Invariante Objekterkennung im visuellen Cortex und die Rolle von Aufmerksamkeit: Simulationen mit dem HMAX-Modell

Diplomarbeit in Biologie

Durchgeführt am
Center for Biological and Computational Learning
Massachusetts Institute of Technology
Cambridge, Massachusetts
USA

Robert Schneider

August 2002

Abstract

Thema der vorliegenden Arbeit ist die Realisierung von visueller Aufmerksamkeit, die auf bestimmte Objekte unabhängig von deren Position gerichtet ist, im Rahmen des von Riesenhuber und Poggio vorgestellten HMAX-Modells der Objekterkennung im visuellen Cortex [55]. Dabei interessiert vor allem, ob Modelle der Aufmerksamkeitseffekte auf neuronaler Ebene, die in physiologischen Experimenten zur objektorientierten Aufmerksamkeit beobachtet wurden [8, 46, 52], in HMAX die aus der Psychophysik bekannte, durch Aufmerksamkeit bedingte Verbesserung der Leistung bei der Objekterkennung erklären können [23, 51]. Als Grundlage dieser Simulationen dient eine systematische Untersuchung zu den Invarianzeigenschaften des Modells gegenüber Translation, Skalierung, Kontraständerung und Hintergrundstimuli für verschiedene Objektklassen. Dabei zeigt sich, daß die Invarianzeigenschaften eines hierarchischen neuronalen Modells wie HMAX für verschiedene Objektkategorien unterschiedlich sein können, in Abhängigkeit davon, wie gut die vom Modell detektierten visuellen Merkmale geeignet sind, die betrachteten Objekte zu beschreiben. Da als Modell für störenden visuellen Hintergrund in den vorliegenden Simulationen ein zweiter Stimulus im visuellen Feld eingesetzt wird, werden außerdem die Interaktionen zweier Stimuli auf neuronaler Ebene in HMAX untersucht und mit den Vorhersagen des Biased Competition Model von Desimone *et al.* [53] sowie experimentellen Daten für dieselben visuellen Stimulationen verglichen. Es erweist sich, daß eventuelle kompetitive Interaktionen von Neuronen im visuellen Cortex sehr wahrscheinlich nicht alleine durch die inhibitorischen Bereiche der rezeptiven Felder der Afferenzen dieser Neuronen erklärt werden können, sondern explizit realisiert sein müßten. Schließlich werden verschiedene experimentell motivierte Mechanismen von Modulationen neuronaler Aktivitäten durch Aufmerksamkeit auf ihre Wirksamkeit für die Objekterkennung untersucht. Dabei werden auch unterschiedliche Methoden, Stimuli in HMAX zu codieren, und die Erkennung individueller Objekte ebenso wie ihre Kategorisierung betrachtet. Die Resultate der Simulationen lassen den Einsatz von Aufmerksamkeit bei der Objekterkennung im Modell als wenig gewinnbringend erscheinen; vielmehr erweisen sich die Effekte modellierter Aufmerksamkeit als wenig spezifisch oder bringen andererseits ein hohes Risiko falsch positiver Resultate mit sich. Zusammen mit der experimentellen Literatur zum Thema legen diese Ergebnisse nahe, daß Aufmerksamkeit in der Objekterkennung selbst keine Rolle spielt, sondern vielmehr einen Mechanismus zur Selektion bereits erkannter Stimuli darstellt.

Danksagung

Mein ganz herzlicher Dank gilt zunächst meinem Betreuer am Center for Biological and Computational Learning, Maximilian Riesenhuber, der es mir ermöglicht hat, ans MIT zu kommen und dort eine spannende Diplomarbeit anzufertigen und der mir stets mit Rat, Tat und inspirierenden Diskussionen zur Seite stand, in neurowissenschaftlichen Fragen ebenso wie in der Theorie und Praxis des Tae Kwon Do. Besonders bedanken möchte ich mich auch bei meinem Betreuer an der Universität München, Prof. Gerhard Neuweiler, ohne dessen unbürokratische und wie selbstverständliche Übernahme der Betreuung diese Arbeit nicht hätte entstehen können, abgesehen davon, daß ich ohne sein Beispiel und Vorbild womöglich gar nicht in der Neurowissenschaft gelandet wäre. Dank gebührt natürlich auch und ganz besonders dem Leiter des CBCL, Prof. Tomaso Poggio, für das freundschaftliche, offene und multidisziplinäre Klima an diesem Institut, an dem ich dankenswerterweise eine Zeitlang teilhaben durfte.

Auf keinen Fall fehlen darf der wärmste Dank an meine Eltern für jede nur denkbare Unterstützung – und natürlich an Julia, für ihre Nähe trotz der großen räumlichen Distanz zwischen uns.

Inhaltsverzeichnis

Kapitel 1

Einführung

Objekterkennung im visuellen System ist ein hochkomplexes Problem, dessen Lösung dem Gehirn[1] erstaunlich leicht zu fallen scheint. Ein Objekt kann als ein und dasselbe wahrgenommen werden, selbst wenn es in verschiedenen Größen, an unterschiedlichen Positionen, in der Bildebene oder auch im Raum rotiert, in verschiedenen Belichtungszuständen sowie vor einem potentiell störenden Hintergrund präsentiert wird. Der Versuch, diese Leistungen am Computer zu modellieren, macht deutlich, wie schwierig es tatsächlich ist, einen derart hohen Grad an Invarianz gegenüber Störungen und zugleich genügend Spezifität zur Unterscheidung individueller Stimuli zu erreichen. Eines der Prinzipien, die das Gehirn zu diesem Zweck verwendet, ist die hierarchische Verschaltung von Neuronen über mehrere Verarbeitungsstufen hinweg, um zunehmend komplexe Stimuluspräferenzen zu generieren – ein Prinzip, das zuerst von Hubel und Wiesel in ihrem klassischen Artikel über einfache und komplexe Zellen im visuellen Cortex der Katze postuliert wurde [22] und das sich auch in anderen Teilen des Cortex wiederzufinden scheint, etwa in den höheren Arealen des ventralen Teils des visuellen Cortex, der allgemein als für die Objekterkennung zuständig angesehen wird [71]. So finden sich bei Makaken im inferotemporalen Cortex (IT) Neuronen, die sehr selektiv auf bestimmte zweidimensionale Ansichten von Gesichtern reagieren und dabei in ihrer Antwortstärke in hohem Maße Invarianz gegenüber Änderungen der Position oder Größe ihrer bevorzugten Stimuli an den

[1] Die Mehrzahl der Erkenntnisse, auf die sich diese Arbeit stützt, wurden an Menschen und Makaken gewonnen, so daß im engeren Sinne deren Gehirne gemeint sind, wenn hier vom „Gehirn“ die Rede ist. Damit soll freilich nicht behauptet werden, daß etwa Objekterkennung oder Aufmerksamkeit nur Leistungen des Gehirns von Primaten wären.

Tag legen [4]. IT gilt als das höchste noch rein mit der Verarbeitung visueller Information befaßte Areal im ventralen visuellen System.

Während die Selektivitäts- und Invarianzeigenschaften von Neuronen gemäß Hubel und Wiesel durch „bottom-up"-Prozesse entstehen, also durch die Verschaltung von Neuronen mit ihren Afferenzen, spielen bei der Wahrnehmung von Stimuli im Gehirn auch „top-down"-Prozesse eine wichtige Rolle, also Signale, die entgegen dem Fluß der sensorischen Information von höheren Arealen ausgehend modulierend auf näher an der Peripherie gelegene Areale einwirken. Visuelle Aufmerksamkeit bedient sich solcher Prozesse, um aus der enormen Menge an Reizen, die in jedem Moment auf das visuelle System einwirken, die für das Verhalten wichtigen Informationen herauszufiltern. Ohne solche Selektionsmechanismen wäre das Gehirn kaum in der Lage, seine Ressourcen effizient einzusetzen und situativ angemessen auf Umwelteinflüsse zu reagieren.

Bereits seit einigen Jahren beschäftigt sich eine zunehmende Anzahl von Studien mit den Effekten von Aufmerksamkeit auf die Wahrnehmung visueller Stimuli. Der Einsatz bewußter Aufmerksamkeit durch eine Versuchsperson oder ein Versuchstier läßt sich dabei im Experiment dadurch kontrollieren, daß mit der Präsentation visueller Stimuli eine Aufgabenstellung verbunden wird, die nur dann mit über dem Zufallsniveau liegender Leistung bewältigt werden kann, wenn die Stimuli bewußt beachtet werden – etwa eine Aufgabe, bei der ein „Zielobjekt" unter mehreren anderen Objekten („Distraktoren") gefunden werden oder ein Urteil über die relative Orientierung zweier Balkenstimuli abgegeben werden soll. Unterschiede etwa in einer Verhaltensreaktion oder in der neuronalen Aktivität, die sich ergeben, wenn ein und derselbe visuelle Stimulus einmal für die zu lösende Aufgabe relevant, einmal irrelevant ist (also einmal Zielobjekt ist und beachtet wird, das andere Mal nicht), können dann als Effekt von Aufmerksamkeit interpretiert werden.

Für das Problem der Objekterkennung relevante Experimente, die gezielt den Einsatz von Aufmerksamkeit steuern, wurden bereits in den 70er Jahren in der Psychologie und Psychophysik durchgeführt. Dabei zeigte sich, daß Versuchspersonen in einer raschen, ununterbrochenen Folge visueller Stimuli (*rapid serial visual presentation*, RSVP) ein bestimmtes Objekt sehr viel besser erkennen bzw. über sein eventuelles Vorhandensein in einer solchen Folge besser urteilen konnten, wenn sie im voraus über dieses Zielobjekt informiert

wurden [51]. Die Trefferquote in einem solchen RSVP-Experiment ist dabei um so höher, je genauer die Vorabinformation über das zu detektierende Objekt ist (wenn also etwa nicht nur bekannt ist, daß „ein Tier“ oder „ein Transportmittel“ erkannt werden soll, sondern „ein Hund“ oder „ein Auto“) [23]. Eine Möglichkeit, diese Ergebnisse zu interpretieren, ist es, einen Einfluß von Aufmerksamkeit auf die Prozesse der Objekterkennung anzunehmen, so daß etwa visuelle Reize, die zum erwarteten bzw. bewußt beachteten Zielobjekt gehören, vom visuellen System bevorzugt selektiert und weiterverarbeitet werden.

Auf der anderen Seite wurden vor allem im vergangenen Jahrzehnt mehr und mehr Untersuchungen zu physiologischen Effekten von Aufmerksamkeit durchgeführt, unter Verwendung von Einzelzellableitungen oder zunehmend auch bildgebenden Verfahren wie fMRI (Überblicke über solche Studien finden sich etwa in [27, 68]). Dabei zeigt sich, daß Aufmerksamkeit die Aktivität von Neuronen meßbar beeinflussen kann. Neuronen feuern stärker, wenn sich in ihrem rezeptivem Feld ein Stimulus befindet, der gerade im Fokus der Aufmerksamkeit ist, als wenn derselbe Stimulus gerade nicht beachtet wird [41, 46]. Weiterhin bestimmt ein bewußt beachteter Stimulus die Antwort eines Neurons auch in Gegenwart anderer Stimuli. Wenn sich zwei Stimuli im rezeptiven Feld eines Neurons befinden, von denen der eine normalerweise eine stärkere Antwort des Neurons hervorruft als der andere, und wird die Aufmerksamkeit auf den von diesem Neuron weniger bevorzugten Stimulus gerichtet, so wird es mit einer Feuerrate antworten, die etwa so hoch ist wie für diesen Stimulus alleine, obwohl der andere, „bessere“ Stimulus weiterhin innerhalb des rezeptiven Felds verbleibt [7, 8, 53]. Derartige Effekte wurden vor allem für die höheren Areale V2, V4 und IT des ventralen visuellen Systems beschrieben, sind aber auch im primären visuellen Cortex V1 [58] sowie im dorsalen Teil des visuellen Systems gefunden worden, der u.a. für die Bewegungserkennung zuständig ist [69]. Aufmerksamkeit kann dabei auf eine bestimmte Position im visuellen Feld gerichtet sein (*spatial attention*, lokale Aufmerksamkeit) oder ortsunabhängig auf ein bestimmtes Objekt (*feature* oder *object attention*, objektorientierte Aufmerksamkeit), das als Zielobjekt eines Verhaltensexperiments vorgegeben wurde. Diese beiden Formen von Aufmerksamkeit unterscheiden sich auch auf der neuronalen Ebene [20]. Objektorientierte Aufmerksamkeit ist dabei diejenige Form, die auch im RSVP-Experiment eingesetzt wird.

Soweit deuten die experimentellen Befunde sowohl aus der Psychophysik wie auch aus

der Physiologie in der Tat darauf hin, daß Aufmerksamkeit Objekte im Gesichtsfeld für die Wahrnehmung selektieren und aus ihrer Umgebung hervorheben kann. Bisher wurde jedoch der Zusammenhang zwischen den Beobachtungen auf Verhaltensebene und auf neuronaler Ebene noch nicht systematisch untersucht. Genauer gesagt wäre es von Interesse, ob Aufmerksamkeit einen Einfluß auf die Objekterkennung hat, ob die erhöhte Leistung bei der Erkennung von Objekten im RSVP-Experiment durch die beobachteten aufmerksamkeitsinduzierten Veränderungen in der neuronalen Aktivität erklärt werden kann und ob objektorientierte Aufmerksamkeit eventuell sogar notwendig ist, um Objekte unter schwierigen Bedingungen noch erkennen zu können – also etwa bei sehr kurzen Expositionszeiten oder in einer natürlichen Umgebung mit vielen Distraktorstimuli.

Ziel dieser Arbeit ist es, diese Fragen mit Hilfe von Modellierung der neuronalen Verarbeitung im visuellen System zu beantworten. Während freilich bis heute kein Modell einen so großen Teil des Gehirns wie das ventrale visuelle System in seiner ganzen Komplexität simulieren kann, können Modelle, die die für die jeweilige Fragestellung relevanten Eigenschaften des untersuchten Systems hinreichend gut abbilden, sehr gut dazu dienen, aus Experimenten abgeleitete Hypothesen zu testen, Rahmenbedingungen für Theorien zu formulieren sowie Impulse für weitere Experimente zu liefern. Einige Computermodelle visueller Verarbeitung mit Aufmerksamkeitseffekten wurden bereits getestet (etwa [72, 73]), üblicherweise aber mit dem Ziel, die beobachteten physiologischen Prozesse nachzubilden, weniger im Hinblick auf das Problem der Objekterkennung. Andererseits wurde das HMAX-Modell der Objekterkennung im visuellen Cortex von Riesenhuber und Poggio [55] genau zu dem Zweck entwickelt, zu einem besseren Verständnis der Leistungen des visuellen Cortex bei der Erkennung von Objekten beizutragen. Basierend auf einer Erweiterung des hierarchischen Modells von Hubel und Wiesel simuliert es explizit die Eigenschaft von Neuronen im Areal IT von Makaken, auf bestimmte zweidimensionale Ansichten komplexer Stimuli unabhängig von deren Größe und Position im visuellen Feld zu antworten [24, 35]. Dies wird durch eine hierarchische Verschaltung von Modellneuronen in aufeinanderfolgenden Schichten erreicht, bei der die Stimuluspräferenzen von Neuronen zunehmend komplex und ihre rezeptiven Felder immer größer werden. Dabei werden abwechselnd zwei verschiedene Mechanismen verwendet, mit denen die Modellneuronen auf Aktivität ihrer Afferenzen antworten: lineare Summierung über die Afferenzen, um Spezifität für zunehmend komplexe Stimuli zu erzeugen, und eine nichtlineare

Maximums-Operation („MAX“), bei der ein Neuron seine eigene Feuerrate entsprechend der Aktivität des am stärksten feuernden afferenten Neurons einstellt. Dieser Mechanismus ist entscheidend für die Invarianzeigenschaften des Modells bezüglich der Größe und Position von Stimuli und zugleich das Kernpostulat von HMAX. Ferner ist das Modell im Aufbau bewußt einfach gehalten, um möglichst nicht nur eine sehr begrenzte Menge experimenteller Daten sehr genau zu modellieren, sondern auf verschiedene experimentelle Paradigmen anwendbar zu sein. Vor allem soll es quantitative Vergleiche zwischen Experiment und Theorie ermöglichen und dadurch Hypothesen liefern können, die experimentell überprüfbar sind.

Es hat sich herausgestellt, daß das HMAX-Modell in der Tat auch einige andere Eigenschaften der Informationsverarbeitung im ventralen visuellen Cortex modellieren kann, für die es nicht explizit konstruiert wurde, etwa Leistungen bei der Kategorisierung von Objekten [15, 29] oder Objekterkennung in der Gegenwart irrelevanter Hintergrundstimuli [44, 54]. Die Verarbeitung visueller Information in HMAX ist bisher allerdings rein „feed-forward“, d.h. verläuft von „einfacheren“ zu „komplexeren“ Zellen ohne Einfluß von „top-down“-Effekten wie Aufmerksamkeit. In der vorliegenden Arbeit wird nun untersucht, inwieweit simulierte Effekte von Aufmerksamkeit in HMAX die Leistungen des Modells bei der Objekterkennung beeinflussen können, mit dem Ziel, daraus Schlußfolgerungen auf die Rolle von Aufmerksamkeit im realen visuellen System zu ziehen. Unter Aufmerksamkeit versteht sich im Rahmen dieser Untersuchung stets objektorientierte Aufmerksamkeit, also diejenige Form, die Versuchspersonen einsetzen, wenn ihnen ein zu detektierendes Zielobjekt bekannt ist, nicht aber, wo im visuellen Feld es auftauchen wird. Lokale Aufmerksamkeit kann mit dem HMAX-Modell sehr intuitiv nachgebildet werden, indem zunächst auf der Basis von Vorabinformationen oder aufgrund besonders auffälliger visueller Reize eine Position im visuellen Feld gewählt wird, die dann bevorzugt oder ausschließlich in die Verarbeitung durch das Modell eingeht. Es wurden bereits entsprechende Simulationen durchgeführt, bei denen die bevorzugte Position mit Hilfe eines Modells der von Itti und Koch postulierten „saliency map“ (eine Art topographischer Repräsentation von Reizen im visuellen Feld, die etwa aufgrund von Kontrast, Farbe oder Helligkeit besonders hervorstechen) gewählt wurde [76]. Es ist jedoch viel weniger klar, wie sich das visuelle System auf die bevorzugte Wahrnehmung bestimmter Stimuli einstellen könnte, von denen womöglich nicht nur ihre Position, sondern auch ihr genaues

Aussehen nicht bekannt sind, und wie eine solche Form von Aufmerksamkeit die Objekterkennung in Gegenwart von Distraktorstimuli beeinflussen kann. Diese Frage wird im Zentrum dieser Untersuchung stehen.

Für die Simulationen zu Fragestellungen der Aufmerksamkeit wurden als Stimuli sowohl die „Büroklammern“ aus der HMAX-Originalpublikation bzw. den Untersuchungen von Logothetis *et al.* [35, 55] als auch computergenerierte Abbildungen von Autos und Gesichtern als Vertreter realistischerer Objektklassen verwendet, die als Zielobjekt oder Distraktor an verschiedenen Positionen innerhalb des Ausgangsbildes präsentiert wurden. Eine genaue Untersuchung der Invarianzeigenschaften von HMAX für verschiedene Stimulusklassen, Positionen und Größen von Stimuli wird daher, nach einer Einführung in den Aufbau des Modells sowie allgemein verwendete Methoden, an erster Stelle dieser Arbeit stehen. Variationen in der Antwort der Modellneuronen und in der Erkennungsleistung bei der Translation und Skalierung von Stimuli sowie die Abhängigkeit dieser Phänomene von der Stimulusklasse werden diskutiert. Im Fall der Translation wird eine quantitative Beschreibung entwickelt. Als weitere grundlegende Eigenschaften von HMAX ohne Verwendung von Aufmerksamkeitseffekten werden außerdem die Veränderungen der neuronalen Antworten bei unterschiedlichem Stimuluskontrast sowie bei Hinzufügen eines Distraktorstimulus beschrieben und mit experimentellen Resultaten verglichen.

Zur vereinfachten Simulation von störendem visuellen Hintergrund wird in den hier vorgestellten Simulationen grundsätzlich ein zweiter Stimulus, in der Regel aus derselben Objektklasse, im visuellen Feld verwendet. Für der Interaktion zweier solcher Stimuli auf neuronaler Ebene sowie für den Einfluß von Aufmerksamkeit auf diese Interaktion ist in der Literatur ein einflußreiches und häufig experimentell untersuchtes Modell beschrieben, das „Biased Competition Model“ („beeinflußter Wettbewerb“) von Desimone und Mitarbeitern [12, 53]. HMAX, unterstützt von neueren experimentellen Daten [16], und das Biased Competition Model machen zum Teil unterschiedliche Vorhersagen über neuronale Antworten in Gegenwart von mehr als einem Stimulus. In einem eigenen Kapitel werden daher die beiden unterschiedlichen Ansätze vorgestellt sowie physiologische Experimente zur Interaktion zweier Stimuli in HMAX simuliert. Die Resultate werden mit den experimentellen Daten verglichen und daraufhin untersucht, ob und inwieweit die beiden Modelle kompatibel sein könnten.

Auf den Erkenntnissen der vorhergehenden Kapitel aufbauend werden schließlich in Kapitel 5 zunächst die von experimentellen Resultaten vorgegebenen Rahmenbedingungen für Modelle von Aufmerksamkeit definiert. Sodann werden verschiedene Implementationen neuronaler Effekte von Aufmerksamkeit in HMAX und deren Auswirkungen auf die Leistungen des Modells bei der Objekterkennung untersucht. Die Ergebnisse dieser Simulationen sollen dann dazu dienen, die Rolle der Aufmerksamkeit bei der Objekterkennung im visuellen System so weit als möglich einzugrenzen.

Die nur in einzelnen Kapiteln verwendeten Methoden werden jeweils dort in eigenen Methodenabschnitten vorgestellt. Ebenso werden die erhaltenen Resultate in jedem Kapitel zusammengefaßt und diskutiert. Ein Ausblick in Kapitel 6 schließt die Arbeit ab und versucht eine zusammenfassende Deutung der Ergebnisse der vorangehenden Kapitel.

Kapitel 2

Methoden

In diesem Kapitel sollen die in der gesamten Arbeit verwendeten Methoden vorgestellt werden, also natürlich das HMAX-Modell selbst, aber auch Stimulusklassen, Methoden zur Messung der Erkennungsleistung, ROC-Kurven oder die verwendete Kontrastskala. Für die einzelnen Kapitel spezifische Methoden werden jeweils dort in eigenen Abschnitten diskutiert.

2.1 Das HMAX-Modell

Das HMAX-Modell der Objekterkennung im visuellen Cortex von Primaten [55] (siehe Abbildung 2-1) verwendet als Input Graustufenbilder in Form zweidimensionaler Felder, die in dieser Arbeit durchweg 128×128 oder 160×160 Pixel groß sind. Diese werden auf der ersten Verarbeitungsstufe einer Faltung mit verschiedenen Filterkernen unterzogen, d.h. an jeder Pixelposition des Bildes werden Filter zentriert und die Produkte von Bild-Pixelwert und Filter-Funktionswert an allen Punkten innerhalb des „rezeptiven Feldes“ des Filters aufsummiert. Diese S1-Filter sind zweidimensionale Gauß-Funktionen. In einer Dimension wird dabei die zweite Ableitung einer Gaußschen Funktion verwendet, um ein rezeptives Feld ähnlich dem einer einfachen Zelle im primären visuellen Cortex zu erhalten, also eine excitatorische Zone etwa in Form eines orientierten Balkens mit einer inhibitorischen Umgebung (siehe Abbildung 2-2). Die Filter werden in den Orientierungen $0°$, $45°$, $90°$und $135°$und Größen von 7×7 bis 29×29 in 2-Pixel-Schritten verwendet.

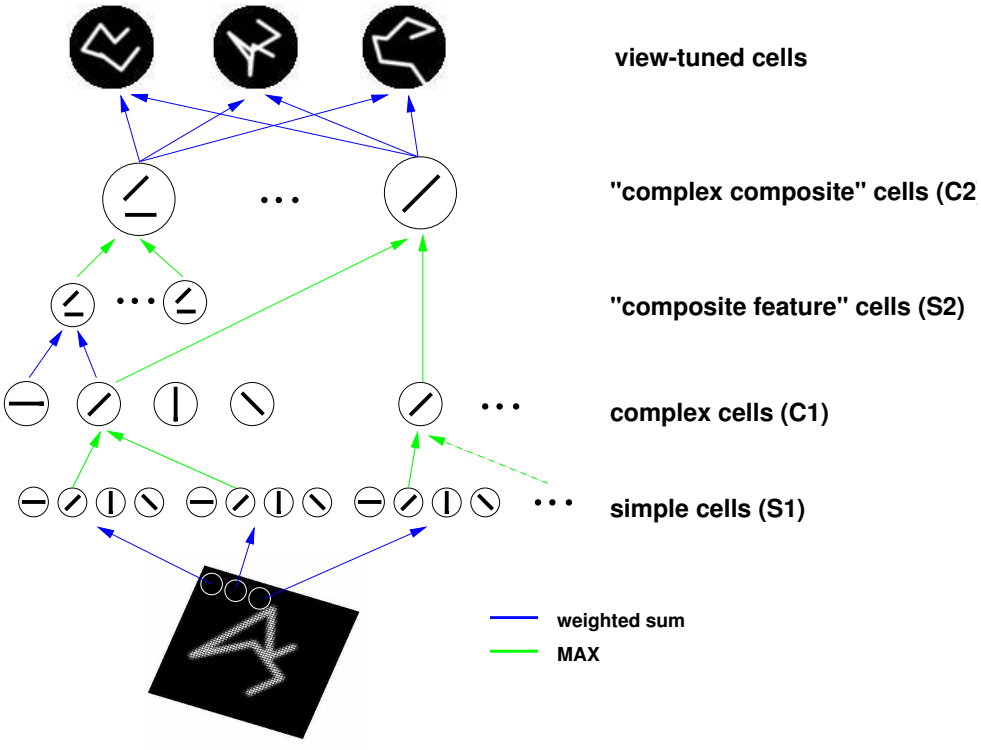

Abbildung 2-1: Schemadarstellung des HMAX-Modells.

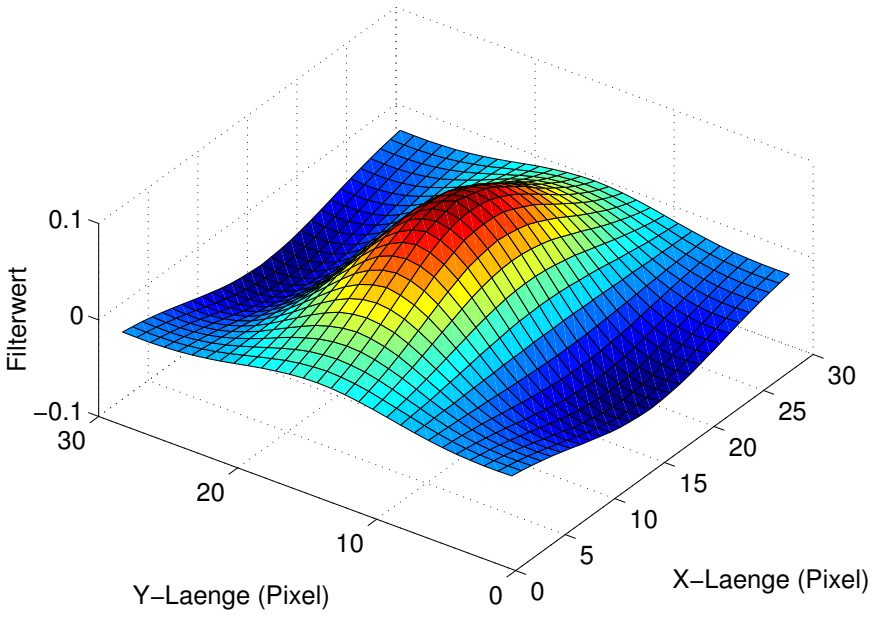

Abbildung 2-2: Rezeptives Feld einer S1-Zelle, 29×29 Pixel.

An jeder Position des Bildes werden Filter aller Größen und Orientierungen zentriert. Die Summe der Filterwerte ist jeweils auf 0 und die Summe ihrer Quadrate stets auf 1 normiert, und das Ergebnis einer Faltung eines Bildausschnitts mit einem Filter wird durch die Summe der Quadrate der Pixelwerte dieses Bildausschnitts geteilt. Dies führt zu S1-Aktivitäten zwischen -1 und 1 (reelle Werte). Dieser Verarbeitungsschritt simuliert die Detektion einfacher Bestandteile des visuellen Inputs, also etwa orientierte Kanten und Kontrastverläufe, im primären visuellen Cortex V1.

Im nächsten Schritt werden die Aktivitäten von S1-Zellen zusammengefaßt und dienen als afferenter Input für C1-Zellen, Modelle der komplexen Zellen in V1. Dazu werden zunächst vier sogenannte „Filterbänder“ definiert, in denen alle S1-Filter verschiedener Größen zusammengefaßt werden, die auf denselben Typ von C1-Zellen projizieren (Filterband 1: 7×7 und 9×9 Pixel große Filter, Filterband 2: von 11×11 bis 15×15 Pixel, Filterband 3: von 17×17 bis 21×21 Pixel; Filterband 4: von 23×23 bis 29×29 Pixel). Für jedes Filterband wird eine Konstante (*„poolRange“*) festgelegt, die bestimmt, wieviele benachbarte S1-Zellen auf eine C1-Zelle projizieren. *poolRange* hat für die Filterbänder 1 bis 4 jeweils die Werte 4, 6, 9 bzw. 12. Ein Wert von 4 für Filterband 1 bedeutet dabei, daß ein Feld von 4×4 benachbarten S1-Zellen der Größe 7×7 Pixel und zugleich ein Feld von ebenfalls 4×4 S1-Zellen der Größe 9×9 Pixel auf ein und dieselbe C1-Zelle projizieren. (Vier benachbarte S1-Zellen sind dabei solche, die auf vier aneinandergrenzenden Pixeln zentriert sind.) Nur S1-Zellen derselben Orientierung projizieren auf eine gegebene C1-Zelle. Das heißt, in den C1-Zellen ist bereits ein größerer Invarianzbereich gegenüber Translation und Skalierung von Stimuli realisiert, weil in ihnen die Aktivitäten von S1-Zellen verschiedener Größen und an verschiedenen Positionen vereinigt wird. Die rezeptiven Felder benachbarter C1-Zellen überlappen außerdem in der Regel um einen Betrag, der im Parameter *„c1Overlap“* festgelegt wird. Ein Wert von 2 (der HMAX-Standard) bedeutet dabei, daß in jeder Richtung die Hälfte der S1-Zellen, die auf eine C1-Zelle projizieren, zugleich auch auf die in dieser Richtung benachbarte C1-Zelle projizieren. (Für eine Illustration dieses Arrangements siehe auch Abbildung 3-3). Zusätzlich modellieren C1-Zellen die Invarianz komplexer Zellen gegenüber Kontrastumkehr, indem sie den Absolutbetrag ihres Inputs verrechnen.

Die C1-Zellen verarbeiten ihre S1-Inputs mit einer Maximums- (*„MAX“*-) Operation, d.h.

die Feuerrate einer C1-Zelle entspricht der Aktivität ihrer am stärksten aktiven S1-Afferenz. Die MAX-Operation entspricht effektiv der sehr viel rechenzeitaufwendigeren „sliding template window"-Operation, die in Computer-Sehsystemen oft verwendet wird: Beide detektieren den Bestandteil eines Bildes, der am besten einer gegebenen Vorlage („template") entspricht. Im Kontext des HMAX-Modells ermöglicht die MAX-Operation das Vergrößern der rezeptiven Felder von Neuronen, ohne dabei ihre Spezifität für bestimmte Stimuli aufgeben zu müssen. Sie verhindert, daß etwa aufgrund einer linearen Summierung von Afferenzen mehrere von einem Neuron nicht bevorzugte Stimuli dieselbe Antwortstärke hervorrufen wie ein einzelner Stimulus, der der Präferenz des Neurons entspricht. Experimentelle Resultate deuten darauf hin, daß in der Tat komplexe Zellen in V1, aber auch Subgruppen von Neuronen in V4 ihre Afferenzen mit einer MAX-Operation verrechnen [16, 33]. Es wurden bereits einige neuronale Mechanismen beschrieben, die *in vivo* die MAX-Operation realisieren könnten [77].

In jedem Filterband projizieren jeweils 2×2 quadratisch angeordnete C1-Zellen auf eine S2-Zelle. Hierbei werden alle Kombinationen von C1-Zellen verschiedener Orientierungspräferenz realisiert, so daß entsprechend den $4^4 = 256$ möglichen Kombinationen von vier C1-Zellen mit jeweils 4 möglichen Orientierungen 256 Typen von S2-Zellen entstehen. Die S2-Zellen sprechen auf ihre vier C1-Eingangswerte gemäß einer vierdimensionalen Gaußfunktion mit Mittelwert $(1|1|1|1)$ und Standardabweichung 1 an, d.h. sie feuern mit einer maximalen Rate von 1, wenn alle ihre C1-Afferenzen ebenfalls ihre maximale Aktivität 1 erreichen. S2-Zellen repräsentieren den Vorrat an komplexen Merkmalen, die HMAX detektieren kann, in diesem Fall also alle möglichen Kombinationen von vier quadratisch angeordneten Balken mit je vier möglichen Orientierungen.

Um Größeninvarianz über alle Filterskalen und Ortsinvarianz über das gesamte verarbeitete Bild zu erreichen, werden schließlich S2-Zellen, wiederum mittels der MAX-Operation, zu C2-Zellen, den Zellen der Ausgabeschicht des HMAX-Kernmodells, zusammengefaßt. Für jeden der 256 Typen von S2-Zellen gibt es genau eine C2-Zelle, auf die alle S2-Zellen dieses Typs, an allen Positionen und in allen Größen, projizieren. Folglich feuert jede der 256 C2-Zellen genauso stark wie die am stärksten aktive S2-Zelle mit demselben bevorzugten Muster aus vier Balken, unabhängig davon, wo sich diese S2-Zelle befindet oder wie groß das von ihr bevorzugte Muster ist. C2-Zellen sind als Modelle von Neuronen im

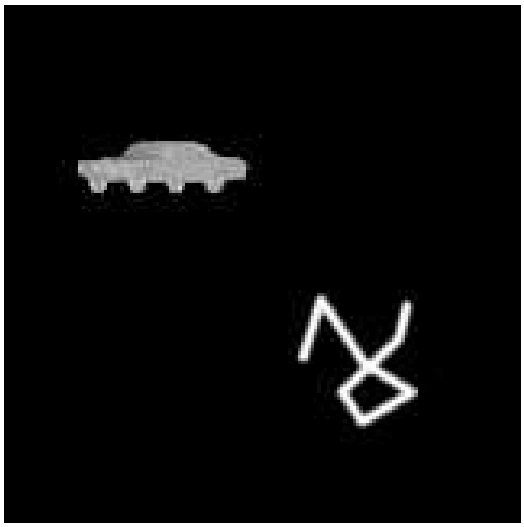

Abbildung 2-3: Beispiele für die verwendeten Auto- und Büroklammer-Stimuli.

visuellen Areal V4 gedacht.

C2-Zellen projizieren wiederum auf die sogenannten *„view-tuned units“* (VTUs), die ihren Namen tragen, weil sie bevorzugt auf eine bestimmte zweidimensionale Ansicht eines komplexen dreidimensionalen Stimulus antworten, in Übereinstimmung mit den von Logothetis *et al.* gefundenen Neuronen im inferotemporalen Cortex [35]. VTUs sind bisher die einzige Stufe, auf der Lernen in HMAX stattfindet. VTUs werden auf einen Stimulus „trainiert“, indem die Aktivitätswerte der 256 C2-Zellen bei Präsentation dieses Stimulus als Mittelwert der 256-dimensionalen Gaußfunktion verwendet werden, mit der VTUs auf ihre C2-Afferenzen ansprechen. Das führt dazu, daß eine VTU genau dann ihre maximale Feuerrate von 1 erreicht, wenn die Aktivität über die C2-Zellen exakt derjenigen bei der Präsentation des Stimulus entspricht, auf den diese VTU trainiert wurde. Um größere Invarianz der VTU-Antwort in Gegenwart von Hintergrund zu erreichen, können auch nur diejenigen C2-Zellen als Afferenzen für eine VTU gewählt werden, die am stärksten auf diesen Stimulus antworten [54]. In dieser Arbeit werden üblicherweise 40, 100 oder alle 256 C2-Zellen als Afferenzen für eine VTU verwendet. Ein weiterer Parameter, der das Antwortverhalten einer VTU bestimmt, ist die Standardabweichung ihrer Gaußschen Antwortfunktion (der σ-Wert). Ein kleinerer Wert für σ führt zu einer spezifischeren Abstimmung der VTU auf ihren bevorzugten Stimulus, da ihre Feuerrate dann schon für kleinere Abweichungen des C2-Aktivitätsmusters vom Optimum stärker abfallen wird.

2.2 Stimuli

Autos. Als Stimuli für das Modell wurden zum einen Autos aus dem in [56] beschriebenen „8-Auto-System“ verwendet. Diese Stimuli wurden mit Hilfe einer von Christian Shelton entwickelten multidimensionalen 3D-Morphing-Software generiert [64]. Das System besteht aus 8 Abbildungen von Autos, den „Prototypen“, und Mischformen („Morphs“) aus je zwei dieser Prototypen, die schrittweise veränderte Merkmalsanteile der beiden Prototypen aufweisen. Jeder Prototyp kann in 10 Schritten durch stufenweise Veränderung seiner äußeren Merkmale in jeden anderen „umgewandelt“ werden. D.h., insgesamt 28 „Morphlinien“, die jeweils in 10 Intervalle unterteilt sind, verbinden je zwei Prototypen, was zu einer Gesamtzahl von 260 Autostimuli führt. Dieses Arrangement induziert eine Ähnlichkeitsmetrik: Je zwei Prototypen sind 10 „Morphschritte“ voneinander entfernt, und ein Autostimulus, der 3 Morphschritte von einem gegebenen Stimulus entfernt ist, ist diesem ähnlicher als einer, der beispielsweise 7 Morphschritte entfernt wäre. Mit diesem System können also Einflüsse kontrollierter Veränderungen der äußeren Form auf die Erkennungsleistung untersucht werden. Alle Autos wurden dabei in dieser Arbeit aus demselben Blickwinkel präsentiert (225°, von vorne links).

Büroklammern. Weiterhin wurden dieselben 200 „Büroklammer“-Stimuli verwendet wie in den Experimenten von Logothetis *et al.* [35] sowie in der HMAX-Originalpublikation [55]. Üblicherweise wurden für eine Simulation 15 Büroklammern als Zielobjekte und 60 als Distraktoren gewählt. Diese wurden ebenfalls jeweils nur aus einem Blickwinkel präsentiert. Ähnlich den Autos könnten auch die Büroklammern parametrisiert und mit der erwähnten Morphing-Software ineinander übergeführt werden. Allerdings können sich nur wenig veränderte Büroklammern äußerlich stark voneinander unterscheiden, etwa wenn durch Änderung eines Winkels zwischen zwei Drahtsegmenten eine neue Überkreuzung entsteht. Deshalb wurden die Büroklammern hier lediglich als individuelle Stimuli ohne parametrische Formveränderungen eingesetzt. Beispiele für die beiden Stimulusklassen sind in Abbildung 2-3 gezeigt.

Kontrast. Die verwendeten Stimuli wurden entweder vor einem schwarzen Hintergrund (Pixelwert 0) präsentiert oder mit einem definierten Kontrastwert. Das hier verwendete

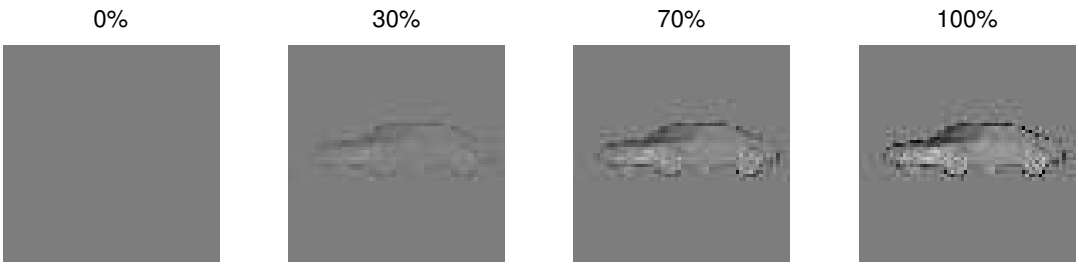

Abbildung 2-4: Auto-Prototyp bei verschiedenen Kontrasten. Kontrastwerte sind über den einzelnen Abbildungen angegeben.

Kontrastmaß ist 0, wenn alle Pixel im Bild den Wert des Bildhintergrunds haben, und erreicht sein Maximum 1 bzw. 100%, wenn die maximale Abweichung eines Pixelwertes vom Hintergrundwert genauso groß ist wie dieser Hintergrundwert, also etwa wenn der geringste im Bild erreichte Pixelwert gleich Null ist. Der Hintergrund-Pixelwert war in Simulationen mit variiertem Kontrast grundsätzlich auf 128 gesetzt. Eine Illustration der so definierten Kontrastniveaus ist in Abbildung 2-4 gezeigt.

2.3 Messung der Erkennungsleistung

Aktivste VTU-Paradigma. Es wurden zwei verschiedene Methoden verwendet, um die Leistung des Modells bei der Objekterkennung zu messen, die zugleich zwei verschiedenen Aufgaben im Verhaltensexperiment entsprechen. Zum einen ist dies das sogenannte *„Aktivste VTU-Paradigma"*: Bei der Präsentation eines Stimulus gilt derjenige Stimulus aus einer Menge als erkannt, dessen VTU (d.h. die VTU, die auf diesen Stimulus trainiert wurde) von allen VTUs, die auf Exemplare dieser Menge Stimuli trainiert wurden, am stärksten feuert. Üblicherweise wurden auf jedes Zielobjekt (d.h. die 8 Auto-Prototypen oder die 15 ausgewählten Büroklammern) eine VTU trainiert, so daß das Objekt, dessen VTU unter den – je nach präsentierter Stimulusklasse – 8 bzw. 15 VTUs am stärksten aktiv ist, als erkannt gilt. 100% Erkennungsleistung wird in dieser Maßeinheit erreicht, wenn für jede Präsentation eines Zielobjekts aus einer Stimulusklasse jeweils die auf dieses Objekt trainierte VTU die am stärksten aktive ist. Zufallsniveau wird andererseits erreicht, wenn dies nur für ein Zielobjekt der Fall ist, denn auch die Wahrscheinlichkeit, daß rein zufällig eine VTU stärker als alle anderen feuert, ist 1 dividiert durch die Anzahl von VTUs (bzw. Zielobjekten, also 12.5% für Autos, 6.7% für Büroklammern). Dieses Paradigma entspricht

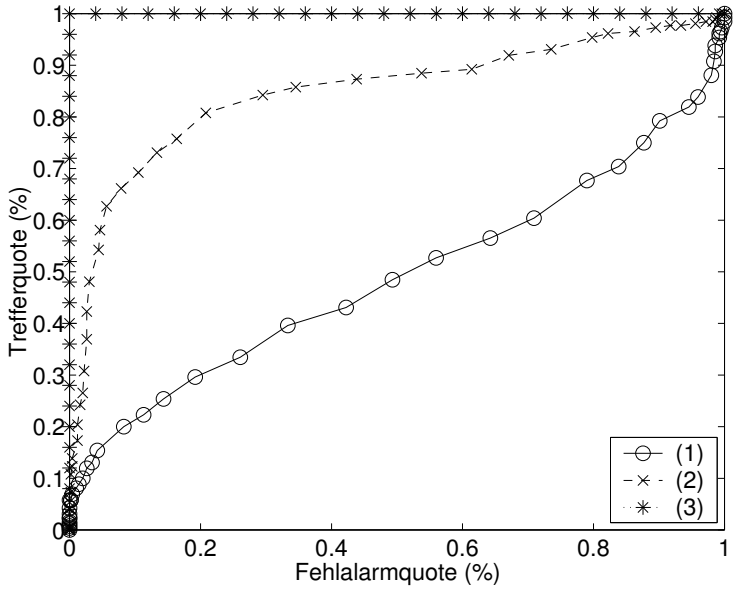

Abbildung 2-5: Beispiele für ROC-Kurven. (1) Zufallsniveau, (2) mittlere Leistung, (3) perfekte Leistung.

einem Verhaltensexperiment, in dem ein Versuchstier oder eine Versuchsperson eine bestimmte Anzahl Stimuli lernt und dann beurteilen muss, welcher dieser Stimuli in einem Bild gezeigt wird – in der Regel in einer anderen Konfiguration, etwa in veränderter Größe oder vor einem störenden Hintergrund.

Stimulusvergleich-Paradigma. Als zweite Möglichkeit der Quantifizierung der Erkennungsleistung wurde das *„Stimulusvergleich-Paradigma“* verwendet. Dabei wird die Aktivität einer VTU bei der Präsentation ihres bevorzugten Stimulus mit ihrer Aktivität bei der Präsentation eines anderen Stimulus verglichen. Feuert sie im ersten Fall stärker, wurde ihr bevorzugter Stimulus erkannt. Eine VTU erreicht 100% Erkennungsleistung, wenn sie mit höherer Aktivität auf ihren bevorzugten Stimulus reagiert als auf alle anderen präsentierten Stimuli. Die Leistung des Modells insgesamt ergibt sich daraus durch Mittelung über die Leistungen der einzelnen VTUs. Das Zufallsniveau entspricht hier einer Leistung von 50%, wenn also die VTUs im Mittel mit gleicher Wahrscheinlichkeit auf einen beliebigen Stimulus stärker oder schwächer antworten als auf ihr trainiertes Zielobjekt. Dieses Paradigma modelliert ein „two alternative forced choice“-Experiment, in dem entschieden werden muß, welcher von zwei als Alternativen präsentierten Stimuli mit einem zuvor gezeigten Musterstimulus identisch ist.

ROC-Analyse. Neben der Angabe, wie oft eine VTU stärker auf ihren bevorzugten Stimulus antwortet als auf einen Distraktor, kann die Erkennungsleistung im Stimulusvergleich-Paradigma auch in Form einer sogenannten ROC-Kurve (*receiver operating characteristic*) [38] dargestellt werden. ROCs messen die Fähigkeit eines Signaldetektionssystems (hier der VTUs), zwischen verschiedenen Signalen (hier den visuellen Stimuli) verläßlich zu unterscheiden, und zwar unabhängig davon, welchen Schwellenwert ein Beobachter des Systems verwenden könnte, um aufgrund der Antwort des Systems zu entscheiden, welches Signal detektiert wurde. Im Kontext des HMAX-Modells sind die Signale, zwischen denen eine VTU unterscheiden können muß, einerseits ihr trainierter Stimulus und andererseits alle anderen Stimuli. Zur Generierung der ROC wird das Intervall zwischen der maximalen und der minimalen Antwort einer VTU in eine beliebig wählbare Anzahl Teilintervalle aufgeteilt. Für jeden Wert an der Grenze zweier Teilintervalle („Schwellenwert“) wird berechnet, wie häufig die Aktivität der VTU bei Präsentation eines Zielobjekts („Trefferquote“) und wie häufig sie bei Präsentation eines Distraktors („Fehlalarmquote“) über diesem Wert liegt. Beide Werte können im Prinzip unabhängig voneinander variieren. Sie werden nicht auf die Anzahl aller betrachteten Stimuli, sondern jeweils auf die Anzahl der Präsentationen eines Zielobjekts bzw. Distraktors bezogen. Die ROC entsteht, indem für jeden Schwellenwert die Trefferquote gegen die Fehlalarmquote aufgetragen wird. Sie enthält stets die Randpunkte $(0\%|0\%)$ und $(100\%|100\%)$ für Schwellenwerte über der maximalen bzw. unter der minimalen auftretenden Aktivität der betrachteten VTU (siehe Abbildung 2-5). In dieser Arbeit gezeigte ROCs sind stets über alle betrachteten VTUs gemittelt. Für perfekte Erkennungsleistung (d.h. alle VTUs feuern stets stärker, wenn ihr jeweiliger trainierter Stimulus gezeigt wird, als sonst) enthält die ROC den Punkt $(0\%|100\%)$, d.h. es gibt mindestens einen Schwellenwert, für den die VTU perfekt zwischen den Präsentationen der Zielobjekte und denen der Distraktoren unterscheidet. Erreicht die Leistung einer VTU nur Zufallsniveau, verläuft ihre ROC als Gerade mit der Steigung 1 zwischen den Punkten $(0\%|0\%)$ und $(100\%|100\%)$, d.h. für jeden Schwellenwert rufen genausoviele Präsentationen des Zielobjekts dieser VTU eine überschwellige Antwort hervor wie Präsentationen von Distraktoren. Je größer also die Fläche unter der ROC-Kurve, desto besser die Leistung der VTU bei der Unterscheidung von Objekten.

Kapitel 3

HMAX ohne Aufmerksamkeit: Einflüsse von Translation, Skalierung, Kontrast und Hintergrundstimuli

Das HMAX-Modell wurde entwickelt, um die über einen weiten Variationsbereich gegenüber Translation und Skalierung invarianten Antworten von Neuronen im inferotemporalen Cortex von Primaten zu modellieren [24]. Es wurden bisher jedoch keine systematischen Untersuchungen über die Abhängigkeit der Leistung von HMAX von den gewählten Werten für die Parameter des Modells oder von der verwendeten Stimulusklasse durchgeführt. Die Experimente zum Einfluß von Aufmerksamkeitseffekten auf die Objekterkennung in dieser Arbeit verwenden Stimuli verschiedener Objektklassen an unterschiedlichen Positionen, einzeln und in Kombination, sowie bei wechselnden Kontrasten. Es war daher erforderlich, als Grundlage für diese Simulationen die Invarianzeigenschaften von HMAX gegenüber Translation, Skalierung, Kontraständerung und Hinzufügen von Hintergrundstimuli für Exemplare der verwendeten Objektklassen zu untersuchen. Die Resultate dieser Analyse im Hinblick auf die darauf folgenden Experimente sowie mögliche Implikationen für die Verarbeitung visueller Information im Gehirn sind Thema dieses Kapitels.

3.1 Methoden

3.1.1 Simulationen

Änderungen der Stimulusposition

Um die Antwort von HMAX-Modellneuronen bei Veränderung der Position eines Stimulus im visuellen Feld zu bestimmen, wurden zunächst VTUs auf jeden der acht Auto-Prototypen sowie auf 15 Büroklammern an einer Position auf der horizontalen Mittellinie des Bildes in der linken Bildhälfte trainiert. Die Stimuli selbst waren 64×64 Pixel groß, das Gesamtbild 160×160 Pixel. Alle Hintergrundpixel hatten den Wert 0. Es wurden dann die Antworten von C2-Zellen und VTUs für alle 8 Auto-Prototypen und 15 Büroklammern sowie für die 252 verbleibenden Autostimuli und 60 weitere Büroklammern als Distraktoren an 60 weiteren Positionen entlang der horizontalen Mittellinie berechnet, im Abstand von jeweils 1 Pixel. Zuvor wurde sichergestellt, daß Verschiebungen von Stimuli entlang einer vertikalen Bildachse qualitativ dieselben Ergebnisse liefern, wie es die für beide Achsen identische Organisation von HMAX nicht anders erwarten läßt und wie auch Abbildung 3-1 a zeigt.

Änderungen der Stimulusgröße

Die Invarianz gegenüber Skalentransformationen wurde untersucht, indem VTUs auf die 8 Auto-Prototypen sowie auf 15 Büroklammern in der Größe 64×64 Pixel in der Bildmitte trainiert wurden. HMAX-Antworten für diese Stimuli sowie alle übrigen 252 Autos und 60 weitere Büroklammern wurden dann für Stimulusgrößen von 16×16 Pixel bis 160×160 Pixel in Halboktavschritten (d.h. 16, 24, 32, 48, 96, 128 und 160 Pixel Seitenlänge), jeweils zentriert in einem Bild von 160×160 Pixel Größe, berechnet. Wiederum waren alle Hintergrundpixel auf den Wert 0 gesetzt.

Ferner wurden für die acht Auto-Prototypen und acht zufällig ausgewählte Büroklammern die C2-Aktivitäten berechnet, wenn nur jeweils eines der vier Filterbänder aktiv war und daher die Aktivität aller C2-Zellen notwendig von der Aktivität der afferenten Zellen in diesem Filterband bestimmt war. Das Filterband, bei dessen Verwendung die Akti-

vität einer C2-Zelle denselben Wert hatte wie bei Verwendung aller vier Filterbänder und Präsentation desselben Stimulus, war dann die „Filterbandquelle“ der Aktivität dieser C2-Zelle für diesen Stimulus und zeigte, in welcher Größe das von dieser C2-Zelle bevorzugte Merkmal im Originalbild am deutlichsten vorhanden war. Für die acht Auto-Prototypen wurden die Aktivitäten der C2-Zellen bei Einsatz nur jeweils eines Filterbandes auch für zwei verschiedene Stimulusgrößen (64×64 und 128×128 Pixel) berechnet.

Kontraständerungen

Zur Untersuchung des Verhaltens von HMAX-Zellen bei Veränderungen im Stimuluskontrast wurden die 260 Autostimuli sowie 75 Büroklammern einzeln bei Kontrasten von 0 bis 100% in Schritten von 10% gemäß der in Kapitel 2.2 definierten Kontrastskala präsentiert und die Antworten von C2-Zellen sowie VTUs, die auf die 8 Auto-Prototypen und auf 15 ausgewählte Büroklammern bei 100% Kontrast trainiert waren, berechnet.

Hinzufügen von Distraktorstimuli

Als Modell für störenden visuellen Hintergrund wurden Zielobjekte, auf die VTUs trainiert worden waren (8 Auto-Prototypen, 15 Büroklammern, jeweils bei 100% Kontrast), zusammen mit einem zweiten Objekt präsentiert. Die beiden Stimuli der Größe 64×64 Pixel befanden sich stets an den Positionen $(50|50)$ und $(110|110)$ innerhalb eines Bildes der Größe 160×160 Pixel, wobei die Position $(1|1)$ die obere linke Ecke des Bildes bezeichnet, d.h. die Stimuli befanden sich auf der Diagonalen von links oben nach rechts unten, symmetrisch zum Bildmittelpunkt (wie in Abbildung 2-3; allerdings wurden keine Präsentationen mit Objekten verschiedener Klassen verwendet). Die Positionen waren so gewählt, daß die Antworten aller Zellen auf denselben Stimulus an beiden Positionen identisch waren und das Zielobjekt an derselben Position erschien wie während des VTU-Trainings, um den Einfluß unterschiedlicher Positionen auf das Resultat auszuschließen (siehe 3.2.1). Ferner wurde durch den gewählten Abstand der beiden Stimuli eine Interaktion auf Ebene der S1-Zellen ausgeschlossen. Der Bildhintergrund war stets auf den Pixelwert 128 gesetzt. Die Kontraste sowohl von Zielobjekt als auch Distraktor konnten unabhängig voneinander variieren. Jeder von 8 Auto-Prototypen wurde als Zielobjekt zusammen mit allen Autosti-

muli präsentiert, die auf von dem jeweiligen Prototypen ausgehenden Morphlinien lagen, inklusive der jeweils sieben anderen Prototypen. Die 15 Büroklammern, für die VTUs trainiert worden waren, wurden jeweils zusammen mit 60 anderen Distraktor-Büroklammern präsentiert.

3.1.2 Auswertung

Änderungen von Stimulusposition und Größe

Zur Auswertung der Simulationen mit Stimuli an verschiedenen Positionen bzw. in unterschiedlichen Größen wurden einerseits die Aktivitätswerte der VTUs und die gemittelte Aktivität aller C2-Zellen betrachtet, andererseits auch die Erkennungsleistung gemäß der in Kapitel 2 definierten Paradigmen. 100% Erkennungsleistung im Aktivste VTU-Paradigma wurden für eine Stimulusgröße oder eine Position erreicht, wenn jede VTU der betrachteten Stimulusklasse jeweils bei Präsentation ihres bevorzugten Stimulus in der betreffenden Größe bzw. an der entsprechenden Position stärker als alle jeweils anderen VTU aktiv war. Im Stimulusvergleich-Paradigma mußte eine VTU bei Präsentation aller Distraktoren (alle 259 anderen Autos, falls die VTU auf ein Auto trainiert war, andernfalls alle 60 Distraktor-Büroklammern) einer Größe bzw. an einer Position mit geringerer Feuerrate antworten als bei Präsentation ihres trainierten Stimulus, um für diese Stimulusgröße bzw. -position 100% Leistung zu erreichen (Abb. 3-1, 3-2, 3-7, 3-8).

Im Fall der Autos konnten außerdem noch die Distraktoren nach ihrer Morph-Distanz zum Zielobjekt der betrachteten VTU gruppiert werden. Die Erkennungsleistung im Stimulusvergleich-Paradigma wurde dann, für jede Stimulusposition bzw. -größe, getrennt für alle Morph-Abstände berechnet. 100% Leistung für eine bestimmte Morph-Distanz zwischen Zielobjekt und Distraktor bedeutet in diesem Fall, daß eine VTU bei Präsentation ihres bevorzugten Stimulus stets stärker feuert als bei Präsentation aller von diesem Zielobjekt eine entsprechende Anzahl von Morphing-Schritten entfernten Autostimuli. Auf diese Weise konnte die Abhängigkeit der Erkennungsleistung von der Ähnlichkeit der verwendeten Distraktoren mit dem Zielobjekt untersucht werden, zusätzlich zu ihrer Abhängigkeit von Translation oder Skalierung des Stimulus (Abb. 3-1 d, 3-8 d).

ROCs wurden für diese Simulationen nicht erstellt, weil mit nur je einem Wert für die

Erkennungsleistung pro Größe und Position des Stimulus die Abhängigkeit der Leistung von diesen Faktoren besser dargestellt werden konnte.

Die Aktivitäten einzelner Filterbänder für zwei verschiedene Skalierungsfaktoren bei den acht Auto-Prototypen wurden verglichen, indem pro Prototyp die euklidischen Abstände zwischen den mit nur einem Filterband berechneten C2-Aktivitätsvektoren für die beiden verwendeten Stimulusgrößen 64×64 und 128×128 Pixel bestimmt wurden. Gemittelt über alle Prototypen zeigt ein geringer euklidischer Abstand zweier solcher Aktivitätsvektoren ähnliche Aktivität in zwei Filterbändern für die zwei verschiedenen Stimulusgrößen an und ermöglicht Schlüsse auf das Verhalten der von HMAX detektierten Merkmale bei Veränderung der Größe eines Stimulus (Abb. 3-9 a). Aus den Daten über die Filterbandquelle der Aktivität der C2-Zellen für Autos und Büroklammern einer Größe wurde berechnet, welcher Anteil der 256 C2-Zellen, gemittelt über alle acht Vertreter einer Stimulusklasse, seine Aktivität von welchem Filterband ableitete. Dies erlaubte dann Rückschlüsse darauf, welche Merkmale der beiden Stimulusklassen für die Erkennung in HMAX am bedeutendsten waren (Abb. 3-9 b).

Kontraständerungen

Aus den Simulationen mit Autos und Büroklammern bei verschiedenen Kontrasten wurden ebenfalls sowohl C2- und VTU-Aktivitäten betrachtet als auch die Leistungen des Modells bei der Erkennung von 8 Auto-Prototypen sowie 15 Büroklammern in Aktivste VTU- und Stimulusvergleich-Paradigma bestimmt. Letzteres wurde dabei in Form von ROC-Kurven ausgewertet. Als Distraktoren wurden für eine Auto-VTU alle Prototypen, auf die sie nicht trainiert wurde, sowie alle Morph-Stimuli zwischen diesen und ihrem bevorzugten Prototypen verwendet, d.h. für 100% Erkennungsleistung mußte eine VTU bei Präsentation ihres trainierten Stimulus stärker feuern als bei allen diesen Distraktorstimuli. Distraktoren für eine Büroklammer-VTU waren wie üblich 60 zufällig ausgewählte Stimuli, auf die keine VTU trainiert worden war.

Hinzufügen von Distraktorstimuli

Die Simulationen mit zwei Stimuli im Bild wurden im Hinblick auf die Erkennungsleistung des Modells in den beiden bekannten Paradigmen ausgewertet, gemittelt über alle Zielobjekte bzw. VTUs. Im Aktivste VTU-Paradigma galt ein Zielobjekt in einer Präsentation als erkannt, wenn die auf dieses Objekt trainierte VTU trotz Anwesenheit des Distraktors noch die höchste Aktivität unter allen VTUs, die auf Stimuli dieser Objektklasse trainiert waren, an den Tag legte. Damit lag das Zufallsniveau wiederum für Autos bei 12.5%, für Büroklammern bei 6.7%. Im Stimulusvergleich-Paradigma wurde die VTU-Antwort auf jede Präsentation ihres Zielobjekts paarweise mit allen oder ausgewählten Antworten auf Abbildungen, die das Zielobjekt nicht enthielten, verglichen; für 100% Erkennungsleistung mußte die VTU auf jede Präsentation ihres bevorzugten Stimulus mit höherer Aktivität antworten als auf alle anderen Präsentationen. Für Büroklammern wurden dabei die 60 Präsentationen des trainierten Objekts einer VTU (d.h. ihre bevorzugte Büroklammer zusammen mit jedem der 60 Distraktoren) mit allen übrigen Präsentationen verglichen. Bei Autos wurden nur solche Präsentationen miteinander verglichen, bei denen der Distraktor dieselbe Anzahl von Morphschritten vom Zielobjekt entfernt war, um den Einfluß ähnlicherer oder verschiedenerer Distraktoren auf die Erkennungsleistung untersuchen zu können. Perfekte Erkennung eines Auto-Prototyps bedeutete in diesem Fall also, daß die auf dieses Auto trainierte VTU in allen 7 Präsentationen dieses Autos (je eine mit einem zweiten Auto in einem bestimmten Morphabstand) stärker feuerte als für eine beliebige Präsentation eines der 7 anderen Prototypen mit einem zweiten Auto im selben Morphabstand.

Die Resultate im Stimulusvergleich-Paradigma werden als ROC-Kurven dargestellt. Eine ROC-Kurve kann im Falle der Autos, entsprechend der Ausführungen im vorigen Absatz, nur für einen bestimmten Morphabstand zwischen Zielobjekt und Distraktor erstellt werden. Der jeweils verwendete Morphabstand wird in der Bildunterschrift angegeben.

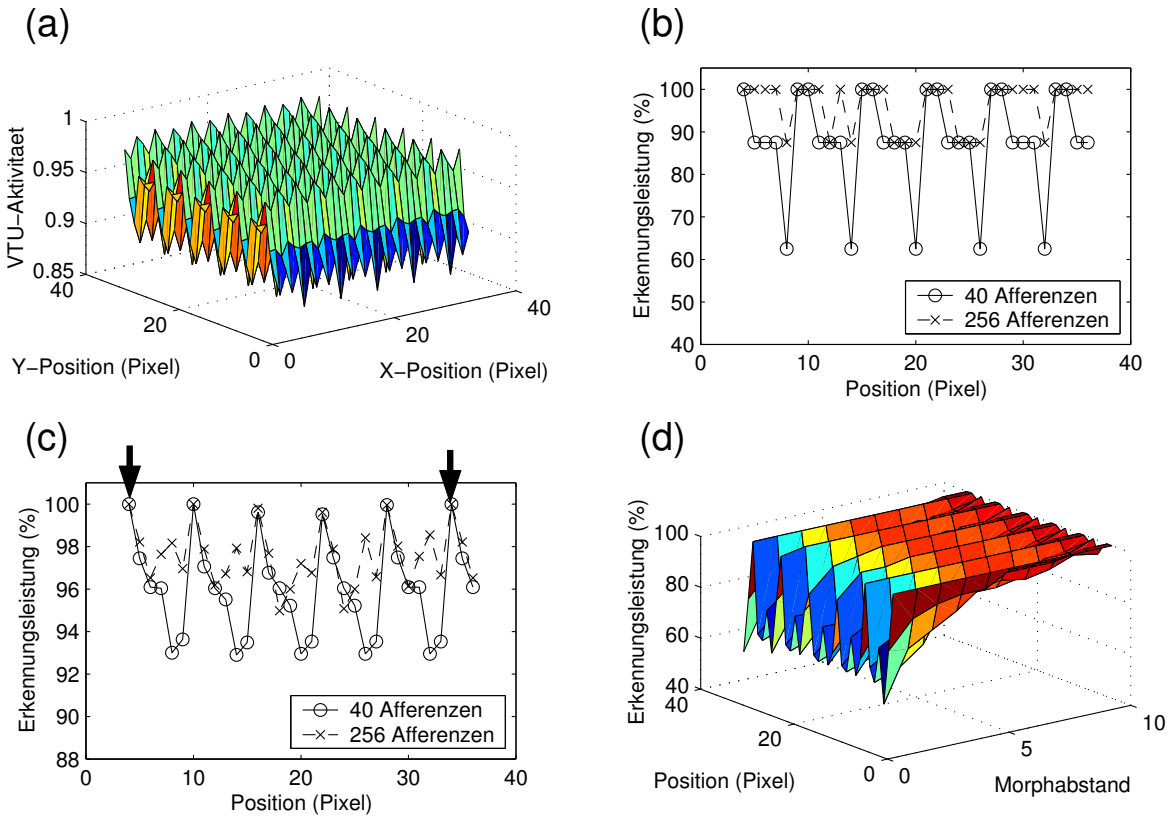

Abbildung 3-1: Effekte der Stimulusposition auf VTUs, die auf Autos trainiert sind. **(a)** Antwort einer VTU (40 Afferenzen, $\sigma = 0.2$) auf ihren bevorzugten Stimulus an verschiedenen Positionen in der Bildebene. **(b)** Mittlere Erkennungsleistung (Aktivste VTU-Paradigma) von 8 Auto-VTUs ($\sigma = 0.2$) für verschiedene Positionen ihrer bevorzugten Stimuli sowie unterschiedlich viele C2-Afferenzen. Zufallsniveau: 12.5%. **(c)** Wie (b), aber für das Stimulusvergleich-Paradigma. Zufallsniveau: 50%. In beiden Paradigmen sind die geringeren Schwankungen bei Verwendung von mehr Afferenzen deutlich sichtbar. Die Periodizität der Schwankungen (30 Pixel) ist durch Pfeile markiert. Die scheinbar kleinere Periodenlänge in (b) geht auf die weniger fein abgestufte Maßeinheit im Aktivste VTU-Paradigma zurück (bei 8 Zielobjekten sind nur Vielfache von $\frac{1}{8} = 12.5\%$ möglich). **(d)** Mittlere Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-VTUs mit 40 Afferenzen und $\sigma = 0.2$, aufgetragen gegen Stimulusposition und Distraktor-Ähnlichkeit („Morphabstand“, siehe Kapitel 2).

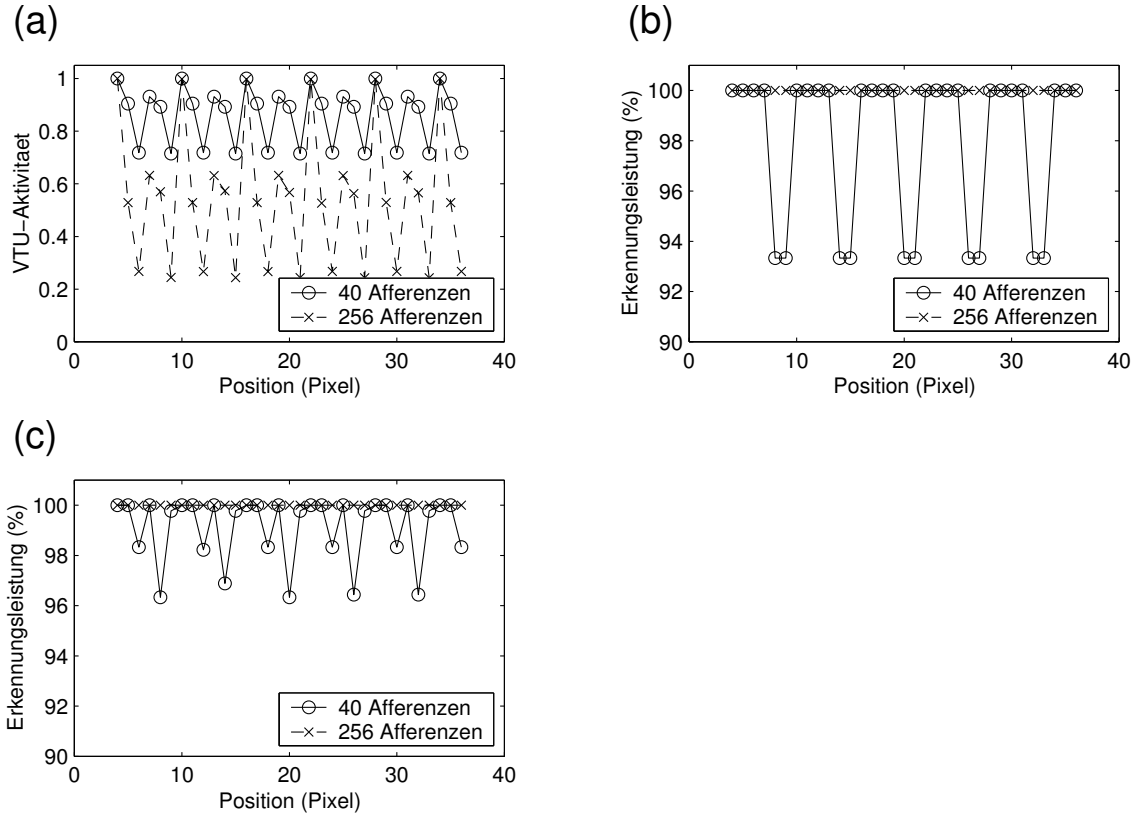

Abbildung 3-2: Effekte der Stimulusposition auf VTUs, die auf Büroklammern trainiert sind. **(a)** Antworten von VTUs mit 40 oder 256 Afferenzen und $\sigma = 0.2$ auf ihre trainierten Stimuli an verschiedenen Positionen. Im Gegensatz zur Erkennungsleistung sind die Schwankungen der VTU-Aktivität für mehr Afferenzen größer. **(b)** Mittlere Erkennungsleistung von 15 auf Büroklammern trainierten VTUs ($\sigma = 0.2$) im Aktivste VTU-Paradigma, in Abhängigkeit von der Position ihrer bevorzugten Stimuli, für unterschiedlich viele C2-Afferenzen. Das Zufallsniveau liegt bei 6.7%. **(c)** Wie (b), aber für das Stimulusvergleich-Paradigma. Zufallsniveau: 50%.

3.2 Ergebnisse

3.2.1 Änderungen der Stimulusposition

Es zeigt sich, daß die Antwort der HMAX-Modellneuronen von der Position eines Stimulus abhängt. Wird ein Stimulus an einer anderen Position gezeigt als an derjenigen, die er während des Trainings einer VTU eingenommen hat, so sinkt die VTU-Antwort (Abb. 3-1 a, 3-2 a). Dies kann zu einem Abfall der Erkennungsleistung führen, d.h. an einer neuen Position kann eine andere VTU als die auf einen Stimulus trainierte bei Präsentation dieses Stimulus maximal aktiv sein (Abb. 3-1 b, 3-2 b), oder eine VTU kann dort auf einen anderen Stimulus stärker antworten als auf den, mit dem sie eigentlich trainiert wurde (Abb. 3-1 c, 3-2 c). Der letztere Fall ist um so wahrscheinlicher, je ähnlicher der andere Stimulus dem trainierten Stimulus ist (Abb. 3-1 d); für größere Morphabstände nimmt die Erkennungsleistung wieder zu. Die Änderungen in der VTU-Antwort und in der Erkennungsleistung hängen periodisch von der Position ab, aber sie sind prinzipiell identisch für Verschiebungen eines Stimulus in x- oder y-Richtung (Abb. 3-1 a) und treten für beide Stimulusklassen auf (vgl. Abb. 3-1 und 3-2). Die „Periodenlänge“ λ der Schwankungen von VTU-Antwort und Erkennungsleistung (also die Distanz, um die ein Stimulus verschoben werden muß, damit beide Werte sich nicht verändern) ist eine Funktion sowohl der Anzahl benachbarter S1-Zellen, die auf eine C1-Zelle projizieren (Parameter *poolRange*) als auch der Überlappungsbreite zweier benachbarter C1-Zellen (Parameter *c1Overlap*, siehe Kapitel 2). Es gilt folgende Gleichung:

$$\lambda = kgV_i \left[ceil \left(\frac{poolRange_i}{c1Overlap_i} \right) \right] \tag{3.1}$$

Dabei läuft i von 1 bis zur Anzahl der Filterbänder (4), kgV steht für das kleinste gemeinsame Vielfache, und $ceil$ bezeichnet das Runden nach oben. Normale HMAX-Parameter (*poolRange*-Werte von 4, 6, 9 bzw. 12 Pixeln in den Filterbändern 1 bis 4 sowie ein *c1Overlap* von 2, siehe Kapitel 2) ergeben also einen Wert für λ, der dem kleinsten gemeinsamen Vielfachen von 2, 3, 5 und 6 Pixeln entspricht, d.h. 30 Pixel. (Scheinbar kleinere Werte für λ in Abbildung 3-2 beruhen auf der Dominanz kleinerer Filter bei der Verarbeitung von Büroklammer-Stimuli in HMAX. Darauf wird in 3.2.2 und Abbildung 3-9 näher eingegangen.)

Die beobachteten Schwankungen in VTU-Antwort und Erkennungsleistung können durch einen „Verlust an Merkmalen“ eines Stimulus bei Verschiebung erklärt werden, der aufgrund der hierarchischen Verschaltung in HMAX auftritt. Dieses Phänomen wird in Abbildung 3-3 für den vereinfachten Fall nur einer einzigen Größe von S1-Filtern illustriert. Ein Merkmal des Stimulus (symbolisiert durch die beiden schwarzen Balken) wird in der Originalposition von zwei benachbarten C1-Zellen detektiert, die auf dieselbe S2-Zelle projizieren. Wird der Stimulus nach rechts verschoben (gestrichelte Balken), kann der rechte Teil des Merkmals außerhalb des rezeptiven Felds der rechten C1-Zelle zu liegen kommen, während der linke Teil noch immer von der linken C1-Zelle detektiert wird. Dieses Merkmal ist für die S2-Zelle, auf die diese beiden C1-Zellen projizieren, also „verloren“. Es kann aber auch nicht von der nächsten sich rechts anschließenden S2-Zelle (deren C1-Afferenzen in Abbildung 3-3 durch gepunktete Umrisse dargestellt sind) erkannt werden, da der linke Teil des Merkmals noch nicht innerhalb des rezeptiven Feldes der linken C1-Afferenz dieser S2-Zelle liegt. Erst wenn der Stimulus so weit verschoben wurde, daß alle Merkmale wiederum vom nächstgelegenen Satz S2-Zellen detektiert werden können, ist die Stimulusantwort von HMAX wieder identisch mit ihrem Wert für die ursprüngliche Position. Wie aus Abbildung 3-3 ersichtlich, ist das genau dann der Fall, wenn der Stimulus um eine Anzahl an Pixeln verschoben wird, die dem Versatz der C1-Zellen relativ zueinander entspricht, also genau dem Quotienten *poolRange/c1Overlap*. Wenn mehrere Filterbänder verwendet werden, von denen alle zur Gesamtantwort der C2-Zellen beitragen, gilt der allgemeinere Wert aus Gleichung 3.1.

Es ist wichtig zu bemerken, daß die Position eines Stimulus während des Trainings einer VTU an sich keineswegs vor anderen Positionen bevorzugt ist. Verschieben eines Stimulus kann ebensogut bewirken, dass neue Merkmale „auftauchen“, die an der ursprünglichen Position nicht detektiert wurden. Da die VTUs aber gemäß einer Gauß-Funktion, deren Maximum für das C2-Aktivitätsmuster während des Trainings erreicht wird, auf ihren C2-Input ansprechen, verursacht *jede* Veränderung in der C2-Aktivität eine geringere VTU-Feuerrate gegenüber dem Trainingswert, unabhängig davon, ob einzelne C2-Zellen stärker oder schwächer feuern.

Um die Erkennungsleistung in Gegenwart von Hintergrundstimuli zu verbessern, kann die Anzahl der C2-Afferenzen von VTUs reduziert werden, und zwar auf diejenigen C2-

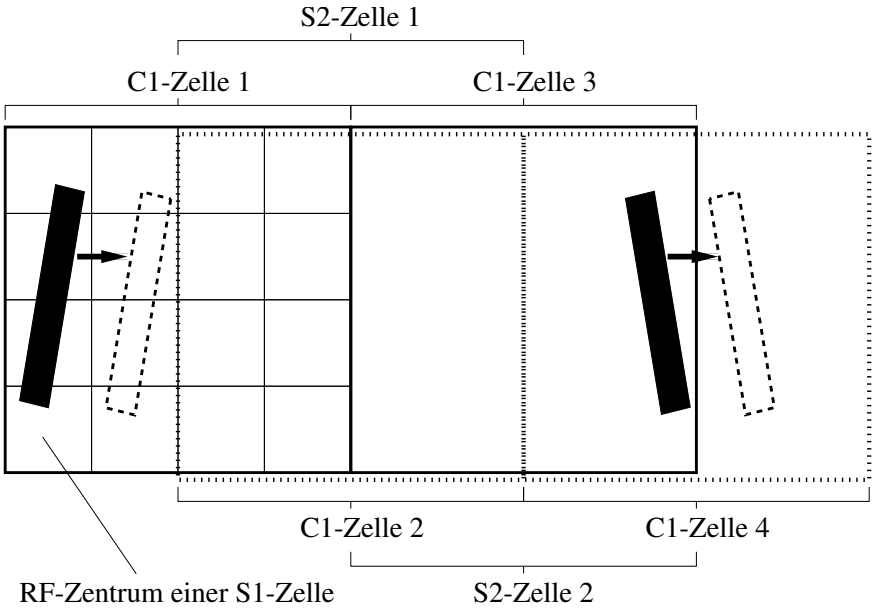

Abbildung 3-3: Schematische Darstellung des „Merkmalsverlusts“ aufgrund einer Veränderung in der Position eines Stimulus. In diesem Beispiel projizieren 4×4 benachbarte S1-Zellen auf eine C1-Zelle (d.h. *poolRange* $= 4$), und die rezeptiven Felder benachbarter C1-Zellen, gemessen in S1-Zellen, überlappen um die Hälfte (d.h. *c1Overlap* $= 2$). Nur 2 von 4 C1-Afferenzen einer S2-Zelle sind dargestellt. Erläuterungen siehe Text.

Zellen, die am stärksten auf den bevorzugten Stimulus der VTU antworten (siehe [54] und Kapitel 2). Dies führt zu einer geringeren Schwankungsamplitude der VTU-Antwort bei Translation eines Stimulus, weil schlicht weniger Terme im Exponenten der Gaußschen Antwortkurve der VTU erscheinen und daher weniger Differenzbeträge vom ursprünglichen Wert der C2-Aktivierung aufsummiert werden (siehe Abb. 3-2 a). Andererseits variiert die VTU-Antwort für kleinere Werte des Parameters σ (die Standardabweichung der Gaußschen Antwortkurve) stärker, wenn der Stimulus verschoben wird, da kleinere σ-Werte zu einer spezifischeren Abstimmung einer VTU auf ein bestimmtes C2-Aktivitätsmuster führen und die VTU-Antwort schon für geringere Abweichungen von diesem bevorzugten Input stärker abfällt. Wie aus Abbildung 3-2 ersichtlich, bedeutet ein größerer Abfall der VTU-Aktivität jedoch nicht notwendigerweise eine entsprechende Reduktion der Erkennungsleistung. Für mehr Afferenzen variiert die VTU-Antwort zwar stärker, die Schwankungen in der Erkennungsleistung sind aber sogar *kleiner*, als wenn weniger Afferenzen verwendet werden (siehe Abb. 3-2 b, c). Dies ist offensichtlich darauf zurückzuführen, daß bei Berücksichtigung einer größeren Anzahl C2-Zellen die zu den einzelnen Zielobjekten gehörenden C2-Vektoren stärker voneinander separiert sind, da in einem

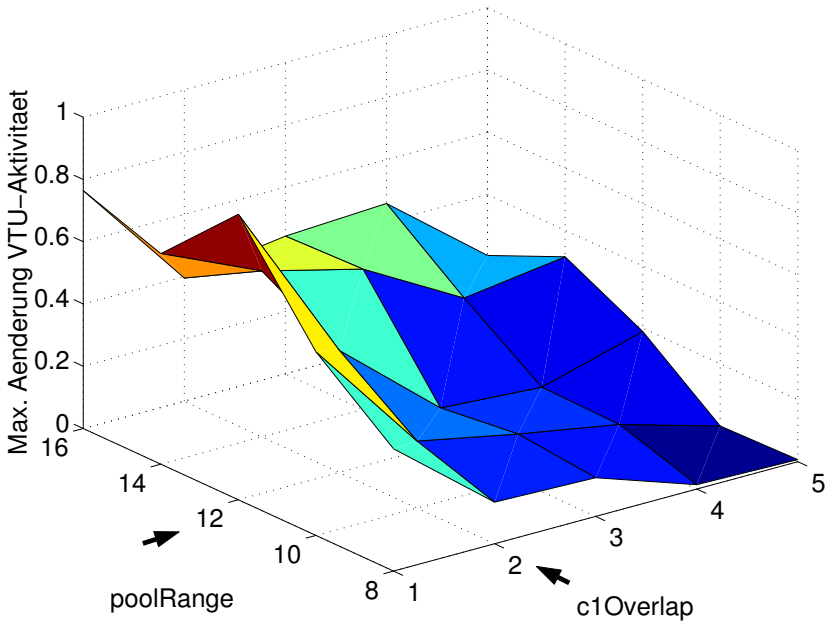

Abbildung 3-4: Abhängigkeit der maximalen bei Translation auftretenden Schwankungen der VTU-Aktivität von den Parametern *poolRange* und *c1Overlap* (siehe Kapitel 2). Die gezeigten Werte wurden für eine auf einen Auto-Prototypen trainierte VTU mit 256 Afferenzen und $\sigma = 0.2$ sowie für nur ein S1-Filterband (mit den größten Filtern, von 23×23 bis 29×29 Pixel) berechnet, damit nicht mehrere verschiedene Werte für *poolRange* pro Simulation (je einer für jedes Filterband) verwendet werden mußten. Pfeile markieren die HMAX-Standardwerte der Parameter. Erläuterungen im Text.

höherdimensionalen Raum operiert wird (hier 256- statt 40-dimensional), d.h. die von den verschiedenen Objekten hervorgerufenen Aktivitätsmuster auf Ebene der C2-Zellen unterscheiden sich stärker voneinander. Daher ist es weniger wahrscheinlich, daß sie durch die Schwankungen bei Verschiebung des Stimulus so stark verändert werden, daß sie eher dem bevorzugten Aktivitätsmuster einer anderen VTU ähneln als der, die auf den gezeigten Stimulus trainiert wurde.

Abbildung 3-4 zeigt, wie stark sich die Antwort einer VTU auf einen Stimulus maximal ändert, wenn er an unterschiedlichen Positionen präsentiert wird, in Abhängigkeit von der Anzahl S1-Afferenzen pro C1-Zelle (*poolRange*) und der Überlappung benachbarter C1-Zellen (*c1Overlap*). Translationsbedingte Aktivitätsänderungen sind im allgemeinen größer für mehr S1-Afferenzen pro C1-Zelle und geringere Überlappungsbreiten, weil auf diese Weise das visuelle Feld auf C1-Ebene gröber abgerastert wird und daher die Wahrscheinlichkeit, daß ein gegebenes Stimulusmerkmal bei Verschiebung „verlorengeht“, größer ist. Aus Abbildung 3-4 ist auch ersichtlich, daß für $poolRange = 16$ und

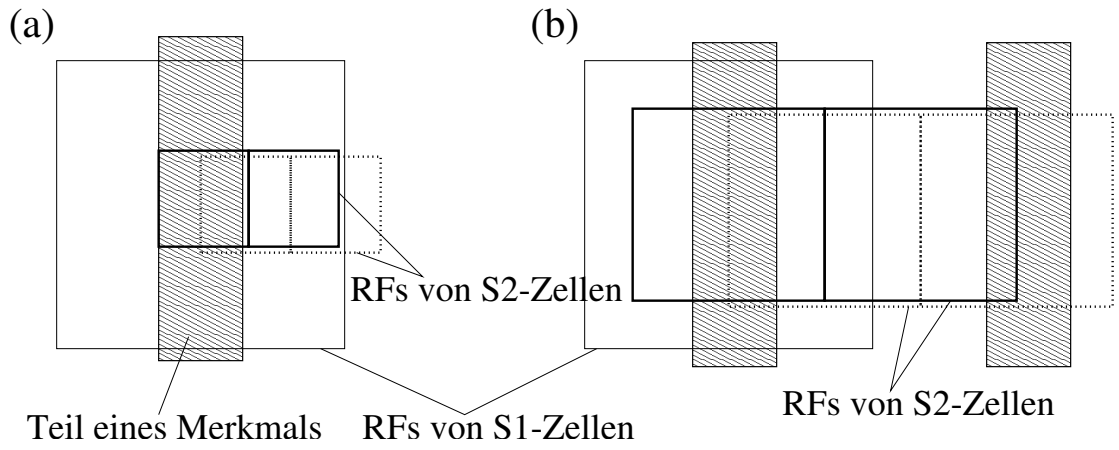

Abbildung 3-5: Illustration der unterschiedlichen Effekte von Translation eines Stimulus bei Verwendung großer S1-Filter und unterschiedlicher Mengen S1-Afferenzen pro C1-Zelle (Parameter *poolRange*). Wie in Abbildung 3-3 sind die rezeptiven Felder von C1- bzw. S2-Zellen in dieser Darstellung um die *Zentren* der rezeptiven Felder von S1-Zellen gezeichnet. Das tatsächliche rezeptive Feld einer S1-Zelle ist jedoch deutlich größer, wie diese Abbildung zeigt. **(a)** Große S1-Filter detektieren große Merkmale im visuellen Feld. Für kleine Werte von *poolRange* projizieren auf eine C1-Zelle nur wenige nebeneinander liegende S1-Zellen, die nahezu dieselben Merkmale detektieren, da sie aufgrund ihrer großen rezeptiven Felder und ihres Abstands von je nur einem Pixel stark überlappen. C1- / S2-Zellen detektieren daher kaum Kontraste im Stimulus, und Verschiebung des Stimulus ändert ihre Aktivität daher wenig. **(b)** Für größere Werte von *poolRange* beeinflussen größere Bereiche des visuellen Feldes die Aktivität einer C1- bzw. S2-Zelle. Daher werden hier trotz der großen S1-Filter eher komplexe Stimulusmerkmale und Kontraste detektiert (illustriert durch zwei schattierte Balken anstelle eines einzelnen in (a)), und eine Verschiebung des Stimulus ändert den visuellen Input von C1-/S2-Zellen und damit ihre Aktivität stärker.

$c1Overlap = 4$ größere Schwankungen in der VTU-Aktivität beobachtet werden als für $poolRange = 8$ und $c1Overlap = 2$, obwohl in beiden Fällen in Übereinstimmung mit Gleichung 3.1 die Periodenlänge λ der Schwankung dieselbe ist. Das läßt sich dadurch erklären, daß in den in Abbildung 3-4 ausgewerteten Simulationen nur die größten S1-Filter verwendet wurden. Wie in Abbildung 3-5 erläutert, detektieren große S1-Filter entsprechend große Merkmale im visuellen Feld, die innerhalb weniger Pixel kaum Helligkeitsänderungen aufweisen. Werden dennoch nur wenige unmittelbar benachbarte S1-Afferenzen pro C1-Zelle verwendet (kleine *poolRange*-Werte), fallen kaum Kontraste innerhalb des von einer C1- bzw. S2-Zelle abgedeckten Bereichs, und Verschiebungen des Stimulus ändern die Aktivitäten dieser Zellen daher wenig (Unterschiede in der absoluten Helligkeit alleine verursachen wegen der Normierung auf S1-Ebene keine Änderungen der Zellaktivität, siehe Kapitel 2). Umgekehrt variiert ein Bild nach der Verarbeitung durch

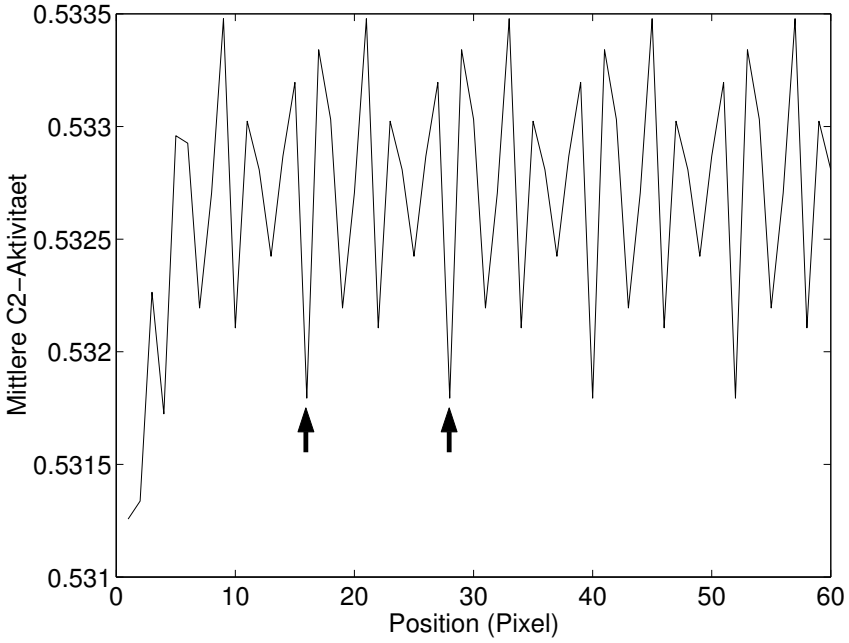

Abbildung 3-6: C2-Aktivität, über alle 256 C2-Zellen gemittelt, bei Präsentation eines Autos an verschiedenen Positionen. In dieser Simulation wurden für die vier Filterbänder weiterhin *poolRange*-Werte von 4, 6, 9 bzw. 12 verwendet, wie in den vorherigen Abbildungen, aber *c1Overlap* wurde auf den Wert 3 gesetzt. In Übereinstimmung mit Gleichung 3.1 verkleinerte sich daher die Periodenlänge der Aktivitätsschwankungen auf 12 Pixel (siehe Pfeile).

kleine S1-Filter über *geringe* Entfernungen in der Helligkeit am stärksten, und entsprechend sind Schwankungen in der Aktivität von C1- bzw. S2- Zellen dann größer, wenn nur *wenige* benachbarte S1-Zellen auf sie projizieren, weil sich kleinräumige Helligkeitsschwankungen nur über größere Entfernungen herausmitteln könnten (nicht abgebildet).

Die Feuerraten von C2-Zellen hängen in derselben Weise von der Position eines Stimulus ab wie die VTU-Aktivität, wie Abbildung 3-6 bestätigt. Der dort sichtbare zusätzliche Abfall der Aktivität für die am weitesten links im Bild gelegenen Stimuluspositionen läßt sich auf einen Randeffekt zurückführen. Zwar wurde in HMAX sichergestellt, daß auch ganz am Rand des modellierten visuellen Feldes gelegene S1-Zellen ebensoviel visuellen Input erhalten wie solche in der Bildmitte (dazu wird das ursprüngliche Bild intern mit einem „Rahmen“ von Pixeln, die auf den Wert 0 gesetzt werden, umgeben). Ein einfaches Merkmal wie ein Balken wird also auch detektiert, wenn es ganz am Rand des Bildes liegt. Eine S2-Zelle jedoch, die etwa ein komplexes Merkmal des präsentierten Stimulus mit ihren oben rechts und unten rechts gelegenen C1-Afferenzen detektiert, kann dieses Merkmal

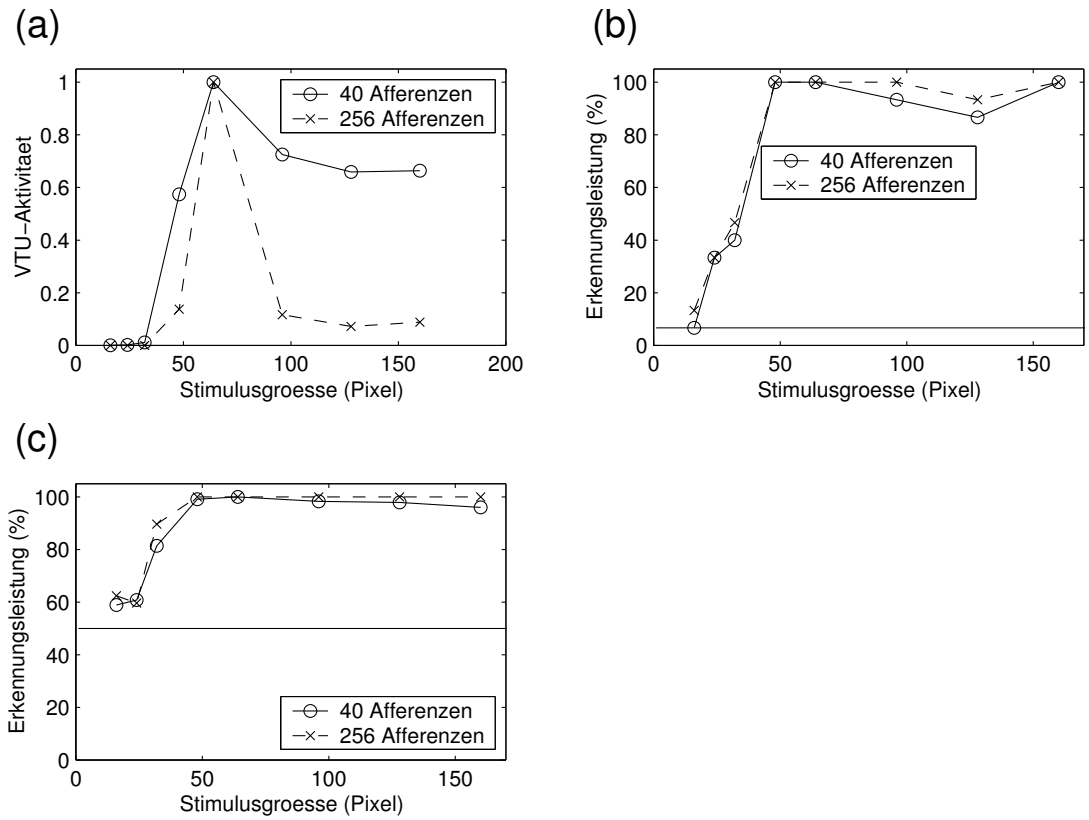

Abbildung 3-7: Auswirkungen von Änderungen der Stimulusgröße auf VTUs, die auf Büroklammern der Größe 64×64 Pixel trainiert wurden. **(a)** Antworten von VTUs mit 40 oder 256 Afferenzen und $\sigma = 0.2$ auf ihre trainierten Stimuli in verschiedenen Größen. **(b)** Erkennungsleistung im Aktivste VTU-Paradigma, gemittelt über 15 auf Büroklammern trainierte VTUs ($\sigma = 0.2$), in Abhängigkeit von der Größe ihrer bevorzugten Stimuli, für unterschiedlich viele C2-Afferenzen pro VTU. Die horizontale Linie zeigt das Zufallsniveau an (6.7%). **(c)** Wie (b), aber für das Stimulusvergleich-Paradigma. Das Zufallsniveau (50%) ist als horizontale Linie eingezeichnet.

womöglich nicht mehr erkennen, wenn es sich ganz am linken Bildrand befindet, und wird daher in diesem Fall schwächer feuern. Finden an anderen Positionen gelegene S2-Zellen des gleichen Typs auch kein gut auf ihre Stimuluspräferenz passendes Merkmal, wird sich die Abschwächung der Aktivität auch auf die C2-Zelle dieses Typs auswirken, wie es offensichtlich in Abbildung 3-6 geschehen ist.

3.2.2 Änderungen der Stimulusgröße

Invarianz der Antworten von C2-Zellen und VTUs gegenüber Änderungen der Größe eines Stimulus wird in HMAX dadurch erreicht, daß afferente Zellen, die auf dasselbe visu-

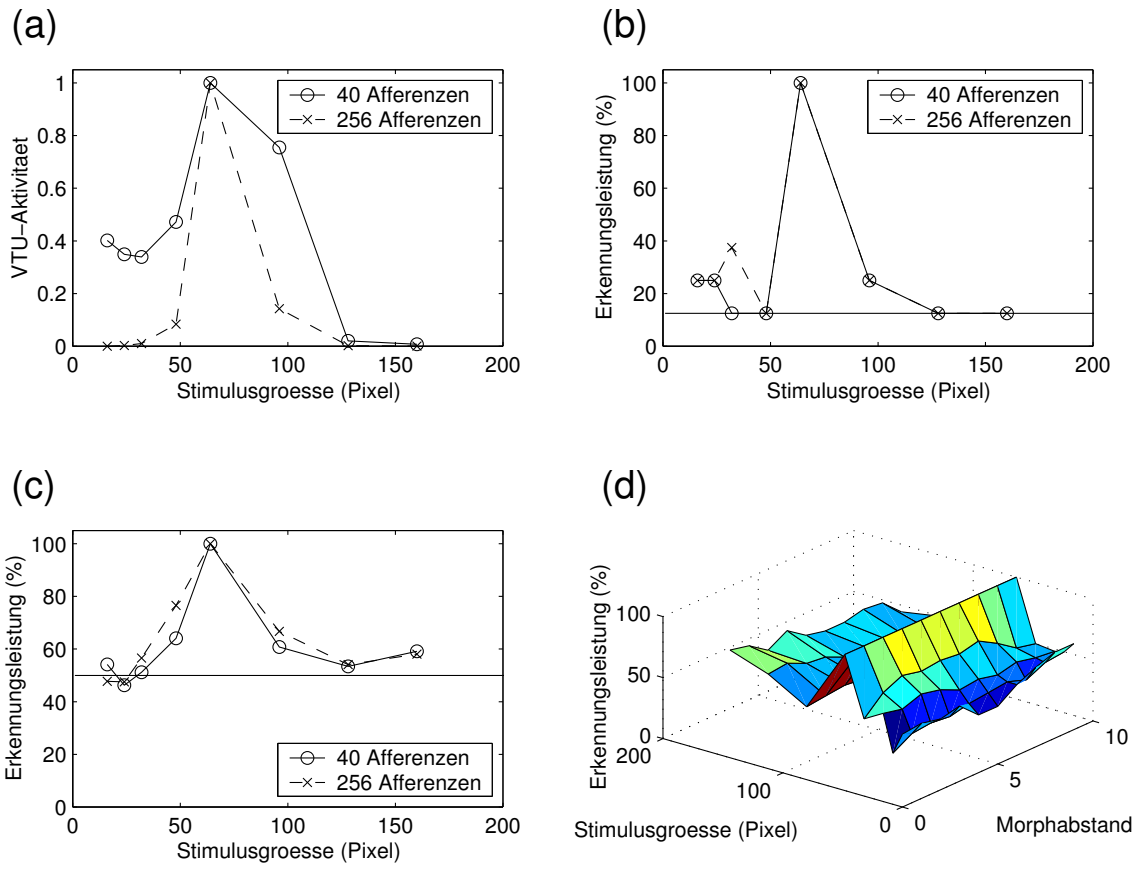

Abbildung 3-8: Auswirkungen von Änderungen der Stimulusgröße auf VTUs, die auf Autos der Größe 64×64 Pixel trainiert wurden. **(a)** Antworten von VTUs mit 40 oder 256 Afferenzen und $\sigma = 0.2$ auf ihre trainierten Stimuli in verschiedenen Größen. **(b)** Erkennungsleistung im Aktivste VTU-Paradigma, gemittelt über 8 auf Autos trainierte VTUs ($\sigma = 0.2$), in Abhängigkeit von der Größe ihrer bevorzugten Stimuli, für unterschiedlich viele C2-Afferenzen pro VTU. Das Zufallsniveau (12.5%) ist als horizontale Linie eingezeichnet. **(c)** Wie (b), aber für das Stimulusvergleich-Paradigma. Die horizontale Linie zeigt das Zufallsniveau an (50%). **(d)** Mittlere Erkennungsleistung im Stimulusvergleich-Paradigma für 8 Auto-VTUs mit 40 Afferenzen und $\sigma = 0.2$, aufgetragen in Abhängigkeit von Stimulusgröße und Ähnlichkeit der Distraktoren („Morphabstand“).

elle Merkmal in verschiedenen Größen bevorzugt antworten, auf dasselbe nachgeschaltete Modellneuron projizieren. Wie bereits in [55] gezeigt, funktioniert dies für Büroklammer-Stimuli über einen weiten Skalierungsbereich sehr gut: VTUs, die auf Büroklammern der Größe 64×64 Pixel trainiert wurden, zeigen sehr robuste Leistungen bei der Objekterkennung für vergrößerte und gute Leistungen auch für nicht zu stark verkleinerte Darstellungen der Stimuli (siehe Abbildung 3-7).

Ein völlig anderes Bild ergibt sich jedoch für Autos. Wie Abbildung 3-8 zeigt, erkennt HMAX sowohl vergrößerte als auch verkleinerte Autos deutlich schlechter als Büroklammern. Abbildung 3-8 d macht deutlich, daß sich die Leistung auch für sehr unterschiedliche Distraktorstimuli (große Werte für den Morphabstand) nicht verbessert. An einer neuen Position kann ein Stimulus dagegen besser von einem unähnlichen als von einem ähnlichen Distraktor unterschieden werden (vgl. Abb. 3-1 d). Da aufgrund der limiterten Auflösung charakteristische Merkmale eines Stimulus bei Verkleinerung rasch unkenntlich werden, ist klar, daß verkleinerte Darstellungen eines Stimulus generell schlechter erkannt werden können, unabhängig von der Stimulusklasse. Abbildung 3-9 zeigt, warum das bei Autos auch für vergrößerte Darstellungen der Fall ist. In Abbildung 3-9 a werden zunächst die Aktivitäten der vier HMAX-Filterbänder für Autostimuli in den Größen 64×64 und 128×128 Pixel miteinander verglichen. Beispielsweise ist die Aktivität in Filterband 4 (große Filter von 23×23 bis 29×29 Pixel) bei Präsentation von Autos der Größe 128×128 Pixel sehr ähnlich der Aktivität in Filterband 2 (Filter von 11×11 bis 15×15 Pixel) für die Präsentation von Autos in der Größe 64×64 Pixel (erkennbar an der vergleichsweise geringen euklidischen Distanz der C2-Vektoren in der Grafik). D.h., Merkmale des ursprünglichen Stimulus, die von Filtern bestimmter Größe detektiert werden, aktivieren nach Verdopplung der Stimulusgröße auch doppelt so große Filter bevorzugt. Das entspricht den Erwartungen an das Verhalten des Modells für verschiedene Stimulusgrößen. Abbildung 3-9 b zeigt jedoch, daß Autos schon in der Größe 64×64 Pixel vor allem die großen Filter in Filterband 4 aktivieren, während Büroklammern derselben Größe hauptsächlich von Filtern bis etwa 15×15 Pixel detektiert werden. Das liegt sehr wahrscheinlich daran, daß die hier verwendeten Autostimuli sehr wenig innere Strukturen aufweisen, so daß ihre hervorstechendsten Merkmale ihre Umrisse sind. In vergrößerten Darstellungen dieser Autos werden daher diejenigen charakteristischen Merkmale, auf die in der Größe 64×64 Pixel die VTUs vor allem trainiert wurden, zu groß sein, um noch von

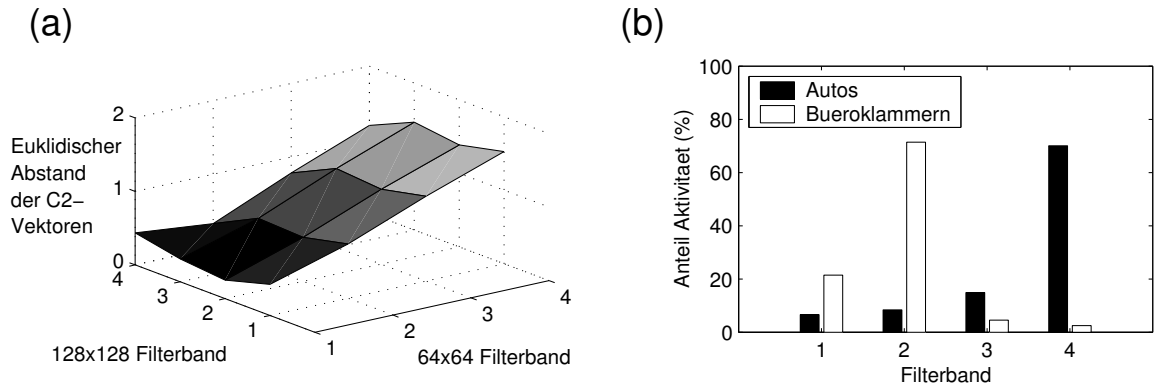

Abbildung 3-9: Aktivität in den einzelnen Filterbändern von HMAX. Filterband 1: Filter 7×7 und 9×9 Pixel; Filterband 2: von 11×11 bis 15×15 Pixel; Filterband 3: von 17×17 bis 21×21 Pixel; Filterband 4: von 23×23 bis 29×29 Pixel. **(a)** Euklidische Abstände der C2-Aktivitätsvektoren, die für nur einzelne Filterbänder und Auto-Prototypen der Größen 64×64 oder 128×128 Pixel berechnet wurden. Kleine Werte entsprechen ähnlichen Aktivierungsmustern der C2-Population in zwei Situationen. **(b)** Filterbandquelle der C2-Aktivität für 8 Auto-Prototypen und 8 zufällig ausgewählte Büroklammern, alle in der Größe 64×64 Pixel. Jeder Balken zeigt den Anteil derjenigen C2-Zellen an der Gesamtpopulation, die in ihrer Aktivität vom betreffenden Filterband bestimmt wurden, wenn ein Auto bzw. eine Büroklammer gezeigt wurde. Das Filterband, das die Antwort einer C2-Zelle bestimmt, enthält die am stärksten aktiven afferenten S2-Zellen unter denen, die auf dasselbe Merkmal wie die betrachtete C2-Zelle selektiv antworten. Es erlaubt Rückschlüsse auf die Größe, in der ein entsprechendes Merkmal im Stimulus vorkommt.

den S1-Zellen des gegenwärtigen Modells detektiert zu werden, und andere Merkmale werden die Antwort der C2-Zellen bestimmen, was als Ursache für die deutlich reduzierte Erkennungsleistung angesehen werden kann.

Es zeigt sich also, daß sich Translation und Skalierung von Stimuli auf die Antwort von HMAX-Modellneuronen unterschiedlich auswirken. In beiden Fällen handelt es sich um affine Transformationen in der Bildebene, deren Auswirkungen auf einen Stimulus – im Gegensatz etwa zu Rotationen dreidimensionaler Stimuli in der Tiefe – exakt durch sein Aussehen an einer Position und in einer Größe bestimmt sind. Außerdem wird in HMAX beiden Transformationen im Prinzip auf dieselbe Weise Rechnung getragen: durch die Projektion von Zellen, die auf verschieden transformierte Versionen desselben Merkmals bevorzugt feuern, auf dieselbe nachgeschaltete Zelle, die aufgrund der MAX-Operation unabhängig davon feuert, von welcher ihrer Afferenzen ihr bevorzugter Stimulus detektiert wurde. Die Veränderungen der neuronalen Aktivität und der Erkennungsleistung des Modells bei Translation von Stimuli sind im Prinzip unabhängig von der Stimulusklasse.

Zwar unterscheiden sich die Beträge der beobachteten Schwankungen in der Erkennungsleistung für die verschiedenen Stimuli; daß sie für Büroklammern generell geringer sind als für Autos (vgl. Abb. 3-1 und 3-2), kann mit dem Wissen aus Abbildung 3-9, daß Büroklammern kleinere S1-Filter bevorzugt erregen als Autos, auf der Basis der Argumentation zu Abbildung 3-5 dadurch erklärt werden, daß die detektierten Merkmale groß im Verhältnis zu den Ausmaßen der am stärksten aktiven Filter und daher weniger anfällig dafür sind, an anderen Positionen nicht mehr detektiert zu werden. Die Invarianzbereiche bei der Translation verschiedener Stimuli, d.h. die Entfernungen, über die ein Stimulus verschoben werden kann, ohne daß seine Erkennung beeinträchtigt wird, sind jedoch für verschiedene Objektklassen jeweils dieselben. Die Leistung von HMAX bei der Erkennung skalierter Stimuli ist dagegen offensichtlich stark davon abhängig, wie gut die in HMAX detektierten komplexen Merkmale tatsächlich die charakteristischen Merkmale dieser Stimuli abbilden. S2- und C2-Zellen antworten bevorzugt auf Kombinationen von 2×2 benachbarten orientierten Balken und sind daher offenkundig gut geeignet zur Erkennung von Büroklammer-Stimuli, da diese effektiv aus solchen orientierten Balken bestehen. Außerdem stimmt die Größe der von den Büroklammern am stärksten erregten S1-Filter gut mit ihrer eigenen Größe überein, so daß bei Skalierung dieser Stimuli der afferente Input aus verschiedenen Filterbändern genutzt werden kann. Beides führt zu guter Erkennungsleistung für skalierte Büroklammern in HMAX. Für Autos sind diese beiden Bedingungen nicht gegeben, so daß die Erkennungsleistung bei Größenänderung stark abfällt.

3.2.3 Kontraständerungen

Abbildung 3-10 zeigt das Verhalten von C2-Zellen und einer exemplarischen VTU bei Veränderungen im Kontrast eines Stimulus. Die mittlere Antwort der C2-Zellen skaliert im Bereich der hier verwendeten Kontraste nahezu linear mit dem Stimuluskontrast (a). Natürlich kann aber, entsprechend der unterschiedlich starken Aktivierung einzelner C2-Zellen für einen bestimmten Stimulus, eine C2-Zelle mehr oder weniger große absolute Aktivitätsänderungen bei Änderung des Stimuluskontrasts an den Tag legen (b). Das lineare Ansprechverhalten von C2-Zellen für verschiedene Kontraste führt zu einem „sigmoidalen“ (genauer: Gaußschen) Aktivitätsprofil bei VTUs, die auf ihre Zielobjekte bei 100% Kontrast trainiert sind, wie es auch in den folgenden Kapiteln stets der Fall sein

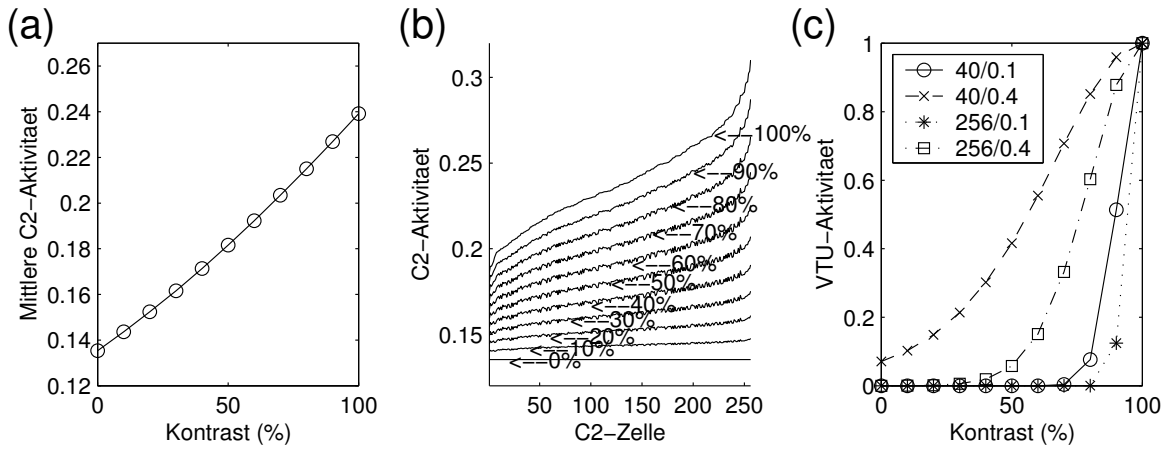

Abbildung 3-10: Typische Kontrastantwort von C2-Zellen und VTUs, hier für Auto-Prototyp Nr. 3. **(a)** Über alle 256 Zellen gemittelte C2-Antwort auf diesen Stimulus bei verschiedenen Kontrasten. **(b)** Antwort aller C2-Zellen für unterschiedliche Kontraste. Gezeigt sind, in jedem Graphen, die Antworten aller C2-Zellen für Auto-Prototyp 3 bei jeweils einem Kontrastwert (siehe Pfeile). Die Zellen sind gemäß ihrer Feuerrate bei 100% Kontrast sortiert, d.h. die Antwort einer bestimmten C2-Zelle ist für alle Kontraste stets an derselben Abszissenposition dargestellt. **(c)** Antwort der auf Auto-Prototyp 3 trainierten VTU bei Präsentation dieses Stimulus mit verschiedenen Kontrastwerten. Die Werte in der Legende bezeichnen die Anzahl der C2-Afferenzen für diese VTU (40 oder 256) und ihren σ-Wert (0.1 oder 0.4).

wird. Die VTU-Antwort sinkt bei geringerem Kontrast schneller, wenn mehr Afferenzen oder ein kleinerer σ-Wert verwendet werden, da beides zu einer spezifischeren Abstimmung der VTU auf ein bestimmtes C2-Aktivitätsmuster führt (c).

In Abbildung 3-11 wird die Erkennungsleistung von HMAX für Stimuli unterschiedlichen Kontrasts untersucht. Es ist offensichtlich, daß in HMAX Stimuli nicht unabhängig von ihrem Kontrast erkannt werden. Vor allem für Autos fällt die Erkennungsleistung bei Kontrasten unterhalb des für das VTU-Training verwendeten Werts in beiden Paradigmen sehr schnell auf Zufallsniveau ab (a, b). Die Leistung verbessert sich auch nicht, wenn als Afferenzen für die VTUs nur die 40 am stärksten auf das Zielobjekt ansprechenden C2-Zellen verwendet werden (a). Büroklammern können hingegen auch bei niedrigeren Kontrastwerten noch gut erkannt und unterschieden werden, zumindest für 40 Afferenzen pro VTU (a, c). Die neuronalen Repräsentationen der Büroklammern auf C2-Ebene skalieren also gleichmäßiger mit dem Stimuluskontrast als die der Autos. Analog den obigen Ergebnissen für verschieden große Stimuli läßt sich auch dieser Effekt dadurch erklären, daß die von HMAX detektierten Stimulusmerkmale besonders gut für Büroklammer-Stimuli ge-

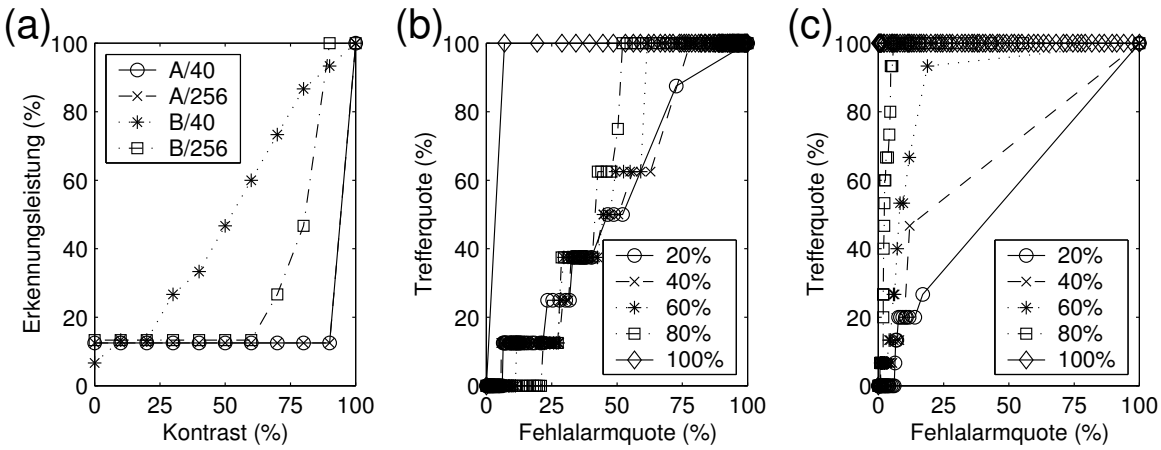

Abbildung 3-11: Erkennung von Autos und Büroklammern bei verschiedenen Kontrasten. **(a)** Erkennungsleistung im Aktivste VTU-Paradigma für einzeln präsentierte Autos und Büroklammern, über alle Zielobjekte (8 Auto-Prototypen bzw. 15 Büroklammern) gemittelt. Die Legende zeigt für jeden Graphen die Objektklasse (A = Auto, B = Büroklammer) und die Anzahl der Afferenzen einer VTU (40 oder 256). Das Zufallsniveau ist für Autos 12.5%, für Büroklammern 6.7%. **(b)** ROC-Kurven für die Erkennungsleistung im Stimulusvergleich-Paradigma für einzeln präsentierte Autos bei unterschiedlichen Kontrasten (siehe Legende) und für 40 Afferenzen pro VTU. Die Aktivität einer VTU bei Präsentation ihres trainierten Prototypen wird mit ihrer Antwort auf alle Autostimuli verglichen, die auf von diesem Prototypen ausgehenden Morphlinien liegen (siehe Kapitel 2). **(c)** Wie (b), jedoch für Büroklammern. VTU-Antworten auf ihre bevorzugten Stimuli werden mit ihren Feuerraten für 60 Distraktor-Büroklammern verglichen.

eignet sind bzw. Büroklammern ebendie Merkmale aufweisen, auf die S2- bzw. C2-Zellen besonders gut ansprechen. Dies wirkt sich nicht so sehr dadurch positiv auf die Erkennungsleistung aus, daß die allgemeine neuronale Aktivität in HMAX bei der Präsentation von Büroklammern höher ist. Das ist zwar der Fall, aber entsprechend sinken die Aktivitäten bei Kontrastabnahme auch schneller. Entscheidend ist vielmehr, daß aufgrund der guten Eignung der von HMAX detektierten Merkmale für die Büroklammern verschiedene Vertreter dieser Stimulusklasse stärker unterschiedliche C2-Aktivitätsvektoren als etwa verschiedene Autos erzeugen (siehe auch 3.2.4) und daher in HMAX auch bei einem kontrastbedingten generellen Verlust an C2-Aktivität noch recht gut voneinander unterscheidbar sind.

Im Gegensatz zur Skalierung von Stimuli werden die Effekte von Kontraständerungen in HMAX nicht durch einen expliziten neuronalen Mechanismus berücksichtigt, d.h. es gibt keine Zellen, die auf dasselbe Merkmal bei verschiedenen Kontrasten bevorzugt antwor-

ten und auf eine einzige nachgeschaltete Zelle projizieren, die gemäß ihrem stärksten Input feuert. Das Modell könnte freilich dennoch gegen Änderungen im Stimuluskontrast invariant gemacht werden, etwa durch Normalisierung eines jeden von einem S1-Filter verarbeiteten Bildausschnitts auf einen mittleren Pixelwert von 0. Alle Änderungen des Kontrasts werden dadurch zu multiplikativen Änderungen von Pixelwerten, die sich durch die von den S1-Zellen durchgeführte Normalisierung ihres Outputs herauskürzen. Allerdings ginge dadurch jegliche Kontrastabhängigkeit der Feuerrate der C2-Zellen verloren, was nicht den experimentellen Daten für Neuronen im Areal V4 entspricht, die die C2-Zellen simulieren sollen [1]. Außerdem werden Effekte von Aufmerksamkeit auf neuronaler Ebene in der Regel bei niedrigen und mittleren Stimuluskontrasten und dementsprechend bei Feuerraten deutlich unterhalb der Saturation beobachtet [40, 52]. Da sich die C2-Zellen im Bereich der verwendeten Kontraste linear gegenüber Kontraständerungen verhalten, also ähnlich wie reale Neuronen, wenn sie durch Aufmerksamkeit modulierbar sind, wird im folgenden die Kontrastabhängigkeit der Aktivität der C2-Zellen beibehalten. Ihr lineares Antwortverhalten kann dann auch dazu genutzt werden, um durch einfache additive oder multiplikative Erhöhung ihrer Aktivität die im Experiment beobachteten Effekte von Aufmerksamkeit zu simulieren (siehe Kapitel 5).

3.2.4 Hinzufügen von Distraktorstimuli

Die Abbildungen 3-12 und 3-13 zeigen, daß das Hinzufügen eines zweiten Stimulus zu einem Zielobjekt in der Tat die Erkennung dieses Objekts beeinträchtigen kann, sofern der Distraktor mit dem gleichen oder höherem Kontrast präsentiert wird. Aufgrund der in 3.2.3 diskutierten geringen Kontrastinvarianz der Erkennung von Autos in HMAX erreicht die Erkennungsleistung für diese Stimulusklasse auch unabhängig von einem Distraktor rasch das Zufallsniveau, wenn der Kontrast des Zielobjekts unter dem Wert im VTU-Training liegt. Für den Fall, daß beide Autos mit 100% Kontrast präsentiert werden, ist die Interferenz der beiden Stimuli größer (d.h. die Erkennungsleistung geringer), wenn der Distraktor *weniger* Ähnlichkeit mit dem Zielobjekt aufweist (d.h. für größere Morphabstände, siehe Abbildungen 3-12 a, b und vgl. c mit d). Bei Autos ist außerdem keine Abhängigkeit der Erkennungsleistung von der Anzahl der Afferenzen einer VTU sichtbar (vgl. Abbildung 3-12 a und b), im Gegensatz zu Büroklammern (siehe Abb. 3-13

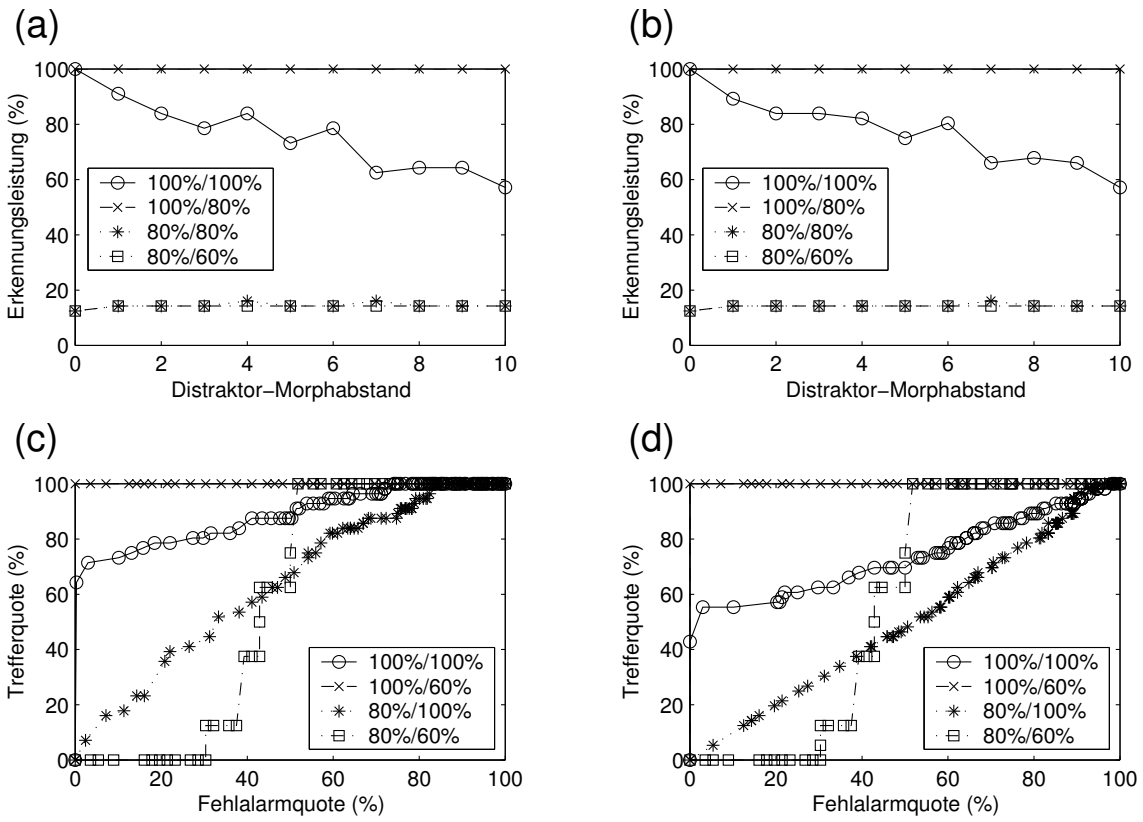

Abbildung 3-12: Erkennung eines Autos in Gegenwart eines zweiten Autos. **(a)** Erkennungsleistung im Aktivste VTU-Paradigma für unterschiedliche Kontrastwerte und 40 C2-Afferenzen pro VTU. Legende: Kontrastwerte für Zielobjekt (erster Wert) und Distraktor (zweiter Wert). **(b)** Wie (a), aber für 100 Afferenzen pro VTU. **(c)** ROC-Kurven für die Erkennungsleistung im Stimulusvergleich-Paradigma. 40 Afferenzen pro VTU. Der Distraktor ist stets 5 Morphschritte vom Zielobjekt entfernt. Legende: Kontrastwerte für Zielobjekt (erster Wert) und Distraktor (zweiter Wert). **(d)** Wie (c), aber für Distraktoren im maximalen Morphabstand (10).

a). Beide Resultate erscheinen paradox. Man würde erwarten, daß ein ähnlicherer Distraktor eher mit der Erkennung eines Zielobjekts interferiert, und für Büroklammer-Stimuli wurde bereits in [54] gezeigt, daß sich die Erkennungsleistung in Gegenwart von Distraktoren verbessern läßt, wenn nur die C2-Zellen, die am stärksten auf einen Stimulus antworten, auf die VTU projizieren, die auf diesen Stimulus trainiert ist. Ein Blick auf die von den 8 Auto-Prototypen am stärksten erregten C2-Zellen zeigt jedoch, daß es für alle Autos zum großen Teil dieselben sind. Werden 40 Afferenzen pro VTU verwendet, so sind alle 8 auf Auto-Prototypen trainierten VTUs gerade einmal mit insgesamt 63 C2-Zellen verbunden, d.h. die verschiedenen Auto-VTUs haben nahezu dieselben Afferenzen. Anstatt *unterschiedliche* C2-Zellen zu aktivieren, erregen verschiedene Autostimuli *dieselben* Zellen

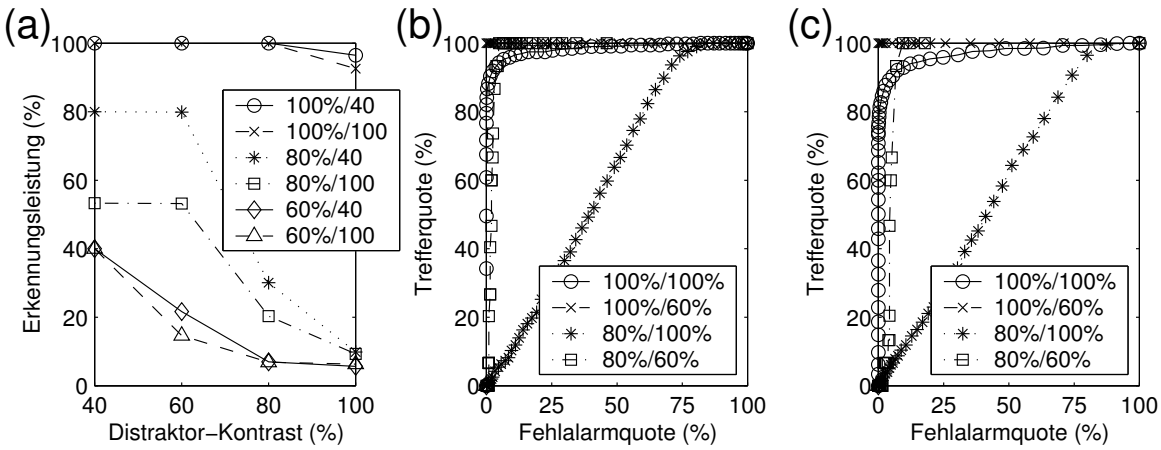

Abbildung 3-13: Erkennung einer Büroklammer in Gegenwart einer zweiten Büroklammer. **(a)** Erkennungsleistung im Aktivste VTU-Paradigma für unterschiedliche Kontrastwerte des Distraktors. Legende: Kontrast des Zielobjekts (erster Wert) sowie Anzahl der Afferenzen der VTUs (zweiter Wert). **(b)** ROC-Kurven für die Erkennungsleistung im Stimulusvergleich-Paradigma. 40 Afferenzen pro VTU. Legende: Kontrast von Zielobjekt (erster Wert) und Distraktor (zweiter Wert). **(c)** Wie (b), aber für 100 Afferenzen pro VTU.

in unterschiedlichem Maß. Deshalb verändert sich das Aktivitätsmuster der Afferenzen einer Auto-VTU um so mehr, je weniger der verwendete Distraktor dem Zielobjekt ähnlich ist. Der hohe Grad an Überlappung zwischen den Mengen der Afferenzen verschiedener VTUs ist auch der Grund, warum für Autos die Erkennungsleistung kaum von der Anzahl der Afferenzen pro VTU abhängt. Selbst wenn eine VTU nur wenige Afferenzen hat, sind es schon zumeist solche, die auch von anderen VTUs verwendet werden und daher auf andere Stimuli ebenfalls stark antworten. Ein Distraktorstimulus beeinflußt also auch die am stärksten feuernden Afferenzen der VTU eines Zielobjektes, d.h. eine geringe Anzahl Afferenzen bringt in diesem Fall keine Vorteile für die Stimuluserkennung.

Für Büroklammern ist die Relation zwischen Anzahl der Afferenzen und Erkennungsleistung dagegen wie erwartet: je mehr Afferenzen eine VTU verwendet, desto größer ist die Wahrscheinlichkeit, daß ein Distraktorstimulus zumindest einige davon ebenfalls beeinflußt, und entsprechend sinkt die Leistung bei der Objekterkennung (siehe Abb. 3-13 a). Dies ist darauf zurückzuführen, daß für Büroklammern die Mengen der Afferenzen verschiedener VTUs sich sehr viel weniger überschneiden. VTUs 1 bis 8 der 15 auf Büroklammern trainierten VTUs haben zusammen beispielsweise 142 C2-Zellen als Afferenzen, wenn 40 Afferenzen pro VTU verwendet werden – deutlich mehr als die 8 auf Autos trai-

nierten VTUs. Daher bringt eine geringere Anzahl Afferenzen hier in der Tat Vorteile für die Erkennungsleistung.

Im Unterschied zum Aktivste VTU-Paradigma zeigt die Erkennungsleistung im Stimulusvergleich-Paradigma (siehe Abb. 3-13 b, c) vergleichsweise wenig Abhängigkeit vom Kontrast des Zielobjektes (sofern der Distraktor nicht mit höherem Kontrast gezeigt wird) oder von der Anzahl Afferenzen. Das macht deutlich, daß die Leistung desselben Systems bei unterschiedlichen Aufgaben in unterschiedlicher Weise von solchen Parametern abhängen kann. Selbst wenn ein Neuron bei niedrigen Kontrasten nicht mehr stärker feuert als andere, wenn sein bevorzugter Stimulus präsentiert wird, kann es möglicherweise im direkten Vergleich zweier Stimuli noch gut zwischen trainiertem Objekt und anderen Objekten unterscheiden (d.h. eine korrekte Antwort im „two alternative forced choice"-Experiment liefern).

Diese Resultate zeigen, daß in HMAX Objekte in Gegenwart anderer Stimuli erkannt werden können, sofern sie selbst kontrastreich genug sind und nicht zusammen mit kontrastreicheren Stimuli gezeigt werden. Vor allem ist für eine gute Erkennungsleistung aber ausschlaggebend, daß verschiedene Objekte auf Ebene der C2-Zellen hinreichend verschiedene Mengen von Zellen aktivieren, wie im Fall von Büroklammern. Da Autos hingegen, als Vertreter einer „realistischen " Objektklasse mit sehr ähnlichen äußeren Umrissen (bzw. einer ähnlichen dreidimensionalen Form), zumeist dieselben C2-Zellen bevorzugt erregen, können sie in Gegenwart eines zweiten Autos deutlich weniger gut erkannt werden. Wie in den vorhergehenden Abschnitten ist auch dies eine Folge der von HMAX auf Ebene der S2- bzw. C2-Zellen detektierten komplexen Objektmerkmale, die sich auch hier als optimal für Büroklammer-Stimuli erweisen, nicht so sehr hingegen für realistischere Stimulusklassen. Eine weitere Bedingung für Invarianz von Objekterkennung, diesmal gegenüber Hintergrundstimuli, ist also, daß verschiedene Objekte möglichst verschiedene Neuronengruppen bevorzugt erregen sollten [54].

3.3 Diskussion

Invariante Erkennung von Objekten an verschiedenen Positionen, in unterschiedlichen Größen, bei verändertem Kontrast und in Gegenwart von störendem Hintergrund gehört

zu den herausragenden Leistungen des visuellen Systems, die zu modellieren alles andere als trivial ist. Während die Erkennung gelernter Objekte freilich am sichersten dadurch erreicht werden könnte, daß für jedes Objekt in jeder möglichen Erscheinungsform (also an allen Positionen, in allen Größen usw.) ein Detektorneuron vorgesehen wird, würde ein solcher Ansatz selbst die im Gehirn zur Verfügung stehenden enormen Verarbeitungskapazitäten übersteigen, aufgrund einer „kombinatorischen Explosion“ der Zellzahl, würde man dieses Prinzip für alle relevanten Stimuli umsetzen wollen. Außerdem wäre es auf diese Weise nicht möglich, einen einmal in einer Größe und an einer Position gesehenen Stimulus anderswo und in einer anderen Größe wiederzuerkennen [54]. In HMAX werden daher, um Objekterkennung von Position und Größe unabhängig zu machen, gemäß des von Hubel und Wiesel [22] zuerst formulierten Prinzips der hierarchischen Verschaltung komplexere Stimuluspräferenzen erst für solche Zellen erzeugt, die größere Invarianzbereiche für Position und Größe aufweisen, während diejenigen Neuronen, die auf eine Position im visuellen Feld und eine Stimulusgröße spezialisiert sind, einfache Stimulusmerkmale detektieren. Die Spezifität der Antwort einer Zelle auf ein bestimmtes Merkmal, selbst für große rezeptive Felder, wird dabei durch die „MAX“-Operation bewahrt, die dafür sorgt, daß jeweils nur der am besten auf die Stimuluspräferenz der Zelle passende Stimulus innerhalb ihres rezeptiven Feldes ihre Aktivität bestimmt. Auf diese Weise kann erreicht werden, daß nicht für jeden neu gelernten Stimulus auf jeder Ebene des visuellen Systems neue Zellen und Verschaltungen angelegt werden müssen, sondern idealerweise das Aktivitätsmuster über die vorhandenen, komplexe Merkmale detektierenden Neuronen ausreicht, um eine große Zahl verschiedener Stimuli zu identifizieren und zu unterscheiden [55]. Das generelle Prinzip, komplexere Simuluspräferenzen durch hierarchische Verschaltung von Neuronen einfacherer Präferenz zu generieren, ist seit Hubel und Wiesel weithin akzeptiert und findet nach wie vor experimentelle Bestätigung in verschiedenen corticalen Arealen [39, 49, 50, 75].

Es stellt sich jedoch zum einen heraus, daß die Diskretisierung des visuellen Feldes in rezeptive Felder sowie die hierarchische Verschaltung der Neuronen eine Antwort des HMAX-Modells erzeugen, die nicht völlig von der Position eines Stimulus unabhängig ist: Ein „Merkmalsverlust“ kann auftreten, wenn ein Objekt an einer anderen Position gezeigt wird als dort, wo es ursprünglich gelernt wurde, mit der möglichen Folge, daß es nicht mehr korrekt erkannt wird. Das kann insofern nicht zu sehr überraschen, als, wie eben

diskutiert, in HMAX keine Detektorneuronen für jeden Stimulus an jedem Ort existieren. Die Variationen, denen die Antwort von HMAX sowie die Erkennungsleistung für verschiedene Positionen eines Objekts unterliegen, können in Abhängigkeit von der Position und den verwendeten Parameterwerten des Modells beschrieben werden. Insbesondere tritt kein Merkmalsverlust auf, wenn ein Objekt an eine in bezug auf die Organisation der C1- und S2-Zellen äquivalente Position verschoben wird. Äquivalente Positionen sind dabei, wie in Gleichung 3.1 beschrieben, eine Anzahl Pixel voneinander entfernt, die sich als das kleinste gemeinsame Vielfache der Quotienten *poolRange* / *c1Overlap* für die verwendeten Filterbänder ergibt.

Ein wichtiges Ergebnis der Simulationen zur Positionsinvarianz, das aber auch generell Gültigkeit hat, ist, daß ein Absinken der Aktivität der Modellneuronen nicht unbedingt mit einem entsprechenden Verlust an Erkennungsleistung einhergehen muß. Objekterkennung hängt im HMAX-Modell von der *relativen* Aktivität einer VTU bei verschiedenen Stimuli oder im Vergleich zu anderen VTUs ab und kann daher auch bei deutlich reduzierter absoluter Aktivität erhalten bleiben. Das entspricht experimentellen Resultaten, die zeigen, daß trotz starker Einbußen in neuronalen Aktivitäten, die in den untersuchten Fällen durch Distraktorstimuli oder eine komplexe natürliche Umgebung zustande kamen, die Leistungen bei der Objekterkennung kaum schlechter wurden [44, 63].

Die Merkmale, die bei Positionsänderung eines Stimulus „verlorengehen“ können, sind wohlgemerkt die auf S2- bzw. C2-Ebene detektierten *komplexen* Merkmale, nicht die einfachen Balken, die von C1-Zellen bevorzugt erkannt werden. Werden nur vier statt 256 C2-Zellen verwendet, die auf dieselben orientierten Balken wie die C1-Zellen selektiv reagieren, so werden keine Modulationen der HMAX-Antwort mit der Position eines Stimulus beobachtet. Im Gegensatz zu einem komplexen Merkmal kann ein solches einfaches Merkmal, wie ein Blick auf Abbildung 3-3 zeigt, nicht durch das Raster der C1-Zellen fallen, solange deren rezeptive Felder nur um einen gewissen Betrag überlappen. Ein solches vereinfachtes Modell dürfte allerdings kaum zur Unterscheidung einer nennenswerten Zahl komplexer Objekte geeignet sein.

Auch im realen visuellen System wird das visuelle Feld diskretisiert, durch die Photorezeptoren in der Retina, und es ist davon auszugehen, daß auch im Gehirn eine hierarchische Verschaltung von Neuronen zum Aufbau komplexer Stimuluspräferenzen, ähnlich

der Prinzipien in HMAX, stattfindet. Beim natürlichen Sehen dürfte ein Verlust von Merkmalen hingegen kein Problem darstellen, da das visuelle Feld ohnehin nicht statisch und nur einmal abgetastet wird wie in HMAX, sondern durchschnittlich etwa dreimal pro Sekunde eine Sakkade ausgeführt wird und daher ein leicht verändertes Bild auf die Retina fällt [51]. Eventuelle „Verluste“ könnten daher leicht ausgeglichen werden. Ob allerdings im visuellen System dasselbe Phänomen eines Merkmalsverlusts auftritt wie in HMAX, kann nicht mit Sicherheit beantwortet werden, da die genaue Verschaltung der Neuronen und ihre Stimuluspräferenzen nicht im Detail bekannt sind. Ähnlich regelmäßige Schwankungen in der Antwort von Neuronen in den Arealen V4 und IT wie in den hier gezeigten Simulationen sind bisher nicht beobachtet worden, wenn ein Stimulus innerhalb ihres rezeptiven Feldes verschoben wurde. Zwar finden sich in der Regel Bereiche des rezeptiven Feldes, an denen derselbe Stimulus eine stärkere oder schwächere Antwort hervorruft als anderswo; diese Phänomene erscheinen allerdings komplexer als die in HMAX beobachteten [13]. Die Periodenlängen der *in vivo* beobachteten Schwankungen, die dem Parameter λ entsprechen würden, sind meist größer als die hier gefundenen, teils sogar größer als die Ausmaße des rezeptiven Feldes eines Neurons [47]. Dies würde in den Parametern des HMAX-Modells für große Werte von *poolRange* (viele S1-Afferenzen pro C1-Zelle) und / oder kleine Werte von *c1Overlap* (geringe Überlappung benachbarter C1-Zellen) sprechen (siehe Gleichung 3.1). Dabei würden aber, wie Abbildung 3-4 zeigt, die Schwankungen der neuronalen Aktivität und der Erkennungsleistung noch weiter zunehmen. Auch im Experiment werden tatsächlich deutliche Veränderungen der Aktivität in Abhängigkeit von der Position eines Stimulus gefunden [47]. Allerdings war von den hier untersuchten Neuronen nicht so gut wie im Fall der von Logothetis *et al.* [35] beschriebenen bekannt, welche Aufgabe sie vermutlich erfüllen. Unterschiedliche Neuronen können in Abhängigkeit von unterschiedlichen Lernprozessen und ihrer jeweiligen speziellen Funktion durchaus verschiedene Invarianzeigenschaften bezüglich der Stimulusposition zeigen. Hier könnten Experimente Klarheit schaffen, die das Training neuer Objekte durch eine zu lösende Aufgabe im Verhaltensexperiment mit dem genauen Kartieren der rezeptiven Felder von Neuronen verbinden, die sich als selektiv für die neu gelernten Objekte herausstellen.

Im visuellen System können Objekte allerdings auch dann gut erkannt werden, wenn sie nur so kurz im Blickfeld sind, daß keine Zeit für Sakkaden bleibt [51, 67]. Ein eventueller Verlust von Merkmalen und Änderungen in der Feuerrate entsprechender Neuronen

würden sich dann also zumindest nicht auf die Objekterkennung auswirken. Im Rahmen des HMAX-Modells könnte das dahingehend interpretiert werden, daß für die Objekterkennung wohl am ehesten solche visuellen Merkmale verwendet werden, die im Verhältnis zur diskreten Organisation der detektierenden Neuronen groß sind, so daß die Schwankungen bei veränderter Position gering sind (vgl. Abb. 3-5). Dagegen sprechen allerdings die großen rezeptiven Felder von Neuronen in Arealen wie V4 [13] sowie die Fähigkeit, auch deutlich kleinere Stimuli zu erkennen. Letztendlich wird man aus den hier präsentierten Resultaten auch und vor allem folgern müssen, daß Objekterkennung im realen visuellen System bei Translation eines Stimulus sehr viel robuster funktioniert als in HMAX.

Die Prinzipien des Merkmalsverlusts bei Translation von Objekten sind dieselben, egal welche Stimuli verwendet werden. Anders verhält es sich dagegen mit der Leistung bei der Skalierung von Stimuli. Es zeigt sich, daß HMAX Büroklammern unterschiedlicher Größe sehr gut erkennen kann, nicht jedoch Autos. Diese Diskrepanz ist auf die von HMAX detektierten komplexen Merkmale zurückzuführen. Eine gute Übereinstimmung der vom Modell detektierten Merkmale und der tatsächlich im zu erkennenden Stimulus vorhandenen Merkmale scheint entscheidend für größeninvariante Objekterkennung zu sein. Die gute Leistung von HMAX bei der Erkennung von Büroklammern unterschiedlicher Größe beruht darauf, daß die auf Ebene der S2- bzw. C2-Zellen erkannten komplexen Merkmale, also Kombinationen von 2×2 benachbarten orientierten Balken, den tatsächlichen Merkmalen von Büroklammern sehr nahekommen, und daß die Größe der rezeptiven Felder der von einer Büroklammer am stärksten aktivierten Merkmalsdetektoren sehr gut mit der Größe des Stimulus selbst übereinstimmt. Insbesondere letzteres ist entscheidend dafür, daß das Modell sich bei einer Veränderung der Größe eines Stimulus die verschiedenen Filtergrößen zunutze machen kann, um die Merkmale des Stimulus unabhängig von ihrer Größe korrekt zu erkennen. Die gute Übereinstimmung zwischen detektierten und tatsächlich vorhandenen Merkmalen sowie zwischen Größe des Merkmals und Größe der am stärksten aktivierten Filter ist für Autos nicht gegeben; daher ist die Leistung von HMAX bei der Erkennung skalierter Autostimuli deutlich geringer als für skalierte Büroklammern.

Obwohl also sowohl Translation als auch Skalierung affine Transformationen in der Bildebene sind und den Auswirkungen beider Transformationen in HMAX im Prinzip auf

dieselbe Weise Rechnung getragen wurde, nämlich durch auf unterschiedliche Positionen und Größen spezialisierte Afferenzen, die von ein und derselben nachgeschalteten Zelle gemäß der „MAX"-Operation verarbeitet werden, sind die Effekte der beiden Transformationen in der hierarchischen Verarbeitung unterschiedlich. Die vom System erreichte Größeninvarianz (also der Bereich, über den die Größe eines Stimulus ohne wesentliche Änderung seiner Erkennbarkeit variieren kann) hängt von den verwendeten Stimuli ab, die Positionsinvarianz hingegen nicht. Ein solcher Effekt könnte auch experimentell beobachtet werden, durch Training eines Versuchstiers mit verschiedenen neuen Objektklassen (ähnlich wie in [35]). Die Vorhersage des HMAX-Modells wäre dann, daß aufgrund der Tatsache, daß sehr wahrscheinlich keine komplett neuen einfachen und komplexen Merkmalsdetektor-Neuronen für die neuen Objektklassen angelegt werden, Neuronen etwa in IT für die verschiedenen Klassen ähnliche Translations-, aber unterschiedliche Größeninvarianzeigenschaften zeigen.

Im Unterschied zu den Effekten von Änderungen in Position und Größe eines Stimulus wurde den Auswirkungen von Kontraständerungen und von Hintergrundstimuli in der Architektur von HMAX nicht explizit Rechnung getragen. Die Resultate der entsprechenden Simulationen ergeben jedoch ein vergleichbares Bild wie für Änderungen der Stimulusgröße: Auch hier scheint die Übereinstimmung der detektierten Merkmale mit den tatsächlichen Merkmalen von Objekten entscheidend für gute Leistungen bei der Objekterkennung zu sein. Für korrekte Erkennung eines Stimulus auch bei niedrigeren Kontrasten und in Gegenwart eines Distraktors ist es erforderlich, daß seine Darstellung auf Ebene der Merkmalsdetektorneuronen (der C2-Zellen) hinreichend verschieden von der anderer Stimuli ist, damit auch bei einer generell geringeren Aktivität aller Zellen aufgrund geringeren Kontrasts der „Abstand" der einzelnen Stimuli im Merkmalsraum noch groß genug ist und sie für nachgeschaltete Neuronen noch unterscheidbar sind. Wenn verschiedene Stimuli eher verschiedene C2-Zellen bevorzugt erregen, wird auch ein Distraktorstimulus die neuronale Repräsentation eines zu erkennenden Zielobjekts nur begrenzt stören, und solange nur die bei Präsentation eines Stimulus am stärksten aktiven C2-Zellen als Afferenzen für eine VTU verwendet werden, kann ein Zielobjekt auch weiterhin gut erkannt werden. Die gute Übereinstimmung zwischen detektierten und tatsächlich vorhandenen Stimulusmerkmalen, die zu robusten und gut unterscheidbaren neuronalen Repräsentationen von Stimuli führt, ist aber wiederum bei den getesteten Stimulusklassen nur für die

Büroklammern in ausreichendem Maße gegeben, nicht jedoch für Autos.

Die hervorragenden Leistungen des visuellen Systems bei der Erkennung nicht nur von translatierten und skalierten Stimuli, sondern auch für wechselnde Kontraste und störende Hintergrundstimuli [1, 63], deuten also unter anderem darauf hin, daß das Gehirn in den verschiedenen Arealen des ventralen visuellen Systems wohl komplexe Merkmale aus dem visuellen Input extrahiert, die besser auf die tatsächlich wahrgenommenen Objekte abgestimmt sind, als das von den durch die S2- bzw. C2-Zellen in HMAX detektierten Merkmalen behauptet werden kann. Die tatsächlichen bevorzugten Stimuli von Neuronen in solchen höheren Arealen des visuellen Systems zu bestimmen ist freilich ein sehr schwieriges Unterfangen, wie eine andere Studie mit dem HMAX-Modell, angewendet auf experimentelle Daten, überzeugend gezeigt hat [30, 70]. Ein vielversprechender Ansatz könnte sein, zu versuchen, möglicherweise geeignete komplexe Merkmale aus statistischen Regelmäßigkeiten üblicher visueller Stimuli aus der natürlichen Umgebung abzuleiten. Ein solcher Prozess des Lernens komplexer Merkmale ist auch im visuellen System wahrscheinlich, eventuell in einer früheren Entwicklungsphase. Untersuchungen mit dem HMAX-Modell zeigen, daß durch Lernalgorithmen gefundene Merkmale, auf Ebene der S2- bzw. C2-Zellen eingesetzt, die Leistung des Modells bei der Gesichtserkennung deutlich verbessern [62]. So gefundene Merkmale könnten auch als Stimuli im physiologischen Experiment eingesetzt werden, um herauszufinden, ob sie in der Tat stärkere Erregungen bei den Neuronen in Arealen wie V4 hervorrufen als üblicherweise verwendete Stimuli.

Andererseits kann aus den suboptimalen Leistungen des HMAX-Modells bei der Erkennung insbesondere von Autostimuli unter verschiedenen Bedingungen gefolgert werden, daß Verbesserungen der Leistung erwartet werden können, wenn Aufmerksamkeit auf bestimmte Objekte eingeführt wird. Insbesondere wenn die Objekterkennung etwa durch niedrigen Kontrast oder Hintergrundstimuli erschwert wird, sollten Aufmerksamkeitseffekte unterstützend eingreifen können. Dies wird in Kapitel 5 näher untersucht, gerade auch für verschiedene Kontraste und Distraktorstimuli. Dabei werden die beiden Stimulusklassen, Büroklammern und Autos, insofern von Nutzen sein, als sie zwei für das verwendete System unterschiedlich gut unterscheidbare Objektklassen repräsentieren können. Es wird interessant sein herauszufinden, ob und wie sich Aufmerksamkeitseffekte unterschiedlich auf solche Objektklassen auswirken können.

Kapitel 4

Interaktion von Stimuli: Das „Biased Competition Model“ und HMAX

Für ein Verständnis der Rolle von Aufmerksamkeit bei der Objekterkennung in komplexen visuellen Szenen mit mehreren Stimuli ist es unabdingbar, zunächst die Interaktion von gleichzeitig präsentierten Stimuli auf neuronaler Ebene ohne Einfluß von Aufmerksamkeit zu untersuchen. Dies wurde zum Teil schon im vorigen Kapitel getan (3.2.4). In der experimentellen Literatur ist aber ein detailliertes und mit zahlreichen Daten untermauertes Modell der Interaktion von visuellen Stimuli im Cortex beschrieben, das sogenannte „Biased Competition Model“ von Desimone *et al.* [12, 53], das eine nähere Betrachtung verdient. Insbesondere wird die Frage zu klären sein, inwieweit HMAX die im Experiment beobachteten Daten im Vergleich zu diesem Modell erklären kann.

Das Biased Competition Model beruht auf den beiden folgenden Annahmen: (1) Wenn mehrere Stimuli zusammen gezeigt werden, aktivieren sie neuronale Populationen, die miteinander konkurrieren. (2) Aufmerksamkeit begünstigt in dieser Konkurrenz diejenige neuronale Population, die vom bewußt beachteten Stimulus aktiviert wird. [53] Eine der Folgerungen des Modells ist, daß ein Neuron auf das Zeigen zweier Stimuli, von denen einer das Neuron stark, der andere schwach aktiviert, mit einer Feuerrate antworten sollte, die *zwischen* den Antwortstärken des Neurons für die beiden isoliert präsentierten Stimuli liegt, weil seine Antwort auf den „guten“ Stimulus durch den Einfluß anderer, für den „schlechten“ Stimulus selektiver Neuronen, abgeschwächt wird.

Effekte in Übereinstimmung mit dem Biased Competition Model, sowohl ohne als auch mit Einfluß von Aufmerksamkeit, wurden in verschiedenen experimentellen Paradigmen und visuellen corticalen Arealen gefunden, von V2 bis IT [7, 8, 53]. Dabei waren die Effekte am stärksten, wenn die beiden Stimuli innerhalb des rezeptiven Feldes des betrachteten Neurons präsentiert wurden; es ist jedoch, trotz teilweise widersprüchlicher Daten, davon auszugehen, daß auch Stimuli jenseits des klassischen rezeptiven Feldes im Prinzip die Antwort eines Neurons auf einen Stimulus innerhalb seines rezeptiven Feldes beeinflussen können [46, 69]. Für die Präsentation zweier Stimuli innerhalb der rezeptiven Felder von V4-Neuronen gibt es jedoch auch Daten, die darauf hindeuten, daß Neuronen in diesem Areal auch eine MAX-Operation realisieren könnten. Gawne und Martin verwendeten statt Balkenstimuli (oder auch anstelle von komplexen Bildern wie in [7, 8]) schwarz-weiße schachbrettartige Muster, die innerhalb des rezeptiven Feldes soweit als möglich voneinander entfernt waren, und erhielten Resultate, die gut mit den Vorhersagen einer MAX-Operation übereinstimmten, d.h. viele Neuronen antworteten auf die Kombination zweier Stimuli nahezu so wie auf den Stimulus, der isoliert eine stärkere Antwort hervorrief, alleine [16].

In HMAX ist keine explizite Konkurrenz zwischen neuronalen Subpopulationen realisiert. Es wäre daher von großem Interesse, herauszufinden, ob die durch das Biased Competition Model beschriebenen Effekte dennoch auch in HMAX beobachtet werden können. In diesem Fall könnten die divergierenden experimentellen Daten eventuell vereinbart werden, ohne eigens konkurrierende Interaktionen zwischen Neuronen postulieren zu müssen. Ein Ansatz dafür könnte sein, daß für zwei nahe beieinander präsentierte Stimuli die Antwort eines Neurons, das eigentlich seinen afferenten Input mit einer MAX-Operation verarbeitet, zwischen den Feuerraten für die einzelnen Stimuli liegen könnte, wie vom Biased Competition Model vorhergesagt – etwa aufgrund von Interaktionen auf der Ebene der Afferenzen dieses Neurons und deren inhibitorischer Subfelder. Für größere Entfernungen, wie sie explizit von Gawne und Martin gewählt wurden, könnte dagegen die MAX-Operation die Antwort des Neurons bestimmen.

In diesem Kapitel werden Simulationen beschrieben, die das Experiment von Reynolds, Chelazzi und Desimone [53] bezüglich der Interaktion zweier Balkenstimuli im rezeptiven Feld von V4-Neuronen in HMAX modellieren. Dabei wurde untersucht, wie sich die

Antwort des Modells in Abhängigkeit vom Abstand der beiden Balken verhielt. Da die zu klärende Frage sich nur auf die Art der Interaktion der beiden Stimuli bezog, wurden hier noch keine Aufmerksamkeitseffekte modelliert.

4.1 Methoden

4.1.1 Simulationen

In Analogie zum Experiment von Reynolds *et al.* [53] wurden Balkenstimuli verschiedener Orientierung (0°, 45°, 90°, 135°) verwendet. Da HMAX keine Farben unterscheidet, wurde auf Balken verschiedener Farben verzichtet; der Grauwert der Balken wurde stets auf 140 gesetzt, der des Hintergrunds auf 0. Jeder Punkt im von HMAX verarbeiteten Gesamtbild (hier 128×128 Pixel) kann die Antwort jeder C2-Zelle beeinflussen. Da die C2-Zellen Modelle für Neuronen im Areal V4 sind, entspricht die Größe des präsentierten Bildes im Modell dem rezeptiven Feld eines V4-Neurons. Rezeptive Felder in V4 variieren in der Größe mit der Exzentrizität ihrer Position im visuellen Feld [13]; für die Simulationen wurde eine durchschnittliche Größe von $4.4° \times 4.4°$ angenommen [31]. Die im Experiment verwendeten Balken der Größe $1° \times 0.25°$ entsprechen im Modell also in etwa Balken der Ausmaße 29×7 Pixel. Zusätzlich wurden auch Resultate für Balken von 13×3 Pixel Größe berechnet. Zwei Balken wurden jeweils auf der Bilddiagonalen von links oben nach rechts unten gezeigt, in Abständen von 8 bis 42 Pixel (29×7- Balken) bzw. 4 bis 38 Pixel (13×3-Balken) in Schritten von 1 Pixel. (Es wurden auch größere Abstände verwendet, die aber für die Auswertung keine neuen Erkenntnisse lieferten.) Ein Abstandswert in Pixel bezeichnet dabei die Distanz, um die der zweite Balken gegenüber dem ersten sowohl nach rechts als auch nach unten verschoben war, also die Länge der beiden Katheten des rechtwinkligen Dreiecks, dessen Hypotenuse die Verbindungslinie der Mittelpunkte der beiden Balken war. Die Position des linken oberen Balkens wurde konstant gehalten. Zu jeder Kombination zweier orientierter Balken an einer Position wurden auch die Antworten auf die beiden Balken einzeln gemessen, um die Interaktionen bei Präsentation beider zugleich mit den Antworten auf die einzelnen Stimuli vergleichen zu können. Es wurde darauf geachtet, daß zwei Balken in keinem Fall überlappend gezeigt wurden; wenn der Abstand zweier Balken so gering war, daß einige Orientierungskombinationen eine überlappende

Darstellung ergeben hätten, wurden diese Kombinationen weggelassen. (Das heißt allerdings, daß bei geringen Abständen insgesamt weniger Daten ausgewertet werden konnten als bei größeren.)

Es wurden auch Simulationen mit einem modifizierten HMAX-Modell durchgeführt, das nur über vier C2-Zellen bzw. vier Typen von S2-Zellen verfügte, die dieselbe Stimuluspräferenz hatten wie S1- bzw. C1-Zellen, also einfache orientierte Balken (siehe Ergebnisse). Dazu projizierte statt 2×2 benachbarten C1-Zellen lediglich *eine* auf eine S2-Zelle, die aber nach wie vor gemäß einer Gaußfunktion auf ihren afferenten Input reagierte (siehe Kapitel 2). In diesen Simulationen wurden in der Regel nur solche Abstände zweier Balken untersucht, die sich in den vorigen Versuchen als interessant für die nähere Untersuchung herausgestellt hatten, und zwar Abstände von 4 bis 9 Pixeln in Schritten von 1 Pixel entlang der Bilddiagonalen oder von 6 bis 15 Pixeln in Schritten von 3 Pixeln entlang der horizontalen Mittellinie des Bildes (im letzteren Fall bezeichnen die Abstände also unmittelbar die Länge der Verbindungslinie zwischen den Mittelpunkten der Balken). Diese Simulationen wurden auch mit verschiedenen Werten der Parameter *s2Sigma* und *s2Target* durchgeführt. *s2Target* bezeichnet dabei den Mittelwert und *s2Sigma* die Standardabweichung der Gaußschen Funktion, die die Antwort der S2- (und damit auch C2-) Zellen auf ihre C1-Afferenzen bestimmt. Beide Werte sind in HMAX normalerweise auf 1 gesetzt. Niedrigere Werte für *s2Target* ergeben schon für schwächeren C1-Input höhere Feuerraten der S2- bzw. C2-Zellen, und kleinere Werte für *s2Sigma* erhöhen die Abstimmschärfe der S2-Zellen auf einen bestimmten C1-Input, d.h. ihre Feuerrate fällt schon für geringere Abweichungen von ihrem bevorzugten Stimulus stärker ab.

4.1.2 Auswertung

Die Auswertung erfolgte in Analogie zu den Methoden von Reynolds *et al.* [53]. Von den beiden Balkenstimuli einer Kombination wurde einer (der linke obere) als „Referenzstimulus“, der andere als „Teststimulus“ bezeichnet. Für jede C2-Zelle und jede Kombination zweier Balken in jedem berechneten Abstand wurden aus den Feuerraten der Zelle bei Präsentation der einzelnen Balken und bei der Kombination der beiden Balken der „Selektivitätsindex“ SE und die „Sensorische Interaktion“ SI berechnet. SE ergibt sich dabei aus der Differenz der Antworten der Zelle auf Test- und Referenzstimulus, SI aus der Differenz

der Antworten auf Kombination und Referenzstimulus, d.h.

$$SE = \textit{Test} - \textit{Referenz} \tag{4.1}$$

$$SI = \textit{Kombination} - \textit{Referenz} \tag{4.2}$$

Die experimentell bestätigte Hypothese von Reynolds *et al.* im Rahmen des Biased Competition Model war dabei, daß Zellen, die auf den Teststimulus stärker antworten als auf den Referenzstimulus (d.h. $SE > 0$), bei Präsentation der Kombination beider Balken eine Feuerrate an den Tag legen müßten, deren Betrag ein gewichtetes Mittel aus den Antwortstärken bei Präsentation der Einzelstimuli ist, also größer als der Betrag der Antwort auf den Referenzstimulus (d.h. $SI > 0$). Analog wurde für $SE < 0$ auch $SI < 0$ erwartet, weil in diesem Fall das Hinzufügen eines von einer Zelle weniger bevorzugten Teststimulus zu einem eine starke Antwort hervorrufenden Referenzstimulus die Feuerrate senken sollte. Die für eine Zelle und einen Referenzstimulus gefundenen $(SE|SI)$-Wertepaare sollten demzufolge gut durch einen linearen Zusammenhang zu beschreiben sein, was sich bei Reynolds *et al.* durch Berechnung von Regressionsgeraden bestätigte.

Aufgrund der Tatsache, daß in den hier vorgestellten Simulationen keine Balken unterschiedlicher Farbe verwendet werden konnten und bei geringen Abständen wegen Überlappung auch nicht alle prinzipiell möglichen 16 Kombinationen zweier orientierter Balken gezeigt werden konnten, waren in einigen Fällen zuwenig Daten vorhanden, um für jede Zelle und jeden Referenzstimulus eine eigene Regressionsgerade zu berechnen. Eine eventuelle Linearität des Zusammenhangs zwischen SE und SI würde aber durch Mittelung über mehrere Referenzstimuli und / oder Zellen nicht verlorengehen (eine lineare Kombination linearer Funktionen ergibt stets wieder eine lineare Funktion), lediglich die Parameter (Steigung und y-Abschnitt) würden sich ändern. Dementsprechend haben auch Reynolds *et al.* in ihrem Experiment 2, wo ebenfalls nicht alle möglichen Kombinationen zweier Balken für jede Zelle getestet wurden, über mehrere Zellen und Referenzstimuli gemittelt und wiederum einen linearen Zusammenhang gefunden. Da in den hier beschriebenen Simulationen nicht die genauen Parameter einer Regressionsgeraden interessieren, sondern lediglich, ob und wie gut der gefundene Zusammenhang linear approximiert werden kann, wurden hier ebenfalls für jeden Abstand zweier Balken nur *ein* SE–SI–Diagramm und *eine* Regressionsgerade für alle Zellen und Referenzstimuli berech-

net.

Die Vorhersage des Biased Competition Model für das SE–SI–Diagramm ist, daß es sich gut durch eine Gerade mit einer Steigung kleiner als 1 approximieren lassen müßte. Wie Reynolds *et al.* zeigen, entspricht die Steigung einer solchen Geraden einem Maß für den Einfluß des Teststimulus auf die Aktivität der Zelle bei der Präsentation der Kombination der beiden Stimuli; aus der Geradengleichung

$$SI = w \cdot SE + \textit{offset} \tag{4.3}$$

mit w als Steigung und *offset* als y-Achsenabschnitt ergibt sich nämlich mit Gleichung 4.1

$$\textit{Kombination} = w \cdot \textit{Test} + (1 - w) \cdot \textit{Referenz} + \textit{offset} \tag{4.4}$$

Damit, wie vom Biased Competition Model vorhergesagt, *beide* Stimuli die Antwort der Zelle bei gleichzeitiger Präsentation beeinflussen und die Antwort auf die Kombination ein gewichtetes Mittel aus den Antworten auf die Einzelstimuli ist, muß $w < 1$ gelten.

Eine Zelle, die eine MAX-Operation durchführt, wird in ihrer Antwort hingegen jeweils nur durch den alleine eine stärkere Antwort hervorrufenden Stimulus bestimmt. D.h. für $SE < 0$ (Antwort auf den Referenzstimulus stärker als auf den Teststimulus) erwartet man $SI = 0$ (Antwort auf die Kombination genauso wie auf den Referenzstimulus alleine), und für $SE > 0$ sollte $SI = SE$ sein (da die Antwort auf die Kombination dann der Antwort auf den Teststimulus alleine entspricht, vgl. Gleichung 4.1). Es würde sich also im SE–SI–-Diagramm ein Graph wie in Abbildung 4-1 d ergeben: die Nullinie für $SE < 0$ und eine Gerade durch den Nullpunkt mit Steigung 1 für $SE > 0$. Die entscheidende Frage für die Simulationen in diesem Kapitel ist also, ob die Antworten der Zellen in HMAX, obwohl sie eine MAX-Operation über ihre Afferenzen durchführen, bei bestimmten Abständen zweier Balken auch ein SE–SI–Diagramm wie von Reynolds *et al.* beobachtet ergeben können.

Neben SE–SI–Diagrammen für ausgewählte Abstände wird in den Ergebnissen auch die Güte des durch Regression errechneten linearen Fits an die Punkte in den SE–SI–Diagrammen für alle berechneten Abstände gezeigt, in Form des Werts der R^2-Statistik in Abhängigkeit vom Balkenabstand. Die R^2-Statistik wird für jeden Balkenabstand zunächst für alle C2-Zellen einzeln berechnet, aus den Daten aller beim jeweiligen Abstand verfügba-

ren Balkenkombinationen, und dann über alle Zellen gemittelt. Sie gibt an, welcher Anteil an der beobachteten Varianz der Datenpunkte durch die Regressionsgerade erklärt werden kann. Ein Wert von 0.6 etwa zeigt, daß die berechneten Regressionsgeraden für einen gegebenen Balkenabstand im Mittel über alle Zellen 60% der Variation der Datenpunkte der SE–SI–Diagramme für diesen Abstand zweier Balken erklären können.

Ein weiterer Wert, der zur Beurteilung der Resultate verwendet wurde, war der sogenannte Sato-Index $k(d)$ in Abhängigkeit vom Balkenabstand d [61]. Der Sato-Index berechnet sich wie folgt:

$$k(d) = \frac{R_d(a+b) - max[R(a), R(b)]}{min[R(a), R(b)]} \tag{4.5}$$

wobei $R_d(a+b)$ die Antwort R einer Zelle auf die Kombination der Balken a und b im Abstand d bezeichnet, $R(a)$ bzw. $R(b)$ die Antworten der Zelle auf die Balken a bzw. b alleine und max bzw. min das Maximum bzw. Minimum. Eine Zelle, die eine Maximumsoperation durchführt, wird einen Sato-Index von oder nahe bei 0 aufweisen. Sato-Werte größer 0 deuten auf eine lineare Summation hin (d.h. die Antwort auf die Kombination ist stärker als auf die Einzelstimuli), während Werte kleiner 0 vom Biased Competition Model vorhergesagt würden (d.h. die Antwort auf die Kombination ist ein gewichtetes Mittel der Antworten auf die Einzelstimuli). Der Sato-Index wurde ebenfalls für einen gegebenen Abstand zweier Balken über alle Zellen und Stimuluskombinationen gemittelt.

4.2 Ergebnisse

Abbildung 4-1 zeigt typische Resultate für die Simulationen mit zwei Balken. Es ist ersichtlich, daß in der Tat die Interaktion zweier Balkenstimuli auf neuronaler Ebene in HMAX vom Abstand der beiden Balken abhängt. Während für große Abstände eindeutig die MAX-Operation dominiert (d), deuten die Daten für geringere Abstände auf eine lineare Summation hin (Sato-Index größer als 0, Punkte oberhalb der MAX-Kurve im SE–-SI–Plot) und für sehr kleine Abstände auf suppressive Interaktionen, d.h. auf eine Antwort von Zellen auf eine Kombination zweier Balken, die betragsmäßig zwischen den Antworten auf die Einzelstimuli liegt, etwa wie es vom Biased Competition Model vorhergesagt wird (Sato-Index kleiner 0, Punkte unterhalb der MAX-Kurve im SE–SI–Plot). Lineare Summation tritt auf, wenn zwei Balken nahe genug zusammen gezeigt werden,

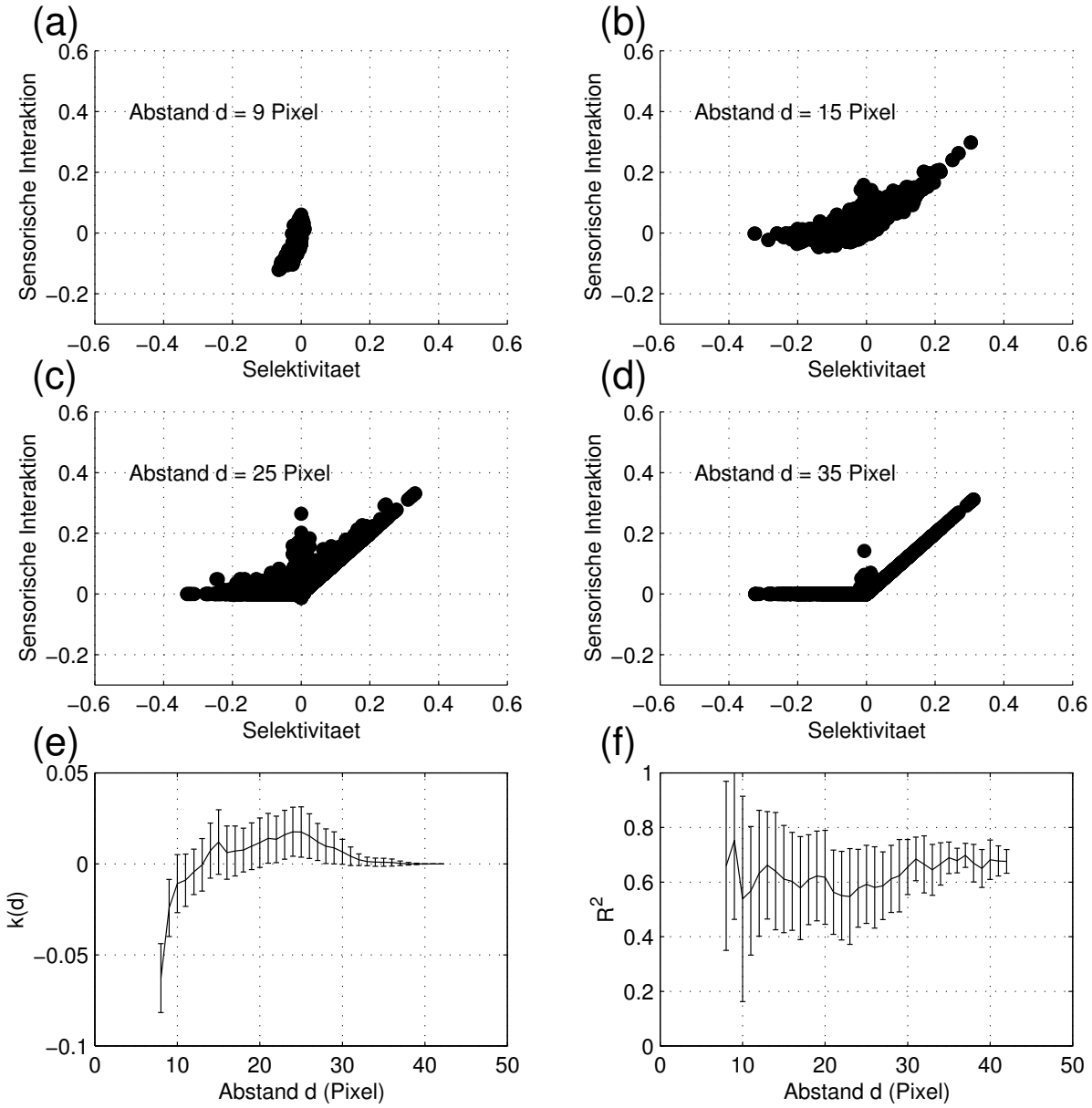

Abbildung 4-1: Interaktion zweier Balkenstimuli (29×7 Pixel). **(a)** – **(d)** SE–SI–Diagramme für alle C2-Zellen und alle verfügbaren Balkenkombinationen bei zunehmenden Abständen zweier Balken. Jeder Punkt entspricht einem $(SE|SI)$-Wertepaar für ein Neuron und eine Kombination zweier Balken beim angegebenen Abstand. Die Punkte in (d) liegen fast vollständig auf der theoretischen Kurve für ein MAX-Neuron. Punkte oberhalb dieser Kurve zeigen lineare Summation, Punkte darunter gewichtete Mittelung von afferenten Stimuli durch die betreffenden Neuronen. **(e)** Sato-Index $k(d)$ für zwei Balkenstimuli in unterschiedlichen Abständen, gemittelt über alle C2-Zellen und Balkenkombinationen bei gegebenem Abstand. Summation: positive Werte; gewichtete Mittelung: negative Werte; MAX-Operation: 0. Fehlerbalken geben die Standardabweichung an. **(f)** R^2-Statistik für die Güte des linearen Regressionsfits an die SE–SI–Diagramme für verschiedene Abstände und einzelne C2-Zellen, gemittelt über alle Zellen. Fehlerbalken geben die Standardabweichung an.

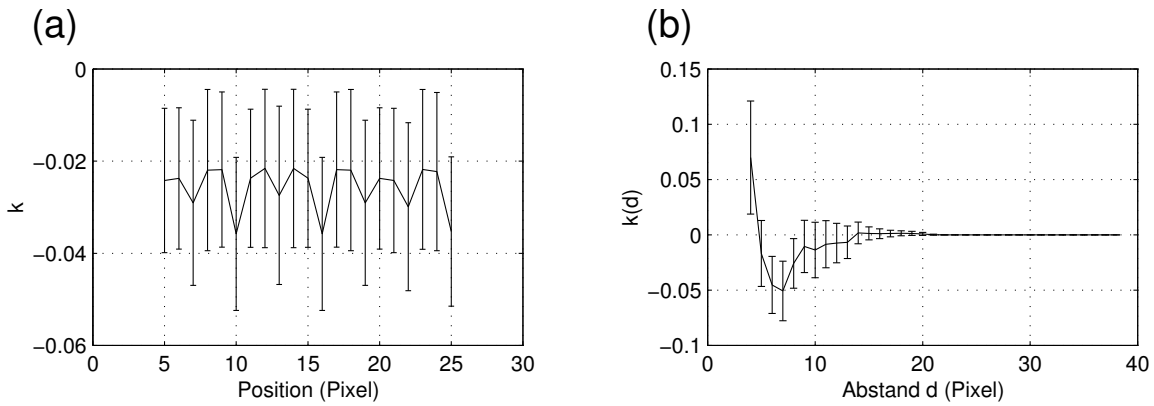

Abbildung 4-2: **(a)** Schwankungen des Sato-Index k bei Verschiebung zweier Balken (29×7 Pixel) in konstantem Abstand. Abszissenwerte geben die horizontale Position des linken von zwei Balken im visuellen Feld an. Ein zweiter Balken wurde stets im Abstand 9 Pixel präsentiert. Die angegebenen Werte sind über alle C2-Zellen und bei diesem Abstand möglichen Kombinationen zweier Balkenorientierungen gemittelt (Fehlerbalken: Standardabweichung). **(b)** Sato-Index $k(d)$ für zwei Balkenstimuli der Größe 13×3 Pixel in unterschiedlichen Abständen, gemittelt über alle C2-Zellen und Kombinationen. Fehlerbalken zeigen die Standardabweichung.

um innerhalb des rezeptiven Feldes einer S2-Zelle zu fallen. Da S2-Zellen die Antworten von vier S1-Zellen im Exponenten ihrer Gaußschen Antwortfunktion summieren, wird in einer solchen Situation in der Regel die Antwort auf zwei Balken stärker sein als auf einen alleine; daraus ergeben sich die positiven Werte für den Sato-Index bzw. die über der MAX-Kurve liegenden Punkte im SE–SI–Diagramm. Wenn zwei Balken hingegen so nahe zusammen präsentiert werden, daß sie ganz oder teilweise in das rezeptive Feld derselben S1-Zelle fallen, kann die Antwort auf zwei Balken zugleich auch kleiner sein als die Antwort auf den von der Zelle eher bevorzugten Balken alleine, wenn nämlich der zweite Balken in einem inhibitorischen Subfeld der Zelle zu liegen kommt (vgl. Abb. 2-2). Daraus ergeben sich dann negative Werte für den Sato-Index bzw. Punkte unter der MAX-Kurve im SE–SI–Diagramm.

Ein konsistenter Trend des Modells, bei geringeren Abständen ein den Ergebnissen von Reynolds *et al.* entsprechendes Verhalten zu zeigen, ist allerdings nicht auszumachen. Zum einen ist, wie in den Methoden erwähnt, die Datenlage für sehr geringe Abstände relativ schlecht, da aufgrund von Überlappungen nicht alle möglichen Balkenkombinationen gezeigt werden konnten. (Zwei überlappende Balken können als neuer Stimulus anstatt als Kombination zweier Stimuli interpretiert werden und wurden daher nicht verwendet.)

Weiterhin sind die Beträge des Sato-Index durchweg relativ klein. Die negativen Werte in Abbildung 4-1 e sind zwar für die ersten fünf Abstände signifikant kleiner als 0 (einseitiger t-Test, $\alpha = 0.05$) und entsprechen also der Vorhersage des Biased Competition Model. Allerdings zeigt Abbildung 4-2 a, daß der Sato-Index alleine aufgrund der in Kapitel 3 diskutierten Schwankungen der Antwort des Modells bei Verschiebung eines Stimulus im Mittel bereits um mehr als 0.01 variieren kann. Für diese Abbildung wurden alle nicht überlappenden Balkenkombinationen in einem konstanten Abstand, aber an verschiedenen Positionen im visuellen Feld gezeigt. Abbildung 4-2 b macht außerdem deutlich, daß für Balken der Größe 13×3 Pixel bei sehr geringen Abständen auch wieder positive Werte des Sato-Index auftreten können. Diese können etwa durch Zellen zustande kommen, deren bevorzugte Orientierung zur Orientierung zweier präsentierter paralleler Balken orthogonal ist. Zwei solche Balken ergeben bei den in HMAX verwendeten S1-Filtern insgesamt mehr excitatorische Stimulation als einer, obwohl sie zum Teil auch im inhibitorischen Subfeld liegen, und daher einen positiven Wert des Sato-Indexes.

Die SE–SI–Diagramme zeigen des weiteren bei Zunahme des Abstands der beiden Balken in der Regel eine rasche Herausbildung einer in etwa der MAX-Operation entsprechenden Kurve. Bei den geringsten Abständen häufen sich die Punkte um einen Selektivitätsindex von 0, weil die häufigsten bei diesen Distanzen noch möglichen Kombinationen aus zwei parallelen Balken bestehen und daher Test- und Referenzstimulus abgesehen von der Position identisch sind. Bei etwas größeren Abständen zeigt schon die visuelle Inspektion in den allermeisten Fällen einen deutlichen Mangel an Punkten im dritten Quadranten ($SE < 0$, $SI < 0$) sowie im ersten Quadranten eine Steigung, die nicht sichtbar kleiner als 1 ist. Beides wäre essentiell für dem Biased Competition Model entsprechende Resultate. Die Auswertung mittels der R^2-Statistik der Regressionsgeraden an die SE–SI–Diagramme bei den verschiedenen Balkenabständen (Abb. 4-1 f) zeigt dann, daß auch bei geringen Abständen die Daten nicht besser durch eine Gerade approximiert werden können als bei solchen Abständen, wo im Diagramm schon eindeutig die MAX-Kurve erkennbar ist (die beste lineare Approximation wurde für den Abstand 9 Pixel erreicht; das zugehörige Diagramm ist in Abbildung 4-1 a gezeigt und entspricht allein aufgrund der großen Steigung ebenfalls nicht den Vorhersagen des Biased Competition Model).

Obwohl also der Abstand zweier Stimuli in HMAX die Art der Interaktion der beiden

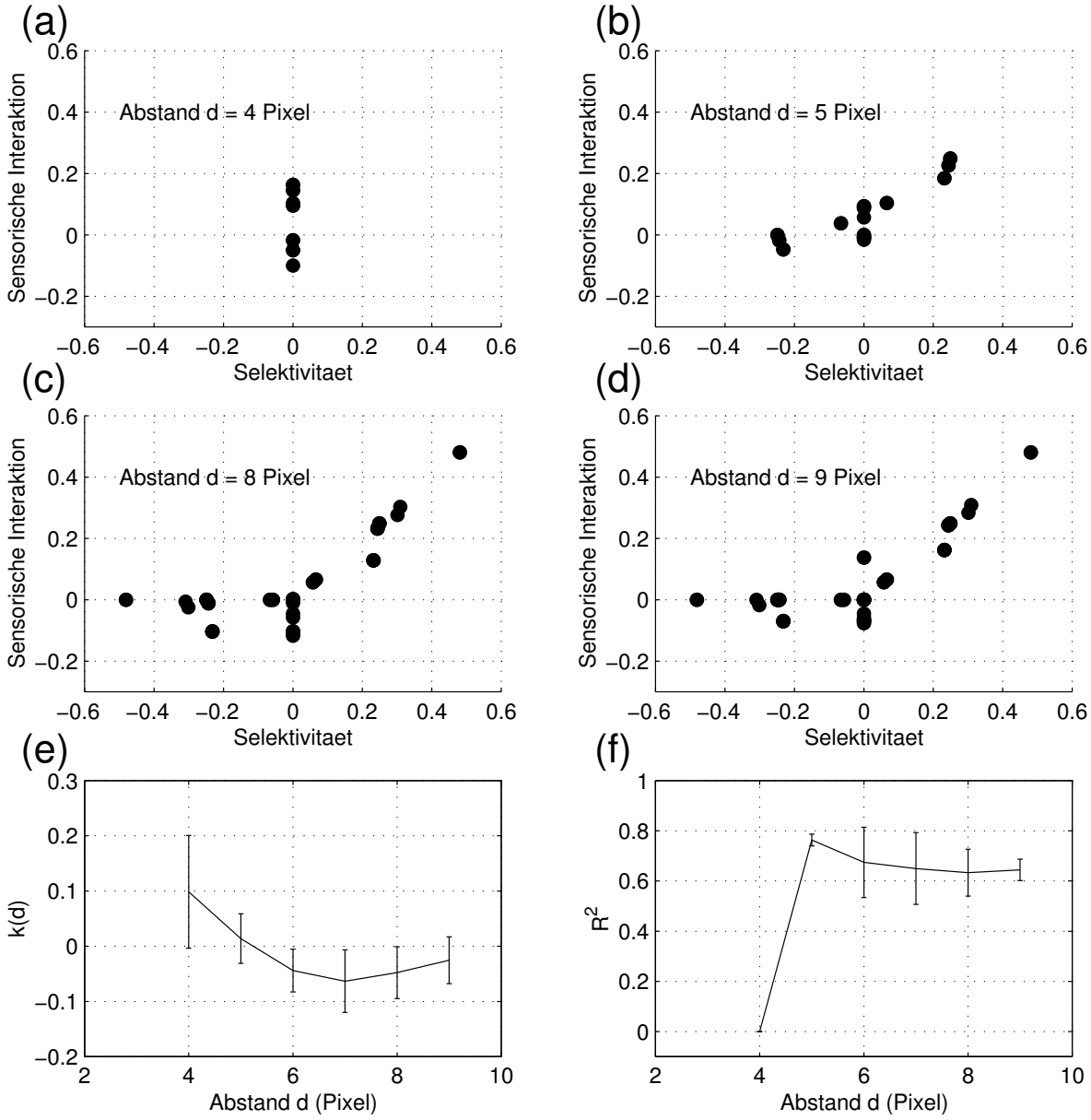

Abbildung 4-3: Interaktion zweier Balkenstimuli (13×3 Pixel) in HMAX mit vier „einfachen“ C2-Zellen. Vgl. Abb. 4-1. **(a)** – **(d)** SE–SI–Diagramme für alle C2-Zellen und alle verfügbaren Balkenkombinationen bei zunehmenden Abständen zweier Balken. **(e)** Sato-Index $k(d)$ für zwei Balkenstimuli in unterschiedlichen Abständen, gemittelt über alle C2-Zellen und Balkenkombinationen bei gegebenem Abstand. Fehlerbalken geben die Standardabweichung an. **(f)** R^2-Statistik für die Güte des linearen Regressionsfits an die SE–SI–Diagramme für verschiedene Abstände und einzelne C2-Zellen, gemittelt über alle Zellen. Fehlerbalken geben die Standardabweichung an.

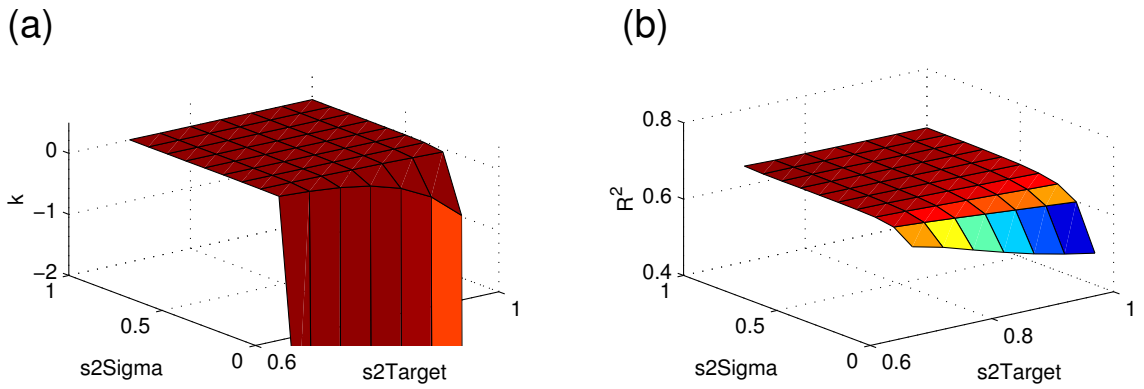

Abbildung 4-4: Sato-Index und R^2-Statistik für zwei Balkenstimuli (13×3 Pixel) im horizontalen Abstand 9 Pixel und HMAX mit vier „einfachen“ C2-Zellen, gemittelt über alle C2-Zellen und Balkenkombinationen, für verschiedene Werte der Parameter *s2Target* und *s2Sigma*. **(a)** Sato-Index k, **(b)** Wert der R^2-Statistik.

auf neuronaler Ebene beeinflußt und im Prinzip Werte wie die von Reynolds *et al.* beschriebenen gefunden werden können, erscheinen die Ergebnisse nicht konsistent und systematisch genug, um behaupten zu koennen, daß Neuronen, die eine MAX-Operation durchführen, alleine durch Verringerung des Abstands zwischen zwei Stimuli ein dem Biased Competition Model entsprechendes Verhalten zeigen könnten. Eine mögliche Ursache für die Differenz könnten unterschiedliche Stimuluspräferenzen bei den Neuronen in HMAX und im Experiment sein. Die Summation von vier C1-Zellen auf Ebene der S2-Zellen etwa begünstigt grundsätzlich einen Stimulus aus zwei (oder mehr) Balken gegenüber einem einzelnen. Reynolds *et al.* hingegen haben Neuronen abgeleitet, die bei der Suche nach Zellen und dem Ausmessen des rezeptiven Feldes gut auf einzelne Balkenstimuli antworteten. Es könnte sich also um Zellen gehandelt haben, deren Stimuluspräferenz besser durch einen einzelnen Balken als durch eine komplexere Kombination von Stimuli beschrieben werden kann. Um die Effekte einer solchen Stimuluspräferenz zu untersuchen, wurde ein modifiziertes HMAX-Modell mit nurmehr vier C2-Zellen bzw. vier Typen von S2-Zellen, die dieselbe Stimuluspräferenz wie C1-Zellen hatten (einzelne Balken in vier Orientierungen), verwendet. Um außerdem die Differenz zwischen den Antworten einer S2- bzw. C2-Zelle auf ihren bevorzugten und auf andere Stimuli zu erhöhen, d.h. die Abstimmung schärfer und spezifischer zu machen, wurden systematisch die Auswirkungen unterschiedlicher Werte für die Parameter *s2Target* und *s2Sigma* untersucht (siehe Methoden) sowie Balken der Größe 13×3 Pixel verwendet, bei denen in der Regel stärke-

re Modulationen der Antwort der C2-Zellen bei verschiedenen Orientierungen festgestellt wurden.

Beispielhafte Ergebnisse für HMAX mit nur vier „einfachen“ C2-Zellen sowie *s2Target* $= 0.7$ und *s2Sigma* $= 0.3$ finden sich in Abbildung 4-3. Auch hier ergibt sich jedoch kein anderes Bild als im Standard-HMAX-Modell. Für die Daten, die am besten durch eine Regressionsgerade erklärt werden können (für den Balkenabstand 5 Pixel, Abbildung 4-3 b, f), ist der Sato-Index eher positiv als negativ (im Widerspruch zum Biased Competition Model, Abb. 4-3 e), und ansonsten dominiert zumeist wieder die MAX-Operation die Antwort der Zellen (sichtbar am MAX-ähnlichen Kurvenverlauf in Abbildung 4-3 c, d). Die systematische Untersuchung der Effekte verschiedener Werte der Parameter *s2Target* und *s2Sigma* (in Abb. 4-4 für den horizontalen Balkenabstand 9 Pixel) ergibt keine Hinweise auf Einstellungen, für die HMAX eindeutig den Vorhersagen des Biased Competition Model entspricht. Für sehr kleine Werte von *s2Sigma* sind zwar teils sogar sehr stark negative Werte des Sato-Index möglich, allerdings ist die Aktivität der C2-Zellen dann auch so extrem gering, daß diese Resultate eher als Artefakte gewertet werden müssen. Zudem wurde im Mittel nirgends ein Wert der R^2-Statistik signifikant über 0.7 (dem Wert, der auch für die reine MAX-Kurve erreicht wird) gefunden.

4.3 Diskussion

Die Resultate dieses Kapitels machen deutlich, daß die unterschiedlichen experimentellen Ergebnisse über das Verhalten von V4-Neuronen bei Präsentation zweier Stimuli von Reynolds *et al.* [53] einerseits und Gawne und Martin andererseits [16] nicht ohne weiteres im Rahmen des hier untersuchten HMAX-Modells vereinbart werden können. Zwar bestätigt sich die Vermutung, daß die Art der Interaktion zweier visueller Stimuli auf neuronaler Ebene von ihrem Abstand abhängen kann. In HMAX wird die Antwort eines C2-Neurons dabei für große Abstände alleine vom stärker bevorzugten Stimulus bestimmt (MAX-Operation), während für geringere Abstände Interaktionen auf S1- und S2-Ebene zu Antworten des Neurons auf zwei Stimuli führen können, die eher einer linearen Summe oder einem gewichteten Mittel der Antworten auf die einzelnen Stimuli entsprechen. Die hier gezeigten Daten lassen es allerdings als unwahrscheinlich erscheinen, daß alleine

solche Interaktionen auf Afferenzenebene die von Reynolds *et al.* beobachteten Effekte erklären können. Auch geringe Stimulusabstände und eine schärfere Abstimmung der Modellneuronen auf ihre bevorzugten Stimuli konnten nicht konsistent den vom Biased Competition Model postulierten linearen Zusammenhang zwischen sensorischer Interaktion (SI-Index) und Selektivität (SE-Index) reproduzieren. Das heißt, es konnten keine Bedingungen gefunden werden, unter denen ein hierarchisches neuronales Modell, dessen Neuronen lediglich Summen- und Maximumsoperationen über ihren Afferenzen ausführen, auf eine Kombination zweier Stimuli generell und systematisch, in einem Maße wie vom Biased Competition Model vorhergesagt, entsprechend einer Mittelungsoperation antworten würde. Auf der Grundlage der Daten von Reynolds *et al.* ist also zu vermuten, daß das Gehirn in der Tat, zumindest in Subpopulationen von Neuronen im visuellen Cortex, eine Form von Wettbewerb zwischen Repräsentationen verschiedener gleichzeitig präsentierter Stimuli realisiert. Die im Cortex weit verbreiteten lateralen inhibitorischen Verbindungen (siehe etwa [32]) dürften dafür geeignet sein.

Die Annahme kompetitiver Interaktionen muß dabei keineswegs im Widerspruch zu den Daten von Gawne und Martin und der neuronalen Realisierung der MAX-Operation stehen. Zum einen ist es sicherlich nicht unrealistisch anzunehmen, daß verschiedene Subgruppen von Zellen etwa in V4 ihren afferenten Input auf unterschiedliche Weise verarbeiten. In der Tat wäre die Realisierung der MAX-Operation auch in höheren visuellen Arealen wie V4 sehr sinnvoll, um die Spezifität von Neuronen für ihre bevorzugten Stimuli zu erhalten, insbesondere wenn ihre rezeptiven Felder bereits so umfangreich sind, wie es in V4 der Fall ist. Es ist nicht unbedingt klar, welchen Zweck eine gewichtete Mittelung verschiedener Stimuli auf neuronaler Ebene für deren Wahrnehmung haben könnte, da dadurch eindeutig Spezifität verlorengeht. Tatsächlich findet man bei genauem Hinsehen auch in den Daten von Reynolds *et al.* [53] (Abb. 4 D und F, S. 1742) Neuronen, deren Antworten mit den Vorhersagen des HMAX-Modells kompatibel sind.

Zum anderen sind durchaus Kombinationen der beiden Verarbeitungsschemata denkbar. Laterale Inhibition etwa kann natürlich auch zwischen Neuronen, die eine MAX-Operation ausführen, implementiert sein. Es wäre sicher vermessen zu behaupten, die Verarbeitung visueller Information im Cortex würde exakt nach den Prinzipien des gegenwärtigen HMAX-Modells ablaufen. Weitere Simulationen mit einem entsprechend mo-

difizierten Modell könnten klären, ob die vorliegenden experimentellen Daten im Rahmen eines solchen „gemischten“ Modells vereint werden können. Dieses Kapitel zeigt also vor allem, daß die Suche nach den Prinzipien der neuronalen Verarbeitung von Information in den höheren Arealen des visuellen Cortex noch längst nicht abgeschlossen ist.

Kapitel 5

Aufmerksamkeit in HMAX: Neuronale Aktivität und Objekterkennung

Daß, wie im vorigen Kapitel gezeigt, HMAX und das Biased Competition Model gegenwärtig unterschiedliche Annahmen über die Verarbeitung afferenten Inputs durch Neuronen im visuellen Cortex machen, bedeutet nicht etwa, daß Aufmerksamkeit in HMAX nicht simuliert werden könnte. Ein expliziter Wettbewerb zwischen neuronalen Populationen ist nicht unbedingt Voraussetzung dafür, daß „top-down", also entgegen dem Fluß der sensorischen Information operierende Aufmerksamkeitseffekte die Wahrnehmung von Stimuli beeinflussen können. Auch in HMAX ist die relative Aktivität von Neuronen oder Neuronengruppen entscheidend dafür, welche Stimuli erkannt werden. Man kann also erwarten, daß Modulationen der Aktivität von Modellneuronen, als Simulation der Effekte von Aufmerksamkeit, hier Wirkung zeigen können. Nach der Grundlagenarbeit der vorangegangenen Kapitel werden nun in diesem Abschnitt Effekte von Aufmerksamkeit im Kontext der Objekterkennung das Thema sein. Die zentrale Frage ist dabei, ob im Modell die simulierten physiologischen Effekte von objektorientierter Aufmerksamkeit, also gezielt veränderte neuronale Aktivitäten, entsprechende Verbesserungen in der Objekterkennungsleistung bewirken können, wie sie im Verhaltensexperiment beobachtet werden. Es geht also darum, ob im Prinzip die neuronalen Effekte von Aufmerksamkeit hinreichend

sein könnten, um die veränderte Leistung bei der Objekterkennung zu erklären. Dafür werden zunächst die aus den verfügbaren experimentellen Daten sowie den Resultaten aus den Simulationen der vorigen Kapitel ableitbaren Rahmenbedingungen formuliert, unter denen Aufmerksamkeit in HMAX modelliert wird. Die Umsetzung dieser theoretischen Überlegungen in Form von Modulationen der Aktivität der Modellneuronen, ihre Auswirkungen auf die Objekterkennungsleistung von HMAX und die Implikationen der erhaltenen Ergebnisse für Aufmerksamkeit im visuellen System sind dann die weiteren Schwerpunkte dieses Kapitels.

5.1 Theoretische Überlegungen zur Aufmerksamkeit

5.1.1 Lokale und objektorientierte Aufmerksamkeit

Es wurde bereits in Kapitel 1 erwähnt, daß zwischen zwei Formen visueller Aufmerksamkeit unterschieden werden kann, je nachdem, ob sie auf eine bestimmte Position im visuellen Feld (lokale Aufmerksamkeit) oder auf ein bestimmtes visuelles Merkmal oder Objekt (merkmals- oder objektorientierte Aufmerksamkeit) gerichtet wird. Experimentelle Daten deuten darauf hin, daß beide Formen sich sowohl in ihren zugrundeliegenden neuronalen Mechanismen als auch in ihren Auswirkungen unterscheiden. Beide rufen grundsätzlich verstärkte Aktivität in Neuronen, die den bewußt beachteten Stimulus repräsentieren, hervor. Lokale Aufmerksamkeit erhöht jedoch in der Regel auch die Spontanaktivität von Neuronen, deren rezeptive Felder mit dem Fokus der Aufmerksamkeit überlappen [36, 52]; sie beeinflußt die stimulusinduzierte Aktivität von Neuronen früher als objektorientierte Aufmerksamkeit [20]; und sie wirkt sich auf die Verarbeitung *aller* Stimuli aus, die an der bewußt beachteten Position erscheinen, unabhängig davon, ob sie im jeweiligen Verhaltenstest Zielobjekte oder Distraktoren sind [21]. Objektorientierte Aufmerksamkeit wirkt sich hingegen nur auf potentielle Zielobjekte aus, kann auch unabhängig von der Position eines Stimulus im visuellen Feld operieren und die Aktivität von Neuronen an mehreren Positionen gleichzeitig beeinflussen, basiert also offensichtlich auf einem parallel arbeitenden Mechanismus [6, 8, 20, 46].

Wie bereits diskutiert, könnte ein Mechanismus lokaler Aufmerksamkeit in einem Mo-

dell der Objekterkennung wie HMAX implementiert werden, indem etwa nur Stimuli aus dem Bereich des Fokus der Aufmerksamkeit verarbeitet werden [76] oder alle Zellen, deren rezeptive Felder an entsprechenden Positionen liegen, eine Erhöhung ihrer Aktivität erfahren. In dieser Arbeit wird es jedoch darum gehen, wie objektorientierte Aufmerksamkeit realisiert werden kann, wie also möglicherweise die Erwartung eines Objekts oder die Suche nach einem Objekt dessen Erkennung erleichtern kann, selbst wenn die genaue Position oder sogar das genaue Aussehen des Objekts im voraus nicht bekannt sind. Es ist anzunehmen, daß eine solche Form von Aufmerksamkeit etwa in RSVP-Experimenten eingesetzt wird.

5.1.2 Frühe und späte Selektion

Eine der großen Diskussionen in der Literatur über Aufmerksamkeit und ihre Funktion im visuellen System dreht sich um die Frage, ob Aufmerksamkeit bereits *vor* der Objekterkennung im eigentlichen Sinn eine Rolle spielt, sich daher auf die neuronalen Repräsentationen von einzelnen visuellen Merkmalen, nicht von Objekten als Ganzes auswirkt und effektiv das visuelle System im voraus passend auf die Verarbeitung erwarteter Stimuli einstellt [5, 21], oder ob Aufmerksamkeit erst auf die neuronalen Repräsentationen von Stimuli einwirkt, deren Verarbeitung bereits bis zur Stufe der Erkennung individueller Objekte fortgeschritten ist. Ersteres wird im allgemeinen als frühe Selektion („early selection“), letzteres als späte Selektion („late selection“) bezeichnet. Viele experimentelle Resultate wie etwa die Verbesserung der Erkennungsleistung im RSVP-Experiment, wenn Vorwissen über das zu detektierende Zielobjekt vorhanden ist, können im Prinzip durch beide Theorien erklärt werden. Die Theorie der frühen Selektion nimmt an, daß das Vorwissen über zu erkennende Stimuli das visuelle System entsprechend auf deren Wahrnehmung einstellt (gewissermaßen ein „Tuningprozess“) und daher Stimuli, die ansonsten der Wahrnehmung entgangen wären, nun erkannt werden können [5]. Späte Selektion geht dagegen davon aus, daß selbst sehr kurz präsentierte Stimuli in der Regel bis zur Erkennung der einzelnen Objekte verarbeitet werden können, daß aber die rasche Folge weiterer Stimuli im RSVP-Experiment üblicherweise dazu führt, daß auch erkannte Stimuli schnell wieder vergessen werden, es sei denn, sie werden durch Aufmerksamkeitsprozesse selektiert und im Gedächtnis gehalten [51]. Auch in der Physiologie finden sich Belege für

beide Ansätze. Während einige Studien Änderungen der neuronalen Aktivität in V4 und IT aufgrund von nicht-lokaler, objektorientierter Aufmerksamkeit schon vor dem Zeigen des eigentlichen Stimulus finden [7, 17], stellen andere einen Einfluß dieser Form von Aufmerksamkeit in Arealen von V1 bis V4 erst ab etwa 150 ms nach Stimulationsbeginn oder später fest [8, 46, 58].

In Anbetracht dieser Diskussion ist klar, daß ein eventueller Einfluß von Aufmerksamkeit auf die Objekterkennung definitionsgemäß nur ein Mechanismus der frühen Selektion sein kann. Wie effektiv ein solcher Mechanismus sein kann, soll in dieser Arbeit untersucht werden. Da es hier zugleich um ortsunabhängige, objektorientierte Formen von Aufmerksamkeit gehen soll, wird dabei die Frage zu beantworten sein, worauf genau Aufmerksamkeit sich richten soll. Genauer gesagt gilt es herauszufinden, wie zur besseren Erkennung eines Stimulus Merkmale, die ihn möglichst hinreichend von anderen Stimuli unterscheiden, im voraus selektiert werden können, um dann vom visuellen System bevorzugt verarbeitet zu werden (etwa dadurch, daß auf diese Merkmale ansprechende Neuronen in elementareren Arealen des visuellen Systems durch modulierende Einflüsse aus höheren Arealen sensitiver reagieren oder stärker spontan aktiv werden). RSVP-Experimente haben gezeigt, daß auch sehr abstrakte Formen von Vorwissen über ein zu erkennendes Zielobjekt dessen Detektion verbessern können, etwa Informationen über die übergeordnete Objektkategorie, der es angehört, oder sogar darüber, welcher Kategorie es im Unterschied zu den anderen präsentierten Objekten *nicht* angehört (wie z.B. „Transportmittel“ bzw. „kein Tier“ für eine Abbildung eines Autos) [23]. Insbesondere in solchen Fällen, in denen kaum etwas über das tatsächliche Aussehen eines zu erkennenden Objekts im voraus bekannt ist, wird zu klären sein, wie ein vor der Objekterkennung operierender Mechanismus von visueller Aufmerksamkeit effizient funktionieren kann.

5.1.3 Zeitlicher Verlauf von Objekterkennungsprozessen

Eine wichtige Einschränkung möglicher Modelle von Aufmerksamkeit ergibt sich aus Versuchen über den zeitlichen Verlauf von Prozessen der Objekterkennung bei Menschen und Affen. Sowohl ERP-Ableitungen (*event-related potential*, ein elektroenzephalographisches Meßparadigma) als auch RSVP-Experimente deuten darauf hin, daß Objekte, selbst in komplexen, realistischen Darstellungen, innerhalb von nur etwa 150 ms nach Beginn der

visuellen Stimulation erkannt und kategorisiert werden können [14, 51, 67]. Diese Zeitspanne ist kaum länger als die Latenz des visuellen Signals im inferotemporalen Cortex (etwa 100 – 130 ms [59, 65]) und läßt faktisch keine Zeit für aufwendige Feedback-Prozesse. Das heißt, Einflüsse von Aufmerksamkeit auf die Objekterkennung selbst können prinzipiell nur in Form von „Tuning"-Prozessen auftreten, die das visuelle System vor oder zu Beginn der Stimulation angemessen vorbereiten und dann die Objekterkennung in einem einzigen „Feedforward"-Durchlauf der sensorischen Information durch die Hierarchie der visuellen Areale ermöglichen. Das entspricht dem für diese Studie gewählten und im vorigen Abschnitt diskutierten Paradigma der frühen Selektion.

5.1.4 Neurophysiologie der Aufmerksamkeitseffekte

In früheren Studien wurde vermutet, daß der Effekt von visueller Aufmerksamkeit auf neuronaler Ebene eine Verkleinerung des rezeptiven Feldes [45] oder eine schärfere Abstimmung auf einen Stimulus sein könnte [18]. Veränderungen in den synaptischen Verbindungen eines Neurons sind in diesem Zusammenhang allerdings sehr unwahrscheinlich, da der Prozeß sehr schnell und reversibel ist [10], und eine Veränderung der Abstimmkurve eines Neurons auf einen bestimmten Stimulus wird nicht beobachtet, wenn die gemessenen Feuerraten um die Spontanaktivität korrigiert werden [41]. Eher wahrscheinlich dürften rasche Veränderungen in synaptischen Stärken sein, die selektiv den bewußt beachteten Stimulus begünstigen, wie im Biased Competition Model angenommen [53], oder excitatorische Stimulation von Zellen, die auf den bewußt beachteten Stimulus ansprechen, durch Neuronen höherer Areale, was häufig in Modellierungsstudien verwendet wird [72, 73] und die Aktivität der betroffenen Zellen entweder direkt oder über Disinhibitionsmechanismen erhöht [43].

Es wurde auch diskutiert, ob die Auswirkung von Aufmerksamkeit auf die Feuerrate eines Neurons eher als multiplikativ beschrieben werden kann (d.h. hohe Feuerraten würden stärker moduliert als niedrige) [41] oder als eine Kontrastverstärkung, also eine Veränderung des Kontrastantwortverhaltens des Neurons (im folgenden auch als Verschiebung der Kontrastantwortkurve bezeichnet), so daß die Antwort des Neurons auf Stimuli geringen Kontrasts am stärksten zunehmen und auch schneller saturieren würde [52]. Beides wurde experimentell beobachtet, allerdings in unterschiedlichen Versuchsansätzen.

Die beiden Mechanismen können daher, wie von Reynolds *et al.* vorgeschlagen, miteinander vereinbart werden, wenn man annimmt, daß die *Abstimmkurve* eines Neurons, die die Antwortstärke auf *verschiedene* Stimuli beim *gleichen* Kontrast beschreibt, durch Aufmerksamkeit multiplikativ verändert wird (d.h. die Antwort auf bevorzugte Stimuli wird stärker erhöht), während die *Kontrastantwortkurve*, die die Antwort des Neurons auf *denselben* Stimulus bei *verschiedenen* Kontrasten beschreibt, zu kleineren Kontrastwerten hin verschoben wird und daher die Feuerrate bei geringen Kontrasten am stärksten zunimmt.

In der Literatur über physiologische Mechanismen der Aufmerksamkeit gibt es weiterhin einen breiten Konsens darüber, daß nicht nur Neuronen, die auf Stimuli im Fokus der Aufmerksamkeit ansprechen, verstärkt feuern, sondern auch, daß andere Neuronen in ihrer Aktivität abgeschwächt werden können [7, 8, 46, 68]. Dies paßt in das Bild von Aufmerksamkeit als einem Mechanismus, der bestimmte Stimuli auf neuronaler Ebene selektiv bevorzugt und sie dadurch in der Wahrnehmung stärker vor anderen hervortreten läßt. Die zitierten Studien berichten jedoch in der Regel nur von abgeschwächter Aktivität bei solchen Neuronen, die auf Stimuli ansprechen, die im jeweils gezeigten Bild tatsächlich vorhanden sind, aber gerade nicht bewußt beachtet werden. Es ist nicht klar, ob das auch für Neuronen mit anderen Stimuluspräferenzen gilt. Allerdings erscheint es als relativ unwahrscheinlich, daß etwa in einem Areal des visuellen Cortex *alle* Neuronen außer denen, die vom gerade bewußt beachteten Stimulus aktiviert werden, supprimiert werden sollten. Dann taucht jedoch in einem Modell der frühen Selektion von Stimuli durch Aufmerksamkeit wie in dieser Arbeit wiederum die Frage auf, welche Neuronen für eine Abschwächung ihrer Aktivität ausgewählt werden sollen. Über visuelle Merkmale von Distraktor- oder Hintergrundstimuli ist im Verhaltensexperiment in der Regel eher noch weniger im voraus bekannt als über die von Zielobjekten, so daß das Problem der frühen Selektion hier noch deutlicher zutage tritt.

Die Quellen der *in vivo* beobachteten Modulationen von neuronalen Aktivitäten durch Aufmerksamkeit werden in einem „verteilten Aufmerksamkeitsnetzwerk“ höherer corticaler Areale, insbesondere im frontalen und präfrontalen Cortex, vermutet. In Arealen wie den frontalen und supplementären Augenfeldern finden sich Neuronen, deren Aktivität die Aufmerksamkeit selbst, nicht etwa von Aufmerksamkeit modulierten sensorischen Input widerzuspiegeln scheint, da sie von der genauen auszuführenden Aufgabe

unabhängig ist und vom Auftauchen des eigentlichen Stimulus im visuellen Feld nicht weiter beeinflußt wird [27]. Im Falle von lokaler Aufmerksamkeit scheint außerdem der posterior parietale Cortex eine besondere Rolle zu spielen [6].

5.1.5 Objektcodierung und Modulationen neuronaler Aktivität

In HMAX wird die Präsenz von Objekten im visuellen Feld durch die Aktivität der 256 C2-Zellen und die von ihnen abgeleitete Aktivität der VTUs codiert. Welchen „Code“ das Gehirn tatsächlich verwendet, ist freilich nicht genau bekannt. Wären etwa, wie in HMAX, die genauen Feuerraten einzelner Neuronen entscheidend für die Stimuluscodierung, könnte man wohl nicht den experimentell gefundenen hohen Grad an Kontrastinvarianz der Objekterkennung erwarten [1]. Außerdem müßte dann ein Mechanismus der frühen Selektion von Stimulusmerkmalen durch Aufmerksamkeit nicht nur die für die richtigen Merkmale codierenden Neuronen im voraus auswählen, sondern auch die passende Stärke der Modulation. Ansonsten könnte die Aktivität der für den bewußt beachteten Stimulus codierenden Zellen eventuell sogar zu stark erhöht werden, etwa wenn der Stimulus bei höherem Kontrast erscheint als erwartet, und die Erkennungsleistung würde wieder absinken.

Ein neuronaler Code, der nicht von genauen Werten der Aktivität einzelner Neuronen abhängt, sondern etwa davon, welche Neuronen aktiv sind und welche nicht (also ein binärer Code), wäre zwar im Prinzip weniger von Variationen im Stimuluskontrast oder vom genauen Betrag einer Modulation der Feuerraten durch Aufmerksamkeit abhängig, andererseits ginge dabei durch den Verzicht auf die in verschiedenen Feuerraten enthaltene Information auch eine erhebliche Menge an Codierungskapazität verloren. Für einen völlig vom Kontrast des Stimulus unabhängigen Code wäre außerdem nicht zu erwarten, daß Aufmerksamkeitseffekte überhaupt einen Einfluß auf die Objekterkennung haben können, sofern sie sich, wie oft angenommen wird, ähnlich wie eine Änderung im Kontrast des Stimulus auswirken [52, 68].

In HMAX zeigt sich, daß VTUs, die für verschiedene Objekte codieren, je nach der verwendeten Stimulusklasse zum Teil eine große Anzahl ihrer Afferenzen gemeinsam haben (siehe Kapitel 3). Auch im Gehirn nehmen Neuronen in der Regel an der Repräsentation

mehrerer Stimuli teil [60]. Damit ergibt sich allerdings das Problem, wie Modulationen neuronaler Aktivität durch Aufmerksamkeit gezielt auf die Repräsentationen bestimmter Stimuli wirken können. Es wird also zu untersuchen sein, welchen Grad an Spezifität solche Modulationen haben können.

Schließlich gilt es zu bedenken, daß bei früher Selektion von Merkmalen durch Aufmerksamkeit die Chance besteht, die falschen Merkmale auszuwählen. Im Verhaltensexperiment entspräche das der Situation, daß ein Zielobjekt vorgegeben wird, das in der eigentlichen Stimuluspräsentation gar nicht auftaucht. Das visuelle System sollte mit einer solchen Situation zurechtkommen können und die Objekte erkennen, die tatsächlich gezeigt werden, nicht etwa die, die erwartet wurden. Der verwendete Code bzw. die verwendeten Mechanismen von Aufmerksamkeit sollten also hohe Fehlalarmquoten vermeiden können. Auch das wird für Modelle von Aufmerksamkeit zu überprüfen sein.

5.2 Methoden

Im folgenden wird nun beschrieben, wie die bisherigen theoretischen Überlegungen zur Aufmerksamkeit in HMAX umgesetzt wurden.

5.2.1 Stimuli

Es ist anzunehmen, daß Aufmerksamkeit eine eventuelle Rolle in der Objekterkennung vor allem unter ungünstigen Bedingungen wie niedrigem Kontrast [52] oder Gegenwart von Hintergrundstimuli spielt. Für die Simulationen in diesem Kapitel wurden daher Stimuluspräsentationen verwendet, in denen sowohl ein Zielobjekt zusammen mit einem Distraktor gezeigt wurde als auch der Kontrast beider unabhängig voneinander variieren konnte. Dabei war der Kontrast des Zielobjektes stets geringer als der für das Training der entsprechenden VTU verwendete, also kleiner als 100%. Das entspricht den experimentellen Resultaten, daß Aufmerksamkeitseffekte nur bei geringen und mittleren Stimuluskontrasten auftreten [52]. Außerdem konnte so verhindert werden, daß ein Effekt von Aufmerksamkeit auf C2-Zellen in jedem Fall ihre Aktivität über den beim Training des betreffenden Stimulus erreichten Wert hinaus erhöhen und damit die Erkennungsleistung

wieder verringern würde.

Wie in Kapitel 3.1 beschrieben waren Zielobjekte wiederum die 8 Auto-Prototypen bzw. 15 Büroklammern, als Distraktoren wurden für Autos aus den 252 übrigen Autos solche in einem definierten Morphabstand vom Zielobjekt gewählt, für Büroklammern 60 zufällig ausgewählte andere Büroklammern.

5.2.2 Auswahl der modulierten Zellen

Modulationen neuronaler Aktivität wurden auf Ebene der C2-Zellen oder der VTUs durchgeführt, da diese Zellen Afferenzen aus dem gesamten modellierten visuellen Feld erhalten sowie komplexe Stimuli repräsentieren und daher am geeignetsten sind, um nichtlokale, objektorientierte Aufmerksamkeit zu simulieren. Außerdem sind C2-Zellen und VTUs als Modelle von Neuronen in den visuellen Arealen V4 bzw. IT konzipiert, in denen die frühesten und stärksten neuronalen Effekte von Aufmerksamkeit beobachtet werden [42]. Da die Ausgabewerte des Modells in Begriffen von Objekterkennung interpretiert werden, repräsentieren Änderungen der Aktivität von Neuronen vor dem Auslesen der Resultate definitionsgemäß Mechanismen der frühen Selektion durch Aufmerksamkeit, d.h. es werden die möglichen Einflüsse der Aufmerksamkeit auf die Prozesse der Objekterkennung selbst untersucht.

Das Problem der Selektion geeigneter Merkmale zur bevorzugten Wahrnehmung eines Zielobjekts wurde im einfachsten Fall dadurch gelöst, daß auf ein einzelnes Zielobjekt mit bekanntem Aussehen gerichtete Aufmerksamkeit simuliert wurde. Dafür wurden die Stimuli gewählt, auf die VTUs trainiert waren (8 Auto-Prototypen oder 15 Büroklammern). Die jeweils verwendeten Modulationen der neuronalen Aktivität wurden dann auf die C2-Afferenzen dieser VTUs angewendet. Dieses Modell entspräche *in vivo* einer „top-down"-Modulation der Aktivität von Zellen in V4 durch für Objekte bzw. Objektperspektiven selektive Zellen im inferotemporalen Cortex. Dabei wurden 40 oder 100 C2-Afferenzen pro VTU verwendet. Für VTUs, die mit allen 256 C2-Zellen verbunden sind, würden sich Modulationen der Aktivität ihrer Afferenzen gleichermaßen auch auf alle anderen VTUs auswirken. Diese Anzahl an Afferenzen wurde daher nur verwendet, um die Leistung in den anderen Fällen mit einer Situation vergleichen zu können, in der von vorneherein

die Modulation durch Aufmerksamkeit nicht gezielt auf bestimmte Stimuli gerichtet sein konnte.

Für jede der 8 auf Auto-Prototypen und der 15 auf Büroklammern trainierten VTUs wurden Simulationen gerechnet, in denen jeweils ihr bevorzugter Stimulus das Zielobjekt der Aufmerksamkeit war, d.h. in denen ihre jeweiligen Afferenzen eine Form von Modulation erhielten. Dabei wurde für jede VTU ein Satz von Antworten auf *alle* Stimuli der jeweiligen Stimulusklasse berechnet, wobei stets nur die C2-Afferenzen *dieser* VTU in ihrer Aktivität moduliert waren. Dadurch wurde für die Auswertung im Stimulusvergleich-Paradigma die Vergleichbarkeit der Antworten einer VTU auf verschiedene Stimuli sichergestellt.

Komplexere Situationen, in denen mehr als eine VTU oder deren Afferenzen Ziele der Modulationen durch Aufmerksamkeit sind, werden unter 5.3.6 und 5.3.7 besprochen.

Weiterhin wurde auch die Wirkung von Suppressionsmechanismen untersucht. Die Zellen, deren Aktivität abgeschwächt wurde, waren zum einen die C2-Afferenzen der am stärksten aktiven unter den VTUs, die nicht auf das jeweilige Zielobjekt trainiert waren. Das entspräche einer Suppression der Repräsentation eines gerade nicht beachteten Stimulus; allerdings handelt es sich hierbei schon nicht mehr um einen Mechanismus der frühen Selektion, da erst nach einem ersten kompletten Durchlauf der eingehenden visuellen Information durch das Modell bestimmt werden kann, welches die am stärksten aktive unter den nicht für das Zielobjekt codierenden VTUs ist. Dennoch wurden die Auswirkungen dieses Mechanismus untersucht, weil die Suppression der Aktivitäten von Zellen, die für einen möglichen oder tatsächlichen Distraktor codieren, biologisch am plausibelsten ist. Als Alternative dazu wurde auch mit der Suppression *aller* C2-Zellen, die nicht auf die VTU des jeweiligen Zielobjekts projizierten, experimentiert. Ein solcher breit wirkender Suppressionsmechanismus erscheint, wie oben bereits diskutiert, nicht sehr realistisch, wurde hier aber als der spezifischste denkbare Mechanismus der Selektion von Merkmalen durch Aufmerksamkeit untersucht.

5.2.3 Art der Modulationen

Die Aktivitäten der für eine Erhöhung der Feuerrate ausgewählten Zellen wurden zum einen durch Addition verschiedener konstanter Werte, zum anderen durch Multiplikation

mit verschiedenen Faktoren (entsprechend der Ergebnisse von McAdams und Maunsell [41] bzw. Motter [46]) modifiziert. Eine dritte Methode war eine Verschiebung der „Kontrastantwortkurve“ der S2- bzw. C2-Zellen, also der Gaußschen Funktion, die die Antwort dieser Zellen auf die Aktivität ihrer afferenten C1-Zellen bestimmt, hin zu niedrigeren Werten der C1-Aktivität. Diese Modifikation entspricht der von Reynolds *et al.* beschriebenen Linksverschiebung der Kontrastantwortkurve von V4-Neuronen, die einer Erhöhung der Kontrastsensitivität und frühere Saturierung zur Folge hat [52]. Die Umsetzung in HMAX geschah durch Verkleinerung des Mittelwertes *s2Target* der Gaußschen Antwortkurve der S2-Zellen gegenüber dem regulären Wert 1, wie bereits in Kapitel 4 beschrieben. (Derselbe Wert wird dabei für alle vier Afferenzen einer S2-Zelle verwendet.) Diese drei Modulationsmethoden wurden hinsichtlich ihrer Wirksamkeit in der Objekterkennung verglichen.

Für die Suppression der Aktivität von Zellen wurde in der Regel multiplikative Verfahren verwendet, d.h. Multiplikationen der Feuerraten mit verschiedenen Faktoren kleiner als 1.

5.2.4 Auswertung

Die Auswertung der vom Modell erzielten Erkennungsleistung erfolgte wie in Kapitel 3.1 für Simulationen mit zwei Stimuli beschrieben, im Aktivste VTU-Paradigma und im Stimulusvergleich-Paradigma. Dabei wurde die Leistung im Stimulusvergleich-Paradigma durch ROC-Kurven dargestellt. Aufgrund der bereits erwähnten Tatsache, daß für jede VTU ihre Antworten auf *alle* Stimuli jeweils nur mit Modulation der Aktivität *ihrer* Afferenzen durch Aufmerksamkeit berechnet wurden, können die Antworten einer VTU auf verschiedene Stimuli sinnvoll verglichen werden. Auf diese Weise ist leicht zu erkennen, ob ein Aufmerksamkeitsmechanismus spezifisch wirkt und die Aktivität einer VTU bei Präsentation ihres Zielobjekts höher ist als sonst, oder ob er unspezifisch wirkt und die Aktivität einer VTU auch bei Präsentation anderer Stimuli gleichermaßen erhöht, also die Fehlalarmrate ebenso steigt wie die Trefferquote. Die Modulation der Afferenzen einer VTU durch Aufmerksamkeit in den Fällen, in denen ihr bevorzugter Stimulus gar nicht gezeigt wird, entspricht dabei der Situation im Verhaltensexperiment, daß ein bestimmter Stimulus erwartet bzw. gesucht, aber nicht präsentiert wird – eine Situation, in der das visuelle System entsprechend die tatsächlich präsenten Stimuli erkennen muß und nicht

die, auf die sich die Aufmerksamkeit gerichtet hatte.

Ein entsprechender Kontrollversuch für Fehlalarme durch Aufmerksamkeit im Aktivste VTU-Paradigma besteht darin, die Aktivität der C2-Afferenzen von VTUs zu erhöhen, deren bevorzugter Stimulus jeweils gerade nicht im Bild ist, und zu überprüfen, ob dennoch aufgrund des Einsatzes dieser Form von Aufmerksamkeit die VTU, deren Afferenzen moduliert wurden, stärker als alle anderen VTUs feuert, also ein Fehlalarm eintritt. Dieser Kontrollversuch wird in 5.3.8 durchgeführt.

5.2.5 Alternative Codierungsmethoden

Neben der üblichen Codierung von Stimuli in HMAX, bei der die VTUs in Abhängigkeit von den genauen Aktivitäten ihrer C2-Afferenzen feuern, wurden zwei alternative Codierungsmethoden verwendet, die die Abhängigkeit der Leistung des Modells von den genauen Feuerraten der C2-Zellen und damit vom Stimuluskontrast sowie von den genauen Stärken einer Modulation durch Aufmerksamkeit verringern sollten. Zum einen wurden in einer Reihe von Versuchen die VTUs so abgestimmt, daß sie auf Feuerraten ihrer C2-Afferenzen, die über den im Training erreichten Werten lagen, nicht wieder mit einem Absinken der Antwort reagierten. Diese hier als *„Gauss-Maximum-Tuning“* bezeichnete Codierungsmethode versieht die VTUs effektiv mit einer sigmoidal-ähnlichen Antwortkurve und Sättigungsverhalten, da für den Fall, daß die Afferenzen einer VTU genauso stark wie oder stärker als während des Trainings dieser VTU feuern, die Antwort dieser VTU stets maximal (also 1) sein wird. Um das zu erreichen, wurde der Exponent der Gaußfunktion, die die Antwort der VTUs auf ihre C2-Afferenzen bestimmt, auf Null gesetzt, sobald alle C2-Afferenzen genauso stark wie oder stärker aktiv waren als während des Trainings mit dem Zielobjekt bei 100% Kontrast.

Zum anderen wurden VTUs auch so abgestimmt, daß stets diejenige VTU am stärksten aktiv sein würde, deren C2-Afferenzen am stärksten feuern, und nicht wie sonst diejenige, mit deren bevorzugter C2-Aktivität die gerade vorhandene C2-Aktivität am genauesten übereinstimmt. Dies wurde dadurch erreicht, daß die Mittelwerte der Gaußschen Antwortkurven, die die Antwort einer VTU auf den Input einer ihrer C2-Afferenzen bestimmen, alle auf einen Wert gesetzt wurden, der größer war als jede von den C2-Zellen

tatsächlich erreichte Aktivität. Jede Erhöhung der Aktivität einer auf eine VTU projizierenden C2-Zelle führte daher auch zu einer Erhöhung der Aktivität dieser VTU. Bei diesem sogenannten *„Aktivste Afferenzen-Tuning“* sind also die absoluten Feuerraten der C2-Zellen nicht entscheidend dafür, welcher Stimulus erkannt wird; im Aktivste VTU-Paradigma würde etwa der Stimulus erkannt, der von derjenigen VTU codiert wird, die über die relativ am stärksten feuernden Afferenzen verfügt. Die Spezifität dieser Codierungsmethode für verschiedene Stimuli ist also nur durch die Auswahl der C2-Afferenzen gegeben.

Für beide alternativen Codierungsmethoden wurden die Leistungen bei der Objekterkennung wie in den Versuchen mit der üblichen HMAX-Codierung bestimmt, mit Autos und Büroklammern als Stimuli, im Aktivste VTU- und Stimulusvergleich-Paradigma sowie für multiplikative Aufmerksamkeitseffekte.

5.2.6 Populationscode

In einem weiteren Experiment wurden mögliche Effekte von Aufmerksamkeit im Kontext eines VTU-Populationscodes untersucht, also für den Fall, daß ein Stimulus durch die Aktivität mehrerer VTUs repräsentiert wird. Diese Art der Stimuluscodierung entspricht den vom Gehirn verwendeten Prinzipien sehr viel eher als die Repräsentation durch einzelne Neuronen [19, 60]. Die Verwendung einzelner VTUs zur Codierung von Stimuli in den übrigen Kapiteln dieser Arbeit ist dabei nicht als Postulat der Existenz von „grandmother neurons“ zu verstehen, vielmehr als Vereinfachung für die Simulationen. Hier soll jedoch explizit die Wirkung eines Populationscodes und der mögliche Einfluß von Aufmerksamkeit darauf untersucht werden.

Dabei galt es zu vermeiden, daß die betrachtete VTU-Population zugleich auf die Stimuli trainiert wurde, die dann als Zielobjekte der Erkennung präsentiert wurden, um die Situation zu simulieren, in der eine Population von Neuronen, die auf Exemplare einer Stimulusklasse selektiv reagiert (wie die Gesichterneuronen im inferotemporalen Cortex von Makaken [11]), mit neuen Exemplaren dieser Klasse konfrontiert wird. Dafür wurde in diesen Simulationen eine neue Stimulusklasse verwendet, und zwar ein Datensatz von 200 Frontalansichten von Gesichtern, der von Thomas Vetter zur Verfügung gestellt wurde [3]. Für 10 dieser Gesichter waren, analog zu den Autos, gemorphte Stimuli verfügbar,

Abbildung 5-1: Beispiel für die als Stimuli verwendeten Gesichter.

die je zwei von ihnen mit einer „Morphlinie“ verbanden, d.h. jedes dieser 10 Gesichter konnte in 10 Schritten in jedes andere dieser 10 Gesichter umgewandelt werden. Auf jedes der 190 übrigen Gesichter wurde dann eine VTU trainiert; diese 190 VTUs stellten die betrachtete Population dar. Jede VTU konnte dabei mit 40 oder 100 C2-Zellen verbunden sein. Dieser Population wurden dann Darstellungen zweier Gesichter aus der Untermenge der 10 übrigen Gesichter und deren gemorphten Varianten präsentiert, in ganz analoger Weise wie im Fall der Autos: eines der 10 ursprünglichen, „prototypischen“ Gesichter als zu erkennendes Zielobjekt an der Position $(50|50)$, bei einem Kontrastwert geringer als 100%, und ein gemorphtes Gesicht als Distraktor an der Position $(110|110)$ innerhalb eines Bildes der Gesamtgröße 160×160 Pixel. Beide Gesichter waren, wie auch die Autos und Büroklammern, je 64×64 Pixel groß.

Ob Zielobjekte durch die VTU-Population erkannt wurden, wurde durch eine zweite Schicht von 10 VTUs bestimmt. Diese VTUs waren trainiert auf das Aktivitätsmuster der VTU-Population bei der isolierten Präsentation eines der 10 als Zielobjekte verwendeten Gesichter, d.h. eine der 10 VTUs in der zweiten Schicht feuerte mit einer maximalen Aktivität von 1, wenn die VTUs in der Population, mit denen sie verbunden war (die 10, 40 oder 100 für das Zielobjekt am stärksten aktiven), dasselbe Aktivitätsmuster zeigten wie

bei der Präsentation des entsprechenden Gesichts aus der Menge der 10 Zielobjekte. Auf diese Weise konnten die Resultate der Simulationen eines Populationscodes ebenso ausgewertet werden wie die Ergebnisse für einzelne VTUs. Die VTUs in der zweiten Ebene waren dabei nicht notwendigerweise als Modelle bestimmter Neuronen im Gehirn gedacht, sondern vielmehr ein Mittel, um die Aktivität der VTU-Population auf einfache Weise auslesen und auswerten zu können. Eines der 10 als Zielobjekte verwendeten Gesichter wurde dementsprechend im Aktivste VTU-Paradigma als erkannt gewertet, wenn die auf dieses Gesicht trainierte VTU in der zweiten Schicht bei seiner Präsentation stärker aktiv war als die 9 übrigen (d.h. die Aktivität der VTU-Population entsprach am ehesten derjenigen bei der Präsentation dieses Gesichts alleine). Im Stimulusvergleich-Paradigma war es für die erfolgreiche Erkennung eines der 10 Gesichter notwendig, daß die auf dieses Gesicht trainierte VTU in der zweiten Schicht auf eine Präsentation dieses Gesichts und eines Distraktors mit einer höheren Feuerrate reagierte als auf eine Präsentation, die dieses Gesicht nicht enthielt (d.h. die VTU-Population konnte zwischen Präsentationen des Zielobjekts und anderer Gesichter verläßlich unterscheiden). Entsprechende Vergleiche wurden, wie auch in den Versuchen ohne VTU-Population, paarweise zwischen allen Präsentationen des jeweiligen Zielobjekts und allen anderen Präsentationen, bei denen die beiden Stimuli denselben Morphabstand hatten wie in der Präsentation des Zielobjekts, angestellt.

Um Mechanismen von Aufmerksamkeit im Rahmen eines Populationscodes zu realisieren, wurden entweder C2-Zellen oder VTUs in ihrer Aktivität moduliert. Zielobjekt der Aufmerksamkeit war dabei jeweils eines der 10 prototypischen Gesichter, auf die die VTUs in der zweiten Schicht trainiert waren. Im Modell wurden dann entweder alle VTUs der Population, die mit der entsprechenden VTU in der zweiten Schicht verbunden waren, in ihrer Aktivität moduliert, oder eine Auswahl der C2-Afferenzen dieser VTUs. Diese Auswahl bestand entweder aus der Vereinigung *aller* C2-Afferenzen dieser VTUs oder nur aus denjenigen unter ihnen, die nicht zugleich auch auf andere VTUs in der zweiten Schicht projizierten. Damit sollten zwei verschieden spezifische mögliche Lösungen des Problems, wie Aufmerksamkeit Neuronen für eine Aktivitätserhöhung selektieren könnte, getestet werden. Auch im Populationscode wurde ferner Suppression von nicht durch Aufmerksamkeit selektierten Neuronen simuliert. Je nachdem, ob VTUs oder C2-Zellen durch Aufmerksamkeit in ihrer Aktivität erhöht wurden, wurden entsprechend alle übrigen VTUs der Population supprimiert oder eine bestimmte Menge C2-Zellen. Die supprimierten C2-

Zellen waren dabei entweder alle nicht von Aufmerksamkeit selektierten Zellen oder die Afferenzen derjenigen VTUs aus der Population, die auf die am stärksten aktive VTU der zweiten Schicht, die nicht für das Zielobjekt codierte, projizierten. Damit wurde die Auswahl der supprimierten C2-Zellen analog gehandhabt wie in den Versuchen ohne VTU-Population.

Analog zu den vorhergehenden Versuchen wurden auch hier die Antworten des Modells auf alle Stimuluspräsentationen je einmal für jedes der 10 Zielobjekte berechnet, wobei die simulierte Aufmerksamkeit jeweils auf das entsprechende Zielobjekt gerichtet war, unabhängig davon, ob es tatsächlich gezeigt wurde oder nicht. Damit ergaben auch hier die ROC-Kurven für die VTUs der zweiten Schicht realistische Bilder von Treffer- und Fehlalarmquoten.

5.2.7 Kategorisierung

Die Kategorisierung von Stimuli und Einflüsse von Aufmerksamkeit darauf, entsprechend der RSVP-Experimente, in denen im voraus nur die übergeordnete Objektkategorie eines Zielobjekts bekannt ist, wurden in HMAX folgendermaßen modelliert. Zunächst wurden VTUs auf die 8 Auto- und die 10 Gesichter-Prototypen in Isolation (alle in der Größe 64×64 Pixel) trainiert sowie ein „Kategorisierungsneuron“ darauf trainiert, zwischen den Aktivierungsmustern dieser 18 VTUs zu unterscheiden, die sich ergaben, wenn dem Modell verschiedene einzelne Autos oder Gesichter (in beiden Fällen eine Auswahl aus den gemorphten Varianten der Prototypen) gezeigt wurden. Das Kategorisierungsneuron bestand dabei aus einem Vektor $\vec{x}$ aus 18 Gewichten, der die überdeterminierte Gleichung $\mathrm{A}\vec{x} = \vec{b}$ im Sinne der Methode der kleinsten Quadrate löste. A war dabei eine Matrix, deren Zeilen aus den Antworten der 18 VTUs auf alle für das Training des Kategorisierungsneurons verwendeten Auto- und Gesichtsstimuli bestanden, und $\vec{b}$ ein Vektor, der für jede Zeile der Matrix A den Wert 1 oder -1 enthielt, je nachdem, ob A in dieser Zeile die Antworten der VTUs auf ein Auto oder ein Gesicht enthielt. Das Kategorisierungsneuron war also darauf trainiert, bei Präsentation eines Autos den Wert 1, bei Zeigen eines Gesichts den Wert -1 auszugeben.

Die zu kategorisierenden Teststimuli bestanden aus Präsentationen je eines Autos oder

Gesichts (Prototypen oder gemorphte Stimuli) bei verschiedenen Kontrasten – die Kontrastwerte von Autos und Gesichtern variierten jedoch gemeinsam, da es sich in beiden Fällen um Zielobjekte der Kategorisierung handelte – , zusammen mit einer zufällig ausgewählten Büroklammer als Modell eines störenden Hintergrundstimulus, ebenfalls in unabhangig vom Auto bzw. Gesicht variierendem Kontrast. Die Kontraste waren dabei so gewählt, daß die mittlere Aktivität der C2-Zellen beim jeweils maximalen Kontrast, der für eine Stimulusklasse verwendet wurde, im selben Bereich lag (zwischen 0.2 und 0.3). Damit wurde sichergestellt, daß nicht eine Objektklasse schon aufgrund ihrer spezifischen Merkmale die Antwort des HMAX-Modells dominierte. Die so erhaltenen maximalen Kontraste waren für Autos 100%, für Gesichter 60% und für Büroklammern 35%.

Die Leistung des Modells bei der Kategorisierung wurde durch eine ROC-Kurve ausgewertet. Perfekte Leistung (und daher eine ROC durch den Punkt $(0\%|100\%)$) wurde für vollständige Separierbarkeit der Antworten des Kategorisierungsneurons auf Stimuli der beiden Klassen erreicht, d.h. wenn das Neuron auf jede Präsentation eines Bildes mit einem Auto und einer Büroklammer stärker antwortete als auf irgendeine Präsentation eines Bildes mit einem Gesicht und einer Büroklammer (trainiert wurde schließlich die Antwort 1 auf Autos und die Antwort -1 auf Gesichter). Zufallsniveau, d.h. eine lineare ROC durch die Punkte $(0\%|0\%)$ und $(100\%|100\%)$, wurde erreicht, wenn das Kategorisierungsneuron nicht zwischen den beiden Stimulusklassen unterscheiden konnte.

Die Rolle von simulierter Aufmerksamkeit in diesem Kontext wäre, die Unterscheidbarkeit der beiden Stimulusklassen bei Präsentation mit Distraktorstimulus und niedrigerem Kontrast zu erhöhen. Ob sie das in der Tat leisten konnte, wurde mit Modulationen der Aktivität von VTUs oder C2-Zellen überprüft, analog der Vorgehensweise für den VTU-Populationscode. Dabei wurden verschiedene Methoden verwendet: Verstärkung der Aktivität aller VTUs einer der beiden Klassen; Verstärkung der Aktivität aller C2-Zellen, die mit VTUs einer der beiden Klassen verbunden sind; Verstärkung der Aktivität nur derjenigen C2-Zellen, die nur mit VTUs einer der beiden Klassen und nicht mit VTUs der anderen Klasse verbunden sind; sowie jede dieser Methoden auch in Verbindung mit Suppression der jeweils übrigen Zellen auf der gerade betrachteten Ebene. Die jeweiligen Modulationen waren stets multiplikativer Natur. Sie sollten in etwa den Einfluß von Informationen über Stimuluskategorien aus Arealen wie dem präfrontalen Cortex [15] auf Neuronen in

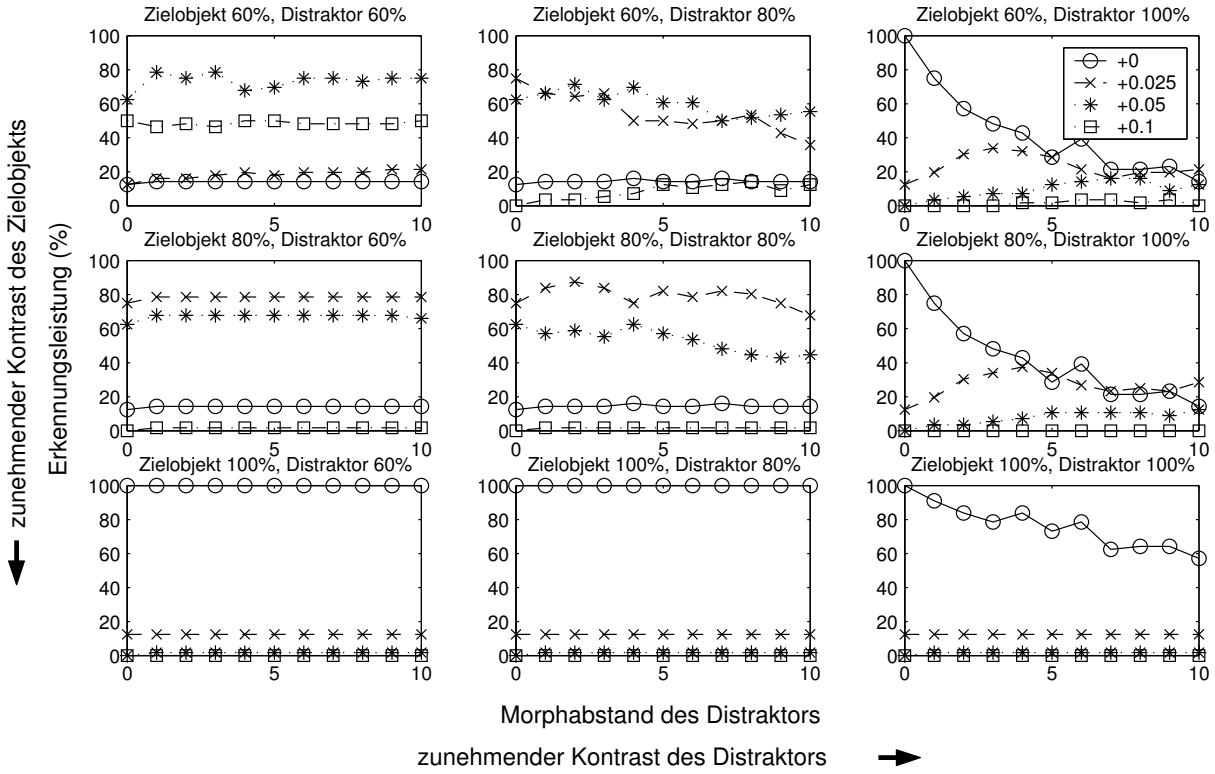

Abbildung 5-2: Additive Aufmerksamkeitseffekte bei C2-Zellen. Erkennungsleistung im Aktivste VTU-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die Werte an, die für die einzelnen Graphen zu den Aktivitäten der C2-Afferenzen der auf das Zielobjekt trainierten VTU addiert wurden. Jede VTU hat 40 C2-Afferenzen.

V4 und IT simulieren.

5.3 Ergebnisse

5.3.1 Additive Aufmerksamkeitseffekte

Abbildungen 5-2 bis 5-5 zeigen, wie sich Aufmerksamkeitseffekte in HMAX auf die Leistung bei der Erkennung von Stimuli auswirken. Hier wurde Aufmerksamkeit modelliert als Addition konstanter Werte zu den Feuerraten der C2-Afferenzen der VTUs, die auf das jeweilige Zielobjekt trainiert wurden. Die Abbildungen zeigen die Erkennungsleistung im Aktivste VTU- bzw. Stimulusvergleichparadigma für Autos und Büroklammern, mit und

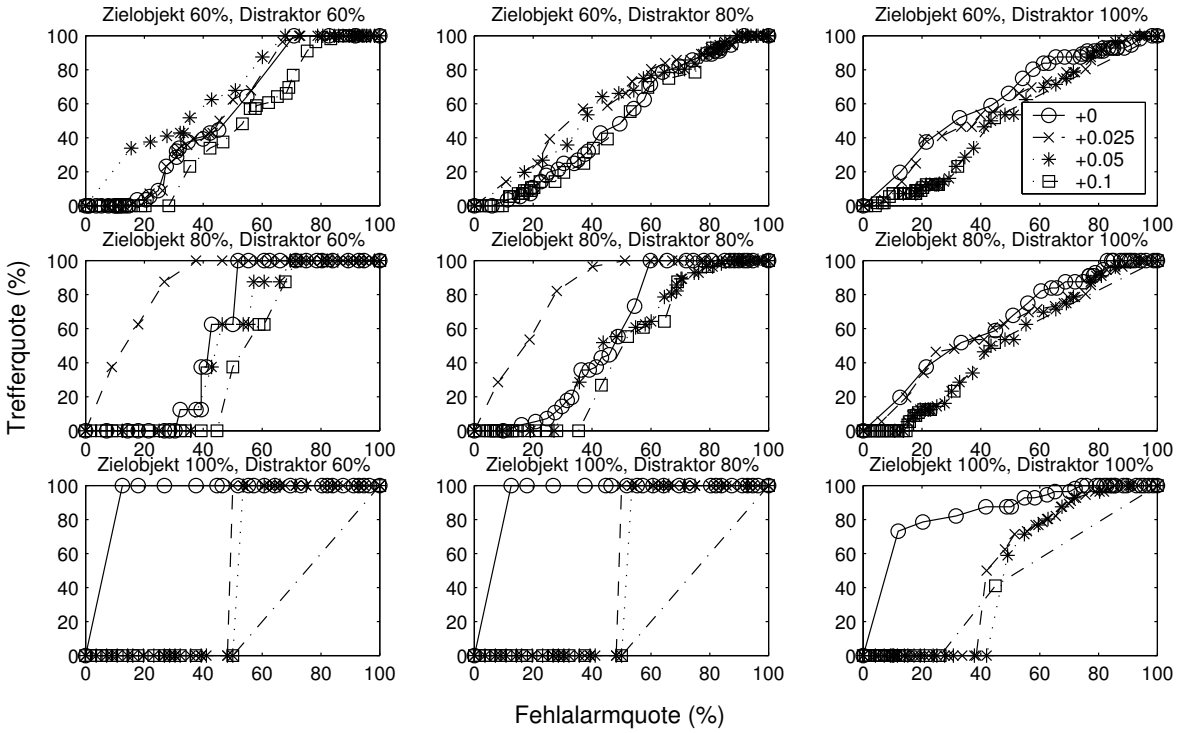

Abbildung 5-3: Additive Aufmerksamkeitseffekte bei C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-2, aber nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

ohne Aufmerksamkeitseffekte. Zur Illustration werden hier auch zwei Fälle präsentiert, die in den Methoden bereits als nicht sinnvoll für die Modellierung von Aufmerksamkeit in HMAX ausgeschlossen wurden: ein Kontrast des Zielobjekts von 100% und VTUs mit 256 Afferenzen. Aufgrund der MAX-Operation, die die C2-Zellen ausführen, kann das Hinzufügen eines zweiten Stimulus ihre Aktivität nur erhöhen, nicht verringern, sofern der Abstand der beiden Stimuli, wie in allen Experimenten in diesem Kapitel, so groß ist, daß keine Interaktionen auf S1-Ebene auftreten (siehe Kapitel 4). Daher werden die C2-Zellen, wenn das Zielobjekt bei 100% Kontrast und zusammen mit einem Distraktor präsentiert wird, genauso stark oder stärker aktiv sein wie während des Trainings der entsprechenden VTU auf dieses Zielobjekt (VTUs werden stets auf Stimuli mit 100% Kontrast trainiert), und eine weitere Erhöhung ihrer Aktivität durch Aufmerksamkeit wird die Aktivität der auf das Zielobjekt trainierten VTU nur senken können. Entsprechend zeigen die Abbildungen 5-2 bis 5-5 stets ein Sinken der Erkennungsleistung, falls Aufmerksam-

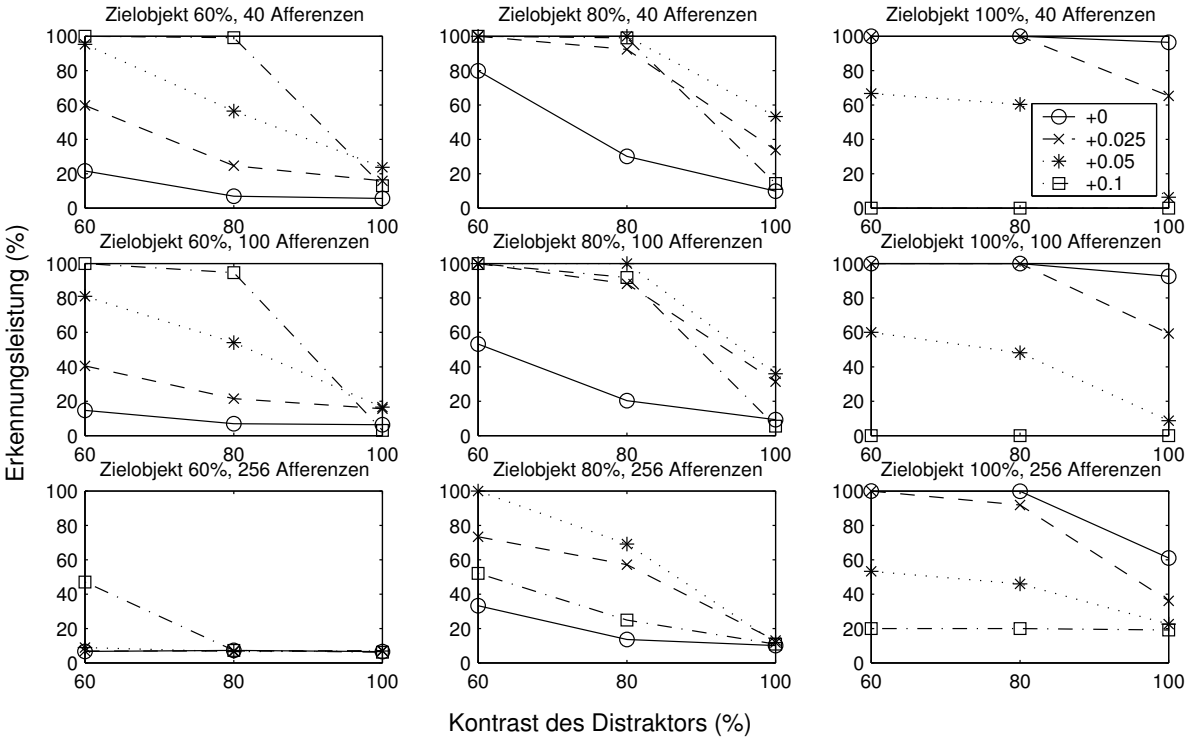

Abbildung 5-4: Additive Aufmerksamkeitseffekte bei C2-Zellen. Erkennungsleistung im Aktivste VTU-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer, gemittelt über alle Zielobjekte bei festem Kontrast des Distraktors (Abszissenwert). Kontrast des Zielobjekts und Anzahl der C2-Afferenzen pro VTU über den Diagrammen. Legende oben rechts: Werte, die als Aufmerksamkeitseffekt zu den Aktivitäten der C2-Afferenzen der auf das Zielobjekt trainierten VTU addiert wurden.

keitseffekte bei der Präsentation von Zielobjekten mit 100% Kontrast angewendet werden. Dies gilt auch für alle anderen möglichen Aufmerksamkeitseffekte, die die Aktivität von C2-Zellen erhöhen, und wird daher im folgenden nicht weiter diskutiert. Aufmerksamkeitseffekte bei hohen Kontrasten des Zielobjekts werden ausserdem im Experiment nicht beobachtet [52]. Weiterhin wird der Fall, daß VTUs mit allen 256 C2-Zellen verbunden sind, üblicherweise ausgelassen, da hier Aufmerksamkeit auf Ebene der C2-Zellen nicht spezifisch auf bestimmte Stimuli wirken kann; jede Modulation ihrer Aktivität wird sich gleichermaßen auf alle VTUs auswirken. Diese Situation wird jedoch zum Vergleich herangezogen, um herauszufinden, ob unspezifische Effekte sich auch auf die Erkennungsleistung auswirken können (s.u.).

Die Abbildungen zeigen zunächst, daß in der Tat die Erkennungsleistung von HMAX durch Anwendung von Aufmerksamkeit auf ein zu erkennendes Zielobjekt verbessert

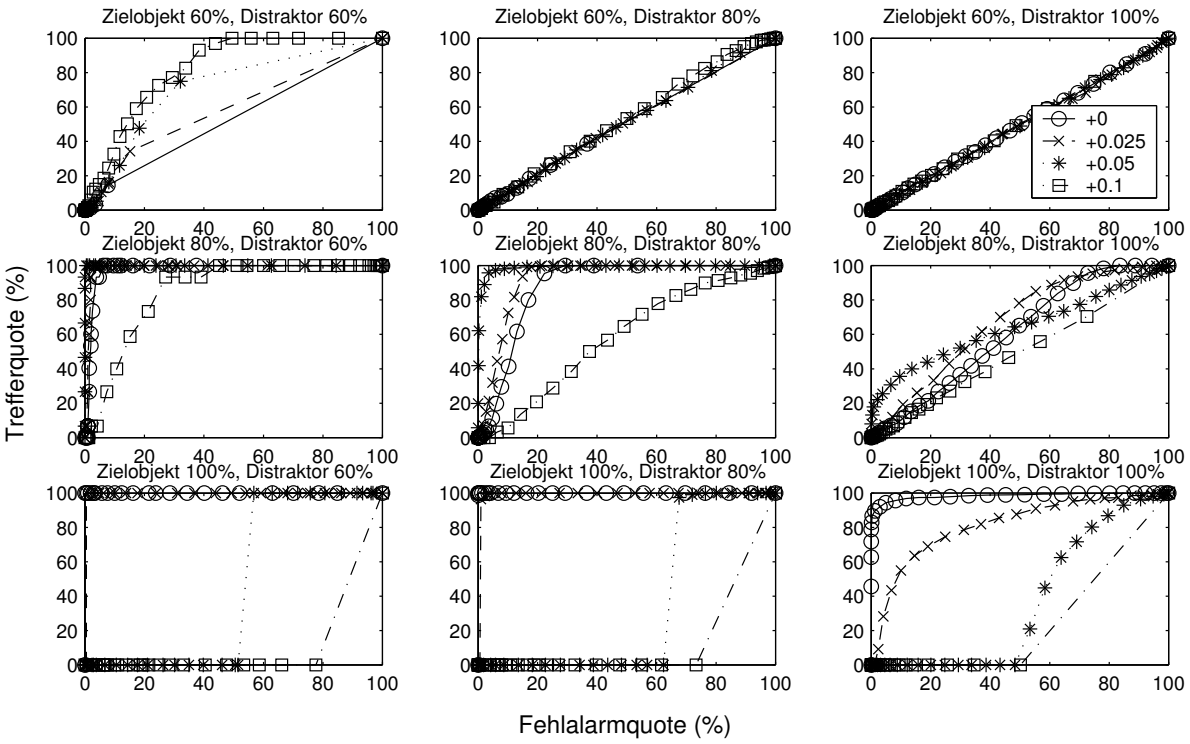

Abbildung 5-5: Additive Aufmerksamkeitseffekte bei C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen). Gleiche Daten wie in Abb. 5-4, aber nur für 40 Afferenzen pro VTU.

werden kann. Für bestimmte Kombinationen aus Stimuluskontrast und Stärke des Aufmerksamkeitseffekts (für Autos z.B. 80% Kontrast und eine Aktivitätserhöhung von 0.025) nehmen sowohl die Werte der Erkennungsleistung im Aktivste VTU-Paradigma als auch die Fläche unter der ROC-Kurve im Stimulusvergleich-Paradigma deutlich zu. Das heißt, mit Aufmerksamkeit auf ein Objekt feuert die auf dieses Objekt trainierte VTU häufiger stärker als alle anderen VTUs, wenn dieses Objekt präsentiert wird, und ihre Antworten sind auch selektiver für ihr Zielobjekt, indem sie eher mit einer höheren Feuerrate auf die Präsentation ihres Zielobjekts als auf das Zeigen eines anderen Objekts reagiert. Es zeigt sich jedoch auch, daß der Erfolg dieses Aufmerksamkeitsmechanismus stark davon abhängt, ob die Stärke der Modulation für das jeweilige Kontrastniveau des Stimulus angemessen ist. Für einen bestimmten Kontrast des Zielobjekts erhöht jeweils nur ein kleiner Bereich möglicher Modulationsstärken tatsächlich die Erkennungsleistung; andere Werte sind entweder zu klein, um die relative Aktivität der VTUs zu verändern, oder so groß,

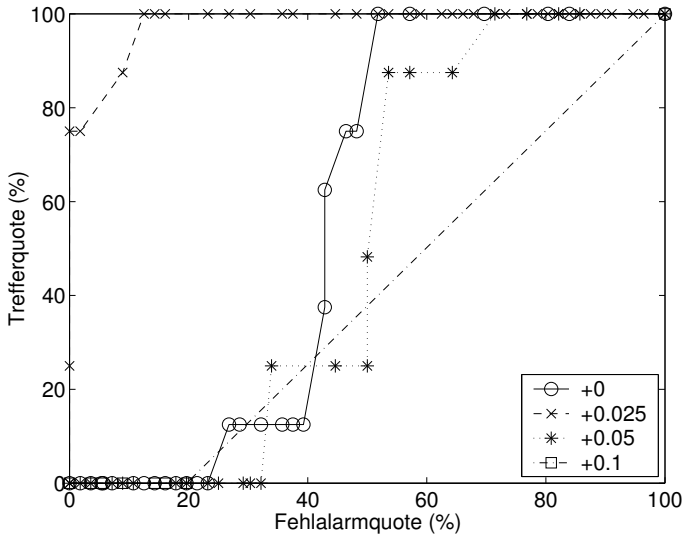

Abbildung 5-6: Additive Aufmerksamkeitseffekte bei C2-Zellen und 256 Afferenzen pro VTU. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte (80% Kontrast) in Gegenwart eines Distraktor-Autos (60% Kontrast). Da jede VTU mit allen 256 C2-Zellen verbunden war, erhielten alle C2-Zellen denselben additiven Aufmerksamkeitseffekt (siehe Legende). Vgl. die ROC für den Wert 0.025 mit der entsprechenden ROC für 40 Afferenzen in Abb. 5-3 (Diagramm Mitte links).

daß die Afferenzen der auf das Zielobjekt trainierten VTU aufgrund des Aufmerksamkeitseffekts stärker feuern als während des Trainings dieser VTU auf ihr Zielobjekt, so daß die Aktivität dieser VTU und die Erkennungsleistung wieder absinken. Da die VTUs in HMAX auf bestimmte Werte der Aktivität ihrer C2-Afferenzen trainiert sind und diese Zellen in Abhängigkeit vom Kontrast des Stimulus feuern, ist klar, daß nicht ein und dieselbe Stärke eines Aufmerksamkeitseffekts für alle Stimuluskontraste gleich gute Ergebnisse liefert. In einem Paradigma der frühen Selektion von Merkmalen durch Aufmerksamkeit bleibt damit aber die Frage offen, wie im voraus entschieden werden kann, wie stark die Aktivität relevanter Neuronen moduliert werden soll.

Weiterhin ist die hier vorgestellte Form von Aufmerksamkeit wenig geeignet, um die Auswirkungen eines Distraktorstimulus zu kompensieren, dessen Kontrast höher ist als der des Zielobjekts. Da aufgrund der MAX-Operation, wie bereits diskutiert, ein Distraktor höheren Kontrasts die Aktivität der C2-Afferenzen, die auf die VTU des Zielobjekts projizieren, nur erhöhen kann, ist eine zusätzliche Erhöhung der Aktivität dieser Zellen nicht in der Lage, den Effekt eines solchen Hintergrundstimulus zu kompensieren. Dennoch er-

geben sich im Aktivste VTU-Paradigma auch Verbesserungen der Erkennungsleistung für höhere Werte des Distraktorkontrasts (siehe Abb. 5-2 und 5-4). Es muß jedoch angemerkt werden, daß in diesem Paradigma nicht die möglichen Fehlalarme durch Aufmerksamkeitseffekte berücksichtigt werden, wie oft also ein Stimulus, auf den die Aufmerksamkeit gerichtet wird, der dann aber nicht im Bild erscheint, dennoch „erkannt“ wird (siehe dazu Kapitel 5.3.8). Die ROC-Kurven andererseits, in deren Berechnung Fehlalarme im Stimulusvergleich-Paradigma explizit eingehen (siehe Methoden), zeigen keine signifikanten Zunahme der Erkennungsleistung für Distraktoren, deren Kontrast den des Zielobjekts übersteigt (siehe dazu aber auch 5.3.3).

Abbildung 5-6 macht schließlich deutlich, daß in manchen Fällen eine unspezifische Erhöhung der Aktivität *aller* C2-Zellen zu besserer Leistung bei der Objekterkennung führt als ein selektiver Aufmerksamkeitseffekt. Hier waren alle VTUs mit allen 256 C2-Zellen verbunden, und alle C2-Zellen wurden jeweils um denselben Betrag in ihrer Aktivität verstärkt. Abbildung 5-4 entspricht zwar der Intuition, daß eine höhere Anzahl Afferenzen und weniger spezifisch wirkende Aufmerksamkeitseffekte die Leistung bei der Objekterkennung verringern. Um zu erreichen, daß eine VTU stärker feuert als andere, worauf es im in Abbildung 5-4 verwendeten Aktivste VTU-Paradigma ankommt, ist es natürlich effektiver, wenn die VTU nur mit einer Auswahl der C2-Zellen verbunden ist und nur diese C2-Zellen verstärkt feuern. Diese Art der Messung von Erkennungsleistung geht aber wiederum nicht auf mögliche Fehlalarme ein. Das Stimulusvergleich-Paradigma in Abbildung 5-6 zeigt dagegen, daß die Selektivität der Antworten einer VTU auf verschiedene Stimuli offensichtlich zumindest in einigen Fällen auch durch eine einfache allgemeine Erhöhung des effektiven Stimuluskontrasts verbessert werden kann. In solchen Fällen scheint ein selektiver Aufmerksamkeitsmechanismus gar nicht erforderlich zu sein.

Insgesamt erscheinen die Resultate für einen Mechanismus von Aufmerksamkeit, der die Feuerraten derjenigen C2-Zellen, die auf Merkmale des bewußt beachteten Stimulus reagieren, um konstante Werte erhöht, nicht allzu vielversprechend. Die Erkennungsleistung kann zwar deutlich erhöht werden, allerdings nur, wenn der Betrag der Modulation der Feuerraten zum jeweiligen Kontrast des Zielobjektes paßt und der Distraktorstimulus nicht zu kontrastreich ist. Außerdem scheint die Selektivität der neuronalen Antworten auch von einer unspezifischen Erhöhung der Feuerraten aller Zellen zu profitieren. Diese Ergeb-

nisse können aber zum Teil auch dadurch bedingt sein, daß hier alle ausgewählten Zellen denselben Betrag an Modulation ihrer Aktivität erhalten haben. Verschiedene C2-Zellen feuern schließlich unterschiedlich stark, wenn ein Stimulus gezeigt wird, und die Addition derselben Konstanten zu allen Aktivitätswerten wird die Auswirkungen von niedrigem Kontrast und Hintergrundstimuli nicht systematisch kompensieren können. Eine Modulation, die besser auf die unterschiedlichen Feuerraten der Zellen bzw. auf den Stimuluskontrast abgestimmt ist, könnte hier bessere Resultate liefern. Dies wird in den folgenden Abschnitten untersucht.

5.3.2 Multiplikative Aufmerksamkeitseffekte

Anstatt die Feuerraten der C2-Zellen durch Addition einer Konstanten zu erhöhen, könnte ihre Aktivität auch mit einem Faktor größer als 1 multipliziert werden. Dies würde die Aktivität von Zellen, die bereits stark aktiv sind, absolut gesehen stärker erhöhen als die von schwächer feuernden Zellen. Während ein solches Modell mit der Hypothese von McAdams und Maunsell, daß Aufmerksamkeit die Aktivität von Neuronen im Areal V4 multiplikativ erhöht [41], übereinstimmen würde, widerspräche es der allgemeineren und experimentell bestätigten Annahme, daß Aufmerksamkeit zwar die *Abstimmkurven* von Neuronen auf verschiedene Stimuli multiplikativ beeinflußt, aber ihre *Kontrastantwortfunktion* zu kleineren Kontrastwerten verschiebt und so die Aktivitäten von schwächer feuernden Neuronen am stärksten zunehmen (siehe 5.1.4 und [52]). Dennoch sollen hier die Auswirkungen von multiplikativ wirkenden Aufmerksamkeitseffekten auf die Objekterkennung in HMAX untersucht werden. Im anschließenden Abschnitt 5.3.3 werden die Resultate dann mit den Ergebnissen für eine Modulation verglichen, die einer Verschiebung der Kontrastantwortkurve entspricht.

Die Effekte einer multiplikativen Verstärkung der Feuerraten von C2-Zellen, die auf die VTU des bewußt beachteten Zielobjekts projizieren, sind in Abb. 5-7 bis 5-10 gezeigt, wiederum für Autos und Büroklammern im Aktivste VTU- und im Stimulusvergleich-Paradigma. Eine Zunahme der neuronalen Aktivität um den Faktor 1.3 entspricht dabei dem von McAdams und Maunsell über eine Stimulusdauer von 500 ms im Mittel beobachteten Effekt von Aufmerksamkeit [41], während Motter Neuronen findet, deren Feuerraten sich durch Aufmerksamkeit bis auf das Doppelte erhöhen [46]. Die Resultate in HMAX sind

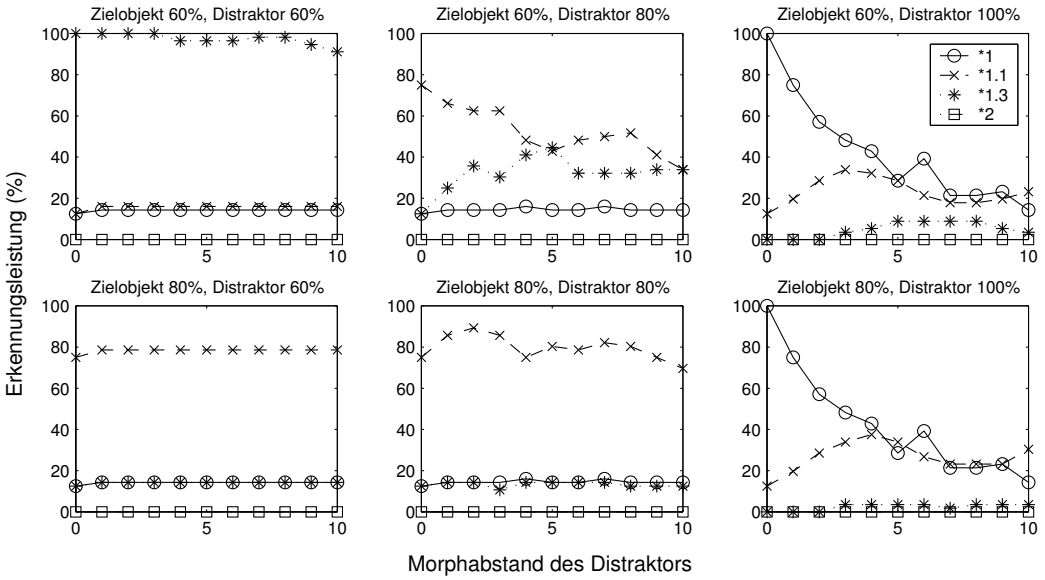

Abbildung 5-7: Multiplikative Aufmerksamkeitseffekte bei C2-Zellen. Erkennungsleistung im Aktivste VTU-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die Faktoren an, um die jeweils die Aktivitäten der C2-Afferenzen der auf das Zielobjekt trainierten VTU erhöht wurden. Jede VTU hat 40 C2-Afferenzen.

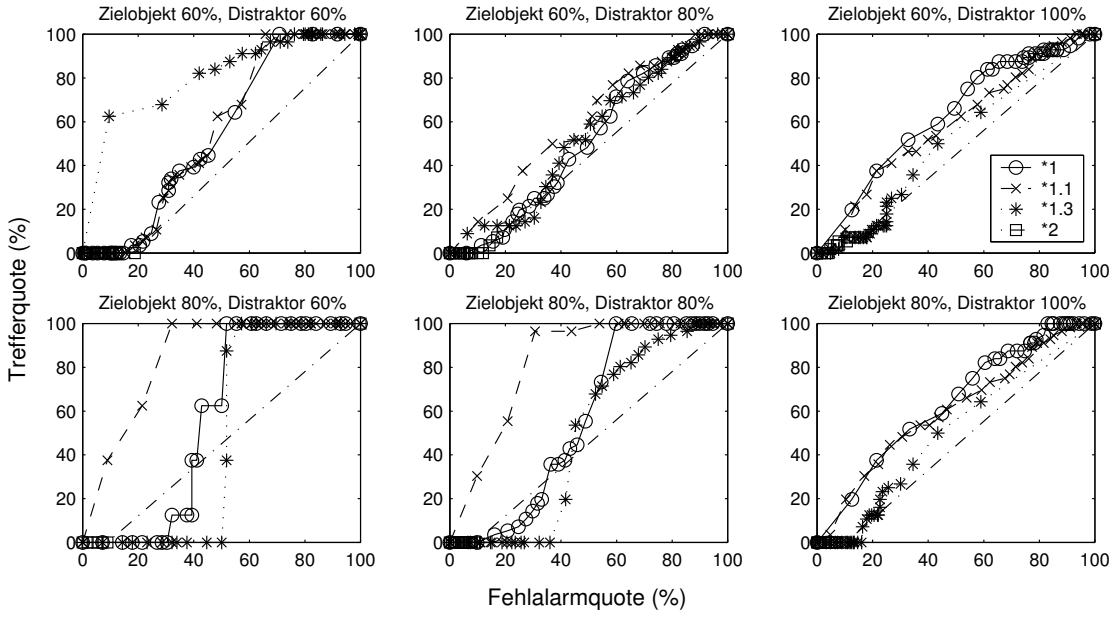

Abbildung 5-8: Multiplikative Aufmerksamkeitseffekte bei C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-7, aber nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

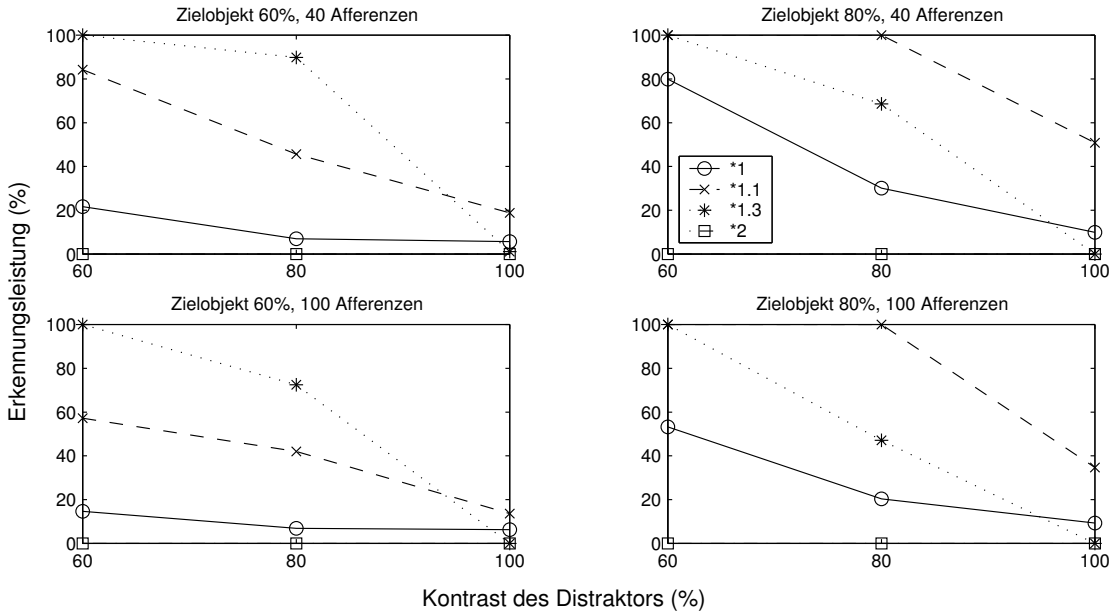

Abbildung 5-9: Multiplikative Aufmerksamkeitseffekte bei C2-Zellen. Erkennungsleistung im Aktivste VTU-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer, gemittelt über alle Zielobjekte bei festem Kontrast des Distraktors (Abszissenwert). Kontrast des Zielobjekts und Anzahl der C2-Afferenzen pro VTU über den Diagrammen. Die Legende oben rechts gibt die Faktoren an, um die die Feuerraten der C2-Zellen erhöht wurden, die auf die VTU des Zielobjekts projizierten.

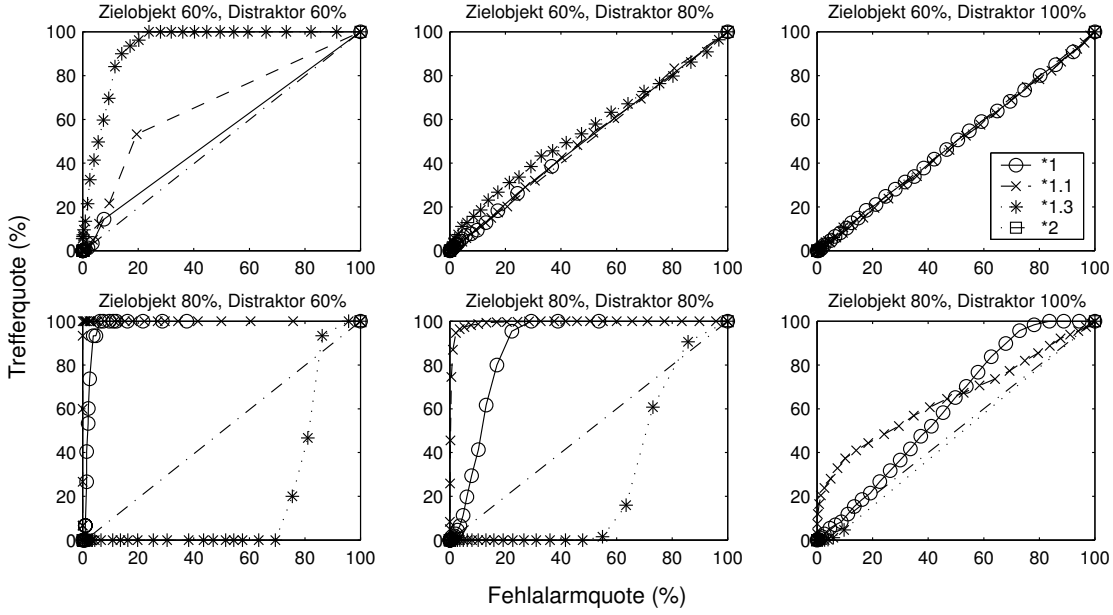

Abbildung 5-10: Multiplikative Aufmerksamkeitseffekte bei C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen). Gleiche Daten wie in Abb. 5-9, aber nur für 40 Afferenzen pro VTU.

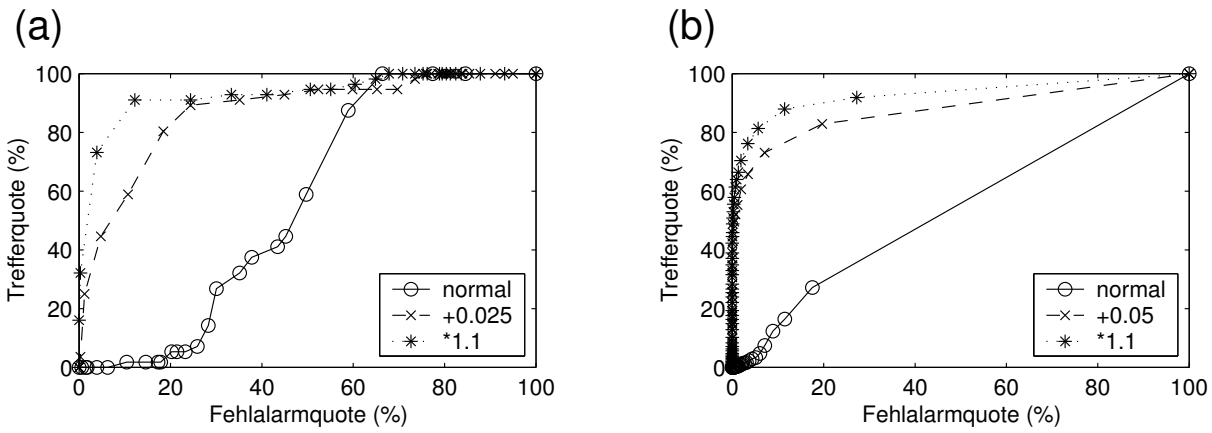

Abbildung 5-11: Auswirkungen von Aufmerksamkeitseffekten auf die Erkennungsleistung, wenn alle 256 C2-Zellen auf alle VTUs projizieren und alle C2-Zellen in ihrer Aktivität verstärkt werden. **(a)** ROCs für die Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (beide mit 80% Kontrast), gemittelt über alle Zielobjekte, für additive und multiplikative Modulationen aller C2-Zellen (siehe Legende). **(b)** Wie (a), jedoch für Büroklammern. Kontrast von Zielobjekt und Distraktor 80%.

qualitativ denen für konstante additive Aufmerksamkeitseffekte sehr ähnlich. Wiederum ergeben bestimmte Kombinationen von Stimuluskontrast und Stärke des Aufmerksamkeitseffekts eine deutlich höhere Leistung des Modells bei der Erkennung beider Objektklassen in beiden Paradigmen (etwa der Faktor 1.3 bei 60% Kontrast und 1.1 bei 80% Kontrast; eine Verstärkung um den Faktor 2 stellt sich für die hier verwendeten Kontraste als zu stark heraus). Es gelten jedoch auch dieselben Einschränkungen wie in 5.3.1. Ebenso wie dort gilt, daß die Stärke der Modulation genau an den jeweiligen Kontrast des Zielobjekts angepaßt sein muß, um die Leistung erhöhen zu können. Die störenden Effekte eines Distraktors, dessen Kontrast höher ist als der des Zielobjekts, können auch hier nur ungenügend wettgemacht werden, zumindest im Stimulusvergleich-Paradigma. Schließlich zeigt sich auch mit multiplikativer Verstärkung der neuronalen Aktivität, daß selbst für VTUs mit allen 256 C2-Zellen als Afferenzen und für eine Erhöhung der Feuerrate aller dieser Zellen die Erkennungsleistung deutlich zunehmen kann, noch deutlicher als für additive Verstärkung oder sogar für spezifischer wirkende Aufmerksamkeitseffekte (siehe Abb. 5-11 im Vergleich zu Abb. 5-8). Dies deutet wiederum darauf hin, daß zur Erhöhung der Selektivität neuronaler Antworten und daher zur Verbesserung der Leistung in einem simulierten „two-alternative forced choice"-Experiment eine unspezifische Kompensation für geringen Stimuluskontrast möglicherweise genauso gut geeignet ist wie ein Mecha-

nismus selektiver Verstärkung der Aktivität nur derjenigen Zellen, die auf Merkmale des bewußt beachteten Stimulus ansprechen.

Eine multiplikative Verstärkung der Aktivität von C2-Zellen scheint insgesamt besser an deren Antworteigenschaften angepaßt zu sein als konstante additive Modulationen. Solange die Feuerraten der C2-Zellen in einem Bereich liegen, in dem eine Zunahme des Stimuluskontrasts eine lineare Zunahme der Aktivität bewirkt, wie es bei den hier verwendeten Stimuli stets der Fall ist (siehe Kapitel 3.2.3), entspricht eine Multiplikation ihrer Feuerraten eher einer Erhöhung des Stimuluskontrasts. Verschiedene C2-Zellen antworten auf einen Stimulus schließlich unterschiedlich stark, und der absolute Zuwachs an Aktivität mit zunehmendem Kontrast ist bei solchen Zellen größer, die auf diesen Stimulus stärker ansprechen. Dasselbe wird durch multiplikative Modulationen erreicht. Wohl deshalb werden mit multiplikativen Aufmerksamkeitseffekten in HMAX etwas bessere Resultate erzielt als durch Addition konstanter Werte (siehe Abb. 5-11). Dennoch bleibt die effektive Anwendung von Aufmerksamkeit weiterhin auf Situationen beschränkt, in denen der Distraktorstimulus nicht kontrastreicher ist als das Zielobjekt und die Stärke der Modulation der neuronalen Aktivität auf den Kontrast des Zielobjekts abgestimmt ist. Ferner erhöhen gezielte Aufmerksamkeitseffekte auch hier die Selektivität der neuronalen Antworten nicht mehr als unspezifische Aktivitätserhöhungen aller C2-Zellen.

5.3.3 Verschiebung der Kontrastantwortkurve als Aufmerksamkeitseffekt

Wie in den Methoden beschrieben, wurde als weiterer möglicher Effekt von Aufmerksamkeit auf die Aktivität von Neuronen die von Reynolds *et al.* beschriebene Verschiebung der Kontrastantwortkurve derjenigen Zellen, die an der Repräsentation des bewußt beachteten Stimulus beteiligt sind, modelliert [52]. Dazu wurde für diese Zellen der Mittelwert *s2Target* der Gaußfunktion, die die Antwort der S2- bzw. C2-Zellen auf ihren C1-Input bestimmt, verkleinert. Diese Modulation dürfte also besser als die Addition einer Konstanten oder die Multiplikation mit einem Faktor eine Zunahme des Stimuluskontrasts nachbilden und sollte daher die Aktivität der C2-Zellen auf eine „natürlichere“ Weise verändern. Die verwendeten neuen Werte für *s2Target* wurden so gewählt, daß die dadurch erreichten Zunahmen der Feuerraten der C2-Zellen im Mittel die Abnahme ihrer Aktivitäten bei einer Senkung des Kontrasts von Autostimuli von 100% auf 60% kompensierten. Analog zu

den Versuchen mit additiven und multiplikativen Aufmerksamkeitseffekten wurde diese Modulation dann auf die C2-Afferenzen der auf das jeweilige Zielobjekt trainierten VTU angewendet. Aufgrund der Wahl der Werte für *s2Target* wurden dabei als Zielobjekte nur Autostimuli mit einem Kontrast von 60% verwendet.

Abbildungen 5-12 und 5-13 zeigen die Resultate. Das Bild, das sich ergibt, ist dem für additive und multiplikative Aufmerksamkeitseffekte sehr ähnlich. Aufgrund der sorgfältig gewählten neuen Mittelwerte der C2-Kontrastantwortkurven erhöht sich die Erkennungsleistung in beiden Paradigmen deutlich. Sogar eine leichte Zunahme der Leistung im Stimulusvergleich-Paradigma für Distraktoren, deren Kontrast höher ist (80%) als der des Zielobjekts (60%), wurde festgestellt (Abb. 5-13 Mitte). Dies ist jedoch kein Widerspruch zu der Behauptung, daß Aufmerksamkeitsmechanismen, die die Aktivität von Neuronen erhöhen, grundsätzlich nicht die Effekte von Hintergrundstimuli mit höherem Kontrast kompensieren können. Solange sowohl Zielobjekt als auch Distraktor mit Kontrasten unter 100% gezeigt werden, erreichen C2-Zellen, die durch beide aktiviert werden, noch nicht ihre Feuerrate für 100% Kontrast. Eine zusätzliche Erhöhung ihrer Aktivität durch Aufmerksamkeit kann sie deshalb nach wie vor weiter dem Aktivitätsniveau annähern, mit dem sie auf die Präsentation des Zielobjekts bei 100% Kontrast antworten, und daher selektiv die VTU, die für das Zielobjekt codiert, begünstigen. Derselbe Effekt kann ebenso für multiplikative Aufmerksamkeitseffekte beobachtet werden, wie Abbildung 5-14 zeigt. Dies verdeutlicht wiederum, daß die Leistungen von HMAX bei der Objekterkennung in der Tat durch Aufmerksamkeitseffekte verbessert werden können, solange die Stärke der Modulation genau an den jeweiligen Stimuluskontrast angepaßt ist.

Ebenso wie für additive und multiplikative Modulationen erweist sich auch bei der Verschiebung der Kontrastantwortkurve die Anwendung des Effekts auf alle 256 C2-Zellen einem Effekt selektiver Aufmerksamkeit als nahezu ebenbürtig (Abb. 5-15). Die verbesserte Leistung für höheren Distraktorkontrast (80% gegenüber 60% beim Zielobjekt) konnte mit dieser unspezifischen Erhöhung des allgemeinen Aktivitätsniveaus allerdings nicht reproduziert werden. Eine Verbesserung der Objekterkennungsleistung für Distraktoren höheren Kontrasts, sofern sie denn möglich ist, erfordert also womöglich gezielter wirkende Mechanismen von Aufmerksamkeit.

Diese Resultate zeigen, daß eine Verschiebung der Kontrastantwortkurve von Neuronen,

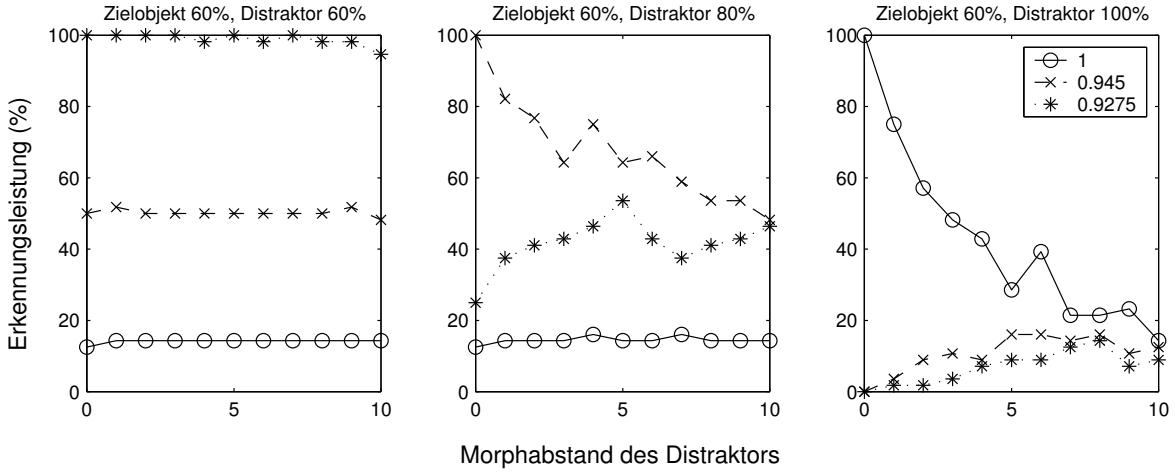

Abbildung 5-12: Verschiebung der Kontrastantwortkurve von C2-Zellen. Erkennungsleistung im Aktivste VTU-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten Diagramm gibt für die einzelnen Graphen in allen Diagrammen die Mittelwerte *s2Target* der Antwortkurven an, die für die 40 C2-Zellen verwendet wurden, die mit der VTU des Zielobjekts verbunden waren. Der normale Mittelwert (also ohne Aufmerksamkeitseffekt) ist 1.

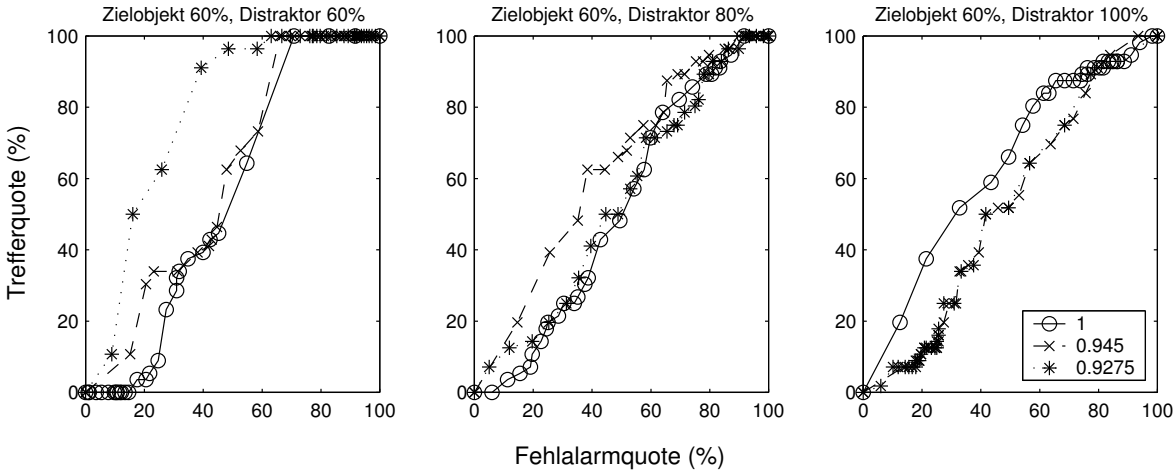

Abbildung 5-13: Verschiebung der Kontrastantwortkurve von C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Die Legende im rechten Diagramm zeigt die jeweils für die C2-Afferenzen der VTU des Zielobjekts verwendeten Werte des Parameters *s2Target*. Gleiche Daten wie in Abb. 5-12, jedoch nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

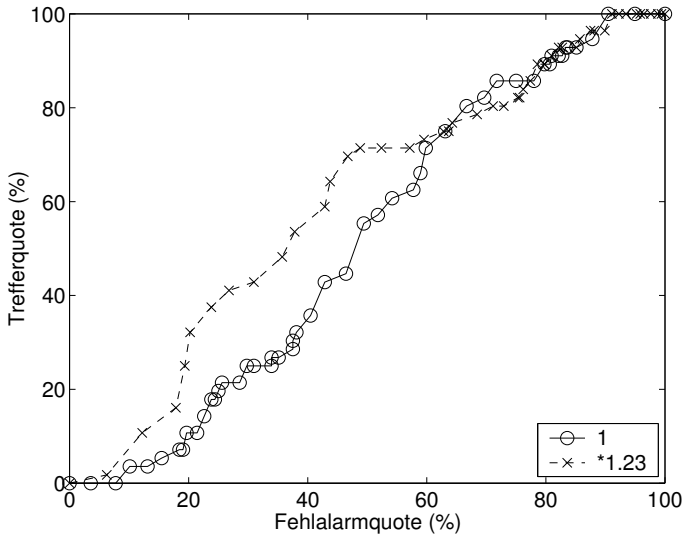

Abbildung 5-14: Erhöhung der Erkennungsleistung durch multiplikative Aufmerksamkeitseffekte für Distraktoren mit höherem Kontrast. ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma, über alle Zielobjekte gemittelt, für Auto-Zielobjekte (60% Kontrast) in Gegenwart eines Distraktor-Autos (80% Kontrast) und 40 Afferenzen pro VTU. Stärke des multiplikativen Aufmerksamkeitseffekts: Faktor 1.23.

wie sie von Reynolds *et al.* für V4 beschrieben wurde [52], auch in HMAX als Aufmerksamkeitseffekt verwendet werden kann und daß diese Modulation auf Ebene der neuronalen Aktivität in der Tat die Leistung des Modells bei der Objekterkennung verbessern kann. Im Modell ergibt sich jedoch kein qualitativer Unterschied zu den beiden zuvor getesteten Methoden, Aufmerksamkeit zu simulieren. Auch die Verschiebung der Kontrastantwortkurve muß zum Kontrast des zu erkennenden Zielobjekts hinreichend gut passen, was die Verwendung auch dieses Mechanismus für die frühe Selektion von Merkmalen, also die Vorbereitung des visuellen Systems auf erwartete Stimuli, sehr erschwert. Keine fundamentalen Unterschiede konnten auch hinsichtlich der Problematik kontrastreicher Distraktorstimuli oder der ähnlich guten Leistung einer unspezifischer Verstärkung der Aktivität aller C2-Zellen gefunden werden. Die bisher diskutierten möglichen Mechanismen der Modulation neuronaler Aktivität durch Aufmerksamkeit verhalten sich also im Kontext des HMAX-Modells in bezug auf ihre Auswirkungen auf die Objekterkennung alle sehr ähnlich. Weil abgesehen davon konstante additive Modulationen weniger gut auf unterschiedlich aktive C2-Zellen abgestimmt sind und die Verschiebung der Kontrastantwortkurve im Modell nur mit deutlich höherem Aufwand zu realisieren ist, werden in den folgenden Abschnitten stets multiplikative Aufmerksamkeitseffekte verwendet. Da

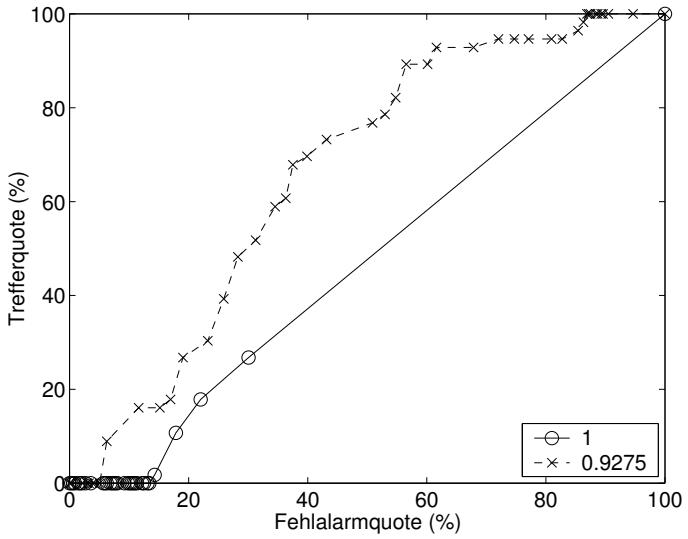

Abbildung 5-15: Verschiebung der Kontrastantwortkurve aller 256 C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontrast beider Stimuli 60%) und 256 Afferenzen pro VTU. Als Aufmerksamkeitseffekt wurde die Gaußsche Kontrastantwortkurve aller C2-Zellen auf den neuen Mittelwert 0.9275 verschoben.

für die Stimuli und Kontraste, die hier zum Einsatz kommen, die Aktivitäten aller C2-Zellen relativ nahe beieinanderliegen und linear mit dem Stimuluskontrast skalieren (siehe auch Kapitel 3.2.3), sind insbesondere die Effekte von Multiplikation und Verschiebung der Kontrastantwortkurve nahezu identisch.

5.3.4 Suppression

Bisher wurde auf ein Zielobjekt gerichtete Aufmerksamkeit in HMAX nur dadurch modelliert, daß die Aktivität von Neuronen, die für Merkmale des Zielobjektes codieren, erhöht wurde. In Anbetracht der in 5.1.4 erwähnten zahlreichen Evidenzen dafür, daß Neuronen, die für jeweils gerade nicht mit Aufmerksamkeit bedachte Stimuli selektiv sind, in ihrer Aktivität abgeschwächt werden, wurden auch entsprechende Suppressionsmechanismen in HMAX untersucht. Dies geschah wiederum vor allem im Hinblick darauf, was ein solcher Mechanismus für die Objekterkennung bedeuten könnte. Bei diesen Simulationen wurde, wie bereits in den Methoden erwähnt, in Kauf genommen, daß der eine der beiden verwendeten Suppressionsmechanismen strenggenommen kein Mechanismus der frühen Selektion sein kann, weil dabei die C2-Afferenzen der neben der VTU des Zie-

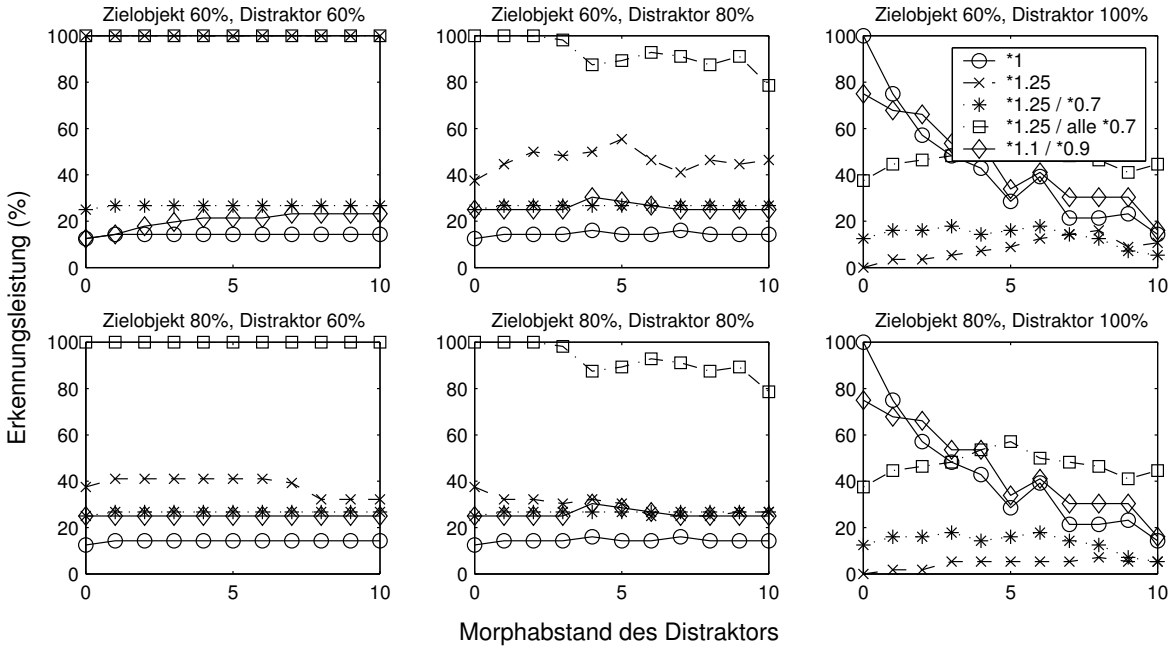

Abbildung 5-16: Multiplikative Aufmerksamkeitseffekte und Suppression bei C2-Zellen. Erkennungsleistung im Aktivste VTU-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die Werte der multiplikativen Verstärkung und der Suppression an (erster Wert: Verstärkung der Aktivität der C2-Zellen, die auf die VTU projizieren, deren bevorzugter Stimulus das Zielobjekt ist; zweiter Wert, wenn vorhanden: Suppression aller übrigen C2-Zellen („alle“) oder der Afferenzen der am stärksten aktiven nicht auf das Zielobjekt trainierten VTU (sonst)). Alle VTUs haben 40 Afferenzen.

lobjekts am stärksten aktiven VTU supprimiert werden, deren Identität erst nach einem vollständigen Lauf des Modells feststeht. Ebenfalls hingenommen wurde, daß der zweite untersuchte Suppressionsmechanismus nur begrenzt realistisch zu sein scheint, weil er sich auf *alle* C2-Zellen auswirkt, die nicht an der Repräsentation des Zielobjekts beteiligt sind. Grundsätzlich anders geartete Suppressionsmechanismen wären jedoch nur schwer vorstellbar, und die beiden hier ausgewählten hatten zudem den Vorteil, daß von ihnen auf die eine oder andere Weise optimale Leistung erwartet werden konnte. Ein Mechanismus der Suppression von neuronalen Aktivitäten, der das Ziel hat, die Repräsentation eines bestimmten Zielobjekts selektiv zu begünstigen, sollte besonders gut funktionieren, wenn entweder bekannt ist, welche Neuronen am stärksten mit der Repräsentation dieses Objekts interferieren, oder wenn alle nicht unmittelbar beteiligten Neuronen in ihrer Aktivität abgeschwächt werden. Wenn also durch die Suppression neuronaler Aktivitäten

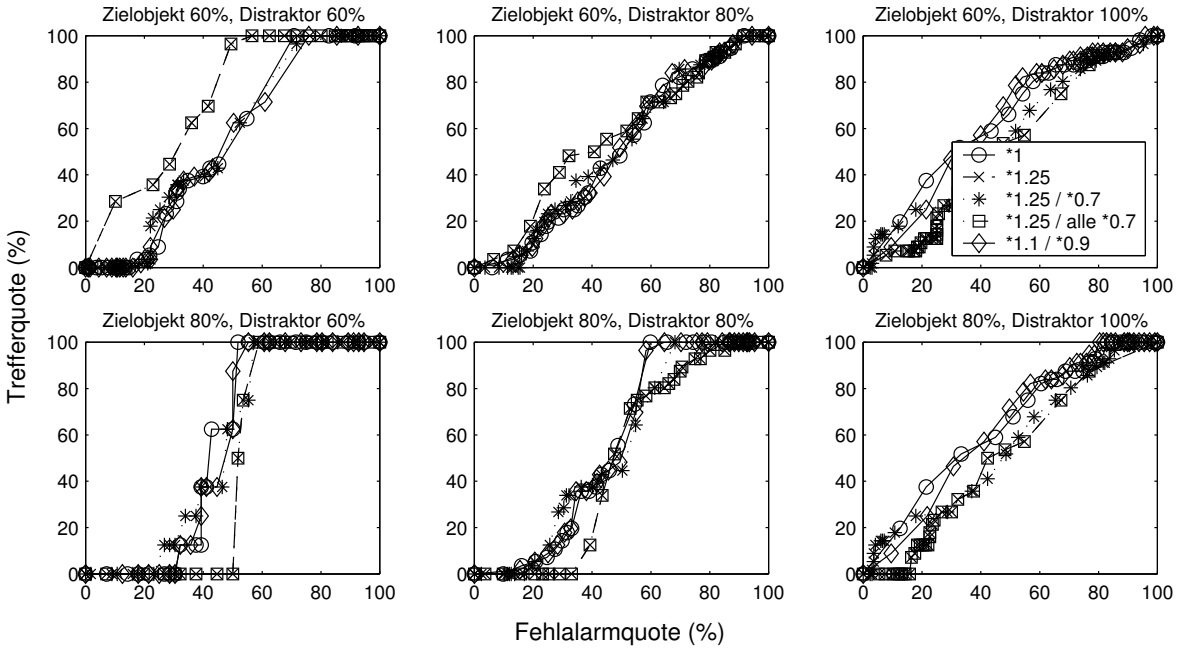

Abbildung 5-17: Multiplikative Aufmerksamkeitseffekte und Suppression bei C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-16, jedoch nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

die Objekterkennungsleistung verbessert werden kann, sollte es an diesen Mechanismen demonstrierbar sein.

Abbildungen 5-16 bis 5-19 machen zunächst deutlich, daß sich die ROC-Kurven identisch sind, unabhängig davon, ob zusätzlich zu einem Aufmerksamkeitseffekt, der die Feuerraten der C2-Afferenzen der VTU des Zielobjekts erhöht, noch die Aktivitäten aller übrigen C2-Zellen abgeschwächt werden oder nicht. Das liegt schlicht daran, daß für die Erkennungsleistung im Stimulusvergleich-Paradigma jeweils nur die Aktivität der auf das Zielobjekt trainierten VTU bei verschiedenen Stimuli entscheidend ist; die Feuerraten von C2-Zellen, mit denen sie gar nicht verbunden ist, haben darauf natürlich keinen Einfluß. Andererseits führt die Verwendung von Suppression, wenn sie auf die C2-Afferenzen der neben der VTU des Zielobjekts am stärksten aktiven VTU angewendet wird, in den meisten Fällen sogar zu schlechterer Leistung im Stimulusvergleich-Paradigma, sowohl für Autos als auch für Büroklammern. Der Grund dafür ist, daß die VTU des Zielobjekts und die am stärksten feuernde unter den übrigen VTUs in der Regel einige C2-Afferenzen ge-

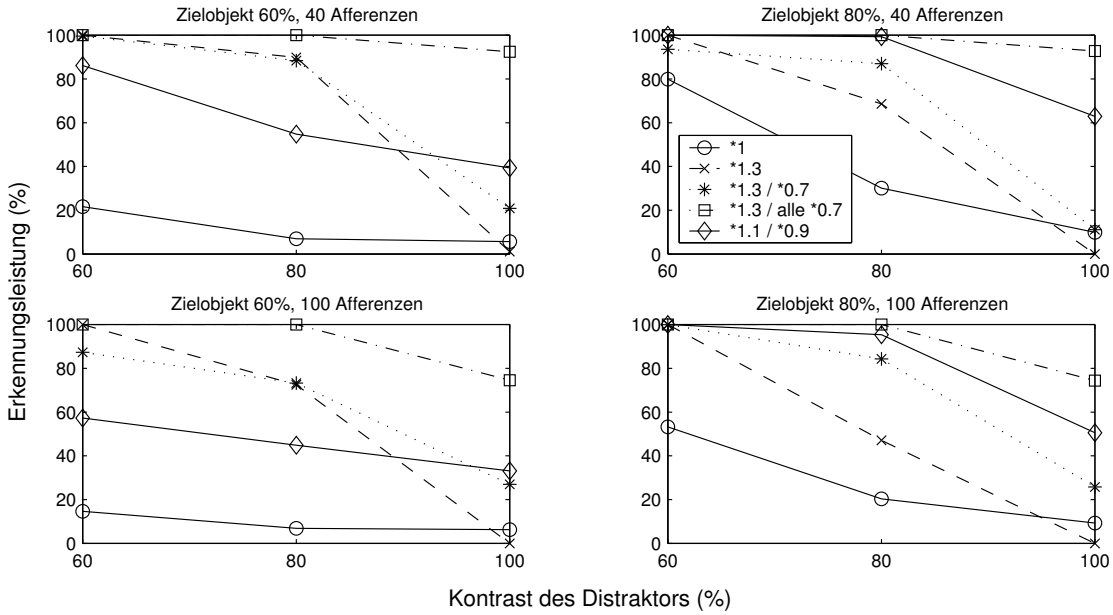

Abbildung 5-18: Multiplikative Aufmerksamkeitseffekte und Suppression bei C2-Zellen. Erkennungsleistung im Aktivste VTU-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer, gemittelt über alle Zielobjekte bei festem Kontrast des Distraktors (Abszissenwert). Kontrast des Zielobjekts und Anzahl der C2-Afferenzen pro VTU über den Diagrammen. Die Legende oben rechts gibt die Werte der jeweils verwendeten multiplikativen Verstärkung und Suppression an, wie in Abb. 5-16.

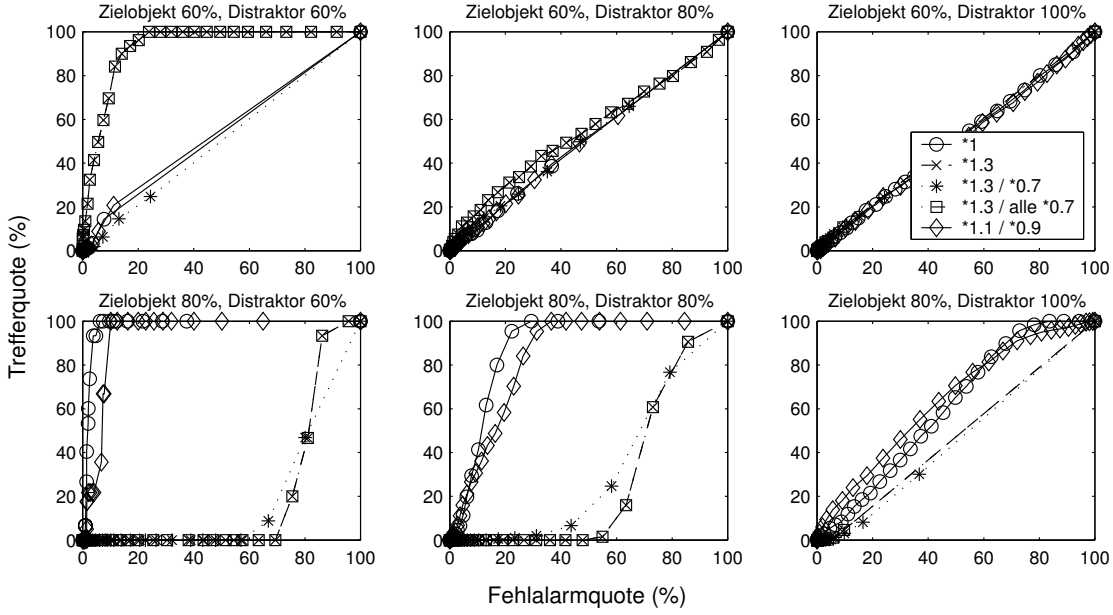

Abbildung 5-19: Multiplikative Aufmerksamkeitseffekte und Suppression bei C2-Zellen. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen). Gleiche Daten wie in Abb. 5-18, jedoch nur für 40 Afferenzen pro VTU.

meinsam haben und daher die Suppression sich auch auf C2-Zellen auswirkt, die mit der VTU des Zielobjekts verbunden sind. Insbesondere für die Autos, wo verschiedene VTUs zum großen Teil dieselben Afferenzen haben, wird dadurch der Aufmerksamkeitseffekt auf die Erkennungsleistung nahezu zunichte gemacht.

Im Aktivste VTU-Paradigma hingegen, wo es auf die Aktivität einer VTU im Vergleich zu *anderen* VTUs beim Zeigen *desselben* Stimulus ankommt, stellt sich die Kombination von Aktivitätserhöhung und Suppression durch Aufmerksamkeit als sehr effektiv heraus. Für die Büroklammern ergibt eine Erhöhung der Feuerrate der auf die VTU des Zielobjekts projizierenden C2-Zellen, vereint mit der Suppression aller übrigen C2-Zellen, überall eine hervorragende Erkennungsleistung. Aber auch die Suppression der Afferenzen der am stärksten aktiven VTU unter den VTUs, die nicht für das Zielobjekt codieren, stellt sich als effektiv heraus. Obwohl auch hier verschiedene VTUs einige C2-Afferenzen gemeinsam haben, sind die Mengen der Afferenzen unterschiedlicher VTUs offensichtlich verschieden genug, um der VTU des Zielobjekts bei dieser Methode insgesamt einen Vorteil zu verschaffen, obwohl einige ihrer Afferenzen sowohl von Aktivitätserhöhung als auch Suppression betroffen sind.

Bei den Autos ist ebenfalls die Methode, alle nicht mit der VTU des Zielobjekts verbundenen C2-Zellen zu supprimieren, sehr effektiv, um die Erkennungsleistung im Aktivste VTU-Paradigma zu erhöhen. Werden jedoch die Afferenzen der neben der VTU des Zielobjekts am stärksten aktiven VTU in ihrer Feuerrate abgeschwächt, so wirkt sich das aufgrund dessen, daß Auto-VTUs sehr viel mehr C2-Afferenzen gemeinsam haben, auch in diesem Paradigma der Erkennungsleistung dahingehend aus, daß die Leistung mit dieser Form von Suppression niedriger ist als mit einem Aufmerksamkeitseffekt auf die Afferenzen der VTU des Zielobjekts alleine. (Eine Ausnahme stellen die Fälle für den Distraktor mit 100% Kontrast in Abb. 5-16 dar; hier heben sich Aktivitätserhöhung und Suppression allerdings soweit gegenseitig auf, daß in etwa die Leistung ohne Aufmerksamkeit wiederhergestellt wird.)

Insgesamt scheint also die Kombination von multiplikativen bzw. kontrasterhöhenden Aufmerksamkeitsmechanismen mit entsprechender Suppression von Neuronen, die nicht an der Repräsentation des Zielobjekts teilnehmen, die Objekterkennung deutlich zu verbessern, wenn es wie im Aktivste VTU-Paradigma darauf ankommt zu entscheiden, wel-

ches Objekt aus einer Reihe von Alternativen in einer Darstellung mit niedrigem Kontrast und störendem Hintergrund gezeigt wird. Dafür muß jedoch angenommen werden, daß der Suppressionsmechanismus eine sehr breite Wirkung entfaltet – in diesen Simulationen auf alle nicht mit der VTU des Zielobjekts verbundenen C2-Zellen. Es ist fraglich, ob solch ein Mechanismus biologisch plausibel sein könnte; möglicherweise denkbar wäre das in dem Fall, daß alle möglichen Stimuli in einem Experiment aus einer begrenzten Menge stammen und so die Suppression zumindest auf eine gewisse Anzahl Neuronen, die für die Codierung dieser Stimulusklasse besonders relevant sind, eingeschränkt werden kann. Eine generelle Abschwächung der Aktivität aller nicht direkt an der Wahrnehmung eines Objekts beteiligten Zellen etwa innerhalb eines visuellen Areals wie V4 scheint nicht realistisch, alleine schon wegen des nötigen hohen Energieaufwands für einen so weitreichenden Mechanismus, und wohl auch nicht notwendig, da auch ohne den Einsatz von Aufmerksamkeit Neuronen im realen visuellen System in der Regel spezifischer aktiv sind als die Modellneuronen in HMAX und daher ohnehin weniger stark feuern, wenn ein Stimulus nicht ihrer Präferenz entspricht.

Abschwächung der Aktivität von Neuronen, die nicht für das Zielobjekt codieren, ist jedoch nicht geeignet, um die Selektivität der Neuronen, die auf das Zielobjekt ansprechen, zu erhöhen, wie an den ROC-Kurven gezeigt wurde. Das Problem, daß die Stärke der Modulation der Aktivität derjenigen Zellen, die an der Repräsentation des Zielobjekts beteiligt sind, im voraus an den Stimuluskontrast angepaßt werden müßte, bleibt auch bei Verwendung von Suppressionsmechanismen erhalten. Und schließlich darf die erfolgreiche Anwendung von Aufmerksamkeitsmechanismen mit Verstärkung und Suppression neuronaler Aktivitäten im Aktivste VTU-Paradigma nicht darüber hinwegtäuschen, daß in diesem Paradigma nicht die Möglichkeit von Fehlalarmen ins Kalkül gezogen wird. Darauf wird in 5.3.8 noch näher eingegangen.

5.3.5 Alternative Codierungsmethoden

Die Abhängigkeit der Erkennungsleistung in HMAX vom Stimuluskontrast sowie die Notwendigkeit, Aufmerksamkeitseffekte genau an den Kontrast des jeweils bewußt beachteten Stimulus anzupassen, um dessen Erkennung zu erleichtern, beruht darauf, daß die VTUs auf bestimmte Feuerraten ihrer Afferenzen trainiert sind. Jede Abweichung der Ak-

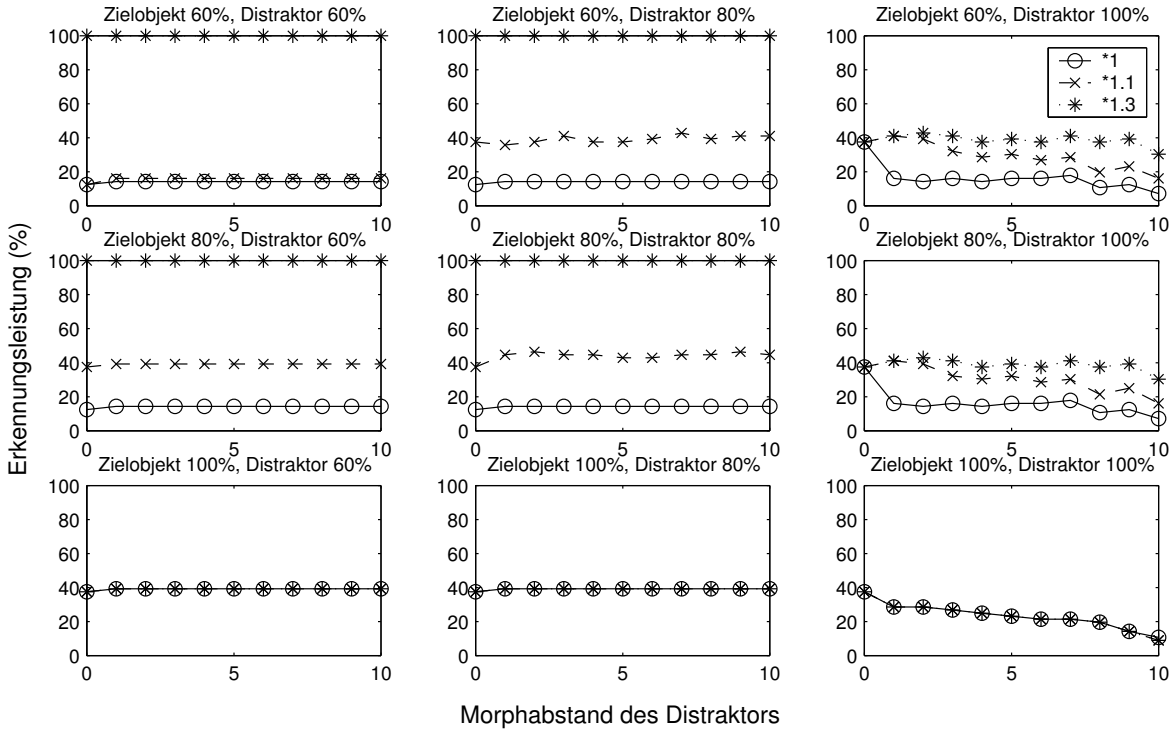

Abbildung 5-20: Gauß-Maximum-Tuning. Erkennungsleistung im Aktivste VTU-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die Faktoren an, um die jeweils die Aktivitäten der C2-Afferenzen der auf das Zielobjekt trainierten VTU erhöht wurden. Jede VTU hat 40 C2-Afferenzen.

tivität der C2-Zellen, die auf eine VTU projizieren, von diesem optimalen Aktivitätsmuster verursacht eine geringere Feuerrate der VTU und damit möglicherweise auch ein Absinken der Erkennungsleistung, unabhängig davon, ob die C2-Zellen schwächer oder stärker als während des VTU-Trainings feuern. Wie in den Methoden erläutert, wurde als eine mögliche Alternative zu dieser Codierungsmethode das „Gauß-Maximum-Tuning“ untersucht, bei dem vermieden wurde, daß hohe Aktivität einer afferenten Zelle wieder zum Absinken der Feuerrate einer VTU führt. Dieses Prinzip ist dabei durchaus auch biologisch motiviert: Merkmale, die den bevorzugten Stimulus einer VTU ausmachen, sollten nicht zu einem Absinken ihrer Aktivität führen, wenn sie „zu deutlich“ dargestellt sind. Außerdem modelliert das Gauß-Maximum-Tuning näherungsweise ein Sättigungsverhalten, je nach Betrachtung entweder der VTUs (weil sie auf Input jenseits einer bestimmten Stärke nicht mehr stärker oder schwächer antworten) oder der C2-Zellen (da es für eine

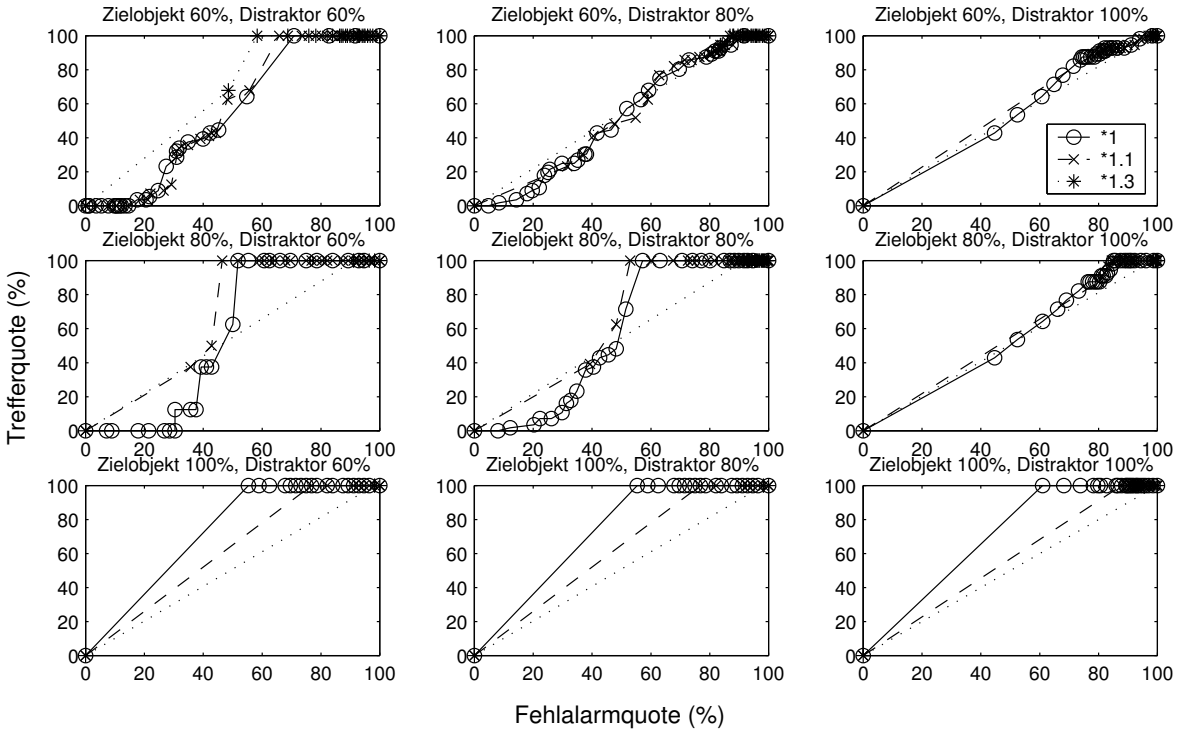

Abbildung 5-21: Gauß-Maximum-Tuning. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-20, aber nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

VTU keinen Unterschied macht, ob die Aktivitäten ihrer afferenten C2-Zellen bestimmte Werte erreichen oder überschreiten, ist der Effekt derselbe, als würden die C2-Zellen bei diesen Werten saturieren). Damit sind auch verschiedene Sättigungspunkte bei verschiedenen Zellen möglich. Das Gauß-Maximum-Tuning zeigt also, was die Verwendung von Zellen, die für ihre bevorzugten Stimuli im Sättigungsbereich feuern (statt wie bisher stets im linearen Bereich), für die Objekterkennung bringen kann. Es gilt aber zu bedenken, daß die ursprüngliche Abstimmung der VTUs in HMAX gewählt wurde, um möglichst hohe Spezifität der Antworten von VTUs in einem abstrakten „Merkmalsraum“ zu erreichen. In dieser Betrachtungsweise bewirkt der Wechsel zum Gauß-Maximum-Tuning einen Verlust an Spezifität, dessen Auswirkungen ebenfalls zu untersuchen sind.

Abbildungen 5-20 bis 5-23 zeigen die Resultate für Gauß-Maximum-Tuning für Autos und Büroklammern, mit und ohne multiplikative Aufmerksamkeitseffekte. Bei niedrigen Kontrasten und ohne Aufmerksamkeit sind keine Änderungen gegenüber der in HMAX übli-

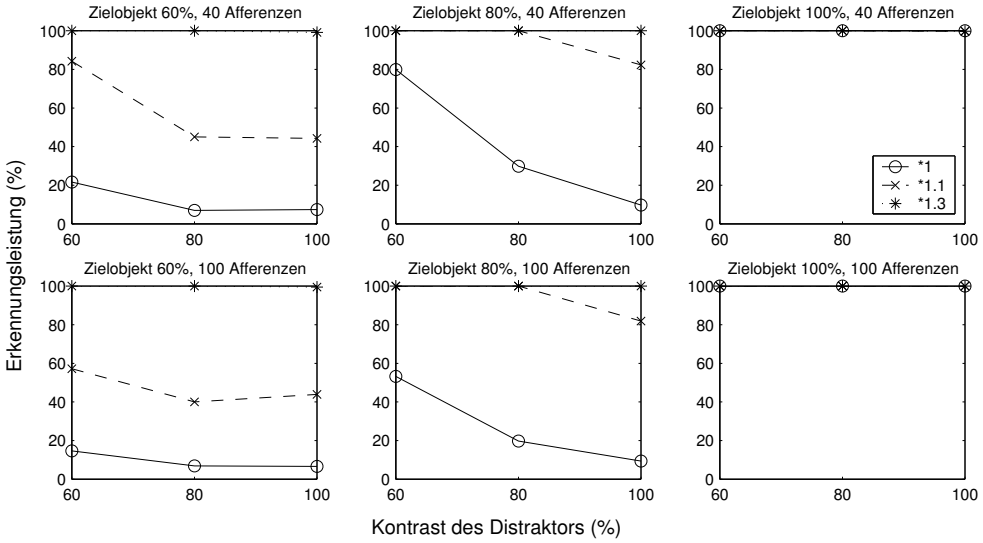

Abbildung 5-22: Gauß-Maximum-Tuning. Erkennungsleistung im Aktivste VTU-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer, gemittelt über alle Zielobjekte bei festem Kontrast des Distraktors (Abszissenwert). Kontrast des Zielobjekts und Anzahl der C2-Afferenzen pro VTU über den Diagrammen. Die Legende oben rechts gibt die Faktoren an, um die die Feuerraten der C2-Zellen erhöht wurden, die auf die VTU des Zielobjekts projizierten.

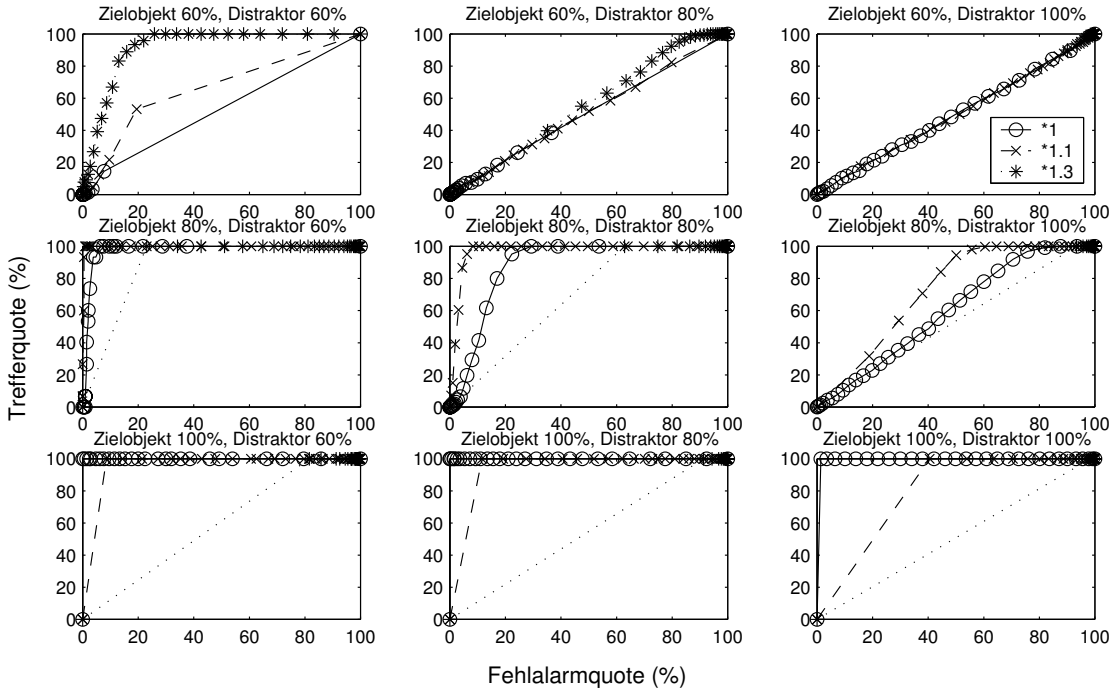

Abbildung 5-23: Gauß-Maximum-Tuning. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Büroklammer-Zielobjekte in Gegenwart einer Distraktor-Büroklammer (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen). Gleiche Daten wie in Abb. 5-22, aber nur für 40 Afferenzen pro VTU.

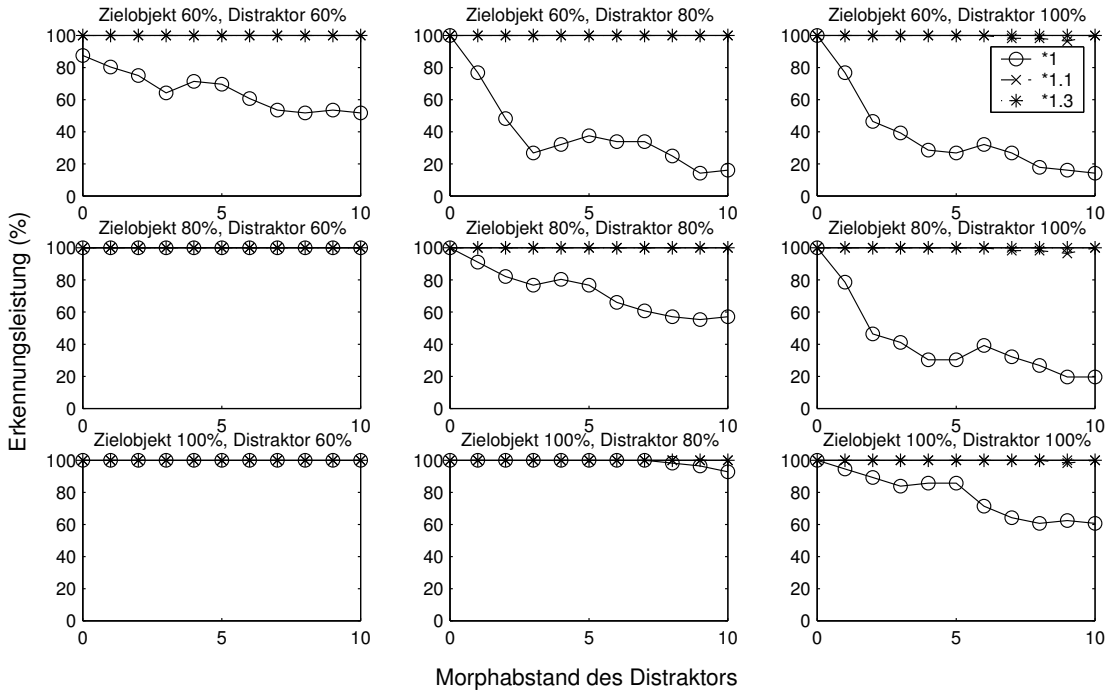

Abbildung 5-24: Aktivste Afferenzen-Tuning. Erkennungsleistung im Aktivste VTU-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die Faktoren an, um die jeweils die Aktivitäten der C2-Afferenzen der auf das Zielobjekt trainierten VTU erhöht wurden. Jede VTU hat 40 C2-Afferenzen.

chen Codierung von Stimuli feststellbar, da der Effekt des Gauß-Maximum-Tuning erst bei höheren C2-Feuerraten zum Tragen kommt. Für Büroklammer-Stimuli erweist sich diese Codierungsmethode als sehr effektiv: mit Hilfe von Aufmerksamkeitseffekten können perfekte Leistungen im Aktivste VTU-Paradigma und Verbesserungen im Stimulusvergleich-Paradigma erzielt werden, während die Spezifität der neuronalen Antworten erhalten bleibt und die Erkennung eines Zielobjekts mit 100% Kontrast, im Gegensatz zu Standard-HMAX, aufgrund der Saturierung der Zellen stets ungestört bleibt. Bei den Autos zeigt sich jedoch, daß das Gauß-Maximum-Tuning in der Tat zu einem Verlust an Spezifität der neuronalen Antworten führt, der für Stimulusklassen, die in sich sehr ähnlich sind, die Erkennungsleistung drastisch reduziert. Da verschiedene auf Autos trainierte VTUs viele C2-Afferenzen gemeinsam haben, feuern bei hohem Kontrast des Zielobjekts oder aufgrund von Aufmerksamkeitseffekten hier gleich mehrere VTUs im Sättigungsbereich, so daß die Unterscheidbarkeit von Stimuli in beiden Paradigmen der Erkennungsleistung

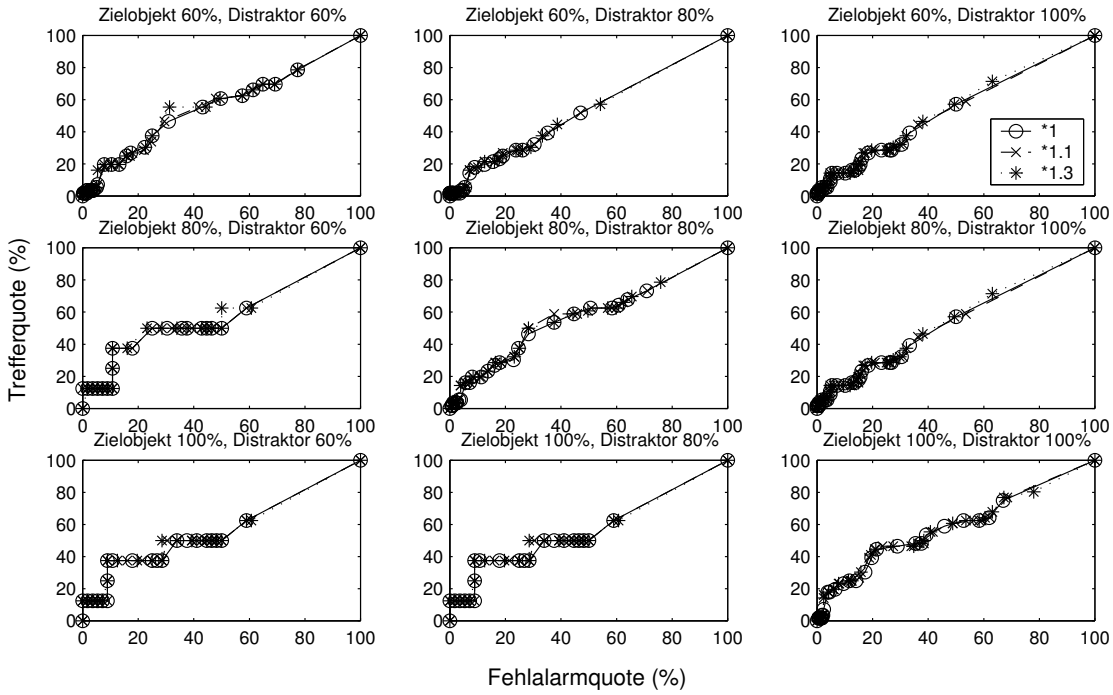

Abbildung 5-25: Aktivste Afferenzen-Tuning. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Auto-Zielobjekte in Gegenwart eines Distraktor-Autos (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-24, aber nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

nicht mehr gegeben ist (siehe Abb. 5-20 und 5-21). Da also das Gauß-Maximum-Tuning sogar die Erkennung von Stimuli bei 100% Kontrast verschlechtern kann, ist es nicht geeignet zur Erkennung untereinander relativ ähnlicher Stimuli, und für andere, stärker unterschiedliche Stimuli wie die Büroklammern ist es nicht unbedingt notwendig, da für solche Objekte auch in Standard-HMAX die Erkennungsleistung hinreichend gut ist (vgl. Abb. 5-9 und 5-10).

Die zweite untersuchte alternative Codierungsmethode, das „Aktivste Afferenzen-Tuning“, hatte noch mehr als das Gauß-Maximum-Tuning das Ziel, die Erkennungsleistung von HMAX stärker vom Kontrast unabhängig zu machen. Bei dieser Art der Stimuluscodierung werden die relativen Aktivitäten verschiedener VTUs nurmehr dadurch bestimmt, wie stark ihre jeweiligen C2-Afferenzen feuern, und nicht dadurch, wie genau sie das Aktivitätsmuster während des Trainings der VTU mit ihrem bevorzugten Stimulus reproduzieren. Dadurch können im Prinzip auch niedrige absolute Aktivitäten von C2-Zellen zur selben relativen Aktivität verschiedener VTUs führen. Abbildung 5-24 zeigt dementspre-

chend für Autos im Aktivste VTU-Paradigma eine deutlich höhere Erkennungsleistung bei niedrigen Kontrasten bereits ohne Aufmerksamkeit sowie eine Erkennungsleistung von 100% bei allen Stimuluskontrasten für Aufmerksamkeitseffekte verschiedener Stärke. Abbildung 5-25 macht jedoch deutlich, daß das Aktivste Afferenzen-Tuning keine realistische Alternative für HMAX sein kann, da die Fähigkeit des Modells, zwischen verschiedenen Stimuli zu diskriminieren, praktisch völlig verlorengeht. Auch mit Aufmerksamkeit auf Zielobjekte wird bei der Erkennung nur das Zufallsniveau erreicht, da ganz offensichtlich durch die Erhöhung der Aktivität der C2-Afferenzen einer VTU die Feuerrate dieser VTU unspezifisch zunimmt, unabhängig davon, ob das Zielobjekt tatsächlich gezeigt wird oder nicht, und daher die Fehlalarmquote ebenso steigt wie die Trefferquote. Die gleichen Resultate ergeben sich für Büroklammern (nicht abgebildet). Der Verzicht auf die in den genauen Feuerraten der Neuronen enthaltene Information führt also zu einem nicht akzeptablen Verlust an Spezifität.

Die untersuchten alternativen Codierungsmethoden scheinen also keine sinnvollen Lösungen für das Problem zu sein, die Objekterkennung bei Hintergrundstimuli und niedrigerem Kontrast zu verbessern sowie die effektive Anwendung von Aufmerksamkeitseffekten auf die Objekterkennung zu ermöglichen. Das Aktivste Afferenzen-Tuning ist insgesamt zu unspezifisch, und während das Gauß-Maximum-Tuning geeignet ist, um die störenden Effekte von Distraktoren mit hohem Kontrast zu kompensieren, ist es für realistische Stimuli, die zahlreiche Merkmale gemeinsam haben, ebenfalls nicht spezifisch genug.

5.3.6 Populationscode

Daß Stimuli im Gehirn durch Populationen von Neuronen codiert werden, ist angesichts der Tatsache, daß Neuronen in aller Regel auf mehrere Stimuli antworten [19, 60], eine realistische Annahme. Dabei existieren aber dennoch auf bestimmte Stimuli hochspezialisierte Populationen wie etwa die bevorzugt auf Gesichter antwortenden Neuronen, die im inferotemporalen Cortex von Makaken gefunden werden können [11]. Es ist anzunehmen, daß solche Populationen auch für die Repräsentation neu gelernter Exemplare der Objektklasse, auf die sie spezialisiert sind, zuständig sind. Eine solche Situation wurde, wie in den Methoden beschrieben, in HMAX mit einer Population von 190 auf Gesichter

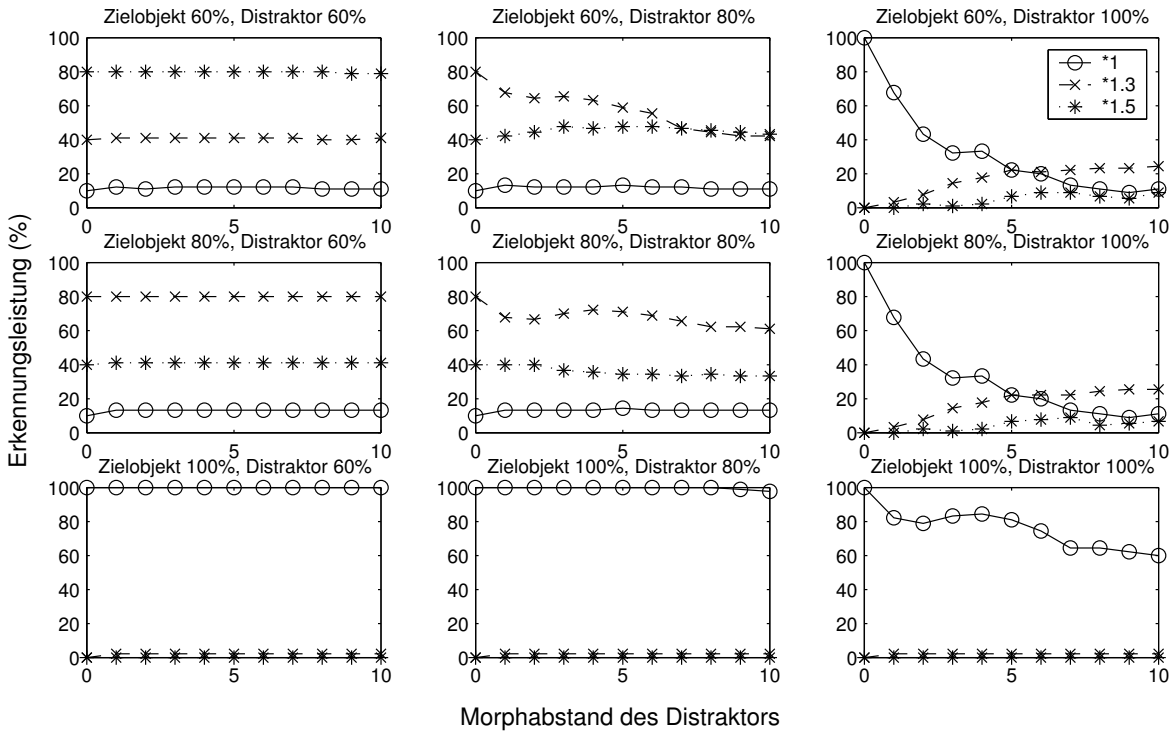

Abbildung 5-26: Erkennung von Gesichtern mit *einer* VTU pro Zielobjekt. Erkennungsleistung im Aktivste VTU-Paradigma in Gegenwart eines Distraktor-Gesichts (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die Faktoren an, um die jeweils die Aktivitäten der C2-Afferenzen der auf das Zielobjekt trainierten VTU erhöht wurden. Jede VTU hat 40 C2-Afferenzen.

trainierten VTUs realisiert, um zu untersuchen, welche Effekte die hier verwendeten Modelle von Aufmerksamkeit im Rahmen eines Populationscodes auf die Objekterkennung haben können. Dabei wurde jeweils nur eine bestimmte Anzahl (10, 40 oder 100) der 190 VTUs der Population betrachtet, nämlich diejenigen, die auf das jeweilige Zielobjekt am stärksten antworteten – analog der Auswahl der bei Präsentation eines Stimulus am stärksten feuernden C2-Zellen als Afferenzen für die zugehörige VTU in den bisherigen Versuchen. Dies entspricht der Annahme des HMAX-Modells, daß zwar aus der Aktivität einer großen Zahl von Neuronen Informationen über einen Stimulus gewonnen werden können, aber für die tatsächliche neuronale Repräsentation dieses Stimulus wohl hauptsächlich die stark aktiven genutzt werden bzw. notwendig sind.

Um die Leistungsfähigkeit eines Populationscodes besser mit der einer Codierung von Stimuli durch einzelne VTUs vergleichen zu können, wurden zunächst wie in den vorher-

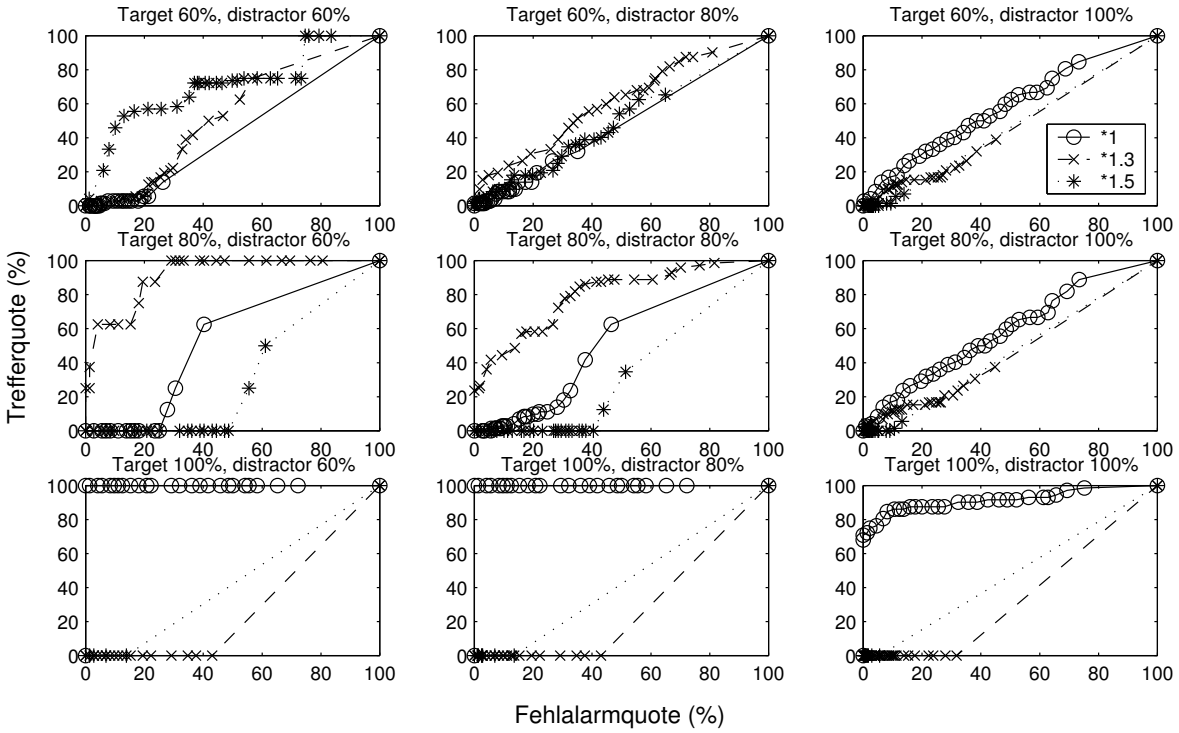

Abbildung 5-27: Erkennung von Gesichtern mit *einer* VTU pro Zielobjekt. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma in Gegenwart eines Distraktor-Gesichts (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-26, aber nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

gehenden Kapiteln einzelne VTUs auf jedes der 10 „prototypischen“ Gesichter (bzw. das von ihnen ausgelöste C2-Aktivitätsmuster) trainiert und die Erkennungsleistung für diese Gesichter mit und ohne multiplikative Aufmerksamkeitseffekte untersucht. Die Ergebnisse in Abb. 5-26 und 5-27 zeigen, daß die Leistung von HMAX für die Gesichter derjenigen für Autos sehr ähnlich ist. Die Erkennung individueller Gesichter ist stark vom Kontrast abhängig und läßt sich durch multiplikative Aufmerksamkeitseffekte auf C2-Ebene verbessern, vorausgesetzt, der Kontrast des Distraktorstimulus ist nicht höher als der des Zielobjekts und die Stärke der Modulation ist an den Kontrast des Zielobjekts angepaßt. Distraktorstimuli, die weniger Ähnlichkeit mit dem zu erkennenden Zielobjekt aufweisen, interferieren stärker mit dessen Erkennung als ähnlichere Stimuli (siehe 5-26 und vgl. Kapitel 3.2.4). Ferner ist, ebenso wie für Autos, die Erkennungsleistung nahezu unabhängig von der Anzahl der für die VTUs verwendeten Afferenzen, und eine unspezifische allgemeine Erhöhung der Aktivität aller C2-Zellen kann die Erkennungsleistung nahezu eben-

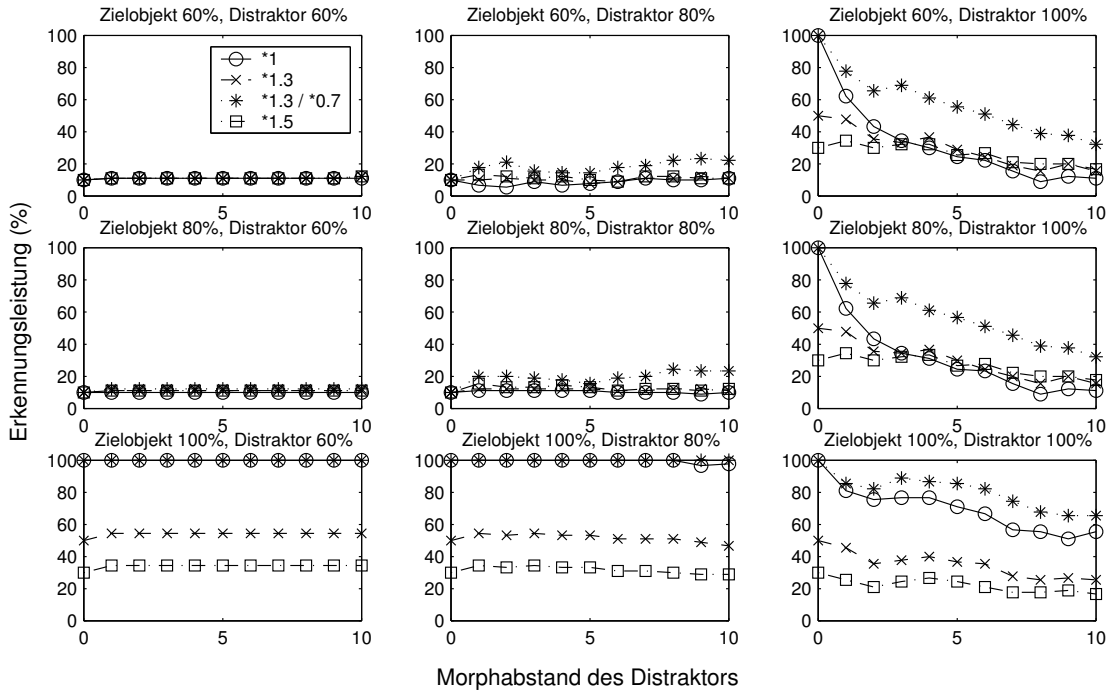

Abbildung 5-28: Erkennung von Gesichtern mit VTU-Populationscode und Aufmerksamkeitseffekten auf VTU-Ebene. Erkennungsleistung im Aktivste VTU-Paradigma in Gegenwart eines Distraktor-Gesichts (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die multiplikative Verstärkung der Aktivität der VTUs an, die am stärksten auf das Zielobjekt ansprechen (erster Wert), sowie den Wert der Suppression aller übrigen VTUs der Population (zweiter Wert, wenn vorhanden). Alle Daten für 10 Populations-VTUs und 40 C2-Afferenzen pro VTU.

so gut verbessern wie gezielte Aufmerksamkeitseffekte (nicht abgebildet).

Abbildungen 5-28 und 5-29 zeigen dann die Leistung der VTU-Population bei der Erkennung individueller neuer Gesichter, mit und ohne Aufmerksamkeitseffekte, die hier direkt auf der Ebene der VTU-Population angewendet wurden, für 10 der 190 VTUs aus der Population, die jeweils 40 C2-Afferenzen hatten. Größere betrachtete Subpopulationen oder mehr C2-Afferenzen pro VTU erbrachten keine bessere Leistung (nicht abgebildet). Wie in den Methoden erläutert, erfolgte die Auswertung der Experimente mit Hilfe einer zweiten Schicht von VTUs und konnte daher ebenso durchgeführt werden wie in den Experimenten, in denen einzelne VTUs individuelle Stimuli repräsentierten. Es wird zunächst deutlich, daß die Erkennungsleistung der VTU-Population ohne Aufmerksamkeitseffekte nicht besser, im Stimulusvergleich-Paradigma sogar etwas niedriger ist als die der ein-

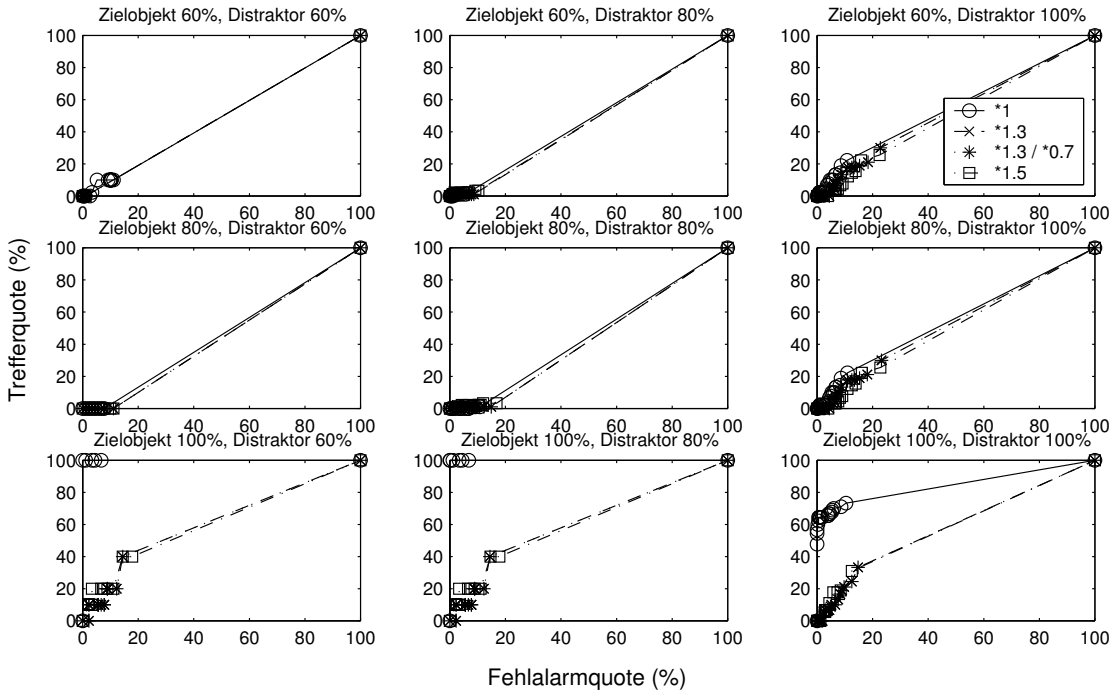

Abbildung 5-29: Erkennung von Gesichtern mit VTU-Populationscode und Aufmerksamkeitseffekten auf VTU-Ebene. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma in Gegenwart eines Distraktor-Gesichts (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen). 10 Populations-VTUs, 40 C2-Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-28, aber nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

zelnen VTUs in Abb. 5-26 und 5-27. Der Grund dafür ist, daß bei der Betrachtung einer ganzen Population von VTUs, im Gegensatz zu einzelnen VTUs, eben auch mehrere VTUs zugleich jeweils eine bestimmte Aktivität haben müssen, um noch hinreichend genau ein für ein bestimmtes Objekt codierendes Aktivitätsmuster zu reproduzieren. Die Spezifität einer VTU-Population für einen bestimmten Stimulus ist also deutlich höher als die einzelner VTUs. Dies ist auch daran erkennbar, daß die VTUs in der zweiten Schicht stets sehr geringe Aktivitätswerte zeigen, sobald ihr jeweils bevorzugter Stimulus bei weniger als 100% Kontrast oder mit einem entsprechend kontrastreichen Distraktor gezeigt wird. Im Kontext des HMAX-Modells scheint ein Populationscode also eher empfindlicher auf störende Einflüsse von Kontrast und Hintergrund zu reagieren.

Aus Abb. 5-28 und 5-29 ist außerdem ersichtlich, daß multiplikative Aufmerksamkeitseffekte auf Ebene der VTUs im allgemeinen die Erkennungsleistung nicht verbessern können. Da VTUs nicht linear auf unterschiedliche Kontraste oder störende Hintergrundstimuli

reagieren, kann nicht erwartet werden, daß eine einfache Erhöhung ihrer Feuerraten das für einen bestimmten Stimulus codierende Aktivitätsmuster wiederherstellt. Nur für den Spezialfall eines sehr ähnlichen Distraktorstimulus mit 100% Kontrast (Abb. 5-28 rechts) ist dieses Aktivitätsmuster offensichtlich noch so wenig gestört, daß ein solcher Aufmerksamkeitseffekt die Erkennungsleistung im Aktivste VTU-Paradigma (das sich hier auf die VTUs in der zweiten Schicht bezieht) etwas verbessern kann. Ansonsten aber und insbesondere im Stimulusvergleich-Paradigma zeigen Modulationen der Aktivität der VTUs in der Population keinen Effekt.

Die Anwendung von Aufmerksamkeitseffekten auf Ebene der C2-Zellen verkompliziert sich im Rahmen eines VTU-Populationscodes dadurch, daß nicht mehr wie bisher eine eindeutig identifizierbare Menge an C2-Zellen die für einen Stimulus wichtigen Merkmale codieren, nämlich die Afferenzen einer einzelnen auf diesen Stimulus trainierten VTU. Eine mögliche Lösung dieses Problems wäre, alle diejenigen C2-Zellen durch Aufmerksamkeit in ihrer Aktivität zu verstärken, die auf die vom bewußt beachteten Zielobjekt am stärksten aktiven VTUs in der Population projizieren, die aber zugleich nicht auch noch mit anderen, für die Repräsentation dieses Stimulus nicht so bedeutenden VTUs verbunden sind. Dadurch könnte auch in einem Populationscode eine gewisse Spezifität des Aufmerksamkeitseffekts gewährleistet werden. Es stellt sich jedoch heraus, daß insbesondere bei Verwendung nur weniger VTUs (10) für die Repräsentation eines Stimulus in der Population nahezu keine C2-Zellen zu finden sind, die nicht auch noch auf andere der 190 VTUs in der Population projizieren. Eine Einschränkung der Aktivitätserhöhung auf diese C2-Zellen ergibt keine sichtbaren Unterschiede in der Erkennungsleistung (nicht abgebildet). Bei Verwendung eines Populationscodes ist es also noch schwieriger als für die Codierung durch einzelne VTUs, Neuronen zu finden, deren Aktivität sowohl charakteristisch für einen Stimulus als auch eindeutig diesem Stimulus zuzuschreiben ist.

Demzufolge wurden für Abb. 5-30 und 5-31 die Feuerraten *aller* C2-Zellen, die auf VTUs projizierten, die zur das Zielobjekt repräsentierenden Subpopulation gehörten, durch multiplikative Aufmerksamkeitseffekte verstärkt, unabhängig davon, ob dadurch etwa auch andere VTUs in der Population beeinflußt wurden. (Für Subpopulationen von 10 VTUs und 40 Afferenzen pro VTU ergaben sich dabei im Mittel etwa 80 von 256 C2-Zellen, deren Aktivität erhöht wurde.) Die Auswirkungen dieses Modells von Aufmerksamkeit stellen

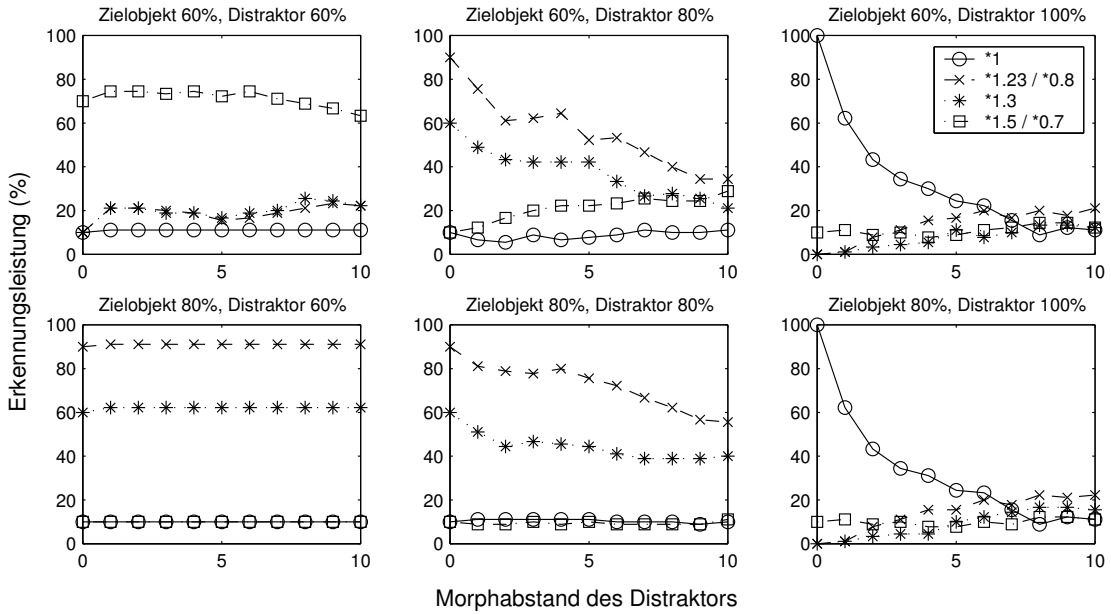

Abbildung 5-30: Multiplikative Aufmerksamkeitseffekte und Suppression auf C2-Ebene im VTU-Populationscode. Erkennungsleistung im Aktivste VTU-Paradigma für Gesichter in Gegenwart eines Distraktor-Gesichts (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen), gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. Die Legende im rechten oberen Diagramm gibt für alle Diagramme die multiplikative Verstärkung der Aktivität der C2-Afferenzen derjenigen VTUs an, die am stärksten auf das Zielobjekt ansprechen (erster Wert), sowie den Wert der Suppression aller übrigen C2-Zellen (zweiter Wert, wenn vorhanden). Alle Daten für 10 Populations-VTUs und 40 C2-Afferenzen pro VTU.

sich als qualitativ sehr ähnlich heraus wie diejenigen von vergleichbaren Modulationen, wenn einzelne VTUs für individuelle Stimuli codieren, sind aber – zumindest im Stimulusvergleich-Paradigma – quantitativ eher schwächer (vgl. 5-26 und 5-27). Wiederum muß die Stärke der Modulation zum Kontrast des Zielobjekts passen, wenn die Leistung sich verbessern soll, und auch dann können die Auswirkungen eines Distraktors mit höherem Kontrast nicht kompensiert werden. Eine Suppression der Aktivität der C2-Zellen, die auf die VTUs der am stärksten aktiven unter den nicht bevorzugt auf den Zielstimulus ansprechenden Subpopulationen projizieren, ist – analog zu den Resultaten in 5.3.4 für Autos – nicht sinnvoll, weil damit auch zahlreiche Afferenzen der für den Zielstimulus selektiven VTU-Subpopulation supprimiert werden und der Effekt von Aufmerksamkeit sich deutlich verringert (nicht abgebildet). Eine Verringerung der Aktivität aller C2-Zellen, die nicht auf VTUs der für das Zielobjekt codierenden Subpopulation projizieren, erhöht – wie auch in 5.3.4 diskutiert – die Erkennungsleistung im Aktivste VTU-Paradigma (wie-

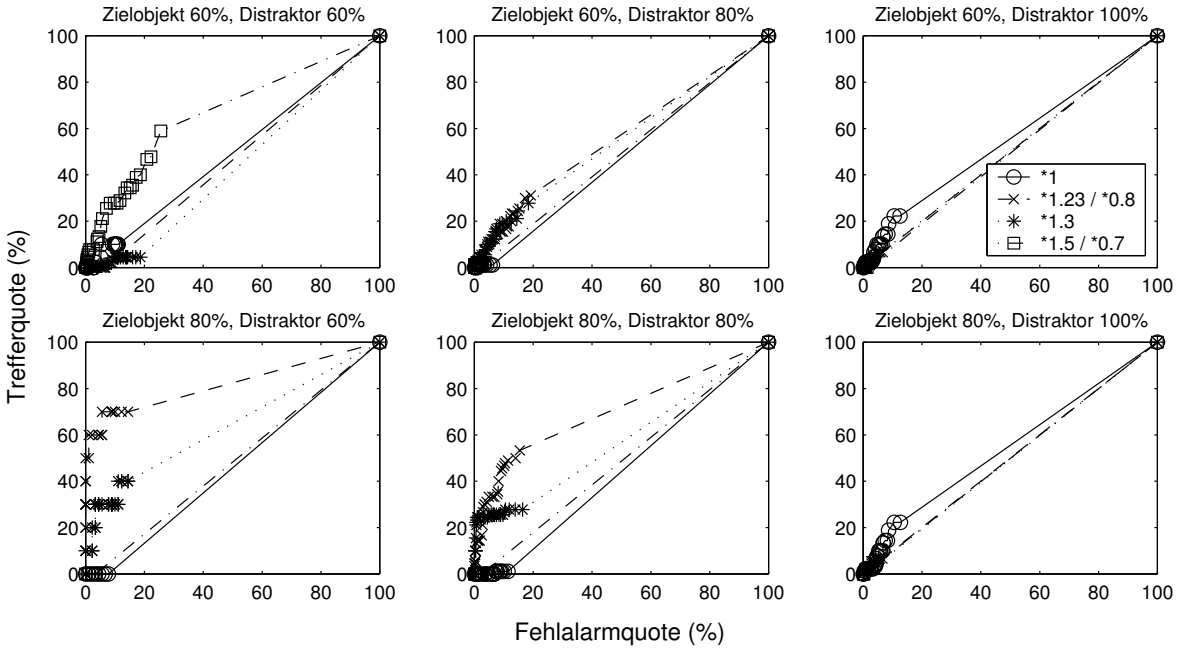

Abbildung 5-31: Multiplikative Aufmerksamkeitseffekte und Suppression auf C2-Ebene im VTU-Populationscode. Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Gesichter in Gegenwart eines Distraktor-Gesichts (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen). 10 Populations-VTUs, 40 C2-Afferenzen pro VTU. Gleiche Daten wie in Abb. 5-30, aber nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

derum bezogen auf VTUs in der zweiten Schicht), aber nicht im Stimulusvergleich-Paradigma. Im Unterschied zur Codierung durch einzelne VTUs verschwinden jedoch eventuelle Verbesserungen der Erkennungsleistung durch Aufmerksamkeit, wenn größere VTU-Subpopulationen und mehr C2-Afferenzen verwendet und alle dieser C2-Afferenzen in ihrer Aktivität verstärkt werden (Abb. 5-32). Auch hierin erweist sich die höhere Spezifität einer Codierung von Stimuli durch eine Population von VTUs; je mehr afferente Information dabei berücksichtigt wird, desto mehr kommt diese Eigenschaft zum Tragen, und desto empfindlicher reagiert die Population auf Störungen ihres bevorzugten afferenten Aktivitätsmusters bzw. desto weniger können Aufmerksamkeitseffekte noch etwas ausrichten.

Insgesamt ergibt die Verwendung eines Populationscodes in HMAX ähnliche Resultate für die Erkennung von Stimuli bei variierenden Kontrasten und in Gegenwart von Distraktorstimuli, wie es für die Codierung von Stimuli durch einzelne VTUs der Fall ist. Ein Populationscode ist offensichtlich effektiv darin, die Spezifität neuronaler Antworten auf

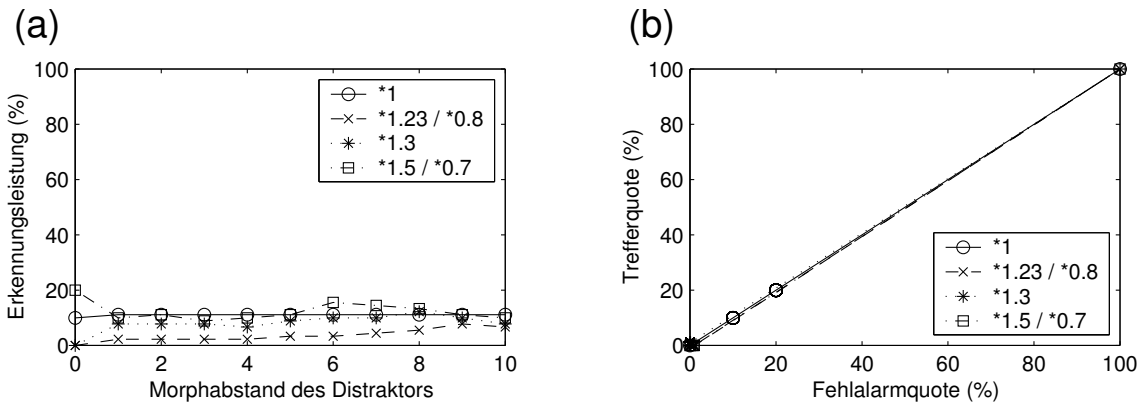

Abbildung 5-32: Auswirkungen von größeren VTU-Subpopulationen und mehr C2-Afferenzen im VTU-Populationscode. Daten für 100 VTUs und 100 C2-Afferenzen pro VTU. Kontrast von Zielobjekt und Distraktor 60%. Die Legende gibt die multiplikative Verstärkung der Aktivität der C2-Afferenzen derjenigen VTU-Subpopulation an, die am stärksten auf das Zielobjekt anspricht (erster Wert), sowie den Wert der Suppression aller übrigen C2-Zellen (zweiter Wert, wenn vorhanden). **(a)** Erkennungsleistung im Aktivste VTU-Paradigma für Gesichter in Gegenwart eines Distraktor-Gesichts, gemittelt über alle Stimuli mit demselben Morphabstand zwischen Zielobjekt und Distraktor. **(b)** Über alle Zielobjekte gemittelte ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für Gesichter in Gegenwart eines Distraktor-Gesichts. Gleiche Daten wie in (a), jedoch nur für Stimuli mit 5 Schritten Morphabstand zwischen Zielobjekt und Distraktor.

Stimuli zu erhöhen. Während das von Vorteil sein dürfte, wenn die einzelnen Neuronen etwa eher breit abgestimmt und spontan aktiv sind, bringt es keine Gewinne für die Unterscheidung von Stimuli, wenn wie in HMAX ideale Modellneuronen verwendet werden. Aufmerksamkeit kann in einem Populationscode in analoger Weise wie für die Codierung durch einzelne VTUs angewendet werden, unterliegt aber denselben Einschränkungen. Außerdem ist im Fall des Populationscodes die Frage, wie ein Mechanismus der frühen Selektion von Merkmalen durch Aufmerksamkeit die „richtigen“ Neuronen für eine Aktivitätserhöhung auswählen kann – also solche, die sowohl möglichst charakteristisch als auch möglichst eindeutig für ein Zielobjekt sind – , noch schwieriger zu beantworten als bei der Codierung durch einzelne VTUs. Aufmerksamkeitseffekte sind deshalb im Kontext eines Populationscodes noch weniger effizient oder andererseits noch weniger spezifisch für ein bestimmtes Objekt.

5.3.7 Kategorisierung

Wie bereits in 5.1 erwähnt, zeigen Experimente im RSVP-Paradigma, daß auch sehr allgemein gehaltene Hinweise auf eine Objektkategorie, der ein zu detektierender Stimulus angehört, dessen Erkennung deutlich verbessern können [23]. Ferner scheint die Beurteilung von nur kurze Zeit (20 ms) präsentierten Stimuli hinsichtlich einer Kategorie, der sie angehören (etwa „Transportmittel“ oder „Tier“), innerhalb von nur 150 ms abgeschlossen werden zu können [14, 74]. Es wurde argumentiert, daß sich vom Standpunkt der Informationsverarbeitung aus betrachtet die Identifikation und Kategorisierung von Objekten im Grunde nicht unterscheidet, da in beiden Fällen gelernt werden muß, einen bestimmten Ausgabewert (also die Identität oder Kategorie eines Objekts) mit einer Reihe verschiedener Eingabedaten (d.h. dasselbe Objekt unter verschiedenen visuellen Bedingungen oder verschiedene Objekte derselben Kategorie) zu assoziieren [57]. Beide Aufgaben könnten also im Prinzip von ähnlichen oder denselben neuronalen Mechanismen gelöst werden, und angesichts der hohen Geschwindigkeit, in der die entsprechenden Prozesse abgeschlossen werden können, laufen sie vermutlich eher parallel als hierarchisch ab. Da HMAX erfolgreich Kategorisierungseigenschaften von Neuronen modellieren konnte, die im präfrontalen Cortex gefunden wurden [15, 29], wurde im Rahmen dieser Arbeit ein Modell für den Einfluß von Aufmerksamkeit auf die Kategorisierung von Stimuli, die bei verschiedenen Kontrasten und in Gegenwart von Distraktoren gezeigt werden, untersucht. Wie in den vorherigen Versuchen handelte es sich damit auch hier um einen Mechanismus der frühen Selektion durch Aufmerksamkeit.

Wie in den Methoden erläutert, wurde ein „Kategorisierungsneuron“ darauf trainiert, Darstellungen einzelner Autos und Gesichter zu unterscheiden. Abbildung 5-33 zeigt, daß dieses Training in der Tat erfolgreich war und sowohl während des Trainings gezeigte als auch neue Stimuli aus der Grundgesamtheit der auch in den anderen Versuchen verwendeten Autos und Gesichter stets korrekt kategorisiert werden konnten. Eine solche perfekte Leistung bzw. eine Fläche unter der ROC-Kurve von 1 ergibt sich, wenn das Kategorisierungsneuron auf Autos stets stärker antwortet als auf Gesichter, da im Training Ausgabewerte von 1 für Autos und -1 für Gesichter „gelernt“ worden waren.

Abbildung 5-34 verdeutlicht zunächst, daß niedrigere Kontraste und Hintergrundstimu-

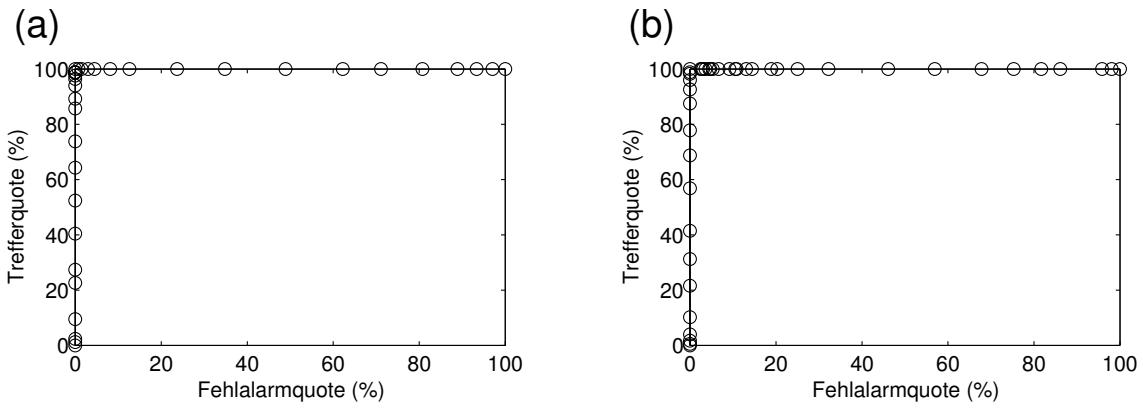

Abbildung 5-33: ROC-Kurven für die Leistung des Kategorisierungsneurons bei der Unterscheidung isolierter Autos und Gesichter bei maximalem Kontrast (Autos 100%, Gesichter 60%). Die VTUs, die auf das Kategorisierungsneuron projizieren, haben je 40 C2-Afferenzen. **(a)** Kategorisierungsleistung für die Trainingsstimuli (alle 8 Auto- und 10 Gesichter-Prototypen sowie zusätzlich zwei gemorphte Stimuli von jeder Morphlinie der beiden Kategorien). **(b)** Leistung für Teststimuli (alle nicht im Training verwendeten einzelnen Autos und Gesichter bei jeweils maximalem Kontrast).

li (verwendet wurden Büroklammern) wie erwartet die Kategorisierungsleistung senken. Aus den Kurven ohne Aufmerksamkeitseffekte (Kreise) ist auch ersichtlich, daß die nominalen maximalen Kontrastwerte, die für die verschiedenen Stimulusklassen unterschiedlich gewählt wurden, um die effektiven Kontraste anzugleichen (siehe Methoden), angemessen waren: Während bei den niedrigsten Kontrasten der Zielobjekte und dem höchsten Distraktorkontrast die Darstellungen der Autos und der Gesichter für das Modell nicht unterscheidbar sind (Abb. 5-34 rechts oben), wird eine mittlere Kategorisierungsleistung erreicht, wenn sowohl Zielobjekte als auch Distraktoren mit ihren jeweiligen maximalen Kontrasten gezeigt werden (Abb. 5-34 rechts unten), und die beiden Klassen sind wiederum perfekt unterscheidbar, wenn die Zielobjekte bei maximalem Kontrast und die Distraktoren bei minimalem Kontrast präsentiert werden (Abb. 5-34 links oben). Die Resultate für die Auswirkungen von Aufmerksamkeitseffekten auf die Kategorisierung der Stimuli sind jedoch bestenfalls mehrdeutig. Die Erwartung wäre gewesen, daß simulierte Aufmerksamkeit auf eine der beiden Kategorien, also Erhöhung der Aktivität der für diese Kategorie relevanten VTUs oder C2-Zellen, die Unterscheidbarkeit der beiden Stimulusklassen erhöhen kann. Während sowohl Modulationen der Feuerraten von C2-Zellen als auch von VTUs teilweise effektiv waren, war ihr Erfolg stark davon abhängig, auf welche Kategorie sich die Aufmerksamkeit richtete, wie hoch die Kontraste von Zielobjekt und

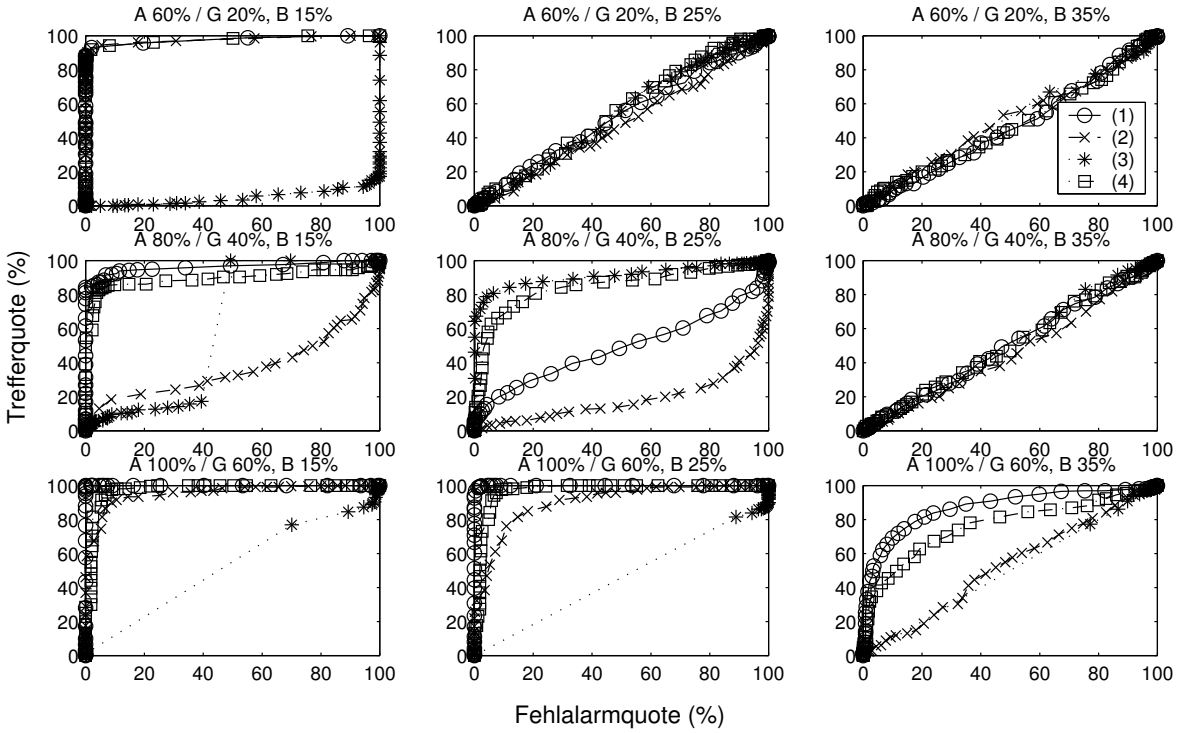

Abbildung 5-34: ROC-Kurven für die Leistung eines Kategorisierungsneurons bei der Unterscheidung von Autos und Gesichtern, die bei verschiedenen Kontrasten und mit einem Distraktorstimulus (Büroklammer) präsentiert wurden (A Auto, G Gesicht, B Büroklammer). Die VTUs, die auf das Kategorisierungsneuron projizieren, haben je 100 C2-Afferenzen. Legende: (1) keine Aufmerksamkeitseffekte, (2) multiplikative Verstärkung (Faktor 1.5) der Aktivität aller C2-Zellen, die *nur* auf Auto-VTUs projizieren, (3) wie (2), aber für *alle* C2-Zellen, die mit Auto-VTUs (und eventuell auch anderen VTUs) verbunden sind, (4) Verstärkung aller C2-Afferenzen der Auto-VTUs um den Faktor 1.3 und Suppression aller C2-Afferenzen der Gesichter-VTUs um den Faktor 0.8. Ähnliche Resultate ergaben sich für Verstärkung der Afferenzen der auf Gesichter trainierten VTUs (nicht abgebildet).

Distraktor waren, wieviele C2-Afferenzen die VTUs hatten, oder auf welche C2-Zellen sich ein Aufmerksamkeitseffekt auswirkte. Bei allen getesteten Aufmerksamkeitseffekten wurden Veränderungen der Kategorisierungsleistung nur für mittlere Kontraste von Zielobjekt und Distraktor gefunden (Auto 80%, Gesicht 40%, Büroklammer 25%). Dieselbe Modulation konnte zum Teil die Leistung verbessern oder verringern, je nachdem, ob sie auf *alle* C2-Afferenzen der VTUs einer Stimuluskategorie angewendet wurde oder nur auf die *eindeutigen* C2-Afferenzen dieser VTUs, also diejenigen C2-Zellen, die nur auf VTUs dieser Kategorie, nicht aber der anderen projizierten (Abb. 5-34 Mitte). Eine ähnliche Veränderung des Effekts einer Modulation, von Verbesserung der Leistung zu einer Verschlechte-

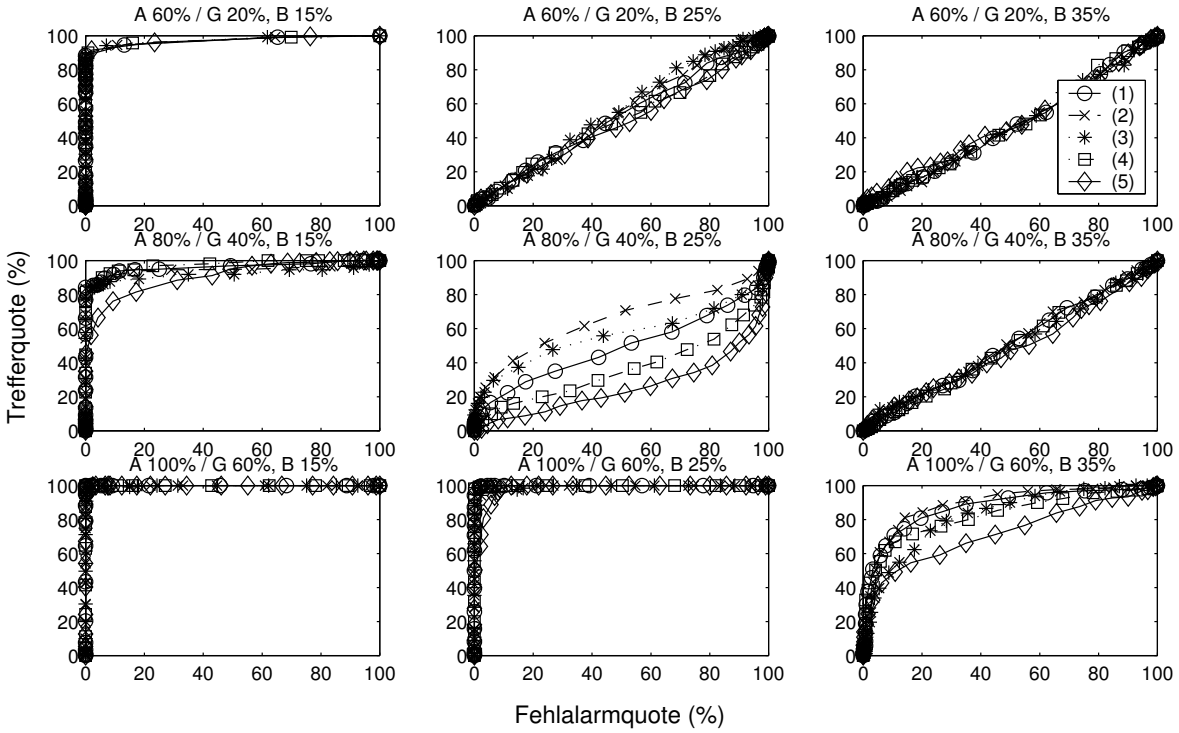

Abbildung 5-35: ROC-Kurven für die Leistung eines Kategorisierungsneurons bei der Unterscheidung von Autos und Gesichtern mit Büroklammern als Hintergrund und bei verschiedenen Kontrasten. Alle VTUs haben 100 C2-Afferenzen. Legende: (1) keine Aufmerksamkeitseffekte, (2) Verstärkung der Aktivität von auf Autos trainierten VTUs um den Faktor 1.1, (3) wie (2), aber für einen Faktor von 1.3, (4) Verstärkung der Aktivität von auf Gesichtern trainierten VTUs um den Faktor 1.1 und (5) wie (4), aber für den Faktor 1.3.

rung, wurde auch beobachtet, wenn für denselben Aufmerksamkeitseffekt die Anzahl der C2-Afferenzen pro VTU von 40 auf 100 erhöht wurde (nicht abgebildet). Verbesserungen der Leistung wurden beobachtet, wenn die Aktivitäten der auf Autos trainierten VTUs verstärkt wurden, dasselbe galt allerdings nicht für die auf Gesichter trainierten VTUs (Abb. 5-35).

Es können also keine konsistenten Effekte von Aufmerksamkeit auf die Kategorisierung von Stimuli durch HMAX gefunden werden. Während Verbesserungen der Leistung im Prinzip möglich scheinen, sind sie zumindest sehr stark von zahlreichen Parametern abhängig und daher auf keinen Fall robust. Es scheint nur schwer oder kaum möglich, zumindest in HMAX, die korrekte Kategorisierung von Stimuli unter schwierigen visuellen Bedingungen durch im voraus operierende Aufmerksamkeitsmechanismen zu erleichtern, selbst für zwei in sich relativ ähnliche und voneinander gut unterscheidbare Stimuluska-

tegorien, wie sie hier verwendet wurden. Die Merkmale einer Stimulusklasse sind nicht notwendigerweise eindeutig für diese Klasse, und die Prozesse der Kategorisierung sind in hohem Maße nichtlinear, so daß Aufmerksamkeitseffekte, wie sie hier modelliert wurden, in ihren Auswirkungen nicht vorhersehbar sind.

5.3.8 Probleme von Aufmerksamkeitsmechanismen

In den bisherigen Abschnitten wurden eine Reihe von Problemen dargestellt, die sich in der Modellierung eines Mechanismus der frühen Selektion von Merkmalen durch Aufmerksamkeit ergaben, wenn also verstärkte Aktivität von Neuronen in HMAX als Modell für einen eventuellen Einfluß von Aufmerksamkeit auf die Objekterkennung verwendet wurde. Neben der Notwendigkeit, die Stärke einer Modulation der Feuerraten von Neuronen an den im voraus unbekannten Kontrast des zu erkennenden Stimulus anzupassen, erwies sich als das bedeutendste Problem, daß die untersuchten Mechanismen von Aufmerksamkeit nicht zugleich effektiv und spezifisch zu sein schienen, daß sie also, wenn sie die Erkennung eines Objekts verbessern konnten, sich möglicherweise auch auf die Erkennung anderer Objekte auswirkten bzw. die Wahrscheinlichkeit von Fehlalarmen erhöhten. Diese Behauptung soll in diesem Abschnitt nochmals gezielt untersucht werden, hier wiederum unter Verwendung des regulären HMAX-Modells mit einzelnen VTUs zur Codierung unterschiedlicher Stimuli und der normalen Gaußschen Abstimmung der VTUs auf ein bestimmtes Aktivitätsmuster ihrer C2-Afferenzen.

Abbildung 5-36 behandelt das Problem der Spezifität. Sie zeigt die Verbesserung der Erkennungsleistung für eine Büroklammer durch einen multiplikativen Aufmerksamkeitseffekt, der eigentlich auf eine andere Büroklammer gerichtet war, d.h. es wurde die Aktivität der C2-Afferenzen einer anderen VTU erhöht als der, deren Objekterkennungsleistung im Stimulusvergleich-Paradigma hier ausgewertet wurde. Das bedeutet, daß selbst für VTUs, die jeweils nur 40 C2-Afferenzen haben und auf Büroklammern trainiert sind, also im Vergleich zu den anderen verwendeten Objektklassen am wenigsten Afferenzen gemeinsam haben (siehe Kapitel 3.2.4), die Auswirkungen von Aufmerksamkeit auf einen Stimulus sich auch auf andere Stimuli übertragen können, also nicht spezifisch sind. Der Effekt auf die Objekterkennung ist sogar ähnlich groß wie für direkt wirkende Aufmerksamkeit (vgl. etwa Abb. 5-10). Entsprechende Resultate wurden auch für die meisten anderen

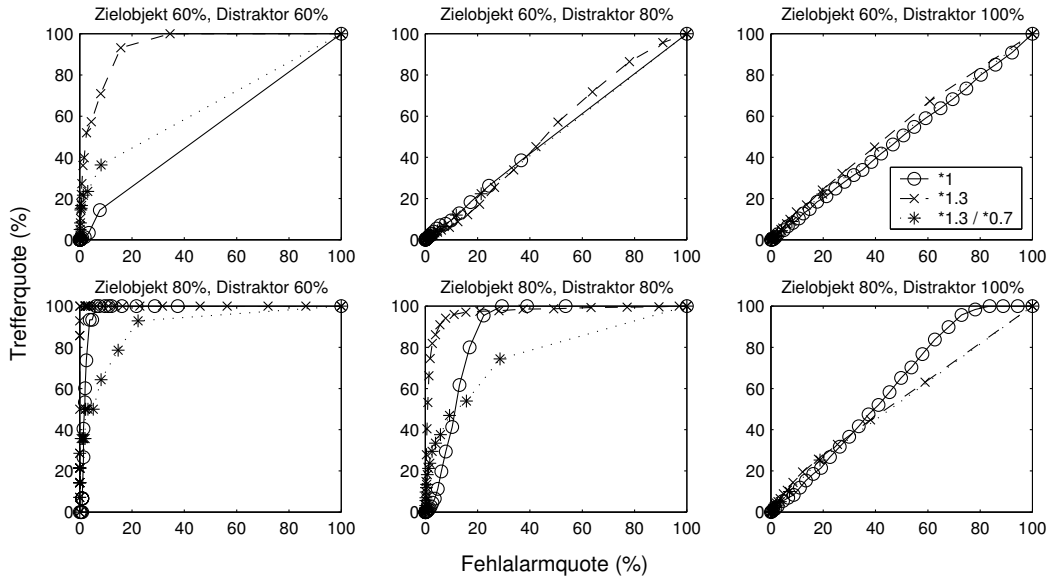

Abbildung 5-36: Verbesserung der Erkennungsleistung durch Aufmerksamkeit für Objekte, auf die die Aufmerksamkeit nicht gerichtet ist. ROCs der Erkennungsleistung im Stimulusvergleich-Paradigma für eine Büroklammer in Gegenwart einer Distraktor-Büroklammer (Kontraste der beiden Stimuli jeweils über den einzelnen Diagrammen) und 40 Afferenzen pro VTU. Die Legende im oberen rechten Diagramm gibt die Verstärkung der Aktivität der C2-Afferenzen einer *anderen* VTU als der, deren Leistung hier ausgewertet wird, an (erster Wert) und die Suppression aller übrigen C2-Zellen (zweiter Wert, wo vorhanden).

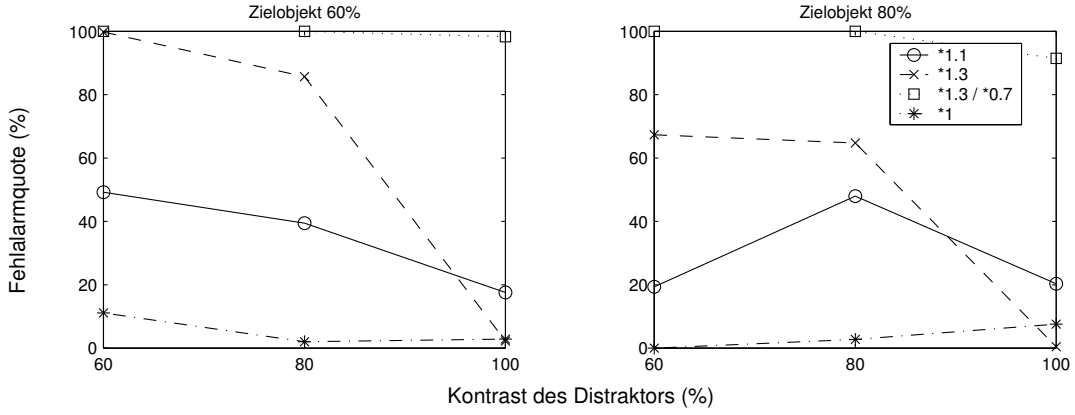

Abbildung 5-37: Fälschliche „Erkennung“ von nicht im Bild gezeigten Büroklammer-Stimuli aufgrund von Aufmerksamkeitseffekten. Die Abbildung zeigt Fehlalarmquoten im Aktivste VTU-Paradigma, d.h. die relative Häufigkeit des Ereignisses, daß die VTU, deren Afferenzen in ihrer Aktivität verstärkt wurden, stärker als alle anderen auf Büroklammern trainierten VTUs feuert, obwohl ihr bevorzugter Stimulus nicht im Bild vorhanden ist. Kontraste des Zielobjekts sind über den Diagrammen angegeben, der Kontrast des Distraktors variiert entlang der x-Achse. Die Legende gibt die Werte der multiplikativen Verstärkung der Aktivität der C2-Afferenzen an (erster Wert) sowie den Wert der Suppression aller übrigen C2-Zellen (zweiter Wert, wenn vorhanden). Jede VTU hat 40 C2-Afferenzen.

Büroklammern gefunden (nicht abgebildet). Das verdeutlicht, wie auch schon die Aktivitätserhöhungen aller C2-Zellen etwa in Abbildung 5-11, daß ein effektiver Aufmerksamkeitsmechanismus nicht notwendigerweise spezifisch ist, oder umgekehrt, daß in diesem Modell keine gezielten, spezifischen Aufmerksamkeitseffekte erforderlich sind, um die Erkennungsleistung zumindest im Stimulusvergleich-Paradigma zu verbessern.

Im Aktivste VTU-Paradigma haben sich andererseits spezifischer wirkende Aufmerksamkeitsmechanismen als geeigneter herausgestellt, um die Erkennungsleistung zu erhöhen, d.h. um die Aktivität der auf das Zielobjekt trainierten VTU über die Aktivitäten der übrigen VTUs hinaus zu erhöhen (siehe etwa 5.3.4). Als der spezifischste Mechanismus von Aufmerksamkeit, der in diesem Kapitel untersucht wurde, kann derjenige betrachtet werden, bei dem die Feuerraten der C2-Afferenzen der auf das Zielobjekt trainierten VTU erhöht werden und alle übrigen C2-Zellen supprimiert werden, da sehr wahrscheinlich alle VTUs außer der, die für das Zielobjekt der Aufmerksamkeit codiert, eine Abschwächung der Aktivität zumindest einiger ihrer Afferenzen erfahren und daher selbst in ihrer Aktivität abgeschwächt werden. Dies wird durch Abb. 5-36 bestätigt: wenn alle außer den verstärkten C2-Zellen supprimiert werden, überträgt sich der Aufmerksamkeitseffekt nicht auf andere Stimuli, auf die sich die Aufmerksamkeit nicht explizit richtet. Abbildung 5-37 macht jedoch die Defizite dieser Form von Aufmerksamkeit sehr deutlich. Hier wurde die Aktivität von C2-Zellen verstärkt, die auf VTUs projizieren, deren bevorzugter Stimulus jeweils *nicht* im Bild gezeigt wurde. Die Prozentwerte auf der Ordinate zeigen, wie oft dennoch die für diesen Stimulus codierende VTU die am stärksten aktive unter den betrachteten VTUs war. Abbildung 5-37 zeigt also Fehlalarme, die im Aktivste VTU-Paradigma durch Aufmerksamkeitseffekte verursacht werden. Es ist offensichtlich, daß bereits für einen multiplikativen Aufmerksamkeitseffekt ohne Suppression, der sich eben (s.o.) als relativ unspezifisch herausgestellt hat, in diesem Paradigma beträchtliche Fehlalarmquoten erreicht werden – für dieselben Kombinationen von Stimuluskontrast und Stärke der Modulation, die zuvor als am effektivsten für die Verbesserung der Erkennungsleistung beschrieben wurden (siehe 5.3.2). Insbesondere wenn zusätzlich noch alle übrigen C2-Zellen supprimiert werden, kann die Fehlalarmquote 100% erreichen, d.h. daß das Objekt, auf das Aufmerksamkeit gerichtet wurde, immer „erkannt“ wird, unabhängig davon, welche Objekte tatsächlich gezeigt werden. Solche Fehlalarmquoten sind natürlich für einen Mechanismus, der die Leistung bei der Objekterkennung erhöhen soll, völlig

inakzeptabel. Hier wurden nur multiplikative Aufmerksamkeitseffekte untersucht. In den Abschnitten 5.3.1 bis 5.3.3 wurde jedoch bereits gezeigt, daß das Addieren konstanter Werte sowie die Verschiebung der Kontrastantwortkurve der S2-Zellen sich als Aufmerksamkeitseffekte faktisch genauso verhalten wie die Multiplikation; und mit den diskutierten alternativen Codierungsmethoden, ebenso wie für Populationscodes und bei der Kategorisierung von Stimuli, traten ähnliche Probleme mit Aufmerksamkeitseffekten auf: wenn sie effektiv waren, erwiesen sie sich als wenig spezifisch oder erhöhten die Wahrscheinlichkeit von Fehlalarmen. Diese Ergebnisse stellen also die Nützlichkeit von Aufmerksamkeit für die Erkennung von Objekten, zumindest im HMAX-Modell, nachhaltig in Frage.

5.4 Diskussion

In diesem Kapitel wurden einige experimentell motivierte Modelle von merkmals- oder objektorientierter Aufmerksamkeit in HMAX daraufhin untersucht, ob sie die Leistung des Modells bei der Objekterkennung für geringen Kontrast und störende Hintergrundstimuli verbessern können. Die primäre Fragestellung war dabei, ob die verbesserte Objekterkennungsleistung, die in psychophysikalischen Experimenten beobachtet wird, wenn genauere Vorabinformationen über einen zu detektierenden Stimulus verfügbar sind, durch die Änderungen der neuronalen Aktivität, wie sie in physiologischen Experimenten mit objektorientierter Aufmerksamkeit beobachtet werden, erklärt werden können. In 5.1 wurde zunächst gezeigt, daß die Annahme, Aufmerksamkeit könne auf die Prozesse der Objekterkennung Einfluß nehmen, einem Paradigma der „frühen Selektion“ von Merkmalen durch Aufmerksamkeit entspricht und daß etwaige solche Einflüsse aufgrund der äußerst kurzen Zeit, in der Objekterkennung abgeschlossen werden kann, nur als Vorbereitungen des visuellen Systems auf die bevorzugte Verarbeitung des erwarteten Stimulus denkbar sind – etwa als Erhöhung der Spontanaktivität oder Verstärkung der synaptischen Gewichte relevanter Neuronen – , um dann die Objekterkennung in einem reinen „Feedforward“-Mechanismus zu ermöglichen. Dafür aber wird hinreichend genaues Vorwissen über den zu erkennenden Stimulus benötigt.

In den hier präsentierten Simulationen mit dem HMAX-Modell konnten in der Tat Verbesserungen der Leistung bei der Objekterkennung durch solche gezielte, im voraus applizier-

te Modulationen der Aktivitäten von Modellneuronen erreicht werden. Allerdings hängt der Erfolg dieser Modulationen im Hinblick auf die erzielte Objekterkennungsleistung für alle untersuchten Mechanismen stark davon ab, wie die Stärke der Veränderung der neuronalen Feuerraten relativ zum Kontrast des Stimulus gewählt wird. Der Kontrast eines zu erkennenden Objekts ist in der Regel aber nicht im voraus bekannt. Weiterhin können die hier diskutierten Aufmerksamkeitseffekte nicht oder nur in geringem Maße die störenden Effekte von Distraktorstimuli kompensieren, deren Kontrast höher ist als der des zu erkennenden Objekts. Während dieses Ergebnis insofern akzeptiert werden kann, als physiologische Experimente nahelegen, daß Aufmerksamkeit die Antworten von Neuronen ohnehin nur bei geringen Kontrasten verändert [52], bliebe dabei allerdings ungeklärt, wie Objekte überhaupt noch erkannt werden können, wenn sie zusammen mit Distraktoren höheren Kontrasts präsentiert werden. Alternative Codierungsmethoden wurden untersucht, um die Abhängigkeit der Erkennungsleistung des Modells vom Stimuluskontrast zu reduzieren und die effiziente Anwendung von Aufmerksamkeitseffekten auch ohne genaue Abstimmung ihrer Stärke mit dem Stimuluskontrast zu ermöglichen; sie erwiesen sich jedoch insbesondere für realistische, relativ ähnliche Stimuli als generell zu wenig spezifisch – selbst unter idealen Bedingungen – und als sehr anfällig für Fehlalarme in der Objekterkennung. Die Codierung von Stimuli durch Populationen von VTUs bzw. die Kategorisierung von Objekten stellten sich als eher weniger bzw. praktisch nicht durch Aufmerksamkeit beeinflußbar heraus. Und schließlich bleibt in einem Mechanismus der frühen Selektion durch Aufmerksamkeit, vor allem für wenig konkrete Vorabinformationen, das Problem der richtigen Auswahl von Merkmalen bzw. zu modulierenden Neuronen. Die hier durchgeführten Simulationen legen allerdings nahe, daß, selbst wenn eine sinnvolle Auswahl getroffen werden kann – wie hier die Wahl der Afferenzen einer VTU, die auf das Zielobjekt trainiert wurde – , in vielen Fällen dieselbe Verbesserung der Leistung durch eine unspezifische generelle Erhöhung der Feuerraten aller Neuronen erreicht werden kann. In solchen Fällen wäre natürlich die Wahl der Neuronen, deren Aktivität durch Aufmerksamkeit verstärkt werden soll, auch für sehr allgemeine Hinweise auf das Zielobjekt unproblematisch, allerdings könnte man dann auch nicht mehr von einem gezielten Effekt von Aufmerksamkeit sprechen, vielmehr von einer Kompensation für geringen Stimuluskontrast. Sehr gezielte Aufmerksamkeitseffekte auf die Objekterkennung neigen andererseits dazu, die Fehlalarmquote deutlich zu erhöhen.

Ein Mechanismus der frühen Selektion von Merkmalen durch Aufmerksamkeit müßte üblicherweise ohne genaue Kenntnisse über den Kontrast eines zukünftigen Stimulus oder sogar über sein genaues Aussehen operieren. Die Erkennung von Objekten sollte auch in Gegenwart von Distraktoren höheren Kontrasts möglich sein. Der Aufmerksamkeitsmechanismus sollte ferner für ein Zielobjekt oder eine Objektklasse selektiv sein, um wirklich die jeweils irrelevante visuelle Information auszublenden, was schließlich eine der Hauptaufgaben von Aufmerksamkeit ist. Dabei dürfen aber die Fehlalarmquoten nicht so stark steigen, daß das Risiko besteht, einen erwarteten Stimulus auf alle Fälle zu „erkennen", auch wenn er gar nicht vorhanden ist. Die hier diskutierten Resultate sprechen also gegen eine Rolle von Aufmerksamkeit, wie sie hier modelliert wurde, in der Objekterkennung, wie sie in HMAX modelliert wird.

Natürlich können diese Ergebnisse nur dann als Hinweise darauf gewertet werden, wie Aufmerksamkeit und Objekterkennung im Gehirn tatsächlich organisiert sein könnten, wenn angenommen wird, daß die Codierung von Stimuli im Gehirn und in HMAX zumindest ansatzweise ähnlich sind. Dem ist nicht notwendigerweise so; schließlich wurden in HMAX etwa keinerlei dynamische Eigenschaften einzelner Neuronen oder neuronaler Netzwerke modelliert. Während experimentell gezeigt wurde, daß Aufmerksamkeit die Aktivität realer Neuronen auf ähnliche Weise erhöht wie eine Zunahme des Stimuluskontrasts [52], könnte der Mechanismus *in vivo* sich sehr von der hier verwendeten einfachen Erhöhung der Feuerrate aller ausgewählten Neuronen, die relativ unabhängig vom tatsächlichen sensorischen Input ist, unterscheiden, und auch die Auswirkungen von Aufmerksamkeit dürften im Kontext eines realen neuronalen Netzwerks andere sein als das Rearrangieren eines Ausgabevektors reeller Zahlen. Beispielsweise werden Disinhibition [43] und Modifikationen synaptischer Stärken [53] als mögliche physiologische Mechanismen von Aufmerksamkeit diskutiert, und die Effekte veränderter Feuerraten von Neuronen auf die Dynamik und stabilen Aktivitätszustände neuronaler Netzwerke können sehr wohl von Bedeutung für die Wahrnehmung von Stimuli sein. Letztendlich ist freilich die genaue Beziehung zwischen neuronaler Aktivität einerseits und der Repräsentation und Wahrnehmung von Stimuli andererseits nicht bekannt. Dennoch können die Resultate dieses Kapitels dazu dienen, die mögliche Rolle von Aufmerksamkeit bei der Objekterkennung einzugrenzen. Das Problem, wie Aufmerksamkeit spezifisch und selektiv wirken kann, ohne dabei zuviele Fehlalarme zu verursachen, wird sehr wahrscheinlich un-

abhängig von der verwendeten Methode, Stimuli zu codieren, auftreten – vorausgesetzt, Aufmerksamkeit beeinflußt tatsächlich die Objekterkennung. Und die Frage, wie die passende Stärke einer Modulation von Feuerraten durch Aufmerksamkeit im voraus gewählt werden kann, wird immer dann zu beantworten sein, wenn absolute und relative Feuerraten von Neuronen von Bedeutung dafür sind, welche Stimuli repräsentiert werden. Es ist sehr wahrscheinlich, daß letzteres im Gehirn der Fall ist, da, wie gezeigt wurde, der Verzicht auf die in den Feuerraten enthaltene Information die Leistung bei der Objekterkennung, insbesondere für ähnlich aussehende Stimuli, drastisch reduziert.

Während es also die hier präsentierten Ergebnisse fraglich erscheinen lassen, ob objektorientierte Aufmerksamkeit sinnvoll auf Prozesse der Objekterkennung einwirken kann, sind sie im Einklang mit einer alternativen Hypothese, nämlich der, daß Aufmerksamkeit einen Prozeß der Selektion von Stimuli darstellt, *nachdem* diese als individuelle Objekte erkannt worden sind. Eine große Zahl experimenteller Resultate legt eine ähnliche Schlußfolgerung nahe. Wie bereits in 5.1.2 erwähnt, finden die meisten entsprechenden Studien bis etwa 150 ms nach Beginn der Stimuluspräsentation keine Effekte von objektorientierter Aufmerksamkeit in visuellen Arealen von V1 bis V4 [8, 20, 46, 58]. Während in IT bereits vor der eigentlichen Stimuluspräsentation eine erhöhte Spontanaktivität von Neuronen beobachtet werden kann, deren bevorzugter Stimulus als zu detektierendes Zielobjekt ausgewiesen wurde, sind die Aktivitäten dieser Neuronen während der ersten Antwortphase mit und ohne Aufmerksamkeit identisch und beginnen erst nach wiederum etwa 150 ms zu divergieren [7]. Da Ergebnisse aus ERP-Studien dafür sprechen, daß bis zu diesem Zeitpunkt Objekte selbst in einer komplexen, realistischen Umgebung erfolgreich erkannt und kategorisiert werden können [67, 74], kann argumentiert werden, daß die neuronale Aktivität in der ersten Antwortphase, die nicht von objektorientierter Aufmerksamkeit moduliert wird, entscheidend für die Objekterkennung ist [66]. Das deutet darauf hin, daß das visuelle System bei der Objekterkennung womöglich gar nicht auf die Unterstützung durch Aufmerksamkeit angewiesen ist, selbst unter schwierigen visuellen Bedingungen. RSVP-Experimente mit gleichzeitigen Einzelzellableitungen im oberen temporalen Sulcus (STS) von Affen zeigen auch, daß die Neuronen in diesem höheren Areal des ventralen visuellen Systems auch dann noch selektiv auf ihre bevorzugten Stimuli reagieren, wenn diese nur für jeweils 14 oder 28 ms in einem kontinuierlichen Bilderstrom gezeigt werden [28]. Wenn also die Selektivität neuronaler Aktivität selbst unter solchen extremen Bedin-

gungen zum großen Teil erhalten bleibt, scheint der Einsatz von Aufmerksamkeit zur Verbesserung der Objekterkennungsleistung in der Tat nicht notwendig zu sein. Ein weiteres Argument könnte darauf aufbauen, daß die Objekterkennung im visuellen System, im Gegensatz zu HMAX, in hohem Maße unabhängig vom Kontrast des jeweiligen Stimulus ist [1]. Andererseits scheint der Effekt von Aufmerksamkeit, unabhängig von den zugrundeliegenden physiologischen Mechanismen, eine Erhöhung des effektiven Stimuluskontrasts zu sein [52]. Wenn aber Objekterkennung zum großen Teil nicht vom Kontrast abhängt, ist es schwer vorstellbar, daß ein Mechanismus, der den effektiven Kontrast eines Stimulus verändert, dessen Erkennung erleichtern könnte.

Ähnliche Folgerungen ergeben sich auch aus in der Psychophysik gewonnenen Daten. Bereits seit langem wird argumentiert, daß kurzzeitig präsentierte Stimuli im RSVP-Experiment tatsächlich erkannt und in einem konzeptuellen Kurzzeitgedächtnis gespeichert werden können, aber aufgrund der Verarbeitung nachfolgender Stimuli normalerweise schnell wieder vergessen werden [9, 51]. Vorabinformation über einen zu detektierenden Stimulus verbessert dessen Erkennung deutlich, verglichen mit der Leistung, die beobachtet wird, wenn Versuchspersonen lediglich den RSVP-Bilderstrom sehen und anschließend gefragt werden, ob ein bestimmter Stimulus darin enthalten war. Weiterhin können, wie bereits erwähnt, auch sehr abstrakte Hinweise auf das zu erkennende Zielobjekt, die keinerlei Informationen über sein genaues Aussehen enthalten, dessen Detektion im RSVP--Experiment signifikant verbessern [23]. Während diese Ergebnisse einen Mechanismus der frühen Selektion von Merkmalen durch Aufmerksamkeit, der im voraus die Verarbeitung der Merkmale des Zielobjekts selektiv begünstigt, nicht ausschließen, sprechen vor allem die Resultate für vergleichsweise vage Vorabinformationen nachdrücklich gegen einen solchen Mechanismus. Schließlich wird in der Regel nur die Erkennung des Zielobjekts selektiv begünstigt, und andere Stimuli der schnellen Bilderfolge im RSVP-Experiment werden eher schlechter erkannt (oder vielmehr erinnert) [51]. Des weiteren erhöht sich die Fehlalarmquote nicht, wenn im voraus das Aussehen des Zielobjekts und nicht nur eine verbale Bezeichnung bekannt ist, obwohl diese Form von Vorabinformation deutlich spezifischer ist [51]. Diese Ergebnisse stehen diametral im Widerspruch zu den Erwartungen, die sich aus der vorliegenden Studie über Mechanismen der frühen Selektion durch Aufmerksamkeit ergeben, nämlich daß die Effekte solcher Mechanismen sich entweder, aufgrund gemeinsamer Merkmale verschiedener Objekte, auch auf andere Stimuli als das Zielob-

jekt übertragen sollten – und zwar um so mehr, je weniger spezifisch die Aufmerksamkeit aufgrund der verfügbaren Vorabinformation wirken kann – , oder daß sich die Anzahl der Fehlalarme erhöhen sollte, wenn die Vorabinformation über das Zielobjekt genauer ist. Allerdings ist dabei anzumerken, daß in RSVP-Studien üblicherweise nicht die Ähnlichkeit von Zielobjekten und Distraktorstimuli kontrolliert wird. Daß die Aufmerksamkeitseffekte sich dabei nicht auf andere Stimuli übertragen bzw. die Fehlalarmquote sich nicht entsprechend der Vorabinformation verändert, könnte also auch eine Folge davon sein, daß die gezeigten Stimuli sich voneinander stark unterscheiden und ihre neuronalen Repräsentationen wenig überlappen. Dies könnte am ehesten in einem RSVP-Experiment mit Morphing-Stimuli wie in dieser Studie überprüft werden. Eine eventuell erhöhte Fehlalarmquote für genauere Vorabinformation und ähnlichere Stimuli alleine könnte jedoch noch nicht als Beleg für Mechanismen früher Selektion gewertet werden. Es sollte nicht erwartet werden, daß das visuelle System auch bei sehr rascher Präsentation noch problemlos zwischen ähnlichen Stimuli in komplexen Darstellungen mit störendem Hintergrund unterscheiden kann. Auch ohne Aufmerksamkeitsmechanismen, die vor der Objekterkennung agieren, wäre es nicht überraschend, wenn die Fehlalarmquote in solchen Situationen zunähme.

Auch andere RSVP-Experimente stärken die Hypothese einer schnellen Objekterkennung ohne Einfluß von Aufmerksamkeit. Wird etwa eine rasche Folge einzelner Wörter präsentiert, in der ein vorher bekanntes Wort detektiert werden soll, so erinnern sich Versuchspersonen anschließend häufiger an solche unter den übrigen gezeigten Wörtern, die semantisch mit dem „Zielwort“ verwandt sind [26]. Ein vor der Erkennung bzw. dem Verstehen der gezeigten Wörter operierender Aufmerksamkeitsmechanismus ist hier noch weniger realistisch als in den bisher diskutierten Fällen, da visuelle Merkmale nicht genutzt werden können, um ein semantisch verwandtes Wort unter anderen Wörtern zu identifizieren. Eine Form von Aufmerksamkeit, die hier die Erkennungsleistung verbessert, muß also auf der Ebene konzeptueller Repräsentationen des visuellen Inputs operieren, d.h. nach der „Erkennung“ seines semantischen Gehalts. ERP-Studien deuten außerdem darauf hin, daß im RSVP-Experiment auch solche Stimuli semantisch verarbeitet werden, die in der Periode unmittelbar (bis etwa 300 ms) nach der Erkennung eines Zielobjekts gezeigt werden, obwohl Versuchspersonen sich später nicht an diese Stimuli erinnern können (dieses Phänomen ist als der „attentional blink“ bekannt) [37]. Dies deutet sehr darauf hin, daß vi-

suell präsentierte Stimuli in der Tat sehr rasch bis zu einem hohen Grad an Abstraktion verarbeitet werden können, auch wenn sie möglicherweise später nicht bewußt abrufbar sind. Daß sich die Leistung im RSVP-Experiment erhöht, wenn spezifischeres Vorwissen über künftige Stimuli verfügbar ist, ist also sehr wahrscheinlich nicht darauf zurückzuführen, daß ansonsten nicht wahrgenommene Stimuli nun erkannt werden können, sondern vielmehr darauf, daß erkannte Stimuli, die mit den Erwartungen übereinstimmen, eher im Gedächtnis behalten werden.

Man könnte jedoch aus den Resultaten der RSVP-Experimente auch ein Argument für einen Mechanismus früher Selektion durch Aufmerksamkeit ableiten. Wie bereits bekannt, führen genauere Vorabinformationen über ein Objekt zu dessen besserer Erkennung (bzw. stabilerer Erinnerung) und auch zu kürzeren Reaktionszeiten im Experiment [23]. Wenn aber die Objekterkennung bei einer gewissen Geschwindigkeit der Bildpräsentation im RSVP-Experiment üblicherweise ohnehin für jedes Bild vollständig abgeschlossen werden kann, und zwar innerhalb von etwa 150 ms, ist kein Grund ersichtlich, weshalb unterschiedliches Vorwissen einen Einfluß auf die Anzahl korrekt erkannter und erinnerter Objekte oder die Reaktionsdauer haben sollte. Wie allerdings schon Intraub betont, wird die Entscheidung, ob eine vor dem Experiment gegebene Beschreibung des Zielobjekts auf einen Stimulus paßt oder nicht, um so schwieriger, je allgemeiner diese Beschreibung ist[1] [23]. Anders gesagt, ist es sicherlich einfacher zu entscheiden, ob ein gegebener Stimulus Element einer wohldefinierten, klar umgrenzten Menge von Objekten ist (wie etwa „Menge aller Hunde“), als etwa zu bestimmen, ob er *nicht* zu einer weniger eindeutig bestimmten, allgemeineren Menge (wie etwa „Transportmittel“) gehört. Insbesondere wenn die Codierung von Stimuli, wie im Gehirn, nicht frei von Störungen und Rauschen ist, dürfte die Lösung einer Aufgabe letzteren Typs längere Integrationszeiten erfordern und anfälliger für Fehler sein, selbst wenn die Objekterkennung an sich erfolgreich war. Die Antwort auf die zu Beginn dieser Arbeit formulierte Frage, ob die von objektorientierter Aufmerksamkeit verursachten unterschiedlichen Aktivitäten von Neuronen die unterschiedlichen Leistungen von Versuchspersonen bei der Detektion von Stimuli mit mehr oder weniger genauem (oder gar keinem) Vorwissen erklären können, lautet also mit hoher Wahrscheinlichkeit, daß die beobachteten neuronalen Aufmerksamkeitseffekte keine Auswirkungen

[1] „The process of deciding that the cue and the target picture match increases in complexity as less specific cues are provided.“ [23], S. 609.

auf die Objekterkennung haben und daß die Leistungsänderungen im Verhaltensexperiment auf die unterschiedliche Komplexität des Abgleichs eines Stimulus mit einer mehr oder weniger genauen Beschreibung zurückzuführen sind.

Natürlich kann die Möglichkeit nicht ausgeschlossen werden, daß das visuelle System in bestimmten Fällen dennoch eine frühe Selektion visueller Merkmale durch Aufmerksamkeit durchführt. Es könnten natürlich im Prinzip ausgereiftere Mechanismen für die Wahl relevanter Merkmale bzw. Neuronen sowie der Stärke der Modulation existieren, als in dieser Untersuchung angenommen wurden. Aufgrund der Resultate dieser und anderer Studien kann man jedoch davon ausgehen, daß ein solcher Aufmerksamkeitsmechanismus wohl nur sehr begrenzt eingesetzt werden könnte. Eine mögliche Anwendung geht etwa aus diesem Kapitel hervor, und zwar die Erkennung eines Zielobjekts aus einer begrenzten Menge trainierter und gut bekannter Stimuli, eventuell mit im voraus bekanntem Kontrast. So könnte etwa ein Affe auf eine Menge für ihn neuartiger Stimuli aus zwei unterschiedlichen Kategorien trainiert werden und lernen, sowohl individuelle Stimuli als auch die beiden Kategorien zu unterscheiden. Man müßte dann nach Neuronen etwa im Areal V4 suchen, die selektiv auf einen dieser Stimuli (oder, wie man vermuten könnte, auf ein komplexes Merkmal dieses Stimulus) ansprechen und überprüfen, ob diese Neuronen schon vor der tatsächlichen Stimuluspräsentation erhöhte Aktivität zeigen, wenn ihr bevorzugter Stimulus oder die Kategorie, der er angehört, im Verhaltensexperiment als zu detektierendes Objekt bzw. zu detektierende Kategorie vorgegeben werden. Bisherige Experimente finden üblicherweise keine derartige vorweggenommene Aktivitätserhöhung in V4 [8]; möglicherweise könnte sich das bei Verwendung weniger, gut trainierter Stimuli ändern. Insbesondere wäre es von Interesse, ob in einem solchen Fall bei der Vorgabe einer Stimuluskategorie für das Detektionsexperiment erhöhte Spontanaktivität bei Zellen in Arealen wie V4 und IT, die auf einzelne Stimuli dieser Kategorie gut ansprechen, beobachtet werden kann. Es ist anzunehmen, daß etwa in V4 und IT noch keine explizite Information über Stimuluskategorien repräsentiert ist [29]. Verstärkte Aktivität solcher Zellen aufgrund von „top-down"-Information über eine relevante Stimuluskategorie, so sie denn gefunden werden kann, wäre ein Hinweis darauf, daß abstraktere Informationen über Objektkategorien auf die konkretere Ebene relevanter Merkmale umgesetzt werden können, was eine entscheidende Voraussetzung für einen Mechanismus früher Selektion von Merkmalen durch Aufmerksamkeit wäre.

Es gibt in der Tat Studien, die Belege für solche im voraus operierenden Aufmerksamkeitsmechanismen finden. Haenny *et al.* [17] berichten von erhöhter Spontanaktivität von Neuronen in V4, wenn eine von diesen Neuronen bevorzugte Orientierung eines Gitterstimulus als Orientierung eines zu detektierenden Zielobjekts vorgegeben wird. Da aber in dieser Untersuchung durchweg nur orientierte Gitterstimuli verwendet wurden und die Aufgabe des Versuchstiers in der Einschätzung ihrer Orientierung bestand, können die Resultate kaum als Belege für eine eventuelle Rolle von Mechanismen früher Selektion durch Aufmerksamkeit in der Erkennung komplexer Objekte gewertet werden. Psychophysikalische Studien, die kurzzeitige Präsentationen *einzelner* Bilder verwenden, finden oft eine erhöhte Erkennungsleistung, wenn das Objekt, das im Bild erkannt werden soll, im voraus bekannt ist [2, 48]. In einer strikten Interpretation der Theorie, daß für kurze Zeit präsentierte Stimuli rasch erkannt werden können, sollte ein einzelner Stimulus entweder erkannt werden oder nicht, unabhängig von eventuellem Vorwissen, und er sollte auch nicht sogleich wieder vergessen werden, da im Gegensatz zum RSVP-Experiment keine nachfolgenden Stimuli die Ressourcen des visuellen Systems in Anspruch nehmen. Änderungen in der Leistung aufgrund unterschiedlichen Vorwissens in einem solchen Experiment könnten daher in der Tat auf einen Einfluß von Aufmerksamkeit auf die Objekterkennung selbst zurückzuführen sein. Die in diesen Experimenten festgestellten Verbesserungen der Erkennungsleistung sind jedoch nur dann substantiell, wenn das genaue Bild des Zielobjekts im voraus präsentiert wurde. Wie sowohl Pachella als auch Potter in Übereinstimmung mit den hier angestellten Überlegungen deutlich machen, sind mögliche derartige Mechanismen der frühen Selektion sehr wahrscheinlich auf solche Situationen beschränkt, in denen sehr spezifische Informationen über den zu erwartenden Stimulus zur Verfügung stehen [48, 51]. Zudem folgen in derartigen Experimenten das Bild des Zielobjekts und die eigentliche Stimuluspräsentation in der Regel unmittelbar aufeinander, und die Aufgabe, die die Versuchsperson zu bewältigen hat, ist normalerweise lediglich eine Einschätzung, ob die beiden Stimuli übereinstimmen oder nicht. Man kann vermuten, daß in solchen Fällen auch auf der Basis ganz elementarer visueller Merkmale der beiden Abbildungen korrekt entschieden werden kann, ob sie sich unterscheiden oder nicht, und daß auch ein Aufmerksamkeitseffekt, der ansonsten leicht zu Fehlalarmen führt, hier effizient sein kann. Diese Studien verdeutlichen also wiederum, daß eventuelle vor der eigentlichen Objekterkennung operierende Mechanismen von Aufmerksamkeit sehr wahrscheinlich nur in

ganz bestimmten Situationen von Nutzen sein können, in denen detaillierte Vorabinformationen verfügbar und die einzelnen Stimuli hinreichend unterscheidbar sind sowie die zu lösende Aufgabe in bezug auf die in dieser Arbeit diskutierten Nachteile solcher Mechanismen der frühen Selektion „fehlerverzeihend“ ist.

Insgesamt scheint eine Theorie, die Aufmerksamkeit auf neuronaler Ebene als einen Mechanismus beschreibt, der bereits erkannte Stimuli selektiert und nach ihrer jeweiligen Bedeutung gewichtet, besser mit den verfügbaren experimentellen Daten ebenso wie mit den Resultaten dieser Arbeit übereinzustimmen als die Annahme, Aufmerksamkeit wirke sich auf die Objekterkennung selbst aus. Die hohe Präzision und Verarbeitungsgeschwindigkeit des visuellen Systems, selbst unter den schwierigen Bedingungen natürlicher Stimulation, machen einen Einfluß von Aufmerksamkeit auf die Erkennung von Stimuli, der ohne adäquates Vorwissen die Leistung sogar eher verschlechtern könnte, offenbar unnötig. Wenn aber die Erkennung der im visuellen Feld präsenten Objekte abgeschlossen ist, können die Mechanismen der objektorientierten Aufmerksamkeit zur Verstärkung und Abschwächung der Aktivität verschiedener Neuronen sehr effizient eingesetzt werden, um bestimmte Stimuli bevorzugt zu repräsentieren und für eventuelle Verhaltensreaktionen zu selektieren. Es hat sich gezeigt, daß Stimuli in der Tat durch Modulationen der Feuerraten von Neuronen in den „Vordergrund“ oder „Hintergrund“ gebracht werden können, wenn es für die momentane Wahrnehmung entscheidend ist, welche Neuronen oder Neuronenpopulationen am stärksten aktiv sind (analog zum Aktivste VTU-Paradigma der Erkennung von Stimuli in HMAX). Wenn bereits bekannt ist, welche Stimuli im visuellen Feld präsent sind, können derartige Mechanismen angewendet werden, ohne dabei Fehler in der Objekterkennung zu riskieren, da die Auswahl von Merkmalen, deren neuronale Repräsentation verstärkt oder abgeschwächt werden soll, sehr viel einfacher sein dürfte, wenn „top-down“-Informationen unmittelbar mit dem sensorischen Input verglichen und einzelne Merkmale den verschiedenen Stimuli zugeordnet werden können. Dies ist ein vielversprechender möglicher Mechanismus für objektorientierte Aufmerksamkeit, der auch bereits im Modell untersucht wurde [73]. Dabei ist anzumerken, daß eine von Aufmerksamkeit induzierte Verringerung der Feuerraten von Neuronen, die für erkannte, aber momentan irrelevante Stimuli codieren, nicht bedeuten muß, daß diese Stimuli nicht mehr „erkannt“ würden. Wenn die Objekterkennung, wie zuvor behauptet, in hohem Maße kontrastinvariant ist, sollte eine solche Veränderung des effektiven Stimu-

luskontrasts die Erkennung dieses Stimulus selbst nicht beeinflussen, und selbst in HMAX, wo die Objekterkennung stark vom Kontrast abhängt, ist es die *relative* Aktivität von VTUs, die für das Vorhandensein von Objekten codiert. Mit einer angemessenen Wahl der Modulationsstärken, wie man sie sich auf der Basis tatsächlicher sensorischer Informationen besser vorstellen kann als nur aufgrund von Vorabinformation, kann die relative Aktivität bei Veränderung der absoluten durchaus erhalten bleiben.

Das Bild der visuellen objektorientierten Aufmerksamkeit, das sich aus dieser Studie ergibt, ist also das eines Mechanismus, der „top-down"- mit „bottom-up"-Information vergleicht und übereinstimmende Aktivitätsmuster selektiert, nachdem die eingehende visuelle Information bis zu einem genügend abstrakten, konzeptuellen Niveau verarbeitet wurde, um auch Vergleiche mit sehr abstrakter „top-down"-Information zu ermöglichen. Nachdem objektorientierte Aufmerksamkeitseffekte ins Spiel gekommen sind, also etwa ab 150 ms nach Beginn der visuellen Stimulation, wenn die Objekterkennung höchstwahrscheinlich bereits abgeschlossen ist, zeigt die neuronale Aktivität in höheren Arealen des ventralen visuellen Systems wie V4 nicht mehr nur, welche Objekte im visuellen Feld vorhanden sind, sondern auch und vor allem, welche davon gerade für das Verhalten des Organismus relevant sind. V4 etwa könnte so als eine Art „saliency map" verstanden werden. Das Konzept der „saliency map" bezieht sich üblicherweise auf eine topographische neuronale Repräsentation derjenigen Positionen im visuellen Feld, die sich in der „bottom-up"-Verarbeitung des sensorischen Inputs als besonders hervorstechend (engl. salient) erwiesen haben, aufgrund von Merkmalen wie Kontrast, Farbe oder Helligkeit [25]. In höheren Arealen des ventralen visuellen Systems geht aufgrund der immer größeren rezeptiven Felder Positionsinformation jedoch mehr und mehr verloren, und welche Stimuli gegenüber anderen herausragen, wird hier nicht nur durch ihre sensorischen Qualitäten, sondern eben insbesondere durch die „top-down"-Selektion durch Aufmerksamkeitsmechanismen bestimmt. Areale wie V4, aber auch schon V1 und V2 [34], zeigen daher Eigenschaften von Objekt-„saliency maps", die sowohl die im visuellen Feld präsenten Stimuli codieren als auch, durch den Einfluß „top-down" agierender Mechanismen objektorientierter Aufmerksamkeit, deren jeweilige Relevanz für das Verhalten.

Kapitel 6

Zusammenfassung und Ausblick

Ziel der vorliegenden Arbeit war es, mit Hilfe des HMAX-Modells der Objekterkennung im visuellen Cortex [55] ein besseres Verständnis der Invarianzeigenschaften hierarchischer neuronaler Systeme gegenüber verschiedenen Stimulustransformationen zu gewinnen sowie die möglichen Auswirkungen neuronaler Effekte von visueller objektorientierter Aufmerksamkeit auf die Leistungen bei der Objekterkennung unter schwierigen visuellen Bedingungen zu untersuchen. Dazu wurden zunächst die Aktivitäten der Modellneuronen sowie die Erkennungsleistung bei der Translation, Skalierung und Kontraständerung von abstrakten (Büroklammern) und realistischen (Autos) Stimuli sowie bei der Präsentation eines Objekts mit einem Distraktorstimulus bestimmt. Es zeigt sich, daß aufgrund der begrenzten neuronalen Ressourcen des Modells sowie der hierarchischen Organisation die Invarianz der Antworten auf Stimuli begrenzt ist. Für verschiedene Stimulusklassen ergeben sich dabei unterschiedliche Invarianzeigenschaften des Modells. Der Invarianzbereich gegenüber Translation wird stets durch Gleichung 3.1 beschrieben, unabhängig vom verwendeten Stimulus, während die Invarianzbereiche gegenüber Skalierung, Kontraständerung und visuellem Hintergrund davon abhängen, wie gut die vom Modell detektierten visuellen Merkmale die tatsächlich bei der jeweiligen Objektklasse vorhandenen Merkmale repräsentieren. Eine gute Übereinstimmung zwischen den detektierten und den tatsächlichen Merkmalen ist Voraussetzung dafür, daß auf verschiedene Objektgrößen spezialisierte Neuronen auch tatsächlich von Stimuli der entsprechenden Ausmaße bevorzugt aktiviert werden. Nur wenn das der Fall ist und auf entsprechend an-

dere Größen selektiv ansprechende Neuronen vorhanden sind, können skalierte Stimuli gut erkannt werden. Die Detektion visueller Merkmale, die gut an die präsentierten Objekte angepaßt sind, führt auch zu neuronalen Repräsentationen individueller Objekte, die sich stärker voneinander unterscheiden und daher robuster gegenüber verringertem Stimuluskontrast oder störendem visuellen Hintergrund sind.

Neuronale Spezifitäten für Merkmale, die gut an den jeweiligen visuellen Input angepaßt sind, kommen im visuellen System sehr wahrscheinlich durch Lernvorgänge zustande. Auch die Invarianzeigenschaften einzelner Neuronen können sich durch Lernen und abhängig von der zu erfüllenden Aufgabe verändern. Das Gehirn verfügt über massive Ressourcen zur parallelen Prozessierung visueller Information und über entsprechendes „Training" mit den relevanten visuellen Stimuli. Es kann daher nicht verwundern, wenn ein neuronales Modell nicht gleich gute Erkennungsleistungen für verschiedene Objektklassen zeigt. In Anbetracht des riesigen „Trainingsvorsprungs" realer visueller Systeme wäre es also durchaus gerechtfertigt, ein Modell wie HMAX jeweils für die in einem Experiment verwendeten Stimulusklassen durch Lernalgorithmen wie in [62] selektiver zu machen. Es wäre sehr interessant zu untersuchen, welche und wieviele Merkmalsdetektoren in HMAX realisiert werden müßten, um die Vielfalt der realen visuellen Stimuli, denen wir ausgesetzt sind, ansatzweise angemessen repräsentieren, erkennen und unterscheiden zu können. Dabei wäre auch zu überprüfen, ob mit entsprechend zahlreicheren und komplexeren detektierten Merkmale immer noch, wie in [54] für die Büroklammer-Stimuli gezeigt, alleine die Lokalität der komplexen Merkmale und die Anordnung der einfacheren Merkmale in ihnen ausreicht, um die Zuordnung einzelner Merkmale zu ganzen Objekten zu ermöglichen oder ob Mechanismen des „Binding" notwendig würden, um zu vermeiden, daß ein besonders hervorstechendes Merkmal eines Stimulus die Repräsentation eines anderen stört. Dies wird in HMAX nicht explizit verhindert und kann etwa bei der Erkennung eines Autostimulus in Gegenwart eines zweiten zu Problemen bei der Erkennung führen.

Die Interaktion visueller Stimuli auf neuronaler Ebene war auch das Thema der Simulationen in dieser Arbeit, die die Experimente von Reynolds *et al.* [53] modellierten. Dabei galt es herauszufinden, wie sich die Antwort von Neuronen auf einzelne Balkenstimuli und Balkenkombinationen, wie von Reynolds *et al.* verwendet, in Abhängigkeit von deren Ab-

stand voneinander ändern kann. Während sich in der Tat das Antwortverhalten ein und desselben Neurons, das seine Afferenzen mit einer MAX-Operation verarbeitet, bei verschiedenen Balkenabständen verändern kann, und zwar lediglich aufgrund unterschiedlicher Interaktionen der Stimuli auf der Ebene der Afferenzen dieses Neurons, konnten keine Belege dafür gefunden werden, daß Neuronen, wie sie in HMAX modelliert werden, unter bestimmten Bedingungen konsistent die von Reynolds *et al.* gefundenen, auf kompetitive Interaktionen hindeutenden Antworteigenschaften an den Tag legen. Andere experimentelle Daten weisen jedoch darauf hin, daß im von Reynolds *et al.* untersuchten Areal V4 auch eine den Vorhersagen des HMAX-Modells entsprechende Maximumsoperation realisiert sein könnte [16]. Die wahrscheinlichste Schlußfolgerung aus diesen Resultaten ist, daß verschiedene Neuronen in Arealen wie V4 verschiedene Aufgaben erfüllen und demzufolge natürlich auch unterschiedliche Operationen über ihrem jeweiligen Input ausführen können. Auch in höheren Arealen wäre die MAX-Operation dabei sehr sinnvoll, um die Spezifität eines Neurons für bestimmte Stimuli auch über die zunehmenden rezeptiven Feldgrößen hinweg, die sich in diesen Arealen finden, zu erhalten.

Schließlich wurden in dieser Arbeit experimentell motivierte neuronale Mechanismen von visueller objektorientierter Aufmerksamkeit, also Veränderungen der Feuerraten von Neuronen in Abhängigkeit davon, ob ihr bevorzugter Stimulus bewußt beachtet wird oder nicht [8, 46, 52], modelliert und daraufhin überprüft, ob sie die in psychophysikalischen Experimenten beobachteten aufmerksamkeitsbedingten Verbesserungen der Objekterkennungsleistung [23, 51] im Rahmen des HMAX-Modells erklären können. Das für die Simulationen gewählte Paradigma der frühen Selektion von Merkmalen durch Aufmerksamkeit, in dem die neuronalen Effekte von Aufmerksamkeit bereits vor der Stimuluspräsentation wirksam werden und das System entsprechend der Erwartung bestimmter Stimuli voreinstellen, ergab sich dabei aus einer Diskussion der Literatur zur Physiologie und Psychophysik der Aufmerksamkeit. Verschiedene Modulationsmechanismen wurden im Kontext unterschiedlicher Codierungsmethoden und simulierter Aufgabenstellungen, nämlich der Erkennung und Unterscheidung einzelner Objekte sowie ihrer Kategorisierung, untersucht. Es stellte sich dabei heraus, daß die Leistung des Modells bei der Objekterkennung zwar in der Tat verbessert werden kann, allerdings in starker Abhängigkeit von Parametern wie dem Stimuluskontrast, die in der Regel nicht im voraus bekannt sind. Weiterhin ergab sich, daß im Modell unter den erwähnten Einschränkungen effiziente Auf-

merksamkeitsmechanismen zumeist entweder zu wenig spezifisch für den jeweils im Zentrum der Aufmerksamkeit befindlichen Stimulus waren oder andererseits ein hohes Risiko bargen, falsch positive Resultate zu liefern, also Stimuli zu „erkennen“, die zwar erwartet, aber nicht präsentiert wurden. In Übereinstimmung mit weiten Teilen der experimentellen Literatur wurde daher gefolgert, daß Stimuli vom visuellen System üblicherweise auch unter schwierigen Bedingungen ohne zusätzliche Effekte von Aufmerksamkeit rasch erkannt werden können und die unterschiedlichen Leistungen mit und ohne bzw. mit unterschiedlich spezifischer Aufmerksamkeit im psychophysikalischen Experiment auf unterschiedlich komplexe Prozesse der Extraktion von Information aus dem visuellen Input zurückzuführen sind. Objektorientierte Aufmerksamkeit ist in der Sichtweise, die sich aus dieser Arbeit ergibt, ein Prozeß der Selektion bereits erkannter Stimuli entsprechend ihrer jeweiligen Relevanz für das Verhalten.

Da objektorientierte Aufmerksamkeit also offensichtlich im Gehirn keine Rolle bei der eigentlichen Objekterkennung spielt, könnte man folgern, daß die Simulation dieser Form von Aufmerksamkeit in einem Modell der Objekterkennung wie HMAX nicht weiter verfolgt werden muß. Es wäre jedoch sehr interessant zu modellieren, wie „top-down“-Informationen über jeweils relevante Stimuli tatsächlich agieren könnten, um in Arealen des ventralen visuellen Systems die neuronale Aktivität so einzustellen, daß die interessierenden Stimuli bevorzugt repräsentiert werden. Dafür würden aber wohl eher erweiterte Modelle benötigt, die auch die dynamischen Aspekte von Neuronen und neuronalen Netzwerken abbilden, ähnlich wie etwa in [72, 73]. Solche Modelle müßten aber wie HMAX die Eigenschaften des visuellen Cortex bei der Objekterkennung berücksichtigen bzw. im Hinblick auf die Objekterkennung interpretierbar sein, um nicht nur die physiologischen Mechanismen der Modulation neuronaler Aktivität zu simulieren, sondern womöglich auch klären zu können, wie trotz des Einflusses von „top-down“-Information die Erkennung von bzw. das Wissen um die tatsächlich präsenten Objekte erhalten bleiben kann. Es ginge letztlich also darum herauszufinden, wie das Gehirn zwischen sensorischer und intern generierter Aktivität unterscheiden kann. Dafür allerdings dürfte noch eine Menge experimenteller und theoretischer Arbeit erforderlich sein.

Literaturverzeichnis

[1] Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., and Malach, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *J. Neurophys.* **87**, 3102–3116.

[2] Biederman, I. (1972). Perceiving real-world scenes. *Science* **177**, 77–80.

[3] Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *SIGGRAPH '99 Proceedings*, 187–194. ACM Computer Soc. Press.

[4] Bruce, C., Desimone, R., and Gross, C. (1981). Visual properties of neurons in a polysensory area in the superior temporal sulcus of the macaque. *J. Neurophys.* **46**, 369–384.

[5] Carr, T. H. and Bacharach, V. R. (1976). Perceptual tuning and conscious attention: Systems of input regulation in visual information processing. *Cognition* **4**, 281–302.

[6] Chelazzi, L. (1999). Serial attention mechanisms in visual search: A critical look at the evidence. *Psych. Res.* **62**, 195–219.

[7] Chelazzi, L., Duncan, J., Miller, E. K., and Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophys.* **80**, 2918–2940.

[8] Chelazzi, L., Miller, E. K., Duncan, J., and Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* **11**, 761–772.

[9] Coltheart, V., editor (1999). *Fleeting Memories: Cognition of Brief Visual Stimuli*. MIT Press, Cambridge, MA.

[10] Connor, C. E., Preddie, D. C., Gallant, J. L., and van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *J. Neurosci.* **17**(9), 3201–3214.

[11] Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *J. Cogn. Neurosci.* **3**, 1–8.

[12] Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Ann. Rev. Neurosci* **18**, 193–222.

[13] Desimone, R. and Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *J. Neurophys.* **57**, 835–868.

[14] Fabre-Thorpe, M., Richard, G., and Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *NeuroReport* **9**, 303–308.

[15] Freedman, D., Riesenhuber, M., Poggio, T., and Miller, E. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316.

[16] Gawne, T. J. and Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to two simultaneously presented stimuli. *J. Neurophys.* In press.

[17] Haenny, P. E., Maunsell, J. H. R., and Schiller, P. H. (1988). State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Exp. Brain Res.* **69**, 245–259.

[18] Haenny, P. E. and Schiller, P. H. (1988). State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. *Exp. Brain Res.* **69**, 225–244.

[19] Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430.

[20] Hillyard, S. A. and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Nat. Acad. Sci. USA* **95**, 781–787.

[21] Hillyard, S. A. and Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Percept. Psychophys.* **36**(2), 185–198.

[22] Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154.

[23] Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *J. Exp. Psych.: Hum. Percept. Perf.* **7**(3), 604–610.

[24] Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophys.* **73**, 218–26.

[25] Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2**(3), 194–203.

[26] Juola, J. F., Prathyusha, D., and Peterson, M. S. (2000). Priming effects in attentional gating. *Mem. Cogn.* **28**(2), 224–235.

[27] Kastner, S. and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Ann. Rev. Neurosci* **23**, 315–341.

[28] Keysers, C., Xiao, D.-K., Földiák, P., and Perrett, D. I. (2001). The speed of sight. *J. Cogn. Neurosci.* **13**(1), 90–101.

[29] Knoblich, U., Freedman, D., and Riesenhuber, M. (2002). Categorization in IT and PFC: Model and Experiments. AI Memo 2002-007, CBCL Memo 216, MIT AI Lab and CBCL, Cambridge, MA.

[30] Knoblich, U. and Riesenhuber, M. (2002). Stimulus simplification and object representation: A modeling study. AI Memo 2002-004, CBCL Memo 215, MIT AI Lab and CBCL, Cambridge, MA.

[31] Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophys.* **71**, 856–867.

[32] Kritzer, M. F., Cowey, A., and Somogyi, P. (1992). Patterns of inter- and intralaminar GABAergic connections distinguish striate (V1) and extrastriate (V2, V4) visual cortices and their functionally specialized subdivisions in the rhesus monkey. *J. Neurosci.* **12**, 4545–4564.

[33] Lampl, I., Riesenhuber, M., Poggio, T., and Ferster, D. (2001). The MAX operation in cells in the cat visual cortex. In *Soc. Neurosci. Abs.*, volume 27, 619.30.

[34] Lee, T. S., Yang, C. F., Romero, R. D., and Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nat. Neurosci.* **5**, 589–597.

[35] Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563.

[36] Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophys.* **77**, 24–42.

[37] Luck, S. J., Vogel, E. K., and Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature* **383**, 616–618.

[38] Macmillan, N. A. and Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge University Press, Cambridge.

[39] Martinez, L. M. and Alonso, J.-M. (2001). Construction of complex receptive fields in cat primary visual cortex. *Neuron* **32**, 515–525.

[40] Martinez-Trujillo, J. C. and Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron* **35**, 365–370.

[41] McAdams, C. J. and Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**(1), 431–441.

[42] Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000). Intermodal selective attention in monkeys. I: Distribution and timing of effects across visual areas. *Cereb. Cortex* **10**, 343–358.

[43] Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000). Intermodal selective attention in monkeys. II: Physiological mechanisms of modulation. *Cereb. Cortex* **10**, 359–370.

[44] Missal, M., Vogels, R., and Orban, G. (1997). Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb. Cortex* **7**, 758–767.

[45] Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784.

[46] Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* **14**(4), 2178–2189.

[47] op de Beeck, H. and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *J. Comp. Neurol.* **426**, 505–518.

[48] Pachella, R. G. (1975). The effect of set on the tachistocopic recognition of pictures. In *Attention and Performance V,* Rabbitt, P. M. A. and Dornic, S., editors (Academic Press, New York).

[49] Perret, D. and Oram, M. (1993). Neurophysiology of shape processing. *Image Vision Comput.* **11**, 317–333.

[50] Pollen, D. A., Przybyszewski, A. W., Rubin, M. A., and Foote, W. (2002). Spatial receptive field orgaization of macaque V4 neurons. *Cereb. Cortex* **12**, 601–616.

[51] Potter, M. (1976). Short-term conceptual memory for pictures. *J. Exp. Psych.: Hum. Learn. Mem.* **2**, 509–522.

[52] Reynolds, J. H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron* **26**, 703–714.

[53] Reynolds, J., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* **19**, 1736–1753.

[54] Riesenhuber, M. and Poggio, T. (1999). Are cortical models really bound by the "Binding Problem"? *Neuron* **24**, 87–93.

[55] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**(11), 1019–1025.

[56] Riesenhuber, M. and Poggio, T. (2000). The individual is nothing, the class everything: Psychophysics and modeling of recognition in object classes. AI Memo 1682, CBCL Paper 185, MIT AI Lab and CBCL, Cambridge, MA.

[57] Riesenhuber, M. and Poggio, T. (2000). Models of object recognition. *Nat. Neurosci. Supp.* **3**, 1199–1204.

[58] Roelfsema, P. R., Lamme, V. A. F., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature* **395**, 376–381.

[59] Rolls, E. T., Judge, S. J., and Sanghera, M. J. (1977). Activity of neurones in the inferotemporal cortex of the alert monkey. *Brain Res.* **130**, 229–238.

[60] Rolls, E. T., Treves, A., Tovee, M. J., and Panzeri, S. (1997). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J. Comp. Neurosci.* **4**, 309–333.

[61] Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake monkeys. *Exp. Brain Res.* **77**, 23–30.

[62] Serre, T., Riesenhuber, M., Louie, J., and Poggio, T. (2002). On the role of object-specific features for real-world object recognition in biological vision. In *BMCV (Biologically Motivated Computer Vision)* (Tübingen). Submitted.

[63] Sheinberg, D. L. and Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *J. Neurosci.* **21**, 1340–1350.

[64] Shelton, C. (1996). *Three-Dimensional Correspondence*. Master's thesis, MIT, Cambridge, MA.

[65] Tamura, H. and Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb. Cortex* **11**, 384–399.

[66] Thorpe, S., Delorme, A., and van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks* **14**, 715–725.

[67] Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* **381**, 520–522.

[68] Treue, S. (2001). Neural correlates of attention in primate visual cortex. *TINS* **24**(5), 295–300.

[69] Treue, S. and Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575–579.

[70] Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.* **4**, 832–838.

[71] Ungerleider, L. G. and Mishkin, M. (1982). Two cortical visual systems. In *Analysis of visual behavior,* Ingle, D. J., editor, 549–586 (MIT Press, Cambridge, MA).

[72] Usher, M. and Niebur, E. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *J. Cogn. Neurosci.* **8**(4), 311–327.

[73] van der Velde, F. and de Kamps, M. (2001). From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. *J. Cogn. Neurosci.* **13**(4), 479–491.

[74] van Rullen, R. and Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *J. Cogn. Neurosci.* **13**(4), 454–461.

[75] Wallis, G. and Rolls, E. (1997). Invariant face and object recognition in the visual system. *Prog. Neurobiol.* **51**, 167–194.

[76] Walther, D., Riesenhuber, M., Poggio, T., Itti, L., and Koch, C. (in press). Towards an integrated model of saliency-based attention and object recognition in the primate's visual system. In *Proc. 2002 Cognitive Neuroscience Society Meeting, San Francisco, CA (Journal of Cognitive Neuroscience Vol. B14)*.

[77] Yu, A. J., Giese, M. A., and Poggio, T. (2000). Neural circuits for the realization of the maximum operation. In *Soc. Neurosci. Abs.*, volume 26, 449.3.