiziert werden muss, um das wahre Optimum zu erhalten. Meine Untersuchungen weisen darauf hin, dass die an das  $1/f$ -Signal optimal angepasste Filterstruktur grob einer logarithmischen Aufteilung folgt, von dieser aber deutlich und systematisch in der Weise abweicht, dass die log-Bandbreite mit zunehmender Mittenfrequenz größer wird. In Abbildung 28 sind die Simulationsergebnisse für ein Intervall von 0.5 bis 16 cyc/deg mit 4

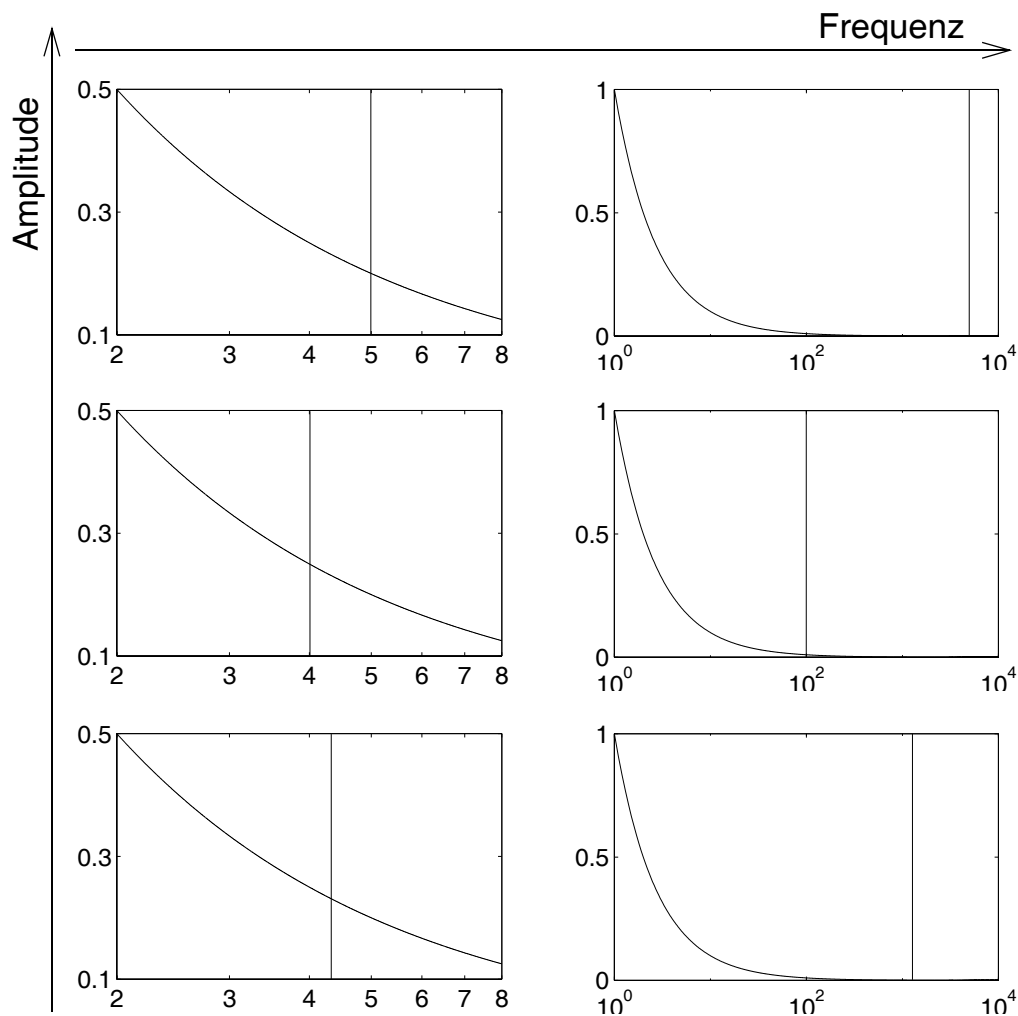


Abbildung 26: Vergleich zwischen linearer, logarithmischer und optimaler Bandaufteilung (von oben nach unten) für unterschiedliche Ausschnitte (links bzw. rechts) eines  $1/f$ -Signals. Die Kanalanzahl wurde für alle Aufteilungen konstant gehalten ( $M=2$ ).

Kanälen auf der im biologischen Kontext üblichen logarithmischen Achse angetragen.

### 2.7.3 Vergleich mit biologischer Kanalstruktur

Zum direkten Vergleich der Simulationsergebnisse mit den biologischen Befunden wurden für Abbildung 29 die experimentellen Ergebnisse aus fünf Studien zusammen mit dem hergeleiteten Optimum angetragen. Die psychophysischen Daten (Blakemore, Campbell, 1969, Georgeson, 1980, Wilson et al., 1983 und Wilson,

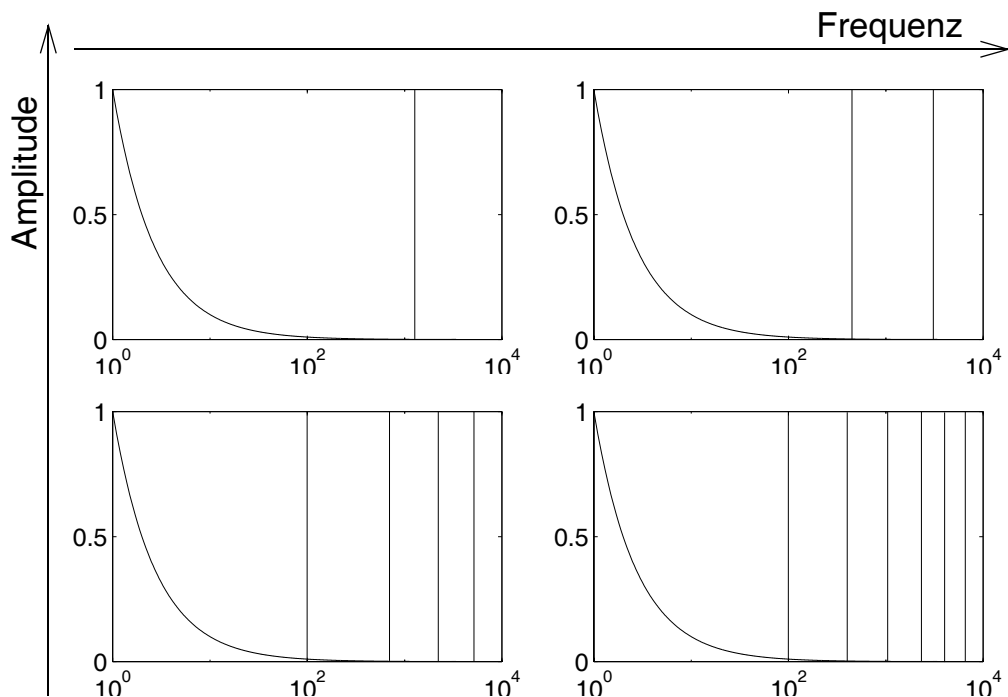


Abbildung 27: Ausgewählte Simulationsergebnisse. Optimale Zerlegung eines  $1/f$ -Signals mit 2, 3, 5 und 7 Kanälen.

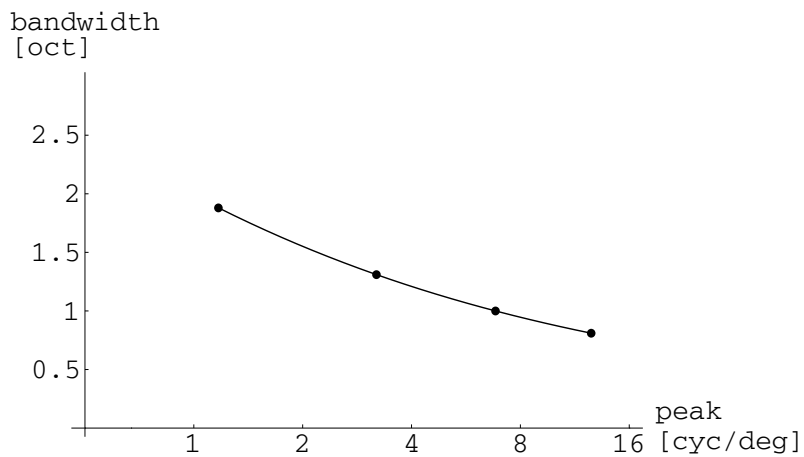


Abbildung 28: Systematische Abweichung der Optimalzerlegung von einer selbstähnlichen Aufteilung. Das kodierte Ortsfrequenzintervall und die Kanalanzahl entsprechen den Ergebnissen aus Abbildung 25.

Gelb, 1984) sind mit unterschiedlichen Dreiecken dargestellt, die neurophysiologischen Daten (De Valois et al., 1982) mit Quadraten und das hergeleitete Optimum mit fünf Kanälen durch Sterne. Die berechnete Bandbreite reichte von 0.25 bis 32 cyc/deg.

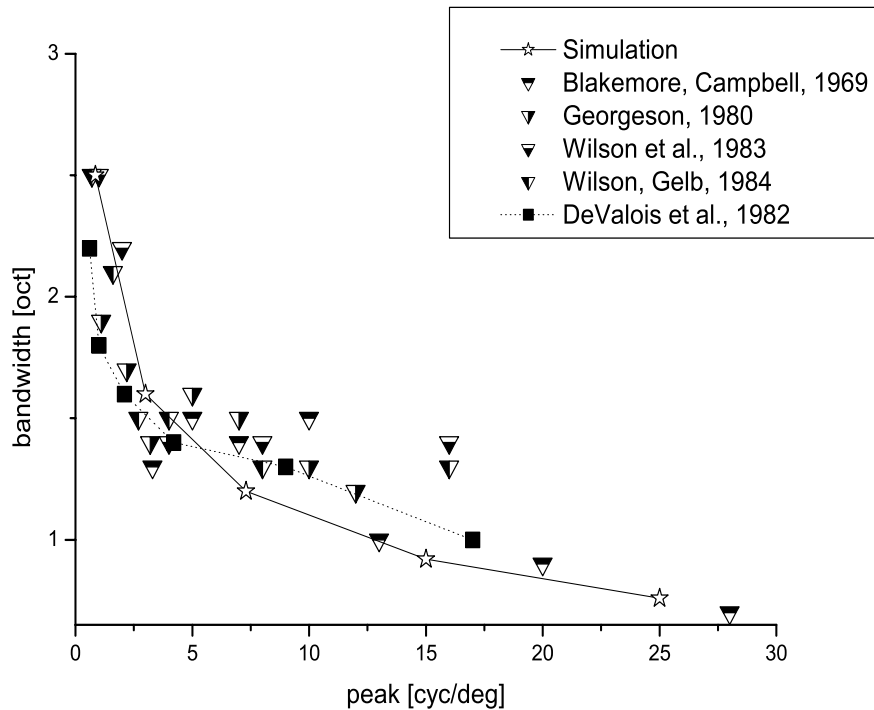


Abbildung 29: Vergleich zwischen der hergeleiteten Optimalzerlegung (Simulation) und der aus unterschiedlichen Experimenten gewonnen biologischen Frequenzaufteilung.

Die aus den Simulationen resultierende Bandpasszerlegung entspricht damit (in Hinblick auf die biologische Variabilität) in guter Näherung der spezifischen Struktur der natürlichen Aufteilung. Sie deckt sich sowohl mit den Befunden zu kortikalen Simple Cells des visuellen Kortex, als auch mit den Beobachtungen aus vielen psychophysischen Experimenten unterschiedlicher Methodik.

## 2.8 Diskussion

Es wurde bisher angemerkt, dass zwischen der näherungsweise selbstähnlichen Bandpasszerlegung in biologischen Sehsystemen und dem typischen "selbstähnlichen"  $1/f^2$ -Abfall des Leistungsdichtespektrums von natürlichen Bildern ein enger Zusammenhang besteht. Intuitiv scheint diese Beziehung gut zu Barlow's Hypothese einer optimalen Anpassung der sensorischen Systeme an die statistischen Redundanzen der Umgebung zu passen, aber bisher konnte keine vollständige formale Rechtfertigung dieser Vermutung angeführt werden. Es gibt auch keine klare Demonstration der Entstehung selbstähnlicher Zerlegungen in künstlichen neuronalen Netzen. Die systematische Abweichung von der konstanten Bandbreite wurde zwar gesehen, konnte aber bisher nicht erklärt werden. So schreibt beispielsweise Field (1993, Seite 163):

*"Although we will treat the visual system as a self-similar code throughout this paper, it should be noted that there is substantial evidence that high frequency cells have somewhat narrower bandwidth (in octaves) at higher spatial frequencies (e.g., Tolhurst and Thompson, 1982). The ramifications of this finding are not yet clear."*

Die Evidenz der Abweichung von einer strikten Selbstähnlichkeit ist in der Tat substantiell und diese ist sogar relativ groß und vor allem systematisch.

Argumente, die vor allem auf die Selbstähnlichkeit der beiden Strukturen verweisen, sind problematisch, da Messungen der Wahrscheinlichkeitsdichtefunktion auf verschiedenen Skalen ergeben, dass diese bei hoher Ortsauflösung eine spärlichen Verteilung aufweist, zu gröberen Skalen hin jedoch gaußischer wird (Zetsche, Krieger, 2001). Diese Tendenz ist im Einklang mit dem zentralen Grenzwertsatz, da bei niedriger Auflösung über größere Bildbereiche, d.h. über eine größere Anzahl von Zufallsvariablen gemittelt wird, spricht jedoch gegen eine wahre Skaleninvarianz.

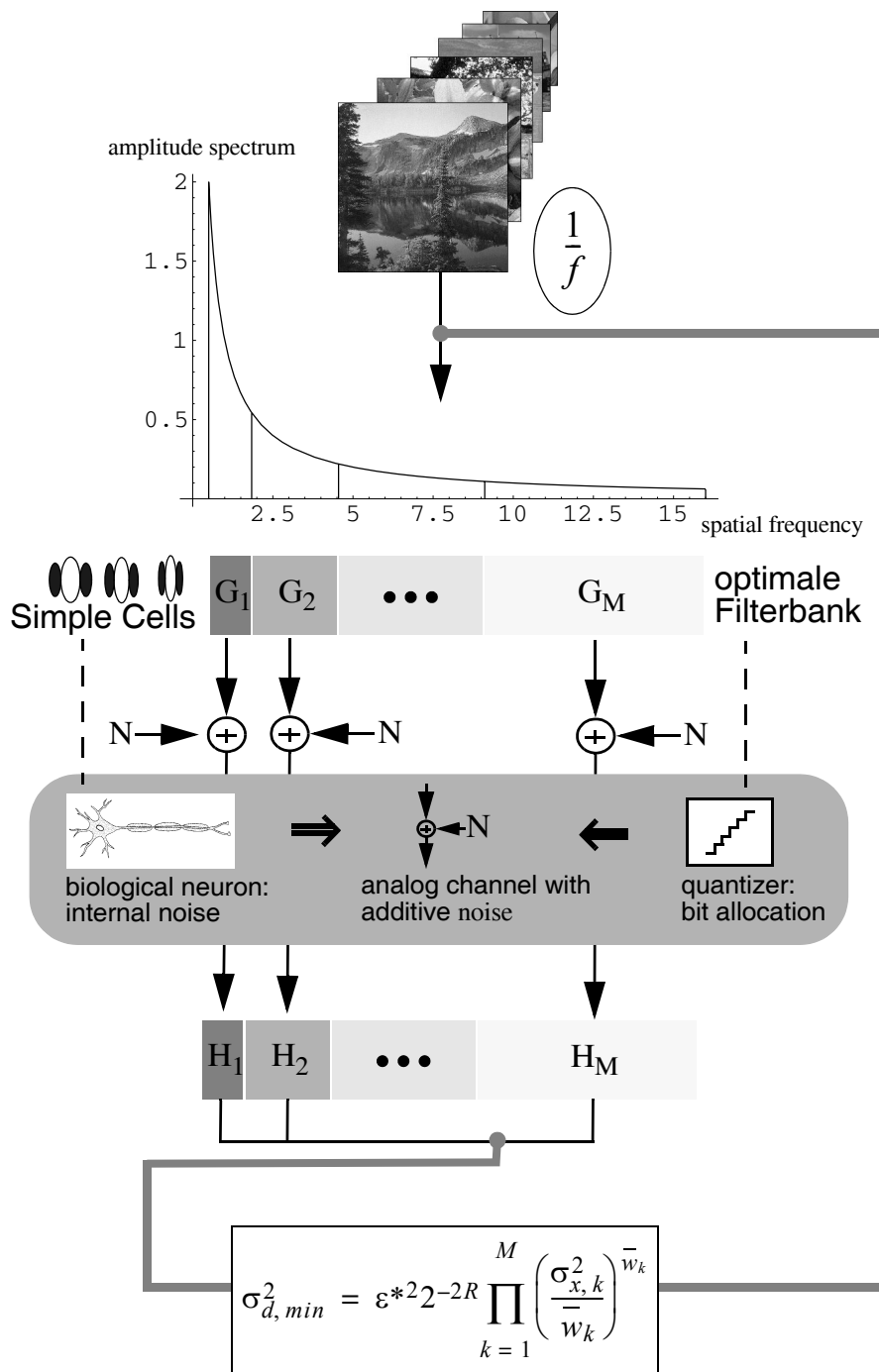


Abbildung 30: Zusammenfassende Darstellung der intrasensorischen Analyse. Zu den wichtigen Konzepten gehören die Statistik natürlicher Bilder, die Bandpasszerlegung biologischer und technischer Systeme und die hergeleitete Minimalverzerrung bei limitierten Ressourcen. Als Resultat lässt sich die spezielle Struktur der biologisch realisierten Bandpässe als eine im Rate-Distortion-Sinne optimale Anpassung an die  $1/f$ -Struktur natürlicher Bilder interpretieren.

In der vorliegenden Analyse wurde die Rate-Distortion-Theorie als informationstheoretischer Rahmen benutzt. Der mittlere Transinformationsgehalt war dabei ein Maß für die zur Beschreibung einer Quelle nötige Information und es war unter allen Codes, welche diese Quelle mit einer Verzerrung  $D$  codieren für denjenigen eine kleinstmögliche Rate  $R$  aufzuwenden, der den geringsten Transinformationsgehalt besaß. In einer Reihe anderer Arbeiten stellt diese Transinformation (mutual information) nun die zu optimierende Kostenfunktion dar. Das Ziel besteht dabei meistens in einer Maximierung dieser Größe, wie etwa bei der klassischen Version des Infomax-Prinzips (Linsker, 1986). Hier wird die Information, die am Systemausgang über den Systemeingang vorhanden ist, maximiert.

In anderen selbstorganisierenden Systemen wird dagegen die Transinformation zwischen benachbarten Bildregionen betrachtet (Becker, Hinton, 1992) oder auch zwischen korrespondierenden Ausschnitten verschiedener Bilder (Ukrainec, Haykin, 1996). Da im letzten Fall die räumliche Verschiebung maximiert werden soll, wird dort die Transinformation minimiert. Verwandt mit dem Infomax-Prinzip ist schließlich die "maximum entropy" Methode (Bell, Sejnowski, 1995) und die des "maximum likelihood" (Cardoso, 1997), bei denen ebenfalls die statistischen Abhängigkeiten zwischen den Ausgabeeinheiten minimiert werden.

Sie alle demonstrieren, dass informationstheoretische Maße geeignet sind, um aus einer statistischen Beschreibung der natürlichen Umgebung wesentliche Eigenschaften der neuronalen Verarbeitung herzuleiten. Diese Ansicht ist auch verwandt mit einer von Craik bereits 1943 beschriebenen Sichtweise, derzufolge im Gehirn ein Modell der externen Welt konstruiert wird, in das die Regelmäßigkeiten und Redundanzen der natürlichen Umgebung eingearbeitet sind.

Es soll betont werden, dass die Anzahl der Kanäle nur ein mögliches Maß für die Komplexität der Zerlegungsschemata darstellt. Abhängig von der genauen Implementierung sind verschiedene andere Maße denkbar. Zum Beispiel könnte in technischen Systemen die Anzahl der Multiplikationen ein essentieller Faktor sein. In biologischen neuronalen Netzwerken könnte die gesamte Verdrahtungs-



länge eine wichtige Rolle spielen, oder auch die Anzahl der Fasern, die auf eine einzelne Ausgangseinheit konvergiert. In der Tat gibt es Hinweise darauf, dass die spezielle Geometrie der Neurone einem "save wire" Organisationsprinzip unterliegt. So lässt sich mit Steinerbäumen eine minimale Gesamtverbindungs­länge durch das optimale Verästelungsvolumen der Dendritenbäume zeigen (Cherniak, 1992, 1994, für einen Überblick über die dendritische Informationsverarbeitung siehe z.B. Mel, 1994). Bezüglich der grundlegenden Reihenfolge der Komplexitäts-Verzerrungs-Kurven, ist das erzielte numerische Ergebnis aber qualitativ für alle Komplexitätsmaße gültig, die monoton mit der Kanalanzahl zunehmen. Die genauen quantitativen Relationen z.B. zwischen der optimalen und der strikt selbstähnlichen Zerlegung für andere denkbare Komplexitätsmaße bleiben dabei natürlich noch offen.

Als weitere Randbedingung ist zu bedenken, dass biologische Systeme unter strikten energetischen Limitierungen operieren, bei denen auch kleinere Einsparungen von großer Bedeutung sein können. So wird der Energieverbrauch von Nervenzellen auf etwa  $10^7$  Adenosin­triphosphat-Moleküle (ATP) pro Bit geschätzt. Aufgrund dieses immensen ATP-Verbrauchs beim Feuern der Neurone sind größere Zellverbände unter Umständen sogar darauf angewiesen, zusätzliche, weniger energieaufwendige Kommunikationskanäle zu benutzen, wie etwa die der nicht-synaptischen Diffusion (Aiello, Bachy-Rita, 2000). Auf das Verhältnis zwischen Energie und Information im Zentralnervensystem soll hier aber nicht weiter eingegangen werden, siehe dazu Laughlin et al (1998), zu den Schwierigkeiten bei der Quantisierung siehe Ames (2000).

Schließlich soll hier auch nicht die Möglichkeit ausgeschlossen werden, dass Statistiken höherer Ordnung zusätzlich eine Rolle spielen, denn das Leistungsdichtespektrum stellt sicherlich keine ausreichende Charakterisierung der strukturellen Eigenschaften natürlicher Bilder dar. Entscheidend ist aber zunächst, dass der hier gewählte Ansatz nicht zu konzeptionellen Inkonsistenzen führt, wie z.B. bei Versuchen, Orientierungsselektivität mit Statistiken zweiter Ordnung zu erklären. Zudem gibt es eine Reihe von Hinweisen darauf, dass gerade auf den frühen Stufen der sensorischen Informationsverarbeitung lineare Verfahren eingesetzt werden, um die in natürlichen

Eingangssignalen vorhandenen hohen Korrelationen auszunutzen. Mit der linearen Methodik wird in Hinblick auf eine effiziente Signalrepräsentation eine weitgehend verlustfreie Umkodierung angestrebt und für das System ist es erstrebenswert, zunächst als Vorverarbeitung eine generische Bildrepräsentation zur Verfügung zu stellen, die möglichst viel Information über das eingehende Signal bewahrt.

## **3. Intra- und intersensorische Verarbeitung**

### **3.1 Erweiterung des kybernetischen Ansatzes**

Im vorangegangenen Kapitel wurde eine für die visuelle Informationsverarbeitung charakteristische neuronale Verknüpfungsstruktur mit spezifischen statistischen Redundanzen in natürlichen Bildern in Beziehung gesetzt. Eine Reihe von intramodalen Eigenschaften konnten mit diesem informationstheoretischen Ansatz erklärt werden, vornehmlich als "Natural Scenes Statistics" innerhalb der visuellen Modalität (Übersicht in Attick, 1992, Ruderman, 1994b, Zetsche, Krieger, 2001), zunehmend aber auch für Eigenschaften auditorischer Neurone (Attias, Schreiner, 1998, Schwartz, Simoncelli, 2001, Lewicki, 2002, Singh, Theunissen, 2003). Ich möchte im Folgenden noch weiter gehen und das Prinzip der Anpassung an die statistischen Eigenschaften der natürlichen Umgebung auf die Interaktion zwischen verschiedenen Modalitäten ausdehnen.

Betrachtet werden nun nicht Korrelationen zwischen Variablen derselben Modalität, wie etwa die Erregungsmuster benachbarter Photorezeptoren auf der Retina, sondern statistische Regelmäßigkeiten zwischen Parametern aus unterschiedlichen Modalitäten, also z.B. akustisch-visuelle Kovariationen. Wenn der kybernetische Ansatz tatsächlich von so universeller Natur ist, sollten sich entsprechende Vorhersagen für die multisensorische Informationsverarbeitung ableiten lassen: So dürften korrelierte Parameter nicht unabhängig voneinander kodiert werden, auch wenn sie aus unterschiedlichen Modalitäten stammen. Stattdessen sollten spezifische Signalkombinationen in Form von intersensorischen Interaktionen bereits auf frühen Verarbeitungsstufen realisiert sein und diese Interaktionen sollten an die multimodale statistische Kovarianzstruktur angepasst sein.

Zugleich rückt damit nach der Analyse einer divergenten Zerlegung in separate Kanäle nun das dazu reziproke Prinzip einer Konvergenz von Informationen aus verschiedenen Quellen in den Vordergrund. Aus Modellierungssicht ist es dabei unerheblich, ob verschiedene Bandpässe des visuellen Systems oder vollständige Sinnessysteme als Kanäle betrachtet werden. Auch innerhalb des

visuellen Systems stellt sich nämlich die Frage, welche Interaktionen zwischen den einzelnen Kanälen stattfinden, z.B. bei der Wahrnehmbarkeit zusammengesetzter Bildmuster. In diesem Kontext wurde das Konzept des signalangepassten Filters entwickelt (Hauske, 1988), das einen adaptiven Mechanismus für bestimmte Elementarsignale annimmt, und es wurden auch nichtlineare, probabilistische Modelle zur Erklärung des kanalweisen Zusammenwirkens einzelner Frequenzkomponenten vorgeschlagen.

Im Folgenden sollen drei grundlegende Interaktionsmodelle vorgestellt werden, die im Kontext der Signaldetektion zur intrasensorischen Modellierung (v.a. des visuellen Systems) entwickelt wurden. Diese werden in Kapitel 4 auf den Fall der intersensorischen Signalintegration und Kanalinteraktion angewendet.

## 3.2 Modelle der Kanalinteraktion

Es können zwei grundlegende Modelle der Interaktion unterschieden werden. Beide nehmen separate, modalitätsspezifische Signale als Eingangsstufe an, die dann zu unterschiedlichen Aktivierungen auf bestimmten psychologischen Dimensionen in den jeweiligen Kanälen führen. Die Modelle unterscheiden sich jedoch darin, wie die Informationen aus den getrennten Kanälen auf der nachfolgenden Verarbeitungsstufe zusammengeführt werden. Die Reize können dabei im Signal-Detektions-Kontext auch als “signal” und “noise” interpretiert werden.

### 3.2.1 Wahrscheinlichkeitssummationsmodell

Das Modell der Wahrscheinlichkeitssummation (“probability summation model” oder “statistical facilitation model”) ist in Abbildung 31 dargestellt. Hier wird für jeden Kanal eine separate Entscheidung getroffen und die Gesamtantwort ergibt sich aus der logischen Verknüpfung der Einzelantworten.

Sind die Kanäle dabei unabhängig, d.h. ist das Überschreiten der Schwelle im einen Kanal unabhängig von einem Überschreiten der Schwellen im anderen Kanal, gilt für die bimodale Verbundwahrscheinlichkeit:

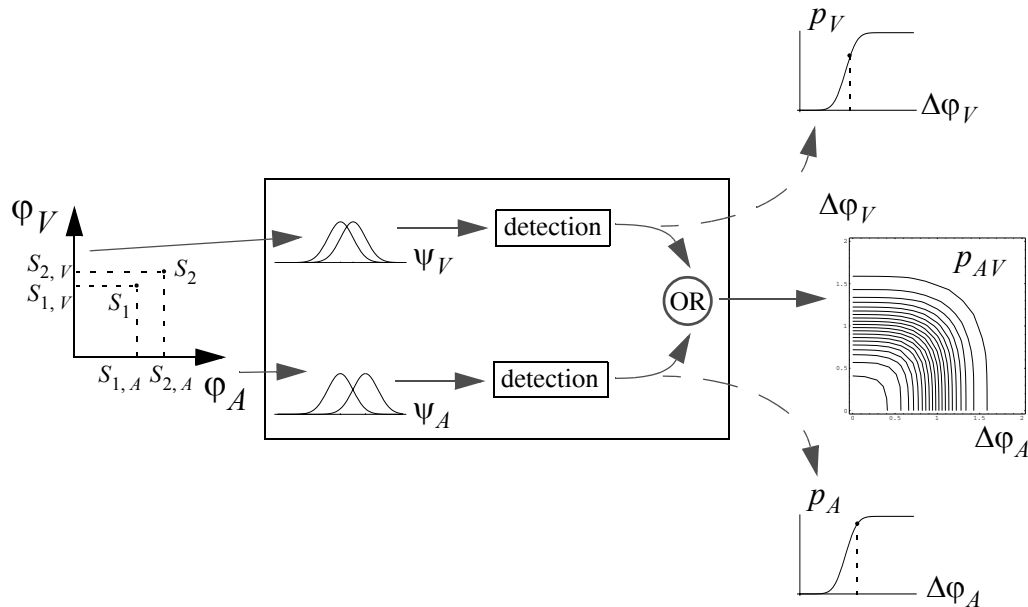
$$p_{AV} = 1 - \prod_{i \in A, V} (1 - p_i).$$

Gegeben sind die Wahrscheinlichkeiten für das Überschreiten der Einzelkanalschwellen  $p(\Delta\varphi_A \geq \theta_A)$  und  $p(\Delta\varphi_V \geq \theta_V)$ . Da es gemäß den Versuchsinstruktionen bereits genügt, in nur einem beliebigen Kanal einen Unterschied zu detektieren, ergibt sich die Gesamtwahrscheinlichkeit aus:

$$p((\Delta\varphi_A, \Delta\varphi_V) \geq \theta) = p((\Delta\varphi_A \geq \theta_A) \vee (\Delta\varphi_V \geq \theta_V)) .$$

Die Wahrscheinlichkeit, dass ein zusammengesetzter Stimulus unerschwellig bleibt, ergibt sich also aus dem Produkt der entspre-

## separate Aktivierung: Wahrscheinlichkeitssummutationsmodell



*Abbildung 31:* Im Modell der Wahrscheinlichkeitssummutation werden die eingehenden sensorischen Daten in getrennten Kanälen verarbeitet und erst auf höherer, kognitiver Ebene logisch miteinander kombiniert. Links sind zwei Reize anhand ihrer Position in einem Koordinatensystem dargestellt, das durch einen akustischen ( $\varphi_A$ ) und einen visuellen Parameter ( $\varphi_V$ ) aufgespannt wird. Die akustisch-visuellen Signale führen zu internen Aktivierungen auf entsprechenden psychologischen Achsen  $\psi_A$ ,  $\psi_V$ . Das Modell macht eine genaue Vorhersage bezüglich des Verlaufs der bisensorischen Diskriminationsschwellen, rechts dargestellt als Funktion des physikalischen Unterschieds.

chenden unimodalen Wahrscheinlichkeiten (Abbildung 32). Die beobachtbare Antwort einer Versuchsperson resultiert hier aus einer Oder-Kombination der beiden kanalspezifischen Entscheidungen.

Die Verbundwahrscheinlichkeit zur Diskrimination des zusammengesetzten Reizes ergibt sich dann aus dem Produkt der Wahrscheinlichkeiten zur Erreichung der Einzelkanalschwellen. Die Gesamtwahrscheinlichkeit für eine korrekte Antwort ist deshalb bei Präsentation eines aus zwei Komponenten bestehenden Reizes etwas höher bzw. der für ein festgelegtes Erkennungsniveau erforder-

liche Signal-Rausch-Abstand fällt etwas kleiner aus. Eine Leistungsverbesserung durch redundante Information führt dieses Modell damit auf einen probabilistischen Summationseffekt zurück.

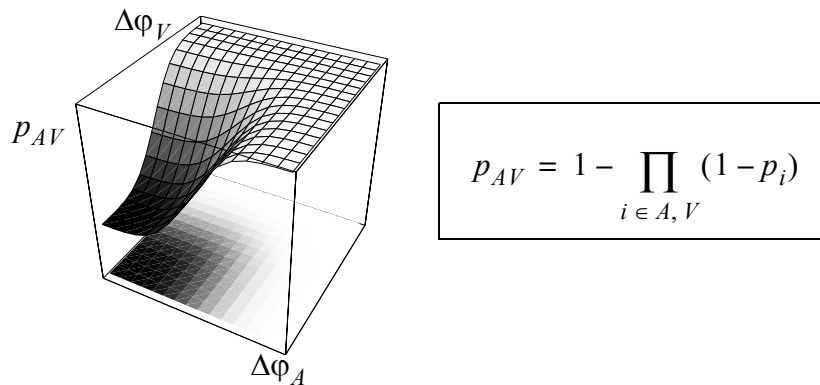


Abbildung 32: Das Wahrscheinlichkeitssummutationsmodell sagt einen kleinen Verbesserungseffekt bei multimodaler Reizung voraus. Die Gesamtwahrscheinlichkeit  $p_{AV}$  für eine korrekte Entscheidung ergibt sich hier aus den unimodalen Wahrscheinlichkeiten  $p_A$  und  $p_V$  und führt zu einem statistischen Verbesserungseffekt.

Andere Namen für dieses Modell sind “seperate activation model” und “extreme detector”. Ist die anhängige Variable nicht die Detektionsleistung, sondern die Reaktionszeit, wird das entsprechende Modell als “race model” bezeichnet.

Da dieses Modell für multisensorische Reize teilweise noch zu große Effekte voraussagt, wurden zwei parametrisierte Modelle vorgeschlagen: Im so genannten “phi model” (Craig et al., 1976) sind die Kanäle nicht mehr unabhängig, ausgedrückt durch einen Korrelationskoeffizienten  $\phi$ ,  $0 \leq \phi \leq 1$ :

$$p_{AV} = p_A + p_V - p_A p_V - \phi \sqrt{p_A(1-p_A)p_V(1-p_V)}$$

Das “mixture model” (Mulligan, Shaw, 1980, Shaw, 1982) dagegen geht davon aus, dass die Aufmerksamkeit in einer probabilistischen

Weise primär nur auf eine Modalität gerichtet wird (Gewichtung mit  $\alpha$ ,  $0 \leq \alpha \leq 1$ ), für den anderen Kanal ist die Schwelle dann erhöht ( $p_A'$  bzw.  $p_V'$ ):

$$p_{AV} = 1 - [\alpha(1 - p_A)(1 - p_V') + (1 - \alpha)(1 - p_A')(1 - p_V)] \quad \text{mit}$$

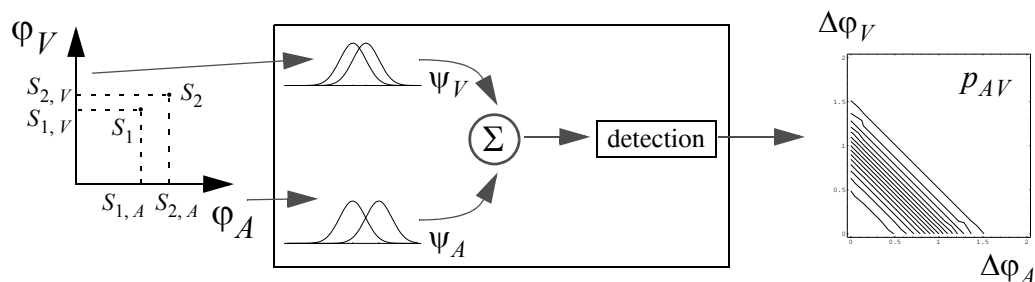
$$p_i \geq p_i' \quad i \in A, V$$

Bei Shaw wird dieses Modell innerhalb der “2nd-order integration” Modelle “sharing mixture model” genannt, bzw. bei  $p_A', p_V' \rightarrow 0$  “all-or-none mixture model”.

### 3.2.2 Neuronales Summationsmodell

Das neuronale Summationsmodell (auch “integration model”, “coactivation model” oder “energy summation model”) nimmt keine Summation von Wahrscheinlichkeiten an, sondern direkte physiologisch-neuronale Summation.

#### Koaktivierung: neuronales Summationsmodell



*Abbildung 33:* Im neuronalen Summationsmodell werden die Eingangssignale bereits vor einer Entscheidungsstufe zu einer gemeinsamen Aktivierung zusammengeführt. Die vorhergesagten zweidimensionalen Schwellen sind rechts als Funktion eines akustischen  $\Delta\phi_A$  und eines visuellen Parameters  $\Delta\phi_V$  dargestellt.

Wie in Abbildung 33 gezeigt, konvergieren hier die Signale auf einer frühen Stufe, bevor dann basierend auf dem resultierenden



Ergebnis eine Entscheidung getroffen wird. Diese Art der Integration entspricht einer Linearkombination der Eingangsvariablen (Abbildung 34) mit Summation der kanalspezifischen Sensitivitätsparameter  $d'$ . Der Index  $d'$  gibt dabei den auf eine Standardabweichung normierten Abstand zwischen den Mittelwerten der beteiligten Verteilungen an.

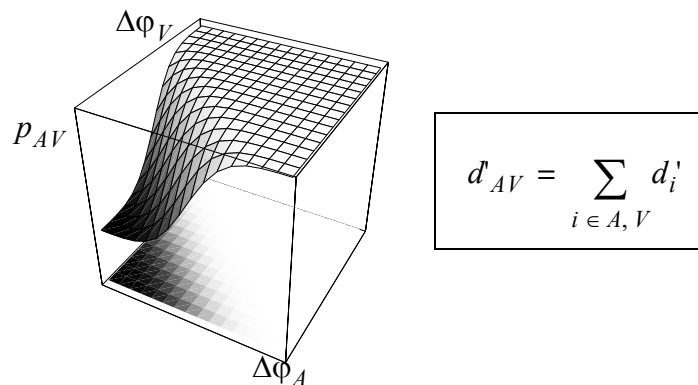


Abbildung 34: Das Summationsmodell nimmt eine Linearkombination der Eingangsvariablen an. Dies entspricht einer Addition der unimodalen Sensitivitätsindizes  $d'$  (rechts) und die daraus resultierende zweidimensionale Schwellenkurve zeigt eine drastische Verbesserung (links).

Während das Modell der Wahrscheinlichkeitssummation dem “high-threshold multiple-analyzers model” von Graham entspricht, findet man das neuronale Summationsmodell unter dem Namen “single-analyzer model” (Graham, 1989).

### 3.2.3 Quick’sches Poolingmodell

Zur Charakterisierung der Größe von Summationseffekten wird häufig das Quick’sche Modell verwendet (Quick, 1974). In diesem wird eine Weibull-Funktion als psychometrische Funktion des Beobachters angenommen:

$$P_{obs} = 1 - 2^{-(c \cdot S_{obs})^k}$$

Diese ist eine gute Näherung an die kumulative Normalverteilung. Daneben enthält das Modell die nichtlineare Summierungsvorschrift

$$S_{obs} = \left( \sum_{i=1}^N S_i^k \right)^{1/k} .$$

Der Summierungskoeffizient  $k$  gibt hier die Art des Informationspoolings an, da die Einzelantworten im Sinne einer Vektornorm zusammenwirken. Für  $k=1$  werden die Anteile linear summiert, bei  $k=2$  liegen die Schwellen auf einem Kreis und für  $k \rightarrow \infty$  wird die Systemantwort zunehmend von dem sensitivsten Mechanismus bestimmt und schließlich treten keinerlei Summationseffekte mehr auf. Mittlere Werte für  $k$ , die zwischen 3 und 5 liegen, sind mit einer Wahrscheinlichkeitssumation konsistent (vgl. Tyler, Chen, 2000, Meese, Andersen, 2002).

Zu beachten ist, dass das Anpassen eines experimentell ermittelten Funktionsverlaufs durch entsprechende Wahl von  $k$  aber auch eine Änderung des Funktionsverlaufs der psychometrischen Funktion nach sich zieht. Ein so genannter "summation-square plot" ist in Abbildung 35 dargestellt.

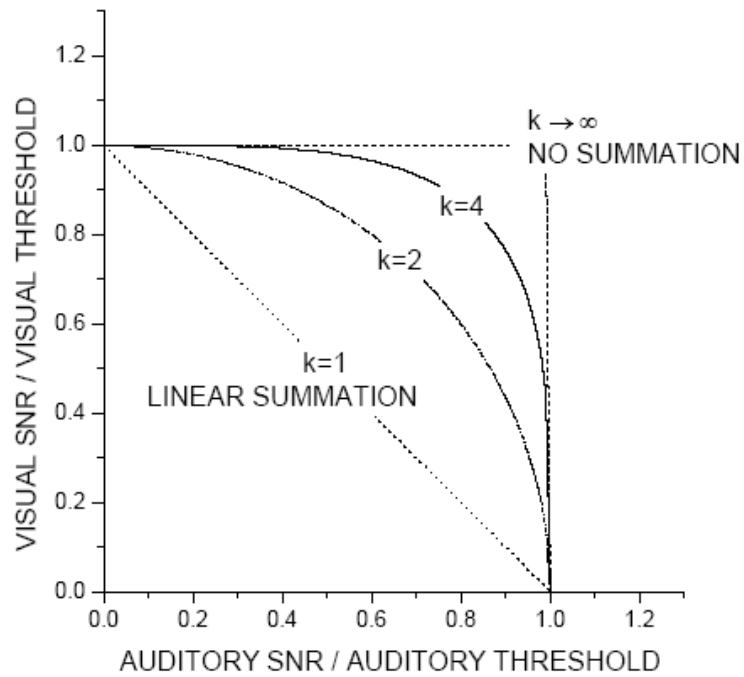


Abbildung 35: Vorhergesagte Schwellenverläufe bei Annahme des Quick'schen Modells. Der Poolingkoeffizient  $k$  reicht von 1, was einer linearer Summation entspricht, bis Unendlich, also zum interaktionslosen Fall (Maximumsuche).

## **4. Interaktion zwischen sensorischen Kanälen**

### **4.1 Einführung**

Bei der Untersuchung sensorischer Prozesse werden die Modalitäten meist voneinander isoliert betrachtet, während wir unter natürlichen Umweltbedingungen in aller Regel multisensorischen Stimuli ausgesetzt sind, die von mehreren Modalitäten parallel verarbeitet werden. Die Sinnessysteme liefern zwar aufgrund ihrer unterschiedlichen peripheren Rezeptorstrukturen zuerst eine Auftrennung in die unterschiedlichen sensorischen Qualitäten, spätestens auf der Ebene der Motorik müssen aber Informationen aus diesen Kanälen fusioniert werden, damit letztendlich eine situationsangepasste Verhaltensantwort generieren werden kann.

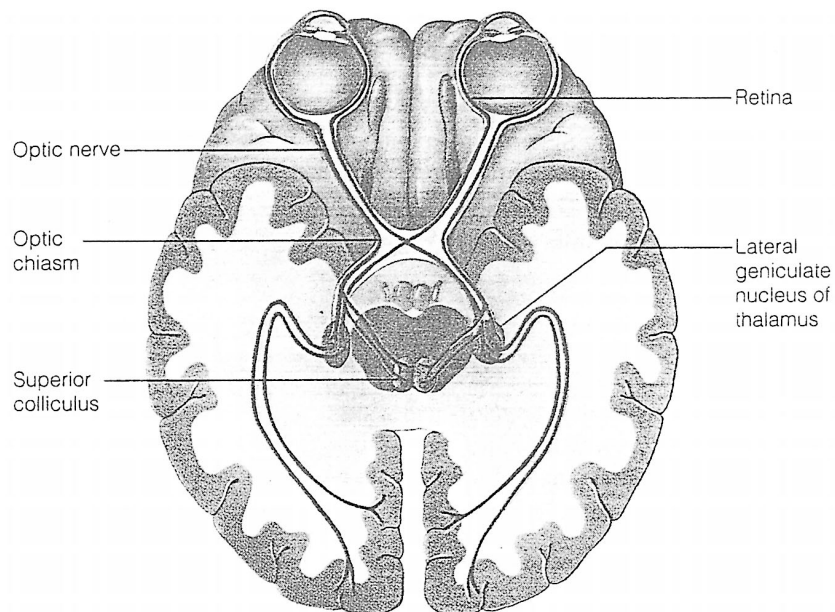
Es stellt sich die Frage, welche Interaktionen zwischen den Modalitäten stattfinden und welcher Verarbeitungsstufe mögliche Interaktionseffekte zuzurechnen sind. Ich habe hier in einer Reihe von psychophysischen Experimenten die intersensorischen Interaktionen bei der Diskrimination von einfachen akustisch-visuellen Stimuli durch Bestimmung von zweidimensionalen Inkrementalschwellen untersucht.

Zunächst aber werden als Erweiterung des gewählten kybernetischen Ansatzes Überlegungen zu polysensorischen Objekten und Ereignissen angestellt, die unter natürlichen Bedingungen intersensorische Redundanzen enthalten oder durch typisches Explorationsverhalten zu solchen führen.

## 4.2 Multisensorische Informationsverarbeitung

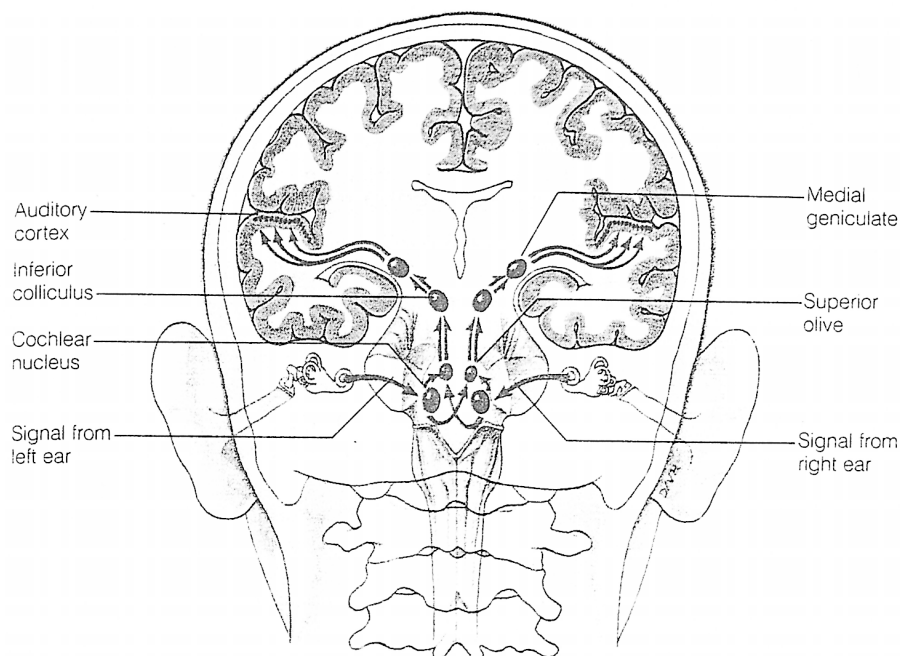
### 4.2.1 Separate Sinnesmodalitäten

Die Evolution hat eine Reihe von sensorischen Mechanismen entwickelt, um Lebewesen mit möglichst vielfältiger Information über ihre gegenwärtige Umwelt auszustatten. Als Informationsquelle steht dabei potenziell das gesamte Spektrum der physikalischen Effekte zur Verfügung. Je nach Art des Informationsträgers, der Energie oder Masse sein kann, lassen sich sechs Signaldomänen unterscheiden (z. B. Middelhoek, Hoogerwerf, 1986): die mechanische, die chemische und die thermische Domäne, die Domäne der elektromagnetischen Wellen sowie die magnetische und die elektrische Signaldomäne. Empfangen werden diese Signale von entsprechenden sinnesphysiologischen Rezeptoren, wie den Mechanorezeptoren, Photorezeptoren usw.



*Abbildung 36:* Die visuelle Informationsverarbeitung erfolgt in der klassischen Sicht in einem anatomisch von anderen Sinnesmodalitäten strikt getrennten Subsystem. Dargestellt ist ein Transversalschnitt durch das menschliche Gehirn, der die primären visuellen Bahnen von der Retina zum Okzipitallappen freilegt.

Diese Sensoren sind in der Regel auf bestimmte physikalische Energien spezialisiert (ihre "adäquaten Reize") und erzeugen aus ihnen die Eingangssignale für hochkomplexe Sinnessysteme, wie etwa das visuelle oder das akustische System (Abbildungen 36 und 37). Diese Teilsysteme konstituieren dann gemäß dem Müller'schen "Gesetz der spezifischen Sinnesenergien" die verschiedenen Sinnesmodalitäten mit ihren entsprechenden Wahrnehmungsempfindungen. Klassisch wird von den fünf Sinnen gesprochen: Sehen, Hören, Schmecken, Riechen, Fühlen. Manche Sinne wie der somatische Sinn werden inzwischen feiner unterteilt und viele weitere Modalitäten sind heute bekannt (z.B. Gleichgewichtssinn, Sinn für Magnetfelder, Infrarotrezeption, Wahrnehmung von Strömungsreizen).



*Abbildung 37:* Ein Frontalschnitt durch das menschliche Gehirn zeigt die wesentlichen Stufen der akustischen Signalverarbeitung. Diese finden in Kerngebieten und Arealen statt, die als rein akustisch beschrieben werden.

Die Einteilung der Wahrnehmung in verschiedene Sinnesmodalitäten ist so selbstverständlich, dass sie auch als das grundlegende Strukturierungsprinzip zur Erforschung der sensorischen Informati-

onsverarbeitung übernommen wurde. In den Titeln der wissenschaftlichen Fachzeitschriften spiegelt sich diese strikte Trennung wieder (etwa *Vision Research*, *The Journal of Auditory Research*, *The Journal of Vestibular Research*) und auch jedes Textbuch zur Wahrnehmungspsychologie oder Neurophysiologie ist kapitelweise nach den einzelnen Sinnen eingeteilt.

Zur Erforschung dieser einzelnen Teilsysteme werden konsequenterweise entsprechende unimodale Reize verwendet: Das visuelle System wird in der Regel ausschließlich mit optischen Reizen untersucht, das akustische System mit auditorischen und das haptische mit haptischen. Unter Laborbedingungen werden Objekte also entweder gesehen, gehört oder gefühlt. Diese Stimuli kontrastieren nun aber stark mit den polymodalen Reizkomplexen, mit denen wir unter natürlichen Bedingungen konfrontiert sind.

#### **4.2.2 Natürliche Reizkomplexe**

Gegenstände, die haptisch wahrgenommen werden, befinden sich in aller Regel auch in unserem Gesichtsfeld, Lebewesen sind meist nicht nur sichtbar, sondern emittieren zugleich auch Schall und der Geschmack von Speisen ist stark mit ihrem Geruch gekoppelt ("Aroma"). Diese natürlichen Objekte erregen mehrere Sinnesmodalitäten über verschiedene physikalische Energien, aber auch bereits ein Signalträger kann dazu ausreichen (Schallwellen können z. B. vom akustischen und haptischen System wahrgenommen werden, Menthol erregt Riechsinneszellen im Riechepithel und Kaltsensoren in der Mundschleimhaut).

Bereits die Tatsache, dass wir zugleich Signale in verschiedenen Modalitäten empfangen können, macht es erforderlich, die Teilsysteme auf Interaktionseffekte hin zu untersuchen. Besonders interessant aber ist, dass durch polymodale Stimuli Information redundant über die modalitätsspezifischen Kanäle übertragen wird. Das Fehlen von Redundanz bemerken wir z. B., wenn die Information aus den Teilbändern nicht übereinstimmt, etwa bei einem Konflikt zwischen visuellem und vestibulärem Signal, oder beim Bauchrednereffekt, der durch die unterschiedliche räumliche Auflösung des akustischen und des visuellen Systems zustande

kommt. Unter natürlichen Bedingungen enthalten die Signale aus den einzelnen Modalitäten jedoch kongruente Information und es ist sehr wenig darüber bekannt, ob und in welchem Ausmaße diese Signalredundanz vom Wahrnehmungsapparat ausgenutzt werden kann.

Welche Informationen können redundant über verschiedene Sinneskanäle übertragen werden? Theoretisch sind alle möglichen Kombinationen in diesem multimodalen Signalraum denkbar, tatsächlich ist dieser Raum wohl eher dünn besetzt. Wir wollen uns im Folgenden auf die beiden Fernsinne konzentrieren. Interessante Interaktionseffekte durch sensorische Konvergenz existieren jedoch entlang der gesamten Neuraxis und lassen sich z.B. auch zwischen dem Geruchs- und dem Geschmacksinn finden (Dalton et al, 2000).

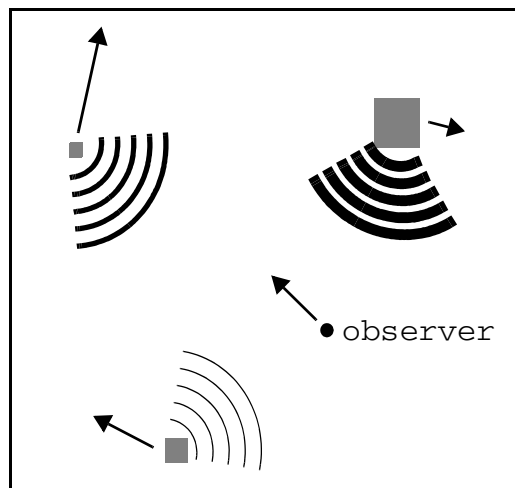
Objekte und Ereignisse, die vom Beobachter gesehen und gehört werden, übertragen also potentiell in beiden Sinneskanälen vielfältige Informationen. Betrachten wir unter dem Redundanzaspekt akustisch-visuelle Reizkomplexe, wie sie in der natürlichen Umgebung vorkommen, so lassen sich diese in mehrere Kategorien einteilen. Zum einen gibt es punktuelle Ereignisse, wie das Zusammenknallen, das Abbrechen oder Klatschen, bei denen die physikalische Ursache für das akustische Signal gesehen wird. Die exakten spatio-temporalen Koordinaten in beiden Modalitäten spielen eine wesentliche Rolle. Länger andauernde akustisch-visuelle Muster, wie Meeresrauschen und Blätterscheln, aber auch Paare von etwa Tierbildern und den entsprechenden Lauten sind gelernte Kombinationen mit hoher Korrelation zwischen geeigneten akustischen und visuellen Parametern. Eine besondere Bedeutung spielen die hochkomplexen Kommunikationssignale, beispielsweise die gesprochene Sprache und die dazu synchronen Lippenbewegungen.

Es gibt jedoch noch eine weitere, entscheidende Quelle für multisensorische Redundanz. Diese liegt begründet in dem natürlichen Explorationsverhalten eines Beobachters in seiner Umgebung. Sowohl die Eigenbewegung des Beobachters als auch die Bewegung des beobachteten Objekts führen nämlich zu statistischen Kovariationen in den sensorischen Eingangskanälen.



### 4.2.3 Multisensorische Signal- und Objektstatistik

Für die akustisch-visuellen Korrelationen bei bewegungsinduzierten Signalen haben wir bereits Modellsimulationen berechnet. Dazu wurden die statistischen akustisch-visuellen Kovarianzen eines Beobachters, der sich in einer Umgebung mit anderen sich bewegenden, schallemittierenden Objekten bewegt, in Computersimulationen untersucht. Dargestellt in Abbildung 38 sind Objekte unterschiedlicher Größe und Position sowie unterschiedlicher Schallrichtung und Lautstärke, die sich mit unterschiedlicher Geschwindigkeit in unterschiedliche Richtungen bewegen.



*Abbildung 38:* Schematische Anordnung der Computersimulation zur Messung akustisch-visueller Kovarianzen mit bewegten Objekten und einem Beobachter. Siehe Text für Details.

Wir haben dabei eine klare statistische Korrelation zwischen wahrgenommener Objektgröße und Lautstärke. Dieses Ergebnis ist weitgehend unabhängig von den statistischen Parametern der Bewegungen und von den statistischen Verteilungen der intrinsischen Objektgrößen und Lautstärkewerten.

Redundanz kann in Parametern zur Identität des Objekts oder des Ereignisses enthalten sein, in verschiedenen zeitlichen Größen, wie

Beginn, Ende, Dauer und Rhythmus und in räumlichen Parametern, wie Position, Anzahl und Ausdehnung. Eine Korrelation lässt sich wahrscheinlich auch für andere akustisch-visuelle Parameterpaare nachweisen. So führen etwa akustische Dämpfungseffekte zu einer frequenzabhängigen Energieabnahme, wodurch weiter entfernt liegende Objekte v.a. durch tiefe Frequenzen wahrnehmbar sind.

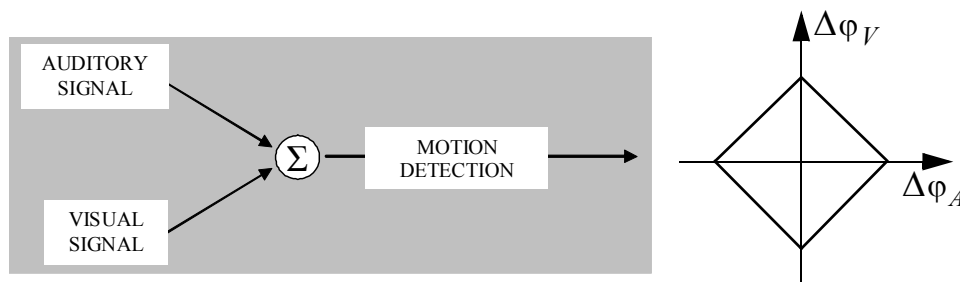
Zur systematischen Analyse wäre eine vollständige akustisch-visuelle Statistik natürlicher Bildsequenzen nötig. Da eine umfassende Berechnung der Wahrscheinlichkeitsdichtefunktion im unisensorischen Fall bereits nahezu unmöglich ist, erscheint hier ein Übergang von der Pixel- zur Objektebene sinnvoll. Durch Betrachtung der statistischen Zusammenhänge auf dieser vergrößerten Ebene verlagert sich die Beschreibungsebene auf eine multisensorische Statistik natürlicher Objekte. Hier können Fragen adressiert werden, wie: Gibt es einen statistischen Zusammenhang zwischen der Größe eines Objekts und seiner Lautstärke? Oder: Hängt die mittlere vertikale Position eines Tieres mit der Frequenzhöhe seiner Laute zusammen? In der Tat gibt es Ergebnisse, die auf eine Interaktion dieser Parameter hindeuten. So wurde in einem Experiment von Ben-Artzi und Marks (1995) gleichzeitig die Frequenz eines Sinustons und die vertikale Position eines einfachen visuellen Reizes variiert. In "kongruenten" Bedingungen wurden hier deutlich kürzere Reaktionszeiten erreicht, wobei dies bisher als ein Effekt auf semantischer Ebene ("natural mapping") erklärt wurde. Der Statistikansatz bietet hier eine fundierte, alternative Interpretation.

#### **4.2.4 Ausnutzung der polymodalen Korrelationen**

Ich habe mir nun die Frage gestellt, ob und wie sich das menschliche Wahrnehmungssystem diese multisensorische Kovarianzstruktur zunutze macht. Wenn der Ansatz des *Natural Scenes Statistics* zur Interpretation von Eigenschaften des visuellen Systems auf das Wahrnehmungssystem insgesamt erweiterbar ist, müsste sich auch auf Verhaltensebene eine Anpassung an natürliche, polymodale Reizkomplexe zeigen lassen.

Eine Ausnutzung der durch Bewegung verursachten polymodalen Korrelationen sollte sich etwa bei einer akustisch-visuellen Detekti-

onsaufgabe in erhöhter Sensitivität niederschlagen. Darüber hinaus wäre im Sinne einer Anpassung an die Statistik zu fordern, dass eine Leistungsverbesserung selektiv nur für ökologisch valide Merkmalsausprägungen zu beobachten ist, d.h. für Stimuluskonfigurationen, die typisch sind für ein Verhalten in natürlichen Umgebungen.



*Abbildung 39:* Links: Summation der neuronalen Signale mit einem multisensorischen Mechanismus zur Bewegungsdetektion. Rechts: Vorhersage dieses Modells für bisensorische Detektionsschwellen.

Zur Verdeutlichung sind hier die in Kapitel 3 eingeführten Interaktionsmodelle für den Kontext der akustisch-visuellen Bewegungsdetektion skizziert: In Abbildung 39 ist das Koaktivierungsmodell dargestellt, das die Eingangssignale einer bisensorisch detektierbaren Bewegung optimal im Sinne einer bestmöglichen Informationserhaltung miteinander verrechnet.

Wenn die aufgestellte Hypothese einer Anpassung an die polysensorische Signalstatistik stimmt, sollten sich Versuchsergebnisse gemäß diesem linearen Summationsmechanismus erklären lassen, wenn die Eingangssignale kongruent zueinander sind, also z.B. die gleiche Bewegungsrichtung aufweisen. Im Zweidimensionalen sagt dieses Modell (als Erweiterung von Abbildung 35) Detektionsschwellen entlang einer rautenförmigen Figur vorher.

Das Modell der Wahrscheinlichkeitssummation (Abbildung 40) sollte dagegen der Verarbeitung inkongruenter Reizkombinationen zu Grunde liegen, da diese nicht die unter natürlichen Bedingungen typischen Reize darstellen. Der durch diesen späten Integrationsmechanismus resultierende Schwellenverlauf ist durch einen Recht-

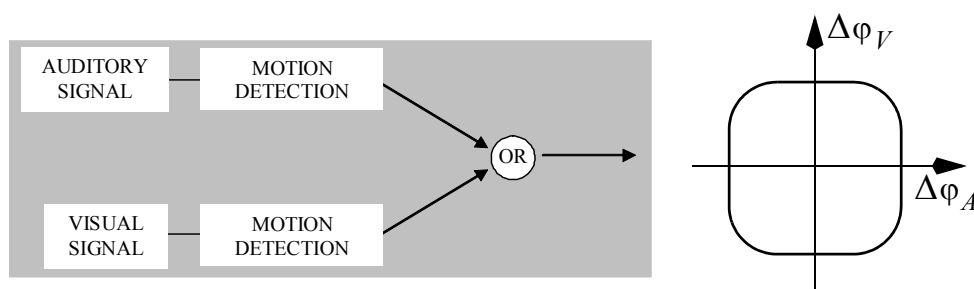


Abbildung 40: Links: Modalitätsspezifische Bewegungsdetektion mit nachfolgender logischer Verschaltung. Rechts: Entsprechende Modellvorhersage für akustisch-visuelle Detektionsschwellen.

eck charakterisiert, das durch den statistischen Summationseffekt abgerundete Ecken aufweist.

Da über multisensorische Integration nur sehr wenige experimentelle Daten verfügbar sind, können wir nicht direkt, wie im ersten Teil der Arbeit, verschiedene neuronale Strukturen in Betracht ziehen und überlegen, in welchem Sinne diese für charakteristische Eigenschaften der visuellen Signalstatistik angepasst sind. Stattdessen müssen wir in einem ersten Schritt diese Daten zur Wahrnehmungsleistung selbst experimentell bestimmen. Die hierzu konzipierten Versuche werden im nächsten Kapitel beschrieben.

### 4.3 Psychophysische Experimente

Zur Untersuchung intersensorischer Interaktionen zwischen der akustischen und der visuellen Modalität habe ich eine Reihe von Detektionsexperimenten konzipiert und durchgeführt.

Der Schwerpunkt dieses experimentellen Teils liegt in der Untersuchung akustisch-visueller Bewegungsinformation. Dafür wurden in drei Experimenten mit unterschiedlichem Versuchsdesign Stimuli mit Frontal- und Seitwärtsbewegung verwendet. Es gab jeweils kongruente Reizkonfigurationen und solche mit widersprüchlicher Information, bei denen die akustische und die visuelle Bewegungskomponente in unterschiedlichen Richtungen verliefen oder in verschiedenen Hemisphären gezeigt wurden. Variiert wurde auch die zeitliche und räumliche Koinzidenz, indem z.B. die relativen Anfangspositionen der modalitätsspezifischen Komponenten systematisch verändert wurden. Die abhängige, zu messende Variable war in allen Versuchen die jeweilige bisensorische Inkrementalschwelle (auch Relativschwelle oder Differenzschwelle genannt), gemessen wurde sie mit adaptiven Prozeduren im 2-Alternativen Forced Choice Paradigma.

Die experimentellen Aufgaben waren jeweils so gestaltet, dass zwischen ökologisch validen Stimulusbedingungen und solchen, die zwar logisch möglich sind, unter natürlichen Bedingungen aber kaum vorkommen, differenziert werden kann. Zusätzlich zu den bisensorischen Bedingungen wurden jeweils die unimodalen Schwellen gemessen. Durch diese unimodalen Variablen wird dann der Raum aufgespannt, in dem mit den gewählten Paradigmen systematische Untersuchungen zur multimodalen Interaktionen durchgeführt werden können.

Die hier konzipierten Experimente schließen die typischen Stimuluskonfigurationen ein, die durch Relativbewegung zwischen einem Objekt und einem Beobachter verursacht werden. Zuerst aber wurde ein Versuch zu einer grundlegenden Dimension aller Modalitäten durchgeführt: der Intensität.

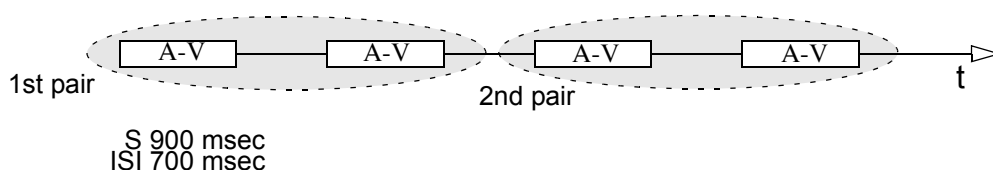
### 4.3.1 Experiment 1: Akustisch-visuelle Intensität

Zur experimentellen Untersuchung der Auswirkung von bisensorischen Kovariationen auf die Unterscheidungsleistung von Menschen habe ich im ersten Versuch die grundlegende Dimension der Reizintensität untersucht.

#### 4.3.1.1 Methode

Die Versuchspersonen saßen in einem halbdunklen Raum mit Kopfhörern in etwa 1.8 m Entfernung vor einem Computermonitor. Sie waren angewiesen, stets sowohl auf die akustische als auch auf die visuelle Komponente zu achten. Die akustische Komponente wurde über Kopfhörer präsentiert und bestand aus einem 1kHz Sinuston von 80dB SPL. Der visuelle Stimulus bestand aus einem hellen Lichtpunkt mit Durchmesser 3.5' und einer Luminanz von 47 cd/m<sup>2</sup>. Er wurde auf einem Monitor mit einer Bildschirmwiederholfrequenz von 80 Hz dargestellt und um ihn herum war zur Schaffung eines homogenen Umfelds eine Blende montiert. Die Linearität im Messbereich wurde durch eine entsprechende Gammakorrektur des Bildschirms sichergestellt.

Sequentiell wurden zwei Paare zusammengesetzter Reize präsentiert, wobei sich ein Reiz von den übrigen bezüglich seiner Intensität unterschied. Diese Variation konnte visuell, akustisch oder bimodal erfolgen, mit unterschiedlichem Vorzeichen und unterschiedlicher Größe.



*Abbildung 41:* Zeitlicher Ablauf von Experiment 1: Sequentiell wurden zwei Paare akustisch-visueller Stimuli präsentiert, wobei die beiden Stimuli eines Paares durch ein Interstimulus-Intervall von 700 msec getrennt waren. Die Präsentationsdauer eines Paares betrug 2,5 sec.

Die Aufgabe der Versuchspersonen bestand darin, zu entscheiden, in welchem von 2 Intervallen sich die Stimuluskonfiguration geändert hat (2AFC). Dabei lautete die Instruktion, dass es für die Antwort unerheblich sei, ob eine Variation des visuellen Signals festgestellt wurde, eine Variation des akustischen Signals oder eine Änderung beider Komponenten. Die Versuchspersonen antworteten durch Drücken von einer von zwei Tasten. Die "Pfeil links"-Taste war zu drücken, wenn ein Unterschied im ersten Paar festgestellt wurde, die "Pfeil rechts"-Taste für einen Unterschied im zweiten Paar. Sie wurden darauf hingewiesen, dass die Reaktionszeit nicht relevant ist.

Die Licht-Ton-Kombinationen wurden für jeweils 900 msec dargeboten, das Inter-Stimulus-Intervall (innerhalb eines Paares) betrug 700 msec. Zwischen den beiden Paaren wurde eine Pause von etwa einer Sekunde Dauer geschaltet. Das verwendete 2-Alternativen forced choice Design ist in Abbildung 41 dargestellt.

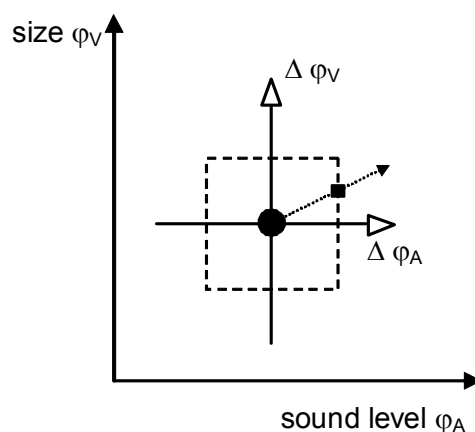


Abbildung 42: Prinzipschema des zweidimensionalen Diskriminationsparadigmas: Ausgehend von einem Referenzstimulus S können bisensorische Inkrementalschwellen in beliebige Richtungen gemessen werden.

Gemessen wurde die Diskriminationsleistung in Abhängigkeit von den visuellen und akustischen Intensitätsunterschieden  $\Delta\phi_V$  und  $\Delta\phi_A$ . Wie in Abbildung 42 ersichtlich, legt das Verhältnis zwischen visueller und akustischer Differenz eine Gerade im zweidimensionalen Schwellen-

lenraum mit Ursprung im Referenzstimulus fest. Diese Gerade gibt die Lage aller möglichen bisensorischen Schwellen unter dieser Versuchsbedingung an, d.h. zur Bestimmung der durch die Gerade festgelegten Schwelle werden genau solche Reize präsentiert, die sich auf dieser Geraden befinden. Wir haben als unterschiedliche Versuchsbedingungen 16 verschiedene Verhältnisse zwischen  $\Delta\varphi_V$  und  $\Delta\varphi_A$  benutzt (Abbildung 43).

	V+	V-
A+	$\frac{\Delta\varphi_A}{\Delta\varphi_V} < 1$	$\frac{\Delta\varphi_A}{\Delta\varphi_V} < 1$
	$\frac{\Delta\varphi_A}{\Delta\varphi_V} = 1$	$\frac{\Delta\varphi_A}{\Delta\varphi_V} = 1$
	$\frac{\Delta\varphi_A}{\Delta\varphi_V} > 1$	$\frac{\Delta\varphi_A}{\Delta\varphi_V} > 1$
A-	$\frac{\Delta\varphi_A}{\Delta\varphi_V} < 1$	$\frac{\Delta\varphi_A}{\Delta\varphi_V} < 1$
	$\frac{\Delta\varphi_A}{\Delta\varphi_V} = 1$	$\frac{\Delta\varphi_A}{\Delta\varphi_V} = 1$
	$\frac{\Delta\varphi_A}{\Delta\varphi_V} > 1$	$\frac{\Delta\varphi_A}{\Delta\varphi_V} > 1$

*Abbildung 43:* Die 12 Bedingungen mit bisensorischer Differenz. Die vorzeichengleichen Bedingungen A+V+ und A-V- finden sich im 1. und 3. Quadranten, während die Bedingungen A-V+ im 2. und A-V- im 4. Quadranten liegen. Zusätzlich wurden die vier unisensorischen Schwellen bestimmt.

Durch dieses Messen in unterschiedliche Richtungen haben wir 16 akustisch-visuelle Differenzschwelle erhalten. Vier Versuchsbedingungen davon nehmen eine Sonderstellung ein, indem ihre Geraden achsenparallel verlaufen. Reize auf diesen Geraden ändern sich rela-



tiv zum Referenzstimulus nur entlang einer Stimulusdimension. Hier werden also unimodale Schwellen gemessen, wenngleich diese Stimuli ebenfalls einen akustischen und einen visuellen Anteil besitzen.

Für die Schwellenmessung wurde eine adaptive Prozedur verwendet (Treutwein, 1997). Diese wurde so eingestellt, dass sie zu etwa 40 Stimuluspräsentationen pro Schwelle führte. Um die Unsicherheit der Versuchspersonen über die aktuell gemessene Schwellenrichtung zu erhöhen und somit die Messresultate zuverlässiger zu machen, wurden jeweils zwei Schwellen in einem Durchgang gemessen. Dabei wurden sorgfältig Kombinationen von Schwellen ausgesucht, so dass zusätzlich sichergestellt war, dass die Versuchspersonen auf beide Modalitäten achten mussten.

#### 4.3.1.2 Ergebnisse

Alle Schwellen wurden zweimal gemessen, wozu 6 bis 7 Sitzungen von je etwa einer Stunde Dauer nötig waren. Fünf bezahlte, naive Versuchspersonen haben an dem Experiment teilgenommen (3 weiblich, 2 männlich).

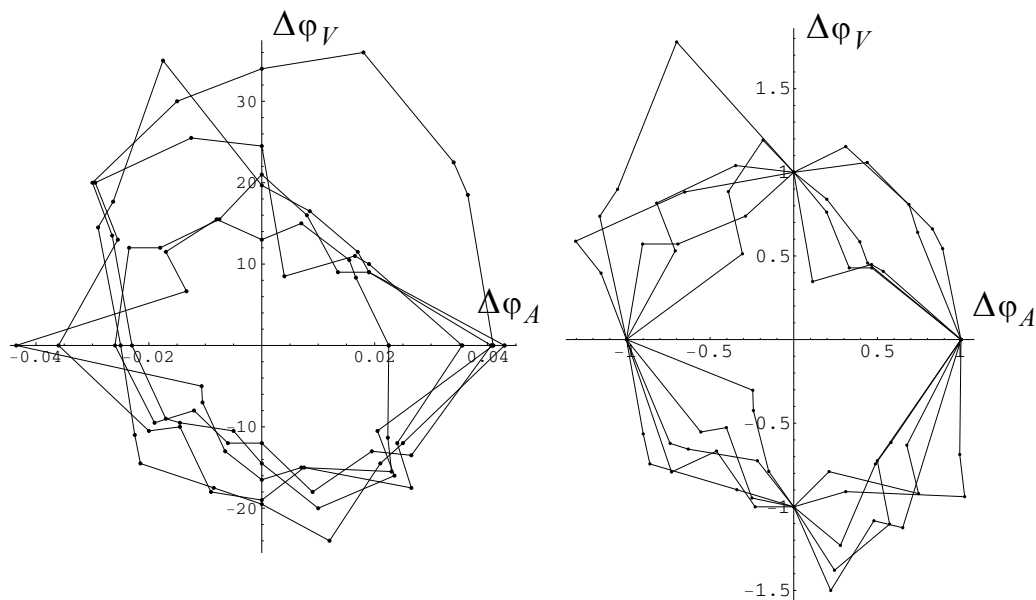


Abbildung 44: Einzelergebnisse der 5 Versuchspersonen: Links sind die zweidimensionalen Inkrementalschwellen in absoluten Größen angetragen, in der Darstellung rechts wurde auf die unimodalen Schwellen normiert.

Die Ergebnisse sind in Abbildung 44 als Differenzschwellenkurven (JNDs - just noticeable differences) um den zusammengesetzten Referenzstimulus herum angetragen. Zur übersichtlicheren Darstellung wurde die Lage des Koordinatensystems so angepasst, dass dieser Stimulus nun genau am Koordinatenursprung liegt. Positive Werte stehen für Amplitudeninkremente, negative für Amplitudendekremente. Da wir an der relativen Leistung interessiert sind, sind die Daten rechts in Abbildung 44 auf die vier gemessenen unimodalen Diskriminationsschwellen auf Abszisse und Ordinate normiert. Dadurch liegen die individuellen Schwellen einer Versuchsbedingung nicht mehr auf einer Geraden, weshalb für die Darstellung in Abbildung 45 vor der Normierung eine Mittelung stattgefunden hat.

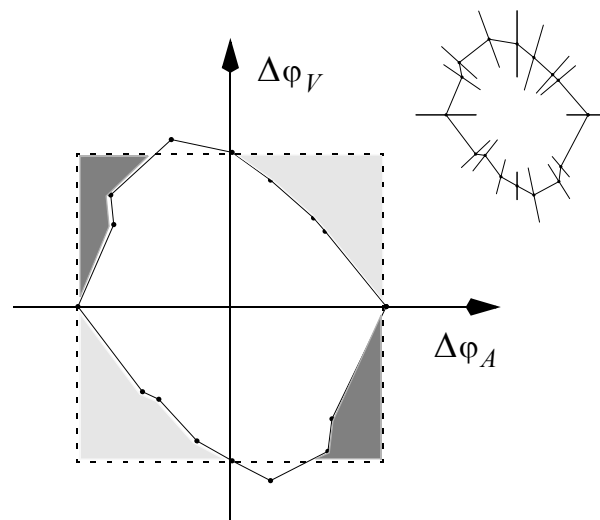


Abbildung 45: Resultate gemittelt über 5 Versuchspersonen, oben rechts ist die Standardabweichung eingetragen. Der Referenzstimulus befindet sich am Koordinatenursprung. Die Flächen zwischen den gemessenen Differenzschwellen und den verlängerten unimodalen Schwellen markieren die quadrantenabhängige Sensitivitätssteigerung.

Zwei wesentliche Versuchsergebnisse lassen sich aus den Abbildungen entnehmen: (1) Die Probanden zeigen substantielle intersensorische Sensitivitätseffekte in Form von erniedrigten Schwellen-

len bei redundanter Amplitudenänderung gegenüber unimodaler Variation. (2) Die Größe der Schwellenerniedrigung hängt interessanterweise von den Vorzeichen der akustischen und der visuellen Differenzen ab. Besonders groß ist der Effekt im 1. und 3. Quadranten, also in Fällen gleichlautender Amplitudenänderung in beiden Sinnesmodalitäten. Wie mit den dunkel unterlegten Regionen nahe gelegt, kann dort der erzielte Redundanzvorteil nicht als rein statistischer Summationseffekt erklärt werden, vielmehr lässt er auf einen direkten neuronalen Summationsmechanismus schließen (vgl. Modellvorhersagen in Kap. 3.2).

#### 4.3.2 Experiment 2: Akustisch-visuelle Frontalbewegung

Im zweiten Versuch wurde die unabhängige, visuelle Variable dahingehend verändert, dass an Stelle von Helligkeitsänderungen nun Änderungen in der Größe des Stimulus detektiert werden mussten. Akustisch wurde wie in Experiment 1 die Intensität variiert. Die so gewählten akustisch-visuellen Parameter besitzen eine spezielle Semantik:

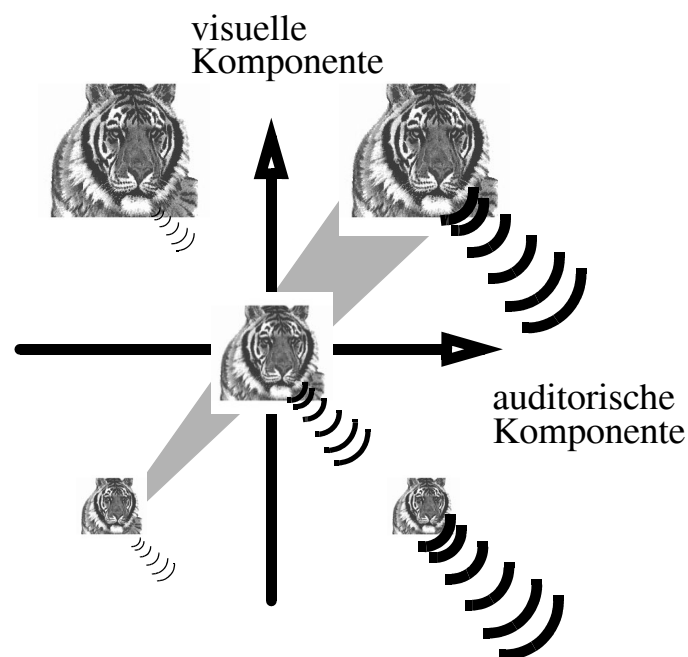


Abbildung 46: Illustration der möglichen akustisch-visuellen Änderungen bezogen auf einen Referenzstimulus im Koordinatenursprung.

Bei einer Größenänderung in der visuellen Stimuluskomponente zusammen mit einer passenden Amplitudenänderung in der akustischen Komponente, gewinnen die Versuchspersonen nämlich den Eindruck eines Schall aussendenden Objekts, das sich ihnen nähert oder sich von ihnen entfernt. Diese Situation ist für Stimuli aus dem 1. und 3. Quadranten gegeben.

Wie in Abbildung 46 dargestellt, sind in dem Versuchsdesign aber auch Situationen enthalten, die unter einer solchen Interpretation nicht kongruent sind. Dies sind Reizänderungen mit ungleichem Vorzeichen, also solchen aus dem 2. oder 4. Quadranten.

### 4.3.2.1 Methode

Der apparative Aufbau und die experimentelle Prozedur gleichen weitgehend denen in Experiment 1. Zur Operationalisierung einer Größenänderung wurde hier aber als visuelle Komponente ein graues Quadrat von  $2.5^\circ$  und  $4.7 \text{ cd/m}^2$  Luminanz vor dunklem Hintergrund mit  $3.6 \text{ cd/m}^2$  gewählt. Die akustische Komponente wurde gegenüber dem ersten Versuch nicht modifiziert.

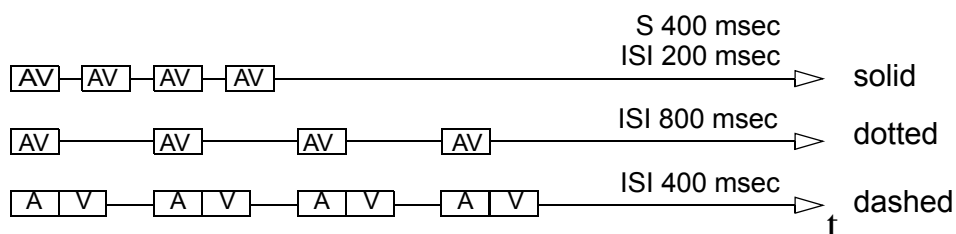


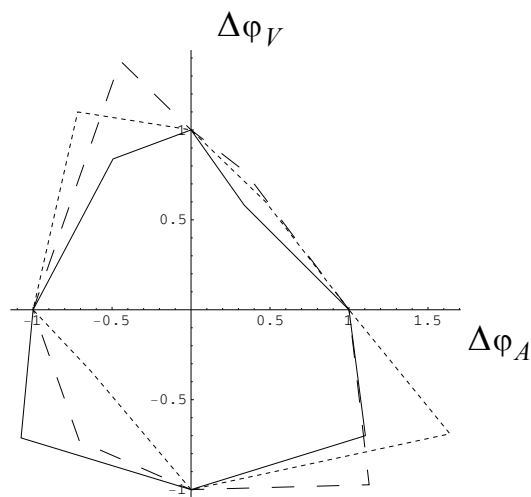
Abbildung 47: Zeitliche Bedingungen in Versuch 2: simultane Präsentation mit kurzem Interstimulusintervall (oben) und mit langem (Mitte). Konsekutive Reizdarbietung mit einem SOA von 0 msec (unten). Die Beschriftung rechts weist auf den Linienmodus in Abbildung 48 hin.

Eine weitere Variation betrifft den zeitlichen Ablauf der Darbietung. Hier gab es zwei zeitgleiche Darbietungsmodi von akustischer und visueller Komponente mit einer Darbietungszeit von 400 msec und einem ISI von 200 oder 800 msec. Daneben gab es noch eine weitere Bedingung, in der wir die zeitliche Koinzidenz aufgegeben haben,

indem die Komponenten nun sequentiell dargeboten wurden (Abbildung 47). Zuerst wurde jeweils die visuelle Komponente für 400 msec präsentiert, anschließend die akustische für ebenfalls 400 msec. Das ISI betrug in dieser Bedingung 400 msec. Für jede dieser zeitlichen Bedingungen wurden 8 bisensorische Schwellen gemessen.

#### 4.3.2.2 Ergebnisse

In 3 Sitzungen von je etwa einer Stunde Dauer wurden alle 8 Schwellen zweimal gemessen. Dazu standen drei bezahlte, naive Versuchspersonen (2 weiblich, 1 männlich) zur Verfügung.



*Abbildung 48:* Versuchsergebnisse für unterschiedliche zeitliche Bedingungen. Die Linienart korrespondiert zu der Legende in Abbildung 47.

In Abbildung 48 sind die gemittelten Ergebnisse für die drei zeitlichen Bedingungen angetragen, jeweils normiert auf die ermittelten unimodalen Schwellen. Die resultierende 2-dimensionale Schwellenkurve zeigt auch hier zum Teil drastische Verbesserungseffekte unter redundanten Stimulusbedingungen. Es konnte dagegen kein Einfluss der Präsentationsart (simultan oder sequentiell) festgestellt werden. Auch die Verlängerung des ISI von 200 msec auf 800 msec hat die Detektionsleistung nicht beeinflusst.

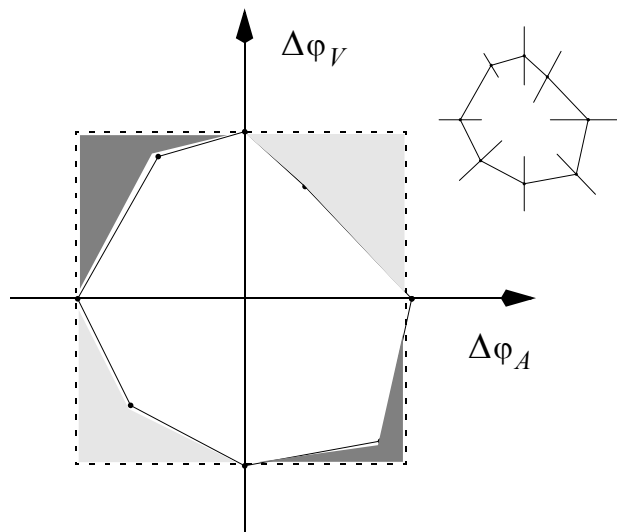


Abbildung 49: Gemessene Inkrementalschwellen, gemittelt über Versuchspersonen und Versuchsbedingungen (rechts oben mit der Standardabweichung). Zur Semantik der Quadranten vgl. Abbildung 46.

Über diese Bedingungen wurde deshalb in Abbildung 49 gemittelt. Deutlich zeigt sich wieder ein quadrantenspezifischer Effekt mit einer erheblichen Sensitivitätssteigerung für Stimuli mit den kongruenten Reizänderungen lauter / größer (erster Quadrant) und leiser / kleiner (zweiter Quadrant) sowie einer weniger stark ausgeprägten Verbesserung in den inkongruenten Bedingungen leiser / größer (dritter Quadrant) und lauter / kleiner (vierter Quadrant). Diese Effekte (Röhrbein, Zetsche, 2000) wurden von Hofbauer bei einer detaillierten Ausmessung mit 28 Schwellen bestätigt (Hofbauer et al., 2001).

### 4.3.3 Multisensorisches ON-OFF Modell

In den ersten beiden Experimenten weichen die ermittelten bisensorischen Inkrementalschwellen für die ökologisch validen Quadranten deutlich von den Vorhersagen des Modells der Wahrscheinlichkeitssummutation (Abbildung 31) ab. Das Ausmaß an Sensitivitätssteigerung kann in diesen Fällen mit dem Modell der neuronalen Integration erklärt werden (Abbildung 33) und ist als Beleg für eine wahre bisensorische Summation der eingehenden Signale auf

einer relativ frühen Stufe im System zu sehen. Dies gilt aber nur für die Quadranten mit gleichzeitigem Inkrement oder Dekrement in der akustischen und in der visuellen Komponente. In den übrigen, inkongruenten Situationen können die Resultate durchaus im Sinne eines statistischen Effektes erklärt werden. Keines der Modelle ist jedoch in der Lage, den quadrantenspezifischen Effekt selber zu erklären.

### selektive Aktivierung: ON/OFF Modell

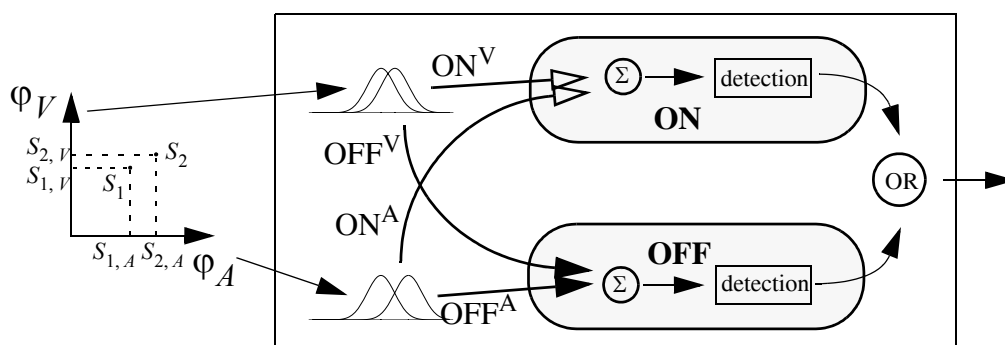


Abbildung 50: Vorschlag für ein Modell zur vorzeichenabhängigen Interaktion. Dieses ON-OFF Modell sieht für gleichlautende Differenzen (Inkrement-Inkrement oder Dekrement-Dekrement) eine neuronale Summation vor, in den übrigen Fällen aber eine Wahrscheinlichkeitssummation.

Deshalb wird hier ein neuartiges selektives Integrationsmodell vorgeschlagen, das beide Sorten von Interaktionseffekten beschreiben kann (Abbildung 50).

Dieses Modell erlaubt die signalspezifische neuronale Integration von Informationen aus dem akustischen und dem visuellen Kanal durch Annahme von zwei Verarbeitungspfaden, einem ON- und einem OFF-Pfad. Es ist an der ON-OFF Struktur des visuellen Kanals orientiert, die durch getrennte Verarbeitung von erregenden und hemmenden Reizen (verbunden mit einer Gleichrichtung) zu getrennten Subsystemen führt (vgl. Kapitel 2.2, siehe Schiller, 1992). Dieses Verarbeitungsprinzip wurde nun von mir zur Modellierung der

quadrantenspezifischen Interaktionseffekte adaptiert. Die Signale werden wie folgt verrechnet:

Bei einem Inkrement  $ON^V$  im visuellen Kanal und einem Inkrement  $ON^A$  im auditorischen Kanal werden die Signale vor einer Entscheidungsstufe im ON-System linear summiert. Im OFF-System liegt kein Signal an und die Auswertung des Oder-Operators ist damit identisch zum Ausgang des ON-Systems. In diesem Fall (und analog im Falle eines Dekrements in beiden Modalitäten  $OFF^A$ ,  $OFF^V$ ) reduziert sich das ON-OFF Modell auf das neuronale Summationsmodell. Besteht der Stimulus jedoch aus einem visuellen Inkrement  $ON^V$  und einem akustischen Dekrement  $OFF^A$ , aktiviert der visuelle Stimulus das ON-System und der akustische Stimulus das OFF-System. Da keine weiteren Eingänge belegt sind, werden beide Signale der jeweiligen Entscheidungsstufe zugeführt und erst danach mit dem Oder-Operator kombiniert. In diesem Fall (und äquivalent für die Signale  $OFF^V$  und  $ON^A$ ) entspricht das ON-OFF Modell dem Wahrscheinlichkeitssummationsmodell (Abbildung 51).

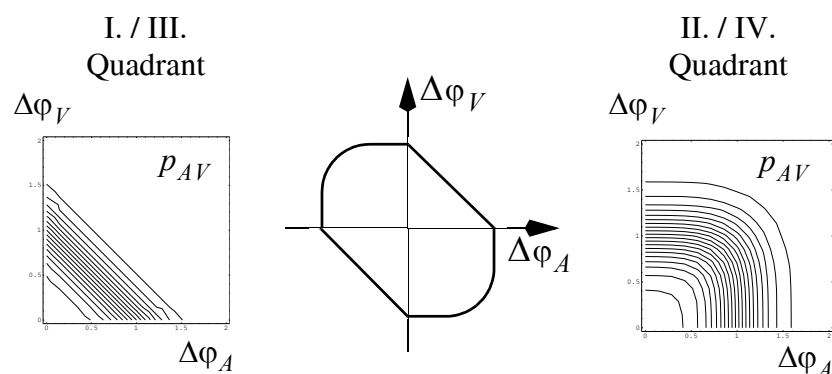


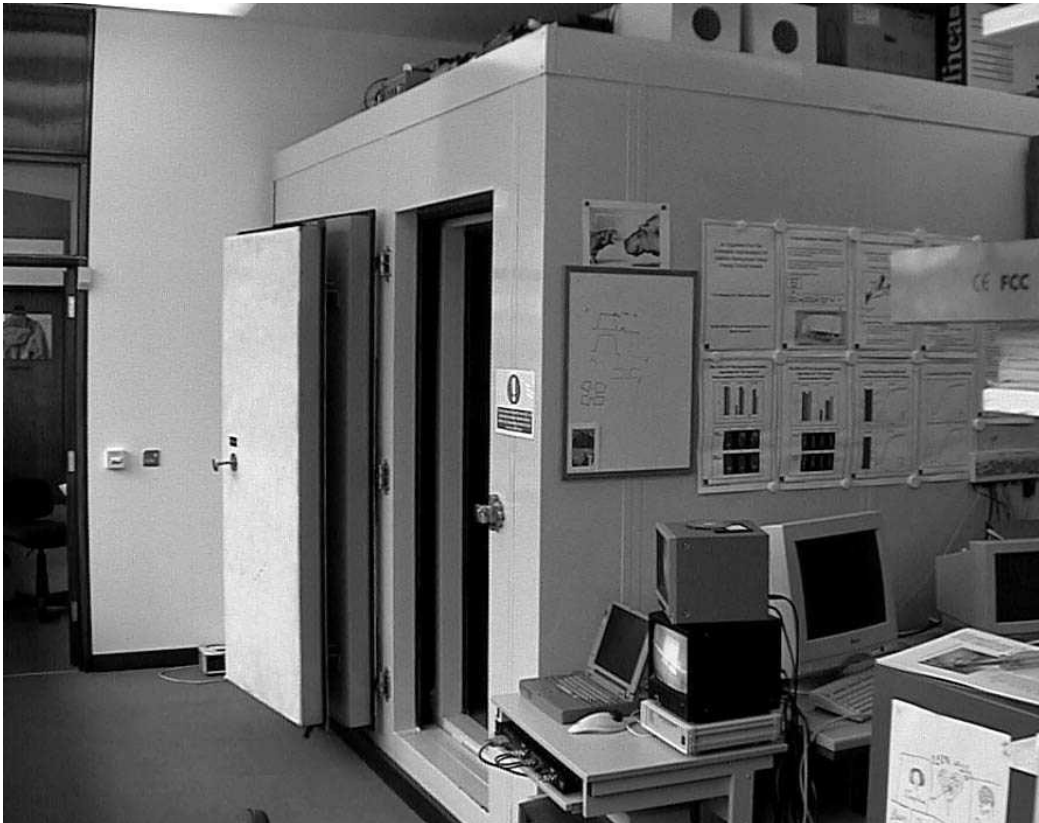
Abbildung 51: Sensitivitätsvorhersagen des ON-OFF Modells: Die vorzeichenabhängige Integration führt für multisensorische Stimuli zu quadrantenspezifischem Systemverhalten.

Im hier vorgeschlagenen Modell ist damit eine vorzeichenabhängige, multisensorische Verarbeitung gewährleistet, bei der nur die ON-ON und die OFF-OFF Kombinationen tatsächlich neuronal integriert werden. Dadurch wird eine optimale Anpassung der sensorischen Informationsverarbeitung an die multimodale statistische Struktur erreicht.



#### 4.3.4 Experiment 3: Akustisch-visuelle Trajektorien

Um die verwendeten Stimuli ökologisch valider zu machen, wurde für die nächsten Experimente ein sehr viel komplexerer apparativer Aufbau benutzt. Sowohl die akustische als auch die visuelle Komponente sollten an unterschiedlichen Positionen im Raum auftauchen können, zudem war eine deutlich längere Bewegungssequenz wünschenswert. Erreicht wurde dies mit einem Versuchsaufbau mit 31 Lautsprechern und 31 LEDs, die um die Versuchsperson herum in einem Halbkreis auf Kopfhöhe montiert wurden.



*Abbildung 52:* Außenansicht des schalldichten Raumes im MacKay Institute of Communication and Neuroscience (Keele University, England). Hierin fanden die Versuche 3 und 4 statt, vor dem Raum befinden sich Module zur Signalgenerierung und zur Steuerung des Versuchsablaufs.

#### 4.3.4.1 Methode

Der Versuch fand in einem abgedunkelten, schalldichten Raum vom Typ IAC1204A statt (Abbildungen 52 und 53). Lautsprecher und LEDs waren auf einem horizontalen Bogen in Abständen von 5 Grad montiert, jeweils ein Lautsprecher und eine LED übereinander (vertikaler Abstand 3 cm). Der Bogen befand sich in 1 m Höhe, sodass die Aktoren auf einer Höhe mit dem Kopf der sitzenden Versuchsperson lagen. Der Abstand der Apparatur zur Versuchsperson betrug etwa 1.3 m.



*Abbildung 53:* Apparativer Versuchsaufbau. Die Armatur zur akustisch-visuellen Reizung befindet sich auf einem Halbkreis in Gesichtshöhe um den Sitz der Versuchsperson herum.

Durch selektive Ansteuerung der 31 Positionen können Objektbewegungen von 180 Grad um den Beobachter herum simuliert werden. In den folgenden Versuchen wurden binäre Signale verwendet, d.h. die visuelle Komponente bestand aus dem Aufleuchten von LEDs,

während akustisch ein Klicken zu hören war. Zur Simulation einer Bewegung lag zu jedem Zeitpunkt an genau einer Position eine Spannung an, womit das SOA zwischen zwei Positionen genau der Einschaltdauer an dieser Position entsprach. Generiert wurden die Signale von einem Echtzeit-Signalprozessor vom Typ Tucker-Davis RP2.

Die Aufgabe der Versuchspersonen bestand nun darin, eine horizontale Bewegung zu detektieren. Dazu wurden gemäß eines 2AFC-Designs hintereinander jeweils zwei Intervalle dargeboten, von dem in randomisierter Weise eines den bewegten Stimulus enthält, das andere eine Zufallssequenz. Bewegungs- und Zufallssequenz waren in Hintergrundrauschen eingebettet, und dieses wurde von Durchgang zu Durchgang je nach vorheriger Antwort der Versuchsperson variiert. Die Antwort erfolgte durch Drücken von einer von zwei Tasten, wobei ein Drücken der oberen (unteren) Taste bedeutete, dass eine Bewegungssequenz im ersten (zweiten) Intervall gesehen wurde.

Die Bewegungssequenz beschrieb stets einen 90 Grad Bogen mit konstanter Geschwindigkeit von etwa 30 deg/sec und konnte im linken oder im rechten Halbfeld vor dem Beobachter erscheinen. Sie konnte vom Fixationspunkt in der Mitte nach außen verlaufen oder von außen nach innen. Da die visuelle und die akustische Komponente unabhängig voneinander manipuliert wurden, konnten die beiden Bewegungskomponenten in die gleiche Richtung oder in verschiedene Richtungen laufen, in demselben Halbfeld präsentiert sein, oder in verschiedenen. Zusammen mit unimodalen Durchgängen gab es insgesamt 24 Bedingungen unterschiedlicher akustisch-visueller Kohärenz (Abbildung 54).

Die Versuchsperson hatte jeweils zu entscheiden, in welchem Intervall sie eine Bewegung wahrgenommen hat, unabhängig von Modalität (visuell / auditorisch / bisensorisch), Bewegungsrichtung (links / rechts) und Halbfeld (links / rechts). Um eine Interaktion gemäß des Simon-Effekts zwischen dem Halbfeld, in der die Bewegung verlief, und der motorischen Reaktion zu vermeiden, wurden die beiden Antworttasten vertikal angeordnet.

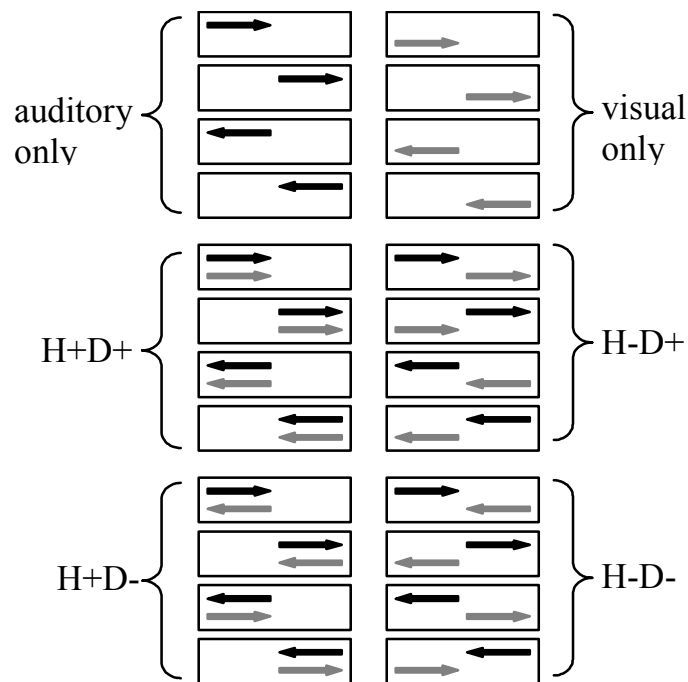


Abbildung 54: Das Experiment umfasst 24 Versuchsbedingungen: acht unimodale Bedingungen (oben) und sämtliche bimodale Kombinationsmöglichkeiten für die Parameter Hemisphäre (H) und Bewegungsrichtung (D). Diese Bedingungen sind nach Kongruenz gruppiert, z.B. steht H+D- für die Klasse der akustisch-visuellen Stimuli, bei denen die akustische und die visuelle Komponente in derselben Hemisphäre gezeigt wurden, jedoch mit unterschiedlicher Bewegungsrichtung.

Um nun die Schwellen für akustisch-visuelle Bewegung in den verschiedenen Bedingungen zu messen, wurde das Hintergrundrauschen variiert, in welchem das Bewegungs- und auch das Zufalls-signal eingebettet waren. Dieses Rauschen bestand aus einer zufällig ausgewählten Menge von Orts-Zeit-Punkten, die zusätzlich zur festgelegten Sequenz aktiviert wurden.

Einen space-time Plot für die Versuchsbedingung H+D- (gleiches Halbfeld, unterschiedliche Richtung) zeigt Abbildung 55: Die beiden sensorischen Signale sind getrennt dargestellt, beide Sequenzen wer-

den in Richtung der Zeitachse von oben nach unten präsentiert. Auf den jeweils rechten Seiten ist das Hintergrundrauschen zu sehen, links ist es überlagert von entweder der Bewegungssequenz (oben, signal trial) oder der Zufallssequenz (unten, noise trial). Dieser “noise trial” setzt sich also aus zweierlei Rauschen zusammen: der Zufallssequenz und dem Hintergrundrauschen. Er enthält damit gleich viel Energie wie der “signal trial”.

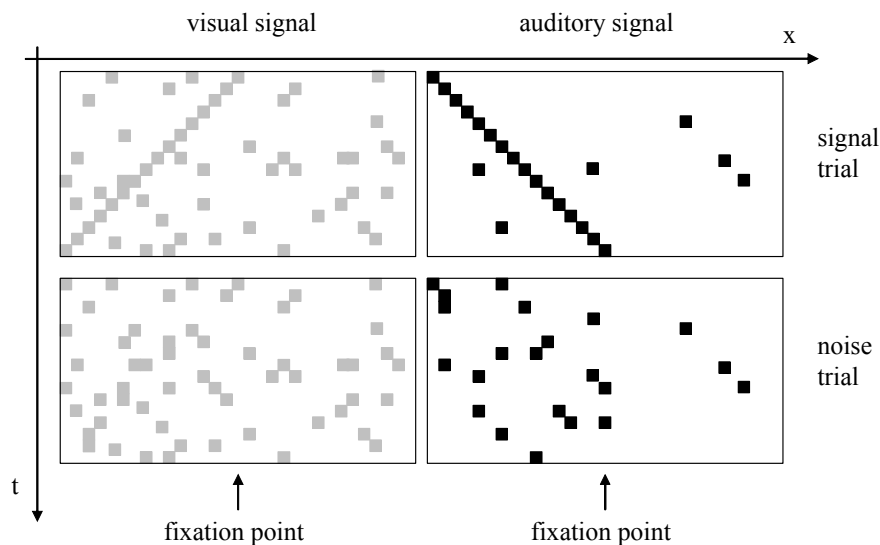


Abbildung 55: Räumliche Struktur und zeitlicher Verlauf einer Stimuluspräsentation. Die räumlichen Positionen sind entlang der x-Achse für beide Modalitäten separat angetragen, der Zeitpfeil läuft dabei nach unten. Dargestellt ist die Reizbedingung H+D-. Für Details siehe Text.

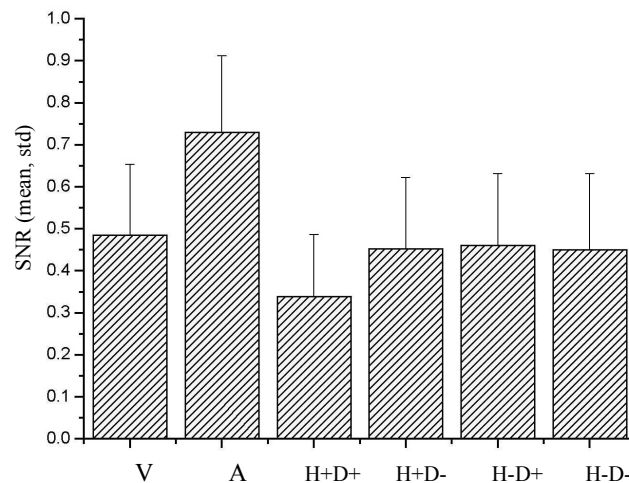
Die Rauschenergie ist die abhängige Variable und wird in Versuch 3 und 4 durch die adaptive Prozedur QUEST (Watson, Pelli, 1983) kontrolliert. Die zurückgelieferten Intensitätswerte können zusammen mit der (konstanten) Signalenergie als Signal-Rausch-Verhältnis (je Halbfeld) interpretiert werden, mit  $0 < \frac{S}{S+N} \leq 1$ .

Da sich das visuelle System durch eine deutlich höhere Sensitivität bezüglich räumlicher Lokalisation auszeichnet, wurde ein Faktor eingeführt, um für den akustischen Kanal eine Rauschenergie zu

erhalten, die zu annähernd vergleichbarer Detektionsleistung führt. Dieser akustisch-visuelle Rauschfaktor wurde auf 0.15 festgesetzt, das akustische Hintergrundrauschen betrug also 15% des visuellen, vergleiche mit Abbildung 55.

#### 4.3.4.2 Ergebnisse

Die 24 zu messenden Bewegungsschwellen wurden für jede Versuchsperson in 6 Sitzungen zu je 30 Minuten bestimmt. Innerhalb einer Sitzung wurden die Schwellenmessungen parallel durchgeführt, indem die Stimulusdarbietungen der verschiedenen Versuchsbedingungen randomisiert wurden. Je Schwellenmessung wurden mindestens 20 solcher Darbietungen generiert.



*Abbildung 56:* Gemessene Bewegungsschwellen für 10 Versuchspersonen. Bisensorische Stimulation führt nur in der Kongruenz-Kongruenzbedingung (H+D+) zu einer deutlichen Leistungsverbesserung.

Die Ergebnisse für alle 10 Versuchspersonen sind in Abbildung 56 dargestellt. Auf der Ordinate ist das Signal-Rausch-Verhältnis (bezogen auf den visuellen Anteil) angetragen, auf der Abszisse die einzelnen Versuchsbedingungen: die beiden unisensorischen Bedingungen (V, A) und die vier bisensorischen Situationen mit gleicher

Hemisphäre und Bewegungsrichtung (H+D+), Übereinstimmung in nur einem dieser Parameter (H+D-, H-D+) und der Bedingung mit ungleicher Hemisphäre und unterschiedlicher Bewegungsrichtung.

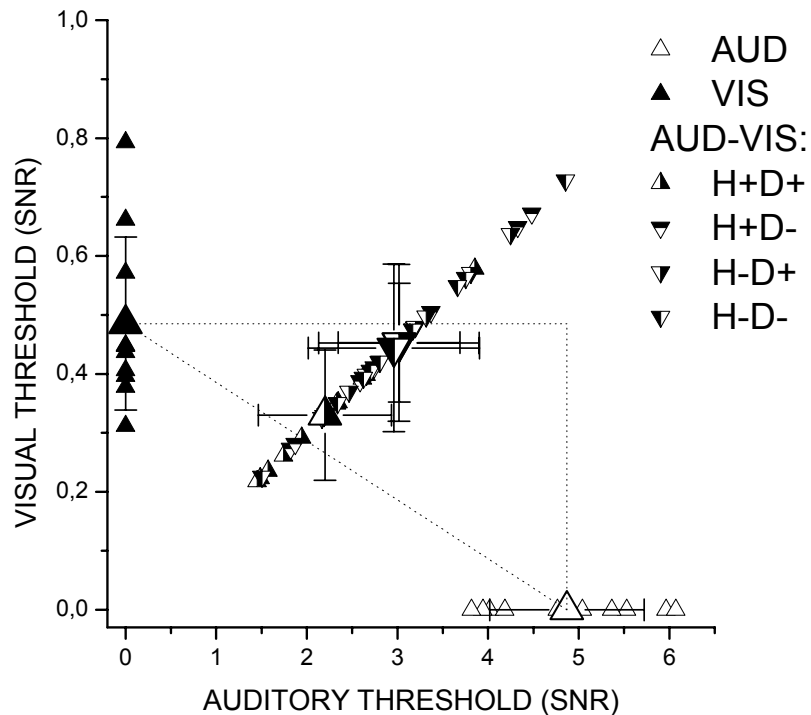


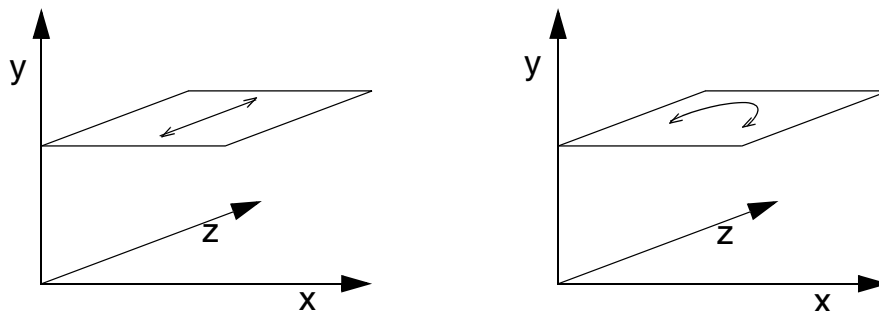
Abbildung 57: Die auf die unimodalen Schwellen normierten Versuchsergebnisse deuten auf einen statistischen Effekt für die inkongruenten Bedingungen hin, während die gemessene Schwelle in der Bedingung H+D+ auf eine lineare Summation schließen lässt.

Da keine Unterschiede innerhalb dieser sechs Gruppen festgestellt wurden, sind die Ergebnisse der jeweiligen vier Unterbedingungen zusammengefasst (vgl. Abbildung 54). Die Bewegungsrichtung (inwärts vs. auswärts) und die Lage des Halbfelds (links vs. rechts) hatten für sich keinen Einfluss auf die Detektionsleistung, entscheidend war allein der Grad der Übereinstimmung dieser Parameter in den beiden Modalitäten.

Die Ergebnisse zeigen klar, dass die beste Detektionsleistung in der kongruenten Bedingung (H+D+) erreicht wird, bei der Bewegungs-

richtung und Halbfeld für die akustische und die visuelle Komponente übereinstimmen. In allen anderen bisensorischen Bedingungen ist keine Leistungsverbesserung gegenüber der visuellen Bedingung feststellbar.

Entscheidend für die Argumentation ist der Vergleich zwischen bi- und unisensorischen Versuchsbedingungen. Da wir an der relativen Leistungsverbesserung interessiert sind, wurden in Abbildung 57 die Daten auf die unisensorischen Detektionsschwellen (Bedingung V und A) normiert. Eine substantielle Verbesserung durch die bisensorische Information ist demnach nur in der kohärenten Bedingung zu beobachten, in allen anderen lassen sich die geringfügig niedrigeren Schwellen als statistische Effekte erklären.



*Abbildung 58:* Schematische Darstellung des in Experiment 2 (links) und Experiment 3 (rechts) adressierten Bewegungsbereichs. Die Achsen beschreiben hier den vor einer Versuchsperson liegenden Raum.

#### **4.3.5 Experiment 4: Akustisch-visuelles Rezeptivfeld**

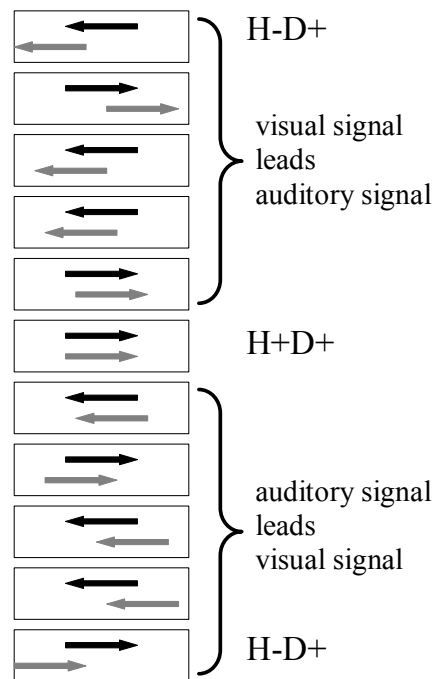
In einem weiteren Versuch wurde der Frage nachgegangen, wie groß der Toleranzbereich ist, über den Bewegungssignale aus verschiedenen Modalitäten effizient integriert werden können. Zur Übersicht sind in Abbildung 58 die möglichen Bewegungsrichtungen aus den beiden vorangegangenen Experimenten noch einmal dargestellt. Hier repräsentieren die Achsen reale Richtungen im Raum: In Experiment 2 erfolgte die Bewegung ausschließlich entlang der y-Achse, während sich der Stimulus im letzten Experiment auf einem



Halbkreis auf einer x-z-Ebene bewegen konnte. Dabei gab es Bedingungen mit kohärenter Bewegungsinformation und solche, bei denen die Bewegung dieselbe Richtung hatte, aber in verschiedenen Halbfeldern stattfand.

#### 4.3.5.1 Methode

In dem folgenden Experiment soll nun untersucht werden, wie sich die Detektionsleistung zwischen diesen beiden "Extrempunkten" (H+D+ und H-D+) verhält. Die akustisch-visuelle Bewegung verlief bei diesem Versuch stets in dieselbe Richtung, jedoch mit unterschiedlich großem Versatz. Dieser wurde in 11 Versuchsbedingungen von 0 bis 75 deg vis variiert (Abbildung 59).



*Abbildung 59:* Zur Untersuchung der Reichweite der Integrationsleistung wurden 11 bisensorische Stimulusbedingungen mit unterschiedlichem räumlichen Versatz verwendet. Über Hemisphäre und Bewegungsrichtung wurde randomisiert.

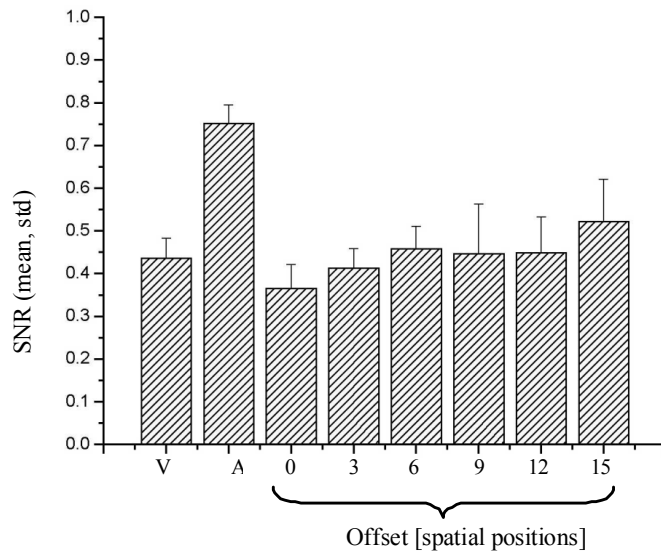
Der räumliche Umfang der Trajektorie wurde auf 60 deg arc reduziert, weshalb sich die Präsentationszeit auf 1.4 sec. verkürzte bei gleich bleibender Geschwindigkeit von 30 deg/sec. Zusätzlich wurden wieder die beiden korrespondierenden unimodalen Schwellen gemessen. Der akustisch-visuelle Rauschfaktor wurde wieder auf 0.15 gesetzt. Die Anfangsposition der Trajektorien wurde zusätzlich randomisiert, um zu vermeiden, dass sich die Versuchspersonen bestimmte LED-Positionen zur Lösung der Aufgabe zu Hilfe nehmen. Da im Experiment 3 keine Unterschiede zwischen links- und rechtsgerichteter Bewegung und zwischen linkem und rechtem Halbfeld gefunden wurden, konnte in diesem Experiment über diese beiden Parameter randomisiert werden.

#### **4.3.5.2 Ergebnisse**

Die resultierenden Detektionsschwellen gemittelt über alle fünf Versuchspersonen sind in Abbildung 60 gezeigt. Deutlich erkennbar ist der Zusammenhang zwischen Größe des räumlichem Versatzes und gemessener Empfindlichkeit: Sind die Trajektorien des visuellen und des akustischen Reizes identisch (Offset = 0), ist die Leistung der Versuchspersonen am besten.

Ein zunehmender Versatz erfordert dagegen ein immer größeres Signal-Rausch Verhältnis, damit der Reiz erkannt werden kann. Eine Verschiebung um 15 Lautsprecherpositionen, also um 75°, hat dabei sogar einen hemmenden Effekt. Die relativ schlechte Leistung in der unimodalen auditorischen Bedingung ist hier nicht von Bedeutung, da sie lediglich ein Hinweis darauf ist, dass die Wahl des akustisch-visuellen Rauschfaktors die überlegene visuelle Leistung nicht vollständig kompensieren konnte.

Vergleicht man die hier erzielten Ergebnisse mit denen aus Experiment 3 (Bedingung H-D+), lässt sich eine leichte Verschlechterung der gemessenen Schwellen feststellen. Dies ist vermutlich darauf zurückzuführen, dass der Umfang der Trajektorie von 90 deg (Experiment 3) auf 60 deg reduziert wurde, um die Anfangspositionen der Trajektorien randomisieren zu können. Eine damit verbundene Reduktion der Stimulusdauer führt bekanntlich zu schwierigeren Detektionsbedingungen. Die Parameter der adaptiven Prozedur



*Abbildung 60:* Ergebnisse von 5 Versuchspersonen in den beiden unimodalen Bedingungen und mit unterschiedlichem Ortsversatz in den bisensorischen Reizvorlagen.

waren für beide Versuchen identisch (Weibullfunktion mit  $\beta=3.5$  und  $\delta=0.05$ ).

Die Ergebnisse lassen sich auch getrennt nach relativer Position darstellen, so dass man zu einer Darstellung im Sinne eines Rezeptivfeldes gelangt (Abbildung 61). Auf der Abszisse ist die Verschiebung in Sehwinkel angegeben, die Werte auf der Ordinate geben die ermittelten Schwellen als visuelles (links) und akustisches (rechts) Signal-Rausch Verhältnis an. Die Linie ist ein Gaußscher Fit an die Daten. Dessen Breite von etwa 20 deg vis kann als Größe des rezeptiven Feldes interpretiert werden.

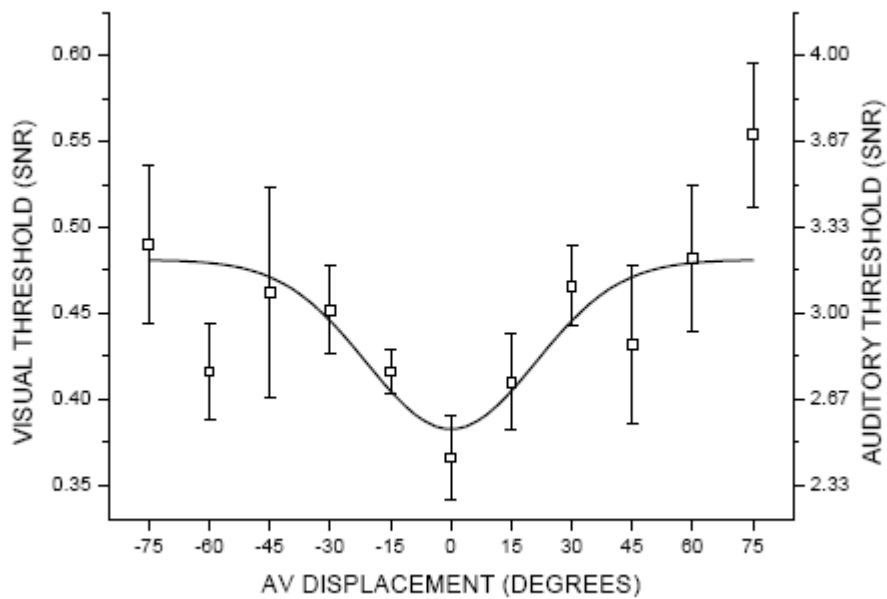


Abbildung 61: Gemessene akustisch-visuelle Schwellen in Abhängigkeit vom spatialen Versatz. Die beiden äußeren Messungen entsprechen den Bedingungen H-D+ (vgl. Abbildung 59). Die unimodalen Werte waren 0.75 (auditorisch) und 0.44 (visuell).

## 4.4 Diskussion

In der Standardreferenz zur sensorischen Interaktion (Welch und Warren, 1986) wird "sensory interaction" folgendermaßen definiert:

*“We define it [intersensory interaction] as the situation in which the perception of an event as measured in terms of one sensory modality is changed in some way by the concurrent stimulation of one or more other sensory modalities.”*

Diese Definition ist meiner Meinung nach unvollständig, da hier der Redundanzaspekt völlig unberücksichtigt bleibt. Die potentielle Beeinflussung einer Modalität durch eine andere ergibt sich bereits durch die Tatsache, dass sich die rezeptiven Felder von Neuronen zweier Modalitäten überlappen können. Deshalb muss natürlich untersucht werden, inwieweit beispielsweise die visuelle Leistung durch simultane Stimulation auditiver Bahnen beeinflusst wird (v.a. im Sinne einer Beeinträchtigung). Die Möglichkeit aber, dass Information durch polymodale Objekte und Ereignisse redundant vorliegt und, dass dieser Umstand vom Organismus ausgenutzt werden kann, ist in obiger Definition nicht enthalten. In der Tat laufen aber viele Versuche unter der Perspektive der gegenseitigen Beeinflussung ab, so beispielsweise auch die Versuche zum so genannten McGurk-Effekt (McGurk, MacDonald, 1976). Dieser tritt beim Sprachverstehen gerade dann auf, wenn die akustische Information (“ba-ba”) mit der visuellen (“ga-ga”) im Widerspruch steht (wahrgenommen wird bei diesem Beispiel “da-da”).

Typischerweise ist in akustisch-visuellen Experimenten auch nur eine Modalität aufgabenrelevant, die andere soll von der Versuchsperson ignoriert werden. Gemessen werden die dennoch auftretenden Interferenzen (siehe z.B. Soto-Faraco, 2003 und Tabelle 4.1 in Soto-Faraco, Kingstone, 2004). Für das Versuchsdesign hier war dagegen entscheidend, dass stets beide Modalitäten aufgabenrelevant sind. Daneben ist es erforderlich, dass den Versuchspersonen eine Aufgabe gestellt wird, für die sie tatsächlich Information aus beiden Sinnen benutzen können. So wurden in Untersuchungen von

Eijkman, Vendrik (1965) und Osborn et al., (1963), bei denen die Versuchspersonen in jeder Modalität eine eigene Aufgabe durchführen mussten, keine Leistungsverbesserung festgestellt. In den Versuchsreihen hier wurde deshalb stets nur nach dem Intervall gefragt, in der der vorher festgelegte Stimulus wahrgenommen wurde, unabhängig davon, ob diese Detektion visuell, akustisch oder bimodal erfolgte.

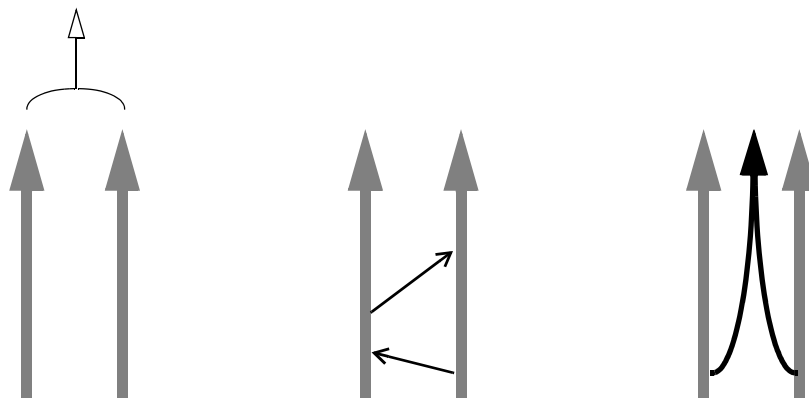
Die hier angewandten Versuchsdesigns haben des Weiteren den Vorteil, dass die Aufmerksamkeit stets auf beide Sinne gerichtet ist. Da Versuchsdurchgänge mit bimodaler Stimulusänderung und solche mit unimodaler Änderung stets vermischt waren, können die gemessenen Schwellen direkt miteinander verglichen werden, ohne dass aufmerksamkeitsbedingte Unterschiede zu berücksichtigen sind.

Bei Durchsicht der vereinzelt Literatur zur akustisch-visuellen Interaktion stellt man fest, dass sich fast alle Arbeiten entweder mit sehr "niederen" oder sehr "hohen" Leistungen befassen. So gibt es einerseits relativ viele Arbeiten im Kontext einfacher Reaktionszeitexperimente, wie z.B. Vigilanzstudien oder Experimente mit Orientierungsreaktionen und andererseits eine Reihe von Untersuchungen zur akustisch-visuellen Sprachwahrnehmung ("speech-reading", Übersicht in Massaro, 1998) und zur Wahrnehmung von Emotionen (z.B. de Gelder, Vroomen 2000). Zwischen diesen Gebieten klafft jedoch eine beträchtliche, empirische Wissenslücke (vgl. dazu Box 1 in Calvert et al., 1998). Die Versuchsaufgaben hier waren deshalb so gewählt, dass Wahrnehmungsprozesse auf einer frühen kortikalen Verarbeitungsstufe angesprochen werden und die Möglichkeit eines Bias und anderer, später Interaktionseffekte ausgeschlossen werden kann.

Auch in dem aktuellen Buchbeitrag von Soto-Faraco und Kingstone (2004) zur multisensorischen Interaktion von dynamischer Information steht die Frage nach der Verarbeitungshöhe im Vordergrund. Sie kommen zu dem Schluss, dass akustische und visuelle Bewegungssignale tatsächlich auf einer frühen Stufe integriert werden können und dass diese Integrationsleistung nicht auf eine Interaktion zwischen statischen Ereignissen reduziert werden kann. Deutlich wird in dieser Übersichtsarbeit wieder, dass ein großer Teil bisheriger Experimente darauf abzielte, wie zunächst unimodale Bewegungsin-

formation durch (statische oder dynamische) Ortsinformation in einer anderen Modalität beeinflusst werden kann. Diese Signale haben dabei ausdrücklich den Status von Distraktoren und werden nicht als redundante Informationskanäle verstanden.

In dem von mir vorgeschlagenen multisensorischen ON-OFF Modell (Abbildung 50) können die Ergebnisse aus Experiment 1 auch als Hinweis auf einen amodalen Verarbeitungsmechanismus für die Intensität von Reizen gedeutet werden. Für diesen Mechanismus ist es irrelevant, in welcher sensorischen Modalität eine Intensitätsänderung stattgefunden hat, entscheidend ist allein das Vorzeichen der Änderung. Die Intensität von Reizen ist neben der Qualität, Dauer und Lokalisation eine grundlegende Wahrnehmungsdimension, die allen Sinnesmodalitäten gemein ist. Amodale, kontinuierliche Repräsentationen, die die physikalischen Eigenschaften von distalen Quellen unabhängig von ihren proximal ausgelösten Empfindungen charakterisieren, spielen u.a. bei der direkt-realistischen Theorie der Wahrnehmung eine wesentliche Rolle (siehe Gibson, 1966).



*Abbildung 62:* Prinzipielle Verschaltungsformen zwischen sensorischen Kanälen: getrennte Verarbeitung mit späterer Integration (links), frühe Interaktion zwischen unabhängigen Kanälen (Mitte) und Fusion zu einem amodalen Kanal (rechts).

Auch bei der Sprache wird das Konzept der Amodalität vertreten. Die modalitätsübergreifende Repräsentation eines Buchstabens etwa kann als direkte audiovisuelle Integration von Phonemen und Graphe-

men ohne eine gemeinsame Zwischenrepräsentation verstanden werden (z.B. Raij et al., 2000). Die übrigen im Bereich der audiovisuellen Sprachwahrnehmung postulierten Modelle lassen sich analog zu der hier bereits getroffenen Einteilung charakterisieren (vgl. Abbildung 62): Modelle mit früher Integration (hier mit einer präphonetischen Repräsentationsebene) und Modelle mit später Integration, d.h. mit einer gemeinsamen symbolischen Repräsentation, wobei das favorisierte Modell das sog. "Fuzzy Logical Model of Perception" (FLMP) zu sein scheint (Massaro, 1999)

Bei psychophysischen Experimenten besteht generell das Problem, wie "tiefe" sensorische Interaktionseffekte von anderen, einfacheren Phänomenen, wie z.B. unspezifische Vorwarnungseffekte, Wahrscheinlichkeitssummation oder späte kognitive Bias-Effekte unterschieden werden können. Gewisse psychophysische Ansätze wie die "Miller race inequality" in Reaktionszeitexperimenten oder die Anwendung von Bias-freien Forced-Choice-Messungen (Bertelson and Aschersleben 1998) oder Maximum-Likelihood-Schätzer (Ernst and Banks, 2002) erweisen sich hier als nützlich, es fehlen jedoch bislang quantitative Beschreibungen der Regeln zur Fusion von modalitätsspezifischer Information.

In technischen Systemen ist der synergistische Gebrauch von multiplen Sensoren durch intelligente Systeme ein seit vielen Jahrzehnten etabliertes Fach: Die Fusion von Sensordaten wird in der klassischen KI-Forschung bei der Entwicklung von Inferenzstrategien berücksichtigt (Garvey, 1987) und in Formalismen zur Wissensrepräsentation (Pau, 1988) integriert. Speziell die Übertragbarkeit von biologischen Konzepten im Umgang mit multisensorischer, redundanter Information wurde in Bereich der Datenfusion vor einiger Zeit erkannt (z.B. Brooks, 1990). In der Robotik spielen die Synthese von Beobachtungen und die Kombination von Evidenzen eine zentrale Rolle (Durrant-Whyte, 1988, Thomopoulos, 1990, Abidi, Gonzalez, 1992) und seit 2000 erscheint für dieses Gebiet das Journal "Sensor Fusion".

Empirische Studien über multisensorische Prozesse machen dagegen nur einen verschwindend geringen Bruchteil vom gesamten sensorischen Forschungsgebiet aus. Es gibt von den geschätzten 150000 Fachzeitschriften (Gaus, 1988, Gröschel, Lügger, 1995)



keine, die sich ausdrücklichen mit dem Themenkomplex der multisensorischen Wahrnehmung befasst. Diese beachtliche empirische Lücke wurde in Abständen immer wieder bemängelt, systematische Untersuchungen fehlen aber nach wie vor. Erst in den letzten Jahren ist das Interesse an multisensorischen Fragestellungen dann auch im Bereich der Wahrnehmungspsychologie und in den Neurowissenschaften deutlich gestiegen (siehe Spence, Driver, 2004 und Calvert et al., 2004).

## 5. Zusammenfassung

In dieser Arbeit wurden zunächst theoretische Überlegungen zur mehrkanaligen Bildverarbeitung durch ortsfrequenzselektive Neurone im visuellen Kortex angestellt.

Dazu habe ich für das  $1/f^2$  Leistungsdichtespektrum natürlicher Bilder unterschiedliche Teilbandzerlegungen im Rahmen der Shannon'schen Rate-Distortion-Theorie untersucht. Meine Ergebnisse zeigen, dass eine logarithmische, selbstähnliche Zerlegung mit sehr viel geringerer Komplexität zu einer gewünschten Signalqualität führen kann, als eine lineare Partitionierungsstrategie. Eine detailliertere Analyse hat erbracht, dass eine leicht modifizierte, optimale Zerlegung existiert, die zu einem noch besseren Tradeoff führt. Diese Kanalaufteilung ist ungefähr selbstähnlich, jedoch mit einer systematischen Abnahme der logarithmischen Bandbreiten mit zunehmender Mittenfrequenz. Exakt dieser Effekt wurde auch in neurowissenschaftlichen Experimenten beobachtet.

Damit kann nun diese elementare, strukturelle Eigenschaft des Sehsystems als Ergebnis einer effizienten Anpassung an die Statistik zweiter Ordnung gedeutet werden (Röhrbein, Zetsche, 1998a und Röhrbein, Zetsche, 1998b). Das Ergebnis unterstützt die Hypothese, dass biologische sensorische Systeme optimal an die statistischen Redundanzen der natürlichen Umgebung adaptiert sind.

Das Prinzip der Angepasstheit gilt darüber hinaus auch für multisensorische Integrationsmechanismen, wie in einer Reihe von psychophysischen Experimenten mit akustisch-visuellem Stimulusmaterial demonstriert werden konnte (Röhrbein, Schill, Zetsche, 2000 und Röhrbein, Zetsche, 2000). Akustisch-visuelle Kovariationen zwischen Größe, Schalldruck und spatio-temporalen Position eines Objekts, aber auch gemeinsame Intensitätsänderungen waren hier Gegenstand der Untersuchung. Es wurde klare Evidenz für einen hochselektiven, multisensorischen Integrationsmechanismus gefunden. Dieser ist durch große Interaktionseffekte für ökologisch relevante Stimuluskombinationen und durch kleine Effekte für statistisch untypische Kombinationen charakterisiert.

Es konnte darüber hinaus gezeigt werden, dass diese Integration kritisch von der Koinzidenz und Kolo­kalisierung der Signalkomponenten abhängt. In der Tat sind die gefundenen Detektionsschwellen nur dann signifikant reduziert, wenn die akustisch-visuellen Bewegungssignale in dieselbe Richtung und in derselben Hemisphäre verlaufen. In einem weiteren Experiment wurde gezeigt, dass die Sensitivität systematisch mit ansteigendem räumlichen Abstand abfällt (Röhrbein, et al. 2003a und Röhrbein et al., 2003b).

Insgesamt liefern die Versuchsergebnisse ein klares Gegenargument zu der klassischen Idee von grundlegend getrennten, modalitätsspezifischen sensorischen Kanälen, die nur in Form von Bias-Effekten oder auf höherer Entscheidungsebene miteinander interagieren. Zudem habe ich gezeigt, dass die modalitätsspezifischen Signale ökologisch valide, in Raum und Zeit kohärente Stimuluskonfigurationen darstellen müssen, um effektiv integriert zu werden. Da inkongruente Reizkomponenten mit hoher Wahrscheinlichkeit von unterschiedlichen Objekten stammen, wird durch die Selektivität der Sensitivität verhindert, dass solche Objektattribute beim "feature binding" aufeinander bezogen werden.

## 6. Literatur

Abidi, M.A., Gonzalez, R.C. (eds.) (1992). *Data Fusion in Robotics and Machine Intelligence*. Academic Press.

Aiello, G.L., Bach-y-Rita, P. (2000). The cost of an action potential. *Journal of Neuroscience Methods*, 103: 145-149.

Akansu, A.N., Haddad, R.A. (1996). Fundamentals and optimal design of subband and wavelet transforms. In: Akansu, A.N., Smith, M.J.T. (eds). *Subband and Wavelet Transforms*. Kluwer Academic Publishers, 33-82.

Alais, D., Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research* 19: 185-194.

Albrecht, D., Hamilton, D.B. (1982). Striate cortex in monkey and cat: Contrast response function. *Journal of Neurophysiology*, 48: 217-237.

Ames, A. (2000). CNS energy metabolism as related to function. *Brain Research Review*, 34: 4-86.

Atick, J.J. (1992). Could information theory provide an ecological theory of sensory processing. *Network*, 3: 213–251.

Atick, J.J., Redlich, A.N. (1990). Towards a theory of early visual processing. *Neural Computation*, 2:308–320.

Attias, H., Schreiner, C.E. (1998). Coding of naturalistic stimuli by auditory midbrain neurons. *Advances in Neural Information Processing Systems*, 10, MIT Press.

Bak, P., Tang, C., Wiesenfeld, K. (1987). Self-Organizing Criticality: An Explanation of  $1/f$  Noise. *Physical Review Letters* 59(4): 381.

Barabell, A., Crochiere, R. (1979). Sub-band coder design incorporating quadrature filters and pitch prediction. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process*: 530–533.

Barrow, H. (1987). Learning receptive fields. In *Proc. IEEE 1st Ann. Conf. Neural Networks*, San Diego, volume 4: 115–121.

Barlow, H., Földiak, P. (1989). Adaptation and decorrelation in the cortex. In Durbin, R., Miall, C., and Mitchison, G., editors, *The computing neuron*, Addison Wesley: 54–72.

Barlow, H., Kaushal, T., Mitchison, G. (1989). Finding minimum entropy codes. *Neural Computation*, 1: 412–423.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M.G., Heinze, H.-J., Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature*, 400: 724-726.

Becker, S., Hinton, G.E. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355: 161-163.

Bell, A.J., Sejnowski, T.J. (1995). An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6): 1129-1159.

Ben-Artzi, E. Marks, L.E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Perception & Psychophysics*, 57(8): 1151-1162.

Berger, T. (1971). *Rate distortion theory: A mathematical basis for data compression*. Prentice-Hall.

Bertelson, P., Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review* 5: 482-489.

Bhaskaran, V., Konstantinides, K. (1995). *Image and video compression standards: Algorithms and architectures*. Kluwer Academic Publishers.

Bialek, W., Ruderman, D.L., Zee, A. (1991). Optimal sampling of natural images: a design principle for the visual system? In Touretzky, D., Moody, J. (eds.), *Advances in neural information processing systems 3*. Morgan Kaufmann: 363-369.

Blakemore, C., Campbell, F.W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology (London)*, 203: 237-260.

Blauert, J. (1983). *Spatial Hearing*. MIT Press.

Borst, A., Theunissen, F.E. (1999). Information theory and neural coding. *Nature neuroscience*, 2(11): 947-957.

Brooks, M. (1990). Highly redundant sensing in robotics - analogies from biology: distributed sensing and learning. In: Tou, J.T., Balchen, J.G. (eds.). *Highly redundant sensing in robotic systems*, NATO ASI Series, Vol. F 58, Springer.

Buracas, G.T., Zador, A.M., DeWeese, M.R., Albright, T.D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20: 959-969.

Burt, P.J., Adelson, E.H. (1983). The Laplacian Pyramid as a Compact Image Code. *IEEE Trans. Commun. COM*, 31: 532-540.

- Burton, G.J., Moorhead, I.R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, 26, 157-170.
- Calvert, G.A., Brammer, M.J., Iversen, S.D. (1998). Crossmodal identification. *Trends in Cognitive Sciences* 2(7): 247-253.
- Calvert, G.A., Spence, C., Stein, B. (eds.) (2004). *The Handbook of multisensory processes*. MIT Press.
- Campbell, F., Cooper, G., Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology*, 203: 223-235.
- Campbell, F., Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London)*, 197: 551-566.
- Cardoso, J.-F. (1997). Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4): 112-114.
- Carlson, C.R. (1978). Thresholds for perceived image sharpness. *Photographic Science and engineering*, 22: 69-71.
- Cherniak, C. (1992). Local optimization of neuron arbors. *Biological Cybernetics*, 66: 503-510.
- Cherniak, C. (1994). Component Placement Optimization in the Brain. *Journal of Neuroscience*, 14(4): 2418-2427.
- Choe, C.S., Welch, R.B., R.M. Gilford, R.M., J.F. Juola, J.F. (1975) The ventriloquist effect: visual dominance or response bias, *Perception & Psychophysics* 18: 55-60.
- Conner, D., Brainard, R., Limb, J. (1972). Intraframe coding for picture transmission. *Proc. IEEE*, 60: 779–791.
- Craig, A., Colquhoun, W.P., Corcoran, D.W.J. (1976). Combining evidence presented simultaneously to the eye and the ear: A comparison of some predictive models. *Perception & Psychophysics*, 19(6): 473-484.
- Craik, K.J.W. (1943). *The Nature of Explanation*. Cambridge University Press.
- Crochiere, R., Rabiner, L. (1983). *Multirate Digital Signal Processing*. Prentice-Hall.
- Crochiere, R., Webber, S., Flanagan, J. (1976). Digital coding of speech in subbands. *Bell System Technical Journal*, 55: 1069–1085. Daugman, J. (1988). Complete discrete 2D gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust. Speech Signal Process. ASSP*, 36: 1169-1179.

Dalton, P., Doolittle, N., Nagata, H., Breslin, P.A.S. (2000). The merging of the senses: integration of subthreshold taste and smell. *Nature Neuroscience*, 3(5): 431-432.

Davissou, L.D. (1972). Rate distortion theory and applications. *Proceedings of the IEEE*, 60: 800-808.

Dealy, R.S., Tolhurst, D.J. (1977). Is spatial frequency adaptation and aftereffect of prolonged inhibition? *Journal of Physiology (London)*, 241: 261-270.

Deriugin, N. (1956). The power spectrum and the correlation function of the television signal. *Telecommunications*, 1(7):1-12.

De Valois, K.K. (1977). Spatial frequency adaptation can enhance contrast sensitivity. *Vision Research*, 17: 1057-1065.

De Valois, R.L., Albrecht, D.G., Thorell, L.G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22: 545-559.

De Valois, R.L., De Valois, K.K., Ready, J., Blanckensee, H. (1975). Spatial frequency tuning of macaque striate cortex cells. *Assoc. Res., Vision Ophthalmol.*, 15: 16.

Downar, J., Crawley, A. P., Mikulis, D. J., Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, 3(3): 277-283.

Driver, J., Spence, C. (eds.) (2000). Multisensory perception: beyond modularity and convergence. *Current Biology* 10: R731-R735.

Durrant-Whyte, H.F. (1988). Integration, coordination and control of multi-sensor robot systems. Kluwer Academic Press.

Eijkman, E., Vendrik, A.J.H. (1965). Can a sensory system be specified by its internal noise? *The Journal of the acoustical society of America*, 37(6): 1102-1109.

Ernst, M.O., Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature*, 415: 429-33.

Felleman, D.J., van Essen, D.C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex* 1:1-47.

Field, D.J. (1987). Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells. *Journal of the Optical Society of America A*, 4: 2379-2394.

Field, D.J. (1993). Scale-invariance and self-similar wavelet transforms: an analysis of natural scenes and mammalian visual systems. In Farge, M., Hunt, J.C.R., Vassilicos, J.C (eds.). *Wavelets, fractals, and Fourier transforms*. Clarendon Press: 151-193.

Field, D.J. (1994). What is the goal of sensory coding? *Neural Computation*, 6: 559-601.

Fischer, T.R. (1992). On the rate-distortion efficiency of subband coding. *IEEE Transactions on Information Theory*, 38: 426-428.

Foster, K.H., Gaska, J.P., Pollen, D.A. (1983). Spatial and temporal frequency selectivity of V1 neurons in the macaque monkey. *Invest. Ophthalmol. Visual Sci. (Suppl.)*, 22: 228.

Frassinetti F, Bolognini N, Ladavas E (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research* 147: 332-343.

Garvey, T.D. (1987). A survey of AI approaches to the integration of information. *SPIE Infrared Sensors and Sensor Fusion*, 782: 68-82.

Gaus, W. (1983). *Dokumentations- und Ordnungslehre*. Springer, 1983.

de Gelder, B., Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, 14 (3): 289-311.

Georgeson, M.A. (1980). The perceived spatial frequency, contrast and orientation of illusory gratings. *Perception*, 9: 695-712.

Gibson, J.J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin Co.

Giray, M., Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology - Human Perception and Performance*, 19(6): 1278-91.

Graham, N. (1989). *Visual Pattern Analyzers*. Oxford Univ. Press.

Graham, N., Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channel models. *Vision Research*, 11: 251-259.

Graham, N., Robson, J.G., Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research*, 18: 815-825.

Gröschel, M., Lügger, J. (1995). *Wissenschaftliche Kommunikation am Wendepunkt. Bibliotheken im Zeitalter globaler elektronischer Netze*. *ZfBB*, 42 (3): 287-301.



Habibi, A., Wintz, P. A. (1971). Image coding by linear transformation and block quantization. *IEEE Trans. Communications*, COM-19: 50–63.

Harrison, C. (1952). Experiments with linear prediction in television. *Bell System Technical Journal*, 31: 764–863.

Hartline, H.K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, 121: 400-415.

van Hateren, J.H. (1992). A theory of maximizing sensory information. *Biological Cybernetics*, 68: 23-29.

Hauske, G. (1988). The visual system as a spatial frequency matched filter. In: Marko, H., Hauske, G., Struppler, A. (eds.): *Processing structures for perception and action*. VCH Verlagsgesellschaft.

Hauske, G. (1994). *Systemtheorie der visuellen Wahrnehmung*. B.G.Teubner.

Hinton, G., McClelland, J., Rumelhart, D. (1986). Distributed representations. In Rumelhart, D., McClelland, J., and PDP research group (eds.), *Parallel Distributed Processing*, MIT Press, Chapter 3: 77–109.

Hofbauer, M., Röhrbein, F., Schill, K., Zetzsche, C. (2001). Visuell-auditive Kovariation und Interaktion. In Bühlhoff, H. H., Gegenfurtner, K. R., Mallot, H. A., Ulrich, R., (eds.), *Beiträge zur 4. Tübinger Wahrnehmungskonferenz*, Knirsch Verlag: 162.

Huang, J., Turcotte, D. (1990). Fractal image analysis: application to the topography of Oregon and synthetic images. *Journal of the Optical Society of America A*, 7: 1124–1130.

Hubel, D.H., Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148: 574–591.

Jayant, N.S., Noll, P. (1984). *Digital coding of waveforms*. Prentice Hall.

Kandel, E.R., Schwartz, J.H., Jessel, T.M. (eds.) (1995). *Neurowissenschaften*. Spektrum Akademischer Verlag.

King, A.J., Cavert, G.A. (2001) Multisensory integration: perceptual grouping by eye and ear. *Current Biology* 11: R322-R325.

Kuffler, S.W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Physiology*, 16: 37-68.

Kulikowski, J.J., Bishop, P.O. (1981). Linear analysis of the responses of simple cells in the cat visual cortex. *Experimental Brain Research*, 44: 386-400.

Laughlin, S.B. (1981). A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung*, 36: 910-912.

Laughlin, S.B. (1992). Retinal information capacity and the function of the pupil. *Ophthalmic and Physiological Optics*, 12: 161-164.

Laughlin, S.B., de Ruyter van Steveninck, R.R., Anderson, J.C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, Vol 1(1): 36-41.

Legge, G.E., Foley, J.M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, 70: 1458-1471.

Lewicki, M.S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, 5(4): 356-363.

Lewis, J.W., Beauchamp, M.S., DeYoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, 10: 873-888.

Linsker, R. (1986). From basic network principles to neural architecture: Emergence of orientation selective cells. *Proceedings of the National Academy of Science USA*, 83: 8390–8394.

Linsker, R. (1994). Sensory processing and information theory. In Grassberger, P., Naadal, J.-P., (eds.). *From statistical physics to statistical inference and back*. Kluwer: 237-247.

Lohscheller, H. (1984). Subjectively adapted image communication system. *IEEE Trans. Commun.*, 32: 1316-1322.

Maffei, L., Fiorentini, A. (1973). The visual cortex as a spatial frequency analyzer. *Vision Research*, 13: 1255–1267.

Mallat, S. (1989). A Theory of Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Trans. PAMI*, 11: 647-693.

Marko, H. (1982). *Kybernetik: Informations- und Kommunikationstheorie*. In: Hoppe, W., Lohmann, W., Markl, H., Ziegler, H., *Biophysik*, Springer.

Massaro, D.W. (1998). Speechreading: Illusion or window into pattern recognition? *Trends in Cognitive Sciences*, 3: 310-317.

Massaro, D.W. (1999). Perceiving talking faces: from speech perception to a behavioral principle. MIT Press.

McDonald J.J., Teder-Sälejärvi W.A., Hillyard S.A. (2000) Involuntary orienting to sound improves visual perception. *Nature*, 407: 906-908.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264: 746-748.

Meese T., Andersen S.J. (2002) Spiral mechanisms are required to account for summation of complex motion components. *Vision Research* 42: 1073-1080.

Mel, B.W. (1994). Information Processing in dendritic trees. *Neural Computation*, 6: 1031-1085.

Menzel, E. (1959). Der Gesichtssinn als linearer Übertragungs-kanal und die Machschen Streifen. *Naturwissenschaften*, 46: 316-317.

Meredith, M.A., Nemitz, J.W., Stein, B.E. (1987) Determinants of multisensory integration in superior colliculus neurones. I. Temporal factors. *Journal of Neuroscience* 10: 3215-3229.

Meredith, M.A., Stein, B.E. (1996). Spatial determinants of multisensory integration in cat superior colliculus. *Journal of Neurophysiology* 75: 1843-1857.

Meyer, G., Wuerger, S. (2001). Cross-modal integration of auditory and visual motion signals. *NeuroReport* 12(11): 2557-2560.

Middelhoek, S., Hoogerwerf, A.C. (1986). Classifying solid-state sensors: the 'sensor effect cube'. *Sensors and Actuators*, 10: 1-8.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Experimental Psychology, Human Perception and Performance* 17(1): 160-169.

Movshon, J.A., Thompson, I.D., Tolhurst, D.J. (1978). Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology (London)*, 283: 101-120.

Mullen, K.T., Sankeralli, M.J. (1998) Evidence for the stochastic independence of the blue-yellow, red-green and luminance detection mechanisms revealed by subthreshold summation,. *Vision Research* 39: 733-745.

Mulligan, R.M., Shaw, M.L. (1980). Multimodal signal detection: independent decisions vs. integration. *Perception & Psychophysics*, 28(5): 471-478.

Nelson, G., Pfeifer, L., Wood, R. (1972). High speed octave band digital filtering. *IEEE Transactions on Audio Electroacoustics*, AU-20(1): 8–65.

Nickerson, S. (1973) Intersensory facilitation of reaction time: energy summation or preparation enhancement? *Psychological Review* 80: 489-509.

Olshausen, B.A., Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381: 607-609.

O'Neal, Jr., J.B. (1966). Predictive quantizing systems (dpcm) for the transmission of television signals. *Bell System Technical Journal*, 45: 689–719.

Osborn, W.C., Sheldon, R.W., Baker, R.A. (1963). Vigilance Performance under conditions of redundant and nonredundant signal presentation. *Journal of Applied Psychology*, Vol. 47(2): 130-134.

Pantle, A., Sekuler, R. (1968). Size detecting mechanisms in human vision. *Science*, 162: 1146-1148.

Pau, L.F. (1988). Sensor Data Fusion. *Journal of Intelligent and Robotic Systems*, 1: 103-116.

Peitgen, H.O., Jürgens, H., Saupe, D. (1992). *Chaos and fractals: new frontiers of science*. Springer.

Perkins, M.G., Lookabaugh, T. (1989). A psychophysically justified bit allocation algorithm for subband image coding systems. *ICASSP, IEEE*, 2833: 1815-1818.

Quick R.F. (1974). A vector magnitude model of contrast detection. *Kybernetik* 16: 65-67.

Raij, T., Uutela, K., Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron* 28: 617-625.

Rao, R.P.N., Ballard, D.H. (1998). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience*, 2: 79-87.

Ramchandran, K., Vetterli, M. (1993). Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. Image Process.*, 2: 160-175.

Reinagel, P. (2001). How do visual neurons respond in the real world? *Current Opinion in Neurobiology*, 11: 437-442.

Röhrbein, F., Hofbauer, M., Meyer, G., Wuerger, S., Schill, K., Zetsche, C. (2003b). The role of co-localization for auditory-visual motion signals. In Bülthoff, H. H., Gegenfurtner, K. R., Mallot, H. A., Ulrich, R. and Wichmann, F. A. (eds.), Beiträge zur 6. Tübinger Wahrnehmungskonferenz, Knirsch Verlag:105.

Röhrbein, F., Schill, K., Zetsche, C. (2000). Intermodal sensory interactions for ecologically valid intensity changes as caused by moving observers or moving objects. In Bülthoff, H. H., Fahle, M., Gegenfurtner, K. R. and Mallot, H. A. (eds.), Beiträge zur 3. Tübinger Wahrnehmungskonferenz, Knirsch Verlag: 62.

Röhrbein, F., Zetsche, C. (1998a). Self-similar bandpass channels and the  $1/f^2$  power spectrum of the natural environment: a rate distortion analysis. *Perception*, 27: 153.

Röhrbein, F., Zetsche, C. (1998b). Exploitation of natural image statistics by biological vision systems:  $1/f^2$  power spectra and self-similar bandpass decompositions. In Proc. CVPR-98, IEEE Computer Society, Los Alamitos, CA: 86–91.

Röhrbein, F., Zetsche, C. (2000). Auditory-visual interactions and the covariance structure generated by relative movements in natural environments. In Guidati, G., Hunt, H., Heiss, A. (eds.), Proceedings of the 7. International Congress on Sound and Vibration: 2427–2434.

Röhrbein, F., Zetsche, C. (2002). The statistics of natural scenes and Weber's law. In Würtz, P., Lappe, M. (eds.), Dynamic Perception, Akademische Verlagsgesellschaft: 233-238.

Röhrbein, F., Zetsche, C., Meyer, G. (2003a). Auditory-visual motion signals and the statistical redundancy of the natural environment. Proceedings IMRF: 15.

Ruderman, D.L. (1994a). Designing receptive fields for highest fidelity. *Network*, 5: 147-155.

Ruderman, D.L. (1994b). The statistics of natural images. *Network*, 5: 517–548.

Ruderman, D.L., Bialek, W. (1994). Statistics of natural images: scaling in the woods. In Cowan, J.D., Tesauro, G., Alspector, J. (eds.). Advances in neural information processing systems 6. Morgan Kaufmann.

Sachs, M.B., Nachmias, J., Robson, J.G. (1971). Spatial frequency channels in human vision. *Journal of the Optical Society of America*, 61: 1176-1186.

Sanger, T.D. (1989). Optimal unsupervised learning in a single-layer feedforward neural network. *Neural Networks*, 2: 459–473.

van der Schaaf, A., van Hateren, J.H. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research*, 28: 2759-2770.

Schade, O.H. (1956). Optical and Photoelectric Analog of the Eye. *Journal of the Optical Society of America*, 46: 721-739.

Schiller, P.H. (1992). The ON and OFF channels of the visual system. *Trends in Neurosciences* 15(3): 86-92.

Schiller, P.H., Finlay, B.L., Volman, S.F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequency. *Journal of Neurophysiology*, 39: 1334-1351.

Schreiber, W.F. (1956). The measurement of third order probability distributions of television signals. *IRE Trans. Information Theory*, IT-2(3): 94–105.

Schwartz, O., Simoncelli, E.P. (2001). Natural sound statistics and divisive normalization in the auditory system. *Advances in Neural Information Processing Systems*, MIT Press, 13: 166-172.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27: 397-423, 623-656.

Shaw, M.L. (1982). Attending to multiple sources of information. *Cognitive Psychology*, 14: 353-409.

Sherrington, C.S. (1906). *The integrative action of the nervous system*. Scribner's Sons.

Singh, N.C., Theunissen, F.E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *Journal of the Acoustic Society of America*, Vol 114(6): 3394-3411.

Snippe, H. (1996). Parameter extraction from population codes: a critical assessment. *Neural Computation*, 8: 511–529.

Soto-Faraco, S., Kingstone, A. (2004). Multisensory integration of dynamic information. In: Calvert, G.A., Spence, C., Stein, B. (eds.). *The Handbook of multisensory processes*. MIT Press.

Soto-Faraco, S., Kingstone, A., Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia* 41: 1847-1862.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research* 14: 139-146.

Soto-Faraco, S., Spence, C., Kingstone, A. (2004). Moving multisensory research along: Motion perception across sensory modalities. *Current Directions in Psychological Science*, 13: 29-32.

Spence, C., Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*: 1000-1030.

Spence, C., Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59: 1-22.

Spence, C., Driver, J. (2004). *Crossmodal space and crossmodal attention*. Oxford University Press.

Spinelli, D.N., Starr, A., Barrett, T.W. (1968). Auditory specificity in unit recordings from cat's visual cortex. *Experimental Neurology*, 22: 75-84.

Srinivasan, M.V., Laughlin, S.B., Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B*, 216: 427-459.

Stein, B.E., Meredith, M.A. (1993). *The merging of the senses*. MIT Press.

Stromeyer, C.F., Julesz, B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America*, 62: 1221-1232.

Thomas, J.P. (1970). Model of the function of receptive fields in human vision. *Psychological Review*, 77: 121-134.

Thomopoulos, S.C.A. (1990). Sensor integration and data fusion. *Journal of Robotic Systems* 7(3): 337-372.

Tolhurst, D.J., Tadmor, Y., Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, 12: 229-232.

Tolhurst, D.J., Thompson, I.D. (1981). On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proc. Royal Society of London B*, 213: 183-199.

Townsend, J.T., Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, 39(4): 321-359.

Treutwein, B. (1997). YAAP: Yet another adaptive procedure. *Spatial Vision*, 11: 129-134.

Tyler, C.W., Chen, C.C. (2000). Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and probability summation. *Vision Research*, 40: 3121-3144.

Ukrainec, A., Haykin, S. (1996). A modular neural network for enhancement of cross-polar radar targets. *Neural Networks*, 9, 143-168.

Vaidyanathan, P.P. (1993). *Multirate systems and filterbanks*. Prentice-Hall.

Wagner, H., Kautz, D., Poganiatz, I. (1997). Principles of acoustic motion detection in animals and man. *Trends in Neurosciences*, 20(12): 583-588.

Wallace, M.T., Meredith, M.A., Stein, B.E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research* 91(3): 484-488.

Wallace, M.T., Meredith, M.A., Stein, B.E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology* 80: 1006-1010.

Wallace, M.T., Stein, B.E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience* 17: 2429-2444.

Watson, A.B. (1982). Summation of grating patches indicates many types of detectors at one retinal location, *Vision Research*, 22: 17-25.

Watson, A.B., Pelli, D.G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33 (2): 113-120.

Webster, M.A., Mollon, J.D. (1991) Changes in colour appearance following post-receptoral adaptation. *Nature* 349: 235–238.

Welch, R.B, Warren, D.H. (1986). Intersensory interactions. In: *Handbook of Perception and Human Performance*, Vol.1: Sensory Processes and Perception, Boff, K.R., Kaufmann, L., Thomas, J.P. (eds.), Wiley, NY, Ch. 25: 1-36.

Wilson, H.R., Bergen, J.R. (1979). A four mechanism model for threshold spatial vision. *Vision Research* 19: 19-32.

Wilson, H.R., Gelb, D.J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America A*, 1: 124-131.



Wilson, H.R., McFarlane, D.K., Phillips, G.C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23: 873-882.

Woods, J.W., O'Neil, S.D. (1986). Subband Coding of Images. *IEEE Transactions on Acoustics, Speech and Signal Processing ASSP*, 43: 1278-1288.

Wuerger, S.M., Hofbauer, M., Meyer, G.F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics* 65: 1188-1196.

Yang, Z., Purves, D. (2003). A statistical explanation of visual space. *Nature Neuroscience* 6: 632-640.

Younberg, J., Boll, S. (1978). Constant-q signal analysis and synthesis. In *Proc. IEEE Int. Conference on Acoustics, Speech, Signal Processing*: 375–378.

Zaidi, Q., Shapiro, A.G. (1993). Adaptive orthogonalization of opponent-color signals. *Biological Cybernetics* 69: 415–428.

Zetsche, C. (1990). Sparse coding: the link between low level vision and associative memory. In Eckmiller, R., Hartmann, G., Hauske, G., (eds.), *Parallel Processing in Neural Systems and Computers*: 273–276.

Zetsche, C., Barth, E., Wegmann, B. (1993). The importance of intrinsically two-dimensional image features in biological vision and picture coding. In Watson, A., (ed.) *Digital images and human vision*, MIT Press: 109-138.

Zetsche, C., Hauske, G. (1989a). Multiple channel model for the prediction of subjective image quality. In Rogowitz, B.E., (ed.), *Human Vision, Visual Processing and Visual Display*, Vol. 1077 *Proceedings SPIE*: 209–216.

Zetsche, C., Hauske, G. (1989b). Principal features of human vision in the context of image quality models. In *Proc. IEE 3rd Int. Conf. Image Proc. and its Applications*: 102–106.

Zetsche, C., Krieger, G. (2001). Nonlinear mechanisms and higher-order statistics in biological vision and electronic image processing: review and perspectives. *Journal of Electronic Imaging*, 10: 56-99.

Zetsche, C., Röhrbein, F. (2000). Cortical complex cells and the statistical structure of the natural environment. In Baratoff, G., Neumann, H. (eds.), *Dynamische Perzeption, Proceedings in Artificial Intelligence*, 9: 57–62.

Zetsche, C., Röhrbein, F. (2001). Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network: Computation in Neural Systems*, 12: 331-350.

*Ich danke Herrn Prof. Hauske herzlich für die vielen freundlichen Hinweise und Anregungen, die meiner Arbeit schließlich zu einem erfolgreichen Abschluss verholfen haben.*

*Mein sehr spezieller Dank gilt meinen unterschiedlichen Familien in München, Bremen und Berlin. Sie haben mich immer unterstützt: wissenschaftlich und materiell, freundschaftlich und intellektuell.*