

Konzeptionelle Modellierung
geometrischer Invarianzen in der
visuellen Wahrnehmung von Primaten

Situativ gesteuerte Complex-Bildung als
Grundlage invarianter Zellantworten

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem

Fachbereich Physik
der Philipps-Universität Marburg

vorgelegt von

Rüdiger Kupper
aus Weilburg a.d. Lahn

Marburg, 22. November 2004

Konzeptionelle Modellierung
geometrischer Invarianzen in der
visuellen Wahrnehmung von Primaten

Situativ gesteuerte Complex-Bildung als
Grundlage invarianter Zellantworten

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem

Fachbereich Physik
der Philipps-Universität Marburg

vorgelegt von

Rüdiger Kupper
aus Weilburg a.d. Lahn

Marburg, 22. November 2004

Vom Fachbereich Physik der Philipps-Universität als Dissertation
angenommen am 24. September 2004.

Erstgutachter: Prof. Dr. Reinhard Eckhorn

Zweitgutachter: Prof. Dr. Edgar Körner

Tag der mündlichen Prüfung: 12. November 2004

Aber was nun geschah, war so erstaunlich, daß Jim Mund und Nase aufsperrte und Lukas an seiner Pfeife zu ziehen vergaß. Der Riese kam Schritt für Schritt näher und bei jedem Schritt wurde er ein Stückchen kleiner. Als er etwa noch hundert Meter entfernt war, schien er nicht mehr viel größer zu sein als ein hoher Kirchturm. Nach weiteren fünfzig Metern hatte er nur noch die Höhe eines Hauses. Und als er schließlich bei Emma anlangte, war er genauso groß wie Lukas der Lokomotivführer. Er war sogar fast einen halben Kopf kleiner.

*»Sie meinen«, fragte Lukas, »Sie werden gar nicht wirklich kleiner, wenn Sie näherkommen? Und Sie sind auch nicht wirklich so riesengroß, wenn Sie weit entfernt sind, sondern es sieht nur so aus?«
»Sehr richtig«, antwortete Herr Tur Tur.
»Deshalb sage ich, ich bin ein Scheinriese.«*

MICHAEL ENDE
Jim Knopf und Lukas der
Lokomotivführer

Inhaltsverzeichnis

Zusammenfassung	i
Kapitel-Übersicht	iii
1 Invarianten der visuellen Wahrnehmung	1
1.1 Die visuelle Wahrnehmung unserer Umwelt	1
1.2 Invariante Erkennung von Sehobjekten	3
1.3 Geometrische Invarianzen	5
2 Klassische Modellierung geometrischer Invarianzen	9
2.1 Statische hierarchische Merkmalsgruppierung	10
2.1.1 Complex-Bildung als Konzept	10
2.1.2 Binde-Problem	13
2.1.3 Überlappende Repräsentantenmengen	15
2.2 Dynamisches Routing	16
3 Situative Steuerung als Invarianz-Strategie	19
3.1 Individuum und Umwelt	19
3.2 Neurophysiologische Grundlagen	22
3.3 Methoden der Modellierung	24
3.3.1 Konzeptionelle Modellierung	24
3.3.2 Funktionelles Neuron	24
3.3.3 Präzisierung des Begriffes „rezeptives Feld“ . .	26
3.4 Situativ gesteuerte Anpassung aktivierender Felder . .	28
3.5 Der externe Steuerparameter	31
4 Der Schwelleneffekt als Beispiel systematischer Feldanpassung	35
4.1 Übertragungsfunktion	36
4.2 Symmetrisches Stimulationsfeld	37
4.3 Raumfrequenzabstimmung mit Hilfe des Schwelleneffektes	42

4.4	Schwelleneffekt und Größeninvarianz	45
4.5	Unsymmetrisches Stimulationsfeld	46
4.6	Diskussion	49
5	Ein konzeptionelles Modell zur retinalen Schlupfkorrektur	51
5.1	Einleitung	51
5.2	Problemstellung: Neuronale Schlupfkorrektur	54
5.3	Konzeptuelle Grundlagen	56
5.3.1	Präsynaptisches Gating	56
5.3.2	Sensitivitätsprofile	58
5.4	Modell	59
5.5	Simulation	60
5.5.1	Methoden	62
5.6	Ergebnisse	63
5.6.1	Einzelfelder	64
5.6.2	Ausgangsfelder	65
5.7	Diskussion	65
5.7.1	Einzelfelder	66
5.7.2	Ausgangsfelder	67
5.7.3	Neuronaler Aufwand	67
5.7.4	Mikrosakkadische Suppression	68
5.7.5	Gewinnung des Blickrichtungssignals	71
5.7.6	Extraktion aus dem visuellen Reiz	73
5.8	Situativ gesteuerte Complex-Bildung	74
6	Ein konzeptionelles Modell zur Entfernungsinvarianz	77
6.1	Einleitung	77
6.1.1	Der Begriff „Größe“	79
6.2	Größen- gegen Entfernungsinvarianz	79
6.3	Entfernungsschätzung im visuellen System	81
6.3.1	Psychophysische Befunde	82
6.3.2	Physiologische Befunde	83
6.3.3	Interpretation	85
6.4	Problemstellung: Entfernungsinvarianz	87
6.5	Modell	89
6.5.1	Entfernungscharakteristik	89
6.5.2	Entfernungs-Complex-Zellen	93
6.5.3	Invariante Objekterkennung	96
6.5.4	Labelled-Line-Code	96
6.6	Simulation	98
6.6.1	Echtweltumgebung	98
6.6.2	Netzwerk	99
6.6.3	Invarianzindex	100

6.7	Ergebnisse und Invarianzleistung	101
6.7.1	Gleichförmige Entfernungseharakteristik	101
6.7.2	Invarianzleistung	104
6.7.3	Angepaßte Entfernungseharakteristik	104
6.7.4	Deaktivierte Entfernungsemodulation	106
6.8	Diskussion	106
6.8.1	Invarianzleistung der Simulation	107
6.8.2	Sehentfernung als Steuerparameter	110
6.9	Experimenteller Nachweis möglicher Entfernungse- Complex-Zellen	119
6.9.1	Feldeigenschaften der Entfernungse-Complex- Zellen	119
6.9.2	Feldeigenschaften der Neuronen in V4A	121
6.9.3	Interpretation	121
6.9.4	Schlußbemerkung	124
7	Zusammenfassende Diskussion	125
7.1	Verlauf der Argumentation	126
7.2	Beurteilung	127
	Zum Phänomen des Scheinriesen	131
	Literaturverzeichnis	133
	Danksagung	143
	Wissenschaftlicher Werdegang	145

Zusammenfassung

Unser Sehsinn vermittelt uns eine stabile Wahrnehmung der Umwelt. Objekte darin erkennen wir unabhängig von der Position, die wir ihnen gegenüber einnehmen. Diese invariante Wahrnehmung ist im Rahmen der verfügbaren neuronalen Modelle nur mit Einschränkungen zu erklären. Die Standardmodelle basieren auf einer hierarchischen Anordnung von Nervenzellen, deren Ziel es ist, spezifische neuronale Antworten für komplexe visuelle Reize aus Antworten auf einfache Reizkomponenten zu konstruieren. Ein wesentliches Konzept ist dabei die neuronale ODER-Bildung (*Complex-Bildung*) durch konvergente Verschaltung. Die Generierung von Invarianz für bestimmte Reizvariationen läuft hierbei der Formierung reizspezifischer Antworten entgegen – auf Ebene des Signalflusses im Netzwerk, wie auch als Denkmodell. Die klassischen Modelle zur invarianten visuellen Formerkennung weisen daher Schwächen auf, etwa das Binde-Problem oder die Ununterscheidbarkeit von Objekten mit überlappenden Repräsentantenmengen. Die vorliegende Arbeit nähert sich dieser Problematik vom Blickwinkel der konzeptionellen Modellierung.

Ein lebendiges Individuum erfährt seine Umwelt aktiv: Äußere physikalisch-körperliche Umstände beeinflussen die Verarbeitung im visuellen System. Ich formuliere hier das Konzept der *situativ gesteuerten Complex-Bildung*, das auf einer Steuerung der Übertragungseigenschaften einzelner Neuronen durch externe Parameter beruht. Seine Leistungsfähigkeit demonstriere ich in zwei Modellen zur invarianten visuellen Verarbeitung, der neuronalen retinalen Schlupfkorrektur und der entfernungsinvarianten Objektrepräsentation. Die Modelle überwinden entscheidende Probleme der klassischen Modellierung, erfordern jedoch einen erhöhten neuronalen Aufwand. Im Falle des Entfernungsinvarianzmodells führt der Einsatz der situativ gesteuerten *Complex-Bildung* zur Vorhersage einer neuartigen Zellklasse, den *Entfernungs-Complex-Zellen*. Neuronen mit teilweise ähnlichen Codierungseigenschaften wurden in jüngster Zeit experimentell nachgewiesen.

In beiden Modellen wird durch die situativ gesteuerte *Complex-Bildung* eine Szenenrepräsentation generiert, die vom verwendeten Steuerparameter unabhängig ist. Es ist zu erwarten, daß auf gleiche Weise Invarianz auch gegenüber anderen äußeren Bedingungen erzeugt werden kann. Die situativ gesteuerte *Complex-Bildung* erweist sich so als universell einsetzbares Werkzeug zur konzeptionellen Modellierung neuronaler Invarianzen. Damit liefert sie auch ein effektives Denkmodell für das weitere Verständnis kortikaler Verarbeitung.

Kapitel-Übersicht

Ziel

Ich skizziere hier das Manuskript der vorliegenden Arbeit. Zu diesem Zweck stelle ich die Kurzzusammenfassungen der individuellen Kapitel im Text zusammen.

Kern:

Die Modellierung invarianter neuronaler Reizverarbeitung kann auf zwei Weisen erfolgen: Über die passive *Integration* der varianten Erscheinungsformen eines Reizes, oder die aktive *Steuerung* ihrer Verarbeitung. Ich argumentiere für eine Steuerung durch Umweltparameter.

Kapitel 1 Dieses Kapitel dient der Hinleitung zum Arbeitsthema. Ich erläutere die Bedeutung invarianter Objektwahrnehmung auf kognitiver und zellphysiologischer Ebene. Ihre neuronale Implementation ist ungeklärt. Als Teilproblem identifiziere ich die *geometrischen Varianzen*, die von der optischen Abbildung im Auge herrühren. Die Modellierung möglicher neuronaler Implementationen ihrer invarianter Verarbeitung ist Gegenstand dieser Arbeit.

Kapitel 2 Ich stelle die wichtigsten existierenden Ansätze zur Modellierung geometrischer Invarianzen vor: hierarchische Gruppierung und dynamisches *Routing*. Beide Ansätze arbeiten auf hierarchisch strukturierten Netzwerken. Ich formuliere das verallgemeinerte Konzept der *Complex*-Bildung. Ich erläutere das Bindeproblem und das Problem überlappender Repräsentantenmengen, die in Modellen dieser Art wesentliche Schwierigkeiten bereiten.

Kapitel 3 Als alternative Sichtweise auf die Erzeugung geometrischer Invarianzen schlage ich die situative Steuerung neuronaler Codierungseigenschaften vor. Diese beruht darauf, daß Informationen

Kapitel-Übersicht

von außerhalb des Sehpfadens über die körperliche Situation des Individuums in seiner Umwelt zum Zwecke der Invarianzgenerierung ausgewertet werden. Ich führe Belege für den Einfluß der Sehsituation auf die Verarbeitung im visuellen System an und bereite dieses Prinzip zum Zwecke der konzeptionellen Modellierung auf. Ich schlage Methoden vor, die rezeptiven Felder einzelner Neuronen in Abhängigkeit äußerer Einflüsse zu verändern und motiviere meine Annahme, daß diese Einflüsse in Form einzelner Zustandsgrößen in die neuronale Verarbeitung eingehen.

Kapitel 4 Die Beeinflussung der Codierungseigenschaften eines einzelnen Neurons durch einen von außen angreifenden Prozeß demonstriere ich hier beispielhaft am Schwelleneffekt. Durch Variation von Antwortschwelle und Verstärkung sind neuronale Aktivierungsgebiete größenskalierbar. Mit Hilfe dieses Prinzips läßt sich ein Spektrum hochinteressanter Effekte erzielen, darunter die Auflösungskontrolle und die Durchstimmung von Raumfrequenzkanälen im visuellen System. Zur Generierung visueller Größeninvarianz erweist sich der Schwelleneffekt als ungeeignet.

Kapitel 5 In diesem Kapitel entwickle ich ein Modell zur neuronalen Kompensation der durch Mikroaugenbewegungen verursachten retinalen Bildverschiebungen. Es beruht auf der externen Steuerung lokaler neuronaler Codierungseigenschaften. Ich definiere den Begriff der *Sensitivitätsprofile* als Erweiterung dieses Prinzips auf verschiedene Zellgruppen. Darauf basierend führe ich das Konzept der *situativ gesteuerten Complex-Bildung* ein. Ich demonstriere die Leistungsfähigkeit des Modells in einer Computersimulation mit biologisch relevanten Randbedingungen. Das Modell erfordert einen erhöhten neuronalen Aufwand, sowie die Verfügbarkeit eines Blickrichtungssignals. Ich diskutiere mögliche Folgeeffekte wie eine mikrosakkadische Suppression.

Kapitel 6 In diesem Kapitel entwickle ich ein Modell zur entfernungsinvarianten Szenenrepräsentation. Es beruht auf der externen Steuerung in Abhängigkeit der Sehentfernung, deren Einfluß auf neuronale Antworten im Sehsystem experimentell belegt ist. Die Verwendung des zuvor eingeführten Konzepts der situativ gesteuerten *Complex-Bildung* führt mich zur Hypothesierung einer neuartigen Zellklasse, den *Entfernungs-Complex-Zellen*. Ich demonstriere die Leistungsfähigkeit des Modells in einer Computersimulation mit biologisch relevanten Randbedingungen, und für verschiedene Modellan-

nahmen. In dem Modell ist eine entfernungsinvariante Erkennung, sowie eine Schätzung der physikalischen Größe betrachteter Objekte möglich. Es löst das Problem der überlappenden Repräsentantemengen, und erlaubt das schnelle Erlernen neuer Objekte. Das Modell erfordert einen erhöhten neuronalen Aufwand, sowie die Verfügbarkeit einer Entfernungsschätzung. Ich diskutiere den experimentellen Nachweis möglicher Entfernungs-*Complex*-Zellen in einer aktuellen Forschungsarbeit.

Kapitel 7 Hier vollziehe ich die Argumentationslinie der vorausgegangenen Kapitel nach, und diskutiere deren wesentliche Ergebnisse für die Bildung geometrischer Invarianzen der Wahrnehmung. Ich identifiziere das Konzept der externen Steuerung neuronaler Codierungseigenschaften und dessen Anwendung in der situativ gesteuerten *Complex*-Bildung als universell einsetzbares Konzept zur Generierung von Invarianzen. Ich fasse die mögliche Bedeutung dieser Modellansätze, sowie deren Vor- und Nachteile zusammen. Die Methode der konzeptionellen Modellierung erweist sich als effektives Werkzeug zum Verständnis kortikaler Verarbeitung.

Kapitel 1

Invarianten der visuellen Wahrnehmung

Ziel

Dieses Kapitel dient der Hinleitung zum Arbeitsthema. Ich erläutere die Bedeutung invarianter Objektwahrnehmung auf kognitiver und zellphysiologischer Ebene. Ihre neuronale Implementation ist ungeklärt. Als Teilproblem identifiziere ich die *geometrischen Varianzen*, die von der optischen Abbildung im Auge herrühren. Die Modellierung möglicher neuronaler Implementationen ihrer invarianten Verarbeitung ist Gegenstand dieser Arbeit.

1.1 Die visuelle Wahrnehmung unserer Umwelt

Eine wesentliche Leistung des menschlichen Sehsystems besteht darin, uns interpretierbare Eindrücke der Außenwelt zu vermitteln. Die Wahrnehmung unserer Umgebung als eine stabile, äußere Welt ermöglicht es, uns zu orientieren, uns selbst in unserer Umwelt zu verorten, und geplante, zielgerichtete Aktionen darin auszuführen. Die Augen, unsere optischen Sensoren, stellen dabei die Schnittstelle zwischen Außenwelt und zentralem Nervensystem dar. Dort werden die äußeren Reize von einem Geflecht aus Nervenzellen, den Neuronen, verarbeitet. Wir glauben heute, daß diese Verarbeitung im Sinne einer Interpretation und *Repräsentation* geschieht, also in Form einer inneren Darstellung der außen vorgefundenen Welt.

Die Art und Weise, in der diese neuronale Repräsentation der Umwelt aufgebaut wird, ist Gegenstand breiter und intensiver For-

Kapitel 1 Invarianten der visuellen Wahrnehmung

schung in Neuro- und Psychophysik, Neurophysiologie und Neurobiologie. Im Zentrum des Interesses steht dabei naturgemäß der Mensch. Da die visuellen Systeme der Primaten untereinander sehr große Ähnlichkeiten aufweisen, bringen sowohl psychophysische Untersuchungen an menschlichen Probanden, wie auch experimentelle Arbeiten mit Affen wichtige Erkenntnisse auf diesem Gebiet.

Stabile Umwelt

Introspektiv erscheint einem Betrachter die „innere Darstellung“ seiner Umgebung ausgesprochen unmittelbar. So sehr sogar, daß wir uns des inneren Vorgangs unserer eigenen Wahrnehmung üblicherweise gar nicht bewußt werden. Wir betrachten uns selbst als außen, in der Welt befindlich. Ungeachtet der dynamischen Verarbeitung der visuellen Reize in unserem Nervensystem erscheint es uns selbstverständlich, unsere Umgebung als stabil und dauerhaft – eben als objektiv vorhanden – zu empfinden.

Invariante Objekteigenschaften

Diese Empfindung einer stabilen Umwelt spiegelt jedoch *keineswegs* den tatsächlich vorhandenen optischen Reiz, nämlich das von der Augenlinse erzeugte Abbild auf unserer Netzhaut wider. Eine Vielzahl äußerer Umstände hat Einfluß auf das Zustandekommen dieses Reizes. Hierzu gehören Sichtverhältnisse wie die Gesamthelligkeit oder verschiedenartige Beleuchtungen, ja sogar die Güte unseres optischen Abbildungsapparates selbst. Wie selbstverständlich empfinden wir das Erscheinen der optischen Reize unter veränderten Bedingungen jedoch nicht als geänderte Eigenschaften der Gegenstände selbst. Oft sind wir sogar außerstande, die Reizvariationen überhaupt wahrzunehmen: Ein weißes Blatt Papier unter einer gelblichen Schreibtischlampe erscheint uns gerade so weiß wie im bläulichen Tageslicht.¹ Richten wir unseren Blick nach rechts, so verschiebt sich das gesamte Bild auf unserer Netzhaut nach links, doch keineswegs haben wir dabei den Eindruck, die ganze Welt habe plötzlich einen Sprung nach links gemacht. Bewegen wir uns auf einen Menschen zu, so scheint dieser seine Körpergröße „natürlich“ nicht zu verändern, obwohl doch sein Abbild auf unserer Netzhaut wächst:

Objekteigenschaften wie die physikalische Größe eines Gegenstandes oder dessen Farbe sind subjektiv für uns unveränderlich. Sie bilden Invarianten der Wahrnehmung.

¹Nicht zufällig muß uns das Sprichwort „Nachts sind alle Katzen grau“ daran erinnern, daß wir in der Dämmerung fast gar keine Farben wahrnehmen, ohne daß uns das auffiele.

1.2 Invariante Erkennung von Sehobjekten

Bei der Betrachtung einer visuellen Szene nehmen wir bestimmte Szenenbestandteile als zusammenhängende, unabhängige Bestandteile wahr. So betrachten wir sämtliche Teilreize, die zusammen beispielsweise den Eindruck einer Tasse oder eines Baumes bilden, als ein zusammengehöriges Objekt, noch bevor wir den eigentlichen Gegenstand als Baum oder Tasse erkennen. Das gleiche gilt für gänzlich unbekannte Objekte.

Gegenstände visueller Wahrnehmung

Diese *Sehobjekte* sind die Gegenstände der visuellen Wahrnehmung. Wir empfinden sie als eigenständige Bestandteile der visuellen Szene. Oft sind die Gegenstände der visuellen Wahrnehmung mit den physikalischen Gegenständen unserer Umwelt identisch, wie beim Beispiel des Baums oder der Tasse. Allerdings kann ein Sehobjekt auch Teile oder Gruppen physikalischer Gegenstände umfassen. Ebenso kann es sich um eine Zeichnung oder um die einzelnen Buchstaben und Wörter auf einer Textseite wie dieser handeln. Sehobjekte werden außerdem durch geometrische Umstände, etwa den räumlichen Zusammenhang oder gleichartige Bewegung definiert. Gesetzmäßigkeiten dieser Art, nach denen Sehobjekte im visuellen Reiz identifiziert und gruppiert werden, wurden schon früh in den *Gestaltgesetzen* formuliert (Köhler, 1920; Koffka, 1935; Metzger, 1953). Bis heute sind die zugrundeliegenden neuronalen Mechanismen jedoch unklar, welche zur Segmentierung einzelner Sehobjekte im visuellen Reiz, zu ihrer Betrachtung als Entitäten und zu ihrer Trennung vom Hintergrund führen.

Die Wahrnehmung von Sehobjekten als unabhängige Bestandteile der visuellen Szene erlaubt es uns, diesen Objekten individuelle Eigenschaften zuzuordnen. Allem voran sind die Gegenstände unserer visuellen Wahrnehmung mit grundlegenden geometrischen Attributen behaftet: Sie befinden sich an einem bestimmten Ort in unserem Sehfeld oder in unserer Umgebung. Sie haben eine bestimmte Form, eine bestimmte räumliche Größe, und sie nehmen eine bestimmte Entfernung von uns ein.² Die geometrischen Eigenschaften Ort, Größe und Entfernung sind allen Sehobjekten eigen. Wir sind außerdem in der Lage, das gleiche Sehobjekt anhand seiner Form wiederzuerkennen, wenn es an verschiedenen Stellen oder in unterschiedlichen Konfigurationen in unserer Sehumgebung auftaucht.

Sehobjekte sind mit Eigenschaften behaftet

²Wie untrennbar diese Attribute mit unserer bewußten Wahrnehmung der Sehobjekte verknüpft sind, ist an der irritierenden Wirkung der hinlänglich bekannten sogenannten „optischen“ Täuschungen abzulesen: Hier weichen die wahrgenommenen Objekteigenschaften von den physikalisch meßbaren ab.

Kapitel 1 Invarianten der visuellen Wahrnehmung

In dem Moment, in dem wir einem Sehobjekt einen real existierenden physikalischen Gegenstand zuordnen, kommt eine Vielzahl weiterer Attribute hinzu, darunter solche wie Gewicht, Geschmack und Oberflächeneigenschaften, oder abstraktere wie Klassenzugehörigkeit und die Ausführbarkeit bestimmter Aktionen und Manipulationen mit dem Objekt (beispielsweise die Aktionen *nehmen* und *essen*). Unsere Fähigkeit, bekannte Gegenstände in den unterschiedlichsten Zusammenhängen wiederzuerkennen und ihnen sinnvolle Bedeutungen zuzuordnen, ist eine kognitive Leistung, die unser Überleben sichert: Man denke beispielsweise an die Fähigkeit, eine bekannte Person auch dann wiederzuerkennen, wenn sie uns in anderer Umgebung begegnet, wenn sie einen ungewohnten Gesichtsausdruck hat, andere Kleidung trägt oder um viele Jahre gealtert ist. Erkennungsleistungen wie diese, insbesondere solche, die mit einer bestimmten Bedeutung des Objektes für unser Verhalten zusammenhängen, involvieren außer dem für unsere visuelle Wahrnehmung verantwortlichen Teil des Nervensystems eine Vielzahl weiterer kortikaler Vorgänge. Die damit verknüpften Fragestellungen gehen weit über den Kontext hinaus, den diese Arbeit behandeln kann.

Hierarchisches Modell der Objekterkennung

Beschränken wir uns jedoch auf das visuelle kortikale System, so muß dieses eine entscheidende Grundlage für kognitive Leistungen dieser Art liefern: Um einen uns bekannten Gegenstand, ein Gesicht, eine Person in einem Sehobjekt (wieder-)erkennen zu können, müssen wir dessen Form wahrnehmen und diese gegen ein internes Wissen abgleichen. Die Art, in der das Wissen über bekannte Objekte in unserem Nervensystem vorliegt, ist weitgehend ungeklärt. Sehr wahrscheinlich ist es über mehrere hierarchische Stufen von Nervenzellen und deren Synapsen verteilt, und manifestiert sich in der Eigenschaft der Neuronen, selektiv auf bestimmte Reizkonfigurationen anzusprechen (sog. hierarchisches Modell der Objekterkennung, vgl. z.B. Riesenhuber und Poggio, 1999b). Die Verarbeitung im visuellen Kortex geschieht dabei verteilt über eine Vielzahl von Arealen, läßt sich jedoch in zwei Hauptverarbeitungsströme aufteilen (vgl. Felleman und van Essen, 1991; Kandel et al., 2000): den *ventralen Pfad* zum temporalen Kortex, entlang dem die Analyse und das Erkennen der Form eines Sehobjektes stattfindet, und den *dorsalen Pfad* zum parietalen Kortex, entlang dem Ort, räumliche Relation und Bewegung verarbeitet werden (Abbildung 1.1). Wir sagen, daß der visuelle Reiz durch die Aktivität der Nervenzellen in den Hirnarealen *verteilt repräsentiert* wird.

Entspricht die verteilte Nervenaktivität, die beim Betrachten einer Form hervorgerufen wird, hinreichend stark einem „abgelegten“ Muster, so identifizieren wir das Sehobjekt mit dem entsprechenden

1.3 Geometrische Invarianzen

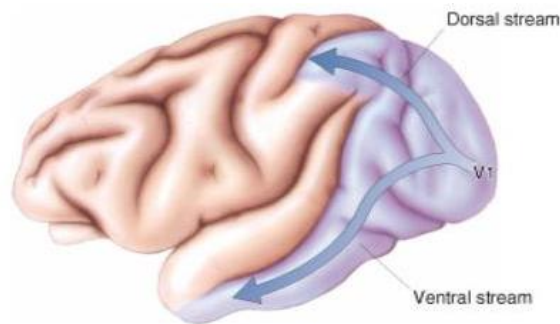


Abbildung 1.1: Dosaler und ventraler Pfad der visuellen Verarbeitung.

Gegenstand: Wir haben das Objekt nicht nur als individuellen visuellen Reiz wahrgenommen, sondern auch in seiner Identität erkannt.

1.3 Geometrische Invarianzen

Bei näherer Betrachtung erweist sich der Vergleich des visuellen Reizes mit unserem internen Wissen über bekannte Objekte als äußerst kompliziert. Ein Grund hierfür ist, daß die konkreten visuellen Reize auf unserer Netzhaut, die als ein und dasselbe Objekt erkannt werden sollen, sehr unterschiedlich aussehen können: Wir erkennen Gegenstände aus verschiedenen Blickwinkeln; wir erkennen sie, wenn sie liegen oder auf dem Kopf stehen; wir erkennen sie, wenn sie sich links oder rechts befinden, dann, wenn sie nah und ihre Netzhautbilder groß, oder wenn sie fern und ihre Netzhautbilder winzig sind. Man spricht in diesem Zusammenhang von Ansichtsinvarianz, von Rotations-, Positions- oder von Größeninvarianz.

*Ansichts-, Rotations-,
Positions-,
Größeninvarianz*

Diese Begriffe werden einerseits im Rahmen psychophysischer Untersuchungen verwendet, wo sie sich auf das Erkennen, also eine kognitive Leistung beziehen. Andererseits bezeichnen sie aber auch die aus dem Netzwerkaufbau resultierende Eigenschaft bestimmter Nervenzellen, auf die entsprechenden visuellen Reize unverändert zu reagieren: Wird eine Zelle im visuellen Kortex von verschiedenen Ansichten des selben Gegenstandes gleichermaßen aktiviert, so bezeichnen wir deren Antwort als ansichtsinvariant. Invariant antwortende Neuronen finden sich vor allem im inferotemporalen Kortex, einem Areal, das in der Hierarchie der visuellen Verarbeitung weit oben im ventralen Formverarbeitungspfad angesiedelt ist. Einzelzelleitungen weisen dort Positions-, Größen- und Ansichtsinvarianz nach (Ito et al., 1995; Booth und Rolls, 1998; Perrett et al., 1991). Zellen mit einem lokal positionsinvarianten Antwortverhalten, die

*Invariante
Zellantworten*

Kapitel 1 Invarianten der visuellen Wahrnehmung

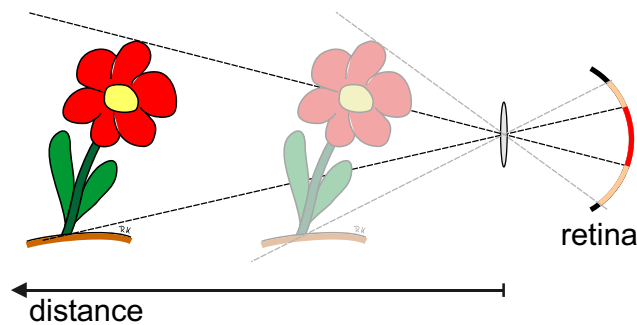


Abbildung 1.2: Geometrische Varianz, hier am Beispiel der Bildgröße: Variation des Sehabstandes verursacht eine Variation der retinalen Projektionsgröße. Der Zusammenhang ist durch die Gesetze der Strahlenoptik gegeben.

Complex-Zellen, sind jedoch auch schon im primären visuellen Areal V1 anzutreffen (zuerst beschrieben an der Katze, Hubel und Wiesel, 1959, 1962). Die invarianten Antworten der Neuronen im visuellen Kortex bauen höchstwahrscheinlich aufeinander auf und nehmen daher im visuellen Verarbeitungsstrom zu den höheren Arealen hin zu (vgl. Abschnitt 2.1). Hierbei geht die Repräsentation vom konkreten, an definierten Stellen im Sehfeld vorliegenden visuellen Reiz nach und nach in eine orts- und konfigurationsunabhängige, dafür objektbezogene Repräsentation über.

Formunabhängige geometrische Transformationen

Es würde unser Erinnerungsvermögen vermutlich überschreiten, sämtliche möglichen Ansichten eines Objektes einzeln abzulegen. Was die *geometrischen Varianzen* betrifft, also diejenigen Variationen der Netzhautbilder ein und desselben Gegenstands, die durch verschiedene Betrachtungspositionen gegeben sind, so erscheint eine andere Annahme viel wahrscheinlicher: Diese Variationen der Netzhautbilder werden durch die Gesetze der optischen Abbildung im Auge bestimmt (Abbildung 1.2). Bewegen wir unsere Augen oder nehmen eine geänderte Betrachterposition gegenüber einem Gegenstand ein, etwa indem wir uns dem Gegenstand nähern, so transformiert sich dessen Netzhautbild meist in einer Weise, die nur von der Art der Blickveränderung abhängt, und nicht von der Form des Gegenstandes.³ Es erscheint sehr plausibel, daß unser visuelles System diesen immergleichen Zusammenhang erlernt und die Fähigkeit entwickelt hat, die immer wiederkehrenden geometrischen Transforma-

³Bei Veränderung des Ansichtswinkels (Objektdrehung im Raum) ist die Formunabhängigkeit nur für hinreichend kleine Variationen gegeben. Sie gilt jedoch vollständig für Positions- und Größenveränderungen, sowie Rotation in der Sichte ebene.

1.3 Geometrische Invarianzen

tionen auszugleichen – und dies für *alle* Sehobjekte.

Um die Modellierung dieses Prinzips geht es in der vorliegenden Arbeit. Abstrakt formuliert ist das Problem, die eindeutige und unveränderliche Identität des Objektes, wie sie in der Außenwelt existiert, auch in unserer inneren Wahrnehmungswelt zu erzeugen. Da diese eindeutige Identität aber durch die variante optische Abbildung verlorengelht, muß sie *zurückgewonnen* (bzw. aktiv rekonstruiert) werden.

Kapitel 2

Klassische Modellierung geometrischer Invarianzen

Ziel

Ich stelle die wichtigsten existierenden Ansätze zur Modellierung geometrischer Invarianzen vor: hierarchische Gruppierung und dynamisches *Routing*. Beide Ansätze arbeiten auf hierarchisch strukturierten Netzwerken. Ich formuliere das verallgemeinerte Konzept der *Complex*-Bildung. Ich erläutere das Bindeproblem und das Problem überlappender Repräsentantenmengen, die in Modellen dieser Art wesentliche Schwierigkeiten bereiten.

Bei der Modellierung neuronaler Netze zur visuellen Objekterkennung erhielten Ansätze mit statischen Netzwerkaufbauten traditionell mehr Aufmerksamkeit als solche, die eine dynamische Anpassung des Netzwerkes verwenden. Das Konzept einer gestaffelten, jedoch festen Netzwerkkonstruktion tauchte in Gestalt der *mehrlagigen Perzeptrons* schon sehr früh in der Geschichte der künstlichen neuronalen Netzwerke auf und wurde mit außerordentlichem Erfolg eingesetzt (vgl. Bishop, 1995). Bis heute sind diese Netzwerke Gegenstand intensiver Forschung.

Beeinflusst von dieser Vorstellung gehen die meisten Modelle des visuellen Systems, ja selbst eine Vielzahl experimenteller Arbeiten, von einer Struktur des Kortex aus, die lediglich während einer begrenzten Lern- oder Entwicklungsphase des Individuums plastisch ausgebildet wird. Nach Abschluß des Lernens, im adulten Zustand, wird jedoch eine feste Konfiguration erreicht. Die Unveränderlichkeit des Netzwerkes wird hier oft als notwendig für eine akkurate und schnelle Repräsentation der Umgebung angesehen. Mit der zunehmenden Erforschung plastischer Restrukturierungseffekte im erwachsenen Gehirn, insbesondere auch auf sehr schnellen Zeitskalen,

erfährt dieses Dogma jedoch einen Umbruch (vgl. Weinberger, 1995, als *Review* zu dieser Thematik).

2.1 Statische hierarchische Merkmalsgruppierung

Hubel und Wiesel's Complex-Zelle

Auf dem Gebiet der invarianten Muster- und Objekterkennung werden mit statischen Ansätzen sehr gute Erfolge erzielt. Sie erzeugen invariante neuronale Antworten über das Prinzip der *hierarchischen lokalen Gruppierung* (engl. *pooling*) der Signale von merkmalsselektiven Neuronen im Netzwerk (siehe z.B. Fukushima, 1980; Perrett und Oram, 1993; Riesenhuber und Poggio, 1999b). Die Idee geht zurück auf den Vorschlag von Hubel und Wiesel (1959, 1962) zur Verschaltung einer lokalen Gruppe positionsspezifischer „Simple“-Zellen auf eine einzige, dann lokal positionsinvariante „Complex“-Zelle (Abb. 2.1). Wie bereits Hubel und Wiesel (1965) feststellten, läßt sich das gleiche Verfahren in einer nächsten Stufe wiederum zur Erzeugung von „Hypercomplex“-Zellen höherer Ordnung anwenden.

2.1.1 Complex-Bildung als Konzept

Aus dem Vorschlag von Hubel und Wiesel ist ein verallgemeinertes Konzept ableitbar, das ich *Bildung von Complex-Eigenschaften* oder kurz *Complex-Bildung* nenne. Zur Bildung von *Complex-Eigenschaften* wird eine Gruppe von Neuronen, welche die verschiedenen Erscheinungsformen eines visuellen Reizes variant repräsentieren, konvergent verschaltet (vgl. Abbildungen 2.1, 2.2). Wird die Gruppenaktivität durch ein nachgeschaltetes Neuron mit geeigneter Schwelle ausgewertet¹, so ist dessen Antwort invariant gegenüber den jeweiligen Erscheinungsformen des Stimulus (neuronale ODER-Bildung):

Ist $\{d_i\}$ eine Menge neuronaler Detektoren² für einen Satz von

¹Je nach Lage der Antwortschwelle kann das nachgeschaltete Neuron die logischen Operationen UND bzw. ODER implementieren.

²Der Begriff „Detektor“ für reizselektiv antwortende Neuronen findet in der Literatur allgemeine Verwendung. Neuronen im visuellen System sind *keine* idealen Detektoren, welche dann und nur dann antworten, wenn ein definierter Reiz anliegt. Ihre charakteristischen Antwortkurven (engl. *تونينجس*) für Stimulusparameter sind oft breit und hängen vom Kontext ab. Gemeint ist stets eine relative Bevorzugung gewisser Stimuluskonfigurationen. Ich werde den Begriff in diesem Sinne weiterverwenden.

2.1 Statische hierarchische Merkmalsgruppierung

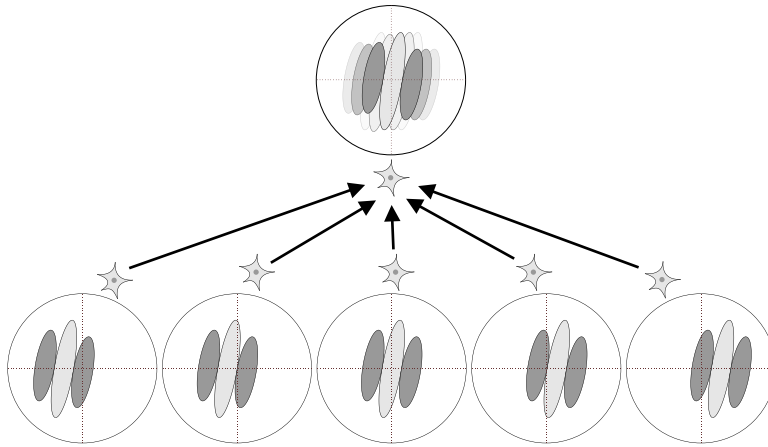


Abbildung 2.1: *Complex-Bildung für Position* (nach Hubel und Wiesel, 1962): Formierung eines lokal positionsinvarianten, formspezifischen Detektors aus einer Reihe positions- und formspezifisch antwortender Neuronen. (Schematische Darstellung. Unter und über den Zellen sind deren bevorzugte Stimuli angedeutet. Die Kreise repräsentieren jeweils einen identischen Ausschnitt des Sehraums.)

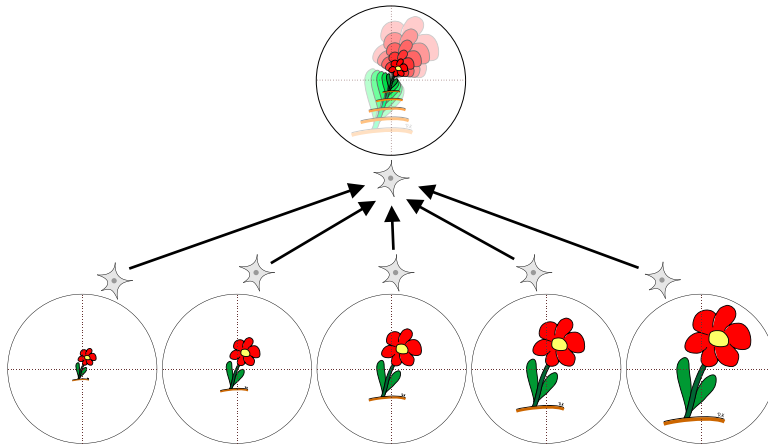


Abbildung 2.2: *Complex-Bildung für Größe*: Formierung eines größeninvarianten, formspezifischen Detektors aus einer Reihe größen- und formspezifisch antwortender Neuronen. (Schematische Darstellung wie Abb. 2.1.)

Kapitel 2 Klassische Modellierung geometrischer Invarianzen

Stimuli $\{s_i\}$ mit

$$d_i : s \mapsto \{0,1\}$$

$$d_i(s_i) = 1$$

und konvergiert die Aktivität der Einzeldetektoren d_i auf einen neuronalen Gruppendetektor D derart, daß

$$D(s) = \bigvee_i d_i(s)$$

so ist

$$D(s_i) = 1 \quad \forall s_i.$$

Das Konzept der Complex-Bildung ist von der Art der zugrundeliegenden Stimuli unabhängig. Die variante Eigenschaft der Stimuli kann Position, Größe, oder jedes andere Attribut sein.

Konjunktion und Disjunktion

Es erweist sich als unpraktikabel, mit diesem Verfahren die insgesamt angestrebte Invarianz in einem einzigen Gruppierungsschritt zu erzeugen: Die Zahl unterschiedlicher Erscheinungsformen einzelner Reize, welche invariant erkannt werden sollen, ist hierfür zu groß. Vielmehr werden in wiederholten Gruppierungsschritten *progressiv invariante* Detektoren aufgebaut, indem in jeder Hierarchiestufe über eine weitere, begrenzte Klasse von Stimuluseigenschaften gruppiert wird. Darüber hinaus werden die Signale bereits teilinvariant reagierender Neuronen kombiniert zu Detektoren, die selektiv bestimmte *Merkmalskombinationen* codieren (Abbildung 2.3). Es ergibt sich eine Hierarchie aus *Konjunktion* (logische UND-Verknüpfung) und *Disjunktion* (*Complex-Bildung*, logische ODER-Verknüpfung³) von Merkmalen (siehe z.B. Fukushima, 1980; Riesenhuber und Poggio, 1999b), an deren oberen Ende Neuronen mit dem erwünschten invarianten Antwortverhalten stehen. Das Verfahren hat Ähnlichkeit mit dem aus der Informatik bekannten Prinzip des *Codierungsbaumes*.⁴

³Riesenhuber und Poggio schlagen zur Implementierung der Disjunktionsstufen statt der traditionell verwendeten Summation mit anschließender Schwellenoperation die Maximumsbildung vor (Riesenhuber und Poggio, 1999b; Yu et al., 2002).

⁴Viele Modellierungsarbeiten zur invarianten Objekterkennung reduzieren die visuelle Verarbeitung auf den ventralen kortikalen Pfad. Sie zielen darauf ab, in einer strengen Hierarchie vollkommene Invarianz für die geometrischen Merkmale Position, Größe und Ansichtswinkel herzustellen. Es existieren jedoch neurophysiologische Arbeiten, die diese Annahmen fragwürdig erscheinen lassen. So zeigen

2.1 Statische hierarchische Merkmalsgruppierung

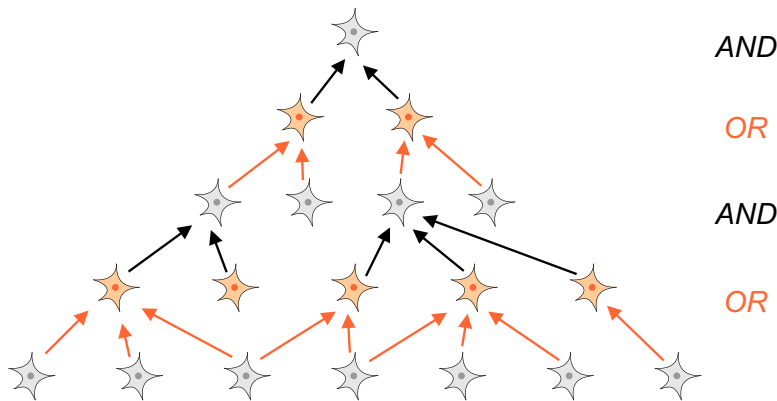


Abbildung 2.3: Hierarchie aus Konjunktion und Disjunktion bei der hierarchischen Gruppierung. UND-Stufen erzeugen Spezifität für Merkmalskombinationen. ODER-Stufen erzeugen Teilinvarianz für eine Gruppe von Merkmalen durch *Complex*-Bildung.

Das Erzeugen von Spezifität für eine bestimmte Stimulusklasse einerseits, und der Invarianz gegenüber gewissen Repräsentanten dieser Klasse andererseits, vollzieht sich bei der statischen hierarchischen Merkmalsgruppierung in abwechselnden Teilschritten.

2.1.2 Binde-Problem

Insbesondere bei der Modellierung geometrischer Invarianzen wie der für Größe oder Position treten hierbei jedoch bedeutende Schwierigkeiten auf. In der Gruppierungs-Hierarchie nimmt von unten nach oben die Invarianzleistung zu. Gleichzeitig damit wachsen die *rezeptiven Felder* der Zellen, also die Regionen im Sehfeld, aus denen die einzelnen Neuronen ihre Stimulation beziehen können. Dieser beabsichtigte Umstand wird jedoch von einem unerwünschten Nebeneffekt begleitet: Da stets bereits teilinvariante Merkmalsdetektoren kombiniert werden, kann im Bereich ihrer rezeptiven Felder die Information über die räumliche Konfiguration einzelner Stimulusbestandteile zueinander verloren gehen. Die Identität eines Sehobjektes ist aber nicht nur durch das bloße Vorhandensein all seiner

Innere Struktur von Sehobjekten

Neuronen im inferotemporalen Kortex zu einer starken Abhängigkeit von der Kontrastpolarität des Stimulus (Ito et al., 1994), was bedeutet, daß ihre Eigenschaften nicht allein hierarchisch auf denen der polaritätsinvarianten *Complex*-Zellen im Areal V1 aufbauen können. Zum anderen weisen sie geringere Positionsinvarianz auf, als bislang angenommen (op de Beek und Vogels, 2000; DiCarlo und Maunsell, 2003). Modelle zur Erreichung umfassender geometrischer Invarianz in einer isolierten UND/ODER-Hierarchie sind daher als artifiziell und wenig biologisch relevant einzustufen. Hingegen ist anzunehmen, daß parallele und rückgekoppelte Verarbeitungsströme im visuellen Kortex sich gegenseitig unterstützen.

Kapitel 2 Klassische Modellierung geometrischer Invarianzen

Teilterkmale gegeben. Ebenso entscheidend ist die korrekte Relation dieser Merkmale, d.h. seine innere Struktur. Was einerseits die gewünschte Insensitivität gegenüber lokalen Variationen herbeiführt, birgt also andererseits die Gefahr einer Verwechslung von Objekten mit ähnlichen Bestandteilen, jedoch in unterschiedlicher Konfiguration.⁵

Lösungsansätze

Durch eine geschickte Wahl der Konjunktions- und Disjunktionsgruppen läßt sich dieses Problem vermindern. So stellen etwa Mel und Fiser (2000) einen aufwendigen Auswahlalgorithmus für die Merkmalsgruppierung vor, der die eindeutige Erkennung eines Objektes aus einer Testmenge gewährleistet. Aus ihren Ergebnissen lassen sich allgemeine Aussagen darüber ableiten, wie Merkmalsgruppen günstig gewählt werden können, um zur Disambiguierung verwechselbarer Stimuli beizutragen. Andere, biologisch plausible Ansätze nutzen die zeitliche Kohärenz natürlicher Stimuli aus, um aus zeitlich benachbarten Erscheinungsformen eines bewegten Stimulus geeignete Gruppierungsbereiche für die *Complex*-Bildung zu lernen (*Trace Rule*, Földiak, 1991; Wallis, 1996; Wallis und Rolls, 1997; Rolls und Milward, 2000) oder zu errechnen (*Slow Feature Analysis*, Wiskott und Sejnowski, 2002).

Verarbeitung in parallelen Pfaden

Auch eine Verarbeitung in parallelen Hierarchien, so wie sie für das visuelle System mit dem ventralen und dorsalen Pfad angenommen wird, könnte die korrekte Darstellung der Reizkonfiguration erlauben: Während der ventrale Verarbeitungspfad eine geometrisch invariante Repräsentation des Sehreizes herstellt, abstrahiert der dorsale Pfad von der Objektform, verarbeitet jedoch die Orts- und Bewegungsinformation. Die Fragestellung, wie verschiedene Stimulusqualitäten im visuellen System zu einem eindeutigen Perzept kombiniert werden, wird in aktuellen Forschungsarbeiten untersucht (Eckhorn et al., Gewaltig et al., persönliche Kommunikation).

Inwieweit aber die statisch-hierarchischen Netzwerkmodelle der invarianten Objekterkennung aufgrund des Verlustes von räumlicher Konfigurationsinformation *prinzipiellen* Einschränkungen unterliegen ist seit Jahren Gegenstand kontroverser und nach wie vor offener Diskussion (z.B. von der Malsburg, 1999; Riesenhuber und Poggio, 1999a; Mel und Fiser, 2000). Die Thematik wird unter dem Begriff *Binde- oder Superpositions-Problem* als ein Standardproblem in der Literatur behandelt, und soll an dieser Stelle nicht in ihrer Tiefe besprochen werden.

⁵Die Buchstaben L, + und T bestehen jeweils sowohl aus einem horizontalen wie einem vertikalen Balken. Würden zur Detektion der einzelnen Balken jeweils positionsinvariante Detektoren eingesetzt, so könnten die drei Buchstaben nicht unterschieden werden.

2.1 Statische hierarchische Merkmalsgruppierung

Anmerkung

Der Ausdruck *Binde-Problem* bezeichnet ebenfalls das Problem, Stimulus-Modalitäten, die in getrennten kortikalen Arealen und entlang verschiedener Pfade verarbeitet werden, zu einer kohärenten Wahrnehmung zu kombinieren (etwa Farbe, Form, Ort, Tiefe und Bewegung). Einesfalls ist also die Zuordnung *innerhalb* eines Verarbeitungspfades gemeint (intra-modal), anderenfalls die Zuordnung *zwischen* Pfaden (inter-modal). Die Begriffe werden in der Literatur leider nicht gegeneinander abgegrenzt. Obwohl beide Probleme verwandt sind, können für sie völlig unabhängige Lösungen existieren (vgl. Abschnitt 2.2).

2.1.3 Überlappende Repräsentantenmengen

Als prinzipielles Problem des statischen Gruppierungsansatzes erweist sich, daß hierbei über die Menge der möglichen Erscheinungsformen (Repräsentanten) eines Objektes generalisiert wird. Was das bedeutet möchte ich am Beispiel der größeninvarianten Erkennung demonstrieren:

In klassischen Modellansätzen zur Generierung von Größeninvarianz mittels statischer Merkmalsgruppierung werden retinale Bilder eines Objektes in verschiedenen Ausdehnungen als identische Objekte erkannt. Dies wird erreicht, indem die Antworten formsелеktiver Neuronen für unterschiedlich große Erscheinungsformen desselben Objektes im Sinne der *Complex*-Bildung zusammengefaßt werden: An der Spitze der Formerkennungshierarchie stehen dann Neuronen, die auf sämtliche verschieden großen Bilder eines Objektes gleichermaßen ansprechen (Abbildung 2.2). Das bedeutet jedoch, daß die Information über die Bildgröße in der Hierarchie verlorengeht: Netzwerke dieser Art werden gerade dafür entworfen, von der Bildgröße zu abstrahieren.

Die Generalisierung über Repräsentanten wird dann zum Problem, wenn physikalisch verschiedene Objekte überlappende Repräsentantenmengen besitzen, also gleiche retinale Bilder hervorrufen. Dies ist etwa bei Gegenständen verschiedener physikalischer Größe, jedoch gleicher Form der Fall – z.B. bei einem echten PKW und seinem maßstabsgetreuen Spielzeugmodell. Diese Gegenstände können in einem streng statischen Gruppierungsansatz⁶ *prinzipiell* nicht sicher unterschieden werden, da die Mengen ihrer möglichen Netzhautbilder (Repräsentanten) überlappen, und in diesem Überlappungsbereich keine eindeutige Zuordnung des Netzhautbildes zum

⁶Die klassische Modellierung geht hier typischerweise von einem isolierten Verarbeitungspfad aus, vgl. Fußnote auf S. 12.

Kapitel 2 Klassische Modellierung geometrischer Invarianzen

Objekt möglich ist: Das Spielzeugauto erzeugt von nahe betrachtet das gleiche Bild auf der Retina wie der echte PKW aus der Ferne.

Objekte mit überlappenden Repräsentantenmengen können in streng statisch-hierarchischen Gruppierungsmodellen prinzipiell nicht sicher unterschieden werden.

Das Problem überlappender Repräsentantenmengen ist in Kapitel 6 von Bedeutung.

2.2 Dynamisches Routing

Router-Circuits

Olshausen, Anderson und van Essen haben eine Variante der hierarchischen Merkmalsgruppierung vorgestellt, bei der die Neuronengruppen während des normalen Netzwerkbetriebes dynamisch verändert werden (*Shifter- oder Router-Circuits*, Anderson und van Essen, 1987; Olshausen et al., 1993). Der Ansatz beruht auf dem Vorhandensein einer großen Zahl simulierter Synapsen zwischen den Netzwerkneuronen, von denen jedoch im normalen Betrieb der überwiegende Teil inaktiv ist. Durch konzertiertes Umschalten umfangreicher Synapsengruppen läßt sich nun ein in Größe und Position variabler Ausschnitt des retinalen Stimulus selektieren und durch die Hierarchie des Netzwerkes „leiten“ (engl. *to route*). Eine Objekterkennung kann dann auf dem so in Größe und Position normalisierten Bildausschnitt operieren. Die Auswahl des passenden retinalen Ausschnittes vollzieht sich dabei *vom retinalen Stimulus getrieben*, d.h., das Netzhautbild wird vorab von einem vereinfachten neuronalen Netzwerk auf hervorstechende Bereiche hin ausgewertet und der Leitmechanismus entsprechend eingestellt.

Objekt/Hintergrund-Trennung

Zu den wesentlichen Kritikpunkten dieses Modells zählt, daß der neuronale Mechanismus zur Auswahl des Bildausschnittes eine zumindest rudimentäre Objekt/Hintergrund-Trennung leisten muß. In Olshausen et al. (1993) wird hierzu auf neuronalem Wege eine Tiefpaßfilterung des visuellen Reizes vorgenommen und der Bereich größter Antwort in Position und Ausdehnung bestimmt. Dieses Verfahren verspricht bei komplexen visuellen Szenen geringen Erfolg. Im allgemeinen ist die neuronale Objekt/Hintergrund-Trennung ein ungelöstes Problem (vgl. z.B. Gail, 2002).

Binde-Problem

Das dynamische *Routing* stellt eine mögliche Vermeidung des Binde-Problems bezüglich der geometrischen Relation von Teilmerkmalen dar (Abschnitt 2.1.2), da hier der retinale Stimulus vor der Erkennung vollständig und unter Erhalt der Topographie in einen

2.2 Dynamisches Routing

objektzentrierten Referenzrahmen überführt wird. Die Objekterkennung kann damit unbelastet von Größen- und Positionsvariationen erfolgen. Invarianzgenerierung und Erkennung finden bei diesem Ansatz in zwei getrennten, aufeinanderfolgenden Schritten statt, und nicht mehrfach abwechselnd wie bei der statischen hierarchischen Merkmalsgruppierung.

Die vom retinalen Stimulus getriebene Steuerung ist andererseits die Ursache dafür, daß das Problem der überlappenden Repräsentantenmengen beim dynamischen *Routing* unverändert bestehen bleibt: Rufen zwei verschiedene physikalische Objekte, z.B. der PKW und sein Spielzeugmodell, unter gewissen Bedingungen das gleiche Netzhautbild hervor, so wird dieses gleiche Netzhautbild dafür sorgen, daß der neuronale Leitmechanismus in beiden Fällen gleich eingestellt wird. Entsprechend werden in beiden Fällen die gleichen Neuronen aktiviert werden, und eine Unterscheidung der beiden Objekte ist, prinzipbedingt, unmöglich.

**Repräsentanten-
Problem**

Objekte mit überlappenden Repräsentantenmengen können auch im dynamischen Routing-Modell prinzipiell nicht sicher unterschieden werden.

Kapitel 3

Situative Steuerung als Invarianz-Strategie

Ziel

Als alternative Sichtweise auf die Erzeugung geometrischer Invarianzen schlage ich die situative Steuerung neuronaler Codierungseigenschaften vor. Diese beruht darauf, daß Informationen von außerhalb des Sehpfadcs über die körperliche Situation des Individuums in seiner Umwelt zum Zwecke der Invarianzgenerierung ausgewertet werden. Ich führe Belege für den Einfluß der Sehsituation auf die Verarbeitung im visuellen System an und bereite dieses Prinzip zum Zwecke der konzeptionellen Modellierung auf. Ich schlage Methoden vor, die rezeptiven Felder einzelner Neuronen in Abhängigkeit äußerer Einflüsse zu verändern und motiviere meine Annahme, daß diese Einflüsse in Form einzelner Zustandsgrößen in die neuronale Verarbeitung eingehen.

3.1 Individuum und Umwelt

Die mehr oder weniger blinde Gleichbehandlung sämtlicher potentiell auftretender Erscheinungsformen eines Objektes – sei es in einem statischen oder einem dynamischen Ansatz – scheint *per se* keine dem Problem angepaßte Methode zu sein. Da ein fester Zusammenhang besteht zwischen der Betrachterposition, die wir einem Objekt gegenüber einnehmen, und dem Bild, das es auf unserer Netzhaut erzeugt, wäre es intuitiv einleuchtender, die aktuelle Sehsituation in Betracht zu ziehen, und die geometrischen Variationen des Objektbildes gezielt neuronal zu kompensieren: Wie ein Fotograf selbstverständlich die Ausrichtung und den Zoom eines Teleobjektivs auf Position und Entfernung des abzubildenden Objektes einstellt, so wird

Kapitel 3 Situative Steuerung als Invarianz-Strategie

auch das Sehsystem diese Informationen verwenden, um die Verarbeitung des Sehreizes in Position und Größe anzupassen. Eine entsprechende Annahme über die sensorische Verarbeitung wurde bereits 1950 durch von Holst und Mittelstädt formuliert (*Reafferenzprinzip*, von Holst und Mittelstädt, 1950).

Relation zur Umwelt

Beide in Kapitel 2 besprochenen Konzepte, die hierarchische Gruppierung (Abschnitt 2.1) wie auch das *Routing* (Abschnitt 2.2), versuchen hingegen, die invariante Erkennung allein ausgehend vom retinalen Stimulus zu realisieren. Sie vernachlässigen dabei den Umstand, daß die Objekterkennungshierarchie nicht in einem isolierten Laborsystem vorliegt, sondern Bestandteil eines komplexen, autonom in seiner Umwelt agierenden Lebewesens ist. Wahrnehmung, auch die visuelle, ist weniger ein Prozeß passiver Projektion als die Art und Weise, in der ein Lebewesen die Umwelt durch seinen Körper und seine Aktionen erlebt.

Augenausrichtung

Beispielsweise wird der zweifellos mächtigste und eleganteste Mechanismus zur Generierung visueller Positionsinvarianz durch eine *körperliche Aktion* realisiert, nämlich durch unsere Körper-, Kopf- und Augenausrichtung: Indem wir interessante Objekte fixieren, bringen wir ihr Netzhautbild stets in das hochauflösende Zentrum der Retina, und damit in eine Normalposition.¹ Neuronale Positionsinvarianz zur Objekterkennung ist daher vom visuellen System nur noch in sehr begrenztem Maße zu leisten – ein Umstand, der in fast allen Modellen gänzlich vernachlässigt wird.² Tatsächlich kommen neuere Untersuchungen zu dem Ergebnis, daß die Neuronen am oberen Ende des ventralen Pfades keineswegs in dem Umfang positionsinvariante Eigenschaften haben, wie frühere Experimente vermuten ließen. Tatsächlich bevorzugen viele Neuronen eine foveale Position des Stimulus (op de Beeck und Vogels, 2000). DiCarlo und Maunsell (2003) ermitteln mit modernen Methoden am aktiv umherblickenden Affen deutlich kleinere rezeptive Felder für Neuronen im

¹Da unsere Erkennungsleistung auf einen kleinen Bereich um den Fixationspunkt beschränkt ist (Henderson et al., 2003), werden Sehobjekte großer Ausdehnung durch eine Folge von Fixationsbewegungen abgetastet, so daß deren charakteristische Bestandteile nacheinander im zentralen Bereich der Retina zu liegen kommen. Eine Objekterkennung auf Basis dieses Vorgangs erfordert ein sakkadenübergreifendes Kurzzeitgedächtnis. Wie die sakkadischen Zielorte ausgewählt und die dort gewonnene Information über die Sakkaden hinweg integriert wird, ist Gegenstand aktueller Forschung (Review in Henderson, 2003).

²Zwar sind wir mit großen Einschränkungen in der Lage, Objekte auch bei Nichtfixation zu erkennen. Der Leser wird jedoch fraglos zustimmen, daß es einen untypischen Betriebsmodus des Sehsystems darstellt, an Gegenständen vorbeizublicken, die man erkennen möchte. Die Modellierung dieser Leistung darf daher als ein Randproblem der Materie gelten.

3.1 Individuum und Umwelt

anterioren inferotemporalen Kortex als zuvor beschrieben. Die Neuronen könnten demnach tatsächlich als positionsabhängige Detektoren angelegt sein, deren Invarianzleistung durch visuelle Erfahrung und Lernen erweitert wird (siehe DiCarlo und Maunsell, 2003, für eine vergleichende Diskussion).

Gleichzeitig reduziert der Mechanismus der aktiven Augenbewegungen das Binde-Problem (Abschnitt 2.1.2) auf ein Minimum, denn die Zusammengehörigkeit multimodal repräsentierter Stimuluskomponenten definiert sich in trivialer Weise über den Ort: Dem fixierten Objekt zugehörig sind all jene Komponenten, die sich in der Mitte des Sehfeldes befinden. Entstehen intramodale Mehrdeutigkeiten bezüglich der räumlichen Konfiguration von Reizbestandteilen, wie sie in Hierarchien aus Konjunktion und Disjunktion auftreten (Abschnitt 2.1.2), so lassen sich diese dadurch auflösen, daß die fragwürdigen Teile des Objektes nacheinander in einer Folge von Fixationsbewegungen (*Sakkaden*) abgetastet werden. Bestehen bleibt allerdings das Problem der Objekt/Hintergrund-Trennung, also die Auswahl der Fixationspunkte, welche zum selben Objekt gehören. O'Regan und Noë (2000, 2001) betrachten die aktive „Erforschung“ der Umwelt durch eine Folge von Sakkaden gar als die Grundlage der bewußten visuellen Erfahrung schlechthin. Psychophysische Untersuchungen zeigen, daß beim freien Sehen die Wahrnehmung von Objektinformationen auf einen kleinen Bereich um den Fixationspunkt beschränkt ist (Henderson et al., 2003).

Die aktive Erforschung der Umwelt durch Augenbewegungen erlaubt es also auf elegante Weise, die Probleme der *Positionsinvarianz* und der *Merkmalsbindung* weitgehend zu vermeiden. Gerade diese beiden gelten jedoch traditionell als zentrale Probleme der visuellen Objekterkennung. Offenbar eröffnet die Sichtweise auf das in seiner Umwelt agierende Individuum eine Ebene der Betrachtung, auf der wesentliche Probleme des isolierten Laborsystems gar nicht erst auftreten.

Umgekehrt bilden die physikalischen Gegebenheiten der Umwelt, so wie auch die des eigenen Körpers, obligatorische äußere Randbedingungen, denen das Lebewesen gerecht werden muß. Sensorik und Motorik des agierenden Individuums unterliegen physikalischen Zusammenhängen und Gesetzen. Es liegt nahe anzunehmen, daß physikalisch-körperliche Bedingungen des Lebewesens durch einen internen Zustand in der neuronalen Verarbeitung reflektiert werden. Er spiegelt die augenblickliche Situation des Individuums und seines Körpers in der Umwelt wider, denn diese ist entscheidend für eine kontextabhängige neuronale Verarbeitung. Gerade im dominanten sensorischen System des Primaten, dem visuellen, ist ei-

**Situationsabhängige
Verarbeitung**

ne solche Situationsabhängigkeit zu erwarten.

Basierend auf diesen Überlegungen schlage ich eine alternative Sichtweise auf das Problem der geometrischen Invarianzbildung vor. Sie beruht auf der Annahme, daß zum Zwecke der Objekterkennung physikalische Zustandsgrößen in die neuronale Verarbeitung einfließen, welche die körperliche Situation des Individuums in seiner Umwelt widerspiegeln.

Ich nehme an, daß Informationen über die physikalisch-körperliche Situation des Individuums in die neuronale Verarbeitung eingehen und zur Generierung geometrischer Invarianzen herangezogen werden.

3.2 Neurophysiologische Grundlagen

Bezogen auf die visuelle Verarbeitung ist eine wichtige physikalische Randbedingung für die Objekterkennung etwa das Verhältnis unseres eigenen Körpers zu den betrachteten Objekten in der Umwelt: In welcher Stellung befinden sich unser Kopf und unsere Augen? Wie weit ist das betrachtete Objekt entfernt? Bewegt sich das Objekt, oder bewegen wir nur unsere Augen? — Sollen diese Größen in die Verarbeitung einbezogen werden, so müssen sie dem visuellen System auf *neuronaler* Ebene verfügbar sein. Als notwendige Voraussetzung für die Richtigkeit der zuvor formulierten Annahme ist also zu prüfen, ob physikalisch-körperliche Randbedingungen in der Aktivität von Nervenzellen im visuellen System abgebildet werden.

Gainfields und Entfernungsmodulation

Messungen der neuronalen Aktivität im Sehkortex belegen indes, daß der Zustand vieler Zellen tatsächlich von der aktuellen Sehsituation beeinflusst wird: Eine Abhängigkeit der Zellaktivität beispielsweise von der Augen- und Kopfstellung ist für Neuronen im parietalen Kortex seit längerem unter dem Begriff *Gainfield* bekannt und wird dort in Zusammenhang mit Koordinatentransformationen gebracht (Andersen und Mountcastle, 1983; Andersen et al., 1985; Brotchie et al., 1995; Salinas und Thier, 2000). Offenbar wird hier die Information über die aktuelle Blickrichtung benutzt, um die Repräsentation unserer Umwelt von retinazentrierten in kopf- oder körperzentrierte Koordinaten umzusetzen – eine wesentliche Invarianzleistung, welche für die Koordination von Sensorik und Motorik erforderlich ist. Im ventralen intraparietalen Areal VIP werden Objekt- und Eigenbewegungen codiert. Bremmer et al. weisen hier die Integration vestibular- und somatosensorischer Information in die visuelle Verarbeitung nach. Sie dient dazu, Objekt- von Eigenbewegungen zu unterscheiden (Bremmer et al., 2002, 2004; Schlack et al.,

3.2 Neurophysiologische Grundlagen

2003). Diese Effekte tragen vermutlich wesentlich dazu bei, uns die Wahrnehmung einer stabilen Welt zu vermitteln.

Blicksituationsabhängige Modulation tritt jedoch auch bei Neuronen im ventralen Sehpfad auf, wie neuere experimentelle Ergebnisse von Dobbins et al. (1998); Rosenbluth und Allman (2002); Brinksmeyer et al. (2002) anschaulich belegen (vgl. Abbildung 6.1 auf Seite 84). Laut Dobbins et al. (1998) weisen etwa zwei Drittel der Neuronen in den Arealen V1, V2 und V4 eine Modulation ihrer Feuerrate in Abhängigkeit der momentanen Fixationsentfernung auf. Das bedeutet, daß in diesen Arealen die Mehrzahl aller Neuronen unterschiedlich auf Sehreize reagiert, welche völlig *identische Netzhautbilder* hervorrufen, jedoch in verschiedenen Entfernungen vom Betrachter präsentiert werden.³ Darüber hinaus stellen Rosenbluth und Allman (2002) in den ventralen Regionen des Sehkortex zusätzlich eine Abhängigkeit der neuronalen Antworten vom Sehwinkel fest, bei wiederum identischen Netzhautbildern.

Somit findet sich in den Kortexarealen des ventralen Sehpfades, welchen im allgemeinen die Leistung der Form- und Objekterkennung zugesprochen wird, in auffällig vielen Neuronen eine Abhängigkeit der Feuerraten von der momentanen Sehsituation in Blickrichtung und Fixationsentfernung.

In ventralen Arealen sind physikalisch-körperliche Randbedingungen wie Augenstellung und Fixationsentfernung neuronal verfügbar. Diese Eigenschaft kann zur Generierung geometrischer Invarianzen herangezogen werden.

Im folgenden werde ich ein Konzept zur Modellierung geometrischer Invarianzen entwickeln, das auf der Modulation neuronaler Codierungseigenschaften in Abhängigkeit physikalisch-körperlicher Parameter beruht. Die beschriebenen blickrichtungs- und entfernungsabhängigen Ratenänderungen in Neuronen des ventralen Sehpfades sind experimentelle Fakten, welche die Grundlage solcher Modelle bilden können. Die Kapitel 5 und 6 werden konkrete Modellbeispiele zu diesen beiden Effekten präsentieren.

³Dieser Entfernungseffekt ist vermutlich auf einen geänderten Vergenzwinkel der Augen zurückzuführen. Zusätzlich kommt eine Reihe weiterer Quellen in Betracht (Abschnitt 6.8.2, *Verfügbarkeit des Steuersignals*). Möglicherweise kann die Entfernungsabhängigkeit auch teilweise auf eine systematische Änderung der Disparität zurückzuführen sein (vgl. Cumming und Parker, 1999). Ein entsprechender Zusammenhang ist jedoch nicht nachgewiesen. Für alle hier gemachten Betrachtungen ist es unerheblich, wodurch die Aktivitätsmodulation verursacht wird, solange diese systematisch mit der Sehentfernung zusammenhängt.

3.3 Methoden der Modellierung

3.3.1 Konzeptionelle Modellierung

Diese Arbeit folgt, was die Wahl der Beschreibungsebene und der Modellkomponenten betrifft, dem Grundsatz der *konzeptionellen Modellierung* (vgl. Wissel, 1989):

| *So einfach wie möglich, so detailliert wie nötig!*

Ziel ist, ein Verständnis für die funktionellen Zusammenhänge zu erlangen, sowie die Bildung möglichst vielseitig wiederverwendbarer Modellierungsprinzipien. Ziel ist *nicht* die detaillgetreue Nachbildung des Nervensystems oder seiner Teile.

**Konzeptionelle und
Simulationsmodelle**

Aus komplexen sogenannten *Simulationsmodellen*, die möglichst viele der bekannten biologischen Fakten inkorporieren, lassen sich nützliche Vorhersagen gewinnen – hinreichend genaue Kenntnis einer genügend großen Zahl von Parametern, Randgrößen und deren Fehler vorausgesetzt. Sie eignen sich jedoch prinzipiell wenig dazu, Verständnis für funktionelle Zusammenhänge zu erlangen. Der Komplexitätsgrad dieser Modelle nähert sich naturgemäß dem des biologischen Systems an, was bedeutet, daß die wechselseitigen Abhängigkeiten ihrer Mechanismen und Parameter ähnlich schwierig zu verstehen sein werden wie die des natürlichen Vorbildes. Den detaillierten Simulationsmodellen stehen am anderen Ende der Komplexitätsskala die *konzeptionellen Modelle* gegenüber. Diese sind oft stark idealisierte, prozeßorientierte Denkmodelle, in die wesentliche Schlüsselfaktoren des vorliegenden Problems eingehen. Ihr hoher Abstraktionsgrad erlaubt es, grundlegende Zusammenhänge des modellierten Systems zu erforschen und zu verstehen. Sie dürfen und sollen jedoch nicht als „Abbild der Natur im Kleinen“ verstanden werden.⁴ Beide Modellierungsansätze sind, nach Maßgabe der Fragestellung, gleichermaßen berechtigt.

Die in dieser Arbeit entwickelten Modelle sind konzeptionell.

3.3.2 Funktionelles Neuron

Ein den in dieser Arbeit entwickelten konzeptionellen Modellen angemessenes Neuronenmodell ist das *Graded-Response-Modell*. Dieses nimmt vereinfachend an, daß die Aktivität einer Nervenzelle durch deren *Feuerrate* dargestellt wird, die im Modell als dimensionslose Größe $r \in \mathbb{R}_0^+$ geschrieben wird. Das Neuron selbst wird in

⁴Mehr als deren „Karikatur“ (Wissel, 1989).

3.3 Methoden der Modellierung

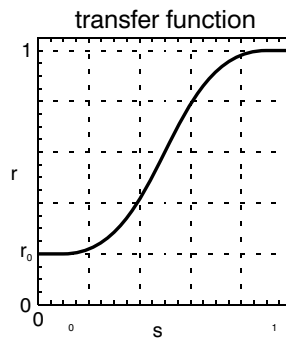


Abbildung 3.1: Beispiel für eine Übertragungsfunktion mit typisch sigmoider Form.

diesem Ratenmodell vollständig durch seine *Übertragungseigenschaften* charakterisiert: Die Aktivität des Neurons hängt lediglich von der geeignet akkumulierten Eingangsaktivität⁵ (Stimulation s), sowie von der *Übertragungsfunktion* Θ ab:

$$r = \Theta(s) \quad (3.1)$$

$$= \Theta(f_{acc}(\vec{r}_{in})) \quad (3.2)$$

mit

$$\Theta : \mathbb{R} \longrightarrow \mathbb{R}_0^+ \quad (3.3)$$

$$f_{acc} : \mathbb{R}^n \longrightarrow \mathbb{R} \quad (3.4)$$

$$f_{acc} : \vec{r}_{in} \longmapsto \sum_i w_i r_i \quad \text{oder} \quad (3.5)$$

$$\vec{r}_{in} \longmapsto \max_i \{w_i r_i\} \quad \text{oder} \quad (3.6)$$

$$\vec{r}_{in} \longmapsto \dots$$

wobei

$$\vec{r}_{in} = (r_i) : \text{präsynaptische Feuerraten}$$

$$(w_i) : \text{synaptische Gewichte}$$

Die Übertragungsfunktion Θ ist eine monotone nichtlineare Funktion von typischerweise sigmoider Form (Abbildung 3.1).

Die funktionelle Beschreibung der Nervenzelle als ratenbasiertes **Spike-Modelle** Ein-/Ausgabe-Element ist kompatibel mit der Großzahl existierender Neuronenmodelle. Naturgemäß lassen sich auf dieser Ebene keine Phänomene darstellen, für die eine explizite Modellierung von

⁵Beispielsweise Summation oder Maximumsbildung (Gleichungen 3.5, 3.6), vgl. Riesenhuber und Poggio (1999b).

Aktionspotentialen auf einer Zeitskala von Millisekunden notwendig ist. Dies wird jedoch für die betrachteten Anwendungen nicht nötig sein. Im allgemeinen wird davon ausgegangen, daß detaillierte dynamische Neuronenmodelle auf Basis von Aktionspotentialen unter der Bedingung langer Zeitkonstanten in das *Graded-Response-Modell* übergehen.⁶

3.3.3 Präzisierung des Begriffes „rezeptives Feld“

Der Begriff des *rezeptiven Feldes* ist als Werkzeug für die Charakterisierung der Codierungseigenschaften einzelner Neuronen von herausragender Bedeutung. Dies gilt auch und insbesondere für Nervenzellen im visuellen System. In seinem ursprünglichen Verständnis bezeichnet hier das rezeptive Feld (RF) den Bereich im Sehraum, von dem aus die Antwort einer Zelle durch Applizieren visueller Reize beeinflusst werden kann. Als ein in der aktiven Forschung weiträumig verwendetes Arbeitsmittel hat der Begriff in verschiedenen Zusammenhängen, und auch im Gebrauch und Verständnis verschiedener Forscher, unterschiedliche und teilweise inkompatible Erweiterungen erfahren. Dies ist für ein Werkzeug der Begriffsbildung nicht ungewöhnlich, kann jedoch allenthalben zu Mißverständnissen führen. Als Beispiel sei hier nur angedeutet, daß das rezeptive Feld in seiner ursprünglichen Definition ein zweidimensionales *Gebiet im Sehraum*, also eine Punktmenge bezeichnet. Es wird jedoch genausooft als Funktion einer zweidimensionalen Ortsvariablen aufgefaßt, wobei der Funktionswert die Art und Stärke des Einflusses beschreibt, der von Reizen am jeweiligen Ort auf die Nervenzelle ausgeübt wird. Eine moderne Definition des Begriffes bezieht außerdem eine zeitliche Entwicklung des rezeptiven Feldes ein. Es wird hier als die raumzeitliche Korrelation zwischen Sehreizen und neuronalem Signal betrachtet (RF-Cinematogramm, Eckhorn et al., 1993). Rezeptive Felder sind typischerweise *nicht* unabhängig vom verwendeten Testreiz, so daß verschiedene Felder für verschiedene Testreize gemessen werden. Wird, abweichend von der ursprünglichen Meßvorschrift, zur Bestimmung des rezeptiven Feldes mehr als ein einzelner visueller Testreiz verwendet, so können Kopplungseffekte auftreten. Sie werden im Begriff *nicht-klassisches rezeptives Feld* zusammengefaßt. Schließlich kann das rezeptive Feld nicht nur im Sehraum definiert werden, sondern ebenso als retinales rezeptives Feld, oder als rezeptives Feld in einem beliebigen topologisch geordneten kortikalen Areal – eine Variante, die in der funktionellen

⁶Der exakte Nachweis wurde jedoch meines Wissens bislang nicht geführt.

3.3 Methoden der Modellierung

Modellierung verbreitet ist. Folglich kann es sich nicht nur auf visuelle Reize, sondern ebenso auch auf Stimulation aufgrund neuronaler Aktivität, oder auf elektrische Stimulation (*Activation Zones*, Schanze et al., 2002) beziehen.

Da dem neuronalen rezeptiven Feld in dieser Arbeit eine zentrale Bedeutung zukommt, ist es wichtig, den Begriff im Rahmen der Modelle zu präzisieren. Es ist darauf hinzuweisen, daß die hier verwendete Terminologie mit der in verwandter Literatur gebrauchten nicht unbedingt übereinstimmen muß. Es erweist sich für meine Betrachtungen als sinnvoll, folgende Feldbegriffe gegeneinander abzugrenzen. Beide können als Spezifizierung des herkömmlichen Begriffes *rezeptives Feld* gesehen werden (vgl. auch Abbildung 3.2).⁷

Stimulationsfeld (stimulierendes Feld, SF) Das stimulierende Feld $SF(\vec{x})$ quantifiziert den stimulierenden Einfluß, den ein geeigneter, einzelner visueller Reiz am Ort \vec{x} im Sehraum oder auf der Retina auf das betreffende Neuron ausübt. Das Stimulationsfeld nimmt positive Werte an, wo dieser Einfluß exzitatorisch, und negative Werte, wo er inhibitorisch ist. Es ist unabhängig von der Antwort des betreffenden Neurons, insbesondere davon, ob das Neuron durch die Stimulation aktiviert, also über die Antwortschwelle hinaus erregt wird.

Neurophysiologisch beschreibt das Stimulationsfeld die Einflüsse auf das Membranpotential des betreffenden Neurons. Es ist experimentell als Einzelsignal durch intrazelluläre Ableitung des Membranpotentials zugänglich, sowie als extrazellulär abgeleitetes Massensignal in Form des lokalen Feldpotentials (LFP).

Aktivierungsfeld (aktivierendes Feld, AF) Das aktivierende Feld $AF(\vec{x})$ quantifiziert den aktivierenden Einfluß, den ein geeigneter, einzelner visueller Reiz am Ort \vec{x} im Sehraum oder auf der Retina auf das betreffende Neuron ausübt. Es bezieht sich also auf die *Antwort* des Neurons als Reaktion auf die Stimulation.

Das Aktivierungsfeld ist experimentell als Einzelsignal durch intrazelluläre Ableitung von Aktionspotentialen zugänglich, sowie als extrazellulär abgeleitetes Massensignal in Form der *Multi Unit Activity* (MUA).

⁷Wir beschränken uns auf den klassischen Feldbegriff, also auf die Beschreibung des Einflusses einzelner visueller Testreize.

Zwischen dem Stimulationsfeld SF und dem Aktivierungsfeld AF besteht im verwendeten *Graded-Response*-Modell definitionsgemäß der Zusammenhang

$$AF \equiv \Theta(SF) \quad (3.7)$$

Beide Begriffe werden als Funktionen einer ein- oder zweidimensionalen Ortsvariablen verstanden. Sie sind nicht explizit zeitabhängig. Wo statt des Begriffes *Feld* der Begriff *Gebiet* steht, (etwa „aktivierendes Gebiet“), sei stets die Teilmenge des Definitionsbereiches gemeint, auf dem die entsprechende Funktion einen Mindestwert ϵ übersteigt:

$$AF \equiv \Theta(SF) \quad \text{aktivierendes Feld} \quad (3.8)$$

$$\Omega_{AF} := \{x \mid AF(x) > \epsilon\} \quad \text{aktivierendes Gebiet} \quad (3.9)$$

Den Ausdruck *rezeptives Feld* werde ich aufgrund seines teils widersprüchlichen Gebrauchs nur dort verwenden, wo eine genauere Unterscheidung nicht notwendig ist.

3.4 Situativ gesteuerte Anpassung aktivierender Felder

In Abschnitt 3.1 habe ich die Annahme formuliert, daß Informationen über die physikalisch-körperliche Situation des Individuums in die neuronale Verarbeitung eingehen, und daß diese im visuellen System zur Generierung geometrischer Invarianzen herangezogen werden. Es ist nun das Ziel, diese Steuerung der neuronalen Verarbeitung durch externe Bedingungen zu modellieren.

Die offenbar zugrundeliegenden Vorgänge manifestieren sich experimentell zugänglich auf der Ebene von Ratenmodulationen einzelner Nervenzellen (Abschnitt 3.2). Im Rahmen eines konzeptionellen Modells dient uns daher das ratenbasierte *Graded-Response*-Modell als geeignete Beschreibung der neuronalen Codierungseigenschaften, soweit sie diese Effekte betreffen (Abschnitt 3.3.2). Wir suchen nach einer Methode, die Codierungseigenschaften eines Neurons in Abhängigkeit einer äußeren Bedingung systematisch zu variieren. Dies soll letztenendes ermöglichen, über das gezielte Einstellen bestimmter Neuronengruppen das visuelle Objekterkennungssystem auf aktuelle physikalisch-körperliche Gegebenheiten wie Sehentfernung oder Blickwinkel zu justieren. Unser Modell muß insbesondere die systematische Veränderung der *aktivierenden Felder*⁸

⁸Zur Terminologie siehe Abschnitt 3.3.3.

3.4 Situativ gesteuerte Anpassung aktivierender Felder

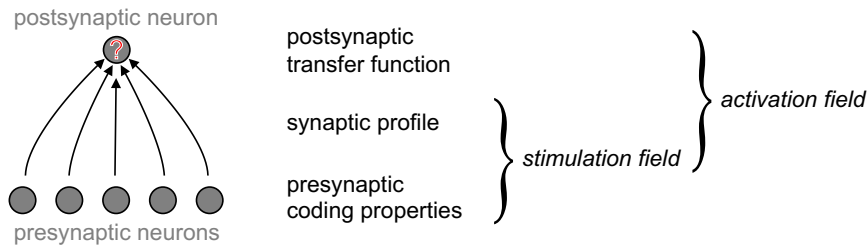


Abbildung 3.2: Externe Angriffspunkte für Feldveränderungen des postsynaptischen Neurons. In das *Stimulationsfeld* gehen die Codierungseigenschaften aller präsynaptischen Neuronen ein, sowie das Profil ihrer synaptischen Verbindungen zum Zielneuron. Das *Aktivierungsfeld* umfasst zusätzlich die Aktivierung des Zielneurons, welche durch dessen Übertragungsfunktion Θ gegeben ist.

ermöglichen, da hierüber grundlegende geometrische Invarianzen realisiert werden können.

Doch wie gelingt es, ein neuronales Aktivierungsfeld zu ändern, obwohl dessen räumliche Struktur durch die synaptischen Verbindungen der Zelle festgelegt zu sein scheint? Zwar haben Olshausen, Anderson und van Essen mit ihrem Modell des dynamischen *Routing* ein Verfahren vorgestellt, das darauf beruht, große Mengen synaptischer Gewichte in konsistenter Weise zu verändern, doch scheint der Aufwand der Koordination zu hoch, um ein akkurates Einstellen eines ganzen neuronalen Systems zu gewährleisten (Anderson und van Essen, 1987; Olshausen et al., 1993, vgl. Abschnitt 2.2).

**Invarianz durch
Veränderung
aktivierender Felder**

Tatsächlich spielen jedoch neben den konkreten synaptischen Gewichten auch andere neuronale Parameter eine entscheidende Rolle für die Ausformung des aktivierenden Feldes einer Nervenzelle. Salinas und Abbott etwa stellten 1997 ein Verfahren vor, mit dem durch systematische Variation der Sensitivität eines Neurons für afferente Signalbestandteile eine Verschiebung des Aktivierungsfeldes realisiert werden konnte – ohne jedoch einen Mechanismus für diese Sensitivitätsänderung auf Ebene des Neurons vorzuschlagen (Salinas und Abbott, 1997). Ein externer Prozeß könnte verschiedene funktionelle Bestandteile des Neurons beeinflussen. Abbildung 3.2 zeigt schematisch, welche Angriffspunkte für einen äußeren Steuerprozeß denkbar sind, um auf das aktivierende Feld eines Neurons einzuwirken.⁹ Die Unterscheidung dreier Wirkungsorte erlaubt die strukturelle Einordnung möglicher Modellierungsansätze:

**Synapsen sind nicht
alles**

⁹Insbesondere auf molekularer Ebene ist natürlich eine sehr viel größere Zahl möglicher Wirkungsorte und -mechanismen vorstellbar. Zur gewählten Ebene der Beschreibung siehe Abschnitt 3.3.2.

Modelle lassen sich nach Angriffspunkten ordnen

Synaptisches Profil des Neurons Hiermit ist die räumliche Struktur und Gewichtsverteilung der Synapsen gemeint, die Signale von präsynaptischen Neuronen an das betreffende Neuron übertragen. Der Begriff bezieht sich im einfachsten Fall auf die tatsächlich anatomisch vorliegende synaptische Struktur. Betrachten wir jedoch das *Routing*-Verfahren von Olshausen et al. (1993), so ist nur ein kleiner Teil der anatomisch vorliegenden Verbindungen für eine bestimmte Konfiguration auch wirklich aktiv. Das *effektive* synaptische Profil kann also deutlich kleiner sein als die Menge anatomischer Verbindungen.

Die Notwendigkeit, große Mengen von Synapsen konzertiert umzuschalten, macht hier ansetzende Verfahren zur Feldveränderung in ihrer neuronalen Umsetzung sehr aufwendig. Hierzu wäre ein Mechanismus vonnöten, der zeitlich exakt und sehr lokalisiert an einer Vielzahl einzelner, räumlich weit verteilter Orte wirken kann. Abhängig vom Grad der Divergenz und Konvergenz der synaptischen Verbindungen kann die Zahl der einzeln anzusteuernenden Synapsen bis zu quadratisch von der Zahl der Neuronen abhängen.

Übertragungsfunktion Θ des Neurons Durch die Beeinflussung der internen *Übertragungseigenschaften* einer Nervenzelle lassen sich sehr viel einfacher Änderungen im räumlichen Codierungsverhalten erreichen. Neuronen agieren als Schwellenelemente (oder allgemeiner: als nichtlineare Elemente). Daher läßt sich durch die Anpassung der Übertragungsfunktion nicht nur ihre Antwortstärke, sondern gerade auch ihr effektiver räumlicher Einzugsbereich, also das Aktivierungsfeld, steuern. Diese anschaulich als „Eisbergeffekt“ zu beschreibende Eigenschaft stellt das nachfolgende Kapitel 4 vor. Es demonstriert neben der Leistungsfähigkeit des Ansatzes auch seine konzeptuell einfache Anwendbarkeit im Modell.

In seiner biologischen Umsetzung würde ein hier ansetzendes Verfahren die Modulation neuronaler Übertragungsvorgänge wie der raum-zeitlichen Integration postsynaptischer Ströme einschließen. Die effektive Form der Übertragungsfunktion ist zudem von der exzitatorischen und inhibitorischen Hintergrundaktivität abhängig, der das Neuron ausgesetzt ist (Abbott et al., 1997; Chance et al., 2002).

Codierungseigenschaften der präsynaptischen Neuronen Für alle Nervenzellen gilt, daß sie die Signale der ihnen vorgeschalteten Zellen weiterverarbeiten. Daher werden stets die Eigen-

schaften der präsynaptischen Zellen das Verhalten der postsynaptischen maßgeblich beeinflussen: Ihre Codierungseigenschaften werden dem postsynaptischen Neuron aufgeprägt. Man denke hier beispielsweise an den auf Hubel und Wiesel (1962) zurückgehenden Vorschlag, daß die phasenunabhängigen Antworten von *Complex*-Zellen durch eine konvergente Verschaltung mehrerer *Simple*-Zellen gleicher Vorzugsorientierung, doch mit verschiedenen bevorzugten Kantenorten zustandekommen (Abbildung 2.1, Seite 11). Die Eigenschaft, Detektor für Kontrastkanten einer bestimmten Orientierung zu sein, geht hier von den *Simple*- auf die *Complex*-Zellen über. Sie wird gewissermaßen „vererbt“. Das impliziert, daß eine Einflußnahme auf Codierungseigenschaften präsynaptischer Neuronen unmittelbare Auswirkung auf das Codierverhalten der postsynaptischen Nervenzellen hat. Dies kann in der Modellierung ausgenutzt werden: Im einfachsten Fall können durch Inhibition, oder durch die Kontrolle der *Verstärkung* (engl. *gain*) vorgeschalteter Neuronen, gezielt Bestandteile eines rezeptiven Feldes ein- oder ausgeblendet werden. Im allgemeinen stehen jedoch alle denkbaren Mechanismen zur Beeinflussung der präsynaptischen Zellen zur Auswahl, inklusive der zwei zuvor beschriebenen.

Die Kapitel 5 und 6 stellen dieses Konzept ausführlich vor und demonstrieren seine Anwendung in zwei Modellen zu aktuellen Fragestellungen der Forschung am visuellen System.

3.5 Der externe Steuerparameter

Den in den folgenden Kapiteln formulierten Modellen zur geometrischen visuellen Invarianzbildung ist gemein, daß zur systematischen Steuerung der Eigenschaften eines kompletten neuronalen Teilsystems ein *einzelner* äußerer Parameter herangezogen wird. Warum ist es sinnvoll, von einem solch einfachen funktionellen Zusammenhang auszugehen? Hierfür sprechen verschiedene Erwägungen:

Warum ein einzelner externer Parameter?

Das Ziel der Steuerung soll es sein, das visuelle System den augenblicklich vorliegenden Sehbedingungen anzupassen, um so eine verlässliche und eindeutige Form- und Objekterkennung zu ermöglichen (vgl. Abschnitt 3.1). Der auf die Netzhaut projizierte optische Reiz ist das Resultat eines Abbildungsvorganges, der vollständig durch die physikalischen Gesetze der Strahlenoptik festgelegt ist. So beschreibt etwa ein einfaches *Abbildungsgesetz* den Zusammenhang

Kapitel 3 Situative Steuerung als Invarianz-Strategie

zwischen der Größe eines in der äußeren Welt vorhandenen Objektes, und seines durch die Linse unseres Auges auf die Retina projizierten Bildes:

$$\frac{B}{G} = \frac{b}{g} \quad (3.10)$$

Dabei stehen G und B für die Gegenstands- und Bildgröße, g und b für die Entfernungen von Gegenstand und Bild, gemessen vom Linsenmittelpunkt. Zwar gilt diese Relation exakt nur für hinreichend dünne Linsen und ebene Projektionsflächen, doch soll hier der prinzipielle Zusammenhang interessieren. Da die Bildweite b im Falle des Auges fix ist, folgt für einen festen Gegenstand

$$B \sim \frac{1}{g} \quad (3.11)$$

Je weiter entfernt vom Auge des Betrachters ein Gegenstand ist, um so kleiner ist sein Netzhautbild.

Kongruenz der Wahrnehmungen

Dieser einfache physikalische Zusammenhang gilt zu jeder Zeit und für alle Gegenstände. Hingegen ist die Unveränderlichkeit der (beispielsweise ertasteten) physikalischen Größe eines starren Körpers für den Betrachter eine *Konstante der Erfahrung*, die der variierenden Größe des Abbildes auf der Retina gegenübersteht. Unter der Annahme, daß die Gegenstandsweite g dem neuronalen System als die momentane Fixationsentfernung verfügbar ist (Kapitel 3.2), ist davon auszugehen, daß der immergleiche Zusammenhang von Gegenstandsgröße, Bildgröße und Gegenstandsweite während der Entwicklung des Individuums erlernt wird. Die Fähigkeit zur Bewegung und aktiven Manipulation von Gegenständen der Umgebung erlaubt die Erforschung in einer Wahrnehmungs-Handlungsschleife (engl. *action-perception-cycle*). Mehr noch, mit hoher Wahrscheinlichkeit wird das visuelle System die einfache Gesetzmäßigkeit ausnutzen, um die variierende Größe des Netzhautbildes neuronal zu kompensieren, und so die subjektive Wahrnehmung des Betrachters der haptisch erfahrenen Konstanz der Objektgröße anzugleichen.

Funktionen einfacher Parameter

Gleiches gilt für die durch Augenbewegungen verursachte retinale Bewegung der Projektion, welche der erfahrenen Stationarität der äußeren Welt widerspricht. Wir werden also erwarten, daß die abbildungsbedingten Reizvariationen im Rahmen der visuellen Objekterkennung ganz oder teilweise vom Sehsystem kompensiert werden. Die hierzu notwendigen geometrischen Transformationen hängen in einfacher, deterministischer Weise von einzelnen Parametern

3.5 Der externe Steuerparameter

wie der Fixationsentfernung ab. Entsprechend müssen große Teile des neuronalen Netzwerkes – möglicherweise sogar das ganze visuelle System – konsistent entsprechend einzelner physikalischer Größen eingestellt werden.

Die wahrgenommene physikalische Größe, die wahrgenommene dreidimensionale Gestalt und Entfernung eines Gegenstandes beruhen offenbar auf ein und derselben Schätzung des Objekt-Betrachter-Abstandes (van Damme und Brenner, 1997; Brenner und van Damme, 1999). Liegen fehlerhafte Werte für diesen Parameter vor, oder erhält unser Sehsystem aus bestimmten Gründen inkonsistente Rückmeldungen über die Sehsituation, dann führt dies zu Fehlwahrnehmungen der Form und Größe von Objekten (Johnston, 1991; Landy et al., 1995; Brenner und van Damme, 1999). Zur Illustration möge das in Abbildung 3.3 dargestellte Selbstexperiment dienen. Hier dient der Vergenzwinkel, also der von den Achsen der Augen eingeschlossene Winkel, als eine Quelle für die Schätzung der Fixationsentfernung (Brenner und van Damme, 1998, 1999; Brenner und Smeets, 2000). Wird die Abbildung in *gekreuzter* oder *ungekreuzter Fusion* betrachtet, so resultiert aus der geänderten Vergenz das subjektive Empfinden kleinerer bzw. größerer Gegenstände als bei normaler Betrachtung.¹⁰ Die Tatsache, daß die geänderte Wahrnehmung großflächig im ganzen Sehfeld auftritt, und nicht etwa auf Teile der Szene oder des einzelnen Objektes beschränkt ist, kann als Hinweis gewertet werden, daß hier große Teile des Sehsystems in konsistenter Weise durch einen einzelnen Parameter angesteuert und eingestellt werden.

Fehlwahrnehmungen

Letztenendes weisen Messungen im visuellen System einen systematischen Zusammenhang der Feuerraten vieler Zellen mit einzelnen abbildungsrelevanten Größen nach (Abschnitt 3.2). Zwar ist daraus nicht zu entnehmen, daß die entsprechenden Parameter in isolierter Form neuronal vorliegen – die ihnen gemäße Anpassung von Teilen des Nervensystems ist hingegen erwiesen.

¹⁰Den gleichen Effekt kann der Leser auch beim Betrachten der beliebten *Magisches-Auge*-Bilder in gekreuzter und ungekreuzter Sichtweise beobachten.

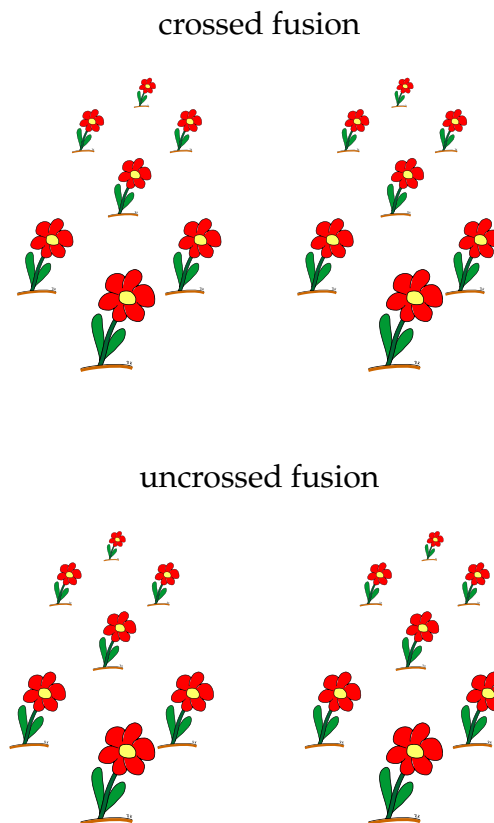


Abbildung 3.3: Einfluß des Vergenzwinkels auf die Größenwahrnehmung. Dargestellt ist ein sich dreidimensional in die Tiefe erstreckendes Blumenfeld. Die obere Abbildung ist in gekreuzter Fusion zu betrachten, d.h., die Augenachsen sind wie beim Schielen stärker zu kreuzen, bis sich die beiden Bilder in der Mitte überlagern. Die untere Abbildung ist in ungekreuzter Fusion zu betrachten, d.h., die Augenachsen sind nach außen voneinander zu entfernen, bis sich die beiden Bilder in der Mitte überlagern. In ungekreuzter Fusion scheinen die Objekte subjektiv größer zu sein als bei normaler Betrachtung. In gekreuzter Fusion erscheinen sie kleiner.

Kapitel 4

Der Schwelleneffekt als Beispiel systematischer Feldanpassung

Ziel

Die Beeinflussung der Codierungseigenschaften eines einzelnen Neurons durch einen von außen angreifenden Prozeß demonstriere ich hier beispielhaft am Schwelleneffekt. Durch Variation von Antwortschwelle und Verstärkung sind neuronale Aktivierungsgebiete größenskalierbar. Mit Hilfe dieses Prinzips läßt sich ein Spektrum hochinteressanter Effekte erzielen, darunter die Auflösungskontrolle und die Durchstimmung von Raumfrequenzkanälen im visuellen System. Zur Generierung visueller Größeninvarianz erweist sich der Schwelleneffekt als ungeeignet.

In Kapitel 3 habe ich die situative Steuerung als Prinzip des Einflusses äußerer Bedingungen auf die interne Verarbeitung im visuellen System formuliert. Auf der Ebene des einzelnen Neurons muß dieser Einfluß eine Veränderung der Codierungseigenschaften zur Folge haben. Im Zusammenhang mit den geometrischen Invarianzen der visuellen Wahrnehmung liegt besondere Aufmerksamkeit auf der Veränderung der räumlichen Codierung, welche durch das aktivierende Feld der Zelle charakterisiert ist (Abschnitt 3.3.3).

Eine Möglichkeit zur Einflußnahme bietet die Manipulation der internen Übertragungseigenschaften einer Nervenzelle. Wir können hierdurch deren aktivierendes Feld gezielt verändern. Der unter 3.4 erwähnte „Eisbergeffekt“, hervorgerufen durch eine Variation der neuronalen Antwortschwelle, kann insbesondere zur Größenskalie-

Kapitel 4 Der Schwelleneffekt als Beispiel systematischer Feldanpassung

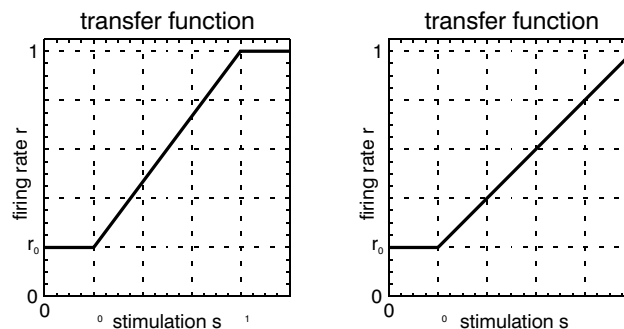


Abbildung 4.1: Einfache Annahmen für eine Übertragungsfunktion im *Graded-Response*-Beschreibungsmodell für die neuronale Aktivität. **Links:** Stückweise lineare Übertragungsfunktion, charakterisiert durch die Parameter *Spontanrate*, *Antwortschwelle*, *Verstärkung* und *Sättigungsschwelle*. **Rechts:** Vereinfachung zu einer zweiteiligen Funktion, charakterisiert durch *Antwortschwelle* und *Verstärkung*.

zung des *Aktivierungsgebietes*¹ eines Neurons eingesetzt werden. Dieser *Schwelleneffekt* soll hier zunächst an einer einzelnen Nervenzelle untersucht werden.

4.1 Übertragungsfunktion

Als Übertragungsfunktion wird im Sinne der konzeptionellen Modellierung (Abschnitt 3.3.1) die einfachste Funktion gewählt, die im *Graded-Response*-Modell noch das typische Verhalten von Neuronen wiedergibt; diese sollte demnach einen sigmoidalen Zusammenhang von *Stimulation* s und *Feuerrate* r darstellen. Abbildung 4.1 (links) zeigt die verwendete Übertragungsfunktion Θ : Die Antwort des Neurons hängt im mittleren Erregungsbereich linear von seiner Stimulation ab. Unterhalb der Antwortschwelle ϑ_0 liefert das Neuron stets die Spontanrate r_0 .² Oberhalb einer Sättigungsschwelle ϑ_1 antwortet die Zelle mit der maximalen Rate 1.

Die konkrete Wahl der Spontanrate ist für das Prinzip des Schwelleneffekts nicht von funktioneller Bedeutung. Gehen wir ferner davon aus, daß das Neuron unterhalb seiner Sättigungsschwelle

Schwelle und Verstärkung

¹Wir verstehen unter dem Begriff *Gebiet* einen Bereich des Sehraumes, ohne dabei Aussagen über die dort bevorzugt codierten visuellen Merkmale zu machen. Das *Aktivierungsgebiet* ist der Bereich im Sehraum, von dem aus durch geeignete Stimulation eine *Aktivierung*, also eine Antwort der Zelle hervorgerufen werden kann. Vgl. hierzu Abschnitt 3.3.3.

²Ein Absinken der Rate unter diesen Wert, wie man es bei der Inhibition echter Nervenzellen beobachtet, wird von dem Modell nicht reproduziert. Es ist für den Schwelleneffekt von geringer funktioneller Bedeutung.

4.2 Symmetrisches Stimulationsfeld

betrieben wird ($s < \vartheta_1$), so wird das Übertragungsverhalten der Zelle wesentlich durch zwei Parameter bestimmt. Es sind dies

- die **Antwortschwelle** ϑ_0
- die **Verstärkung** (*Gain*) im linearen Antwortbereich, $\frac{\Delta r}{\Delta s}$.

Diese vereinfachte Form der Übertragungsfunktion ist in Abbildung 4.1 (rechts) dargestellt.

Ein hypothetischer Prozeß, der von außen an ein solches Neuron angreift, kann allein durch die Veränderung dieser zwei Parameter dessen aktivierendes Feld innerhalb eines gegebenen Rahmens einstellen. Die gezielte Beeinflussung sowohl von Antwortschwelle wie auch der neuronalen Verstärkung läßt sich in realistischen Netzwerken durch inhibitorische und exzitatorische Hintergrundaktivität realisieren (Chance et al., 2002).

4.2 Symmetrisches Stimulationsfeld

Betrachten wir ein Neuron mit einfachem stimulierenden Feld.³ Wir wollen uns hier auf den eindimensionalen Fall beschränken; alle Ortsfunktionen können jedoch als Schnitte durch entsprechende zweidimensionale Funktionen gedacht werden. Das stimulierende Feld $SF(x)$ sei wie in Abbildung 4.2 als Gaußfunktion gegeben und fest. Als Beispiel für Neuronen im visuellen System mit entsprechenden Feldern denke man etwa an Ganglienzellen ohne *Surround*-Profil im lateralen Geniculatum (*Bistratified Ganglion Cells*, vgl. Rodieck, 1998), oder im allgemeinen an jede *Complex*-artige Zelle im Sinne des Hubel-und-Wiesel-Modells (Abschnitt 2.1.1), die mit ihrem bevorzugten Reiz stimuliert wird (s.u.).

Das *aktivierende Feld* $AF(x)$, definiert als

$$AF(x) = \Theta(SF(x)) \quad (4.1)$$

entspricht in seiner Form dem Stimulationsfeld, abgesehen von der unteren Begrenzung durch die Spontanrate r_0 , siehe Abbildung 4.3 rechte Seite. Durch Anpassung der Antwortschwelle ϑ_0 des Neurons kann nun der durch Θ übertragene Teil des stimulierenden Feldes eingestellt werden: Die „Spitze des Eisbergs“ ragt einmal mehr und einmal weniger über die „Wasserlinie“ der Spontanaktivität hinaus. Der Bereich auf der Ortsachse, in dem die Antwort des Neurons auf einen Testreiz von der spontanen Aktivität verschieden

³zur Terminologie vgl. Abschnitt 3.3.3.

4.2 Symmetrisches Stimulationsfeld

ist, wird beim Anheben der Antwortschwelle sukzessiv kleiner. Er scheint symmetrisch um den Mittelpunkt des stimulierenden Feldes zu schrumpfen.

Es ist wichtig, sich die Natur dieser Feldverkleinerung vor Augen zu führen: Da der Schwelleneffekt ein neuronaler Mechanismus ist, wird auf diese Weise der *Signaleinzugsbereich* der jeweiligen Zelle einstellbar. Durch den Schwelleneffekt *unverändert* hingegen die Codierungseigenschaften der präsynaptischen Zellen. Man denke hier beispielsweise an visuelle *Complex*-Zellen und deren Äquivalente bei *Complex*-Bildung in verschiedenen Kortexarealen (vgl. Abbildung 2.1, S. 11). Käme der Schwelleneffekt bei einer *Complex*-Zelle nach Hubel und Wiesel (1962) zur Anwendung, so bliebe deren Eigenschaft erhalten, selektiv auf Kontrastkanten einer bevorzugten Raumfrequenz und Orientierung zu antworten, denn diese Eigenschaft wird von den präsynaptischen Neuronen unverändert „vererbt“. Variiert würde hingegen die Ausdehnung des *klassischen rezeptiven Feldes* der *Complex*-Zelle, also des Gebietes im Sehraum, von dem aus einzelne, bevorzugte Testreize die Zelle aktivieren können.⁴

**Integrationsbereich
und bevorzugter Reiz**

Wir werden in Abschnitt 4.3 sehen, daß sich mit dem gleichen Verfahren in einfacher Weise auch eine Skalierung des bevorzugten Reizes selbst realisieren läßt. Beide Effekte, Anpassung des Integrationsbereiches und Anpassung des bevorzugten Reizes, haben sinnvolle Anwendungen für die Modellierung des visuellen Systems, sind jedoch konzeptionell völlig verschieden. Die Tatsache, daß sie gleichermaßen mit dem Begriff „Größenänderung des rezeptiven Feldes“ belegt werden, trägt zur Erschwerung der Diskussion über invariante Reizverarbeitung bei.

Bezogen auf ein Neuron im visuellen System bedeutet eine Größenänderung des *Aktivierungsgebietes*, daß der Bereich im Sehfeld, von dem aus das betreffende Neuron durch geeignete visuelle Reizung aktiviert werden kann, wächst oder schrumpft. Einen solchen

**Visuelle Auflösung
und Schwelleneffekt**

⁴Führen *Complex*-Zellen eine lineare Summation über die afferenten Signale aus, so können sie durch ausgedehnte Texturreize (etwa Gitterreize im Falle einer *Complex*-Zelle nach Hubel und Wiesel (1962)) besser aktiviert werden als durch einzelne lokalisierte Testreize. Der Begriff des klassischen rezeptiven Feldes als Ortsfunktion, und damit auch der Schwelleneffekt, verlieren hier ihre Bedeutung. Mit zunehmender Komplexität der bevorzugten Merkmale wird eine Texturbildung allerdings erschwert. (Man versuche, eine Textur anzugeben, welche die auf S. 11 angedeutete *Complex*-Zelle für Größe optimal stimuliert!) Führt eine *Complex*-Zelle die von Riesenhuber und Poggio (1999b) vorgeschlagene Maximumsbildung aus (Gleichung 3.6), so wird die Zellantwort in jedem Fall vom maximal stimulierenden Einzelreiz bestimmt.

Kapitel 4 Der Schwelleneffekt als Beispiel systematischer Feldanpassung

Effekt beschreiben Wörgötter et al. (1998) für die Form aktivierender Felder im primären visuellen Kortex anästhesierter Katzen. Er wird dort mit einer möglichen dynamischen Kontrolle der visuellen Auflösung in Zusammenhang gebracht. Die „Restrukturierung“ der Aktivierungsfelder von Einzelzellen verlief schnell (im Bereich weniger hundert Millisekunden) und korreliert zum Gesamtzustand des Gehirns, wie er durch das Elektroenzephalogramm (EEG) charakterisiert wird. Die Ergebnisse weisen darauf hin, daß sich neuronale Codierungseigenschaften im visuellen Kortex der Katze konsistent und in Abhängigkeit makroskopisch meßbarer Zustandsgrößen ändern. Der Schwelleneffekt stellt einen möglichen Mechanismus hierfür dar. Eine spätere Arbeit (Suder et al., 2000) schlägt ein solches „Eisbergprinzip“ als ein Erklärungsmodell für die experimentellen Ergebnisse vor.

Kompensatorische Variation der Verstärkung

Ein Blick auf Abbildung 4.3 macht die Bedeutung des oft verwendeten Begriffes „Eisbergeffekt“ unmittelbar deutlich. Tatsächlich ist der Mechanismus in dieser Weise jedoch kaum geeignet, eine effektive Größenänderung des aktivierenden Gebietes herbeizuführen, da zusammen mit dessen Verkleinerung auch die maximal übertragene Signalamplitude absinkt. Ein geeigneter Stimulus, der im Zentrum des aktivierenden Gebietes anliegt, wird in den drei abgebildeten Fällen unterschiedlich starke Zellantworten hervorrufen. Eine Schrumpfung des Aktivierungsgebietes der Zelle geht einher mit einer insgesamt erniedrigten Antwort. Um lediglich das aktivierende Gebiet des Neurons zu verkleinern, ohne dabei die Effektivität eines zentral anliegenden Reizes zu mindern, kann parallel zur Verschiebung der Antwortschwelle ϑ_0 auch die Verstärkung im linearen Antwortbereich angepaßt werden. Diese entspricht der Steigung $\frac{\Delta r}{\Delta s}$ der Übertragungsfunktion Θ im Bereich oberhalb ϑ_0 . Die Anpassung kann dabei beispielsweise so geschehen, daß zentral im Feld anliegende Reize stets die gleiche neuronale Antwort hervorrufen.⁵ Abbildung 4.4 stellt dar, wie die Ergebnisse in diesem Fall aussehen: Wie zuvor schrumpft der Bereich auf der Ortsachse, in dem die Antwort des Neurons von der spontanen Aktivität verschieden ist, symmetrisch zusammen. Gleichzeitig bleibt jedoch das Maximum des Aktivierungsfeldes unverändert. Lediglich die Ausdehnung des räumlichen Übertragungsbereiches ändert sich, nicht aber die Stärke der Antwort auf einen optimal positionierten Reiz.

Das Aktivierungsgebiet eines visuellen Neurons würde in diesem

⁵Chance et al. (2002) demonstrieren, daß die Verstärkung eines Neurons vom Gleichgewicht exzitatorischer und inhibitorischer Hintergrundaktivität abhängt und in weiten Bereichen einstellbar ist.

4.2 Symmetrisches Stimulationsfeld

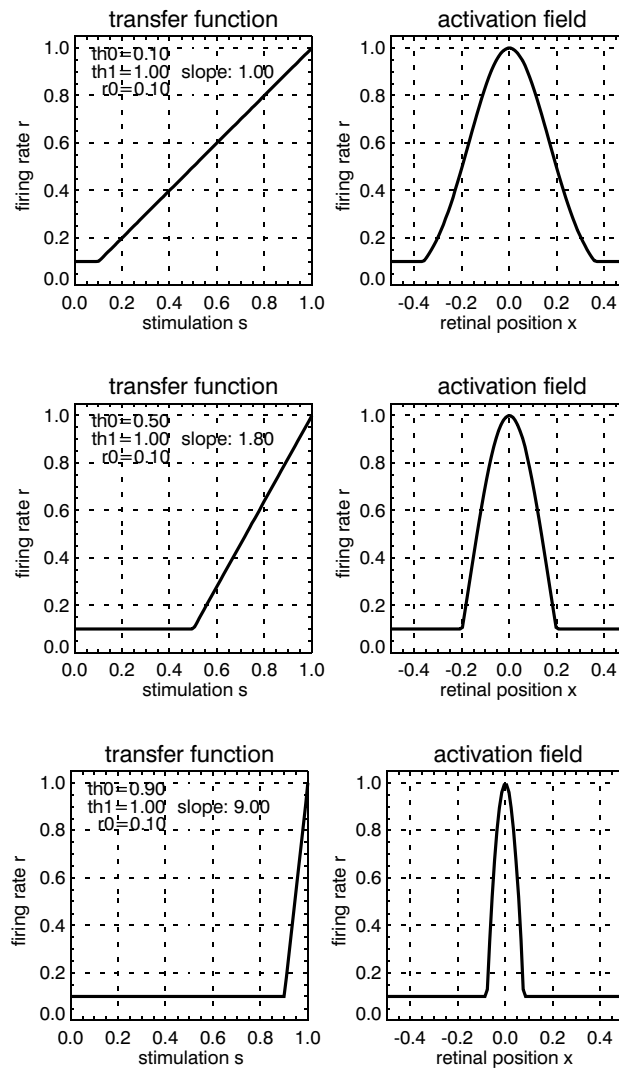


Abbildung 4.4: Der erweiterte Schwelleneffekt beim symmetrischen Stimulationsfeld. Parallel zur Antwortschwelle wird die Verstärkung im Antwortbereich angehoben.

Kapitel 4 Der Schwelleneffekt als Beispiel systematischer Feldanpassung

Fall im Sehraum zusammenschrumpfen, wobei die Zelle durch visuelle Testreize nach wie vor optimal aktiviert werden kann, solange diese nur innerhalb des richtigen Gebietes zu liegen kommen. Der Schwelleneffekt ist damit ein Konzept, das hervorragend zur Modellierung einer Auflösungsanpassung im visuellen System geeignet ist, wie sie von Wörgötter et al. (1998) und Suder et al. (2000) angenommen wird.

4.3 Raumfrequenzabstimmung mit Hilfe des Schwelleneffektes

„Restrukturierung“
von Simple-Zell-RF

Wörgötter et al. haben die Form der Aktivierungsfelder über die raumzeitliche Korrelation der neuronalen Antworten mit visuellen Testreizen bestimmt (RF-Cinematographie, vgl. Eckhorn et al., 1993). Sie waren daher in der Lage, den „Restrukturierungsprozeß“ bei einzelnen Subfeldern von Simple-Zellen zu untersuchen: On- und Off-Subfelder einer Simple-Zelle veränderten ihre Ausdehnung jeweils in gleichem Maße. Der Schwelleneffekt erlaubt auch hierfür eine einfache mögliche Erklärung im *Difference-of-Gaussians*-Modell für das *Center/Surround*-Profil visueller Aktivierungsfelder. Hierbei wird davon ausgegangen, daß die typische mexikanerhutartige Gestalt der Felder durch die subtraktive Überlagerung zweier gaußförmiger Profile zustandekommt. Dies würde einer neuronalen Verschaltung entsprechen, bei der die Signale einer exzitatorischen und einer inhibitorischen Zelle mit gaußförmigen Feldern verschiedener Breite auf ein und dieselbe Zelle konvergieren (Abbildung 4.5).⁶ Kommt nun bei beiden präsynaptischen Zellen der Schwelleneffekt zum Einsatz, so ergeben sich für das postsynaptische Neuron aktivierende Felder, wie sie in Abbildung 4.6 dargestellt sind.

Frequenzdurchstim-
mung

Auf das *Difference-of-Gaussians*-Modell angewandt liefert der Schwelleneffekt somit ein denkbar einfaches Modell zur räumlichen Frequenzdurchstimmung visueller thalamischer, Simple- oder Complex-Zellen. Hierzu ist keinerlei Eingriff in die synaptische Struktur notwendig, allein eine lokale Anpassung der Übertragungsfunktionen der beteiligten Neuronen ist ausreichend. Eine derartige Abstimmung der räumlichen Frequenzkanäle könnte eine entscheidenden

⁶In der zweidimensionalen Erweiterung führt dies zunächst zu konzentrischen *Center/Surround*-Profilen, die dann beispielsweise über geeignete Vorwärtsverschaltung das linear ausgedehnte Simple-Zell-Feld erzeugen können (vgl. etwa Hubel und Wiesel, 1962). Durch die Überlagerung räumlich gegeneinander verschobener zweidimensionaler gaußförmiger Felder lassen sich außerdem Simple-Zell-artige Felder direkt erzeugen.

4.3 Raumfrequenzabstimmung mit Hilfe des Schwelleneffektes

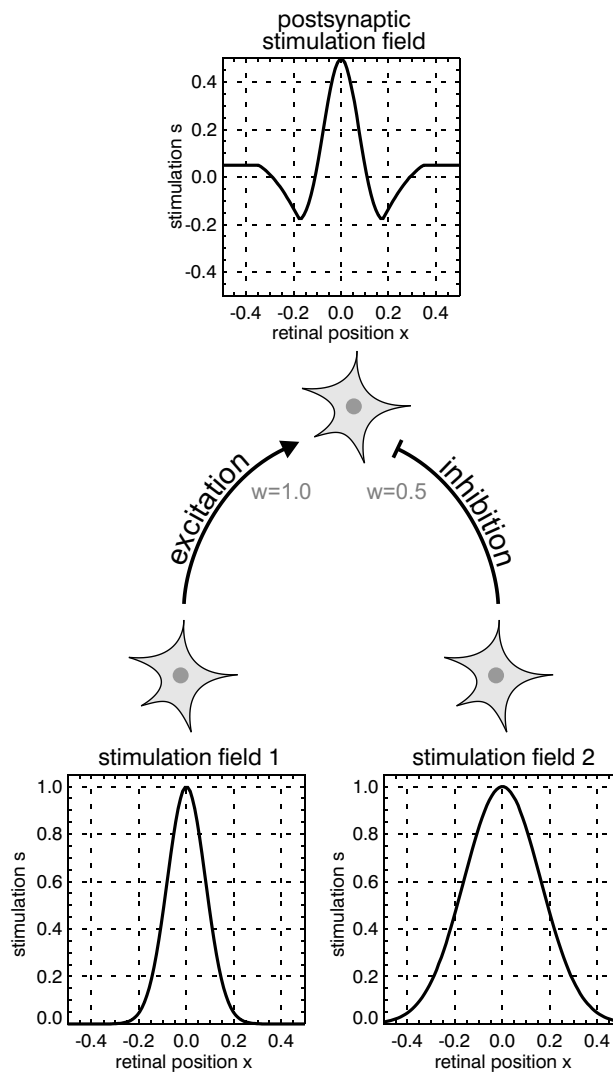


Abbildung 4.5: *Difference-of-Gaussians-Modell zur Erzeugung eines Center/Surround-Profiles. Zwei präsynaptische Neuronen mit gaußförmigen Stimulationsfeldern werden exzitatorisch und inhibitorisch auf ein Zielneuron verschaltet. Das resultierende Stimulationsfeld der postsynaptischen Zelle hat die typische Mexikanerhutform.*

Kapitel 4 Der Schwelleneffekt als Beispiel systematischer Feldanpassung

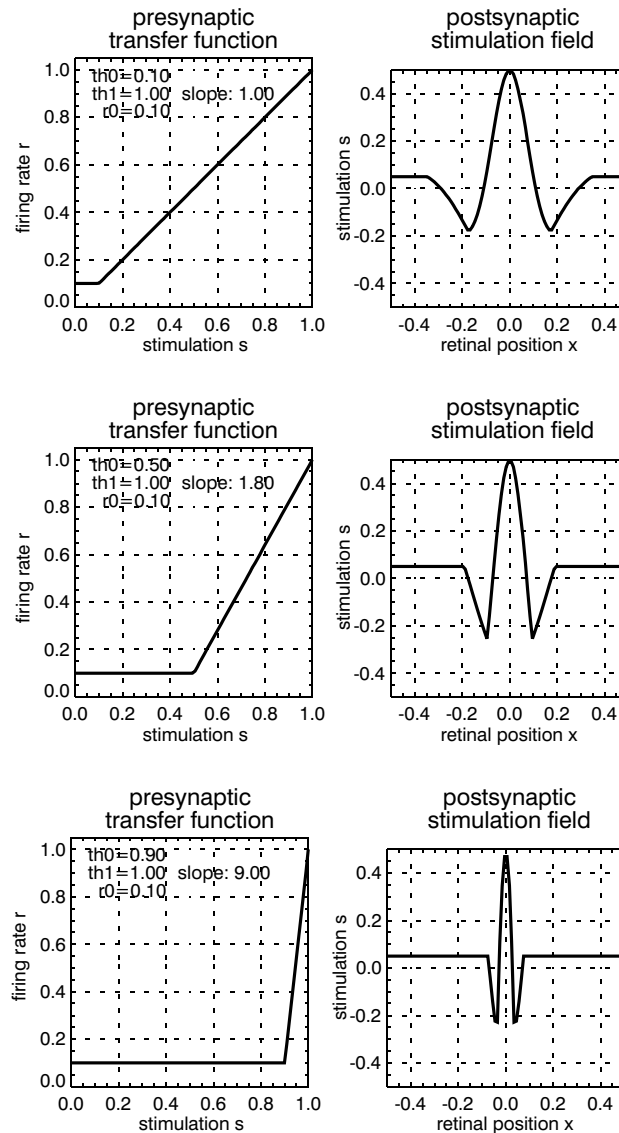


Abbildung 4.6: Der Schwelleneffekt im *Difference-of-Gaussians*-Modell. Verwendet wurde die Verschaltung nach Abbildung 4.5. Bei Variation der Übertragungsfunktion der beiden präsynaptischen Zellen (links) erfährt das Stimulationsfeld der postsynaptischen Zelle (rechts) eine systematische Änderung. Bei Variation der präsynaptischen Übertragungsfunktionen wurde wie zuvor die Verstärkung im Antwortbereich parallel zur Antwortschwelle angepaßt.

4.4 Schwelleneffekt und Größeninvarianz

de Rolle für die entfernungsinvariante Objektwahrnehmung spielen. So wäre es denkbar, daß entferntere (und daher kleiner abgebildete) Objekte durch einen Satz höherauflösender Kontrastdetektoren abgetastet werden, um zu einer invarianten Repräsentation zu gelangen. Brinksmeyer et al. (2002) untersuchen diese Fragestellung mit Hilfe von Ableitungen am Kortex des wachen Affen, und unter abstandsvarianter visueller Stimulation. Die Ergebnisse dieser Arbeit zeigen indes *keine* solche Abhängigkeit der bevorzugten Raumfrequenz primärvisueller Neuronen von der Fixationsentfernung. Die Annahme, daß situationsabhängige Änderungen der Aktivierungsfelder primärer visueller Neuronen dazu dienen könnten, die Auflösung des Sehsystems anzupassen, wird hier also nicht bestätigt.

4.4 Schwelleneffekt und Größeninvarianz

Wäre eine Frequenzdurchstimmung in der beschriebenen Weise prinzipiell dazu geeignet, das Netzhautbild zu skalieren, um so dessen entfernungsabhängige Größenänderung zu kompensieren? Die Angleichung der Aktivierungsfelder visueller Neuronen in Gesamtgröße und räumlicher Frequenzselektivität stellt nur einen Teil der hierzu notwendigen geometrischen Operationen dar. Um dies einzusehen, wenden wir uns Abbildung 4.7 zu.

Für eine Größenskalierung des Netzhautbildes, und zwar über das gesamte Sehfeld hinweg, wäre es nötig, das ganze Bild mit reduzierter Auflösung abzutasten. Erst in zweiter Linie bezieht sich das jedoch auf die Frequenzanteile der repräsentierten visuellen Merkmale. Bedenken wir, daß die Neuronen im primären Sehkortex die Position eines lokalen räumlichen Kontrastsprunges anzeigen: Jeder Zelle ist eindeutig ein bestimmter Reizort zugeordnet. Eine bestimmte Konstellation aktivierter Zellen spiegelt daher eine feste räumliche Konstellation visueller Reize wider. Mithin ist es entscheidend, bei der Skalierung der Repräsentation diesen Ortscode so zu erhalten, daß eine *relationale Codierung* von Objektmerkmalen ermöglicht wird: Um eine entfernungsunabhängige Objekterkennung zu erlauben, muß das gleiche Neuron in den unterschiedlich ausgedehnten Netzhautbildern desselben Sehobjektes ein und dieselbe *objektinterne* Position codieren. Dies ist nur zu erreichen, wenn neben einer Frequenz- und Größenabstimmung die aktivierenden Bereiche selbst im Sehraum verschoben werden, so wie in Abbildung 4.7 illustriert.

**Relationale Codierung
von Objektmerkmalen**

Kapitel 4 Der Schwelleneffekt als Beispiel systematischer Feldanpassung

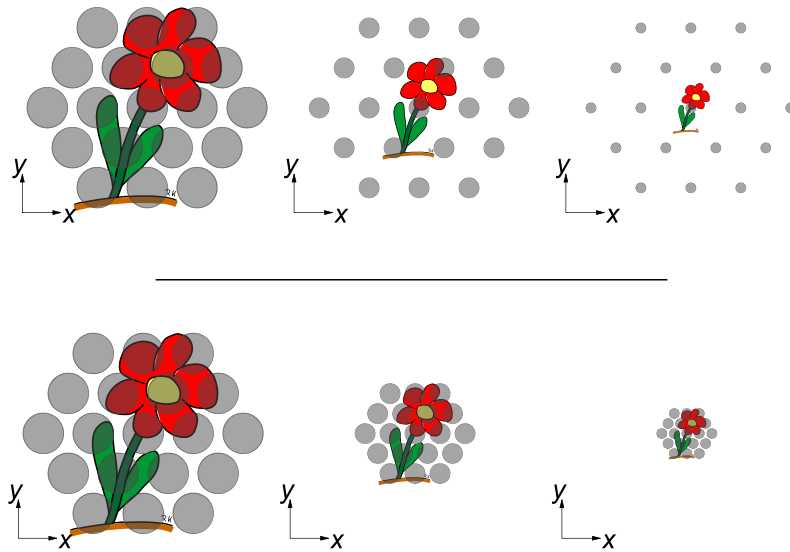


Abbildung 4.7: Auflösungskontrolle durch Skalierung der retinalen aktivierenden Felder (Kreise), ohne und mit gleichzeitiger Verschiebung der Feldmittelpunkte. Ein Satz von Merkmalsdetektoren tastet das Bild einer Blume in verschiedenen Projektionsgrößen ab. **Oben:** Ohne Verschiebung codieren gleiche Detektoren verschiedene Objektteile. **Unten:** Mit Verschiebung codieren gleiche Detektoren gleiche Objektteile.

4.5 Unsymmetrisches Stimulationsfeld

Stellt nun der Schwelleneffekt auch hierfür ein geeignetes Konzept dar? Wie kann durch eine lokale Einflußnahme auf das Neuron eine Verschiebung des Aktivierungsgebietes erreicht werden? Ein solcher Prozeß ist denkbar, wenn von unsymmetrischen Stimulationsfeldern ausgegangen wird. Abbildung 4.8 zeigt ein solches unsymmetrisches („windschiefes“) Stimulationsfeld. Wie sich der Schwelleneffekt bei diesem Ansatz darstellt veranschaulicht Abbildung 4.9.

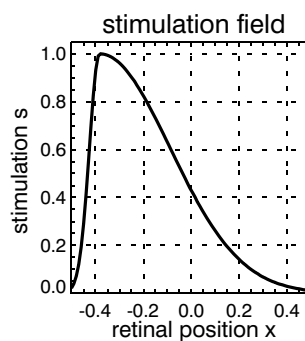


Abbildung 4.8: Unsymmetrisches Stimulationsfeld.

4.5 Unsymmetrisches Stimulationsfeld

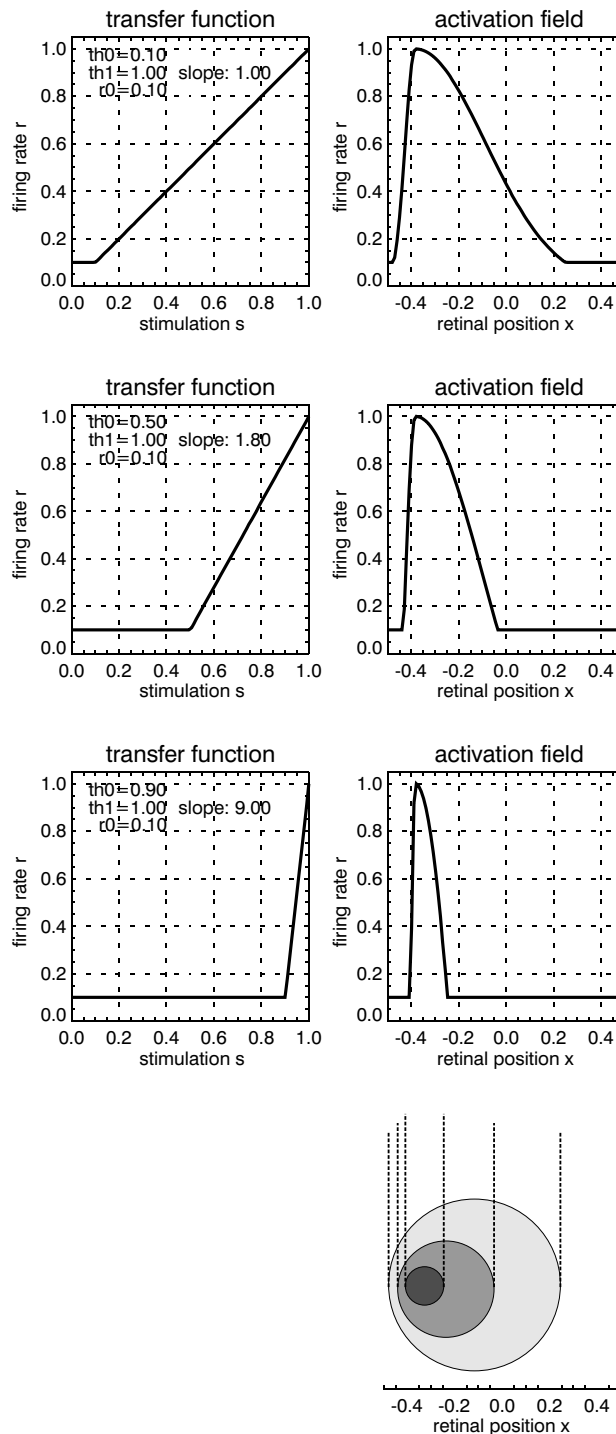


Abbildung 4.9: Der Schwelleneffekt beim unsymmetrischen Stimulationsfeld. Während das Aktivierungsfeld um sein Maximum zusammenschrumpft (rechte Spalte), verschiebt sich dessen *Mittelpunkt* auf der Retina nach links (Schema unten). Auch hier wurde die Verstärkung im Antwortbereich parallel zur Antwortschwelle angepaßt, um gleiche Maximalraten zu erhalten.

Kapitel 4 Der Schwelleneffekt als Beispiel systematischer Feldanpassung

Die Verfahrensweise ist parallel zu der beim symmetrischen Stimulationsfeld. Wie schon zuvor zieht sich beim Anheben der Antwortschwelle das Aktivierungsfeld um sein Maximum hin zusammen. Wegen der Asymmetrie des stimulierenden Feldes liegt dieses Maximum nun aber nicht mehr im Zentrum des Feldes. Vielmehr verlagert sich der Mittelpunkt des Aktivierungsfeldes mehr und mehr von seinem ursprünglichen Ort weg (Abbildung 4.9, unten). Unter diesen Bedingungen wird das Aktivierungsgebiet eines visuellen Neurons im Sehraum wandern.

Lokalitätsprinzip

Dieser Vorgang unterliegt jedoch einer entscheidenden Einschränkung. Einerseits wird das stimulierende Feld eines Neurons nur in einem lokalen Bereich von 0 verschieden sein. Sein Stimulationsgebiet Ω_{SF} ist daher lokal begrenzt:

$$\Omega_{SF} := \{x \mid SF(x) \neq 0\} \quad (4.2)$$

$$\exists x_{min}, x_{max} : x_{min} \leq x \leq x_{max} \quad \forall x \in \Omega_{SF} \quad (4.3)$$

Andererseits handelt es sich bei den für Θ in Frage kommenden Übertragungsabbildungen um monotone Funktionen. Hieraus folgt zunächst, daß auch das resultierende Aktivierungsfeld $AF \equiv \Theta(SF)$ nur in einem lokalen Bereich Ω_{AF} von seinem Minimum $\Theta(0)$ (i.d.R. also von der Spontanrate) verschieden sein kann:

$$AF \equiv \Theta(SF) \quad (4.4)$$

$$\Omega_{AF} := \{x \mid AF(x) > \Theta(0)\} \quad (4.5)$$

$$\exists x_{min}, x_{max} : x_{min} \leq x \leq x_{max} \quad \forall x \in \Omega_{AF} \quad (4.6)$$

Insbesondere gilt jedoch

$$\Omega_{SF} \supset \Omega_{AF} \quad (4.7)$$

Ω_{SF} wird stets Ω_{AF} umfassen. Damit beschränkt das Stimulationsgebiet Ω_{SF} sowohl die maximale Größe des Aktivierungsgebietes Ω_{AF} , als auch dessen maximale Verschiebung durch den Schwelleneffekt. Die realisierbaren Aktivierungsgebiete bilden auf der Retina geschachtelte Gebiete (Abbildung 4.9, unten).

Beim Schwelleneffekt ist die maximal mögliche räumliche Verlagerung des Aktivierungsfeldes von der Größenordnung des Felddurchmessers.

Hierbei handelt es sich um eine prinzipielle Beschränkung, die im Zusammenhang mit dem Schwelleneffekt nicht überwunden werden kann. Die zur Generierung echter visueller Größeninvarianz

notwendigen Verschiebungen über Distanzen, die den Durchmesser der aktivierenden Felder um ein Vielfaches übertreffen (vgl. Abbildung 4.7), lassen sich hiermit nicht realisieren. Wir werden jedoch in den beiden folgenden Kapiteln ein Modellierungskonzept kennenlernen, das dieser Einschränkung nicht unterliegt.

4.6 Diskussion

Der Ansatz, durch einen von außen angreifenden Prozeß das Codierungsverhalten von Neuronen gezielt zu beeinflussen, wird beim Schwelleneffekt beispielhaft umgesetzt. Bei aller modellhaften Einfachheit des Verfahrens (Variation nur zweier makroskopischer Kenngrößen der Übertragungsfunktion, Schwelle und Verstärkung) finden wir, daß damit räumliche Feldveränderungen realisiert werden können, die ein Spektrum hochinteressanter Effekte abdecken:

- Größenanpassung des räumlichen Aktivierungsgebietes von Neuronen
- Kontrolle des Integrationsbereiches bei *Complex*-Bildung
- Damit auch Kontrolle der visuellen Auflösung
- Gleichzeitige Verlagerung des aktivierenden Gebietes im Bereich des Felddurchmessers
- Durchstimmung von Raumfrequenzkanälen

Es ist unklar, inwieweit diese Mechanismen tatsächlich im Sehsystem Verwendung finden, und ob sie eine Rolle insbesondere für die entfernungsinvariante Verarbeitung visueller Reize spielen. Für eine neuronale Raumfrequenz- oder Auflösungskontrolle in Abhängigkeit der Sehentfernung existieren teils widersprüchliche Befunde (vgl. Brinksmeyer et al., 2002; Wörgötter et al., 1998; Bennett und Cortese, 1996; Rainville und Kingdom, 2002). In jedem Fall jedoch bekommt der Modellierer mit dem Schwelleneffekt ein wichtiges Werkzeug zur Hand. Es erlaubt ihm, die oben genannten, sowie alle strukturell verwandten Vorgänge auf einfache Art zu simulieren. Darüber hinaus erleichtert es als Denkmodell das Verständnis für eine Klasse sehr grundlegender Mechanismen in der visuellen Informationsverarbeitung, allen voran der *Complex*-Bildung. Der Schwelleneffekt bewährt sich also als Simulations- und Verständnisbaustein im Sinne der konzeptionellen Modellierung.

Kapitel 5

Ein konzeptionelles Modell zur retinalen Schlupfkorrektur¹

Ziel

In diesem Kapitel entwickle ich ein Modell zur neuronalen Kompensation der durch Mikroaugenbewegungen verursachten retinalen Bildverschiebungen. Es beruht auf der externen Steuerung lokaler neuronaler Codierungseigenschaften. Ich definiere den Begriff der *Sensitivitätsprofile* als Erweiterung dieses Prinzips auf verschiedene Zellgruppen. Darauf basierend führe ich das Konzept der *situativ gesteuerten Complex-Bildung* ein. Ich demonstriere die Leistungsfähigkeit des Modells in einer Computersimulation mit biologisch relevanten Randbedingungen. Das Modell erfordert einen erhöhten neuronalen Aufwand, sowie die Verfügbarkeit eines Blickrichtungssignals. Ich diskutiere mögliche Folgeeffekte wie eine mikrosakkadische Suppression.

5.1 Einleitung

Die sensorische Wahrnehmung ermöglicht es dem Menschen, sich in seiner Umwelt zu orientieren und planvoll mit ihr in Beziehung zu treten. Mehr als alle anderen Sinne vermittelt ihm dabei der Sehsinn ein scheinbar lückenloses Bild seiner sichtbaren Umgebung. Dieses „innere Spiegelbild“ der äußeren Welt deckt sich mit deren physikalischen Gegebenheiten: Unsere Umwelt erscheint uns als räumlich stabil, und insbesondere unabhängig von unserer eigenen Bewegung darin, oder von der Richtung unseres Blickes.

Dieser scheinbar so triviale Umstand wird erstaunlich, wenn wir **Sakkaden**

¹Teile dieses Kapitels sind als Konferenzbeitrag erschienen (Brinksmeyer et al., 2001).

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

die Arbeitsweise unseres Sehapparates in Betracht ziehen. Tatsächlich ruht unser Blick beim freien Umherblicken niemals über längere Zeit auf einem Gegenstand. Mehrere Male pro Sekunde vollführen die menschlichen Augen sprunghafte Blickrichtungsänderungen, sogenannte Sakkaden. Die Blickziele (*Points of Interest*) sind hierbei solche Bestandteile der visuellen Szene, deren Betrachtung entweder dem Erkennungsvorgang dient, oder die verhaltensrelevante Reize darstellen (Henderson, 2003; Pauly, 2000; Kinser, 1999). So sind etwa periphere Blink- oder Bewegungsreize hervorragend geeignet, einen sakkadischen Reflex auszulösen.

Unsere innere Darstellung der Außenwelt kommt also dadurch zustande, daß wir die Umgebung in einer schnellen Folge einzelner Blicke erfassen. Dabei ist der zentrale Bereich unseres Sehfeldes, in dem hochaufgelöste Verarbeitung stattfindet, äußerst begrenzt. Inwieweit der lückenlose und detailreiche Eindruck der uns umgebenden Welt daher eine Illusion sein muß ist Gegenstand wissenschaftsphilosophischer Betrachtungen (Noë, 2002; O'Regan und Noë, 2000, 2001). Zumindest stellt sich jedoch die Frage, wie wir die äußere Welt als unbewegt und stabil empfinden können, obgleich sich der an der Netzhaut anliegende Reiz mit jeder Sakkade ändert.

Residuale Augenbewegungen

Selbst bei willentlicher Fixation eines festen Punktes, bei der die sonst reflexhaft verlaufenden Sakkaden unterdrückt werden, stehen unsere Augen nicht still. Vielmehr vollführen sie unablässig Bewegungen von bis zu 2° Sehwinkel, (Median bei ca. $\frac{1}{2}^\circ$ Sehwinkel, beim Menschen, Henderson, 2003), die nicht dem Willen des Betrachters unterliegen.² Sie bestehen aus einer Überlagerung langsamer Drifts und schneller Sprünge (*Mikrosakkaden*), sowie hochfrequentem Augentremor. Ein typisches Blickrichtungssignal vom Affen ist in Abbildung 5.2 dargestellt. Man spricht zusammenfassend von *residualen Augenbewegungen* oder *Mikroaugenbewegungen*. Ihre Rolle beim Sehvorgang ist weitgehend ungeklärt; sie wird in der Vermeidung des perzeptuellen Ausgrauens (*Fade-Out*) vermutet, der durch die retinale Adaptation bei retinafixen Stimuli auftritt (Riggs und Ratliff, 1952; Ditchburn und Ginsborg, 1952; Rodieck, 1998; Bridgeman und Palca, 1980; Martinez-Conde et al., 2000, 2002). Möglicherweise sind sie auch der Informationsverarbeitung in der zeitlichen Domäne dienlich, indem sie Zellantworten synchronisieren und damit eine verlässlichere Codierung ermöglichen (Greschner et al., 2002). Der Leser kann die Mikroaugenbewegungen mithilfe der Abbildung 5.1

²Diese Mikroaugenbewegungen treten jedoch in reduzierter Zahl auf, wenn Probanden visuelle Aufgaben mit höchster Genauigkeit vollführen (Bridgeman und Palca, 1980).

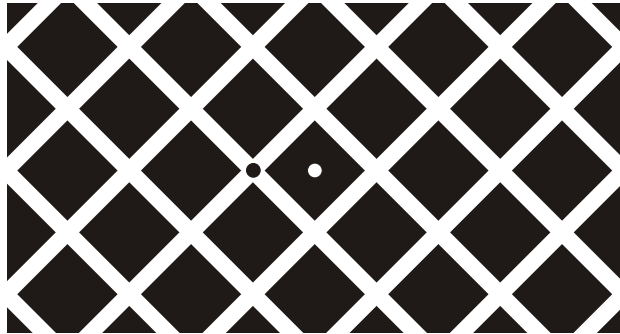


Abbildung 5.1: Selbstexperiment zu Demonstration residualer Augenbewegungen. Der Leser fixiere für etwa eine Minute den kleinen weißen Kreis. Danach lasse er seinen Blick zum schwarzen Kreis wandern. Das in die Netzhaut „ingebrannte“ Nachbild der schrägen Streifen vollführt bei Fixation deutlich wahrnehmbare Bewegungen. Diese rühren von den residualen Augenbewegungen des Betrachters her.

im Selbstexperiment erfahren.

Der durch die residualen Augenbewegungen verursachte Schlupf des auf die Retina projizierten Bildes wird vom Betrachter nicht wahrgenommen; ihm erscheint das Bild ruhig, scharf und stabil.³ Offenbar ist unsere Wahrnehmung, und daher vermutlich ein großer Teil der visuellen Reizverarbeitung, invariant gegenüber residualen Augenbewegungen (eine spezielle Form der Positionsinvarianz). Die typische menschliche Sehschärfe liegt im Bereich der Fovea Centralis bei ca. einer Winkelminute (für Visusbestimmung mithilfe des Landolt-Ringes, Schmidt und Thews, 1997). Sie löst Details auf, die weit unterhalb der Amplitude dieser Augenbewegungen liegen. Wieso bleibt die visuelle Wahrnehmung von den Mikrobewegungen unbeeinflusst?

Eine mögliche Erklärung hierfür ist, daß das visuelle System die Mikrobewegungen des auf die Netzhaut projizierten Bildes durch einen neuronalen Mechanismus kompensiert. Dies müßte auf niedriger Ebene in der Hierarchie der visuellen Verarbeitung geschehen, vor Ablauf des Erkennungsvorganges, und bevor das Bild in das Bewußtsein des Betrachters dringt. Jede weitere visuelle Verarbeitung würde dann auf dem bereits korrigierten Signal aufbauen und wäre damit von den residualen Augenbewegungen unbeeinflusst. Das macht den primären visuellen Kortex zum vermuteten Wirkort die-

Neuronale Korrektur

³Introspektiv ist dem Leser möglicherweise das unbestimmte Gefühl bekannt, einen sehr kleinen Gegenstand oder ein kleines Detail „mit den Augen einfach nicht recht festhalten“ zu können, obgleich sich das Bild keineswegs zu bewegen scheint. Es besteht die Möglichkeit, daß dieses Gefühl seinen Ursprung in den residualen Augenbewegungen hat.

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

ses Prozesses. Für die Existenz eines solchen neuronalen Korrekturgliedes finden sich sowohl bestätigende (Motter und Poggio, 1990) wie auch widersprechende (Gur und Snodderly, 1997) experimentelle Befunde.

Die exakte Vermessung visueller Aktivierungsfelder⁴ von Neuronen im primären visuellen Kortex unter gleichzeitiger Beobachtung der residualen Augenbewegungen scheint eine geeignete Methode, um Aufschluß über einen vermuteten Kompensationsmechanismus zu erlangen. Brinksmeyer et al. (2001) verwenden das RF-Cinematographie-Verfahren (Eckhorn et al., 1993) beim wachen Affen in Kombination mit einer hochauflösenden Blickrichtungsüberwachung durch einen Infrarot-Eyetracker, um zu entscheiden, ob das visuelle System eine neuronale Korrektur durchführt.⁵ Ihre Ergebnisse stehen im Einklang mit der Annahme, daß in der Tat eine dynamische Verschiebung visueller Aktivierungsgebiete *in der Retina* stattfindet, die dem retinalen Schlupf des projizierten Bildes entgegenwirkt. Indem die aktivierenden Gebiete in der Netzhaut mit der Bewegung des Bildes mitgeführt werden, erhalten Neuronen im visuellen Kortex feste Aktivierungsfelder *in der äußeren Welt*.⁶

Ich stelle hier ein Modellierungskonzept im Rahmen der situativen Steuerung von Codierungseigenschaften vor (Kapitel 3 und Abschnitt 3.4), das über die gesteuerte Anpassung aktivierender Felder erlaubt, diesen Vorgang nachzuvollziehen.

5.2 Problemstellung: Neuronale Schlupfkorrektur

Der gesuchte neuronale Mechanismus soll in der Lage sein, die aktivierenden Felder visueller Neuronen in der Retina zu verschieben. Dies muß in Abhängigkeit der momentanen Augenstellung geschehen, gleichförmig über mindestens das zentrale Sehfeld hinweg, und mit einer möglichst geringen Verzögerung.

Ich will im folgenden davon ausgehen, daß der Fixationsfehler, also die momentane Abweichung der tatsächlichen Augenstellung von der beabsichtigten Fixationsrichtung, neuronal verfügbar ist.

⁴Zur Terminologie siehe Abschnitt 3.3.3.

⁵Ein Anlaß für ihre Untersuchungen war die beabsichtigte Verfeinerung der cinematographischen Auflösung durch *explizite* rechnerische Korrektur der Augenbewegung (Brinksmeyer, persönliche Kommunikation).

⁶Für die Dauer der Fixation, ohne makroskopische Augen- oder Kopfbewegungen.

5.2 Problemstellung: Neuronale Schlupfkorrektur

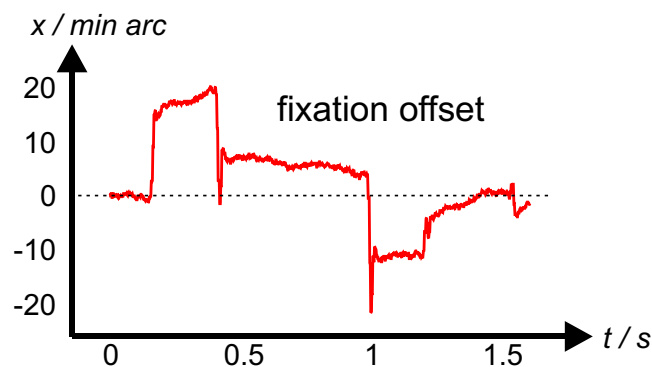


Abbildung 5.2: Typischer Verlauf des Fixationsfehlers (Horizontalkomponente). Originales Meßsignal aus einem Affenexperiment, Brinksmeyer et al. (2001, persönliche Kommunikation).

Damit liegt eine physikalisch-körperliche Größe, nämlich die „Fehlstellung“ der Augen, zum Zwecke der situativen Steuerung vor. Mögliche Quellen für dieses Signal sind Efferenzkopien der neuronalen Steuersignale zu den Augenmuskeln, oder die propriozeptive Wahrnehmung ihrer augenblicklichen Stellung. Ferner kann der Fixationsfehler auch aus dem bewegten visuellen Reiz selbst extrahiert werden, eine Methode, die Abschnitt 5.7.5 näher beleuchtet wird. Ein typisches Blickrichtungssignal ist in Abbildung 5.2 zu sehen. Es handelt sich dabei um ein originales Meßsignal aus den Erhebungen von Brinksmeyer et al. (2001, persönliche Kommunikation).

Durch die Schlupfkorrektur soll ausschließlich eine *Translation* des Stimulus ausgeglichen werden. Alle weiteren Parameter der neuronalen Verarbeitung, insbesondere die Form und Größe der aktivierenden Felder und der bevorzugten Reize müssen von dem Vorgang unbeeinflusst bleiben. Die neuronal codierten visuellen Merkmale dürfen sich bei der Verschiebung nicht ändern, da sonst das wahrgenommene Bild Verzerrungen erfährt. Wir wissen aus unseren Untersuchungen in Kapitel 4, daß eine Feldverlagerung beim Schwelleneffekt mit einer Größenänderung verbunden ist. Der Schwelleneffekt ist also keine geeignete Methode zur Modellierung der Schlupfkorrektur. Da das gesuchte Verfahren robust und mit geringer Verzögerung arbeiten soll, ist es jedoch sinnvoll, weiterhin von einer strukturell einfachen neuronalen Vorwärtsverschaltung auszugehen. In Abschnitt 3.4 haben wir uns klargemacht, daß sich in einem solchen System die Merkmale der präsynaptischen Zellen dem postsynaptischen Neuron aufprägen (Seite 30). Diese Eigenschaft wollen wir jetzt ausnutzen.

**Schwelleneffekt
ungeeignet**

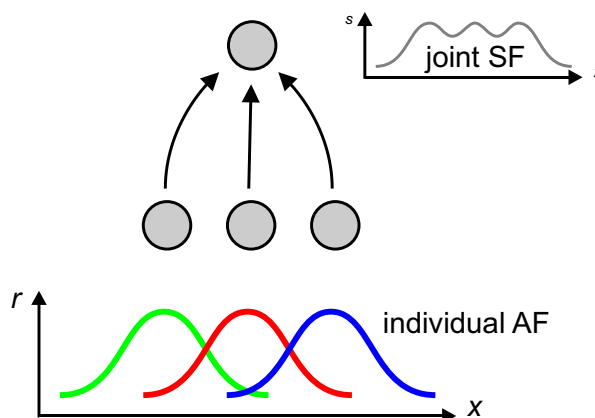


Abbildung 5.3: Neuronaler Gating-Baustein. Im neutralen Zustand ist das Stimulationsfeld des Zielneurons eine Überlagerung der präsynaptischen aktivierenden Felder.

5.3 Konzeptuelle Grundlagen

5.3.1 Präsynaptisches Gating

Betrachten wir ein zweistufiges neuronales System mit konvergenter Verschaltung. Die Neuronen in der ersten Stufe haben identische aktivierende Felder, die sich nur durch eine relative Verschiebung zueinander unterscheiden. (Abbildung 5.3; wie schon zuvor gehen wir von einem eindimensionalen Ansatz aus.) Weiterhin haben alle diese Zellen gleichstarke synaptische Afferenzen zum Neuron in der zweiten Stufe. Das Stimulationsfeld dieses Neurons wird dann eine Überlagerung der aktivierenden Felder der präsynaptischen Neuronen sein (Abbildung 5.3, oben rechts).

Gating

An dieser Stelle ist die Trennung der in Abschnitt 3.3.3 definierten Feldbegriffe entscheidend: Die visuellen Reize werden durch die stimulierenden Felder der Neuronen in der ersten Stufe gewichtet. (Diese sind durch die synaptischen Profile und Übertragungsfunktionen aller vorgeschalteten Neuronen bestimmt.) Die hervorgerufene Zellstimulation wird anschließend entsprechend der Übertragungsfunktion der Zellen in die Ausgangsrate umgesetzt. Erst diese Zellaktivität konvergiert auf das Neuron in der zweiten Stufe.

Das stimulierende Feld des Zielneurons ist eine Überlagerung der aktivierenden, nicht der stimulierenden Felder der Quellneuronen.

Gating und Complex-Bildung

Die relative Gewichtung, mit der das jeweilige Aktivierungsfeld in das Gesamtstimulationsfeld des Zielneurons eingeht, kann über

5.3 Konzeptuelle Grundlagen

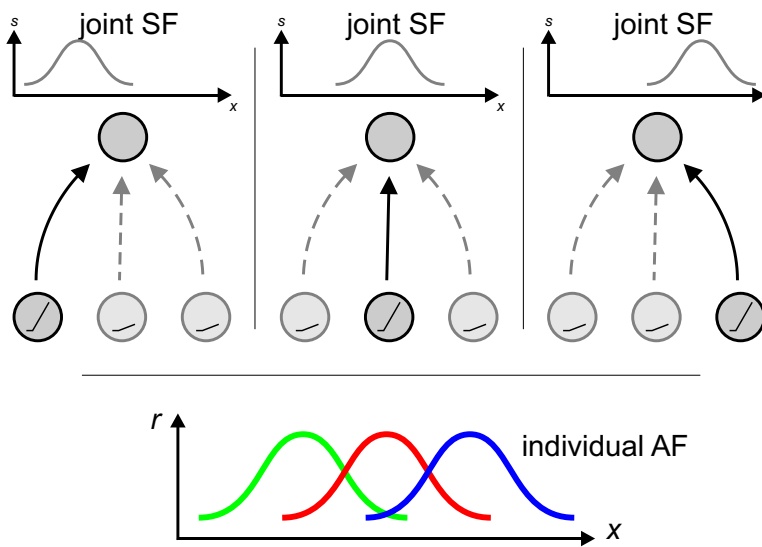


Abbildung 5.4: Verschiebung des Stimulationsfeldes für das Zielneuron durch präsynaptisches *Gating*. Das Gesamtfeld kommt durch die gewichtete Überlagerung der als Quellen dienenden präsynaptischen Aktivierungsfelder zustande. Die jeweils funktionslosen Neuronen sind in der Abbildung hell dargestellt.

die Verstärkung der einzelnen präsynaptischen Neuronen eingestellt werden. Bei dem hier vorgestellten neuronalen Modul handelt es sich also um eine spezielle *gesteuerte* Form der in Abschnitt 2.1.1 beschriebenen *Complex*-Bildung. Der Vorgang ist in Abbildung 5.4 illustriert. Im einfachsten Fall können auf diese Weise Bestandteile des kombinierten Feldes völlig ausgeblendet werden (*Gating*), indem die Verstärkung der betreffenden Quellneuronen soweit reduziert wird, daß ihre Ausgangssignale keinen Einfluß mehr auf das Verhalten des Zielneurons haben. Die betreffenden Zellen sind im neuronalen System dann funktionslos, also *effektiv nicht vorhanden*.

In unserem Beispiel der linearen Anordnung einer Reihe von Neuronen mit identischen doch verlagerten Stimulationsfeldern kann das präsynaptische *Gating* zur Generierung von Positionsinvarianz eingesetzt werden: Werden die Signale einer Zelle gegenüber allen anderen deutlich stärker gewichtet, so wird das Aktivierungsfeld dieser Zelle das Gesamtfeld dominieren. Durch Auswahl des jeweils „eingebledeten“ Quellneurons kann damit das Stimulationsfeld der Zielzelle – unter Erhaltung seiner Form – linear verschoben werden.

Im Unterschied zum zuvor besprochenen Schwelleneffekt setzt bei diesem Verfahren ein potentieller äußerer Steuerprozeß nicht am

betreffenden Neuron selbst, sondern an den Übertragungsfunktionen der präsynaptischen Neuronen an.

5.3.2 Sensitivitätsprofile

Das Konzept der situativ gesteuerten Feldanpassung basiert darauf, daß ein globaler Parameter, der eine von außerhalb des neuronalen Systems gegebene Größe widerspiegelt, ein ganzes neuronales Teilsystem geeignet einstellt (Abschnitt 3.4). Betrachten wir das in Abschnitt 5.3.1 skizzierte Beispiel der linearen Feldverschiebung durch präsynaptisches *Gating*, so wird klar, daß sich das globale Steuersignal auf die einzelnen beteiligten Neuronen *unterschiedlich* auswirken muß: Bei einigen von ihnen soll die Verstärkung der Übertragung angehoben, bei anderen reduziert werden, um eine Feldverschiebung zu erzielen (Abbildung 5.4). Der durch die äußeren Umstände vorgegebene globale Steuerparameter muß also lokal für die einzelnen Neuronen geeignet umgesetzt werden.

Rezeptives Feld für den Steuerparameter

Ich will hierfür die Existenz individueller *Sensitivitätsprofile* annehmen. Diese definieren den Zusammenhang des global vorliegenden Steuerparameters mit dessen Einfluß auf die Codierungseigenschaften eines einzelnen Neurons. Dieser Einfluß wird typischerweise modulatorischer Natur sein, und das Neuron für afferente Signale sensitivieren oder desensitivieren. Die in Abschnitt 3.2 besprochenen *Gainfields* für Position und Entfernung stellen solche Sensitivitätsprofile dar.

Im verwendeten Ratenmodell kann als *Sensitivität* einer Zelle deren Verstärkung, also die Steigung der Übertragungsfunktion Θ verstanden werden. Der für den Effekt blickrichtungsmodulierter Zellantworten geprägte Begriff *Gainfield* („Verstärkungs-Feld“) beschreibt einen solchen Zusammenhang. Das Konzept des Sensitivitätsprofils ist jedoch weder auf eine *örtliche* Größe als Steuerparameter beschränkt, noch ist die Variation der Verstärkung der einzig denkbare Mechanismus zur Zellsensitivierung.⁷ Insofern stellt das *Sensitivitätsprofil* eine Erweiterung des *Gainfield*-Konzeptes dar.

Das Sensitivitätsprofil eines Neurons ist von dessen visuellem rezeptiven Feld zu trennen. Jedoch gilt:

⁷Die effektive Variation der Antwortschwelle ϑ_0 , etwa durch inhibitorische Einflüsse, ist ein weiterer denkbarer Mechanismus.

Das Sensitivitätsprofil eines Neurons kann als rezeptives Feld für den Steuerparameter verstanden werden. Es definiert den Zusammenhang zwischen Steuerparameter und Zellsensitivität, gerade so wie das visuelle rezeptive Feld den Zusammenhang zwischen visuellem Stimulationsort und Zellerregung definiert.

5.4 Modell

Zur Modellierung eines neuronalen Schlupfkorrekturgliedes im primären Sehkortex gehen wir von der in Abbildung 5.3 skizzierten Basisanordnung aus, die uns hier als Modellbaustein dient. Der relative Versatz der drei rezeptiven Felder, ausgedrückt in Sehwinkel, liege im Bereich der Amplitude residualer Augenbewegungen. Wir wollen annehmen, daß der gesamte der Korrektur unterliegende Sehbereich mit Hilfe solcher neuronalen Module abgetastet wird. Die Sensitivitätsprofile seien für alle diese Module identisch vorgegeben, entsprechend Abbildung 5.5.

Im Falle des Modells zur retinalen Schlupfkorrektur beschreibt das Sensitivitätsprofil, wie der momentane Fixationsfehler (als globaler Steuerparameter, Abschnitt 5.2) in die Übertragungsverstärkung der drei einzelnen Neuronen (als lokaler Einfluß) übersetzt wird (Abbildungen 5.3, 5.5).

Die an der Ordinatenachse aufgetragene *Sensitivität* S beeinflusst die Verstärkung des betreffenden Neurons multiplikativ. S nimmt Werte im Bereich zwischen 0 (komplette Insensitivität) und 1 (maximale Verstärkung) an. Der Zusammenhang zwischen Sensitivität S und aktueller sowie maximaler Verstärkung $\frac{\Delta r}{\Delta s}$ der Übertragungsfunktion eines Neurons im verwendeten Ratenmodell ist wie folgt:

$$\frac{\Delta r}{\Delta s} = S \cdot \left(\frac{\Delta r}{\Delta s} \right)_{max} \quad (5.1)$$

Die Abtastung des zu korrigierenden Sehbereichs mit identischen *Gating*-Modulen dieser Art liefert drei komplette Sätze gegeneinander verschobener visueller Detektoren, wie sie in Abbildung 5.6 schematisch dargestellt sind. Da die einzelnen Module identische Sensitivitätsprofile nach Abbildung 5.5 besitzen, ist jeweils ein kompletter Satz von Detektoren sensitiv. Nur dieser, also entweder der zentrale, der nach links oder der nach rechts verlagerte Satz (in Abbildung 5.6 farblich gekennzeichnet) leitet zu einer gegebenen Zeit seine Aktivierung an die Ausgangsneuronen der *Gating*-Module weiter. Die Sensitivitätsprofile sind dabei so gewählt, daß eine residuale Augen-

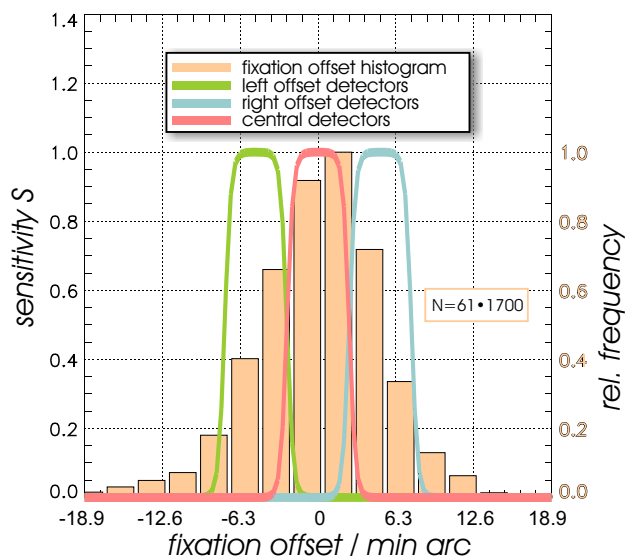


Abbildung 5.5: Sensitivitätsprofile der drei Detektorpopulationen im Schlupfkorrektur-Modell. Der nach links verschobene Detektorsatz ist sensitiv für einen Fixationsfehler nach links (retinale Verschiebung ebenfalls nach links, vgl. Diskussion Abschnitt 5.7.1), und umgekehrt. Die Einheiten der Abszisse sind Winkelminuten. Hellrot hinterlegt die Verteilung der experimentell ermittelten Fixationsfehler.

bewegung, die eine Verschiebung des Bildes auf der Netzhaut nach rechts verursacht, auch zur Sensitivierung der nach rechts versetzten Detektoren führt, und umgekehrt. Die resultierenden Stimulationsfelder der Detektormodule werden dadurch kohärent über das ganze Sehfeld hinweg den residualen Augenbewegungen nachgeführt. Abbildung 5.6 zeigt die einzelnen Modellkomponenten in der Übersicht.

5.5 Simulation

Zum Test des vorgeschlagenen Modells vollziehe ich die Untersuchungen von Brinksmeyer et al. (2001) exemplarisch nach. Die Autoren hatten in einer Fixationsaufgabe beim wachen Affen die rezeptiven Felder primärvisueller Neuronen aus Einzelzell- und Massensignalen erhoben. Dabei kam die RF-Cinematographie-Methode zum Einsatz (Eckhorn et al., 1993). Gleichzeitig wurde die Blickrichtung registriert, um den Fixationsfehler durch Mikroaugenbewegungen bestimmen zu können.

*Simulierte
RF-Cinematographie*

In einer Computersimulation des vorgeschlagenen Modells setze ich ein ähnliches Verfahren ein, indem ich am Bildeingang (entspre-

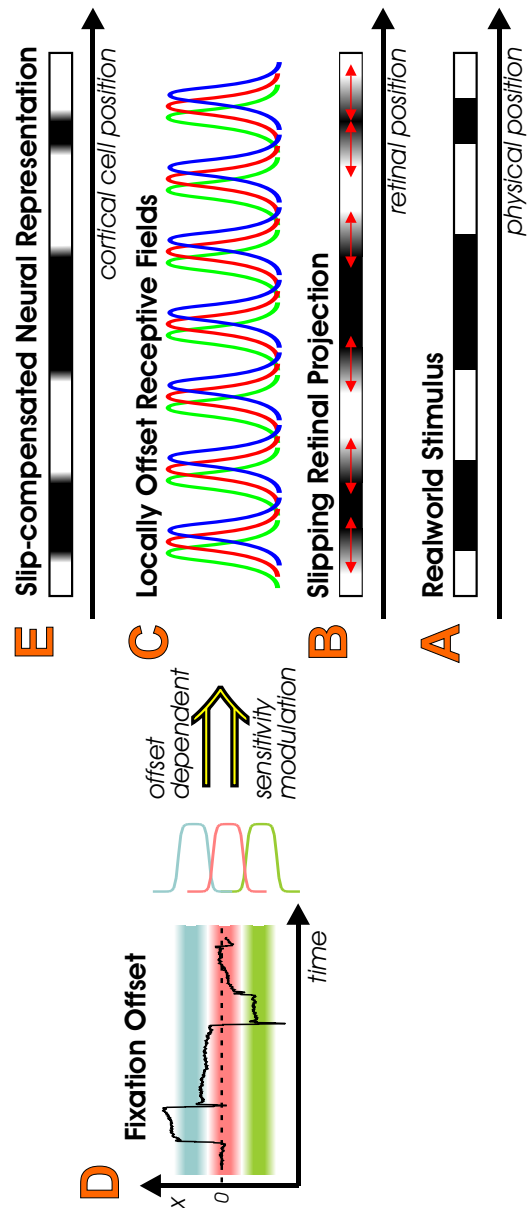


Abbildung 5.6: Schematische Darstellung des Simulationsmodelles zur retinalen Schlupfkorrektur. **A:** Echtweltstimulus, der durch die Augenlinse auf die Retina projiziert wird. **B:** Aufgrund von Mikroaugenbewegungen ist die retinale Projektion des Stimulus instabil. **C:** Die retinale Projektion wird von drei gegeneinander verschobenen Detektorsätzen abgetastet. (Die rezeptiven Felder der Detektoren liegen wesentlich dichter als hier dargestellt und überlappen.) **D:** Die Sensitivität der Detektorsätze wird entsprechend der Sensitivitätsprofile durch das Blickrichtungssignal gesteuert. **E:** Das Ausgangssignal der *Gating*-Module liefert dann eine schlupfkompenzierte Repräsentation des Stimulus.

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

chend Retina, *corpus geniculatum laterale* oder primärem Sehkortex, im folgenden als *simulierte Retina* bezeichnet) Testreize in Form kleiner heller Punkte appliziere. Der Stimulationsort unterliegt dabei Schwankungen, die Mikroaugenbewegungen imitieren. Gleichzeitig wird die Aktivität aller Neuronen des Korrekturgliedes in simulierten Einzelzellmessungen registriert. Die Korrelation der Zellantworten mit dem momentanen Eingangsreiz liefert dann eine Schätzung des Aktivierungsfeldes der einzelnen Neuronen (vereinfachtes RF-Cinematogramm (s.u.), Eckhorn et al., 1993).

5.5.1 Methoden

Stimulation

Die Simulation arbeitet auf einem räumlichen Gitter der Auflösung $1,26' \text{ arc/pixel}$. Die Testreize sind einzelne helle Pixel in der simulierten Retina, und mit $1,26' \text{ arc}$ Durchmesser deutlich kleiner als die von Brinksmeyer et al. verwendeten Stimuli. In Affenexperimenten sind den verwendbaren Testreizen geometrische Randbedingungen gesetzt, hinsichtlich ihrer Eignung zur Erregung der Neuronen im retino-kortikalen Pfad. In der Simulation kann die Reizhelligkeit soweit angepaßt werden, daß auch verglichen mit den RF-Durchmessern kleine Reize aktivierend wirken. Im Sinne einer möglichst hochauflösenden cinematographischen Abtastung werden daher kleinstmögliche Teststimuli gewählt.

Cinematographie

In der Simulation kommt ein vereinfachtes Cinematographieverfahren zum Einsatz, da im Modell keine neuronalen Laufzeiten zu berücksichtigen sind. Die Ausgangsneuronen sprechen ohne zeitliche Verzögerung an. Da ferner das Ausgangssignal bei konstantem Eingang keine zeitliche Dynamik aufweist, enthalten die berechneten Cinematogramme nur räumliche und keine zeitlichen Komponenten. Sie stellen Schätzungen der aktivierenden Felder der Neuronen dar.

Mikrobewegungen

Residuale Augenbewegungen werden in Form zeitabhängiger örtlicher *Offsets* auf den Positionen der Testreize simuliert: Die Stimulationsorte werden dem augenblicklichen Fixationsfehler entsprechend auf der simulierten Retina verschoben. *Diese Verschiebungen werden als externe Störung aufgefaßt und gehen nicht in die Berechnung der Aktivierungsfeld-Schätzungen ein.* Als Datenbasis dienen 61 Blickrichtungsmessungen von je 3,4s Länge (je 1700 Datenpunkte, Auflösung 2ms) aus Originalexperimenten von Brinksmeyer et al. (2001, persönliche Kommunikation).

Detektoren

Simuliert wird eine lineare Anordnung von 64 *Gating*-Modulen, wie sie in den Abbildungen 5.3 und 5.4 dargestellt sind. Jedes

der *Gating*-Module verwendet drei Eingangszellen, so daß die Testreize von insgesamt $3 \cdot 64$ Detektoren abgetastet werden. Der Modell-Ausgang besteht aus den 64 schlupfkorrigierten Zellantworten der Zielneuronen. Zum Einsatz kommt das *Graded-Response*-Neuronenmodell mit der in Abschnitt 4.1 eingeführten Übertragungsfunktion. Spontanaktivität und Antwortschwelle der beteiligten Neuronen werden als fest und vernachlässigbar klein angenommen ($r_0, \vartheta_0 \approx 0$).⁸

Die Breite der als gaußförmig angesetzten stimulierenden Einzelfelder liegt mit $\sigma = 3,8'$ arc im Bereich der kleinsten von Brinksmeyer et al. (2001) gemessenen Felder (Zellen in V1 des Affen mit parafovealen Feldpositionen).⁹ Der relative Versatz der drei Neuronenpopulationen entspricht der Standardabweichung der gemessenen Blickrichtungssignale.

Die Sensitivitätsprofile der drei Detektorsätze sind in Abbildung 5.5 dargestellt. Ihr relativer Versatz entspricht gleichfalls der Standardabweichung der gemessenen Blickrichtungssignale. Um eine annähernd konstante Abdeckung des ganzen Blickbereiches zu erhalten, wird für die Profile eine Funktion mit quasi-rechteckigem Verlauf gewählt (*Fermi-Gate*):

$$\mathcal{FG}^T(x) : x \mapsto 1 - \mathcal{F}^T\left(|x - m| - \frac{w}{2}\right) \quad \text{Fermi-Gate} \quad (5.2)$$

mit

$$\mathcal{F}^T(x) : x \mapsto \frac{1}{1 + e^{-\frac{x}{T}}} \quad \text{Fermifunktion}$$

m : Mittelwert

w : Fensterbreite

T : Temperaturkoeffizient

Die Simulationsdauer beträgt $61 \cdot 1700$ Zeitschritte, entsprechend einer Gesamtfixationsdauer von ca. $3\frac{1}{2}$ min.

5.6 Ergebnisse

Abbildung 5.7 zeigt die Ergebnisse der Simulation. Dargestellt sind (links) die mittleren Aktivierungsfeld-Cinematogramme sowohl der

⁸Ein „Eisbergeffekt“ durch Variation der Antwortschwelle wäre ein weiterer möglicher Mechanismus zur relativen Zellsensitivierung. Hier soll nur die Übertragungsverstärkung zur Einstellung der Sensitivität herangezogen werden.

⁹Bei den Massenzellableitungen von Brinksmeyer et al. (2001, persönliche Kommunikation) wurden typischerweise annähernd gaußförmige Feldprofile ohne negative Subfelder gemessen. Alle Modellergebnisse gelten ebenso für komplexer strukturierte rezeptive Felder.

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

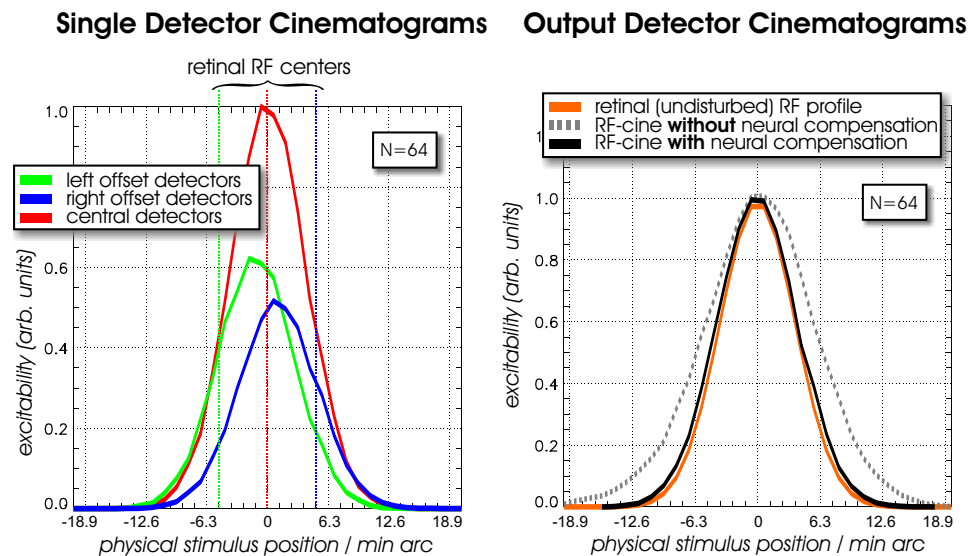


Abbildung 5.7: Simulationsergebnis. Aktivierungsfeld-Cinematogramme der simulierten visuellen Detektoren. **Links:** Die mittleren Felder der drei einzelnen Detektorpopulationen werden nicht an dem Ort gemessen, der aufgrund ihrer retinalen Position zu erwarten wäre. Die gemessenen Felder sind zum Zentrum hin verschoben. **Rechts:** Das mittlere Aktivierungsfeld der Neuronen in der Ausgangsstufe ist bei deaktivierter neuronaler Schlupfkorrektur deutlich verbreitert (graue Kurve). Bei aktivierter Korrektur ist es weitgehend deckungsgleich mit dem tatsächlich einprogrammierten, retinalen Stimulationsfeld des zentralen Detektorsatzes (orange Kurve). Die Einheiten der Abszisse sind Winkelminuten.

Einzeldetektoren in der Eingangsstufe des Korrekturgliedes, als auch (rechts) die der Ausgangsstufe. Der Mittelwert wurde über die 64 identischen *Gating*-Module gebildet. Die Farben der Kurven links entsprechen denen in Abbildung 5.6.

5.6.1 Einzelfelder

In Abbildung 5.7 links sehen wir das mittlere Aktivierungsfeld-Cinematogramm über alle 64 Zellen einer jeden Population. Auf der Abszisse ist der Ort des Teststimulus *im Sehraum* aufgetragen. Die Ordinate gibt die Zellantwort in arbiträren, für die drei Kurven jedoch gleichen Einheiten wieder. Als örtlicher Nullpunkt der Mittelung diente für alle drei Detektoren eines jeden Moduls die retinale Position des stimulierenden Feldes¹⁰ des mittleren Detektors. Wir erwarten also, den retinalen Versatz der beiden anderen Felder in der

¹⁰Hier und im folgenden ist mit *Position des Feldes* dessen Mittelpunkt gemeint.

Grafik wiederzufinden. Als Orientierungspunkte sind hierzu die retinalen Positionen der verschobenen Felder links und rechts als gestrichelte Linien eingezeichnet. (Der Versatz der Detektoren in der simulierten Retina betrug 4 Pixel, entsprechend $5,04'$ arc.)

Wir sehen, daß die Cinematogramme der beiden äußeren Detektorsätze deutlich zum Zentrum hin versetzt sind: Sie wurden durch die Cinematographiemethode nicht an der Stelle gemessen, die ihrer retinalen Position entspricht. Weiter fällt auf, daß die gemessenen Profile der beiden äußeren Detektorsätze gegenüber dem zentralen Satz geringere Amplituden aufweisen.

*Verschobene
Einzelfelder*

5.6.2 Ausgangsfelder

In Abbildung 5.7 rechts findet sich das mittlere Cinematogramm der resultierenden aktivierenden Felder der Ausgangsneuronen dargestellt (**schwarze Kurve**). Dies sind die Neuronen in der oberen Stufe der *Gating*-Module, vgl. Abbildung 5.3. Zur Orientierung sind in dem Diagramm zwei weitere Kurven eingetragen:

- Die **orange Kurve** gibt das *retinale* Stimulationsfeld des mittleren (rot dargestellten) Detektorsatzes wieder, so wie es als synaptisches Ortsprofil in der Simulation einprogrammiert ist.
- Die **gepunktete Kurve** zeigt das Cinematogramm für einen veränderten Simulationslauf: Hier war für die gesamte Dauer der Simulation außschließlich der mittlere (rot dargestellte) Detektorsatz sensitiv, *unabhängig von der Blickrichtung*. Die beiden anderen Sätze waren während der gesamten Simulationsdauer inaktiviert. Das Cinematogramm ist wie unter diesen Umständen zu erwarten über den Bereich des Fixationsoffsets hin „verschmiert“, also gegenüber der orangen Kurve deutlich verbreitert.

5.7 Diskussion

Erinnern wir uns, daß bei der Berechnung der Cinematogramme der Fixationsfehler durch residuale Augenbewegungen nicht berücksichtigt wurde. Die Bewegungen sind jedoch in die visuelle Reizung eingegangen, und sie lagen als Steuersignale an den *Gating*-Modulen an. Dies entspricht der Annahme, daß die mit Fixationsfehlern behafteten Signale im neuronalen Korrekturglied bereinigt werden. Nachfolgende Verarbeitungsstufen operieren auf diesen bereinigten Signalen, unbeeinflusst von Mikroaugenbewegungen. Die

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

berechneten Cinematogramme erlauben uns abzuschätzen, welche Signale die nachfolgenden Verarbeitungsstufen erreichen, und wie effektiv die neuronale Korrektur arbeitet.

5.7.1 Einzelfelder

Die Cinematogramme der drei Detektorsätze am Eingang (Abbildung 5.7, links) zeigen, wie sich die Ausgangssignale der Neuronen in der unteren Ebene des Modells für eine aufbauende Verarbeitungsstufe darstellen: Offenbar scheinen die rezeptiven Felder aller drei Detektoren an fast identischen Orten im Sehraum zu liegen. Wie ist das zu erklären?

Aufgrund der durch die Sensitivitätsprofile vorgegebenen Kopplung der Zellempfindlichkeiten an die momentane Blickrichtung (Abbildung 5.5) ist jeweils nur einer der Detektorsätze sensitiv:

- Der **mittlere** Satz, wenn die Blickrichtung ungefähr mit der beabsichtigten Fixationsrichtung übereinstimmt.
- Der **linke** Satz, wenn die Blickrichtung weiter als eine halbe Standardabweichung des typischen Fixationsfehlers nach *links* versetzt ist. (Wir wählen als retinales Bezugssystem die Aufsicht von der Pupille auf die Netzhaut, d.h. ein Fixationsfehler nach links im Sehraum verursacht eine Verschiebung der retinalen Projektion ebenfalls nach *links*.)
- Der **rechte** Satz, wenn die Blickrichtung weiter als eine halbe Standardabweichung nach *rechts* versetzt ist. (Verursacht eine Verschiebung der Projektion nach *rechts*.)

In allen anderen Fällen sind die Detektoren insensitiv, also funktionslos und damit effektiv nicht vorhanden. Sie liefern dementsprechend auch keinen Beitrag zum Cinematogramm. Da dies für die äußeren Detektorsätze häufiger geschieht als für den zentralen Satz (das ist dem in Abbildung 5.5 hinterlegten Histogramm der Blickrichtungen zu entnehmen), haben die Cinematogramme für diese Detektoren eine deutlich kleinere Amplitude. Alle Cinematogramme wurden auf die Gesamtzahl der Testreize normiert.

Im übrigen trägt jeder Detektor genau nur dann Anteile zum Cinematogramm bei, wenn das retinale Bild fehlerhaft in Richtung des Feldversatzes verschoben ist. Als Resultat liegen alle drei *cinematographisch gemessenen* Felder an etwa gleicher Position *im Sehraum*.

5.7.2 Ausgangsfelder

Das stimulierende Feld der Zellen in der Ausgangsstufe der *Gating*-Module setzt sich wie in den Abbildungen 5.3 und 5.4 dargestellt zu gewichteten Teilen aus den Aktivierungsfeldern der einzelnen Quellneuronen zusammen. Da die Aktivierungsfelder der einzelnen Detektorsätze an fast gleichen Stellen im Sehraum gemessen werden, überlagern sie zu einem kombinierten Feld, welches dem retinalen (ungestörten) Feld eines Einzeldetektors sehr ähnlich ist (vgl. schwarze und orange Kurve in Abbildung 5.7, rechts). Verglichen mit der Verbreiterung, die ohne ein neuronales Kompensationsglied auftritt (gestrichelte Kurve), kann von einer erfolgreichen neuronalen Korrektur gesprochen werden.

Jeder der Einzeldetektoren ist für einen beträchtlichen Teil der Zeit inaktiviert, nämlich immer dann, wenn die momentane Blickrichtung nicht in seinen Sensitivitätsbereich fällt. Im Gegensatz dazu werden in den Ausgangsneuronen der *Gating*-Module unabhängig von der Blickrichtung immer Signale registriert, vorausgesetzt, sie werden geeignet visuell stimuliert. Ihre Antwort weist also keine zeitlichen Ausfälle auf (vgl. jedoch Abschnitt 5.7.4). Eine an dieses Glied anschließende Verarbeitungsstufe kann demnach auf neuronalen Signalen aufbauen, die von residualen Augenbewegungen weitgehend unbeeinflusst sind. Dies erlaubt, die neuronale Repräsentation der visuellen Szene, und vor allem deren subjektive Wahrnehmung, stabil und scharf zu halten.

Zeitliche Kontinuität

5.7.3 Neuronaler Aufwand

Im vorliegenden Fall ist offenbar bereits mit *drei* gegeneinander verschobenen Detektorsätzen eine sehr gute Korrekturleistung zu erreichen. Maßgeblich sind hierbei die Größenverhältnisse von residualen Augenbewegungsamplituden und dem Durchmesser der rezeptiven Felder. In der Simulation waren diese beiden Größen miteinander vergleichbar; die simulierten Augenbewegungen lagen im Größenbereich des Stimulationsfeld-Durchmessers ($\sigma_{fix} = 5,27' \text{ arc}$, $\sigma_{SF} = 3,78' \text{ arc}$). Die Verteilung der Fixationsfehler sollte über ihre gesamte Breite hinweg durch Sensitivitätsprofile abgedeckt sein, um neuronale Ausfälle zu vermeiden (also Blickrichtungen, für die sämtliche Detektoren insensitiv sind, s. Abschnitt 5.7.4). Bei der Wahl dreier Detektorsätze betrug in der Simulation daher die Breite der einzelnen Sensitivitätsprofile ca. $\frac{1}{3}$ des Stimulationsfeld-Durchmessers (vgl. Abbildungen 5.5 und 5.7 rechts, orange Kurve).

Um über das gesamte menschliche Sehfeld hinweg eine neuro-

RF-Größe und Exzentrizität

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

nale Korrektur konstanter Güte zu ermöglichen, müßte dieses Verhältnis von 1 : 3 der Breiten von Sensitivitätsprofil und rezeptivem Feld des Einzeldetektors beibehalten werden. Da die Verteilungsbreite der Fixationsfehler konstant ist, die Größe der rezeptiven Felder jedoch von der retinalen Exzentrizität abhängt, können hierfür verschieden viele Sätze von Detektoren nötig sein: Das *Fan-In* der einzelnen *Gating*-Module muß entsprechend der Exzentrizität angepaßt werden. Insbesondere im zentralen Sehfeld wäre eine mehr als dreifache Multiplizität der Detektoren anzusetzen. Der neuronale Aufwand für eine Kompensationsschaltung dieser Art ist daher nicht unerheblich.

Für rezeptive Felddurchmesser weit oberhalb der Augenbewegungsamplitude ist allerdings nicht zu erwarten, daß sich die residualen Bewegungen noch negativ auf die neuronale Verarbeitung auswirken.¹¹ Auf eine neuronale Korrektur kann dann verzichtet werden. Ein Mechanismus wie der hier vorgeschlagene würde seine Wirkung also auf den zentralen und mittleren Sehfeldbereich beschränken. Neuronen, welche mit der Verarbeitung von Reizen im peripheren Sehfeld betraut sind, werden möglicherweise direkt von Zellen mit retinafesten rezeptiven Feldern gespeist. Die maximalen Bewegungen überstreichen in den Messungen von Brinksmeyer et al. (2001) ca. 1° Sehwinkel. Wird angenommen, daß der Einfluß der Augenbewegungen vernachlässigt werden kann, wenn deren Weite unterhalb des halben rezeptiven Felddurchmessers liegt, so wäre eine neuronale Korrektur bis etwa 10° retinaler Exzentrizität nötig (Felddurchmesser nach Murakami und Cavanagh, 2001).

5.7.4 Mikrosakkadische Suppression

Mikrosakkadeninduzierte Blindheit

Bei Betrachtung der Abbildung 5.5 fällt auf, daß die angesetzten Sensitivitätsprofile die extrem weiten, aber sehr selten auftretenden Fixationsfehler nicht abdecken. Da in diesen Fällen sämtliche Detektoren insensitiv sind, ist das Netzwerk bei extremen Blickschwankungen effektiv blind. Eine solche *mikrosakkadeninduzierte Blindheit* ist eine Konsequenz des vorgestellten Modells, wobei die Wahl der $1\frac{1}{2}$ -fachen Standardabweichung als Abschneidewert der Bewegungsamplitude rein willkürlich ist. Es liegen verschiedene Modellvariationen auf der Hand, um durch eine bessere Abdeckung der Blickrichtungsverteilung den Blindheitseffekt zu mindern:

¹¹Möglicherweise ist außer der Größe der rezeptiven Felder auch deren räumliche Dichte und Überlapp maßgebend für die visuelle Auflösung (Eurich und Schwegler, 1997).

Verbreiterung der individuellen Sensitivitätsprofile Mit breiteren Sensitivitätsprofilen könnte die gleiche Zahl von Eingangsneuronen einen größeren Amplitudenbereich abdecken. Dies ginge zu Lasten des Breitenverhältnisses von Sensitivitäts- und rezeptivem Feldprofil der Einzeldetektoren, und damit zu Lasten der Korrekturgüte.

Erhöhung des *Fan-Ins* Durch eine größere Zahl von Eingangsneuronen könnte die Abdeckung eines größeren Amplitudenbereichs durch zusätzlichen neuronalen Aufwand realisiert werden, bei gleichbleibend breiten Sensitivitätsprofilen und daher gleichbleibender Korrekturgüte.

Ausdehnung der beiden äußeren Sensitivitätsprofile Die in geringer Zahl auftretenden weiten Mikrobewegungen könnten dem Sensitivitätsbereich des jeweiligen äußeren Detektorsatzes zugeschlagen werden. Die Korrekturgüte im Bereich nahe des Fixationspunktes bliebe in diesem Fall unverändert, und wäre lediglich beim Auftreten eines entsprechenden Ereignisses herabgesetzt. Besonders großamplitudige Fehlfixationen würden dann möglicherweise vom Betrachter als (seltene) Bildschwankungen wahrgenommen, unter denen auch die Erkennungsleistung litte.

Es existieren Hinweise auf eine eingeschränkte Wahrnehmungsleistung bei residualen Augenbewegungen (vgl. Bridgeman und Palca, 1980). Möglicherweise stellt eine mikrosakkadeninduzierte Blindheit jedoch ein durchaus geeignetes Mittel dar, das visuelle System vor Beeinträchtigungen durch retinale Bildverschiebungen zu bewahren, welche die Leistung des neuronalen Korrekturapparates übersteigen. Der visuell-sensorische Datenstrom würde kurzfristig unterbrochen und so eine ungestörte Weiterbearbeitung des letzten „validen“ Reizes ermöglicht. Eine solche Unterbrechung oder Unterdrückung der neuronalen Verarbeitung ist in Verbindung mit Makrosakkaden nachgewiesen (sakkadische Suppression, Burr et al., 1994). Sie wirkt sich auf die Neuronen des magnozellulären Pfades aus. Diese generieren den vorherrschenden Eingang für Areale, die mit Bewegungswahrnehmung beschäftigt sind. Obgleich während einer Sakkade für eine Dauer von 10 bis 40 ms starke Bewegungsreize anliegen, nimmt ein Betrachter in Verbindung mit der Sakkade normalerweise keine Bewegung wahr.¹² Ein schnell über die Retina

*Sakkadische
Suppression*

¹²Eine Bewegung kann jedoch unter ungewöhnlichen Gegebenheiten wahrgenommen werden, etwa beim Betrachten eines einzelnen Lichtpunktes in der Dun-

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

hinwegfegender visueller Reiz, wie er auch während der Sakkade auftritt, aktiviert nur dann die bewegungssensitiven Nervenzellen, wenn er *nicht* durch eine Bewegung der Augen hervorgerufen wird (Thiele et al., 2002). Findet die Bewegung des Reizes jedoch zusammen mit einer Sakkade statt, so wird die Aktivität der bewegungssensitiven Neuronen unterdrückt.

Sehr wahrscheinlich ist die makrosakkadische Suppression einer der Mechanismen, die es dem Betrachter erlauben, seine Umwelt als stabil und unabhängig von seiner eigenen Augenbewegung wahrzunehmen. Auf neuronaler Ebene entlastet sie das visuelle System davon, das während der Sakkade unvermeidlich auftretende irritierende Bewegungssignal verarbeiten zu müssen. Eine ganz ähnliche Aufgabe käme im untersuchten Modell der mikrosakkadeninduzierten Blindheit zu: Durch sie würde der irritierende, die Korrekturleistung übersteigende fehlerbehaftete Stimulus einfach ausgeblendet. Mikrosakkadische Suppression wäre vor allem bei Neuronen des parvozellulären Pfades mit ihrem hohen räumlichen Auflösungsvermögen zu erwarten, und sie sollte vorwiegend bei weiten Mikrosakkaden auftreten. Sofern ein solcher Mechanismus existiert, sollte er sowohl in Zellableitungen wie auch in psychophysischen Untersuchungen feststellbar sein.

Leopold und Logothetis (1998) untersuchen die Aktivität von Nervenzellen im Kortex des Affen im Zusammenhang mit Mikroaugenbewegungen. Sie berichten von einer Ratenreduktion vieler Neuronen des striaten Kortex (V1) im Anschluß an eine Sakkade; ein Ergebnis, das mit dem vorgeschlagenen *Gating*-Mechanismus gut vereinbar ist (vgl. jedoch Martinez-Conde et al., 2000, 2002). Zeitlich parallel zu der verminderten Aktivität in V1 registrieren Leopold und Logothetis eine Erregung von Zellen in den extrastriaten Arealen V2 und V4. Bedenkt man die typische zeitliche Charakteristik neuronaler Antworten, die nicht in das Modell eingehen (Abschnitt 3.3.2), so sind starke transiente Reaktionen auf mikrosakkadisch verursachte Bildsprünge in der Retina nicht verwunderlich. Greschner et al. (2002) demonstrieren, daß eine durch Mikroaugenbewegungen erhöhte Synchronisation von Nervenzellsignalen die verlässlichere Schätzung des visuellen Reizes ermöglichen kann. Signaltransienten würden auch von den jeweils aktiven Komponenten der hypothetischen *Gating*-Module zu höheren Verarbeitungsstufen übertragen, ebenso wie die Modulation der neuronalen Verstärkung selbst transiente Antworten hervorrufen könnte.

kelheit (Burr et al., 1994). Der Leser mag den Effekt beim Betrachten einer Straßenlaterne in der Nacht beobachten.

Den residualen Augenbewegungen käme damit eine funktionelle Bedeutung im Sinne eines Auffrischungseffektes zu: mit jeder Mikrosakkade würde die visuelle Verarbeitung neu angestoßen. Im Zusammenhang mit Modellen, die eine neuronale Verarbeitung in der zeitlichen Domäne mittels eines *Spike-Rang-* oder *Latenzcodes* vorschlagen (Thorpe et al., 2001; van Rullen und Thorpe, 2001; Körner et al., 1999), ist das Setzen eines zeitlichen Referenzrahmens von entscheidender Bedeutung. Kupper et al. (2004) zeigen in Computersimulationen, daß sowohl sakkadengekoppelte Inhibition, wie auch kurzzeitige Relaxation von Neuronen geeignete Mechanismen sind, einen verwertbaren Spikelatenzcode im visuellen System aufrechtzuerhalten. Kommen die im Modell vorgeschlagenen *Gating-Module* in einem *Spike*-basierten Modell zum Einsatz, so setzen sie diesen Mechanismus ideal um: Die Neuronenpopulationen relaxieren während ihrer insensitiven Phase gegen das Ruhepotential. Werden sie anschließend durch eine entsprechende Mikroaugenbewegung sensitiviert, so befinden sich alle Neuronen des betreffenden Detektorsatzes in einem ähnlichen Zustand und können so einen verlässlichen Latenzcode generieren.

Funktion: zeitliche Referenz

5.7.5 Gewinnung des Blickrichtungssignals

In das Simulationsmodell gingen als wesentlicher Bestandteil experimentell ermittelte Daten über die während der Fixation vollführten Blickschwankungen ein. Es wurde explizit angenommen, daß dieser Parameter, eine makroskopische Zustandsgröße des Organismus, in einer Repräsentation als neuronales Signal lokal verfügbar ist. Als Ursprung eines solchen Signals kommen grundsätzlich drei Quellen in Betracht (von Helmholtz, 1910). Es sind dies

- eine **Efferenzkopie** (engl. *corollary discharge*) der motorischen Signale zur Steuerung der Augen (vgl. z.B. Wurtz und Sommer, 2004)
- **propriozeptive Rückmeldung** über die momentane Beanspruchung der Augenmuskeln
- der visuelle Reiz, genauer die Aktivität neuronaler **Bewegungsdetektoren**, ausgelöst durch die Verschiebung der retinalen Projektion als Folge residualer Augenbewegungen (Barbur, 1985; Murakami und Cavanagh, 1998, 2001)

Wir wollen diese Möglichkeiten vom modelltheoretischen Standpunkt aus beleuchten. Dabei soll es uns um die Eignung zur Steue-

Kapitel 5 Ein konzeptionelles Modell zur retinalen Schlupfkorrektur

rung eines neuronalen Kompensationsgliedes gehen. Für eine kritische Zusammenfassung experimenteller Befunde zu den einzelnen Punkten siehe Murakami und Cavanagh (2001).

Informationsgehalt

Eine Efferenzkopie kann nur dann zur Steuerung dienen, wenn die neuronalen Motorsignale auch wirklich Informationen über die residualen Augenbewegungen enthalten. Sollten diese ihre Ursache ganz oder teilweise in einer ungenauen Umsetzung durch die Augenmuskeln haben, so sind Motorsignale als Datenquelle ungeeignet. Für die propriozeptive Rückmeldung gilt, daß sie detailliert genug sein muß, um Schwankungen im Bereich einiger Winkelminuten zu übertragen.

Verarbeitungszeit

Eine prinzipielle Überlegung spricht aber für die beiden ersten Punkte: Hier kann die relevante Information bereits vorliegen, noch ehe die von der Augenbewegung beeinträchtigten Zellantworten über den retino-kortikalen Pfad eintreffen. Damit steht hinreichend Zeit zur Einstellung des Kompensationsmechanismus zur Verfügung. Chance et al. (2002) demonstrieren, daß eine Regulation der neuronalen Verstärkung über eine Anpassung der Hintergrundaktivität im Netzwerk erfolgen kann. Ein solcher Vorgang würde eine endliche, wenn auch vermutlich kurze „Einschwingzeit“ bis zur Erreichung eines quasistationären Zustands benötigen.

Ressourcen

Hingegen erscheint die Generierung der *Stell-* aus der *Istgröße*, nämlich die Extraktion der momentanen Blickrichtung aus dem davon beeinflussten visuellen Reiz, auf einen ersten Blick ungünstig. Wird ein einziger Verarbeitungsstrom angenommen, so müßte die Bewegungsauswertung auf einer dem Korrekturglied vorgeschalteten Stufe erfolgen, um gleichzeitig mit der Bildinformation verfügbar zu sein. Da als wahrscheinlicher Wirkort einer Schlupfkorrektur nur visuelle Areale mit hoher räumlicher Bildauflösung in Betracht kommen, diese also vermutlich im primären Sehkortex oder noch früher vollzogen werden muß, sind die verfügbaren Ressourcen begrenzt. Interessanterweise werden beim Hasen komplexe Mechanismen zur Bewegungs- und Richtungsdetektion bereits in retinalen Ganglienzellen gefunden (Vaney und Taylor, 2002; Fried et al., 2002; Taylor et al., 2000; Taylor und Vaney, 2002).

Tatsächlich verläuft die Verarbeitung im visuellen System jedoch schon früh in parallelen Strömen. Der schnelle magnozellige Pfad signalisiert hierbei bevorzugt transiente Komponenten des visuellen Reizes, die gut zur Bewegungsdetektion geeignet sind. Deren Auswertung in bewegungssensitiven kortikalen Arealen kann parallel zur Verarbeitung des hochauflösenden, aber langsameren parvozellulären Datenstroms erfolgen. Murakami und Cavanagh (2001) schließen aus ihren Untersuchungen eines Bewegungsnacheffektes

beim Affen auf eine Beteiligung des parietalen Pfades, insbesondere des Areals MT, bei der Verarbeitung des retinalen Schlupfes. Eine schnelle Integration der Aktivität aus bewegungssensitiven Arealen in die hochauflösende Repräsentation der visuellen Szene im primären Sehkortex stellt aus modelltheoretischer Sicht eine optimale parallele Aufgabentrennung dar. Sie wäre eine gute Quelle zur Steuerung der neuronalen Schlupfkorrektur.

Schließlich unterscheiden sich die drei möglichen Signalquellen in der Form, in der sie die nötige Information verfügbar machen: Während motorische Efferenzkopie und Propriozeption die momentane Blickrichtung vermutlich als absolute, wie auch als differenzielle Größe übermitteln können, beruht die stimulusbasierte Methode generell zunächst auf einer Messung der *Bewegung* und nicht der *Position*. Da als Steuersignal für eine Schlupfkorrektur die *Blickposition* relativ zum Fixationsziel benötigt wird, muß hier zusätzlich zeitlich integriert werden. Neuronale zeitliche Integratoren sind im visuellen System im Zusammenhang mit makroskopischen Sakkaden bekannt und können erfolgreich modelliert werden (Sklavos und Moschovakis, 2002; Goldman et al., 2002). Es ist sehr wahrscheinlich, daß vergleichbare Strukturen die zeitliche Integration residualer Augenbewegungen leisten können.

Bewegungsintegration

5.7.6 Extraktion aus dem visuellen Reiz

Im Falle der Bewegungsextraktion aus dem visuellen Reiz tritt als zusätzliche Schwierigkeit hinzu, solche Bewegungskomponenten, die von Augenbewegungen herrühren, zu unterscheiden von den Eigenbewegungen der Szenenbestandteile. Augenbewegungen sind dabei charakterisiert durch homogene Translation des gesamten Sehfeldes, während die Bewegung einzelner Objekte auf einen Ausschnitt begrenzt ist. Eine globale Auswertung von Bewegungen über das gesamte Sehfeld hinweg dürfte neuronal jedoch kaum zu bewerkstelligen sein. Die Untersuchung eines durch residuale Augenbewegungen verursachten Nacheffektes veranlaßt Murakami und Cavanagh (2001), einen lokal arbeitenden Mechanismus hierfür vorzuschlagen. Dieser beruht auf der Annahme, daß die langsamste auftretende Bewegung innerhalb einer Region des Sehfeldes eine Schätzung für die von Mikroaugenbewegungen herrührenden Verschiebungen ist. Diese werden *in der Regel* an Stellen auftreten, an denen keine äußeren Objektbewegungen vorliegen.

In Abbildung 5.8 skizziere ich ein neuronales Modul zur Gewinnung des Blickrichtungssignals, das auf diesem Vorschlag beruht.

Die Idee der Schätzung über die langsamste Bewegung wird darin über eine asymmetrische inhibitorische Kopplung regionaler Gruppen von Bewegungsdetektoren umgesetzt. Liegt hinreichende Aktivität in einer der Gruppen vor, so führt dies zur Inhibition aller Gruppen für höhere Geschwindigkeiten. Schließlich kann Aktivität nur in der Neuronengruppe überleben, welche die geringste vorliegende Geschwindigkeit repräsentiert, da diese als einzige keiner Inhibition unterliegt. Die so gewonnene Schätzung für die *Blickgeschwindigkeit* (und Bewegungsrichtung) wird anschließend von einem Ensemble neuronaler zeitlicher Integratoren in ein amplitudenkontinuierliches *Blickrichtungssignal* umgesetzt. Dieses Signal ist zur Ansteuerung des beschriebenen Schlupfkorrekturgliedes geeignet.¹³

5.8 Situativ gesteuerte Complex-Bildung

Die Existenz einer neuronalen Schlupfkorrektur selbst wird kontrovers diskutiert (Motter und Poggio, 1990; Gur und Snodderly, 1997; Brinksmeyer et al., 2001). Das in diesem Kapitel vorgestellte konzeptionelle und Simulationsmodell erlaubt diesbezüglich keine Entscheidung. Obgleich es mit den Ergebnissen von Brinksmeyer et al. kompatibel ist, kann weder seine neuronale Relevanz, noch seine strukturelle Übereinstimmung mit potentiellen Mechanismen im Sehsystem garantiert werden.

Seine Bedeutung liegt jedoch auch in der Formulierung wertvoller Ideen und Begriffe, mittels derer unser Verständnis zugrundeliegender neuronaler Steuerprozesse gebildet und ausgebaut werden kann. Ein wesentlich neues Konzept des vorgestellten Modells ist die situativ gesteuerte *Complex*-Bildung. Diese vereint hier die positiven Eigenschaften der klassischen *Complex*-Bildung sowie des dynamischen *Routings*, nämlich die Bildung invarianter Zellantworten durch hierarchische Merkmalsgruppierung einerseits, sowie andererseits die Vermeidung des Binde-Problems durch eine über das Sehfeld hinweg kohärente Kontrolle über die ausgeführten neuronalen Transformationen (vgl. Kapitel 2). Die situativ gesteuerte *Complex*-Bildung wird uns im folgenden Kapitel als neuronaler Modellbaustein von großem Nutzen sein. Darüber hinaus werden wir sehen, daß sie, im Gegensatz zu den klassischen Verfahren, das Problem der überlappenden Repräsentantenmengen aufzulösen vermag.

¹³Eine Computersimulation dieses Schätzmoduls erfolgte nicht. Ein Test sollte möglichst an Ganzfeldaufnahmen realer bewegter Szenen erfolgen, die reichhaltige Objekteigenbewegungen enthalten.

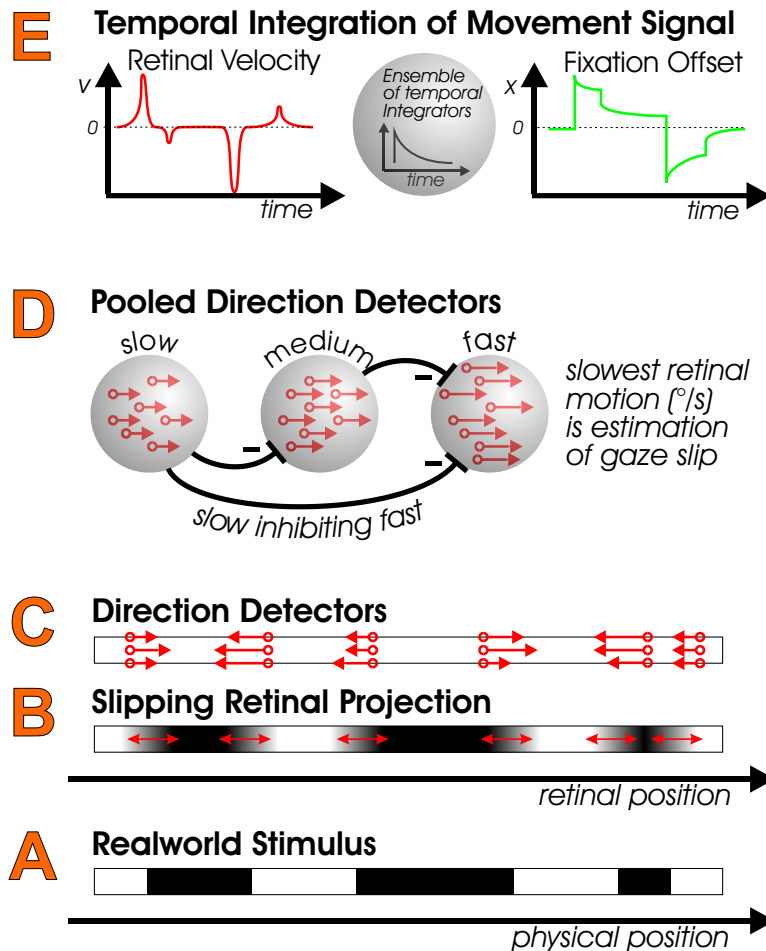


Abbildung 5.8: Modellskizze: Entwurf eines neuronalen Moduls zur Gewinnung des Fixationsfehlers aus dem bewegten retinalen Stimulus. Nach einem Vorschlag von Murakami und Cavanagh (2001). Verarbeitungsreihenfolge von unten nach oben: **A:** Echtweltstimulus, der durch die Augenlinse auf die Retina projiziert wird. **B:** Aufgrund von Mikroaugenbewegungen ist die retinale Projektion des Stimulus instabil. **C:** Der bewegte retinale Stimulus aktiviert bewegungssensitive Neuronen. **D:** Gruppenweise gerichtete Inhibition extrahiert den langsamsten Geschwindigkeitsanteil der Bewegung. Dieser stellt eine Schätzung für den durch Augenbewegungen verursachten retinalen Schlupf dar. **E:** Durch zeitliche Integration kann aus dem Bewegungssignal der Fixationsfehler rekonstruiert werden. Die Integration kann durch eine zeitliche Tiefpaßfilterung, etwa durch die elektrischen Eigenschaften der neuronalen Membran, angenähert werden.

Kapitel 6

Ein konzeptionelles Modell zur Entfernungsinvarianz¹

Ziel

In diesem Kapitel entwickle ich ein Modell zur entfernungsinvarianten Szenenrepräsentation. Es beruht auf der externen Steuerung in Abhängigkeit der Sehentfernung, deren Einfluß auf neuronale Antworten im Sehsystem experimentell belegt ist. Die Verwendung des zuvor eingeführten Konzepts der situativ gesteuerten *Complex*-Bildung führt mich zur Hypothesierung einer neuartigen Zellklasse, den *Entfernungs-Complex-Zellen*. Ich demonstriere die Leistungsfähigkeit des Modells in einer Computersimulation mit biologisch relevanten Randbedingungen, und für verschiedene Modellannahmen. In dem Modell ist eine entfernungsinvariante Erkennung, sowie eine Schätzung der physikalischen Größe betrachteter Objekte möglich. Es löst das Problem der überlappenden Repräsentantenmengen, und erlaubt das schnelle Erlernen neuer Objekte. Das Modell erfordert einen erhöhten neuronalen Aufwand, sowie die Verfügbarkeit einer Entfernungsschätzung. Ich diskutiere den experimentellen Nachweis möglicher *Entfernungs-Complex-Zellen* in einer aktuellen Forschungsarbeit.

6.1 Einleitung

Die invariante Objekterkennung ist eine bedeutende Leistung des menschlichen Sehsystems. Sie ermöglicht es, Bestandteile der visuellen Szene wiederzuerkennen, unabhängig davon, in welcher Konfiguration oder konkreten Erscheinung sie auftreten. Dies umfaßt die

¹Teile dieses Kapitels sind erschienen in Kupper und Eckhorn (2002b), sowie Kupper und Eckhorn (2002a).

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

unterschiedlichsten Sehbedingungen, im Bereich der geometrischen Invarianzen jedoch vor allem die Position im Sehfeld, die Größe von Objekten oder derer Bilder, sowie ihre Ansicht im dreidimensionalen Raum.

Optische und neuronale Transformation

Beim freien Sehen kann die Positionsinvarianz zweifellos zu ihrem größten Teil durch die Ausrichtung des Blickes auf das Objekt erreicht werden. Ein Beispiel dafür, daß diese Blickpositionierung nicht perfekt geleistet wird, doch neuronal berichtigt werden kann, haben wir im vorausgegangenen Kapitel gesehen. Für eine Anpassung der Bildgröße und des Ansichtswinkels bietet der optische Abbildungsapparat des Menschen hingegen keinerlei Hilfsmittel, abgesehen von einer Bewegung des gesamten eigenen Körpers im Raum, auf das Objekt zu oder um es herum. Eine weitverbreitete Annahme ist, daß die Erkennung im Sehsystem auf einem (impliziten oder expliziten) Vergleich mit wenigen gelernten Prototypen der Sehobjekte beruht. Ist diese Annahme korrekt, dann muß eine Abstimmung des visuellen Reizes auf die Sehbedingungen des gelernten Musters erfolgen: Beispielsweise muß der visuelle Reiz intern im gleichen räumlichen Maßstab wie der erlernte Prototyp eines Objektes repräsentiert sein, um einen direkten Vergleich zu erlauben. Diese Abstimmung des visuellen Reizes zum Zwecke des internen Vergleichs muß auf rein neuronalem Wege erfolgen. Ein solcher Vorgang stellt immense Anforderungen an das visuelle System, wenn man bedenkt, wie sehr sich verschiedene Erscheinungsformen desselben Objektes voneinander unterscheiden.

Neuronale Größentransformation

Entlang des ventralen visuellen Pfades muß eine äußerst effiziente neuronale Berechnung geometrischer Transformationen vonstatten gehen. Unter diesen nehmen die Größentransformationen eine Sonderstellung ein: Sie sind charakterisiert durch Veränderung der gesamten *Ausdehnung* eines visuellen Reizes, ohne typischerweise seine *Gestalt* zu betreffen, also die Anordnung seiner geometrischen Bestandteile. Die innere Struktur des Reizes bleibt erhalten, es ändert sich lediglich die Größenskala, auf der die einzelnen Bestandteile selbst und ihre relative Lage zueinander beschrieben werden.

Ich werde hier ein Modell der entfernungsinvarianten Verarbeitung im visuellen System entwickeln. Dabei werden die in den vorangegangenen Kapiteln gebildeten Konzepte der situativen Steuerung Verwendung finden. Im Mittelpunkt steht die neuronale Transformation, welche die Größenskala zum Mustervergleich des visuellen Reizes den äußeren Sehbedingungen anpaßt – vergleichbar mit einer Art „neuronalem Teleobjektiv“.

6.1.1 Der Begriff „Größe“

Im Folgenden wird der Begriff „Größe“ in verschiedenen Zusammenhängen Verwendung finden. Um den dabei typischerweise auftretenden Mehrdeutigkeiten aus dem Wege zu gehen, will ich hier zwei verschiedene Größenbegriffe voneinander abgrenzen.²

Physikalische Größe Die physikalische Größe oder räumliche Größe eines Objektes ist dessen Ausdehnung in der Umwelt. Sie ist die Gegenstandsgröße G im optischen Abbildungsgesetz (3.10). Die physikalische Größe ist die Meßgröße, die durch das Anlegen eines Maßbandes an das jeweilige Objekt ermittelt werden kann. Sie ist eine Länge, ihre Einheit ist Meter.

Retinale Größe Die retinale Größe eines Objektes ist die Ausdehnung seines Abbildes auf der Retina des Betrachters. Sie ist die Bildgröße B im optischen Abbildungsgesetz (3.10). Aufgrund der festen Augenform (Bildweite b) kann die retinale Größe entweder als eine Länge auf der Retina in Zentimetern angegeben werden, oder als optischer Öffnungswinkel in Grad Sehwinkel.

Zur Demonstration der beiden Begriffe kann wiederum die Abbildung 5.1 auf Seite 53 herangezogen werden. Der Leser fixiere für eine Minute den weißen Punkt, mit einem Sehabstand, der etwa der halben Leseentfernung entspricht. Danach richte er seinen Blick auf die Textseite in gewohnter Leseentfernung. Das „eingebrennte“ Bild behält dabei seine *retinale* Größe bei, das gedruckte Bild auf dem Papier seine *physikalische*.³

Selbstversuch

6.2 Größen- gegen Entfernungsinvarianz

Bei der Diskussion der größeninvarianten Objekterkennung bleibt oft unbeachtet, daß eine unterschiedlich große Projektion einer bekannten Form auf die Netzhaut von zwei verschiedenen Ursachen herrühren kann:

²Die Unterscheidung mag trivial erscheinen. Als abschreckendes Beispiel dafür, welche subtile Verständnisfehler bei einer inkonsequenten Trennung beider Begriffe auftreten, können hier jedoch die unzähligen falschen, wiewohl vordergründig einsichtigen Erklärungen der *Mondtäuschung* genannt werden. Eine ebenso kritische wie lesenswerte Übersicht liefert McCready (1999).

³Viele Betrachter würden spontan der Aussage „das Nachbild ist größer geworden“ zustimmen. Diese intuitive, ungenaue Verwendung des Begriffes „Größe“ sollte offensichtlich vermieden werden. Sie trägt zur Erschwerung der Diskussion über visuelle Größenwahrnehmung bei.

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

Sehabstand Ein Grund dafür, daß Sehobjekte eine variierende retinale Bildgröße besitzen ist, daß dasselbe Objekt in verschiedenen Abständen vom Betrachter auftaucht. Eine übliche Quelle für diese Art der Bildgrößenänderung ist eine Eigenbewegung des Betrachters oder der Objekte. Die dadurch hervorgerufenen Netzhautbilder werden als Instanzen ein und desselben realen Gegenstandes wahrgenommen. Selbst enorme Unterschiede in der Ausdehnung der retinalen Projektion kommen dem Betrachter nicht zu Bewußtsein. Vielmehr hat er eine Wahrnehmung der physikalischen Objektgröße, welche nicht von der Sehentfernung abhängig ist.⁴

Ich will für diese Art des invarianten Erkennens den Begriff *Entfernungsinvarianz* verwenden. Sie wird gänzlich unbewußt geleistet.

Physikalische Objektgröße Es gibt eine weitere Bedingung, die unterschiedlich ausgedehnte retinale Projektionen mit gleicher Form hervorruft: Es ist das Erblicken von Gegenständen mit wirklich verschiedener Größe. Erscheinen nacheinander verschiedene Gegenstände im Sehfeld des Betrachters, die zwar die gleiche Gestalt, jedoch unterschiedliche physikalische Größen besitzen, so werden diese stets als individuelle Gegenstände wahrgenommen. Als Beispiel hierfür möge ein echtes Fahrzeug und dessen Spielzeugmodell dienen. Wir erkennen darin völlig bewußt zwei physikalisch verschieden große Repräsentanten der Objektklasse *Fahrzeug* – zwei verschiedene Objekte.⁵

Unter gewöhnlichen Sehbedingungen werden diese beiden Typen der Größenvariation perzeptuell unterschieden. Das retinale Bild eines nahen Spielzeugautos kann durchaus identisch sein mit dem eines entfernten echten Fahrzeugs, und dennoch werden wir diese beiden Objekte nicht miteinander verwechseln. Erscheinen sie uns nacheinander in gleicher Entfernung, so schließen wir aus der einmal mehr und einmal weniger ausgedehnten retinalen Projektion, daß es sich um zwei Objekte unterschiedlicher Größe handelt. Sehen wir hingegen ein und dasselbe Auto nacheinander von nahe und von fern, so werden wir es als dasselbe Objekt erkennen und annehmen, daß es sich bewegt hat.

⁴Wir können jedoch durch einiges Training erlernen, den Anteil unseres Sehfeldes zu beurteilen, den ein betrachtetes Objekt einnimmt.

⁵Unter artifiziell kontrollierten Sehbedingungen kann diese bewußte Wahrnehmung eingeschränkt sein. Hierauf beruhen die im Film verwendeten Trickeffekte mit verkleinerten Modellen.

6.3 Entfernungsschätzung im visuellen System

Invariante Erkennung wird für variierenden Sehabstand geleistet. Sie tritt hingegen nicht für variierende physikalische Größe auf.

Eine entsprechende Unterscheidung muß sich auch in unserer Modellvorstellung der beteiligten neuronalen Mechanismen wiederfinden: Ein Modell, das ausschließlich auf der Verarbeitung *retinaler* visueller Merkmale aufbaut, kann nicht in der Lage sein, diese Trennung zu leisten.⁶ Das gilt sowohl für die Domäne der Modelle, die dynamisches *Routing* verwenden, um zweidimensionalen Bildtransformationen neuronal zu implementieren (Anderson und van Essen, 1987; Olshausen et al., 1993; Salinas und Abbott, 1997), wie auch für die der Modellierung mittels statischer hierarchischer Merkmalsgruppierung (etwa Fukushima, 1980; Riesenhuber und Poggio, 1999b; Mel und Fiser, 2000).

Die räumliche Größe ist eine inhärente physikalische Eigenschaft aller Objekte unserer Umwelt. Sie ist daher ebenfalls eine inhärente Eigenschaft der Objekte unserer Begriffswelt: Zwei physikalisch verschieden große Gegenstände können niemals identisch sein. Diese Tatsache findet offenbar auch bei der größeninvarianten visuellen Wahrnehmung Beachtung. Ich formuliere daher die folgende Hypothese als Grundannahme für ein Modell zur größeninvarianten Wahrnehmung durch situative Steuerung:

**Räumliche Größe als
Objekteigenschaft**

Die räumliche Ausdehnung ist eine physikalische Objekteigenschaft. Sie ist daher gleichfalls eine inhärente Eigenschaft aller Objekte unserer visuellen Wahrnehmung, und wir berücksichtigen sie bei der Identifikation von Gegenständen. Dabei gewinnen wir die räumliche Größe eines Seobjektes unter Berücksichtigung der momentanen Sehentfernung.

Meines Wissens existiert kein Modell der größeninvarianten Erkennung, das dieser Annahme gerecht wird.

6.3 Entfernungsschätzung im visuellen System

Die beiden Typen der Größenvariation (verursacht durch veränderte Sehentfernung einerseits, und durch verschiedene physikalische

⁶Wir werden allerdings in Abschnitt 6.8.2 sehen, daß basierend auf dem retinalen Reiz eine Entfernungsschätzung möglich ist, die dann eine Trennung erlaubt.

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

Größe andererseits) sind dann für ein Verarbeitungssystem leicht voneinander zu unterscheiden, wenn zusätzlich zum retinalen Abbild eines Sehobjektes auch dessen Entfernung vom Betrachter bekannt ist: Der Zusammenhang zwischen Bildgröße, Objektentfernung und Objektgröße folgt dem einfachen Proportionalitätsgesetz in Gleichung 3.10 (optisches Abbildungsgesetz, Seite 31).

Hinweise darauf, daß auch das visuelle System die zusätzliche Entfernungsinformation für den Seh- und Erkennungsvorgang verwendet, liefern sowohl psychophysische Experimente am Menschen, wie auch Zelleitungen beim Affen.

6.3.1 Psychophysische Befunde

*Größe, Tiefe,
Entfernung*

Psychophysische Untersuchungen weisen darauf hin, daß an der Beurteilung der Größe, der dreidimensionalen Gestalt, sowie der Entfernung visueller Objekte separate neuronale Module mitwirken (Brenner und van Damme, 1999; van Damme und Brenner, 1997). Beim Sehen in natürlicher Umgebung sind die Einschätzungen dieser physikalischen Eigenschaften gemeinhin konsistent und entsprechen den tatsächlichen Gegebenheiten. Aus visuellen Schätzaufgaben unter kontrollierten Laborbedingungen schließen Brenner und van Damme jedoch, daß die einzelnen neuronalen Module zur Beurteilung von Größe, Tiefe und Entfernung weitestgehend unabhängig voneinander arbeiten. Die Übereinstimmung der Schätzungen unter gewöhnlichen Sehbedingungen rührt ihrer Ansicht nach von der Konsistenz der im visuellen Reiz vielfältig vorliegenden Hinweise über die dreidimensionale Struktur der Szene und einzelner Objekte her.

Entfernung

Die Experimente von Brenner und van Damme lassen darüberhinaus darauf schließen, daß die getrennt verlaufenden Schätzungen auf *gemeinsamen neuronalen „Meßgrößen“* aufbauen. Eine dieser Meßgrößen ist offenbar eine neuronale Repräsentation der Sehentfernung (van Damme und Brenner, 1997). Informationen über die Entfernung eines Gegenstands vom Betrachter kann das visuelle System aus einer Vielzahl von Quellen gewinnen. Zu den Hinweisen (engl. *cues*) auf die Objektentfernung zählen beispielsweise die Vergrößerung, also der von den Achsen der Augen eingeschlossene Winkel, die Akkomodation der Linsen oder der Sehwinkel relativ zum Horizont (Ooi et al., 2001). Sie können jedoch auch als *bildhafte Hinweise* aus Merkmalen der zweidimensionalen Projektion gewonnen werden, etwa aus der perspektivischen Verkürzung der Längen, dem Kontrastverlauf, dem räumlichen Frequenzgradienten von Oberflä-

6.3 Entfernungsschätzung im visuellen System

chentexturen, aus der gegenseitigen Verdeckung oder aus der Parallaxe bewegter Szenenbestandteile.

Möglicherweise liegt die Information über die reale, physikalische Größe eines Sehobjektes schon in frühen visuellen Verarbeitungsstufen vor. In einer verzögerten Vergleichsaufgabe mit menschlichen Versuchspersonen kann ein als Zielreiz dienendes Gittermuster durch die Präsentation eines ebenfalls gitterförmigen Zwischenreizes maskiert werden. Bennett und Cortese (1996) finden, daß der Effekt der Maskierung stets dann am geringsten ist, wenn beide Gitter die gleiche *räumliche* Ortsfrequenz aufweisen, unabhängig von ihrer individuellen Präsentationsentfernung. Entscheidend für den Maskierungseffekt ist die Gleichheit der Frequenzen gemessen in absoluten Raumkoordinaten (*cyc/m*), *nicht* in retinalen Koordinaten (*cyc/°*). Die Autoren schließen daraus, daß eine Repräsentation wirklicher, physikalischer Längen in visuellen Stimuli bereits auf primärvisueller Ebene, also im gestreiften Kortex (V1) vorliegt. Zwar muß diese Aussage als eine Vermutung gewertet werden, da ein zeitlich verzögerter Vergleich zweier Gitterreize sicher auch höhere visuelle Verarbeitungsstufen involviert (der *Reverse Hierarchy Theory* zufolge nimmt die bewußte Wahrnehmung ihren Ausgang am oberen Ende des Verarbeitungspfades, Hochstein und Ahissar, 2002). Andererseits findet sich hier jedoch ein klarer Hinweis auf die Relevanz physikalischer Größe als Parameter der neuronalen Verarbeitung.

*Physikalische Größe
als neuronaler
Parameter*

Psychophysische Befunde wie diese können als Hinweis darauf gewertet werden, daß eine Art *Kalibrierung* von Teilen des visuellen Systems stattfindet, abhängig von der Entfernung des betrachteten Objektes. Diesem Effekt könnte auf Neuronenebene die situative Steuerung von Codierungseigenschaften zugrundeliegen, wie sie in Kapitel 3 besprochen wurde. Tatsächlich bestätigen jüngere Untersuchungen auch bei elektrischen Zellableitungen die Vorstellung einer entfernungsabhängigen Verarbeitung im Sehsystem.

*Kalibrierung des
Sehsystems?*

6.3.2 Physiologische Befunde

Distanzabhängige Modulation der Feuerraten von Einzelzellen wird in hohem Umfang (64–85 % der gemessenen Neuronen) in den visuellen Kortexarealen V1, V2 und V4 des Affen gefunden (Dobbins et al., 1998; Rosenbluth und Allman, 2002, s. Abbildung 6.1). Diese Entfernungsmodulation ist damit eine äußerst weitverbreitete neuronale Eigenschaft im ventralen Sehpfad. Während die Abhängigkeit der Zellantwort (engl. *tuning*) von zweidimensionalen Stimuluseigenschaften wie Kontrast, Kantenorientierung und Raumfrequenz

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

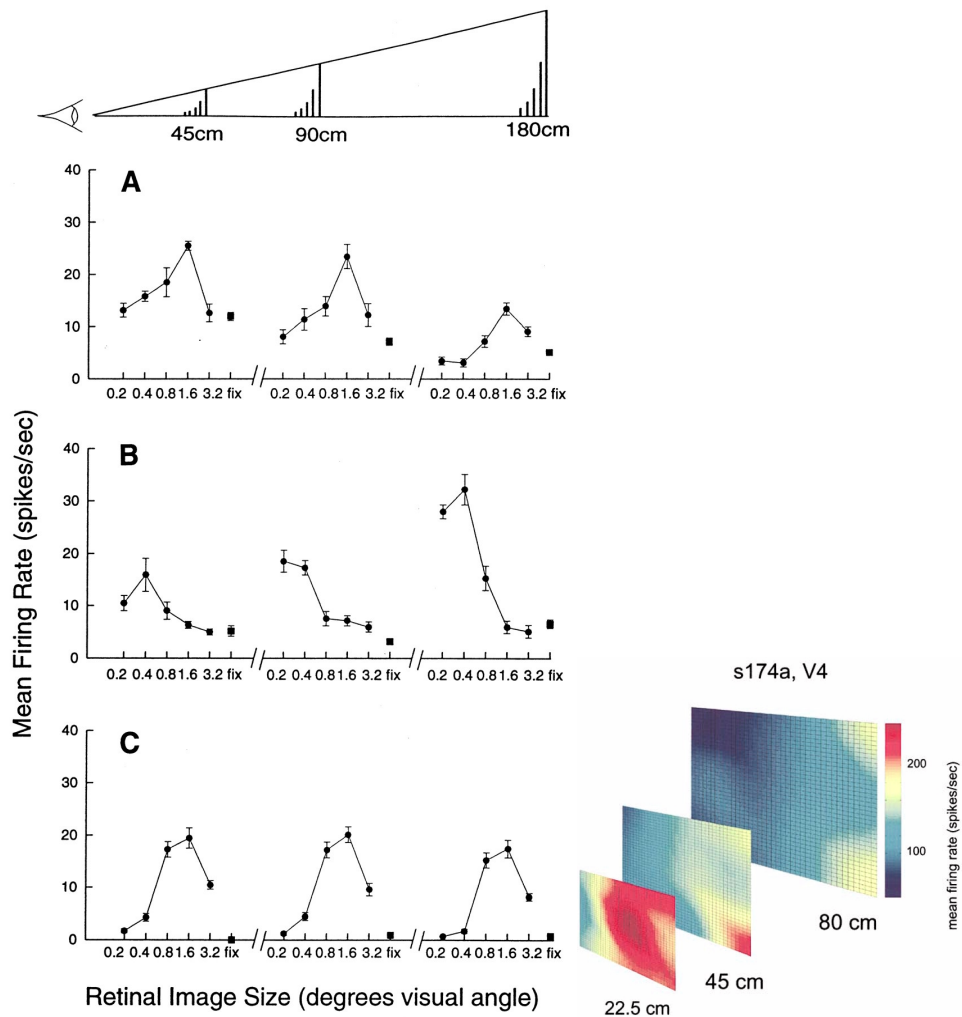


Abbildung 6.1: Distanzmodulation der Feuerraten von Einzelneuronen im visuellen Affenkortex. **Links:** Typische Charakteristik je eines als *Nahzelle* (A), *Fernzelle* (B) oder als *unmoduliert* (C) klassifizierten Neurons (aus Dobbins et al., 1998). Als Rechteck dargestellte Meßpunkte geben die Spontanrate bei Fixation ohne visuelle Reizung an. **Rechts:** Typische distanz- und blickrichtungsabhängige Modulation eines Einzelneurons (aus Rosenbluth und Allman, 2002). Distanzmodulation wird für 64–85 % der Neuronen beobachtet.

6.3 Entfernungsschätzung im visuellen System

unverändert bleibt, weisen die Neuronen eine zusätzliche Modulation ihrer Feuerrate mit der Fixationsentfernung auf (Dobbins et al., 1998; Brinksmeyer et al., 2002; Rosenbluth und Allman, 2002). Selbst in Abwesenheit eines visuellen Stimulus zeigt die Spontanaktivität der Nervenzellen ein solches Verhalten.

Dies gilt für mehr als zwei Drittel aller Neuronen in V1, V2 und V4 des Affen. Welche grundlegende Funktion könnte die Modulation der Antwort einer solchen Vielzahl von Neuronen im ventralen Sehpfad erfüllen?

6.3.3 Interpretation

Vermutlich wegen der erhöhten Schwierigkeit bei der Stimulation im dreidimensionalen Raum testen alle experimentellen Untersuchungen nur drei verschiedene Präsentationsentfernungen im Nahraum bis 180 cm.⁷ Aufgrund der Lage des dortigen Antwortmaximums werden die entfernungsmodulierten Zellen üblicherweise in drei Kategorien eingeordnet, Zellen mit Präferenz für nahe, mittlere oder ferne Stimuli. Multimodale Entfernungscharakteristiken wurden nicht beobachtet (Dobbins et al., 1998; Rosenbluth und Allman, 2002; Brinksmeyer et al., 2002, persönliche Kommunikation).

Einteilung in Nah-, Mittel- und Fernzellen

Die ermittelten Daten sind bei weitem nicht detailliert genug, um eine Aussage über die genaue Form der Distanzabhängigkeit zu erlauben. Ich möchte hier drei zusätzliche Annahmen machen, die mit den verfügbaren Informationen kompatibel sind, jedoch darüber hinausgehen. Sie sind als Interpretation der experimentellen Befunde anzusehen.

1. Die Entfernungscharakteristik (*Entfernungs-Tuning*) einer Zelle weist ein einzelnes Maximum begrenzter Breite bei einer bevorzugten Sehentfernung auf.
2. Das Phänomen der Entfernungsmodulation ist nicht auf den Raum unterhalb von 180 cm Sehentfernung beschränkt.
3. Die Verteilung der bevorzugten Entfernungen über die Zellpopulation hinweg ist unbekannt. Die experimentelle Klassifizierung als Nah-, Mittel- oder Fernzellen ist eine Folge der Wahl dreier Testentfernungen. Insbesondere kommen mehr als drei bevorzugte Distanzen in Betracht.

Vergleiche hierzu Abbildung 6.2. Über Form, Lage und Breite der

⁷Dobbins et al. (1998): 45, 90, 180 cm, Rosenbluth und Allman (2002): 22,5, 45, 80 cm, Brinksmeyer et al. (2002): 45, 90, 180 cm sowie 45, 78, 135 cm.

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

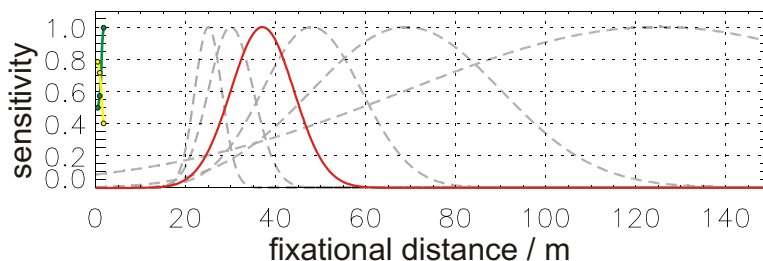


Abbildung 6.2: Beispiel für eine der im Modell verwendeten Entfernungskarakteristiken (rot). Eine Zelle mit dieser Charakteristik bevorzugt Fixation in der Entfernung 37 m. Für Reize in Entfernungen unterhalb 17 m oder oberhalb 57 m ist sie praktisch insensitiv. In grau hinterlegt sind die angenommenen Entfernungspräferenzen anderer Zellen der Population.

Maxima der so angenommenen Entfernungskarakteristiken lassen die verfügbaren Daten keine näheren Rückschlüsse zu (Messpunkte aus Dobbins et al., 1998, sind am linken Rand der Abbildung eingetragen). Sie stellen Modellparameter dar, die nach funktionellen Erwägungen gewählt werden. Weiteres hierzu behandelt der Abschnitt 6.5.1.

Erweitertes Merkmalskonzept

Die obigen Annahmen erlauben eine Interpretation der experimentellen Ergebnisse, die das klassische Konzept der visuellen Merkmalsdetektoren erweitert, nach dem einzelne Nervenzellen im visuellen System selektiv auf bestimmte visuelle Teilstimuli reagieren:

Die Ergebnisse können dahingehend interpretiert werden, daß visuelle Neuronen neben ihren zweidimensional bevorzugten Stimuli zusätzlich auch eine bevorzugte Sehentfernung codieren.

Abbildung 6.3 veranschaulicht dieses Prinzip. Ihm kann im Rahmen der Generierung visueller Entfernungsinvarianz eine entscheidende funktionelle Bedeutung zukommen:

Vermutete Funktion

Ich vermute, daß eine neuronale Distanzinformation dazu verwendet wird, aus den verfügbaren visuellen Merkmalsdetektoren eine Untermenge mit geeigneter Entfernungspräferenz auszuwählen. Die Repräsentation des jeweiligen Sehobjektes baut dann auf den Signalen nur solcher Neuronen auf, die auch die aktuelle Sehentfernung bevorzugen.

Im Rahmen des Konzeptes der situativ gesteuerten Anpassung aktivierender Felder (Abschnitt 3.4) schlage ich ein Modell zur entfer-

6.4 Problemstellung: Entfernungsinvarianz

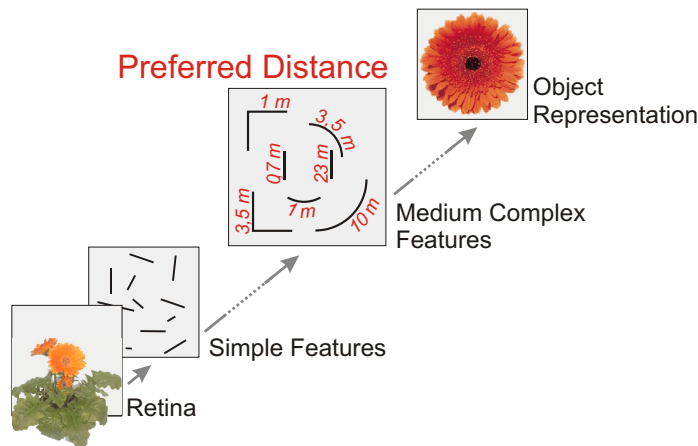


Abbildung 6.3: Interpretation der experimentell beobachteten Distanzmodulation: Visuelle Merkmalsdetektoren codieren neben ihren zweidimensionalen bevorzugten Stimuli auch eine bevorzugte Sehentfernung, also etwa einen Kreisbogen bei Fixation in 3,5 m Entfernung.

nungsinvarianten Objektrepräsentation vor, das auf diesem Prinzip beruht. Die beschriebenen psychophysischen und elektrophysiologischen Befunde werden in dieses Modell einfließen.

6.4 Problemstellung: Entfernungsinvarianz

Gesucht ist ein Modellvorschlag für ein Teilsystem der visuellen Objekterkennungshierarchie, das eine Distanzkalibrierung im erwähnten Sinne erlaubt. Aufbauend auf entfernungsmodulierten Signalen visueller Merkmalsdetektoren soll es eine Repräsentation der visuellen Szene erzeugen, welche eine entfernungsunabhängige Objekterkennung ermöglicht. In dieser Repräsentation soll eine Schätzung der physikalischen Größe eines betrachteten Gegenstandes möglich sein.

Es geht an dieser Stelle nicht um die Ausarbeitung einer kompletten Objekterkennungshierarchie. Vielmehr wollen wir eine Systemkomponente entwerfen, die den neuronalen Übergang einer entfernungsvarianten zu einer invarianten Repräsentation implementiert. Die Zielrepräsentation soll insbesondere genügend generisch sein, um die Ankopplung einer separaten Objekterkennungsstufe zu erlauben, ohne die Auswahl der hierfür zur Verfügung stehenden Mo-

**Wirkungsort und
Rahmen des Modells**

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

delle zu sehr einzuschränken. Im Hinblick auf den hierarchischen Charakter der Großzahl existierender Modellvorstellungen erscheint es sinnvoll, das gesuchte Verfahren funktionell auf der Ebene einfacher oder mäßig komplexer visueller Merkmale anzusiedeln. Gelingt es nämlich, Detektoren für grundlegende Konstituenten der visuellen Verarbeitung wie etwa orientierte Kanten oder Texturbestandteile mit entfernungsinvarianten Eigenschaften auszustatten, so wird diese Charakteristik an alle nachfolgenden Verarbeitungsstufen „vererbt“. Auf diese Weise arbeitet ein beliebiges darauf aufsetzendes Erkennungsverfahren entfernungsinvariant. Das im folgenden vorgestellte Modell ist für diese Form der Invarianzgenerierung geeignet.

Neuronale Kompensation

Eine entfernungsunabhängige Erkennung von Sehobjekten ist erreichbar, wenn die erzeugte Szenenrepräsentation derart von der Fixationsentfernung abhängt, daß die vom optischen Abbildungsvorgang herrührende Entfernungsabhängigkeit der retinalen Projektion gerade kompensiert wird (vgl. hierzu Abschnitt 3.5).

Erprobte Konzepte

In diesem Sinne ist das gesuchte System der im letzten Kapitel vorgestellten Schlupfkorrektur ähnlich. Wir haben nun die Möglichkeit, die dort bereits erprobten und verstandenen Modellbausteine einzusetzen: Die Konzepte des gesteuerten neuronalen *Gatings* und der Sensitivitätsprofile werden uns auch zur Implementation der gesuchten Entfernungskalibrierung dienlich sein.

Der Schwelleneffekt ist aus den in Abschnitt 4.4 beschriebenen Gründen kein geeignetes Konzept zur Modellierung von Entfernungsinvarianz. Diese erfordert die Verschiebung aktivierender Gebiete über große Distanzen im Sehfeld hinweg, die vom Schwelleneffekt nicht geleistet werden kann.

Entfernung als Steuergröße

Wie zuvor gehen wir auch hier davon aus, daß der Steuerparameter – in diesem Fall die aktuelle Fixationsentfernung – als neuronale Größe explizit vorliegt. Als mögliche Quellen für dieses Signal kommen eine ganze Reihe physiologischer oder visuell-bildhafter Hinweise (engl. *cues*) in Betracht (vgl. Abschnitt 6.3.1). Die Frage, auf welche Art diese einzelnen Hinweise ausgewertet und über verschiedene Sinnesmodalitäten hinweg zu einem Gesamtsignal integriert werden, ist Gegenstand aktueller Forschung (Landy et al., 1995; Atkins et al., 2001; Hillis et al., 2002). Hier wollen wir lediglich annehmen, daß neben der Aktivierung einfacher visueller Merkmalsdetektoren auch die Sehentfernung als eindimensionale Größe geeignet neuronal verfügbar ist. Wir werden sie wie vorgeschlagen zur Auswahl einer Untermenge von Detektoren einsetzen.

6.5 Modell

Ausgangspunkt für das Modell sind visuelle Merkmalsdetektoren wie sie typisch im ventralen Sehpfad gefunden werden. Ihre rezeptiven Felder seien über den zweidimensionalen Sehraum verteilt. Die genaue Art der bevorzugten Merkmale muß im Modell nicht vorgegeben werden. Damit müssen hinsichtlich der Ebene der visuellen Verarbeitungshierarchie, in der das Modell angesiedelt sein kann, keine Beschränkungen auferlegt werden. Als Kandidaten für diese Nervenzellen kommen sowohl einfache orientierungs- und richtungsselektive Detektoren in Betracht, wie auch Neuronen, die Merkmalskombinationen mittlerer Komplexität codieren, etwa Kantenstops, Konturelemente wie Ecken und Bögen, oder Oberflächen- und Texturbestandteile. Es soll sich bei den repräsentierten Merkmalen jedoch um Elemente handeln, die deutlich unterhalb der Ebene ganzer visueller Objekte liegen.

Merkmalsdetektoren

Diese Nervenzellen sollen die beschriebene Ratenmodulation mit der Fixationsentfernung aufweisen. Wir werden diese Eigenschaft im Modell explizit erzeugen, indem wir uns des bereits in Abschnitt 5.3.2 eingeführten Konzeptes der *Sensitivitätsprofile* bedienen.

Distanzmodulation

Als Beschreibung des neuronalen Antwortverhaltens dient weiterhin das funktionale *Graded-Response-Modell* mit der einfachen stückweise linearen Übertragungsfunktion in Abbildung 4.1, S. 36 rechts. Sie wird durch die zwei Parameter Antwortschwelle und Verstärkung charakterisiert.

Übertragungsfunktion

6.5.1 Entfernungskarakteristik

Experimentelle Untersuchungen weisen neuronale Distanzmodulation im ventralen Pfad für Sehentfernungen unterhalb 1,80 m nach, machen jedoch keine Aussage über größere Entfernungen. Ein Modell der Entfernungsinvarianz sollte sinnvollerweise auch die invariante Erkennung weiter entfernter Objekte einschließen. Einerseits ist für die Ausarbeitung des Modellierungskonzeptes die konkrete Wahl der Modellparameter von untergeordneter Bedeutung; im Blickpunkt sollte eine überschaubare Struktur des Modells stehen. Andererseits ist es wünschenswert, das Modell in einem Arbeitsbereich zu betreiben, der normalen Sehbedingungen entspricht. Das Modell wird auf Sehentfernungen unterhalb 150 m operieren.

Operationsbereich

Bei der Entfernungskarakteristik (engl. *tuning*) der Zellen handelt es sich um *Sensitivitätsprofile*, wie sie in Abschnitt 5.3.2 besprochen wurden. Das Konzept kam bereits im Modell der retinalen Schlupfkorrektur zum Einsatz.

Sensitivitätsprofile

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

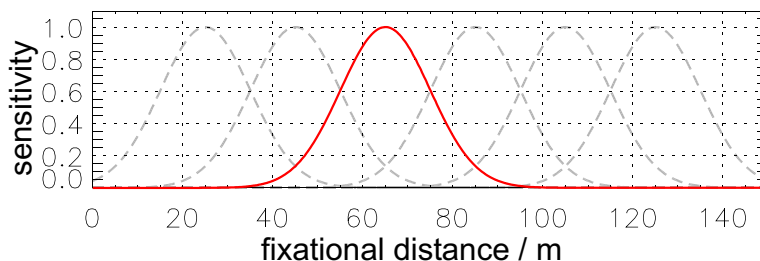


Abbildung 6.4: Sensitivitätsprofil: Einfache regelmäßige Verteilung der Entfernungseigenschaften für sechs Zellklassen über den betrachteten Entfernungsbereich. Abstand der Maxima: 20 m, Breite: Standardabw. $\sigma = 10$ m. Rot hervorgehoben die Entfernungseigenschaft für die dritte Zellklasse. (Hervorhebung nur zur besseren Übersicht, alle Klassen sind gleichwertig.)

Die Entfernungseigenschaft eines visuellen Neurons ist ein Sensitivitätsprofil. Der zugehörige globale Steuerparameter ist die momentane Fixationsentfernung.

Diskrete Entfernungspräferenzen

Aus Gründen der Anschaulichkeit und strukturellen Einfachheit nehmen wir an, daß sich die verwendeten Merkmalsdetektoren hinsichtlich ihrer Entfernungspräferenz in eine kleine Zahl von Klassen einteilen lassen. Dies entspricht dem Vorgehen in experimentellen Arbeiten (Abschnitt 6.3.3). Die Neuronen innerhalb einer jeden Klasse haben identische Entfernungseigenschaften, also identische Sensitivitätsprofile. Dies stellt nur eine der möglichen Annahmen für die Verteilung der Entfernungspräferenzen dar, auf die wir uns hier aber beschränken werden (vergleiche auch Abschnitt 6.8.2). Eine einfache Verteilung der Profile für sechs diskrete Zellklassen ist in Abbildung 6.4 dargestellt. Dabei unterscheiden sich die Charakteristiken der einzelnen Klassen nur in ihrer bevorzugten Fixationsentfernung, haben jedoch identischen Verlauf. Die Form der charakteristischen Kurven wird so einfach wie möglich gewählt. Entsprechend der Lokalitätsannahme in Abschnitt 6.3.3 werden sie als einfache Gaußfunktionen mit Überlapp angesetzt.

Sensitivität

Die Ordinate in Diagramm 6.4 gibt die *Zellsensitivität* S an. Hierbei handelt es sich um die multiplikative Modulation der neuronalen Verstärkung, entsprechend Gleichung 5.1, S. 59. Sie liegt im Bereich zwischen 0 (völlige Insensitivität) und 1 (maximale Verstärkung).

Eine Insensitivierung bis hin zum Faktor 0 für alle Neuronen ist eine vereinfachende Modellannahme. Dobbins et al. (1998, vgl. Abbildung 6.1) finden Modulationen um etwa den Faktor 2 ($S \in [0,5, 1]$). Rosenbluth und Allman (2002, selbe Abbildung) finden für den reinen Entfernungseffekt mindestens Faktor 2,5 ($S \in [0,4, 1]$). Für

die zusätzliche Modulation mit der Blickrichtung fällt die Rate hier sogar fast auf 0 ab ($S \in [0, 1]$).⁸ Für Neuronen, die keine vollständige Insensitivierung erreichen, kann ein entsprechender Effekt durch Annahme einer nachgeschalteten Schwellenoperation, etwa im postsynaptischen Neuron, erzeugt werden. Ich verzichte hier auf eine explizite Modellierung dieses Vorgangs und gehe von einer Sensitivität S im Bereich $[0, 1]$ aus.

Die Distanzmodulation der Merkmalsdetektoren stellt im Modell den grundlegenden Mechanismus dar, auf dessen Basis entfernungs-invariante Zellantworten generiert werden. Daher wird die konkrete Form der Entfernungscharakteristiken einen Einfluß auf die Leistung des Modells haben. Die in Abbildung 6.4 gezeigte Verteilung stellt eine Annahme dar; die Anzahl der Klassen, sowie Form und Verteilung der Kurven sind Modellparameter. Ich verwende die dargestellte und ähnliche Verteilungen von Entfernungscharakteristiken in einer Computersimulation des beschriebenen Modells (Abschnitt 6.6). Ein Vergleich der Modelleistung für verschiedene Verteilungen findet in Abschnitt 6.7 statt.

*Verschiedene
Charakteristiken*

Abbildung 6.5 stellt die bis hierhin beschriebenen Modellkomponenten schematisch zusammen. Sie ist die erste Abbildung der Serie 6.5–6.9, welche die wesentlichen Modellbestandteile präsentiert. Die Merkmalsdetektoren sind in Abbildungsteil B durch einige ihrer bevorzugten Merkmale angedeutet (Kantenelemente; die grauen Quadrate bedeuten identische Ausschnitte des Sehfeldes). Sie werden anhand ihrer bevorzugten Fixationsentfernung in diskrete Klassen eingeteilt. Die experimentell beobachtete Distanzmodulation wird im Modell erzeugt, indem für jede dieser Klassen die Verstärkung der Übertragungsfunktion entsprechend ihres Sensitivitätsprofils eingestellt wird. Die Sensitivitätsprofile der einzelnen Klassen sind in Abbildungsteil A angedeutet; sie haben die Form gegeneinander verschobener Gaußfunktionen, vgl. Abbildung 6.4. Ihr globaler Steuerparameter ist die momentane Fixationsentfernung. Der Verlauf der Sensitivitätsprofile entspricht der zu erzeugenden Distanzcharakteristik der Merkmalsdetektoren. In der Abbildung ist außerdem beispielhaft eine aktuelle Sehentfernung vorgegeben: Die Zellklasse mit halbnaher Entfernungspräferenz ist bei der aktuellen Sehentfernung sensitiviert, alle anderen Zellklassen sind insensitiv. Aufgrund dieser entfernungsabhängigen Sensitivierung wird stets diejenige Klasse von Merkmalsdetektoren die visuelle Szene repräsentieren, welche die augenblickliche Sehentfernung bevorzugt.

⁸Die Werte sind den Originaldiagrammen in Abbildung 6.1 entnommen.

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

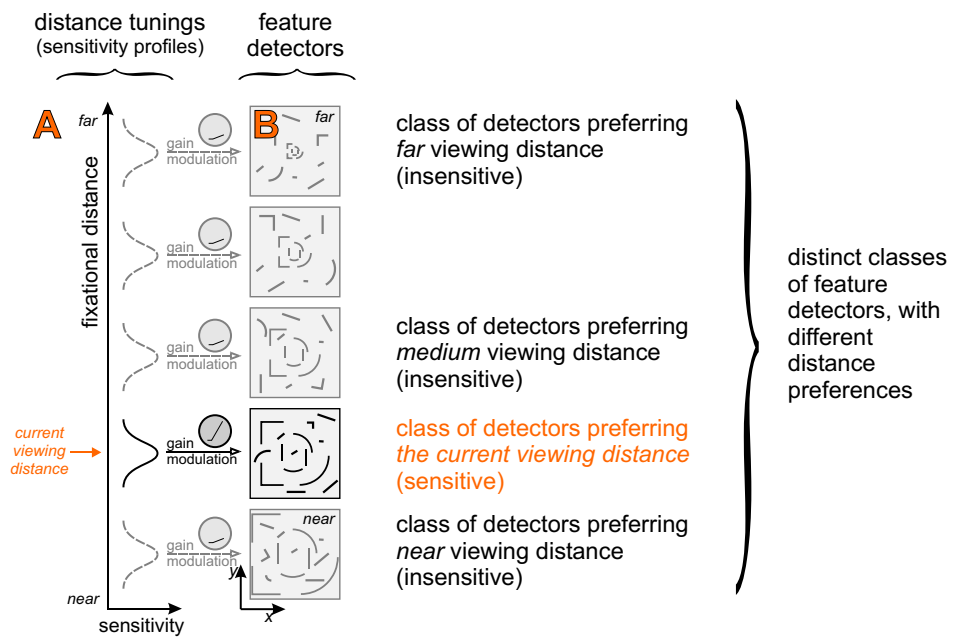


Abbildung 6.5: Generierung der Distanzmodulation im Modell (schematisch). **A:** Sensitivitätsprofile der Merkmalsdetektoren, entsprechend der zu erzeugenden Distanzmodulation. **B:** Visuelle Merkmalsdetektoren, symbolisiert durch Kantenelemente. Graue Quadrate bedeuten identische Ausschnitte des Sehfeldes. Die Zellklasse mit halbnaher Entfernungspräferenz ist aufgrund der aktuellen Fixationsentfernung sensitiviert, alle anderen Zellklassen sind insensitive. (Die Abbildungen 6.5–6.9 bilden eine Serie.)

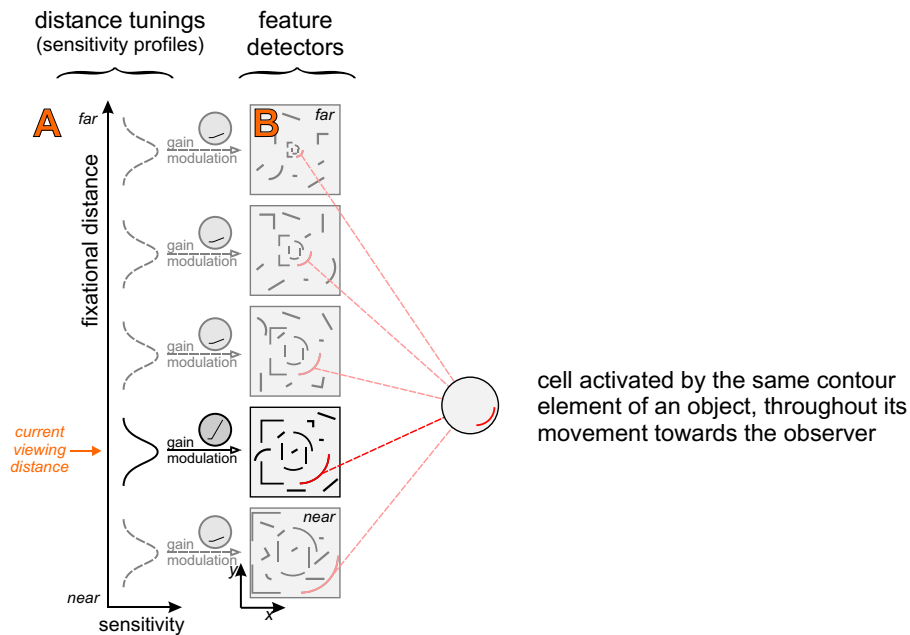


Abbildung 6.6: Entfernung-Complex-Zelle. Merkmalsdetektoren, die einen identischen Objektbestandteil in verschiedenen Sehtiefen repräsentieren, konvergieren auf ein gemeinsames Zielneuron (B). Der Vorgang ist ein Beispiel für situativ gesteuerte Complex-Bildung.

6.5.2 Entfernung-Complex-Zellen

Die bisher beschriebenen Modellkomponenten generieren Detektoren für visuelle Merkmale einfacher oder mittlerer Komplexität, die aufgrund ihrer Sensitivitätsprofile eine Präferenz für bestimmte Fixationsentfernungen aufweisen. Diese Merkmalsdetektoren mit Entfernungspräferenz dienen uns nun als Signalquellen für eine Gruppe von Neuronen in der nächsten Hierarchiestufe unseres Modells.

Wie in Abbildung 6.6 dargestellt konvergieren die Signale der entfernungspräferenten Merkmalsdetektoren systematisch auf nachfolgende Neuronen. (In der Abbildung ist zunächst nur ein solches Neuron dargestellt.) Die Auswahl der Quellneuronen folgt dabei einer festen Regel:

Auswahlregel Denken wir uns, daß der Betrachter ein Objekt fixiert, das sich aus einiger Entfernung geradlinig auf ihn zubewegt. Das retinale Abbild des Objektes wird je nach Abstand verschieden groß sein, jedoch stets die gleiche Gestalt besitzen, also aus gleich geformten Teilreizen zusammengesetzt sein. Das Abbild ein und desselben Objektteils – etwa eines Kontur- oder Oberflächenelements – wird sich aufgrund der Objektbewegung über die Retina verschie-

Auswahlregel

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

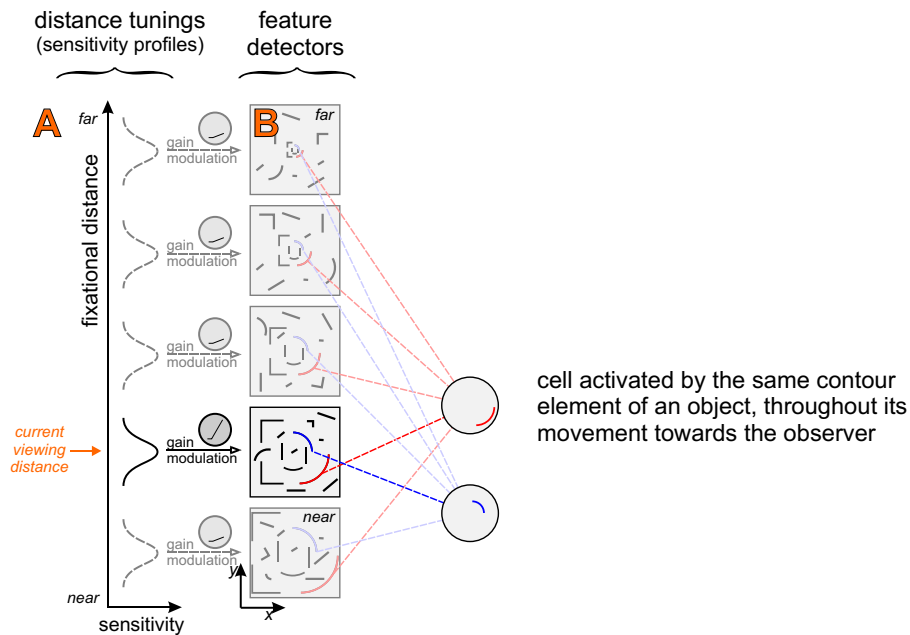


Abbildung 6.7: Eine zweite Entfernungs-Complex-Zelle wird nach der gleichen Auswahlregel gebildet (B).

ben und dort, je nach Objekt-Abstand, verschiedene Merkmalsdetektoren aktivieren. Es sind dies die Detektoren, die einerseits das entsprechende zweidimensionale visuelle Merkmal codieren, sowie andererseits eine Präferenz für die entsprechende Sehentfernung besitzen. Es konvergieren nun gerade diese Merkmalsdetektoren auf ein gemeinsames Zielneuron, die im Verlauf der Objektbewegung vom Abbild ein und desselben Objektteils aktiviert würden. Die in Abbildung 6.6 dargestellte Nervenzelle wird von der rechten unteren Kontur eines runden Gegenstandes aktiviert, und zwar unabhängig davon, in welcher Entfernung dieser Gegenstand betrachtet wird.

Entfernungs-Complex-Zelle

Dieses Neuron ist damit eine *Complex-Zelle* im Sinne der erweiterten Definition aus Abschnitt 2.1.1, denn sie integriert über verschiedene variante Erscheinungsformen ein und desselben Reizes (hier ein und desselben Objektes). Die von den einzelnen Merkmalsdetektoren über die Form ihrer rezeptiven Felder codierten Merkmale gehen jedoch nicht mit gleichem Gewicht in das Gesamtfeld der Zielzelle ein: Die Zielzelle bildet zusammen mit den als Quellen dienenden Neuronen und deren Sensitivitätsprofilen *Gating-Module*, wie sie in Abschnitt 5.3.1 eingeführt wurden, vgl. Abbildungen 5.3, 5.4. Wir erinnern uns, daß in diesem Fall das stimulierende Feld des Zielneurons eine Überlagerung der aktivierenden, nicht der stimulierenden Felder der Quellneuronen ist. Bei passend gewählten

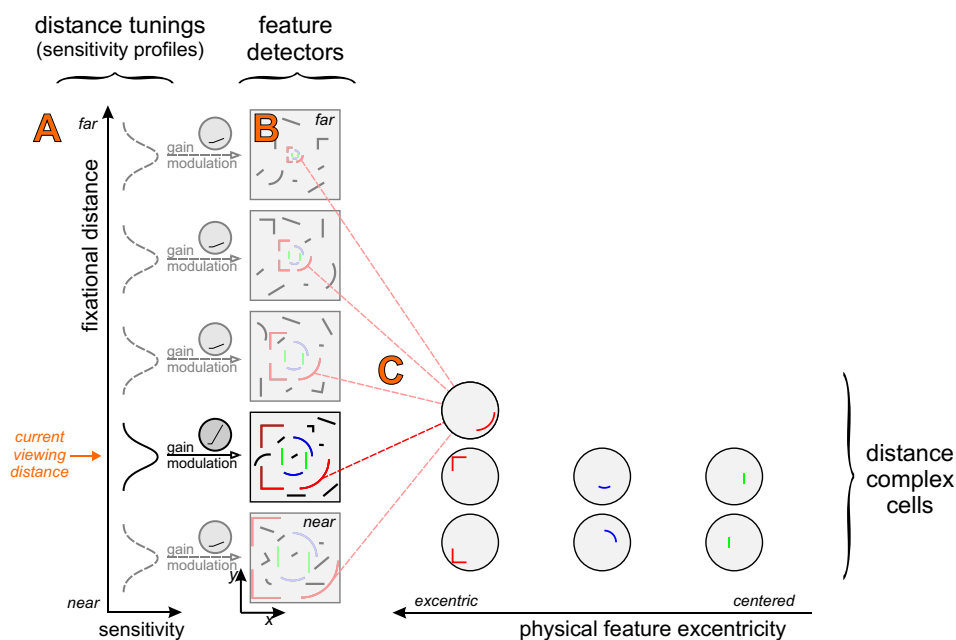


Abbildung 6.8: Das Verfahren führt zu einem kompletten Satz von Entfernung-Complex-Zellen (C). Dargestellt ist eine Auswahl. Projektionen der Merkmalsdetektoren werden beispielhaft nur für eine Complex-Zelle gezeigt. Die Zellen bilden eine sehentfernungsinvariante Szenen-Repräsentation. Sie etablieren außerdem einen *Labelled-Line-Code* für die räumliche Exzentrizität der Merkmale (Abschnitt 6.8.2).

Sensitivitätsprofilen wird aus dem gesamten Satz von Quellneuronen nur derjenige Merkmalsdetektor Einfluß ausüben, dessen Entfernungspräferenz der aktuellen Fixationsentfernung entspricht. Ich möchte für Zellen dieser Art den Begriff *Entfernungs-Complex-Zelle* verwenden.

Wir finden hier also ein weiteres Beispiel der situativ gesteuerten *Complex-Bildung*, wie sie auch schon im Modell der retinalen Schlupfkorrektur vorlag.

Entsprechende Entfernung-Complex-Zellen können nun über alle verfügbaren Merkmalsdetektoren gebildet werden. In Abbildung 6.7 ist dieser Vorgang für ein zweites Neuron dargestellt. Auf gleiche Weise erhalten wir einen vollständigen Satz von Entfernung-Complex-Zellen für sämtliche durch die Merkmalsdetektoren repräsentierten visuellen Merkmale. Abbildung 6.8 stellt einen Satz von Entfernung-Complex-Zellen schematisch dar. Aus Gründen der Übersichtlichkeit sind hier nur die Netzwerkverbindungen für eine einzelne Complex-Zelle eingetragen; die Verbindungen der restlichen Zellen verlaufen gemäß o.g. Auswahlregel.

Jede Entfernung-Complex-Zelle integriert über die entfernungs-

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

varianten Erscheinungsformen eines einzelnen visuellen Merkmals (in den Abbildungen in den Kreisen eingetragen). Das jeweilige Neuron wird von jedem Objekt, dessen Ansicht das entsprechende Merkmal enthält, aktiviert werden – und zwar unabhängig von der Sehentfernung, solange nur der Betrachter das Objekt fixiert.

Die Entfernung-Complex-Zellen bilden einen Satz entfernungsinvarianter Merkmalsdetektoren. Sie generieren eine Repräsentation des Sehgegenstandes in entfernungsinvarianten Teilmerkmalen.

6.5.3 Invariante Objekterkennung

Auf den Ausgangssignalen der Entfernung-Complex-Zellen kann eine beliebige Objekterkennungshierarchie aufsetzen. Zur Auswahl stehen hier uneingeschränkt sämtliche Modelle, die eine Erkennung basierend auf visuellen Teilmerkmalen wie Kontur- oder Oberflächenelementen implementieren. Es entspringen lediglich die Eingangssignale zu diesen Modellen nun nicht mehr direkt den varianten Merkmalsdetektoren, sondern deren entfernungsinvarianten Pendants. Auf diese Weise wird jede Objekterkennung – ganz gleich, welches Modell zu ihrer Umsetzung herangezogen wird – entfernungsinvariant.

Jede Objekterkennung, die auf den Signalen der Entfernung-Complex-Zellen aufsetzt, arbeitet entfernungsinvariant.

Die genaue Form des aufsetzenden Objekterkennungsmechanismus steht nicht im Blickpunkt unserer Betrachtungen. In Abbildung 6.9 ist eine übersimplifizierte Variante dargestellt, bei der sich die Objektrepräsentationen in trivialer Weise unmittelbar aus den Einzelmerkmalen zusammensetzen. Kein realistisches Modell wird die Objektrepräsentationen in einem einzigen Schritt erzeugen. Die gezeigte Variante ist jedoch für alle unsere Überlegungen ausreichend. Die erlangten Objektrepräsentationen sind entfernungsinvariant.

6.5.4 Labelled-Line-Code

Jede Entfernung-Complex-Zelle integriert über die entfernungsvarianten Erscheinungsformen eines einzelnen visuellen Merkmals. Nach der definierten Auswahlregel sind dies die varianten Erscheinungsformen eines Merkmals, das in verschiedenen Sehentfernungen auftritt, sowie in gleichbleibendem *räumlichen* (nicht retinalen!)

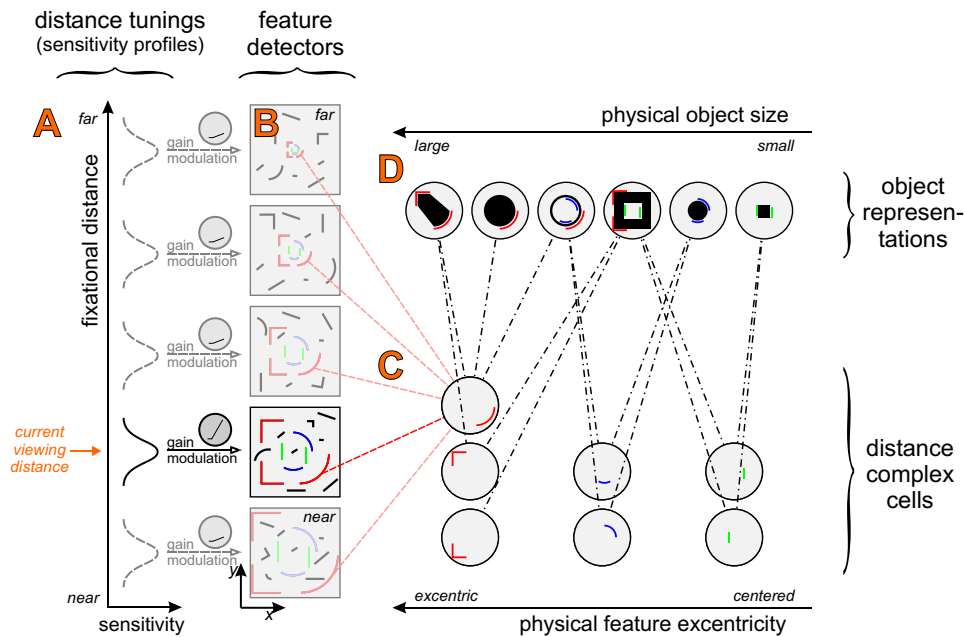


Abbildung 6.9: Beispiel für eine Generierung von Objektrepräsentationen aus den Signalen der Entfernungs-Complex-Zellen (D). Die erlangten Objektrepräsentationen sind entfernungsinvariant.

Abstand vom fixierten Punkt in der Umwelt – ganz so, als sei das Merkmal Teil eines starren Körpers.

Die Entfernungs-Complex-Zellen können damit als visuelle Detektoren für *physikalische Objektbestandteile* gelten. Sie codieren keine Bestandteile des retinalen Bildes mehr, sondern sprechen auf die Sichtbarkeit eines bestimmten physikalischen Objektteils an, unabhängig von dessen Entfernung vom Betrachter. Da das visuelle Merkmal zusätzlich in korrekter *räumlicher Lage* zum fixierten Punkt auftreten muß, etablieren sie einen sog. *Labelled-Line-Code* für die räumliche Position von Objektbestandteilen, d.h., jeder Entfernungs-Complex-Zelle ist eindeutig ein bestimmter Objektbestandteil mit fest definierter räumlicher Lage zugeordnet. Sein Vorhandensein zeigt die betreffende Zelle durch ihre Aktivität an.

Die Entfernungs-Complex-Zellen lassen sich bezüglich der **Ordnbarkeit** räumlichen Exzentrizität der von ihnen codierten Merkmale ordnen (Abbildung 6.8 B und C, Farbe kennzeichnet die räumliche Exzentrizität). Diese Eigenschaft ermöglicht die Schätzung der räumlichen Größe von Sehobjekten. Abschnitt 6.8.2 wird diesen Umstand diskutieren.

6.6 Simulation

In einer Computersimulation des vorgeschlagenen Modells teste ich dessen Fähigkeit zur Generierung einer entfernungsinvarianten Szenenrepräsentation (Abschnitt 6.5.2). Die Simulation umfaßt die Modellkomponenten A–C aus Abbildung 6.9, also einen Satz visueller Merkmalsdetektoren (B), deren entfernungsabhängige Sensitivitätsmodulation (A), sowie einen vollständigen Satz Entfernung-*Complex*-Zellen (C). Eine Objekterkennungsstufe (D) wird nicht simuliert. Als visuelle Reize dienen computergenerierte virtuelle dreidimensionale Szenen. Die Güte der Invarianzgenerierung auf Ebene der *Complex*-Zellen wird quantifiziert. Dies erlaubt, die Modelleistung für verschiedene Verteilungen der Entfernungseigenschaften (A) zu vergleichen.

6.6.1 Echtweltumgebung

Zur Überprüfung des Modellkonzeptes ist die explizite Verfügbarkeit der Parameter *räumliche Entfernung zum Betrachter* und *räumliche Ausdehnung eines Objektes* zwingend notwendig. Dies erfordert, die dreidimensionale Umgebung des „Betrachters“ in die Simulation einzubeziehen. Anders als in konventionellen Simulationen des visuellen Systems kann die Basis der Reizverarbeitung in meinem Modell also nicht der retinale Stimulus sein. Vielmehr ergibt sich dieser als die durch die Augenlinse vermittelte optische Projektion eines räumlichen Gegenstandes.

Virtuelle Umwelt

Ich verwende das verbreitete 3d-Rendering-System *OpenGL* zur Projektion einer virtuellen dreidimensionalen Umgebung auf die simulierte Retina. Die Abbildungsparameter wie Linsenabstand von der Retina und Größe des Sehfeldes wurden dem menschlichen Augendurchmesser angepaßt (Linsenabstand sowie wahrnehmungsrelevanter Retinadurchmesser 2,7 cm, entsprechend 53° Öffnungswinkel).

Die virtuelle Umwelt des Betrachters besteht aus eine Reihe dreidimensionaler Objekte, die als Polygonzüge definiert sind (Abbildung 6.10). Einzelne Objekte können frei in der Szene plaziert werden. Die Definition der Objekte, sowie die Plazierung geschieht in Echtwelt-Koordinaten (m). Das Abbildungssystem errechnet die retinale Projektion in retinalen Koordinaten (cm) und liefert deren Abtastung als Rasterbild beliebiger Auflösung. Die Simulation verwendet eine Auflösung von 1280 Pixeln pro 2,7 cm Retina.

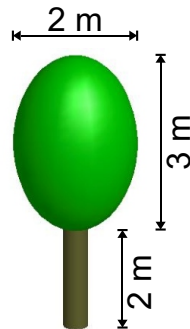


Abbildung 6.10: Echtwelt-Objekt „Baum“. Die Ausmaße des dreidimensionalen Objektes, sowie seine Position in der Umwelt werden in Echtwelt-Koordinaten (m) angegeben.

6.6.2 Netzwerk

Das Netzwerk verarbeitet einen zentralen Ausschnitt des monokularen retinalen Reizes der Kantenlänge $\frac{2,7\text{cm}}{5} = 0,54\text{cm}$, entsprechend $11,4^\circ$ Öffnungswinkel.⁹ Der Ausschnitt gelangt in Form von 256×256 Grauwertpixeln in das simulierte Netzwerk. Zusätzlich liegt dem Netzwerk die aktuelle Fixationsentfernung als eindimensionale Größe vor. Außer diesen beiden Eingängen werden keine weiteren Informationen verarbeitet.

Netzwerkeingang

Die Abtastung des retinalen Bildes geschieht mittels D Sätzen von je vier Schichten aus 128×128 Kantendetektoren in retinotoper Anordnung (vier Orientierungspräferenzen 0° , 45° , 90° und 135°). D ist hierbei die Zahl der Zellklassen mit diskreten Entfernungscharakteristiken. Die stimulierenden Felder¹⁰ der Kantendetektoren sind als zweidimensionale symmetrische Gaborfunktionen der retinalen räumlichen Wellenlänge $0,27\text{mm}$ ($12,8\text{Pixel}$) implementiert. Als Neuronenmodell dient wie zuvor das in Abschnitt 3.3.2 beschriebene ratenbasierte *Graded-Response-Modell*.

Merkmalsdetektoren

Eine Distanzcharakteristik der Kantendetektoren wird über das in Abschnitt 5.3.2 eingeführte Konzept der Sensitivitätsprofile erzeugt. Die Sensitivitätsmodulation erfolgt wie beschrieben über eine Anpassung der Verstärkung der neuronalen Übertragungsfunktion Θ . Für die Entfernungscharakteristiken der D Detektorsätze werden verschiedene Verteilungen angesetzt (Abschnitt 6.7).

Distanzmodulation

⁹Zum einen beschränkt dies die Bilder auf einen Größenbereich, in dem zu vermuten ist, daß größeninvariante Erkennung ohne serielle Fovealisierung einzelner Objektteile geleistet wird. Zum anderen treten bei der Projektion für sehr nahe, in der Tiefe ausgedehnte Objekte Verzerrungen der Umriss auf, die vermieden werden sollen.

¹⁰Zur Terminologie vgl. Abschnitt 3.3.3.

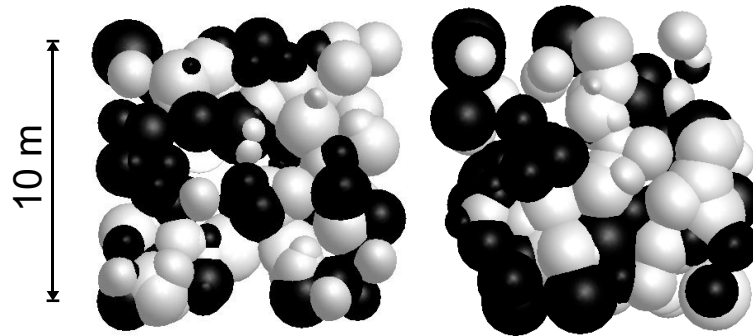


Abbildung 6.11: Zwei der 50 verwendeten Testobjekte zur Bestimmung des Invarianzindexes.

Complex-Zellen

Die D Sätze von je $4 \times 128 \times 128$ Kantendetektoren konvergieren auf vier *echtweit-topologische* Schichten von 128×128 kantenselektiven Entfernungs-Complex-Zellen (Orientierungspräferenzen 0° , 45° , 90° und 135°). Die Quellneuronen zu jeder Complex-Zelle werden so bestimmt, daß sie aufgrund ihrer bevorzugten Orientierung, ihrer Entfernungspräferenz und ihrer retinalen rezeptiven Feldposition die definierte Auswahlregel erfüllen. Der erforderliche Größenzusammenhang ist durch die optische Abbildungsgleichung (3.10) gegeben.

6.6.3 Invarianzindex

Testreize

Um die Invarianzleistung des Netzwerks beurteilen zu können, wird eine größere Zahl von Echtweltobjekten zur Verarbeitung benötigt. Diese sollen verschieden sein, jedoch vergleichbar gute Reize für die Merkmalsdetektoren darstellen. Die Wahl fiel hier auf zufällig generierte Anordnungen schwarzer und weißer kugelförmiger Objekte, wie sie in Abbildung 6.11 dargestellt sind. Ihre retinalen Projektionen enthalten starke räumliche Helligkeitskontraste aller Orientierungen. Fünfzig dieser Testobjekte werden dem Netzwerk nacheinander in 50 verschiedenen Entfernungen von 25–125 m präsentiert. Der Steuerparameter *Fixationsentfernung*, der neben dem retinalen Reiz in die Verarbeitung eingeht, wird jeweils auf den Abstand des Objektes vom simulierten Betrachter eingestellt.

Invarianzindex

In Anlehnung an Booth und Rolls (1998) definiere ich den Invarianzindex eines Neurons mit Feuerrate r als

$$II := \frac{\langle var_d(r) \rangle_{obj}}{\langle var_{obj}(r) \rangle_d} \quad \text{Invarianzindex} \quad (6.1)$$

Das Antwortverhalten der Nervenzelle kann als entfernungsinvariant gelten, wenn die Ratenvarianz var_d über alle Präsentationentfer-

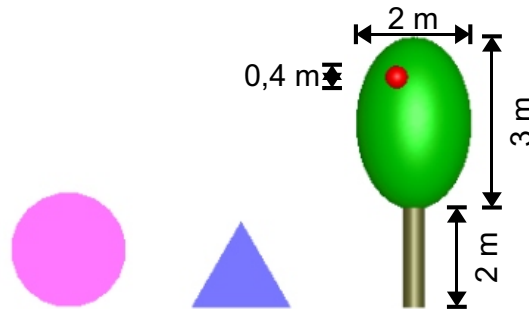


Abbildung 6.12: Echtwelt-Szene „Obstbaum“.

nungen eines Testobjektes klein ist, während zugleich die Ratenvarianz var_{obj} über die Präsentationen aller Testobjekte in gleicher Entfernung groß ist, die Zelle also ihre Objektspezifität behält. Der Invarianzindex II nimmt mit zunehmender Invarianzleistung ab. Er erreicht seinen Minimalwert 0 für ideale Entfernungsinvarianz.

6.7 Ergebnisse und Invarianzleistung

6.7.1 Gleichförmige Entfernungscharakteristik

Abbildung 6.13 zeigt die Ergebnisse von Simulationsläufen, die zur Illustration mit einer einfachen Echtwelt-Szene durchgeführt wurden. Die Szene enthielt ein Echtwelt-Objekt „Obstbaum“, auf das fixiert wurde, sowie zwei tiefenlose Referenzobjekte¹¹ (Abbildung 6.12). Die linke Spalte in Abbildung 6.13 zeigt den retinalen Eingangszreiz für die Sehentfernungen 30, 60 und 90 m. Die Achsen sind in retinalen Koordinaten beschriftet. Die zweite und dritte Spalte zeigen die durch den Reiz hervorgerufenen Aktivierungen der Entfernungs-*Complex*-Zellen. Die Antworten der Neuronen aller vier *Complex*-Schichten sind im Diagramm als orientierte Linien ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) dargestellt. Ihre Länge entspricht der Feuerrate der jeweiligen Zelle. Die Achsen sind in Echtwelt-Koordinaten beschriftet. Dabei sind die Zellaktivitäten an der Stelle eingetragen, die den Neuronen aufgrund des *Labelled-Line*-Codes für Echtweltposition zugeordnet ist (Abschnitt 6.5.4).

Illustratives Beispiel

Die zwei rechten Spalten in Abbildung 6.13 zeigen die Simulationsergebnisse für zwei unterschiedlich angesetzte Verteilungen von Entfernungscharakteristiken für die D Klassen von Merkmalsde-

¹¹Sie dienten zur Kontrolle der im *OpenGL*-System aufgrund der ebenen Projektionsfläche auftretenden Verzerrungen.

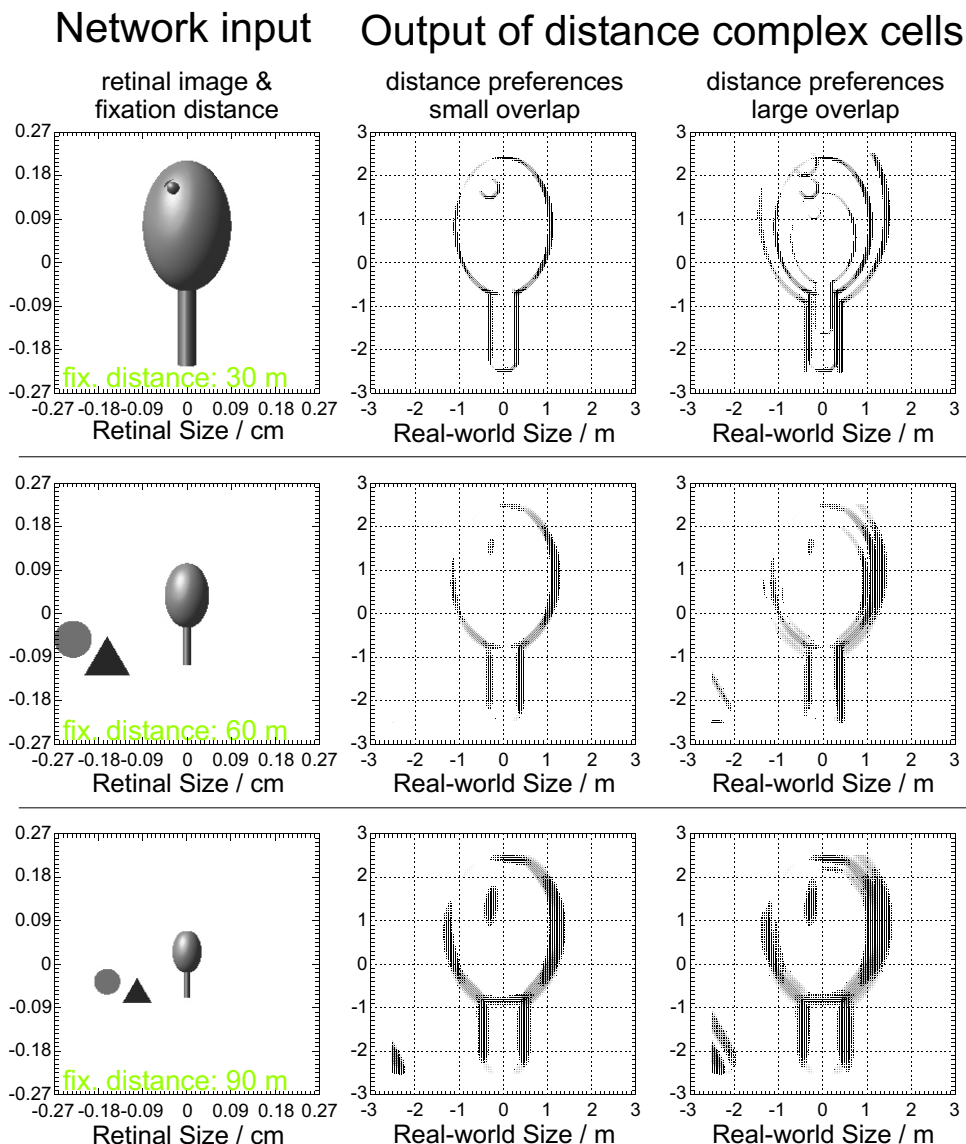


Abbildung 6.13: Netzwerkeingang und -ausgang für Präsentation der Echtwelt-Szene „Obstbaum“ (Abb. 6.12) in 30, 60 und 90 m Sehentfernung (Zeilen von oben nach unten). **Linke Spalte:** Der auf die Netzhaut abgebildete Reiz variiert stark mit der Sehentfernung. Diese größenvariante retinale Projektion dient, zusammen mit der aktuellen Sehentfernung, als Netzwerkeingang. Der Ausschnitt entspricht einem Öffnungswinkel von $11,4^\circ$. **Mittlere und rechte Spalte:** Die Zellaktivitäten der vier *Complex-Schichten* sind überlagert in Form verschieden langer orientierter Kantenelemente dargestellt. Die Diagramme sind in Echtweltkoordinaten beschriftet. Dabei sind die Zellaktivitäten an der Stelle eingetragen, die den Neuronen aufgrund des *Labelled-Line-Code*s für Echtweltposition zugeordnet ist (Abschnitt 6.5.4). Die erzeugten Repräsentationen sind weitgehend entfernungsinvariant. Dargestellt sind die Simulationsergebnisse für gleichförmige Entfernungscharakteristiken mit geringem Überlapp (Mitte), und starkem Überlapp (rechts).

6.7 Ergebnisse und Invarianzleistung

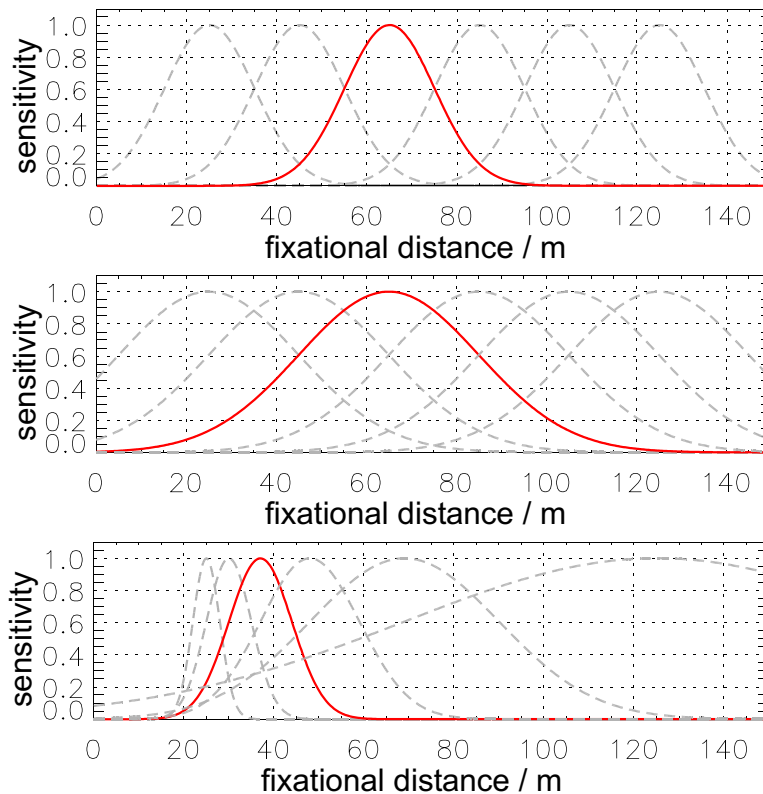


Abbildung 6.14: Verschiedene Entfernungscharakteristiken für $D=6$ Zellklassen. Abgedeckte Distanz 25–125 m. **Oben:** Gleichförmige Verteilung, Abstand der Maxima 20 m, $\sigma=10$ m. **Mitte:** Gleichförmige Verteilung, Abstand der Maxima 20 m, $\sigma=20$ m. **Unten:** Auf konstante Invarianzlast angepaßte Verteilung. Entfernungspräferenzen bei 25, 29,76, 36,76, 48,08, 69,44 und 125 m. Breiten der Gaußkurven im gleichen Verhältnis angepaßt. Rot hervorgehoben die Entfernungscharakteristik für die dritte Zellklasse. (Hervorhebung nur zur besseren Übersicht, alle Klassen sind gleichwertig.)

tektoren. Die Distanzmodulation der Merkmalsdetektoren stellt im Modell den grundlegenden Mechanismus zur Generierung entfernungsinvarianter Zellantworten dar. Daher wird die konkrete Form der Entfernungscharakteristiken einen Einfluß auf die Leistung des Modells haben.

Eine einfache Annahme ist, daß die D Klassen von Merkmalsdetektoren Entfernungspräferenzen besitzen, die gleichförmig über die abzudeckende Distanz verteilt sind. Jede einzelne Klasse besitzt eine charakteristische Kurve, die in Ermangelung detaillierter Erkenntnisse als eine Gaußfunktion angesetzt wird (siehe Abschnitt 6.5.1). Abbildung 6.14 (oben) zeigt eine solche Verteilung von Entfernungscharakteristiken für $D=6$. Die sechs Maxima sind gleichförmig über

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

eine Distanz von 25–125 m verteilt, ihr Abstand beträgt 20 m. Die Breite der Gaußkurven wurde mit $\sigma=10$ m so gewählt, daß die Kurven an ihren Wendepunkten überlappen. Dies soll eine hinreichende Abdeckung der gesamten Distanz gewährleisten. Der Modellausgang für ein äquivalentes Paradigma ist in der mittleren Spalte von Abbildung 6.13 dargestellt (geringer Überlapp der Entfernungskarakteristiken, hier: $D=10$, Abstand der Maxima 10 m, $\sigma=5$ m).

Die Breite der Gaußkurven ist indes ein frei wählbarer Modellparameter. Abbildung 6.14 (Mitte) zeigt eine abgewandelte Verteilung der Entfernungskarakteristiken für $D=6$. Die Maxima der Kurven und damit die bevorzugten Distanzen der Merkmalsdetektoren liegen an den gleichen Stellen wie im oberen Diagramm. Die Kurven weisen jedoch mit einer Breite von $\sigma=20$ m einen deutlich stärkeren Überlapp auf. Der Modellausgang für ein äquivalentes Paradigma ist in der rechten Spalte von Abbildung 6.13 dargestellt (starker Überlapp der Entfernungskarakteristiken, hier: $D=10$, Abstand der Maxima 10 m, $\sigma=10$ m).

6.7.2 Invarianzleistung

Zur Quantifizierung der Invarianzleistung des simulierten Netzwerks wurde der Invarianzindex nach Gleichung (6.1) bestimmt. Hierzu wurden 50 der computergenerierten Testreize nacheinander in 50 Entfernungen zwischen 25 m und 125 m präsentiert. Die $D=6$ Klassen entfernungsselektiver Merkmalsdetektoren waren mit gleichförmig verteilten Sensitivitätsprofilen nach Abbildung 6.14 (oben, $\sigma=10$ m), bzw. Abbildung 6.14 (Mitte, $\sigma=20$ m) versehen.

Gleichung (6.1) liefert individuelle Werte für jede der $4 \times 128 \times 128$ Entfernungs-Complex-Zellen. Ihr Mittelwert $\langle II_n \rangle_n$ charakterisiert die Gesamtleistung des Systems. Der mittlere Invarianzindex liefert für eine gleichförmige Abdeckung mit kleiner Kurvenbreite $\sigma=10$ m den Wert 0,71, für gleichförmige Abdeckung mit großer Kurvenbreite $\sigma=20$ m den Wert 0,67, der eine erhöhte Invarianzleistung anzeigt:

Distanzcharakteristiken	$\langle II \rangle \pm \sigma(II)$
gleichförmig, $\sigma=10$ m	$0,71 \pm 0,14$
gleichförmig, $\sigma=20$ m	$0,67 \pm 0,15$

Eine Übersicht über alle ermittelten Invarianzindizes zeigt Abbildung 6.15.

6.7.3 Angepaßte Entfernungskarakteristik

Entfernungsabhängige Invarianzlast Wie der Abbildungsgleichung (3.10) zu entnehmen ist, erfährt das re-

6.7 Ergebnisse und Invarianzleistung

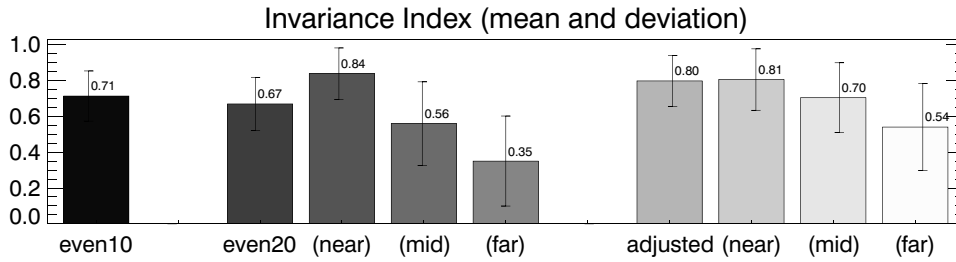


Abbildung 6.15: Vergleich der Invarianzleistung für verschiedene Distanzcharakteristik-Paradigmen. *Even10* und *even20*: Gleichförmige Entfernungseigenschaften mit $\sigma=10$ m bzw. $\sigma=20$ m. *Adjusted*: Auf konstante Bildgrößenänderung angepasste Distanzcharakteristiken. Dargestellt ist Mittelwert und Standardabweichung des Invarianzindex für alle *Complex*-Zellen.

tinale Abbild eines Gegenstandes bei einer Entfernungsänderung Δg eine Größenänderung ΔB , die von der Sehentfernung g abhängt:

$$B = \frac{bG}{g} \quad (6.2)$$

$$\begin{aligned} \Delta B &= B_1 - B_2 \\ &= \frac{bG}{g_1} - \frac{bG}{g_2} \\ &= -bG \frac{g_1 - g_2}{g_1 g_2} \\ &\approx -bG \frac{\Delta g}{g^2} \end{aligned} \quad (6.3)$$

Die vom Netzwerk zu kompensierende Größenänderung ΔB ist daher für nahe Objekte größer als für ferne Objekte.

Um die Auswirkung dieser unterschiedlichen Invarianzlast zu untersuchen, wurden für das $\sigma = 20$ m-Paradigma getrennte Simulationsläufe für nahe, mittlere und ferne Objektdistanzen durchgeführt. Hierzu wurde die gesamte Distanz von 100 m in drei gleiche Teile zu je $33\frac{1}{3}$ m aufgeteilt. Als Invarianzindizes ergeben sich:

Entfernungsbereich	$\langle II \rangle \pm \sigma(II)$
gesamte Distanz	$0,67 \pm 0,15$
nah	$0,84 \pm 0,14$
mittel	$0,56 \pm 0,23$
fern	$0,35 \pm 0,25$

Die Invarianzleistung ist für entferntere Objekte höher. Dies legt nahe, daß eine gleichförmige Verteilung der Entfernungspräferenzen

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

zen wie in den Abbildungen 6.14 (oben) und (Mitte) keine geeignete Abtastung der Entfernungsdimension darstellt. Vielmehr sollten die Entfernungspräferenzen der Merkmalsdetektoren so verteilt sein, daß für je zwei aufeinanderfolgende Präferenzentfernungen $\Delta B = const.$ gilt. Abbildung 6.14 (unten) zeigt eine entsprechend angepaßte Verteilung für $D=6$ Zellklassen, die diese Forderung erfüllt. Mit dieser Verteilung wurden gleichfalls Simulationsläufe über die gesamte Distanz sowie die drei Teildistanzen durchgeführt. Als Invarianzindizes ergeben sich in diesem Fall:

Entfernungsbereich	$\langle II \rangle \pm \sigma(II)$
gesamte Distanz	$0,80 \pm 0,14$
nah	$0,81 \pm 0,17$
mittel	$0,70 \pm 0,20$
fern	$0,54 \pm 0,24$

Die Ergebnisse für alle Simulationsläufe sind in Abbildung 6.15 zum Vergleich gegenübergestellt.¹²

6.7.4 Deaktivierte Entfernungsmodulation

Als interessanter Sonderfall sei schließlich zur Illustration dargestellt, wie sich der Netzwerkausgang verhält, wenn keinerlei Entfernungsmodulation der Merkmalsdetektoren stattfindet, etwa deshalb, weil keine Entfernungsschätzung vorliegt. Hierzu wurden die Entfernungscharakteristiken aller Detektorclassen auf konstant 1 gesetzt ($\sigma \rightarrow \infty$). Damit sind zu jeder Zeit sämtliche Merkmalsdetektoren sensitiv. Die Aktivität der Entfernung-Complex-Zellen für die Echtwelt-Szene „Obstbaum“ und $D=10$ Detektorsätze ist unter diesen Bedingungen in Abbildung 6.16 dargestellt ($D=10$, Abstand der Maxima 10 m, $\sigma \rightarrow \infty$).

6.8 Diskussion

Aus Abbildung 6.13 ist ersichtlich, daß das beschriebene Modell in der Computersimulation eine weitgehend entfernungsinvariante Szenenrepräsentation erzeugt. Für verschiedene Sehentfernungen, und damit für unterschiedliche retinale Reize, werden ähnliche

¹²Animierte Simulationsergebnisse für gleichförmige Entfernungscharakteristiken mit geringem und großem Überlapp, sowie für die angepaßte Entfernungscharakteristik sind als Zusatzmaterial einsehbar unter <http://neuro.physik.uni-marburg.de/~kupper/supplement/>.

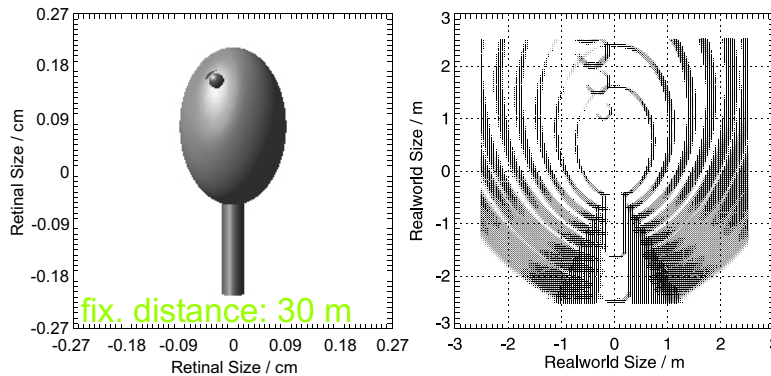


Abbildung 6.16: Aktivität der Entfernungs-Complex-Zellen im Falle deaktivierter Entfernungsmodulation.

Entfernungs-Complex-Zellen aktiviert. Die den Complex-Zellen zugeordneten Echtweltgrößen (siehe Achsenbeschriftung) entsprechen den physikalischen Ausmaßen des Objektes (Abbildung 6.12).

6.8.1 Invarianzleistung der Simulation

In den Aktivitäten der Complex-Zellen verbleibt eine Restvarianz, die von der Wahl der Entfernungscharakteristiken (Sensitivitätsprofile) der Merkmalsdetektoren abhängt. Abbildung 6.15 stellt die für verschiedene Paradigmen in der Simulation ermittelten Invarianzindizes zusammen.

Überlapp der Sensitivitätsprofile

Aus den beiden rechten Spalten in Abbildung 6.13 ist ersichtlich, daß die erzeugten entfernungsinvarianten Repräsentationen für verschiedene Entfernungscharakteristiken qualitativ unterschiedlich aussehen können: Für stark überlappende Entfernungscharakteristiken (rechte Spalte) sind stets mehrere Klassen von Merkmalsdetektoren sensitiv, während für wenig überlappende Profile maximal zwei – bei passend gewählter Schwelle genau eine – Zellklasse sensitiv ist.

Mehrfachbilder

Es könnte vermutet werden, daß die gleichzeitige Sensitivierung möglichst weniger Zellklassen vorteilhaft ist, um eine eindeutige Repräsentation zu erhalten. Der Vergleich der beiden Paradigmen mit gleichförmigen Distanzcharakteristiken verschiedener Breite in Abschnitt 6.7.2 zeigt hingegen, daß die Invarianzleistung des Gesamtnetzes leicht ansteigt, wenn der Überlapp der Kurven erhöht wird. Offenbar ist die Restvarianz in der Aktivität der Complex-Zellen hier geringer.

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

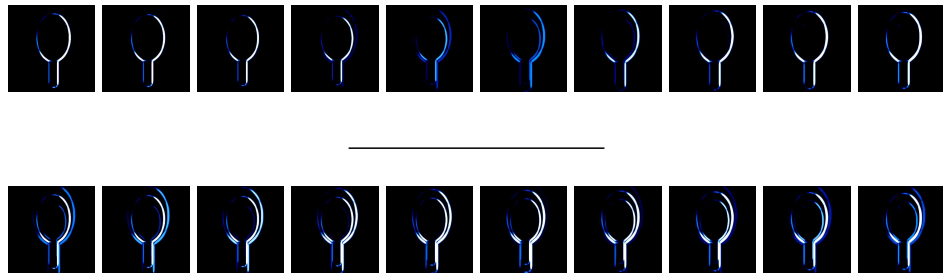


Abbildung 6.17: Aktivität der Entfernung-Complex-Zellen, hellkeitscodiert, alle Orientierungen überlagert. Zwei Sequenzen bei Präsentation über eine Distanz von 10 m in Schritten von 1 m (v.l.n.r.). **Oben:** Entfernungseigenschaften mit geringem Überlapp (Abstand 10 m, $\sigma=5$ m). **Unten:** Entfernungseigenschaften mit großem Überlapp (Abstand 10 m, $\sigma=10$ m). Die dargestellten Bilder sind Ausschnitte einer animierten Sequenz, einsehbar unter <http://neuro.physik.uni-marburg.de/~kupper/supplement/>.

Um diesen Umstand zu beleuchten, betrachten wir die Sequenzen in Abbildung 6.17: Gezeigt ist die hellkeitscodierte Aktivität der Complex-Zellen für Objektentfernungen über eine Distanz von 10 m. Diese Distanz entspricht gerade dem Abstand zweier Entfernungspräferenzen. In der oberen Sequenz (geringer Überlapp der Distanzcharakteristiken) sind wenige Zellklassen gleichzeitig sensitiv. Die Repräsentation ist meist scharf, erfährt aber noch gewisse Größenschwankungen. In der unteren Sequenz (großer Überlapp) sind stets mehrere Zellklassen sensitiv. Die Repräsentation besteht aus der Überlagerung ihrer Aktivitäten. Dieses „Mehrfachbild“ ist indes über die Sequenz hinweg deutlich homogener als im oberen Fall.

Das Auftauchen der „Mehrfachbilder“ ist damit der Invarianzgenerierung zuträglich, denn die Repräsentation erfährt geringere Veränderungen mit der Entfernung. Dies gilt natürlich nur dann, wenn die Gesamtrepräsentation hinreichend spezifisch bleibt, um einzelne Objekte voneinander unterscheiden zu können. Obwohl eine entsprechende Bedingung in die Definition des Invarianzindexes eingeflossen ist, ist die Entscheidung hierüber von der Objektrepräsentation im aufsetzenden Objekterkennungsmechanismus abhängig und auf Einzelzellebene nicht entscheidbar.

Frequenzgehalt

Die simulierten Kantendetektoren wiesen zwar verschiedene Entfernungspräferenzen auf, ihre stimulierenden Felder waren jedoch alle als Gabor-Funktionen gleicher retinaler Wellenlänge implementiert.

Dies trägt dem Umstand nicht Rechnung, daß sich die im retinalen Reiz enthaltenen Raumfrequenzanteile mit zunehmender Entfernung des Objektes zu höheren Frequenzen hin verschieben. Dieser Effekt äußert sich in der Simulation in einer zunehmenden „Verschmierung“ der erzeugten Repräsentation für große Entfernungen. Sie ist in Abbildung 6.13 gut sichtbar. Wie bereits im Fall der „Mehrfachbilder“ festgestellt, ist eine gewisse Verschmierung der Invarianzgenerierung zuträglich, solange eine hinreichende Objektspezifität erhalten bleibt. Ist die Aktivierung hinreichend spezifisch, so verbessert dieser Effekt zusätzlich die Invarianzleistung im Fernbereich.

Die Verwendung von Gabor-Feldern verschiedener retinaler Wellenlängen würde den Verschmierungseffekt mindern. Weiterhin muß an dieser Stelle aber die Frage nach der geeigneten Ebene in der visuellen Hierarchie wiederholt werden, in der ein solcher Invarianzmechanismus ansetzen kann (vgl. Abschnitt 6.5). Obgleich es von Vorteil wäre, bereits sehr grundlegende visuelle Merkmale mit entfernungsinvarianten Eigenschaften auszustatten, so sind hierfür Kantendetektoren mit Gabor-artigen rezeptiven Feldern wegen ihrer starken Kopplung an den Raumfrequenzgehalt des retinalen Stimulus keine guten Kandidaten. In der Simulation fiel die Wahl auf diese Merkmale aufgrund ihrer einfachen Implementierbarkeit, und um die Zahl der verschiedenen Merkmale gering zu halten. Es ist jedoch zu vermuten, daß der Mechanismus auf Merkmalsdetektoren mittlerer Komplexität erfolgreicher arbeiten kann, da diese bereits abstraktere Bestandteile des Stimulus bis hin zu Objektteilen codieren. Für eine Objektrepräsentation in Merkmalen mittlerer Komplexität spricht auch deren besondere Eignung für die Klassifikation von Objekten (Ullman et al., 2002), sowie eine Verminderung des potentiell auftretenden Bindeproblems (Mel und Fiser, 2000).

Entfernungsabhängigkeit

In der Simulation wurde versucht, die Entfernungsabhängigkeit der Invarianzleistung zu kompensieren, indem die Distanzcharakteristiken der Merkmalsdetektoren auf konstante Bildgrößenänderung angepaßt wurden (Abschnitt 6.7.3). Abbildung 6.15 (Paradigma *adjusted*) zeigt, daß gegenüber der gleichförmigen Verteilung zwar die Änderung des Invarianzindexes mit der Entfernung geringer geworden ist, allerdings ist dies auf höhere Invarianzindizes im Mittel- und Fernbereich zurückzuführen. Wie erklärt sich die insgesamt schlechte Invarianzleistung bei angepaßten Distanzcharakteristiken?

Abbildung 6.14 (unten) zeigt die in diesem Simulationslauf verwendete Verteilung von Distanzcharakteristiken. Maxima und Breite

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

der individuellen Profile wurden konstanter Größenänderung angepaßt. Die Zahl von $D=6$ diskreten Zellklassen wurde beibehalten, um einen Vergleich mit den anderen beiden Paradigmen zu ermöglichen. Es fällt jedoch auf, daß die Abdeckung des zweiten und dritten Distanzdrittels (58,58–125 m) nun nur noch von im wesentlichen zwei Zellklassen geleistet wird. Damit kann in diesem Entfernungsbereich keine invariante Repräsentation hoher Güte erwartet werden. Die Tatsache, daß die Invarianzindizes im zweiten und dritten Distanzdrittel gegenüber dem ersten noch immer niedriger sind, mag auf die beiden zuvor beschriebenen Faktoren *Überlapp* und *Frequenzgehalt* zurückzuführen sein. Die geringe Zahl von Zellklassen in den oberen Distanzbereichen macht die Ergebnisse dieser Simulationsläufe jedoch fragwürdig.

Zahl der Zellklassen

Die Zahl der simulierten Zellklassen mit verschiedenen Entfernungspräferenzen wurde in der Simulation mit $D=6$ für einen Entfernungsbereich von insgesamt 100 m recht gering angesetzt. Grund war eine Beschränkung des Simulationsaufwandes in Speicher und Laufzeit. Die Anzahl der im visuellen System des Primaten verfügbaren entsprechenden Zellklassen ist auf Basis der verfügbaren experimentellen Untersuchungen nicht abzuschätzen (Abschnitt 6.3.3). Für eine Diskussion des mit der Anzahl der Zellklassen verbundenen neuronalen Aufwands siehe Abschnitt 6.8.2.

6.8.2 Sehentfernung als Steuerparameter

Im vorgestellten Modell werden entfernungsinvariante neuronale Antworten über das Prinzip der situativ gesteuerten *Complex*-Bildung erzeugt. Dieser erweitert das klassische auf Hubel und Wiesel (1962) zurückgehende Prinzip der *Complex*-Bildung um die Steuerung durch einen externen Parameter. Dieser spiegelt eine physikalisch-körperliche Randbedingung wider (hier die Sehentfernung), auf die das neuronale System kalibriert wird.

Das Verfahren der situativ gesteuerten *Complex*-Bildung hat gegenüber der klassischen ungesteuerten *Complex*-Bildung deutliche Vorteile. Seine Nachteile bestehen vor allem in der großen Zahl benötigter Neuronen.

Größenschätzung

Wie in Abschnitt 6.5.4 erklärt, generieren die Entfernungs-*Complex*-Zellen im Modell einen *Labelled-Line*-Code für die räumliche Position der durch sie codierten Objektmerkmale: Das entsprechende

Merkmal muß in korrekter *räumlicher Lage* zum in der Umwelt fixierten Punkt auftreten, um die zugeordnete *Complex-Zelle* zu aktivieren. Das impliziert, daß sich die codierten Merkmale nach ihrer räumlichen Exzentrizität ordnen lassen. In Abbildung 6.9 (C) ist dies bereits dargestellt. Die räumliche Exzentrizität der durch die *Complex-Zellen* repräsentierten Merkmale ist dort durch Farben gekennzeichnet. Sie nimmt von links nach rechts ab.

Werden nun Objektrepräsentationen auf den Signalen der *Complex-Zellen* gebildet (Abbildung 6.9, D), so geht diese Eigenschaft auf die Objektrepräsentationen über (Prinzip der Merkmalsvererbung in neuronalen Hierarchien). Diese lassen sich dann ebenso ordnen – beispielsweise nach der größten räumlichen Exzentrizität der enthaltenen Merkmale. Diese kann gut als ein Maß für die räumliche Größe eines Objektes dienen. Die in Abbildung 6.9 (D) gezeigten Objektrepräsentationen haben diese Ordnung. Ihre räumliche Größe nimmt von links nach rechts ab.

Aufgrund der Ordenbarkeit der repräsentierenden Neuronen ist die Schätzung der räumlichen Objektgröße auch als neuronales Signal verfügbar, beispielsweise als Position des Aktivierungsmaximums in einer topologischen Anordnung von Neuronen.

In dem vorgestellten Modell ist die räumliche Größe eines fixierten Objektes schätzbar. Sie ist gegeben als die größte räumliche Exzentrizität der an der Repräsentation beteiligten Objektmerkmale.

Auf gleiche Art ermöglicht die Repräsentation eine getrennte Schätzung der räumlichen Objekthöhe oder -breite.¹³

Das Erzeugen einer entfernungsunabhängigen Repräsentation, in der die Schätzung physikalischer Objektgrößen möglich ist, war die in Abschnitt 6.4 formulierte Problemstellung. Sie kann hiermit als gelöst gelten.

Neuronaler Aufwand

Die Erzeugung von Invarianz durch situativ gesteuerte *Complex-Bildung* wird auch in diesem Modell durch einen erhöhten neuronalen Aufwand realisiert. Die benötigte Multiplizität D , mit der Merk-

¹³Eine Schätzung der Objekttiefe erlaubt die Auswertung binokularer Disparität, die in diesem Modell nicht behandelt wird. Ferner ist sie aus bildhaften Hinweisen wie Helligkeitsverteilung oder Bewegungsparallaxe ableitbar. Im Modell wäre sie durch sequentielle Fixation verschieden entfernter Objektbestandteile zugänglich.

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

malsdetektoren im Modell vorhanden sein müssen, hängt von verschiedenen Faktoren ab. Hier geht der insgesamt invariant zu repräsentierende Entfernungsbereich ein. Von entscheidender Bedeutung wird jedoch auch die Art der repräsentierten visuellen Merkmale sein, wie in Abschnitt 6.8.1 unter *Frequenzgehalt* angesprochen.

Selbst die Wahl disjunkter Zellklassen mit wenigen diskret verteilten Entfernungspräferenzen ist eine Idealisierung des Modells. Die Verteilung der bevorzugten Sehentfernungen kann ebensogut als kontinuierlich über den gesamten Distanzbereich angenommen werden. Dies erfordert keineswegs vollständige Detektorsätze für alle visuellen Merkmale in allen retinalen Positionen und sämtlichen Entfernungen. Vielmehr können die Entfernungscharakteristiken verschiedener Neuronenpopulationen oder individueller Neuronen unterschiedlich verteilt sein. Letztlich müssen nicht sämtliche Merkmale einer Objektrepräsentation entfernungsvariant repräsentiert sein, um eine Entfernungs- und Größenschätzung für das betreffende Objekt zu erlauben. Möglicherweise ist die Beschränkung auf wichtige *Schlüsselmerkmale* ausreichend. Diese würden dann die Basis für die Größenschätzung des gesamten Objektes bilden.¹⁴

Inhomogener Aufbau der Retina

Der Durchmesser der dendritischen Felder retinaler Ganglienzellen¹⁵ nimmt beim Primaten linear mit der retinalen Exzentrizität zu (vgl. Rodieck, 1998). Eine hochauflösende Repräsentation visueller Merkmale ist daher vornehmlich für die zentralen Bereiche des Sehfeldes zu leisten. Das periphere Sehfeld ist durch eine weitaus geringere Dichte von Neuronen repräsentierbar. Da zudem ausgedehnte Sehobjekte in einer Folge von Fixationsbewegungen abgetastet werden, könnte in der Peripherie auch die Auflösung des Entfernungsparameters, und damit die Multiplizität der Merkmalsdetektoren, deutlich geringer ausfallen. Dies würde eine entfernungsinvariante visuelle Verarbeitung mit einem deutlich geringeren neuronalen Aufwand erlauben.

Ein linearer Zusammenhang zwischen retinaler dendritischer Feldgröße und Exzentrizität entspricht zudem gerade der Größen- und Raumfrequenzskalierung visueller Teilmerkmale mit geänderter Sehentfernung. Der inhomogene Aufbau der Retina würde dem-

¹⁴Das hätte die interessante Konsequenz, daß Reize konstruierbar sein müßten, deren Größenschätzung erschwert oder unmöglich ist, da sie die entsprechenden Schlüsselmerkmale nicht enthalten.

¹⁵Der Durchmesser der dendritischen Felder kann als ein Hinweis auf die Größe der retinalen rezeptiven Felder der betreffenden Zellen gelten.

nach die Ausbildung der Entfernungs-*Complex*-Zellen unterstützen, welche die entsprechenden Reiztransformationen vermitteln.¹⁶

Transfer zwischen Objekten

Objektrepräsentationen, die auf den Signalen der Entfernungs-*Complex*-Zellen aufbauen, sind sofort entfernungsinvariant: Die entsprechende Eigenschaft der *Complex*-Zellen wird in der neuronalen Hierarchie „vererbt“ (Abschnitt 6.5.3).

Dies gilt ebenfalls für jede neu erlernte Objektrepräsentation, gleich, auf welche Weise sie erlernt wurde. Insbesondere ist es ausreichend, eine Objektrepräsentation an einem einzelnen Repräsentanten, also in einer einzigen Sehentfernung erlernt zu haben. Das Wissen über die (physikalisch determinierte) distanzabhängige Transformation einzelner Objektmerkmale ist in den Netzwerkverbindungen der Merkmalsdetektoren auf die Entfernungs-*Complex*-Zellen fest abgelegt. Somit kann jedes neu erlernte Objekt sofort in allen Entfernungen wiedererkannt werden. Es muß nicht in sämtlichen Distanzen trainiert werden, um eine verlässliche Erkennung zu erlauben.

Das vorgestellte Modell ermöglicht das Wiedererkennen eines einmal erlernten Objektes in allen Sehentfernungen. Dazu ist es ausreichend, die Objektrepräsentation in einer einzigen Sehentfernung zu erlernen.

Überlappende Repräsentantenmengen

Konventionelle Ansätze zur Generierung visueller Invarianz, wie die ungesteuerte *Complex*-Bildung oder das dynamische *Routing*, leiden unter dem Problem der Ununterscheidbarkeit von Objekten mit überlappenden Repräsentantenmengen (Abschnitt 2.1.3). Im Zusammenhang mit entfernungsinvarianter Wahrnehmung sei hier an das Beispiel *Modellauto und echtes Auto* erinnert: Beide physikalisch sehr wohl unterschiedlichen Objekte besitzen in entsprechenden Sehentfernungen identische Netzhautbilder. Kein Erkennungssystem, das allein retinale visuelle Merkmale auswertet, kann diese beiden Objektklassen verlässlich unterscheiden.¹⁷

¹⁶Eine unmittelbare Bedeutung des retinalen Aufbaus für die größen- und rotationsinvariante visuelle Verarbeitung, auch im Zusammenhang mit einer möglichen mathematisch *log-polaren* Projektion des Sehreizes durch die retino-thalamo-kortikalen Verbindungen, wurde u.a. bei Schwartz (1981); Reitboeck und Altmann (1984); Hartmann et al. (1993); Busemann et al. (1993) diskutiert.

¹⁷Sofern nicht eine Entfernungsschätzung aufgrund des Bildinhaltes generiert wird, s. folgende Abschnitte *Verfügbarkeit des Steuersignals* und *Extraktion aus dem*

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

Das im Modell verwendete Prinzip der situativ gesteuerten *Complex*-Bildung mit der Sehentfernung als Steuerparameter erlaubt eine Lösung dieses Problems. Wir haben uns bereits klargemacht, daß dieses Verfahren eine Schätzung der physikalischen Objektgröße erlaubt. Damit ist offensichtlich, daß die beiden physikalisch verschiedenen großen Objekte unterscheidbar bleiben, selbst wenn sie identische Netzhautbilder besitzen.

Zum besseren Verständnis will ich erklären, wie sich diese Unterscheidung auf der Ebene einzelner neuronaler Antworten im Modell darstellt. Die in Form des Steuerparameters eingehende Zusatzinformation über die äußeren Sehbedingungen wird hier zur Unterscheidung herangezogen. Zur Illustration dient Abbildung 6.18. Dargestellt ist die Verarbeitung zweier physikalisch verschieden großer runder Objekte in Sehentfernungen, die identische Netzhautbilder verursachen:

1. Sollen zwei physikalisch verschieden große Objekte gleiche Netzhautbilder besitzen, so müssen sie sich in verschiedenen Sehentfernungen befinden (A).
2. Ist dies der Fall, so werden sie bei Fixation durch verschiedene Klassen von Merkmalsdetektoren (an gleichen retinalen Positionen) repräsentiert (B).
3. Aufgrund der durch die Auswahlregel in Abschnitt 6.5.2 gegebenen Systematik müssen Merkmalsdetektoren an gleichen retinalen Stellen, jedoch mit verschiedener Entfernungspräferenz, auf *verschiedene Entfernung-Complex-Zellen* konvergieren (C).
4. Verschiedene Verteilungen von *Complex*-Zellen aktivieren unterschiedliche Objektrepräsentationen (D).

Die Verschiedenheit der beiden Objekte ist damit wahrnehmbar. Die räumliche Größe des jeweils betrachteten Objektes ist aufgrund der Ordenbarkeit der Objektrepräsentationen zugänglich.

Das Problem überlappender Repräsentantenmengen wird durch das Prinzip der situativ gesteuerten Complex-Bildung mit der Sehentfernung als Steuerparameter gelöst.

retinalen Reiz.

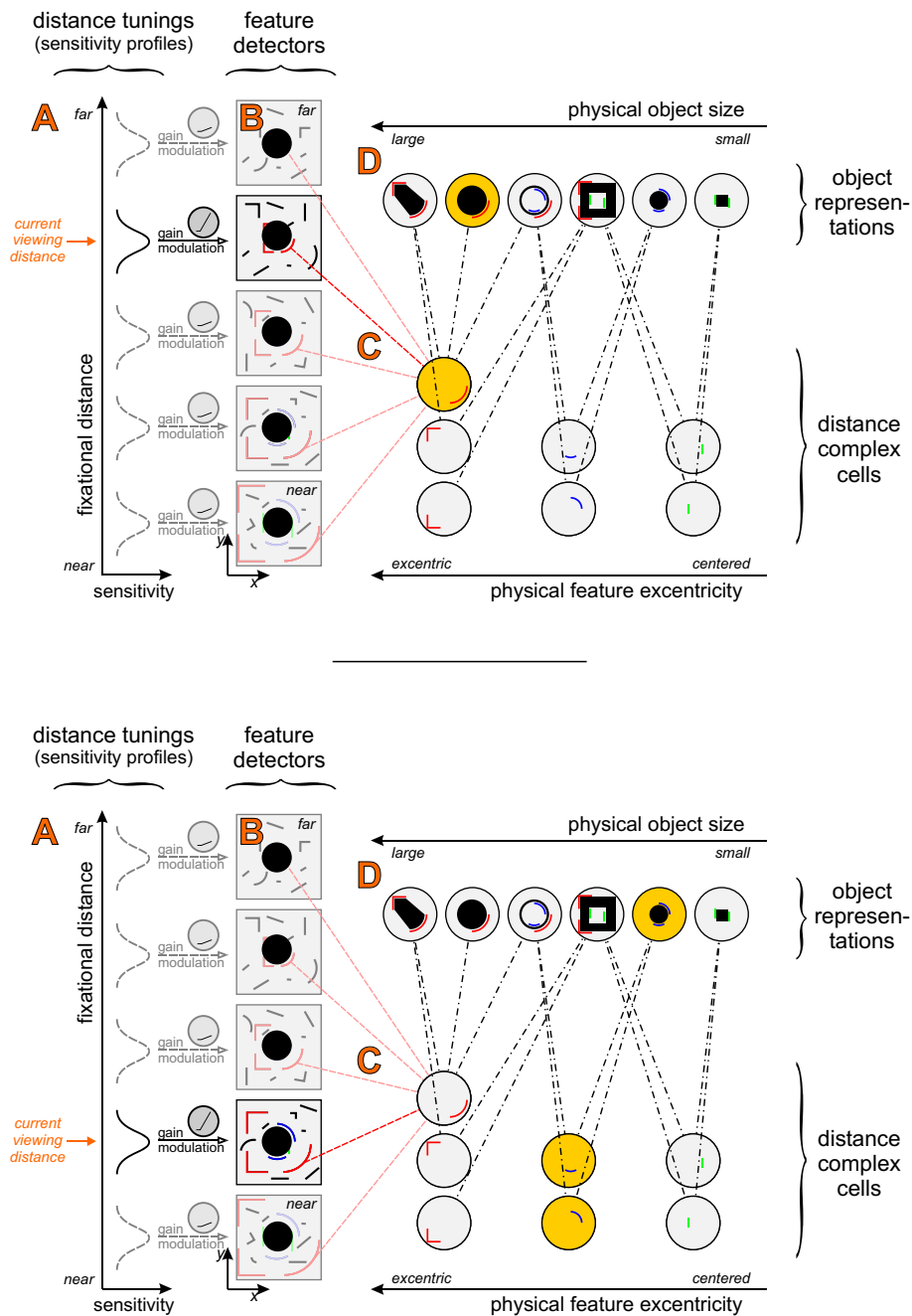


Abbildung 6.18: Lösung des Repräsentantenproblems im Modell. **Oben:** Großes rundes Objekt in weiter Sehentfernung. **Unten:** Kleines rundes Objekt in naher Sehentfernung. Beide Objekte besitzen identische retinale Projektionen, dargestellt als identische Kreise im Bereich der Merkmalsdetektoren (B). Die momentane Sehentfernung ist links eingetragen (A). Insensitive Merkmalsdetektoren sind mit niedrigem Kontrast dargestellt (B). Aktive Entfernungs-Complex-Zellen und Objektrepräsentationen sind orange dargestellt (C und D).

Verfügbarkeit des Steuersignals

In das Modell geht als Steuerparameter der neuronalen Sensitivitäten eine Schätzung der aktuellen Sehentfernung ein.

Ein entsprechendes Signal kann das visuelle System aus einer Vielzahl von Quellen gewinnen. Hierzu zählen die zur Fixation erforderliche Vergenz, die Akkomodation der Augenlinsen oder der vertikale Sehwinkel relativ zum Horizont (Ooi et al., 2001). Auch *bildhafte Hinweise* wie perspektivische Längenverkürzung, Kontrastverlauf, räumliche Frequenzgradienten von Oberflächentexturen, Verdeckung, oder Parallaxe bewegter Szenenbestandteile liefern eine entsprechende Schätzung.

Betrachten von Bildern

Dieser Umstand ist dann von entscheidender Bedeutung, wenn eine Objekterkennung *gerade nicht* auf realen Objekten in der dreidimensionalen Umwelt operieren soll – sondern auf deren zweidimensionalen Projektionen, wie etwa auf Bildern, Zeichnungen oder im Film. Beim Betrachten einer stark verkleinerten Photographie eines Gegenstandes hat der Betrachter normalerweise *nicht* die Wahrnehmung eines „Miniaturgegenstandes“. Eine mögliche Erklärung hierfür wäre, daß die unter diesen Bedingungen irreführenden Hinweise *Vergenz* und *Akkomodation* nicht verwendet werden, sondern die Kalibrierung des Systems allein aufgrund bildhafter Merkmale geschieht. Erst wenn der Betrachter bewußt versucht, etwa auf eine Aufforderung hin, die Größe des *Abbildes auf der Photographie* zu beurteilen, wird das System, z.B. unter Verwendung des Hinweises *Vergenz*, auf den tatsächlichen Leseabstand kalibriert und erlaubt dann eine korrekte Beantwortung dieser (neuen) Fragestellung.

Aktuelle Forschungsergebnisse zeigen, daß bei der Bildung einer Entfernungsschätzung einzelne Entfernungshinweise je nach Aufgabenstellung, Situation oder Verlässlichkeit mit unterschiedlicher Gewichtung einfließen können (Landy et al., 1995; Atkins et al., 2001; Hillis et al., 2002).

Mondtäuschung

Die Verwendung bildhafter Hinweise zur Entfernungskalibrierung im vorgestellten Modell erlaubt nicht nur das Betrachten von Photographien und Filmen, sondern kann auch zur Erklärung einiger sogenannter „optischer“ Täuschungen herangezogen werden. So ist ein Erklärungsansatz für die Wahrnehmung eines übergroßen Mondes am Horizont (*Mondtäuschung*) eine Fehleinschätzung der Mondentfernung aufgrund bildhafter Hinweise in den Projektionen der auf der Erdoberfläche verteilten Gegenstände.¹⁸

¹⁸Die korrekte Erklärung dieser Täuschung ist möglicherweise komplizierter, vgl. McCready (1999).

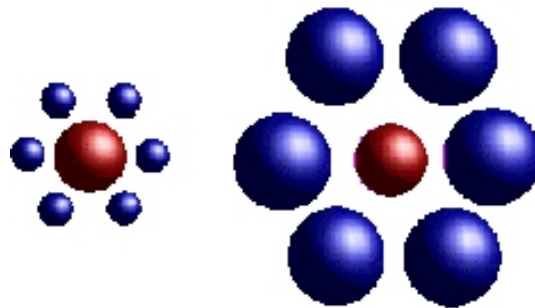


Abbildung 6.19: Ebbinghaus-Täuschung. Die mittleren Kugeln sind auf dem Papier identisch groß. Abdecken der Außenringe bringt die Täuschung zum Verschwinden.

Auch die in Abbildung 6.19 dargestellte *Ebbinghaus-Täuschung* ist so erklärbar: Obwohl die mittleren Kugeln in der zweidimensionalen Abbildung gleichen Durchmesser besitzen, erscheint die rechte Kugel kleiner. Eine mögliche Ursache hierfür ist, daß die umgebenden Kugeln aufgrund eines Frequenzgradienten den Eindruck vermitteln, die linke Anordnung sei weiter vom Betrachter entfernt als die rechte. In diesem Fall würde der linken roten Kugel – bei identischem retinalen Reiz – eine größere räumliche Ausdehnung zugeordnet als der rechten. Zum besseren Verständnis kann Abbildung 6.18 und deren Erläuterung im Text herangezogen werden. Die dort dargestellten Reize entsprechen gerade dieser Konstellation. Sie ist eine Erscheinungsform des Repräsentantenproblems, das von diesem Modell gelöst wird.

Ebbinghaus-Täuschung

Weitere Beispiele für Täuschungen, die durch „fehlerhafte“ Entfernungsschätzungen erklärt werden können, sind die *Müller-Lyer*- und die *Ponzo-Täuschung*.

Schließlich mögen sogar interne Faktoren in die verwendete Steuergröße eingehen, und so eine aktive, willentliche Größentransformation des visuellen Reizes ermöglichen. Beispielsweise dann, wenn wir gebeten werden, aus einer Reihe von Modellautos dasjenige auszusuchen, das einem echten Vergleichswagen entspricht.

Aktive Kontrolle

Selbst das Ausschalten jeglicher Distanzmodulation, etwa im Falle, daß keine verlässlichen Entfernungshinweise vorliegen, könnte dem Erkennungssystem noch eine grobe Orientierung erlauben. Wie aus Abbildung 6.16 ersichtlich ist, mag die erzeugte Repräsentation noch spezifisch genug sein, um eine eingeschränkte Erkennung zu ermöglichen. Ein solcher Arbeitsmodus könnte entscheidend sein für die schnelle Generierung einer ersten Hypothese über die betrachtete Szene (Körner et al., 1997, 1999). Diese könnte noch vor ei-

Schnelle Hypothese

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

ner korrekten Einstellung des Vergenzwinkels oder der Auswertung bildhafter Hinweise gebildet werden.

Extraktion aus dem retinalen Reiz

Bildhafte Hinweise wie Kontrast- und Frequenzgradienten, Parallaxe oder perspektivische Verkürzung können in aller Regel nur Information über die *relative* Entfernung von Gegenständen zueinander liefern. Eine *absolute* Schätzung für die auf einer Photographie dargestellten Sehentfernungen kann dagegen aus dem Bildinhalt selbst extrahiert werden. Dies wird möglich, da im Modell den einzelnen Objektrepräsentationen explizite räumliche Größen zugeordnet sind. Einzelne Schlüsselobjekte mit sehr charakteristischer Form, oder von großer Verhaltensrelevanz, etwa menschliche Körper oder Gesichter, können sehr wahrscheinlich auch ohne eine konkrete Entfernungsschätzung erkannt werden. Das Netzwerk darf sich hierfür in einem „neutralen Zustand“ ähnlich Abbildung 6.16 befinden. Gelingt es, ein beliebiges bekanntes Objekt in der Szene wiederzuerkennen, so kann aus dessen bekannter physikalischer Größe und dem retinalen Reiz eine Entfernungsschätzung implizit generiert werden. Die Netzwerkhierarchie wird hierzu in umgekehrter Reihenfolge durchlaufen. Zur Verdeutlichung dieses Vorgangs kann Abbildung 6.18 herangezogen werden, die bereits die Lösung des Repräsentantenproblems illustrierte:

Ausgehend von der Repräsentation eines einzigen erkannten Objektes (D) können die zugehörigen Entfernungs-*Complex*-Zellen aktiviert werden (C). Diese wiederum erzeugen eine Voraktivierung der mit ihnen verbundenen Merkmalsdetektoren (B). Wie man sich leicht klarmacht, wird die Voraktivierung dieser Merkmalsdetektoren nur in einer einzigen der *D* entfernungs sensitiven Zellklassen (B) durch den retinalen Reiz unterstützt. Neuronale Aktivität wird in jenen Zellen vorherrschen, die sowohl durch die Objekterwartung (*top down*) voraktiviert, als auch durch den retinalen Reiz (*bottom up*) getrieben werden. Unter der plausiblen Annahme, daß zwischen den Neuronen einer Entfernungsklasse laterale unterstützende Verbindungen existieren, zwischen Neuronen verschiedener Entfernungsklassen jedoch hemmende¹⁹, wird sich durch die in einer bestimmten

¹⁹Dies entspräche einem *Center/Surround*-Profil für die Entfernung. Verschaltungsprofile dieser Art bilden ein allgegenwärtiges Designprinzip in der kortikalen Verarbeitung. Die entsprechende Verschaltung kann unter normalen Sehbedingungen erlernt werden, unter denen Aktivität durch die externe Steuerung nur jeweils in einer Zellklasse (oder angrenzenden Klassen) vorliegt.

6.9 Experimenteller Nachweis möglicher Entfernung-Complex-Zellen

Zellklasse vorherrschende Aktivität eine kohärente *implizite* Schätzung der Fixationsentfernung über das gesamte Sehfeld einstellen.

Ist eine absolute Entfernungsschätzung für ein Sehobjekt der Szene gelungen, so können vielfältige bildhafte Hinweise zur Schätzung der übrigen Objektentfernungen herangezogen werden.

6.9 Experimenteller Nachweis möglicher Entfernung-Complex-Zellen

Die Existenz der im Modell beschriebenen Entfernung-Complex-Zellen ist rein hypothetisch. Ihre Postulierung wurde notwendig, um mittels des Prinzips der situativ gesteuerten *Complex*-Bildung eine entfernungsinvariante Repräsentation erzeugen zu können. Unter der Voraussetzung der Entfernungsmodulation von Merkmalsdetektoren im visuellen System, für die experimentelle Nachweise existieren (Abschnitt 6.3.2), ergeben sie sich als die Zielneuronen der *Gating*-Module nach Abbildung 5.4.

Hochinteressant ist in diesem Zusammenhang der kürzlich gelungene Nachweis einer bisher unbekanntem Zellklasse im visuellen Kortex des Affen (Pigarev et al., 2002). Diese findet sich in einem kortikalen Gebiet, das bislang als ein Bestandteil des visuellen Areals V4 klassifiziert wurde. Pigarev et al. identifizieren die Neuronen in diesem mit V4A bezeichneten Areal aufgrund ihrer rezeptiven Feldeigenschaften. Diese weisen bemerkenswerte Ähnlichkeit mit den Eigenschaften auf, die für die im beschriebenen Modell postulierten Entfernung-Complex-Zellen zu erwarten sind.

Neu klassifiziertes Areal V4A

6.9.1 Feldeigenschaften der Entfernung-Complex-Zellen

Aus dem schematischen Aufbau des Modells in den Abbildungen 6.6–6.8 sind folgende rezeptiven Feldmerkmale für die Entfernung-Complex-Zellen abzuleiten. Sie sind in Abbildung 6.20 (oben) illustriert:

1. Die individuellen Stimuli, die zur Aktivierung der *Complex*-Zellen führen, haben die retinale Form einfacher, lokalisierter Merkmale, etwa die kurzer Konturelemente.
2. Für eine gegebene *Complex*-Zelle decken diese individuellen Stimuli in ihrer Gesamtheit einen sektorförmigen Ausschnitt des zweidimensionalen Sehfeldes ab (siehe Abbildung 6.8, A).

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

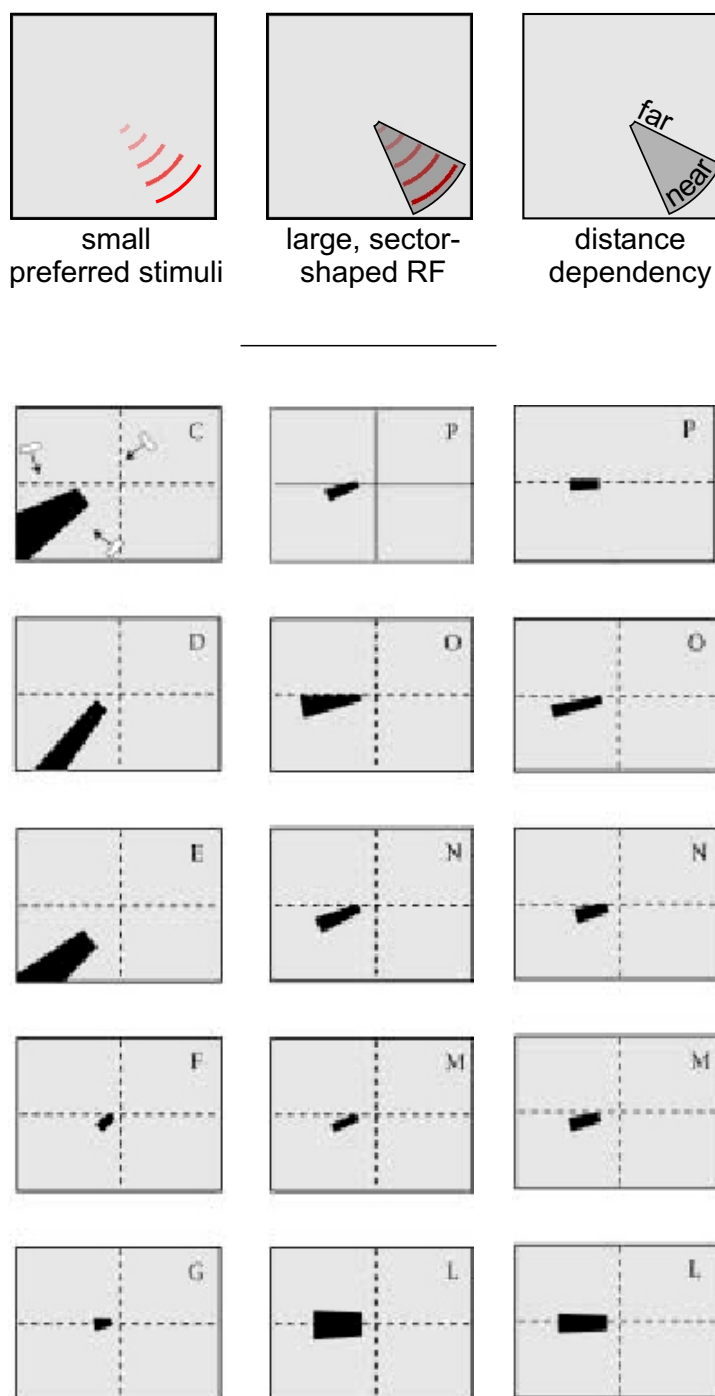


Abbildung 6.20: Retinale receptive Felder und ihre Eigenschaften. **Oben:** Entfernung-Complex-Zellen, wie aus dem Modellaufbau erwartet. **Unten:** Neuronen in V4A. Receptive Felder (schwarz) und Beispiele für Testreize (weiß). Ausschnitte entsprechen $40^\circ \times 30^\circ$ Sehwinkel. (Abbildung aus Pigarev et al., 2002).

6.9 Experimenteller Nachweis möglicher Entfernungs-Complex-Zellen

3. Dieser sektorielle Ausschnitt ist radial orientiert und erheblich größer als die individuellen Merkmale. Er wird in keinem Fall die Fovea abdecken.
4. Die individuellen Stimuli aktivieren die entsprechende *Complex-Zelle* nur dann, wenn sie in geeigneter Sehentfernung präsentiert werden. Reize in fester Sehentfernung aktivieren die *Complex-Zelle* nur in einem radialen Ausschnitt des Sektors, der durch die Breite der Sensitivitätsprofile gegeben ist.

6.9.2 Feldeigenschaften der Neuronen in V4A

Pigarev et al. (2002) beschreiben in ihrer Arbeit die rezeptiven Feldeigenschaften der neu klassifizierten Nervenzellen als ungewöhnlich für das visuelle Areal V4: Die Neuronen besitzen große, radial orientierte rezeptive Felder, deren Form als „kometenartig“ beschrieben wird. Keines der gemessenen Felder überdeckte die Fovea. Einige der Meßergebnisse sind in Abbildung 6.20 (unten) wiedergegeben.

Die rezeptiven Felder wurden mit Testreizen in Form orientierter Balken ausgemessen. Die bevorzugten retinalen Balkenreize waren stets viel kleiner als das gesamte rezeptive Feld. Ihre Ausdehnung war geringer als die des schmalen Endes des „kometenartigen“ Feldes. Viele Neuronen waren sensitiv für radial bewegte Reize. Die inneren „Köpfe“ der rezeptiven Felder hatten scharfe Begrenzungen, die der peripheren „Schwänzen“ waren diffus.

Zur Präsentation der Testreize kam ein Monitor im festen Sehabstand von 57 cm zum Einsatz. Für einen großen Anteil der untersuchten Zellen (29 von 50 bei einem Tier) konnte mit dieser Methode keinerlei Aktivierung hervorgerufen werden. Die entsprechenden Neuronen zeigten jedoch Aktivierung im Zusammenhang mit Augenbewegungen des Affen in den Phasen zwischen den Fixationsaufgaben. Pigarev et al. schlossen daraus, daß die auf dem Monitor präsentierten Reize suboptimale Stimuli darstellten. Sie bestätigten diese Vermutung, indem sie *reale Objekte* aus der Laborumgebung zur Stimulation einsetzten, z.B. einen Schwamm. Diese realen Objekte riefen in vielen der zuvor nicht aktivierten Neuronen „heftige Antworten“ hervor. Die Antworten waren besonders stark, wenn die Stimuli in größeren Entfernungen präsentiert wurden.

Stimulationsentfernung

6.9.3 Interpretation

Die von Pigarev et al. (2002) beschriebenen neuronalen Eigenschaften entsprechen weitgehend denen, die für Entfernungs-*Complex-*

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

Zellen zu erwarten wären. Leider wurde eine wesentliche erwartete Eigenschaft, nämlich der systematische Zusammenhang zwischen Stimulationsentfernung und retinaler Exzentrizität, nicht untersucht. Testreize zur Bestimmung der rezeptiven Felder wurden nur in einer Sehentfernung von 57 cm präsentiert. Die als „kometenförmig“ beschriebenen Felder haben allerdings gerade die Form, die für Stimulation in einer festen Entfernung zu erwarten ist, nämlich radiale Ausschnitte retinaler Sektoren.

Ich möchte im folgenden die Annahme machen, daß es sich bei den betreffenden Zellen tatsächlich um Entfernungs-Complex-artige Zellen handelt. Dieser Schluß ist nicht zwingend. Ich werde die Beobachtungen von Pigarev et al. (2002) unter dieser Annahme diskutieren.

Radiale Ausdehnung und Sehentfernung

Viele der von Pigarev et al. bestimmten „kometenförmigen“ Felder haben eine große radiale Ausdehnung. Sollte es sich bei den betreffenden Neuronen tatsächlich um Entfernungs-Complex-artige Zellen handeln, so müßten diese stark überlappende Distanzcharakteristiken besitzen. Eine andere, vielleicht plausiblere Erklärung für die große radiale Ausdehnung wäre, daß sich während der Bestimmung des Feldes der Steuerparameter verändert hat, der die Sensitivierung der Zellklassen beeinflusst: Pigarev et al. (2002) beschreiben, daß die Messung der rezeptiven Felder interaktiv verlief, und pro Feld bis zu 15 Minuten dauern konnte. Obgleich der Affe während der Feldbestimmung den Zielpunkt am Monitor fixieren mußte (erlaubte Abweichung 1° Sehwinkel), so ist doch keineswegs gewiß, daß alle Parameter, welche die Zellsensitivität beeinflussen, über eine so große Zeitspanne hinweg konstant waren. Die Güte der Fixation wurde mit Hilfe der *Search-Coil*-Technik bestimmt. Es ist anzunehmen, daß die Implantation nur an einem Auge vorgenommen wurde. Betreffend den Hinweis *Vergenz* auf die Sehentfernung kann somit nicht ausgeschlossen werden, daß der tatsächlich binokular fixierte Punkt vor oder hinter dem Monitor lag. Selbst wenn die Stellung beider Augen überprüft wurde, so entspricht eine Abweichung von max. 1° Sehwinkel pro Auge in 57 cm Entfernung einem Distanzbereich von 36–132 cm (bezogen auf einen Augenabstand von 3,5 cm bei *macaca fascicularis*). Zudem hatten wir in Abschnitt 6.8.2 vermutet, daß die Vergenz bei der Betrachtung eines flachen Monitorbildes möglicherweise nicht als Hinweis in die Entfernungsschätzung eingeht, sondern eine Schätzung aufgrund bildhafter Hinweise oder erkannter

6.9 Experimenteller Nachweis möglicher Entfernungs-Complex-Zellen

Objekte erfolgt. Zu welchem Ergebnis das vermutlich übertrainierte Versuchstier in der Anwesenheit eines einzelnen Fixationskreuzes und einiger Balkenreize hier gelangte, dürfte kaum abzuschätzen sein. Tatsächlich ist eine nicht unplausible Annahme, daß sich das Sehsystem in einem neutralen Zustand ähnlich Abbildung 6.16 befand. In diesem Fall sind radial sehr weit ausgedehnte rezeptive Felder für die Entfernungs-*Complex*-Zellen zu erwarten.

Bevorzugte Stimuli

Die stark erhöhte Zellantwort für die Reizung mittels realer Objekte wie einem Schwamm ist in der Tat ein Hinweis darauf, daß die verwendeten Balkenreize mangelhafte Stimuli zur Aktivierung der betreffenden Neuronen darstellten. Ein Grund hierfür wäre, daß die Entfernungs-*Complex*-Zellen günstigerweise auf den Signalen von Merkmalsdetektoren für Merkmale höherer Komplexität gebildet werden, als es orientierte Kontrastkanten sind (Abschnitt 6.8.1). Eine weitere Erklärung ist, daß mit rellen, dreidimensionalen Objekten in der räumlichen Sehumgebung des Betrachters eine ausgeprägtere Entfernungsschätzung verknüpft ist, als bei der Betrachtung eines Monitorbildes. Die Tatsache, daß diese Objekte vor allem in größerer Präsentationsentfernung Zellen aktivierten, die in 57 cm Sehentfernung nicht ansprachen, ist ein zusätzlicher Hinweis auf eine distinguierte Distanzselektivität der beteiligten Neuronen.

Distanzcharakteristik

Sollten die Antworten der von Pigarev et al. (2002) beobachteten Neuronen wirklich auf Nervenzellen mit Entfernungscharakteristiken ähnlich Abbildung 6.14 beruhen, so weisen die scharfen zentralen Begrenzungen der gemessenen Felder, im Gegensatz zu diffuseren Grenzen im peripheren Bereich, auf Sensitivitätsprofile mit in der Entfernung zunehmender Breite und Überlapp hin. In diesem Fall wäre eine angepaßte Verteilung nach Abbildung 6.14 (unten) möglich. Aussagen hierüber sind jedoch rein hypothetisch.

Radiale Bewegungsselektivität

Hinsichtlich der in Abschnitt 6.5.2 formulierten Auswahlregel ist es bemerkenswert, daß viele der untersuchten Neuronen für Reizbewegungen in radialer Orientierung empfindlich waren. Eine plausible Hypothese für die Ausbildung von Entfernungs-*Complex*-Zellen der beschriebenen Form ist das Lernen auf zeitkontinuier-

Kapitel 6 Ein konzeptionelles Modell zur Entfernungsinvarianz

lichen Stimuli (Földiák, 1991; Wallis, 1996; Wallis und Rolls, 1997; Rolls und Milward, 2000; Wiskott und Sejnowski, 2002, vgl. auch Abschnitt 2.1.2). Ein entsprechendes Reizparadigma wäre in diesem Fall die Beobachtung eines realen Gegenstandes, der sich auf den Betrachter zu oder von ihm weg bewegt. Bewegte Reize dieser Art enthalten ausschließlich radiale retinale Geschwindigkeitsanteile.

6.9.4 Schlußbemerkung

Die Beobachtungen von Pigarev et al. (2002) sind nicht hinreichend, um auf die Existenz von Entfernung-*Complex*-Zellen der im Modell geforderten Art im Primatenkortex zu schließen. Die Übereinstimmung mit den erwarteten Eigenschaften dieser Neuronen ist jedoch zweifellos groß genug, um die Modellhypothese zu unterstützen, und um weitere Untersuchungen auf diesem Feld zu motivieren. Hier schließt sich eine Überprüfung des vermuteten systematischen Zusammenhangs zwischen retinaler Reizexzentrizität und Fixationsentfernung bei der Aktivierung von Neuronen in V4A an.

Die Entfernungsabhängigkeit der neuronalen Antworten im Areal V4A wird derzeit an der Katze mit Hilfe frei positionierbarer Echtweltstimuli ausgewertet (Pigarev 2004, persönliche Kommunikation). Die Ergebnisse dieser Untersuchungen können weiteren Aufschluß über eine Rolle der Nervenzellen in V4A bei der entfernungs-invarianten visuellen Verarbeitung geben.

Kapitel 7

Zusammenfassende Diskussion

Ziel

Hier vollziehe ich die Argumentationslinie der vorausgegangenen Kapitel nach, und diskutiere deren wesentliche Ergebnisse für die Bildung geometrischer Invarianzen der Wahrnehmung. Ich identifiziere das Konzept der externen Steuerung neuronaler Codierungseigenschaften und dessen Anwendung in der situativ gesteuerten *Complex*-Bildung als universell einsetzbares Konzept zur Generierung von Invarianzen. Ich fasse die mögliche Bedeutung dieser Modellansätze, sowie deren Vor- und Nachteile zusammen. Die Methode der konzeptionellen Modellierung erweist sich als effektives Werkzeug zum Verständnis kortikaler Verarbeitung.

Die invariante visuelle Wahrnehmung von Sehobjekten, wie sie sich täglich in unserem Nervensystem vollzieht, gibt aus Sicht der theoretischen Neurophysik nach wie vor Rätsel auf. Sie ist im Rahmen der verfügbaren neuronalen Modelle nur mit Einschränkungen zu erklären. Diese Einschränkungen nehmen typischerweise die Form subtiler Unentscheidbarkeiten in der neuronalen Repräsentation an, wie die des Bindeproblems, oder des Problems der überlappenden Repräsentantenmengen. Die Standardmodelle basieren auf einer hierarchischen Anordnung von Nervenzellen, deren Ziel es ist, spezifische neuronale Antworten für komplexe visuelle Reize aus Antworten auf einfache, universelle Reizkomponenten zu konstruieren. Die Generierung reizspezifischer Zellantworten läuft jedoch gerade der (verallgemeinernden) Invarianzbildung entgegen – und zwar sowohl auf Ebene des konkreten physikalischen Signalflusses im Netzwerk, wie auch als Denkmodell und Konzept der Modellierung. Es

war der Gegenstand dieser Arbeit, sich diesem Problem vom Blickwinkel der konzeptionellen Modellierung aus zu nähern.

7.1 Verlauf der Argumentation

Zentraler Ausgangspunkt meiner Arbeit ist der Umstand, daß das kortikale Sehsystem nicht isoliert im Labor vorliegt, sondern in ein lebendiges Individuum eingebettet ist, das seine Umwelt aktiv wahrnimmt, und zwar „aktiv“ im Sinne echter körperlicher Aktion. Es ist völlig plausibel anzunehmen, daß das Körper-Umwelt-Verhältnis, in dem sich das Individuum befindet, hochinformativ Parameter für die (visuelle und sonstige) Wahrnehmung beinhaltet, und daß diese in die neuronale Signalverarbeitung einfließen.¹ In Verbindung mit psychophysischen und neurophysiologischen Befunden zur entfernungs- und blickrichtungsabhängigen Verarbeitung visueller Reize führt mich dies zur Formulierung der Annahme, daß äußere physikalisch-körperliche Umstände die Verarbeitung im visuellen Formerkennungssystem beeinflussen, indem Übertragungseigenschaften einzelner Neuronen durch eine externe Steuergröße moduliert werden.

Diese Prämisse dient der Auflösung des Konfliktes zwischen der gegenläufigen Bildung von Reizselektivität und Invarianz – auf beiden Ebenen, konzeptionell wie physikalisch: Ein disambiguierendes Steuersignal geht auf unterer oder mittlerer Ebene in der neuronalen Hierarchie ein, und ermöglicht die invarianzbildende Auswahl von Teilmerkmalen, auf deren Basis dann reizspezifische Zellantworten generiert werden.

Um zu demonstrieren, welche Auswirkung die Variation eines einzelnen neuronalen Modellparameters auf die Codierungseigenschaften von Neuronen haben kann, widme ich mich mit dem *Schwellen-* oder „*Eisberg*“-Effekt einem Modellierungsprinzip, das bereits die Umsetzung einer Reihe interessanter Aspekte für die visuelle Verarbeitung, wie die Auflösungskontrolle und Durchstimmung von Frequenzkanälen erlaubt. Die Modellierung von Positions- oder Größeninvarianz erfordert allerdings die systematische Modulation ganzer Neuronengruppen. Die beiden konzeptionellen Modelle zur retinalen Schlupfkorrektur und zur Generierung von Entfernungsinvarianz verfolgen diesen Ansatz.

¹Eine ähnliche Prämisse hat unter dem Begriff *Embodiment* in den letzten Jahren Einzug in die neurobiologische Modellbildung gehalten. Sie geht davon aus, daß wesentliche neuronale Leistungen, bis hin zu Bewußtsein und Intelligenz, zu ihrer Ausbildung den physischen Körper des Individuums erfordern.

Die Notwendigkeit, verschiedenen Neuronengruppen aufgrund der gleichen externen Bedingung unterschiedlich modulieren zu können, führt mich zur Postulierung klassenspezifischer *Sensitivitätsprofile*. Diese haben konzeptuelle Ähnlichkeit mit den im parietalen Kortex nachgewiesenen *Gain Fields*. Meine Annahme wird gestützt durch die experimentelle Klasseneinteilung der blickrichtungs- und entfernungsmodulierten Zellen in den Arealen V1, V2 und V4. Basierend auf den individuellen Sensitivitätsprofilen konstruiere ich vielseitig einsetzbare hierarchische Modellbausteine, sogenannte *Gating-Module*, und führe damit das Konzept der *situativ gesteuerten Complex-Bildung* ein. Dieses Konzept verwende ich mit Erfolg in zwei Modellen zur invarianten neuronalen Verarbeitung im Sehsystem. Deren Leistungsfähigkeit demonstriere ich in Computersimulationen unter biologisch relevanten Randbedingungen.

Im Modell der retinalen Schlupfkorrektur erlaubt der Einsatz der situativ gesteuerten *Complex-Bildung* die Generierung einer stabilisierten Repräsentation der durch Mikroaugenbewegungen verwackelten retinalen Projektion. Der hierzu verwendete externe Steuerparameter ist der als momentane Blickrichtung gegebene Fixationsfehler.

Im Modell zur Erzeugung von Entfernungsinvarianz hat die Verwendung des Konzeptes der situativ gesteuerten *Complex-Bildung* die Formierung von Neuronen mit neuartigen Codierungseigenschaften zur Folge. Diese *Entfernungs-Complex-Zellen* vermitteln den durch die optische Abbildung im Auge determinierten Zusammenhang zwischen der räumlichen Entfernung physikalischer Objektbestandteile und ihrer entfernungsvarianten zweidimensionalen Projektion auf die Netzhaut. Neuronen mit sehr ähnlichen Codierungseigenschaften wurden in jüngster Zeit experimentell nachgewiesen. Der verwendete externe Steuerparameter ist eine Schätzung der augenblicklichen Sehentfernung, die aus einer Vielzahl von Hinweisen gewonnen werden kann. Die *Entfernungs-Complex-Zellen* erzeugen eine Repräsentation der Szene in entfernungsinvarianten Teilmerkmalen. Auf dieser Repräsentation kann eine Objekterkennungshierarchie aufsetzen, in der eine neuronale Schätzung der räumlichen Größe wahrgenommener Objekte möglich ist. Das Problem der Unterscheidung von Objekten mit überlappenden Repräsentantenmengen wird von diesem Modell gelöst.

7.2 Beurteilung

In beiden Modellen wird durch die situativ gesteuerte *Complex-* **Verallgemeinerte Invarianzgenerierung**

Kapitel 7 Zusammenfassende Diskussion

Bildung eine Szenenrepräsentation generiert, die vom verwendeten Steuerparameter unabhängig ist. Es ist zu erwarten, daß auf gleiche Weise Invarianz auch gegenüber anderen äußeren Bedingungen generiert werden kann. Dazu ist es lediglich erforderlich, die entsprechende externe Größe als neuronales Steuersignal verfügbar zu machen, und eine Entsprechung der *Auswahlregel* in Abschnitt 6.5.2 für diesen Steuerparameter zu formulieren.

Die situativ gesteuerte Complex-Bildung erweist sich als universell einsetzbares Konzept zur Generierung von Invarianz.

Die so konstruierten Modelle haben Eigenschaften, die sie zu idealen Modulen zur Verwendung in größerskaligen kortikalen Modellen machen:

Neutralzustand

Im Falle der Nichtverfügbarkeit des neuronalen Steuerparameters sind sie in der Lage, einen Neutralzustand einzunehmen. Dies kann eine „generische“ Repräsentation durch Überlagerung aller Zellantworten ähnlich Abbildung 6.16 sein. Alternativ kann sich eine bevorzugte Repräsentation einstellen; im Falle der retinalen Schlupfkorrektur wäre ein sinnvoller Neutralzustand in Abwesenheit eines verlässlichen Steuerparameters etwa die Sensitivierung der Zellklasse mit „zentralen“ rezeptiven Feldern. Für das Entfernungsinvarianzmodell käme außer der überlagerten Repräsentation auch die Sensitivierung der Zellklasse mit erfahrungsgemäß größter Relevanz in Betracht – etwa jener Klasse, die Reize in einer Entfernung codiert, in der typischerweise feindliche Angreifer, Partnertiere, oder Nahrung auftauchen (Entfernungserwartung).

Serialisierbarkeit

Die gesteuerten *Gating*-Module sind hierarchische Modellbausteine und können als solche auf allen Ebenen der neuronalen Verarbeitungshierarchie eingesetzt werden. Insbesondere ist es möglich, Invarianz gegenüber verschiedenen äußeren Bedingungen *nacheinander* zu erzeugen. Beispielsweise sind beide vorgestellten Modelle problemlos miteinander kombinierbar: Auf den erzeugten schlupfkorrigierten Zellantworten könnte eine Entfernungsinvarianz-Stufe der beschriebenen Art aufsetzen, um so eine stabilisierte, entfernungsinvariante Szenenrepräsentation zu erzeugen. Die entsprechenden invarianzgenerierenden Hierarchiestufen lassen sich in jedes hierarchisch organisierte Netzwerkmodell einfügen. Sie sind damit zu fast jedem verfügbaren Modell der visuellen Verarbeitung kompatibel.

Zellzahl

Größter Nachteil der situativ gesteuerten *Complex*-Bildung ist unbestreitbar die sehr hohe Zahl benötigter Neuronen. Obgleich

die entsprechenden Zellklassen nicht unbedingt vollständige Sätze von Merkmalsdetektoren bilden müssen, ist eine Vervielfachung des Zellbedarfs zu erwarten, die mit der erwünschten Güte der Invarianzgenerierung wächst. Inwieweit im visuellen System genügend neuronale Ressourcen zur Verfügung stehen, um die Generierung aller benötigten Invarianzen in ausreichender Güte nach dem vorgestellten Konzept zu ermöglichen, ist schwer zu beurteilen. Swindale (2000) versuchen abzuschätzen, wie viele unabhängige Stimulusmerkmale in miteinander verwobenen neuronalen Karten im Kortex prinzipiell repräsentiert sein können. Sie kommen zu dem Ergebnis, daß die kortikale Zelldichte eine Größenordnung von etwa zehn unabhängigen Karten erlaubt. Eine etwaige Repräsentation von Steuerparametern in kortikalen Karten erlaubte demnach nur eine begrenzte Zahl von Invarianzeigenschaften. Eine vollständige Kartenrepräsentation der Steuerparameter wird jedoch in vielen Fällen nicht erforderlich sein. Einen großen Vorteil stellt hier auch die Kombinierbarkeit des vorgeschlagenen Konzeptes mit der Großzahl der Alternativmodelle dar. So sind optimierte Lösungen für bestimmte Invarianztypen möglich.

Die hier formulierten Modellbestandteile sind konzeptionell. Sie wurden auf der Basis experimenteller Befunde entwickelt; ihre konkreten Ausprägungen sind jedoch als Hypothesen zu werten, deren reelle Relevanz nur aufgrund weiterführender zellphysiologischer Untersuchungen entschieden werden kann. Als ein ermutigender Fortschritt kann hier die Arbeit von Pigarev et al. (2002) gewertet werden. Sollten sich die neu klassifizierten Neuronen im kortikalen Areal V4A tatsächlich als Zellen mit Entfernungs-*Complex*-artigen Eigenschaften erweisen, so wäre dies ein ausgesprochen starker Hinweis auf eine tatsächliche Implementierung des beschriebenen Modells im visuellen Primatenkortex. In diesem Fall bildete die situativ gesteuerte *Complex*-Bildung das konzeptuelle Bindeglied zwischen zwei experimentellen Befunden: der entfernungsabhängigen Merkmalscodierung in V1, V2 und V4 (Dobbins et al., 1998; Rosenbluth und Allman, 2002), und den neuronalen Codierungseigenschaften in V4A. Sie wäre damit ein echter Verständnisbaustein.

In jedem Fall aber erweist sich die konzeptionelle Modellierung als ein Werkzeug zur Bildung effektiver Denkmodelle. Sie ermöglicht uns ein verbessertes Verständnis kortikaler Verarbeitung.

**Konzeptionelle
Modellierung**

Epilog

Zum Phänomen des Scheinriesen

Das in Kapitel 6 vorgestellte Modell zur Entfernungsinvarianz bietet einen Erklärungsansatz für das in Ende (1960) beschriebene Phänomen des *Scheinriesen*. Für die Figur *Herr Tur Tur* wird dort eine Anomalie der wahrgenommenen Körpergröße belegt. Mit zunehmender Sehentfernung scheint die Körpergröße Herrn Tur Turs zu wachsen, so daß dieser aus der Ferne, nicht aber aus der Nähe, einen riesenhaften Eindruck vermittelt. Der Effekt kann als eine gestörte situative Steuerung auf seiten des Betrachters interpretiert werden. Ich möchte hier eine Erklärung im Rahmen meines Modells skizzieren:

1. Aufgrund der optischen Abbildung im Auge wird das retinale Abbild eines jeden Objektes mit zunehmender Entfernung kleiner (perspektivische Längenverkürzung).
2. Diese perspektivische Verkürzung wird vom vorgeschlagenen Invarianzmodell gerade kompensiert. Die Verteilung der Sensitivitätsprofile gewährleistet zusammen mit der beschriebenen Auswahlregel für die Afferenzen der Entfernungs-Complex-Zellen die Aktivierung der stets gleichen Objektrepräsentation. Der retinale Reiz wird bei Entfernung des Objektes zwar noch immer kleiner, die aktivierte Objektrepräsentation und damit die Wahrnehmung ist jedoch die konstanter Größe.
3. Wird für die situative Steuerung ein systematischer Fehler angenommen, *etwa ein zu schnelles Anwachsen des Steuerparameters*, so führt dies zu einer ebenso systematischen Störung der Größenwahrnehmung. Wächst der Steuerparameter mit der Sehentfernung übernormal stark an, so wird die Verteilung der

Zum Phänomen des Scheinriesen

Sensitivitätsprofile schneller als notwendig durchlaufen. Die perspektivische Verkürzung wird damit überkompensiert.

Der retinale Reiz wird bei Entfernung des Objektes zwar noch immer kleiner, die aktivierte Objektrepräsentation und damit die Wahrnehmung ist jedoch die *wachsender Größe*. Das beobachtete Objekt verhielte sich gerade so wie der Scheinriese in Ende (1960).

Fraglich bleibt, wieso Herr Tur Tur das einzige Objekt in ganz Mandala ist, das beim Betrachter eine spontane Überbewertung der Steuergröße verursacht. Der Effekt setzt eine Fernwirkung ungeklärten Ursprungs auf das Nervensystem des Betrachters voraus, zumal Herr Tur Tur davon berichtet, daß sein Scheinriesentum während der Kindheit wesentlich geringer ausgeprägt war, nur etwa halb so stark (Ende, 1960, Seite 133). Zudem sollte die Wahrnehmung von Gegenständen in der Umgebung Herrn Tur Turs in Mitleidenschaft gezogen werden — ein möglicher Grund für seine Wahl der Wüste als bevorzugtem Lebensraum.

Das Phänomen des Scheinriesen bleibt damit weiterhin rätselhaft.

Literaturverzeichnis

- L. F. Abbott, J. A. Varela, K. Sen und S. Nelson (1997). Synaptic depression and cortical gain control. *Science* **275**:220–224.
- R. A. Andersen, G. K. Essick und R. M. Siegel (1985). Encoding of spatial location by posterior parietal neurons. *Science* **230**:450–458.
- R. A. Andersen und V. B. Mountcastle (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **3**:532–548.
- C. H. Anderson und D. C. van Essen (1987). Shiftercircuits: A computational strategy for dynamic aspects of visual processing. *Proc. Natl. Acad. Sci. USA* **84**:6297–6301.
- J. E. Atkins, J. Fiser und R. A. Jacobs (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vis. Res.* **41**:449–461.
- J. L. Barbur (1985). Speed discrimination and its relation to involuntary eye movements in human vision. *Neurosci. Lett.* **54**:7–12.
- P. J. Bennett und F. Cortese (1996). Masking of spatial frequency in visual memory depends on distal, not retinal frequency. *Vis. Res.* **36**:233–238.
- C. M. Bishop (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK. ISBN 0198538642.
- M. C. A. Booth und E. T. Rolls (1998). View-invariant representations of familiar objects by neurons in the inferior temporal cortex. *Cereb. Cortex* **8**:510–523.
- F. Bremmer, F. Klam, J. R. Duhamel, S. Ben Hamed und W. Graf (2002). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.* **16**:1569–86.

Literaturverzeichnis

- F. Bremmer, A. Schlack, W. Graf und J.-R. Duhamel (2004). Multisensory self-motion encoding in parietal cortex. *Vis. Cogn.* **11**:161–172.
- E. Brenner und J. B. J. Smeets (2000). Comparing extra-retinal information about distance and direction. *Vis. Res.* **40**:1649–1651.
- E. Brenner und W. J. M. van Damme (1998). Judging distance from ocular convergence. *Vis. Res.* **38**:493–498.
- E. Brenner und W. J. M. van Damme (1999). Perceived distance, shape and size. *Vis. Res.* **39**:975–986.
- B. Bridgeman und J. Palca (1980). The role of microsaccades in high acuity observational tasks. *Vis. Res.* **20**:813–817.
- H. J. Brinksmeyer, A. Gail, E. Andreu, R. Kupper und R. Eckhorn (2001). Neural correction for residual eye movements during ocular fixation? In N. Elsner und G. W. Kreutzberg, Hg., *The Neurosciences at the Turn of the Century. Proceedings of the 4th Meeting of the German Neuroscience Society 2001. 28th Göttingen Neurobiology Conference*, Bd. II, S. 584. Thieme.
- H. J. Brinksmeyer, F. Michler, A. Gail und R. Eckhorn (2002). Stimulus sensitivity in monkey visual cortex is modulated by viewing distance while spatial frequency tuning and receptive field size are not. In R. P. Würtz und M. Lappe, Hg., *Dynamic Perception*, S. 209–214. Workshop of GI section 1.0.4 “Image Understanding” and the European Networks MUHCI and ECOVISION, infix. Ruhr-University Bochum, Germany.
- P. R. Brotchie, R. A. Andersen, L. H. Snyder und S. J. Goodman (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* **375**:232–235.
- D. C. Burr, M. C. Morrone und J. Ross (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature* **371**:511–513.
- M. Busemann, J. Dunker, G. Hartmann, K.-O. Kräuter, E. Seidenberg und H. Wiemers (1993). Building an artificial retina for distance- and orientation-invariant pattern recognition. In S. Gielen und B. Kappen, Hg., *Proceedings of the International Conference on Artificial Neural Networks ICANN '93*, S. 888–893. Springer-Verlag.
- F. S. Chance, L. F. Abbott und A. D. Reyes (2002). Gain modulation from background synaptic input. *Neuron* **35**:773–782.

- B. G. Cumming und A. J. Parker (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.* **19**:5602–5618.
- J. J. DiCarlo und J. H. Maunsell (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J. Neurophysiol.* **89**:3264–3278.
- R. W. Ditchburn und B. L. Ginsborg (1952). Vision with a stabilized retinal image. *Nature* **170**:36–37.
- A. C. Dobbins, R. M. Jeo, J. Fiser und J. M. Allman (1998). Distance modulation of neural activity in the visual cortex. *Science* **281**:552–555.
- R. Eckhorn, F. Krause und J. I. Nelson (1993). The RF-Cinematogram. *Biol. Cybern.* **69**:37–55.
- M. Ende (1960). *Jim Knopf und Lukas der Lokomotivführer*. Thienemann Verlag.
- C. W. Eurich und H. Schwegler (1997). Coarse coding: calculation of the resolution achieved by a population of large receptive field neurons. *Biol. Cybern.* **76**:357–363.
- D. J. Felleman und D. C. van Essen (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**:1–47.
- P. Földiak (1991). Learning invariance from transformation sequences. *Neur. Comp.* **3**:194–200.
- S. I. Fried, T. A. Münch und F. S. Werblin (2002). Mechanisms and circuitry underlying directional selectivity in the retina. *Nature* **420**:411–414.
- K. Fukushima (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**:193–202.
- A. Gail (2002). *Neuronal basis of perceptual binding processes in visual object coding*. Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.), Fachbereich Physik, Philipps-Universität Marburg. <http://archiv.ub.uni-marburg.de/diss/z2002/0420/>.

Literaturverzeichnis

- M. S. Goldman, C. R. Kaneko, G. Major, E. Aksay, D. W. Tank und H. S. Seung (2002). Linear regression of eye velocity on eye position and head velocity suggests a common oculomotor neural integrator. *J. Neurophysiol.* **88**:659–665.
- M. Greschner, M. Bongard, P. Ruhan und J. Ammermüller (2002). Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. *Nature Neurosci.* **5**:341–347.
- M. Gur und D. M. Snodderly (1997). Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. *Vis. Res.* **37**:257–265.
- G. Hartmann, S. Drüe, K.-O. Kräuter und E. Seidenberg (1993). Simulations with an artificial retina. In *Proceedings of the World Congress on Neural Networks (WCNN '93)*, Bd. 3, S. 689–694.
- J. M. Henderson (2003). Human gaze control during real-world scene perception. *Trends Cogn. Sci.* **7**:498–504.
- J. M. Henderson, C. C. Williams, M. S. Castelhana und R. J. Falk (2003). Eye movements and picture processing during recognition. *Percept. Psychophys.* **65**:725–734.
- J. M. Hillis, M. O. Ernst, M. S. Banks und M. S. Landy (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science* **298**:1627–1630.
- S. Hochstein und M. Ahissar (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron* **36**:791–804.
- D. H. Hubel und T. N. Wiesel (1959). Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.* **148**:574–591.
- D. H. Hubel und T. N. Wiesel (1962). Receptive fields, binocular interaction and functional architecture in the cat's striate cortex. *J. Physiol.* **160**:106–154.
- D. H. Hubel und T. N. Wiesel (1965). Receptive fields and fundamental architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* **28**:229–289.
- M. Ito, H. Tamura, I. Fujita und K. Tanaka (1995). Size and position invariance of neural responses in monkey inferotemporal cortex. *J. Neurophysiol.* **73**:218–226.

- M. Ito, I. F. A. H. Tamura und K. Tanaka (1994). Processing of contrast polarity of visual images in inferotemporal cortex of the macaque monkey. *Cereb. Cortex* **5**:499–508.
- E. B. Johnston (1991). Systematic distortions of shape from stereopsis. *Vis. Res.* **31**:1351–1360.
- E. R. Kandel, J. H. Schwartz und T. M. Jessell, Hg. (2000). *Principles of Neural Science*. McGraw-Hill, 4. Aufl.
- J. M. Kinser (1999). Foveation by a pulse-coupled neural network. *IEEE Transact. Neur. Netw.* **10**:621–625.
- K. Koffka (1935). *Principles of Gestalt Psychology*. Harcourt Brace, New York.
- W. Köhler (1920). Physical Gestalten. In W. D. Ellis, Hg., *A Source Book of Gestalt Psychology*, S. 17–54. Humanities Press, New York.
- E. Körner, M.-O. Gewaltig, U. Körner, A. Richter und T. Rodemann (1999). A model of computation in neocortical architecture. *Neur. Netw.* **12**:989–1005.
- E. Körner, H. Tsujino und T. Masutani (1997). A cortical-type modular neural network for hypothetical reasoning. *Neur. Netw.* **10**:791–814.
- R. Kupper und R. Eckhorn (2002a). A neural mechanism for viewing-distance-invariance. In R. P. Würtz und M. Lappe, Hg., *Dynamic Perception*, S. 277–282. Workshop of GI section 1.0.4 “Image Understanding” and the European Networks MUHCI and ECOVISION, infix. Ruhr-University Bochum, Germany.
- R. Kupper und R. Eckhorn (2002b). A neural network model generating invariance for visual distance. In J. R. Dorronsoro, Hg., *Artificial Neural Networks – ICANN 2002. International Conference. Madrid, Spain, August 2002. Proceedings*, Bd. 2415 von *Lecture Notes in Computer Science*, S. 93–98. Springer. ISSN 0302-9743. <http://link.springer-ny.com/link/service/series/0558/bibs/2415/24150093.htm>; <http://link.springer-ny.com/link/service/series/0558/papers/2415/24150093.pdf>.
- R. Kupper, M.-O. Gewaltig, U. Körner und E. Körner (2004). Spike-latency codes and the effect of saccades. *Neurocomp.* Accepted.

Literaturverzeichnis

- M. S. Landy, L. T. Maloney, E. B. Johnston und M. Young (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vis. Res.* **35**:389–412.
- D. A. Leopold und N. K. Logothetis (1998). Microsaccades differentially modulate neural activity in the striate and extrastriate visual cortex. *Exp. Brain Res.* **123**:341–345.
- S. Martinez-Conde, S. L. Macknik und D. H. Hubel (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nature Neurosci.* **3**:251–258. Erratum in: *Nature Neurosci.* **3**(4):409.
- S. Martinez-Conde, S. L. Macknik und D. H. Hubel (2002). The function of bursts of spikes during visual fixation in the awake primate lateral geniculate nucleus and primary visual cortex. *Proc. Natl. Acad. Sci. USA* **99**:13920–13925.
- D. McCready (1999). The moon illusion explained. Web site. Rev. Dez. 2002, Febr. 2003. <http://facstaff.uww.edu/mccreadd/index.html>.
- B. W. Mel und J. Fiser (2000). Minimizing binding errors using learned conjunctive features. *Neur. Comp.* **12**:247–278.
- W. Metzger (1953). *Gesetze des Sehens*. Verlag Waldemar Kramer, Frankfurt am Main.
- B. C. Motter und G. F. Poggio (1990). Dynamic stabilization of receptive fields of cortical neurons (V1) during fixation of gaze in the macaque. *Exp. Brain Res.* **83**:37–43.
- I. Murakami und P. Cavanagh (1998). A jitter after-effect reveals motion-based stabilization of vision. *Nature* **395**:798–801.
- I. Murakami und P. Cavanagh (2001). Visual jitter: Evidence for visual-motion-based compensation of retinal slip due to small eye movements. *Vis. Res.* **41**:173–186.
- A. Noë (2002). Is the visual world a grand illusion? *J. Consc. Stud.* **9**:1–12.
- B. A. Olshausen, C. H. Anderson und D. C. van Essen (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* **13**:4700–4719.

- T. L. Ooi, B. Wu und Z. J. He (2001). Distance determined by the angular declination below the horizon. *Nature* **414**:197–200.
- H. op de Beeck und R. Vogels (2000). Spatial sensitivity of macaque inferior temporal neurons. *Comp. Neurol.* **426**:505–518.
- J. K. O'Regan und A. Noë (2000). Experience is not something we feel but something we do: A principled way of explaining sensory phenomenology, with change blindness and other empirical consequences. Talk given at the ASSC conference: The Unity of Consciousness: Binding, Integration, and Dissociation, Brussels. <http://nivea.psych.univ-paris5.fr/ASSChtml/ASSC.html>.
- J. K. O'Regan und A. Noë (2001). A sensorimotor account of vision and visual consciousness. *Beh. Brain Sci.* **24**:939–1011.
- M. L. Pauly (2000). *Ein aufmerksamkeitsgestütztes, biologienahes Objekt-Erkennungs- und Verfolgungssystem mit impulsodierenden Neuronen*. Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.), Fachbereich Physik, Philipps-Universität Marburg. <http://archiv.ub.uni-marburg.de/diss/z2001/0062/>.
- D. I. Perrett und M. W. Oram (1993). Neurophysiology of shape processing. *Im. Vis. Comp.* **11**:317–333.
- D. I. Perrett, M. W. Oram, M. H. Harries, R. Bevan, J. K. Hietanen, P. J. Benson und T. S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp. Brain Res.* **86**:159–73.
- I. N. Pigarev, H.-C. Nothdurft und S. Kastner (2002). Neurons with radial receptive fields in monkey area V4A: Evidence of a subdivision of prelunate gyrus based on neuronal response properties. *Exp. Brain Res.* **145**:199–206.
- S. J. M. Rainville und F. A. A. Kingdom (2002). Scale invariance is driven by stimulus density. *Vis. Res.* **42**:351–367.
- H. J. Reitboeck und J. Altmann (1984). A model for size- and rotation-invariant pattern processing in the visual system. *Biol. Cybern.* **51**:113–121.
- M. Riesenhuber und T. Poggio (1999a). Are cortical models really bound by the “binding problem”? *Neuron* **24**:87–93.
- M. Riesenhuber und T. Poggio (1999b). Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**:1019–1025.

Literaturverzeichnis

- L. A. Riggs und F. Ratliff (1952). The effects of counteracting the normal movements of the eye. *J. Opt. Soc. Am.* **42**:872–873.
- R. W. Rodieck (1998). *The first steps in seeing*. Sinauer Associates, Inc., Sunderland, Massachusetts.
- E. T. Rolls und T. Milward (2000). A model of invariant object recognition in the visual system: Learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neur. Comp.* **12**:2547–2572.
- D. Rosenbluth und J. M. Allman (2002). The effect of gaze angle and fixation distance on the responses of neurons in V1, V2, and V4. *Neuron* **33**:143–149.
- E. Salinas und L. Abbott (1997). Invariant visual responses from attentional gain fields. *J. Neurophysiol.* **77**:3267–3272.
- E. Salinas und P. Thier (2000). Gain modulation: A major computational principle of the central nervous system. *Neuron* **27**:15–21.
- T. Schanze, M. Wilms, M. Eger, L. Hesse und R. Eckhorn (2002). Activation zones in cat visual cortex evoked by electrical retina stimulation. *Graefe's Arch. Clin. Exp. Ophthalmol.* **24**:947–954.
- A. Schlack, K.-P. Hoffmann und F. Bremmer (2003). Selectivity of macaque ventral intraparietal area (area VIP) for smooth pursuit eye movements. *J. Physiol.* **551**:551.
- R. Schmidt und G. Thews (1997). *Physiologie des Menschen*. Springer, Heidelberg, 27. Aufl.
- E. L. Schwartz (1981). Cortical anatomy, size invariance, and spatial frequency analysis. *Perception* **10**:455–468.
- S. G. Sklavos und A. K. Moschovakis (2002). Neural network simulations of the primate oculomotor system IV. A distributed bilateral stochastic model of the neural integrator of the vertical saccadic system. *Biol. Cybern.* **86**:97–109.
- K. Suder, F. Wörgötter und T. Wennekers (2000). Neural field model of receptive field restructuring in primary visual cortex. *Neur. Comp.* **13**:139–159.
- N. V. Swindale (2000). How many maps are there in visual cortex? *Cereb. Cortex* **10**:633–643.

- W. R. Taylor, S. He, W. R. Levick und D. I. Vaney (2000). Dendritic computation of direction selectivity by retinal ganglion cells. *Science* **289**:2347–2350.
- W. R. Taylor und D. I. Vaney (2002). Diverse synaptic mechanisms generate direction selectivity in the rabbit retina. *J. Neurosci.* **22**:7712–7720.
- A. Thiele, P. Henning, M. Kubischik und K.-P. Hoffmann (2002). Neural mechanisms of saccadic suppression. *Science* **295**:2460–2462.
- S. Thorpe, A. Delorme und R. van Rullen (2001). Spike-based strategies for rapid processing. *Neur. Netw.* **14**:715–725.
- S. Ullman, M. Vidal-Naquet und E. Sali (2002). Visual features of intermediate complexity and their use in classification. *Nature Neurosci.* **5**:682–687.
- W. J. M. van Damme und E. Brenner (1997). The distance used for scaling disparities is the same as the one used for scaling retinal size. *Vis. Res.* **37**:757–764.
- R. van Rullen und S. J. Thorpe (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neur. Comp.* **13**:1255–1283.
- D. I. Vaney und W. R. Taylor (2002). Direction selectivity in the retina. *Curr. Opin. Neurobiol.* **12**:405–410.
- C. von der Malsburg (1999). The what and why of binding: The modeler's perspective. *Neuron* **24**:95–104. <http://cogprints.ecs.soton.ac.uk/archive/00001488/>; <http://cogprints.ecs.soton.ac.uk/archive/00001488/00/cvdm.pdf>; <http://cogprints.ecs.soton.ac.uk/archive/00001488/01/cvdm.ps>.
- H. von Helmholtz (1910). *Handbuch der physiologischen Optik*, Bd. 3. Voss, Hamburg, Leipzig, 3. Aufl.
- E. von Holst und H. Mittelstädt (1950). Das Reafferenzprinzip. Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften* **37**:464–476.
- G. Wallis (1996). Using spatio-temporal correlations to learn invariant object recognition. *Neur. Netw.* **9**:1513–1519.

Literaturverzeichnis

- G. Wallis und E. T. Rolls (1997). Invariant face and object recognition in the visual system. *Progr. Neurobiol.* **51**:167–194.
- N. M. Weinberger (1995). Dynamic regulation of receptive fields and maps in the adult sensory cortex. *Annu. Rev. Neurosci.* **18**:129–158.
- L. Wiskott und T. J. Sejnowski (2002). Slow feature analysis: Unsupervised learning of invariances. *Neur. Comp.* **14**:715–70.
- C. Wissel (1989). *Theoretische Ökologie: Eine Einführung*. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong. ISBN 3-540-50848-1.
- F. Wörgötter, K. Suder, Y. Zhao, N. Kerscher, U. T. Eysel und K. Funke (1998). State-dependent receptive-field restructuring in the visual cortex. *Nature* **396**:165–168.
- R. H. Wurtz und M. A. Sommer (2004). Identifying corollary discharges for movement in the primate brain. In C. A. Heywood, A. D. Milner und C. Blakemore, Hg., *The Roots of Visual Awareness*, Bd. 144 von *Progress in Brain Research*, Kap. 3, S. 47–60. Elsevier. DOI: 10.1016/S0079-6123(03)14400-3.
- A. J. Yu, M. A. Giese und T. A. Poggio (2002). Biophysically plausible implementations of the maximum operation. *Biol. Cybern.* **14**:2857–2881.

Danksagung

Ich danke Professor Eckhorn für die Betreuung meiner Arbeit und für die langjährige erfolgreiche Zusammenarbeit in der Arbeitsgruppe Neurophysik. Professor Körner danke ich für die Möglichkeit, meine Promotion in der anregenden wissenschaftlichen Umgebung des Honda Research Institute zuendezuführen.

Herzlich danken möchte ich meinen geschätzten Freunden und Arbeitskollegen in der Marburger Arbeitsgruppe Neurophysik. Meine Zeit dort wird mir als eine der besten meines Studiums in Erinnerung bleiben. Meine jetzigen Kollegen arbeiten tatkräftig daran, die Zeit in Offenbach zu einer ähnlich guten Erfahrung zu machen.

Hans-Jörg Brinksmeyer danke ich für die Bereitstellung seiner Meßdaten und viele fruchtbare Diskussionen bei der Entwicklung des Schlupfkorrekturmodells.

Für die Unterstützung und ihre beherzten Anmerkungen als Korrekturleser danke ich Alex Gail (der sich meine ewige Hochachtung mit der Durchsicht einer haarsträubenden Rohfassung meiner Dissertation verdient hat), sowie Andreas Thiel, Martina Hasenjäger und Reinhard Eckhorn.

Meinen Mitstreitern auf dem Gebiet der IDL-Programmierung gebührt außerordentlicher Dank für die gemeinsame Entwicklung und Wartung einer prächtigen Sammlung von IDL-Routinen. Mirko Saam, Andi Gabriel und Andreas Thiel, sowie Andreas Bruns waren die NASEn der ersten und zweiten Stunde.

Nicht unerwähnt bleiben soll auch, daß Mareike Overhof mit ihrer Erwähnung des Scheinriesen einen wunderschönen Aufhänger für meine Dissertation lieferte.

Schließlich danke ich meinen Eltern ganz besonders herzlich für ihre kontinuierliche Unterstützung während meines ganzen Studiums.

Wissenschaftlicher Werdegang

Persönliche Daten

Name	Rüdiger Kupper
Geburtsdatum	24. Juni 1972
Geburtsort	Weilburg a.d. Lahn
Familienstand	ledig
Nationalität	deutsch

Schulbildung

1978–1982	Besuch der Grundschule Aumenau
1982–1991	Besuch des Gymnasiums Philippinum in Weilburg a.d. Lahn Abschluß: Allgemeine Hochschulreife, Gesamtnote 1,1

Hochschulbildung

10/1992–11/1998	Studium der Physik an der Philipps-Universität Marburg Abschluß: Diplom, Gesamtnote sehr gut Titel der Arbeit: <i>Selbstorganisation in lokalen Populationen gekoppelter Neuronen</i>
12/1998–11/2004	Promotion in der AG Neurophysik an der Philipps-Universität Marburg

Wissenschaftliche Tätigkeit

seit 01/2003	Senior Scientist am Honda Research Institute Europe, Offenbach Grundlagenforschung und konzeptionelle Modellierung zur kortikalen neuronalen Verarbeitung beim Menschen
--------------	--

