Abteilung Neuroinformatik
Universität Ulm

Das auditorische System: Modelle und Simulationen

Dissertation zur Erlangung des Doktorgrades
Dr.rer.nat. der Fakultät für Informatik der Universität Ulm



von

Marcus Borst

aus Wilhelmshaven

2004

Amtierender Dekan :	Prof. Dr. Friedrich von Henke
1. Gutachter :	Prof. Dr. Günther Palm
2. Gutachter :	Prof. Dr. Günter Ehret
Tag der Promotion :	22. Juli 2004

Danksagung

Danke! Das faßt im wesentlichen zusammen, was in den folgenden Zeilen geschrieben steht. Mein Dank gilt allen, die mir bei der Erstellung dieser Dissertation zur Seite standen. Wie immer können leider nicht alle Personen einzeln aufgeführt werden, allerdings sollen doch einige wenige direkt beim Namen genannt werden. Da wäre als erstes Herr Palm, in dessen Abteilung ich diese Dissertation schreiben durfte, dessen großes Interesse an dieser Thematik mich immer wieder aufs Neue angespornt hat und der trotz vieler anderer Verpflichtungen immer für Diskussionen von neuen Ideen Zeit fand. Von ähnlich großer Bedeutung sind auch die Arbeiten von Herrn Langner und seiner Gruppe in Darmstadt, denen ich das Modell verdanke, welches die Basis für diese Arbeit ist. Obwohl Darmstadt von Ulm einige Autostunden weit weg ist, konnte eine rege Kommunikation aufgebaut und unterhalten werden, durch die viele Anregungen für diese Arbeit entstanden sind. Besonders möchte ich an dieser Stelle auch gleich Michael Ochse danken, der mir freundlicherweise einige Ableitungen von Neuronen zur Verfügung gestellt hat, mit denen ich die Ergebnisse der Dissertation abrunden konnte. Von meinen Kollegen in Ulm möchte ich ganz besonders Andreas Knoblauch danken, der mit mir einen großen Teil der Zeit im gleichen Zimmer sitzen mußte, mich bei der Softwareerstellung unterstützt hat, und immer diskussionsbereit und interessiert war. Weniger für fachliche Unterstützung als vielmehr für die Abwechslung in den Pausen möchte ich mich bei vielen Freunden und Kollegen wie zum Beispiel Veronica Albrecht, Sandra Blome, Ralph Schönfelder, ... bedanken, die mich in den letzten Jahren begleitet haben.

Weitere fachliche Unterstützung konnte ich von Herrn Günter Ehret und seiner Gruppe, insbesondere Thomas Niessner, erfahren, mit dessen Hilfe ich einen tieferen Einblick in den Aufbau des Gehörs von Mäusen bekommen konnte. Für die Modellierung kortikaler Areale waren die Treffen mit der Gruppe von Herrn Henning Scheich von großer Bedeutung. An dieser Stelle soll auch die VW-Stiftung erwähnt werden, denen ich die notwendige Finanzierung meiner Arbeit verdanke, und durch die eine enge Kooperation mit Herrn Langner und Herrn Scheich möglich wurde.

Einige oft vergessene Danksagungen sollen hier endlich erwähnt werden. Als erstes ein Dankeschön an Herrn Linus Torwald für die geniale Idee, Linux ins Leben zu rufen, und somit eine echte Alternative zu den Produkten des Marktführers zu schaffen. Ohne Linux wären Simulationen, die sich teilweise über mehrere Wochen hingezogen haben, nur schwer möglich gewesen. Auch die Angestellten der Uni, die mich in der Cafeteria mit Essen und Trinken versorgten oder mein Zimmer saubergehalten haben, sollen nicht vergessen werden. Und auch den Autoren der Perry Rhodan Serie soll mein Dank gelten für die allwöchentliche Ablenkung, die mich seit langer Zeit begleitet.

Aber das Wichtigste überhaupt in meinem Leben ist meine Familie, insbesondere mein (noch) kleiner Lukas, der mich in den Tagen des Zusammenschreibens feinfühlig motiviert hat ('Was, erst auf Seite 107? Geht das nicht schneller?') und meine Frau Ilona, die mich in solchen Situationen zuverlässig wieder aufgebaut hat. Besonders hilfreich waren auch die Korrekturvorschläge meiner Mutter, die diese Arbeit massiv verbessert haben.

Inhaltsverzeichnis

1 Einleitung **7**

2 Grundlagen **11**
2.1 Einführung 11
2.2 Die biologischen Vorgaben 11
2.3 Modellierung des peripheren auditorischen Systems 14
2.4 Modellierung kortikaler Areale 21
2.5 Das komplette Modell 23

3 Ergebnisse **25**
3.1 Ergebnisse zum ursprünglichen Langnermodell 25
3.2 Ergebnisse zum variierten Langnermodell 38
3.3 Ergebnisse zur Kammfilterdiskussion 49
3.4 Analyse von Ableitungen an wachen Tieren 51
3.5 Aus der Periodizitätskartierung ableitbare Karten 54
3.6 Ergebnisse unter Einbeziehung kortikaler Areale 59
3.7 Quellenlokalisation anhand der Spikes aus der IC Karte 73
3.8 Psychophysik 75

4 Diskussion **85**
4.1 Überlegungen zu den peripheren Arealen (Langnermodell) 85
4.2 Einbeziehung kortikaler Areale 100
4.3 Lokalisations-Experiment 104
4.4 Psychophysik 105

5 Anwendungen **109**
5.1 Einsatz in der Technik 109
5.2 Einsatz in der Medizin 113

6 Hardware **117**
6.1 Der Versuchsaufbau 117
6.2 FPGA Implementierung des variierten Modells 118

7 Zusammenfassung **123**
7.1 Verwendete Parameter für die Simulationen 125

Kapitel 1

Einleitung

Dein Wort in Gottes Ohr! Dieses Sprichwort hat wohl fast jeder schon einmal zu hören bekommen. Doch was im einzelnen mit diesem *Wort* passiert, wenn es dann in *Gottes Ohr* ist, das ist eine offene Frage. Für menschliche Ohren gibt es allerdings einige Thesen, welchem Zweck sie dienen könnten. In den frühen Zeiten der menschlichen Entwicklung war vermutlich ein wichtiger Aspekt der, daß man öfters *jemanden das Fell über die Ohren ziehen* wollte. Später, als die Behaarung weniger wurde, mußte man sich mit anderen Methoden behelfen und hat lieber *die Ohren lang gezogen* oder in Verhandlungen *den anderen übers Ohr gehauen*, wenn dieser *auf seinen Ohren gesessen* ist. Dagegen wurde schon vielen Kindern ohne schlechte Absicht *ein Floh ins Ohr gesetzt*, wenn sie *noch feucht hinter den Ohren* waren. Dann war es die Aufgabe der Eltern, *immer ein offenes Ohr zu haben*, wobei dies nicht immer funktioniert hat, wenn sie *über beide Ohren verliebt* waren. In diesem Fall mußten die Kinder ihren Eltern erst *die Ohren spitzen*. Und wenn die Eltern dann anderer Meinung waren, so ist früher oft die Regel *wer nicht hören will, muß fühlen* in Kraft getreten - glücklicher waren da die Kinder, denen nur *etwas hinter die Ohren geschrieben* wurde. In beiden Fällen haben sie dann das nächste mal *ihre Ohren aufgesperrt*.

Aus diesen Anwendungen entwickelten sich allerdings grundlegendere Fragestellungen: Wie funktioniert das Ohr? Oder - allgemeiner gesagt - wie funktioniert das Hören? Wie man heute weiß, ist das Ohr nur ein kleiner Teil des auditorischen Systems, das alle Bereiche im Gehirn umfaßt, welche sich mit Hören beschäftigen. Die Aufgaben des Hörsystems sind im allgemeinen etwas anders gelagert als oben beschrieben, aber mindestens ebenso vielfältig: Für die Evolution von Bedeutung waren zunächst zwei Aufgaben: Zum einen die Kommunikation mit seinesgleichen und zum anderen die Wahrnehmung der Umgebungsgeräusche. Dabei spielte für letzteren Fall sicher die Warnung vor (Freß)-Feinden eine große Rolle. Für diese Aufgabe ist das Ohr besonders geeignet, da es im Gegensatz zum visuellen System die Signale aus allen Richtungen aufnimmt. Während die Augen viele exakte Informationen über einen Ausschnitt der Welt liefern, geben die Ohren die komplette Umwelt wieder. Damit diese Warnfunktion in allen Lebenslagen gut funktioniert, gibt es auch keine Möglichkeit das Ohr abzuschalten, oder ähnlich wie das Auge mit 'Ohrlidern' zu bedecken. Zusätzlich ist das Ohr auch bei völliger Dunkelheit in der Lage die Umgebung wahrzunehmen, bei vielen nachtaktiven Tieren geschieht die Navigation komplett mit dem auditorischen System.

Generell sind die konzeptionellen Unterschiede zwischen Auge und Ohr sehr unterschiedlich: Das Auge besitzt eine Vielzahl von Rezeptoren (die Stäbchen und Zäpfchen), die durch die vorgeschaltete Optik völlig unterschiedliche Signale verarbeiten müssen, welche sich auch vergleichsweise langsam ändern - schneller als 10 ms ist hier nicht notwendig. Dagegen analysiert das Ohr nur ein Signal, nämlich die Mixtur aller Umgebungsgeräusche. Diese beinhaltet dafür sehr schnelle Signaländerungen, die Grenze des vom Menschen Wahrnehmbaren bewegt sich zwischen 10 kHz und 20 kHz. Dies entspricht 0.1 bis 0.05 ms, also sind die Signaländerungen, die das auditorische System verarbeiten soll, um mehr als Faktor 100 schneller als es im Falle des visuellen System notwendig ist.

Daher unterscheidet sich der Aufbau des auditorischen Systems stark von dem Aufbau des

visuellen Systems. Das akustische Signal muß vielfach vorverarbeitet werden, bevor es von den relativ langsamen Neuronen in höheren Arealen sinnvoll abgebildet werden kann. Diese Arbeit wird in unterschiedlichen Ebenen durchgeführt: Bevor das Schallsignal neuronal aufgenommen wird, wird es mehrfach gefiltert. Diese Filterung fängt bereits mit der Form des Kopfes und der Kopfbehaarung an, wird von der Ohrmuschel fortgeführt und im Innenohr vollendet. Im Gegensatz zu den frühen Untersuchungen, die der Ohrmuschel nur eine Schutzfunktion zugesprochen haben, ist heute bekannt, daß die Ohrmuschel und der Gehörgang eine komplexe Filterung durchführen. Mit Hilfe dieser Filterung und etwas Erfahrung können sogar Menschen, die auf einem Ohr taub sind, eine grobe Schalllokalisation durchführen. Dies ist technisch schwer machbar, die Position einer Schallquelle auch nur ganz grob mit nur einem einzelnen Mikrofon zu berechnen, ist bisher nicht möglich.

Im Innenohr wird der Schall durch die Schnecke (Cochlea) in unterschiedliche Frequenzbänder zerlegt, die einzelnen Filterkanäle werden dann in dem Nucleus Cochlearis weiterverarbeitet. Bis zu diesem Punkt ist die Funktion der einzelnen Komponenten weitgehendst geklärt, wobei selbst auf diesem relativ einfachen Niveau noch viele Detailfragen ungeklärt sind. Im weiteren Verlauf wird das Schallsignal durch verschiedene Gehirnregionen transportiert, bis es im auditorischen Kortex ankommt. Bis dahin ist es vielfach vorverarbeitet und analysiert worden, so daß es im auditorischen Kortex völlig anders repräsentiert wird als auf den unteren Ebenen. Diese Arbeitsschritte sind teilweise völlig unklar, teilweise existieren verschiedene Theorien, an welcher Stelle welche Verarbeitungsschritte vorgenommen werden. Selbst wenn in einer Ebene die Datenrepräsentation bekannt ist, kann es sein, daß es viele unterschiedliche Theorien gibt, mit welchen neuronalen Strukturen diese Repräsentationen berechnet werden.

Auch wenn die genaue Arbeitsweise unbekannt ist, das Ergebnis ist, daß der Mensch in der Lage ist, aus einer Mischung von Geräuschen die einzelnen Geräusche zu extrahieren, festzustellen, wo sie herkommen, und jede Schallquelle zu klassifizieren. Dies sind die drei Grundaufgaben des auditorischen Systems, die durch das komplexe Zusammenspiel der einzelnen Komponenten gelöst wird. Auch wenn viele Funktionen dieses komplexen Systems unbekannt sind, ist das Ziel dieser Arbeit, ein möglichst umfassendes Modell zur Klassifikation und Separation von Schallsignalen zu entwickeln, welches sich so nahe wie möglich an dem biologischen Vorbild orientiert. Dieses Modell besteht aus zwei Teilen, einem Vorverarbeitungssystem, basierend auf dem Langnermodell [16] und mehreren Assoziativspeichern, die Teile des Kortex modellieren. Während die Assoziativspeicher bereits für das visuelle System eingesetzt und analysiert wurden, existierten für das Langnermodell keine fertigen Implementierungen, auf die aufgebaut werden konnte. Daher stellt die Realisierung und Analyse des Langnermodells einen größeren Teil der Arbeit dar, als die Analyse einzelner Assoziativspeicher. Das Zusammenspiel des peripheren Langnermodells mit den Assoziativspeichern und deren Nutzen ist dagegen wieder ausführlich analysiert.

Die Arbeit gliedert sich in die Bereiche *Grundlagen* (ab Seite 11), in denen sowohl die wichtigsten biologischen Strukturen als auch die abstrahierten Modelle, mit denen die Biologie modelliert wird, dargestellt sind. In dem umfangreichsten Kapitel *Ergebnisse* (ab Seite 25) sind die einzelnen Untersuchungen und Simulationen mit den jeweiligen Ergebnissen dargestellt. Diese umfassen sowohl die mit dem Computer simulierten Modelle als auch einige Auswertungen von Messungen an Neuronen im Colliculus inferior, die von Biologen zur Verfügung gestellt wurden. Zusätzlich werden einige Experimente der Psychophysik vorgestellt und mit den Simulationen verglichen. In dem sich anschließenden Kapitel *Diskussion* (ab Seite 85) werden wichtige Teile der Ergebnisse nochmals herausgegriffen und ihre Bedeutung diskutiert. Das nachfolgende Kapitel *Anwendungen* (ab Seite 109) stellt einige Ideen vor, wie die gefundenen Ergebnisse weiterverwertet werden könnten, während das Kapitel *Hardware* (ab Seite 117) zum einen die speziell für die Versuche entwickelte Hardware vorstellt. Zum anderen wird eine integrierte Hardware, welche das Langnermodell in Echtzeit auf einem Monitor darstellt, beschrieben. Diese Hardware wurde mehr für Demonstrationszwecke als für die wissenschaftlichen Analysen entwickelt. Abschließend sind der Vollständigkeit wegen im Anhang (ab Seite 125) noch einige weniger interessante Detailinformationen zu den verwendeten Neuronparametern aufgeführt.

Da diese Arbeit sich sowohl an Biologen richtet, die sich für die Modellierung des auditorischen Systems interessieren, als auch an Informatiker, die Gefallen an biologischen Methoden der

Informationsverarbeitung finden, wird versucht, beiden Parteien die gewünschten Daten zu bieten. Dennoch ist es leider ebensowenig möglich, im Rahmen dieser Dissertation beide Aspekte komplett bis in das letzte Detail zu analysieren, als auch alle Begriffe einem unbedarften Leser jeweils neu zu erklären. Allerdings sollten selbst Leser, die keine Vorkenntnisse haben, aber an der Thematik interessiert sind, anhand der Grundlagen und jeweiligen Erklärungen die interessantesten Ergebnisse verstehen können. In diesem Sinne wünsche ich allen Personen, die diese Arbeit mehr oder weniger freiwillig lesen, eine interessante und verständliche Lektüre.

Kapitel 2

Grundlagen

2.1 Einführung

Dieses Kapitel ist in mehrere Teile untergliedert: Zuerst werden die biologischen Grundlagen, welche für das weitere Verständnis wichtig sind, erläutert. Das auditorische System wird dabei in zwei Bereiche eingeteilt: Die peripheren- und Mittelhirnareale und die kortikalen Areale.

Der nächste Abschnitt beschreibt die Modellierung dieser Areale und welche technischen Geräte dazu notwendig sind.

Danach wird dargestellt, wie die kortikalen Areale modelliert werden.

In einem abschließenden Abschnitt werden einige Methoden der Psychophysik erläutert, die für spätere vergleichende Experimente wichtig sind.

2.2 Die biologischen Vorgaben

Im folgenden werden die Teile des auditorischen Systems beschrieben, die für Mustererkennung und Separationsaufgaben wichtig sind. Die Kerne, die für Schalllokalisation zuständig sind, werden nicht dargestellt.

2.2.1 Peripheres auditorisches System

Zur Vereinfachung werden alle Bereiche, die vor dem Kortex in der Hörbahn sind, mit *peripherem auditorischem System* zusammengefaßt (siehe Abb. 2.1). Dazu gehören sowohl die Bereiche, die für die Schallaufnahme und Wandlung in elektrische Signale verantwortlich sind als auch der Nucleus Cochlearis und der Colliculus Inferior, der schon zum Mittelhirn gehört.

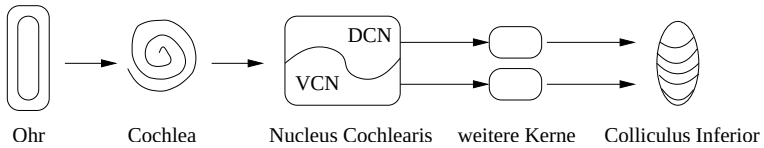

Abbildung 2.1: Schematische Darstellung des peripheren auditorischen Systems.

Das Ohr

Das Ohr dient der Filterung und Umsetzung der Luftdruckschwankungen in elektrische Impulse. Die erste Filterung erfolgt bereits durch die Form der Ohrmuschel und des Gehörganges. Durch das Trommelfell und der kleinsten Knochenkonstruktion des Menschen (Hammer, Amboß und

Steigbügel) werden die Luftdruckschwankungen in wellenförmige Bewegungen einer Flüssigkeit umgesetzt. Ohne die Umsetzung durch das Trommelfell und die kleinen Knochen würden die Schallwellen nur stark abgeschwächt in Flüssigkeitswellen umgesetzt werden, da die Flüssigkeit eine wesentlich höhere Dichte als die Luft hat. Die Auswertung der Flüssigkeitswellen erfolgt anschließend innerhalb der Cochlea (Schnecke).

Die Cochlea

Die Cochlea kann vereinfacht als ein schneckenförmig aufgewickelter Schlauch mit einer Trennmembran dargestellt werden. Diese Trennmembran teilt den Schlauch entlang der Längsachse in eine untere und eine obere Hälfte und heißt Basilarmembran. Die Flüssigkeitswellen wandern auf der oberen Hälfte entlang und werden als Wanderwelle bezeichnet. Durch die Konstruktion der Cochlea ergibt sich ein Maximum der Wanderwelle an einer von der Frequenz abhängigen Stelle. Tiefe Frequenzen erzeugen ein Maximum am hinteren Ende der Cochlea, hohe Frequenzen dagegen am Anfang der Cochlea.

Die Amplitude der Wanderwelle wird von Haarzellen gemessen, die sich entlang der Basilarmembran befinden, und wird in eine elektrische Kodierung umgesetzt: Je größer die Amplitude ist, desto stärker ist die Auslenkung der Haarzellen und desto mehr elektrische Impulse werden erzeugt. Allerdings funktioniert die Kodierung aus verschiedenen Gründen in eine Richtung besser als in die andere, sodaß die Schallwellen nicht komplett als elektrische Impulse repräsentiert sind, sondern hauptsächlich die positiven Halbwellen.

Diese elektrischen Impulse (Spikes) werden vom Hörnerv in den Nucleus Cochlearis weitergeleitet.

Der Nucleus Cochlearis

Im Nucleus Cochlearis (CN) werden die Spikes des Hörnervs erstmalig von komplexen Neuronen verarbeitet [11]. Von den vielen Neurontypen, die sich in diesem Kern finden, werden nur einige - für die Arbeit relevante - genauer erklärt. Generell ist der CN in verschiedene Bereiche unterteilt:

- Ventraler Nucleus Cochlearis:

  Im ventralen Nucleus Cochlearis befinden sich Onset Zellen, die zeitlich sehr exakte Spikes am Anfang eines Reizes liefern, aber auch Choppernеurone, die mit einer Eigenschwingung auf einen Reiz reagieren [37]. Während eine Onset Zelle nur einen oder wenige Spikes pro Reiz erzeugt, feuern Chopper mehrere Spikes, machen eine Pause, feuern wieder, machen Pause, usw. . Dabei werden die Pausen/Feuerzeiten sehr exakt eingehalten. Es existieren noch andere Neurontypen, die aber hier nicht näher behandelt werden.

- Dorsaler Nucleus Cochlearis:

  Im dorsalen Nucleus Cochlearis (DCN) finden sich Pauserneurone und Buildupneurone, die sich dadurch auszeichnen, daß der größte Teil der Aktivität erst mit einer Verzögerung einsetzt [38]. Im Gegensatz zu den Buildupneuronen, die zuerst gar keine Aktivität zeigen und dann nach einiger Zeit anfangen zu spiken, feuern die Pauserneurone erst einige wenige Spikes. Danach ist keine Aktivität mehr vorhanden, und weitere Spikes werden erst nach einer Pause erzeugt.

Die Ausgänge des CN projezieren zu verschiedenen Bereichen und sind meistens tonotopisch organisiert. Das heißt, daß die Frequenz, auf die ein Teil der Ausgänge antwortet, sich systematisch mit dem Ort der Nervenbahnen verändert. Diese Organisation beginnt bereits an der Cochlea, da die beste Antwortfrequenz der Haarzellen sich entlang der Basilarmembran systematisch verändert. Auch viele der Neuronen im CN sind bereits tonotopisch organisiert, so daß daraus die tonotopische Organisation der Ausgänge resultiert.

Ein Bereich, der mit den Ausgängen des CN verbunden ist, ist der Komplex der oberen Olive, der vermutlich Informationen für die Lokalisation von Schallquellen bereitstellt. Weiterhin ist der

Lemniscus Lateralis mit dem CN verschaltet, dessen Ausgänge stark inhibitorisch in den Colliculus Inferior projizieren. Die Funktion dieser Inhibition ist bisher ungeklärt.

Die Ausgänge des Nucleus Cochlearis gehen auch direkt zum Colliculus Inferior, der im folgenden beschrieben wird.

Der Colliculus Inferior

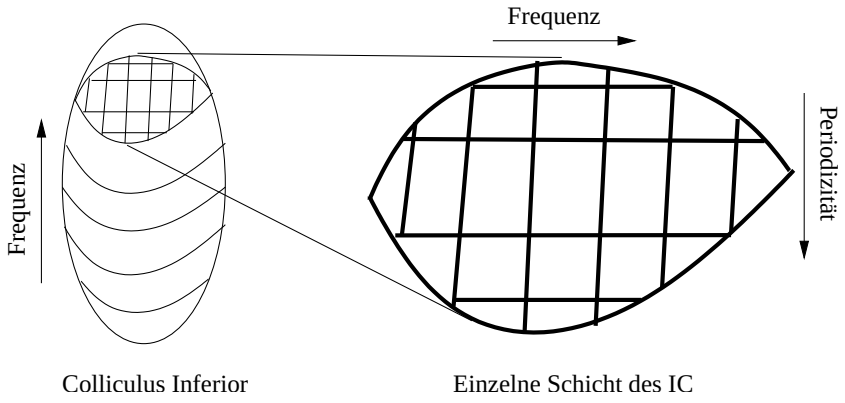

Abbildung 2.2: Schematische Darstellung des Colliculus Inferior. Jede Schicht besitzt einen Periodizitätsgradienten und einen Frequenzgradienten. Die Schichten selber könnten einen gröberen Frequenzgradienten darstellen.

Der Colliculus Inferior (IC) stellt eine zentrale Schaltstelle des Mittelhirns dar [14, 13]. Viele Nervenbahnen aus verschiedenen Kernen werden hier auf eine weitgehendst unbekannte Art verknüpft, und es entstehen vielfältige und komplexe Aktivitätsmuster. Der zentrale Bereich des IC (ICC) besteht aus mehreren Lagen von Neuronen [42]. In diesem Bereich wurden verschiedene Karten gefunden [13]. Zum Beispiel existiert eine Latenzkarte (Zeitdauer von Reizbeginn bis zur neuronalen Antwort). Auch eine möglicherweise konzentrische Amplitudenkarte wurde gefunden, bei der kleine Amplituden in der Mitte der Karte Aktivität erzeugen, große Amplituden dagegen eher außen [19]. Besonders interessant ist die Periodizitäts/Frequenzkarte [31, 28], die am besten anhand eines amplituden modulierten (AM) Reizsignals erklärt werden kann (siehe Abb. 2.2). Dieses wird durch die Trägerfrequenz und die Modulationsfrequenz definiert: ein AM Signal mit 100 Hz Modulationsfrequenz und 5000 Hz Trägerfrequenz ist ein Sinus mit 5000 Hz, dessen Lautstärke 100 mal pro Sekunde sinusförmig von 0 auf 100% ansteigt und wieder zu 0% abfällt. Die Periodizitäts/Frequenz Karte im ICC ist eine orthogonale Karte, wobei entlang der 1. Achse die Trägerfrequenz (Frequenz) und entlang der 2. Achse die Modulationsfrequenz (Periodizität) aufgetragen ist. Jeder Punkt der Karte entspricht demnach einem bestimmten AM Signal. Die verschiedenen Schichten könnten eine grobe Trägerfrequenzunterteilung darstellen, so daß innerhalb jeder Schicht eine feinere Frequenzaufteilung möglich wäre. Anfangs wurde angenommen, daß diese Karte nicht orthogonal sondern circulär aufgebaut ist. Durch weitere Messungen wurde dann jedoch die aktuelle Meinung geprägt, daß es sich um eine orthogonale Karte handelt.

2.2.2 Höhere Verarbeitungseinheiten

Das periphere Hörsystem dient der Extraktion von relevanten Merkmalen, die in den weiteren Ebenen weiterverarbeitet werden. Diese Verarbeitung findet im auditorischen Kortex [12] statt, der aus mehreren Einheiten besteht (siehe Abb. 2.3).

Der primäre auditorische Kortex

Als erstes kortikales Areal ist der primäre auditorische Kortex mit den peripheren Arealen verbunden. Diese Verbindung besteht nicht direkt, sondern über den Corpus geniculatum mediale. Der primäre auditorische Kortex besteht wie der IC aus verschiedenen Karten [9, 10], besonders wichtig ist wieder die Periodizitäts/Frequenzkarte. Im Gegensatz zum IC ist bisher noch nicht klar, ob diese Karte orthogonal [30] oder circulär [43] aufgebaut ist. Allerdings kann die Frage, ob es eine solche Karte prinzipiell gibt, mit *ja* beantwortet werden.

Weitere corticale Areale

Der primäre auditorische Kortex stellt nur einen kleinen Bereich des auditorischen Kortex dar. Es gibt noch viele weitere Areale [34], von denen zum großen Teil unklar ist, welche Funktion sie haben, und wie sie diese erfüllen.

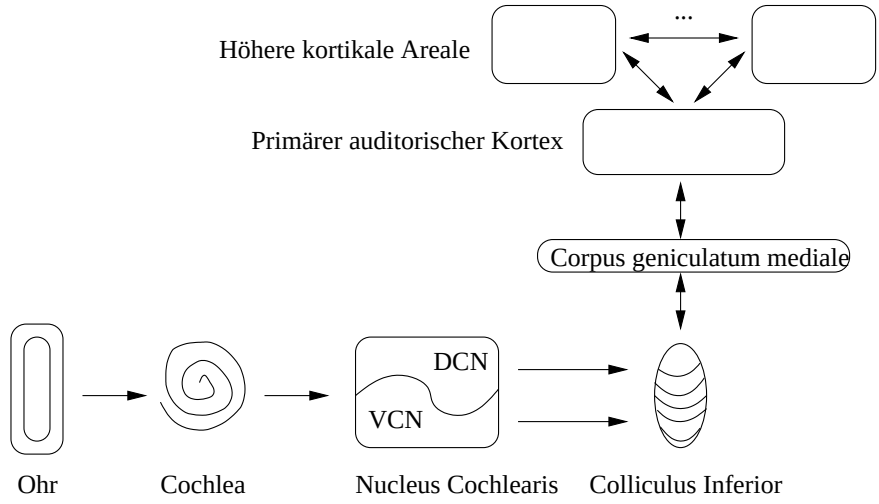

Abbildung 2.3: Schematische Darstellung des peripheren Hörsystems mit einigen kortikalen Arealen. Mit dem primären auditorischen Kortex sind viele weitere Areale verbunden.

2.2.3 Rückprojektionen aus höheren Arealen

Die kortikalen Areale sind stark miteinander verbunden. Es existieren auch viele Rückkopplungen aus höheren Areale in den primären auditorischen Kortex. Zusätzlich projiziert der primäre auditorische Kortex auch zurück in den IC, wobei die Verbindung zwar exitatorisch ist, aber dann durch Interneurone im IC zu inhibitorischen Effekten führen kann. Die genauen Verhältnisse sind bisher noch nicht bekannt.

2.3 Modellierung des peripheren auditorischen Systems

Die Modellierung des peripheren auditorischen System hat in dieser Arbeit als Hauptziel die Erzeugung einer Periodizitäts/Frequenzkarte ähnlich zu der Karte im IC. Dabei sollen die biologischen Rahmenbedingungen - soweit bekannt - eingehalten werden. Bevor allerdings die neuronalen Strukturen modelliert werden, sollen zuerst einige technische Geräte zur Schallaufnahme vorgestellt werden.

Danach werden die neuronalen Strukturen, wie sie für die Realisierung der gewünschten Karte notwendig sind, dargestellt.

2.3.1 Signalaufnahme

Um reale Schallsignale per Computer zu verarbeiten ist eine Vorrichtung notwendig, die den Schall in digitale Daten umwandelt. Im weitesten Sinne kann dies mit der Mechanik des Innenohrs und der Cochlea verglichen werden. Der gesamte Aufbau ist in Abbildung 2.4 dargestellt.

Mikrofon

Die Wandlung von Schallwellen in elektrische Signale geschieht mit einem kostengünstigen elektret Kapselmikrofon. Dieses liefert ein sehr schwaches Signal, welches mit einem Stereovorverstärker weiterverarbeitet wird. Für Stereoaufnahmen war dies notwendig, da die verwendete Soundkarte des PCs nur einen Mono-Mikrofoneingang besitzt. Das Stereo Mikrofonsignal konnte dann an den Line-In-Eingang der Soundkarte angeschlossen werden.

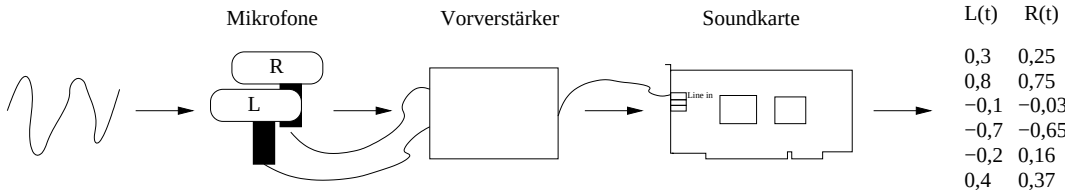

Abbildung 2.4: Aufbau zur Aufnahme und Digitalisierung der Schallsignale. Aus einem akustischen Reiz werden zwei Zahlenfolgen $R(t)$ und $L(t)$.

Soundkarte

Die Soundkarte ist zur Schallaufnahme notwendig, um die analogen Spannungswerte in Folgen von digitalen Null/Eins Werten zu konvertieren, da der Computer nur diese verarbeiten kann. Die verwendete Soundkarte ist eine preiswerte Soundblaster PCI Karte. Die Daten werden mit 16 Bit bei einer Rate von 44,1 kHz in Stereo digitalisiert. Für die meisten Simulationen reicht ein Mono-Signal aus, daher wird oft nur ein Kanal verwendet. Von den zwei Datenströmen $R(t)$ und $L(t)$, die die akustischen Signale repräsentieren, wird einer ausgewählt und durch eine Variable $I(t) = R(t)$ dargestellt. Die Funktionen R, L und somit auch I haben einen Wertebereich von -1 bis 1, eine Auflösung im Wertebereich von $\frac{1}{32768}$ (16 Bit), und können sich 44 100 mal pro Sekunde ändern.

2.3.2 Vorverarbeitung

Bevor die digitalisierten Daten $I(t)$ von neuronalen Netzen analysiert werden, finden zwei Vorverarbeitungsschritte statt (vgl. Abb. 2.6). In den biologischen Systemen entsprechen diese der Filterung auf der Cochlea und der Weiterleitung der Signale durch den Hörnerv.

Die Gammaton Filterbank

Die akustischen Signale werden in der Cochlea gefiltert. Diese Filterung wird in den Simulationen mit Hilfe der Gammaton Filterbank [36, 44] durchgeführt (vgl. Abb. 2.5). Dadurch wird das akustische Signal $I(t)$ auf 50 Filterausgänge $A_i(t)$ (i=1..50) aufgeteilt, deren Frequenzbereich sich von ca. 400 Hz (i=1) bis ca. 4000 Hz (i=50) erstreckt. Dieser Frequenzbereich ist für Sprache besonders interessant, da die Formantfrequenzen - also die wichtigen Frequenzbereiche, mit denen man Vokale unterscheiden kann - in diesem Bereich liegen.

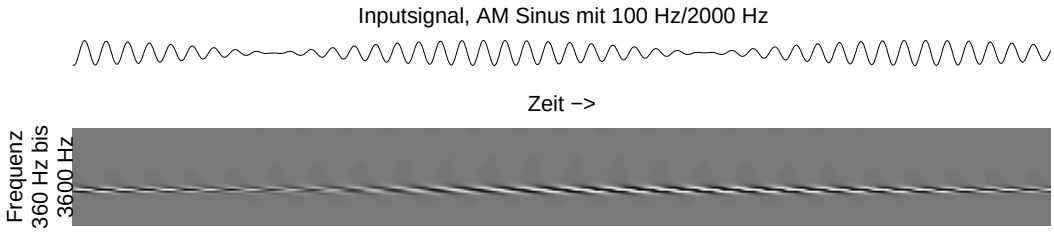

Abbildung 2.5: Die Arbeitsweise der Gammaton Filterbank: Ein Signal (hier ein amplituden modulierter Sinus) wird als Eingabesignal verwendet. Das Ergebnis sind 50 Filterkanäle, die hier übereinander dargestellt sind. Die unterste Zeile des Bildes zeigt die Filterbankantwort für 360 Hz, die oberste die Antwort für 3600 Hz. Die Aktivität, die sich durch den AM Sinus ergibt, sind die Zeilen, wo die Bestfrequenzen der Filterkanäle etwa 2000 Hz sind.

Die Lautstärkentransferfunktion

Nachdem der akustische Reiz mit Hilfe der Gammaton Filterbank in 50 Frequenzkanäle unterteilt wurde, werden diese mit einer logarithmischen Transferfunktion weiter verarbeitet. Dadurch werden mehrere Eigenschaften der Haarzellen innerhalb der Cochlea berücksichtigt. Zuerst wird das Ergebnis der Filterung $A_i(t)$ mit einer um 2 entlang der Y-Achse verschobenen sigmoiden Transferfunktion an die Auslenkung der Haarzellen angepaßt:

$$H_i(t) = \frac{1.6}{e^{-((a_i(t)-0.14)*8}} - 2$$

Das Ergebnis $H_i(t)$ wird danach logarithmiert und gleichgerichtet, ähnlich der Transferfunktion des Hörnervs:

$$S_i(t) = |\log(1 + H_i(t))|$$

Die Vorverarbeitung und die Transferfunktion ist nur eine Näherung der in der Realität wesentlich komplizierteren Effekte. Allerdings ist dies für den Schwerpunkt der Arbeit ausreichend, da die für die weitere Untersuchung mit neuronalen Netzwerken wesentlichen Auswirkungen modelliert sind.

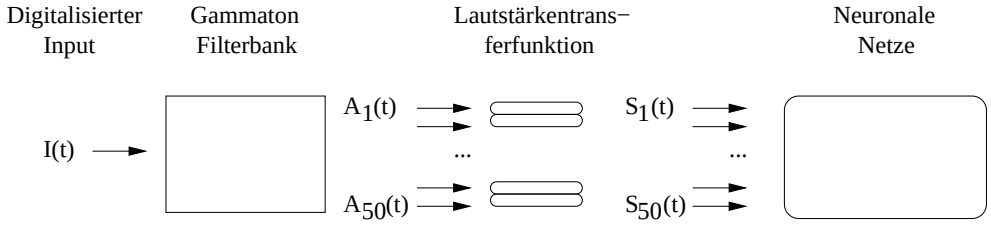

Abbildung 2.6: Die weiteren Vorverarbeitungsschritte: Die digitalisierten Daten werden mit der Gammaton Filterbank in 50 Frequenzkanäle aufgeteilt, die dann mit Hilfe einer logarithmischen Transferfunktion für die weitere Verarbeitung durch neuronalen Netze vorbereitet werden.

2.3.3 Neuronmodelle

Für die Simulation neuronaler Netze existieren verschiedene Modelle, die sich in ihrer Komplexität und dem Berechnungsaufwand stark unterscheiden. Die einfachen Modelle sind lineare, graduelle Neuronen, wogegen Multikompartment-Modelle eine sehr hohe Komplexität aufweisen, da bei diesen das Neuron in verschiedene, unabhängige Bereiche eingeteilt wird. Für diese Arbeit wird ein

Leaky-Integrate-and-Fire Modell verwendet, welches zum einem eine hohe Berechnungsgeschwindigkeit hat. Zum anderen gehört es zu der Klasse der spikenden Neuronen und entspricht den biologischen Modellen somit besser als graduelle Neuronen, deren Ausgabe als Durchschnittsmenge der Spikes interpretiert werden kann.

Leaky Integrate and Fire Modell

Das verwendete Neuronenmodell basiert auf McGregor [33] und kann mit folgenden Differentialgleichungen beschrieben werden:

Das Potential des Neurons	x	$\tau_x * x' = -x + g_{ex} - g_{in}$
Die erregende Leitfähigkeit	g_{ex}	$\tau_{ex} g'_{ex} = -g_{ex} + \delta * input_{ex}$
Die hemmende Leitfähigkeit	g_{in}	$\tau_{in} g'_{in} = -g_{in} + \delta * input_{in}$
Die Ausgabe	y	$y = 1_{[x \geq \theta]}$

Die Variable g_{ex} und g_{in} werden jeweils auf einen Bereich $[0, g_{ex}max]$ bzw. $[0, g_{in}max]$ eingeschränkt. Nach einem Spike wird das Neuronpotential x um x_{ahp} verringert, dies entspricht dem 'after hyper polarization' Potentialsprung eines realen Neurons. Für jedes Neuron existieren also die folgenden Parameter:

Die Zeitkonstanten der Variablen	τ_x, τ_{ex}, τ_{in}
Die maximalen Leitfähigkeiten	$g_{ex}max, g_{in}max$
Der Spike Schwellwert für das Potential	Θ
Die Abnahme des Potentials nach einem Spike	x_{ahp}

Für die Simulation der Neuronen müssen die Differentialgleichungen diskretisiert werden. Dazu wird eine Schrittweite von 25 μs verwendet, dies entspricht 40 kHz.

2.3.4 Das Langnermodell

Ein Ziel dieser Arbeit ist die Modellierung und Simulation der Karte, die von Langner gefunden wurde [31]. Von Langner wurde ein biologienahes Modell zu diesen Ergebnissen entwickelt [16], welches die Basis für diese Arbeit darstellt. Im folgenden werden die einzelnen Komponenten des Modells erklärt, die exakten Parameter sind im Anhang auf Seite 125 aufgelistet.

Einleitung

Das im folgenden 'Langnermodell' genannte Periodizitäts-Detektions-Modell besteht - grob gesagt - aus zwei Pfaden (vgl. Abb 2.7). Entlang des einen Pfades wird der akustische Reiz nahezu unverzögert weitergeleitet, wohingegen entlang des zweiten Pfades der akustische Reiz um eine feste Zeitdauer verzögert weitergeleitet wird. Da periodische Signale sich mit einer festen Zeitdauer wiederholen, wird am Ende beider Pfade dasselbe Signalmuster zu finden sein, wenn die Verzögerungsdauer der Pfade der Periodizität des akustischen Reizes entspricht. Wie dies realisiert wird, ist im folgenden vereinfacht dargestellt, um ein erstes Verständnis zu vermitteln. Genauere Details werden anschließend herausgegriffen und kurz diskutiert, da durch diese die Modifikationen motiviert werden.

Das Triggerneuron

Das Triggerneuron erhält Input vom Hörnerv, und erzeugt phasengekoppelte Spikes. Bei einem AM Signal würde das Triggerneuron mit hoher Präzision einmal pro Modulationsperiode spiken.

Dieses komplexe Verhalten kann mit dem verwendeten einfachen 'Leaky Integrate and Fire'-Neuron nicht erreicht werden. Daher wurde das Triggerneuron für die Simulationen durch einen kleinen Schaltkreis aus zwei 'Leaky Integrate and Fire'-Neuronen ersetzt, der als Eingabe den

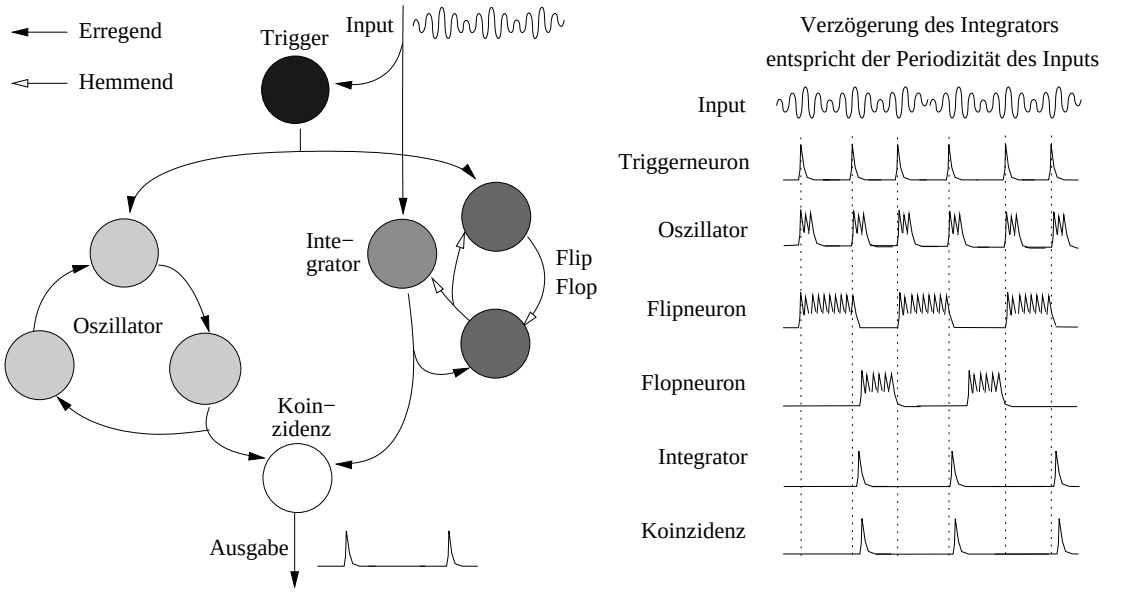

Abbildung 2.7: Der Basisschaltkreis des Langnermodells. Über den Oszillator laufen die akustischen Reize nahezu unverzögert, wohingegen der Integrator das Signal um eine konstante Zeitspanne verzögert. Das Koinzidenzneuron zeigt Aktivität, wenn die Periodizität des akustischen Reizes der Verzögerung des Integrators entspricht. Die jeweilige Neuronaktivität ist idealisiert auf der rechten Seite dargestellt.

graduellen akustischen Reiz S erhält und als Ausgabe die Spikes eines Triggerneurons erzeugt. Wenn ein komplexeres Neuronmodell verwendet werden würde, könnte dieses Verhalten auch mit einem einzelnen Neuron erreicht werden.

Wichtig ist, daß das Triggerneuron die Spikes passend zu der Einhüllenden des akustischen Reizes erzeugt und nicht zu der Trägerfrequenz. Das Triggerneuron hat also die Aufgabe, die Periodizität aus dem Stimulus zu extrahieren. Die weiteren Neuronen werden dann dazu verwendet die Periodizität zu klassifizieren - also festzustellen, welchen Wert die Periodizität hat.

Der Integrator und der Flip/Flop Schaltkreis

Der Integrator und die beiden Flip/Flopneuronen erzeugen eine Verzögerung der Spikes des Triggerneurons. Das wird dadurch realisiert, daß der Integrator eine feste Zeitdauer benötigt, um einen Spike zu erzeugen. Damit diese Verzögerung synchron mit den Spikes des Triggerneurons gestartet wird, ist der Flip/Flop-Schaltkreis nötig. Dieser hemmt das Integratorneuron solange, bis das Triggerneuron feuert. Durch den Spike des Triggerneurons wird das Flipneuron aktiviert, welches seinerseits das Flopneuron inhibiert. Dadurch wird die Hemmung des Flopneurons zum Integrator beendet, und dieser kann mit seiner Integration beginnen. Zur Vereinfachung wird das Integratorneuron mit konstantem Input beschaltet, so daß die Integrationsdauer immer gleich ist. Eigentlich soll der Integrator im Langnermodell das Inputsignal aufintegrieren, dieser Aspekt wird später diskutiert. Bis zu dem Moment, wo der Integrator feuert, bleibt das Flipneuron aktiv, und das Flopneuron wird dadurch gehemmt. Erst durch den Spike des Integrators wird das Flopneuron wieder aktiviert und hemmt jetzt das Flipneuron und den Integrator.

Dieser Vorgang wiederholt sich immer dann, wenn das Triggerneuron einen Spike erzeugt und das Flopneuron aktiv ist. Wird dagegen vom Triggerneuron ein Spike erzeugt, wenn der Integrator noch integriert und das Flipneuron aktiv ist, so wird dieser Spike einfach ignoriert - der Integrator wird dadurch nicht zurückgesetzt.

Der Oszillator

Der Oszillator wurde von Langner eingeführt, weil es im Nucleus Cochlearis Neuronen gibt, die eine ähnliche Antwort aufweisen, und auch die Onsetantworten im IC Indizien für eine ähnliche Verschaltung liefern. Für ein erstes Verständnis könnte er auch weggelassen werden, da er nur die Spikes des Triggerneurons vervielfacht und weiterreicht. Später wird gezeigt werden, wozu er tatsächlich hilfreich sein kann (siehe Abschnitt 4.1.3).

Das Koinzidenzneuron

Das Konzidenzneuron bekommt Input vom Oszillator und vom Integrator. Wenn das Triggerneuron gefeuert hat, erzeugt der Oszillator nach kurzer Zeit ein paar Spikes, wohingegen der Integrator eine längere Zeit benötigt, um den Spike des Triggerneurons weiterzuleiten. Falls der akustische Reiz eine Wiederholungsfrequenz besitzt, die dieser Zeitdifferenz entspricht, so kommen die Spikes vom Oszillator gleichzeitig mit den Spikes vom Integrator am Koinzidenzneuron an. Der Unterschied ist nur, daß die Spikes vom Oszillator durch einen Spike vom Triggerneuron zum Zeitpunkt t ausgelöst sind, während der Spike vom Integrator durch einen Triggerspike zum Zeitpunkt $t - \tau$ erzeugt wird (τ ist die Verzörgerung der Integrators).

Das Koinzidenzneuron erzeugt einen Spike, genau dann, wenn die Spikes vom Oszillator und vom Integrator gleichzeitig ankommen. Dies ist aber nur dann der Fall, wenn die Verzögerung des Integrators der Periodizität des akustischen Reizes entspricht.

Nachteile des Langnermodells

Das Langnermodell besitzt einige Nachteile (Kammfilterantworten, Robustheit, exaktes Antwortverhalten), die im Ergebnisteil illustriert und in der Diskussion näher erläutert werden. Daher wird im folgenden ein leicht modifizierter Schaltkreis zur Periodizitätsdetektion vorgestellt.

2.3.5 Das variierte Langnermodell

Um die vorhandenen Nachteile des Langnermodells auszubessern, wurde der Detektionschaltkreis leicht modifiziert. Für diese Modifikationen wurden von mir mehrere Varianten des Langnermodells implementiert und getestet. Motiviert wurden diese Variationen durch die Nachteile des Langnermodells, wobei versucht wurde, möglichst nur die in den biologischen Vorbildern vorhandenen neuronalen Strukturen zu verwenden. Die besten Eigenschaften hatte die im folgenden mit *variierte* Langnermodell bezeichnete Variation: der Integrator und Flip/Flop Teil wird durch drei anders verschaltete Neuronen ersetzt. Die anderen Teile (Triggerneuron, Oszillator und Koinzidenzneuron) bleiben identisch (vgl. Abb 2.8), die exakten Parameter sind im Anhang auf Seite 125 aufgeführt.

Das IntegStopneuron

Die Funktion des Flip/Flop Schaltkreises wird durch das IntegStopneuron ersetzt. Dieses sorgt dafür, daß bei jedem Triggerneuronspike der Integrator zurückgesetzt wird, und somit durch die Triggerspikes synchronisiert wird. Im Gegensatz zu dem ursprünglichen Langnermodell wirken die Trigger Spikes also auch, wenn der Integrator noch nicht gefeuert hat.

Der Integrator

Der Integrator hat die gleiche Funktion wie im Originalmodell. Durch die andere Verschaltung mit den umliegenden Neuronen ergibt sich allerdings, daß er nicht nur einen Spike erzeugt (denn durch diesen Spike wurde die Hemmung wieder aktiviert) sondern danach spontan weiter Spikes erzeugt.

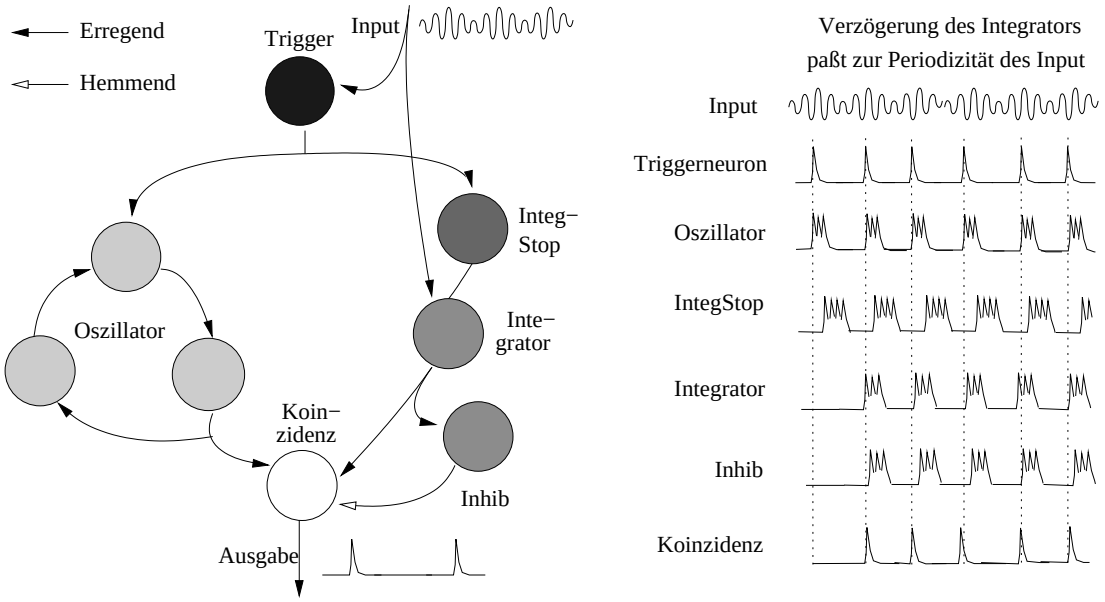

Abbildung 2.8: Das variierte Langnermodell. Der rechte Pfad (Integrator, Flip/Flop) wurde durch drei andere Neuronen ersetzt, die das Signal wie im ursprünglichen Modell eine konstante Zeit verzögern. Die Neuronantworten sind rechts idealisiert dargestellt.

Das Inhibneuron

Das Inhibneuron bekommt die Spikes des Integrators und feuert daher im Prinzip wie der Integrator selber, nur etwas verzögert. Dadurch bekommt das Koinzidenzneuron zuerst Spikes vom Integrator und erst etwas später Spikes vom Inhibneuron. Dies verhindert, daß das Koinzidenzneuron durch die Spikes des Integrators allein schon überschwellig wird und feuert. Nur wenn der erste Spike des Integrators mit den Spikes des Oszillators zusammenfällt, erzeugt das Koinzidenzneuron Spikes. Danach wird es von dem Inhibneuron solange gehemmt, bis diese Hemmung wieder aufhört, weil das Triggerneuron einen Spike erzeugt und somit der Integrator zurückgesetzt wird.

2.3.6 Zuordnung der Modellneurone zu biologischen Neuronen

An dieser Stelle wird nur kurz zusammengefaßt, wo die verwendeten Neuronen im biologischen Vorbild vorkommen. Ein genauer Vergleich der Modellneuronen mit den biologischen Neuronen - inklusive einer Abschätzung der notwendigen Anzahl von Neuronen - wird in der Diskussion ab Abschnitt 4.1.6 auf Seite 95 durchgeführt.

Das Triggerneuron und die Oszillatorneurone besitzen biologische Äquivalente im Nucleus Cochlearis.

Im IC befinden sich Neuronen, die sich ähnlich zu dem Koinzidenzneuron verhalten.

Die Integratorneurone können mit Neuronen im DCN verglichen werden, wobei diese nicht einfach nach wenigen Spikes aufhören zu feuern.

Die Flip/Flopneurone des ursprünglichen Langnermodells müßten ebenfalls im Nucleus Cochlearis zu finden sein, wobei bisher keine entsprechenden Neuronantworten gemessen wurden. Jedoch zeigen anatomische Untersuchungen, daß eine ähnliche Verschaltung wie des Flip/Flop Schaltkreises vorhanden sein könnte.

Das IntegStopneuron paßt dagegen gut zu Neuronen im Nucleus Cochlearis.

Das Inhibneuron könnte entweder im lateralen Lemniscus liegen, von dem bekannt ist, daß er hemmend in den IC projektiert. Alternativ könnte es sich um Interneurone handeln, die direkt im IC liegen.

2.3.7 Erzeugung der tonotopen/periodotopen Karte im IC

Die beiden bisher beschriebenen Schaltkreise ermöglichen die Detektion einer bestimmten Periodizität. Dagegen sind im IC sowohl die Trägerfrequenz als auch die Periodizität kartiert. Dies wird mit dem vorliegenden Schaltkreis erreicht, indem viele einzelne Schaltkreise mit unterschiedlichen Verzögerungen des Integrators repliziert werden. Für jede Kombination der Trägerfrequenz und Periodizität wird ein eigener Schaltkreis zur Verfügung gestellt. Dabei wird die Trägerfrequenz durch die Gammaton Filterbank detektiert, die Periodizität dagegen durch den Langner Schaltkreis (vgl. Abb. 2.9). Für die Simulationen in dieser Arbeit besteht die Karte aus 50x50 modifizierten Langnerschaltkreisen.

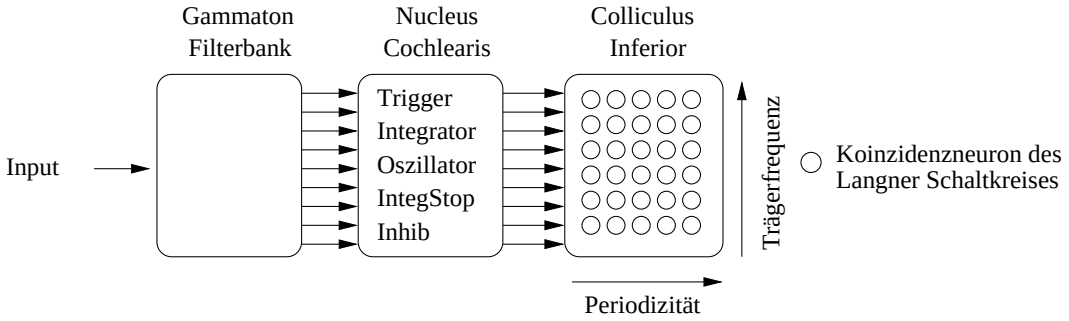

Abbildung 2.9: Das komplette Langnermodell besteht aus einer Vielzahl von einzelnen Langner Schaltkreisen. Die meisten Neuronen des Langnermodells sollen im Nucleus Cochlearis liegen, nur das Koinzidenzneuron befindet sich im IC. Die Karte ergibt sich zum einen durch die unterschiedlichen Filterkanäle und zum anderen durch die unterschiedlich eingestellten Integratoren der Schaltkreise.

2.4 Modellierung kortikaler Areale

Die peripheren Areale dienen der Vorverarbeitung der Daten, wohingegen die kortikalen Areale komplexe Aufgaben wie Klassifikation und Separation lösen sollen. Um diese Areale zu modellieren werden spikende Assoziativspeicher eingesetzt. Diese werden im folgenden kurz beschrieben, eine genaue Beschreibung mit mathematischen Analysen zu Assoziativspeichern im allgemeinen findet sich in [35] und speziell zu spikenden Assoziativspeichern in [25].

2.4.1 Assoziativspeicher

Assoziativspeicher im Vergleich zu Standardspeichern

Im Gegensatz zu normalen Speichern wie z.B. im PC, bei denen Daten an bestimmten Adresspositionen gespeichert werden und nur wieder ausgelesen werden können, wenn die Adressen bekannt sind, werden bei Assoziativspeichern Daten so gespeichert, daß sie nicht mit einer Adresse, sondern mit anderen Daten verknüpft sind. Ein vereinfachtes Beispiel kann dies deutlich machen: Zum speichern eines Liedtextes - z.B. *Help* von den Beatles - würden in einem normalen Speicher die Wörter hintereinander z.B. ab Adresse 0 gespeichert werden. Um den Liedtext wieder zu bekommen, wird der Speicher einfach der Reihe nach ab Adresse 0 wieder ausgelesen. Im Assoziativspeicher dagegen könnte der Liedname und der Text der Reihe nach miteinander verknüpft gespeichert werden: Der Liedtext *Help, I need somebody, ...* würde folgendermaßen gespeichert: *Help* (der Liedname) mit *Help I need* (der Liedanfang), *I need* (der Liedanfang) mit *somebody* (den weiteren Text), usw. Wenn dieser Speicher ausgelesen wird, so würde man ihn mit dem Liednamen *Help* adressieren und bekäme als Antwort *Help, I need.* Damit kann man den Assoziativspeicher wieder adressieren und erhält so der Reihe nach den Liedtext wieder.

Der Vorteil gegenüber dem Standardspeicher ist, daß nicht bekannt sein muß, wo etwas gespeichert ist. Wenn viele Lieder gespeichert sind, müßte beim normalen Speicher zu jedem Lied die Startadresse im Speicher bekannt sein. Ein Nachteil von Assoziativspeichern ist aber, daß der Speicher nur entlang der gespeicherten Assoziationen ausgelesen werden kann. Dies bedeutet zum Beispiel, daß der Liedtext nur vorwärts ausgelesen werden kann. Ist dagegen nur der Schluß des Liedes bekannt, so könnte man es nicht einfach rückwärts aus dem Assoziativspeicher abrufen.

Heteroassoziation verglichen mit Autoassoziation

Normalerweise würde man - wie im obigen Beispiel - Daten mit anderen Daten verknüpfen, um sie in Assoziativspeichern zu speichern. Dies bezeichnet man mit Hetereoassoziation, man bekommt beim Auslesen immer neue Daten, die man wieder zum Adressieren verwenden kann.

Unter Autoassoziation versteht man dagegen den Fall, wo die Daten zum Abspeichern mit sich selbst verknüpft werden. Wenn man einen solchen Speicher ausliest, so bekommt man das, was man zum Adressieren verwendet, wieder zurück. Dies macht auf dem ersten Blick nicht viel Sinn, unter bestimmten Vorraussetzungen zeigt sich aber der Nutzen: Die Assoziativspeicher ermöglichen es, daß man mit unvollständigen Mustern adressiert und trotzdem die gewünschten Daten wiederbekommt. Wenn zum Beispiel Adressen (Name, Straße, Wohnort) autoassoziativ gespeichert werden, so könnte man durch das Adressieren mit einem Namen und einem Wohnort den kompletten Datensatz wieder auslesen, und so auch die Straße wieder bekommen. Diese Eigenschaft wird Mustervervollständigung genannt.

Realisierung von Assoziativspeichern

Für die Realisierung von Assoziativspeichern stehen verschiedene Modelle zur Verfügung. Im folgenden werden spikende Assoziativspeicher verwendet, da diese zum einen dem biologischen Vorbild besser entsprechen und zum anderen noch eine weitere Fähigkeit besonders gut unterstützen. Mit Hilfe der Spikes wird die Mustertrennung stark verbessert. Wenn z.B. Kochrezepte autoassoziativ gespeichert wären, so könnte man mit einer Mischung aus zwei Kochrezepten adressieren, und würde abwechselnd das eine und danach das andere Originalrezept als Antwort bekommen.

Datenrepräsentation im spikenden Assoziativspeicher

Im Gegensatz zu einer früheren Theorie, bei der angenommen wird, daß einzelne Neuronen eine bestimmte Information repräsentieren - wie zum Beispiel eine Großmutterzelle, die immer aktiv wird, wenn die Großmutter zu sehen ist - werden die Daten in Assoziativspeichern als Assemblies gespeichert [48]. Dies bedeutet, daß nicht ein einzelnes Neuron eine einzelne Information repräsentiert, sondern die gleichzeitige Aktivität von Neuron Gruppen einer bestimte Information entspricht (vgl. Abb. 2.10). Zum Beispiel könnten Gruppen mit je 50 Neuronen immer ein Person repräsentieren. Dabei dürfen sich die Gruppen auch überlappen, so daß ein paar Neuronen zwei gleiche Personen zugeordnet sind. Eine solche Gruppe von Neuronen nennt man Assembly.

In einem spikenden Assoziativspeicher werden die Daten auch als Assemblies repräsentiert. Die Aktivitätsmuster einzelner Neuronen enthalten für sich alleine genommen keine Informationen. Erst die Betrachtung der synchronen Aktivitätsmuster aller Neuronen erlaubt eine Aussage darüber, welche Informationen gerade vom Assoziativspeicher repräsentiert werden.

2.4.2 Verwendetes Assoziativspeichermodell

Das für diese Arbeit verwendete Assoziativspeichermodell ist nahezu identisch zu dem Modell von Knoblauch [23]. Es handelt sich um spikende Assoziativspeicher, bei der jeder Speicher für sich autoassoziativ funktioniert. Die Verbindung mehrerer Speicher dagegen ist hetereoassoziativ.

In dieser Arbeit werden zwei Assoziativspeicher zur Modellierung des Kortex eingesetzt. Jeder Speicher besteht aus drei Populationen, wobei zwei 40x40 Neuronen groß sind und eine 20x20 Neuronen. Die Details sind in [23] ausführlich erklärt. Jede Assembly besteht aus 50 Neuronen.

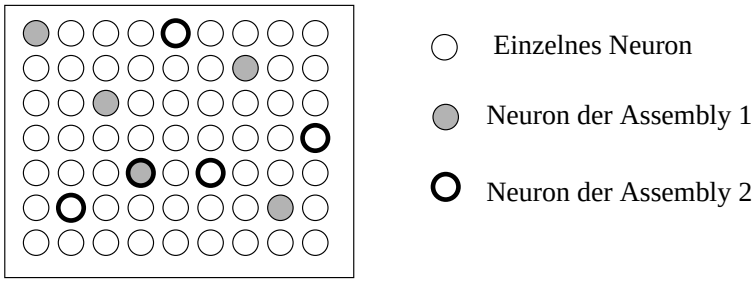

Abbildung 2.10: Beispiel eines neuronalen Assoziativspeichers. Die Informationen sind in Form von Assemblies gespeichert, jede Assembly besteht hier aus 5 Neuronen. Zwei Assemblies sind beispielhaft hervorgehoben. Ein Neuron gehört zu beiden Assemblies. Erst durch die Betrachtung aller Neuronen kann festgestellt werden, welche Assembly gerade aktiv ist.

Areal C

Das Areal C modelliert den primären auditorischen Cortex. Es ist ähnlich zu der Karte im IC aufgebaut: Sowohl die Frequenz als auch die Periodizität ist kartiert. Im Gegensatz zu der Ebene des ICs, wo allgemein anerkannt ist, daß diese Karte orthogonal aufgebaut ist, ist im primären auditorischen Kortex nicht klar, ob die Karte nicht auch eine zirkuläre Struktur aufweisen könnte. Dagegen sprechen mehrere Ergebnisse [30] und der unbekannte Nutzen, den die zirkuläre Struktur mit sich bringen würde. Daher ist die Karte im Areal C auch orthogonal wie im IC modelliert.

Areal D

Das Areal D repräsentiert ein höheres Areal, in dem keine systematische Kartierung der Informationen vorhanden ist. Vielmehr werden hier die Neuronen, die zu einer Assembly gehören, zufällig ausgewählt.

2.5 Das komplette Modell

Die Verbindung der peripheren Areale und der kortikale Areale erfolgt durch die Verbindung der Karte im IC mit dem Assoziativspeicher C (vgl. Abb 2.11). Zusätzlich existiert noch eine inhibitorische Population auf der Ebene des ICs, die es ermöglicht, laterale Inhibition innerhalb des ICs und inhibitorische Rückkopplungen aus dem kortikalen Areal C zu simulieren.

2.5.1 Neuronanzahl

Außer der Gammaton Filterbank werden alle Berechnungen mit Neuronen simuliert. Diese verteilen sich auf die unterschiedlichen Schichten wie folgt:

- Nucleus Cochlearis:

  50x50x2=5000 Neuronen für das Triggerneuron.

  50x50x3=7500 Neuronen für den Oszillatorschaltkreis.

  50x50x3=7500 Neuronen für den Integrator/Flip/Flop-Schaltkreis des originalen Langnermodells.

  oder

  50x50x2=5000 Neuronen für den Integrator und das IntegStopneuron

- Colliculus Inferior:

  50x50x2=5000 Neuronen für die Koinzidenzneurone und die zusätzliche inhibitorische Schicht.

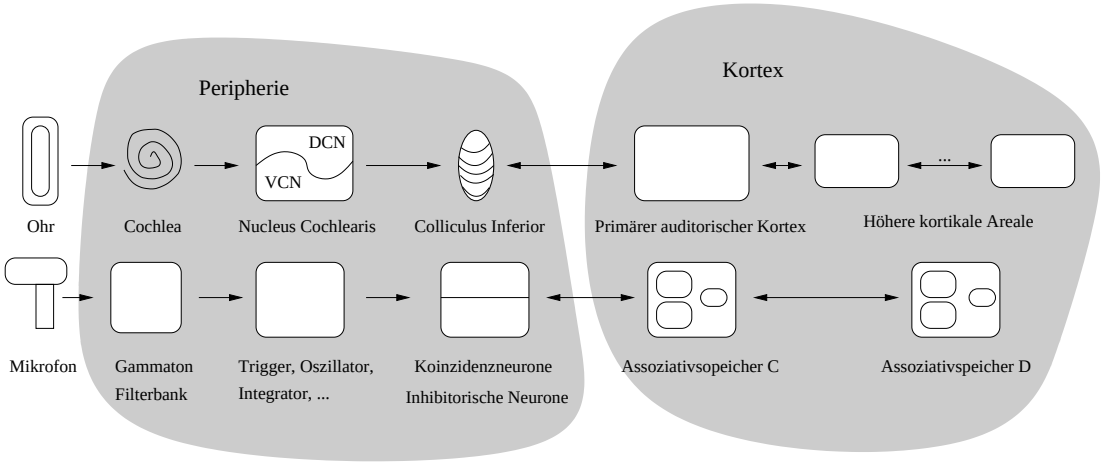

Abbildung 2.11: Das komplette Modell (untere Reihe) mit den jeweiligen biologischen Vorbildern (obere Reihe).

50x50x1=2500 Neuronen beim variierten Modell für das Inhibneuron (könnte auch im lateralen Lemniscus sitzen).

- Kortikale Areale

  40x40x2=3200 Neuronen für zwei Populationen des C Assoziativspeichers.

  20x20x1=400 Neuronen für die dritte Population des C Assoziativspeichers.

  40x40x2=3200 Neuronen für zwei Populationen des D Assoziativspeichers.

  20x20x1=400 Neuronen für die dritte Population des D Assoziativspeichers.

- Insgesamt ergibt dies für beide Modelle je 32200 Neuronen, die simuliert werden müssen.

Die Simulationsprogramme werden mit Felix II [24] erstellt, dies ist eine C++ Erweiterung zur Simulation großer neuronaler Netze. Alle peripheren Bereiche, also die Gammatone Filterbank und die Langnermodelle wurden erstmals für diese Arbeit implementiert. Durch die Verwendung der Simulationsumgebung Felix II konnte der Programmieraufwand reduziert werden. Dagegen mußten die Parameter der Neuronen und deren Verbindungen durch gezieltes Ausprobieren in teilweisen aufwendigen Simulationen eingestellt werden. Für die kortikalen Areale konnten die Programme einer anderen Arbeit weitgehendst übernommen werden, allerdings waren auch hier einige Parameter neu einzustellen.

Kapitel 3

Ergebnisse

Im folgenden werden die Ergebnisse dargestellt, die sich aus den Untersuchungen ergeben haben. Zuerst sind die Ergebnisse des ursprünglichen Langnermodell aufgeführt. Durch diese werden dann die Modifikationen motiviert, die zum variierten Langnermodell geführt haben.

Mit dem variierten Langnermodell werden dann die Assoziativspeicher verbunden, so daß sich das komplette Modell daraus ergibt.

Abschließend werden noch einige Ergebnisse aus der Psychophysik dargestellt.

Dabei stellen die Graphen typische Ergebnisse dar, die sich auch ergeben würden, wenn die Parameter des Modells leicht variiert werden, oder ein etwas veränderter Stimulus verwendet werden würde. Da eine Vielzahl von Parametern variiert werden können, ist es ein Test, der alle Parameter unabhängig variiert, nicht möglich. Die Überzeugung, daß die Schaltkreise stabil gegenüber kleinen Veränderungen sind, ergibt sich aus vielen einzelnen Simulationen, bei denen zum einem unterschiedliche Parameter per Hand variiert wurden, und zum anderen aus automatisierten Simulationsläufen, bei denen der Reihe nach auf alle Neuronpotentiale und auf den Stimulus Rauschen addiert wurde, und anhand der erzeugten Spikes die Stabilität des Schaltkreises festgestellt wurde [7]. Die Ergebnisse wurden also nicht mit einem unstabilen Modell erzeugt, sondern ergeben sich auch unter Variation der Randbedingungen.

Wenn eine Abbildung nicht typische Ergebnisse repräsentiert, so wird dies im Text genauer erläutert.

3.1 Ergebnisse zum ursprünglichen Langnermodell

Zuerst werden die Ergebnisse der Simulationen mit einem einzelnen Langner Schaltkreis dargestellt. Für diese wurden künstliche Stimuli verwendet, die eine genaue Analyse des Schaltkreisverhaltens ermöglichen.

Anschließend werden viele Schaltkreise zu einer Karte verbunden, die der Karte im IC entspricht. Für diese Simulationen werden auch natürliche Stimuli verwendet, wie Musik oder Sprache.

3.1.1 Ergebnisse der Simulationen eines einzelnen Langner Schaltkreises

Für die Darstellung der folgenden Ergebnisse wird oft eine Darstellung verwendet, die in der Biologie üblich ist, um Antworten von einzelnen Neuronen zu visualisieren. Es handelt sich um einen Graphen, bei dem auf der X-Achse die Zeit aufgetragen ist, und auf der Y-Achse entweder die Frequenz oder die Periodizität des Stimulus - abhängig davon, ob ein Sinussignal oder ein AM Signal als Stimulus verwendet wird. Immer wenn das untersuchte Neuron einen Spike generiert, so wird ein Punkt an dieser Stelle eingetragen.

Der Stimulus wird jeweils zum Zeitpunkt 10 ms mit einer Rampe von 5 ms angeschaltet, und bleibt dann 70 ms aktiv. Danach wird die Amplitude mit einer Rampe von 5 ms auf 0 reduziert, die Simulation läuft dann weitere 10 ms. Insgesamt werden also 100 ms simuliert (vgl. Abb. 3.1).

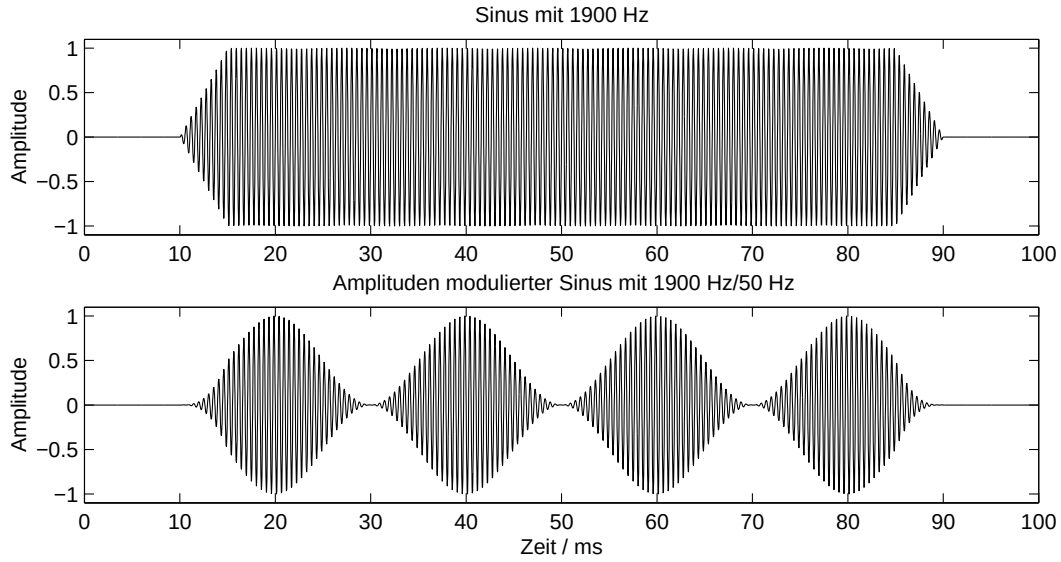

Abbildung 3.1: Zwei typische Stimuli. Der obere ist ein einfacher Sinus mit 1900 Hz, der untere ein AM Signal mit 1900 Hz Trägerfrequenz und 50 Hz Periodzität. (vgl. die Vergrößerungen auf der rechten Seite)

Analyse eines einzelnen Schaltkreises

Für die folgenden Versuche wird ein einzelner Langner Schaltkreis verwendet. Zuerst werden die Parameter des Stimulus bestimmt, auf die der Schaltkreis eingestellt ist. Dies kann direkt aus dem Aufbau des Schaltkreises geschehen, oder mit der Methodik, die ein Neurobiologe verwenden muß, wenn er einen unbekanntes Neuron analysiert. Diese Vorgehensweise wird im folgenden dargestellt und durchgeführt.

Bestimmung der Frequenz, auf die das Triggerneuron eingestellt ist

Zuerst wird die Trägerfrequenz bestimmt. Dazu wird untersucht, bei welchen Frequenzen das Triggerneuron antwortet. Als Stimulus wird ein Sinus Signal verwendet, dessen Frequenz von 500 Hz bis 15 kHz in Schritten zu 5 Hz variiert wird (vgl. Abb. 3.2). Man sieht, daß das Trigger um 1900 Hz antwortet. Das Antwortverhalten ist meistens phasisch (also Spikes nur am Anfang des Sinus), an den Grenzen auch tonisch (also Spikes während des gesamten Stimulus). Das Triggerneuron zeigt somit, daß der zugehörige Cochleafilter ein Bandpaß mit einer Mittelfrequenz von ca. 1900 Hz ist.

Bestimmung der Periodizität, auf die der gesamte Schaltkreis eingestellt ist

Nachdem die Trägerfrequenz von 1900 Hz festgestellt wurde, wird jetzt ein AM Signal erzeugt, dessen Modulationsfrequenz von 30 Hz bis 230 Hz mit einer Schrittweite von 1 Hz variiert wird. In Abb. 3.3 sind die Antworten des Triggerneurons und des Koinzidenzneurons dargestellt. Man sieht, daß die Antwort des Triggerneurons zwar von der Modulationsfrequenz abhängt, aber nur insofern, als daß die Anzahl der Spikes bei steigender Modulationsfrequenz zunimmt, was auch zu erwarten ist, da mehr Modulationsperioden in denselben Zeitraum passen, wenn die Frequenz steigt.

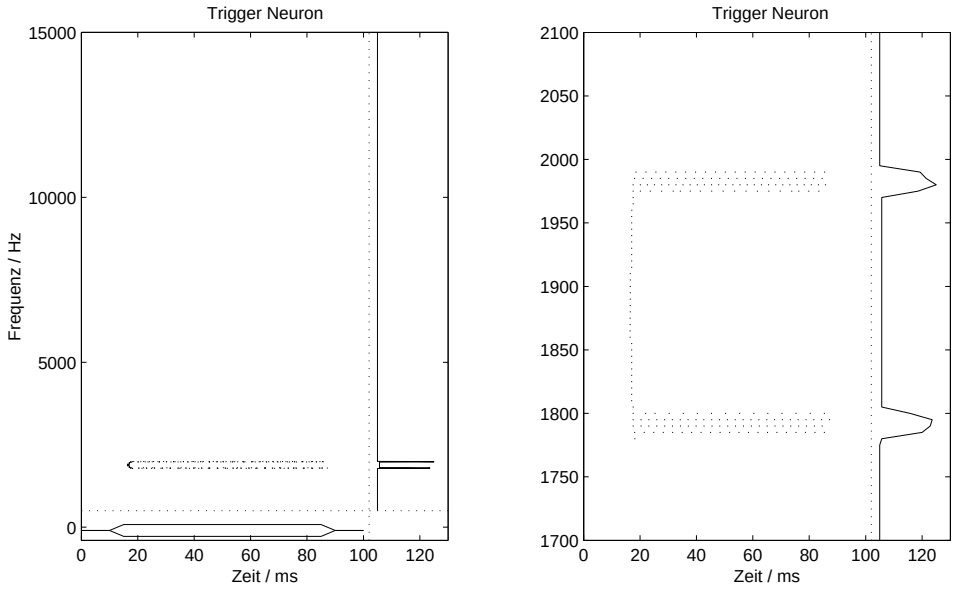

Abbildung 3.2: Links: Die Reaktion des Triggerneurons auf einen Sinus, dessen Frequenz von 500 Hz bis 15 kHz mit einer Schrittweite von 5 Hz variiert wird. Rechts: Eine Ausschnittsvergrößerung des linken Graphes im Bereich von 1700 Hz bis 2100 Hz.

Das Koinzidenzneuron dagegen zeigt eine deutliche Selektion der Periodizitäten. Bei 50 Hz und den vielfachen von ca. 50 Hz zeigt sich Aktivität, dazwischen werden keine Spikes erzeugt. Der Schaltkreis ist also auf 50 Hz Periodizität eingestellt, die Antworten auf die vielfachen von 50 Hz werden Kammfilterantworten genannt. Auf diese wird später noch genauer eingegangen.

Weiterhin zeigt sich in Abb. 3.3, daß ein einzelner ursprünglicher Langner Schaltkreis höchstens bei jeder zweiten Periode der Modulationsfrequenz antwortet. Der Grund dafür wird später diskutiert.

Bestimmung der passendsten Lautstärke

Nachdem die Parameter Trägerfrequenz und Modulationsfrequenz bestimmt wurden, wird jetzt noch die Lautstärke variiert, um festzustellen, wie sich der Schaltkreis verhält, und dann eine geeignete Lautstärke gewählt. Die Lautstärke wird dafür in Prozent zu einem Stimulus angegeben, dessen Maximum bei 1 liegt. Das Ergebnis ist in Abbildung 3.4 dargestellt. Es zeigt sich, daß eine minimale Amplitude von ca. 19 % für das AM Signal notwendig ist. Zwischen ca. 31 % und ca. 43 % ist die Antwort mit allen Kammfilterantworten komplett. Mit zunehmender Lautstärke werden dann die Kammfilterantworten geringer. Die passendste Lautstärke liegt demnach zwischen 31 % und 43 %, daher wird 37 % als Mittelwert von beiden als geeignetste Lautstärke gewählt. Interessant ist das Verschwinden der Kammfilterantworten für laute Signale. Dies wird im folgenden genauer untersucht:

Der Grund für das Verschwinden wird in Abb. 3.5 deutlich: Die Kammfilterantworten fallen nicht weg, weil es sich um Kammfilterantworten handelt, sondern weil das Triggerneuron in einem begrenzten Lautstärkebereich für alle Periodizitäten funktioniert (erstes Bild von links). Ohne die Spikes des Triggerneurons kann aber der gesamte Schaltkreis nicht korrekt funktionieren. Daher sieht man im zweiten Bild von links, daß die Antworten des Koinzidenneurons sich entsprechend auch nicht über den gesamten Lautstärkebereich erstrecken, sondern nur dort sind, wo das Triggerneuron Spikes erzeugt hat. Daß die Antwortbereiche exakt übereinstimmen, zeigt das dritte Bild, welches eine Überlagerung der Antworten des Koinzidenzneurons und des Triggerneurons ist. Zur Verdeutlichung wurde der rechte Graph abgebildet, der eine einzelne Zeile der Überlagerung als

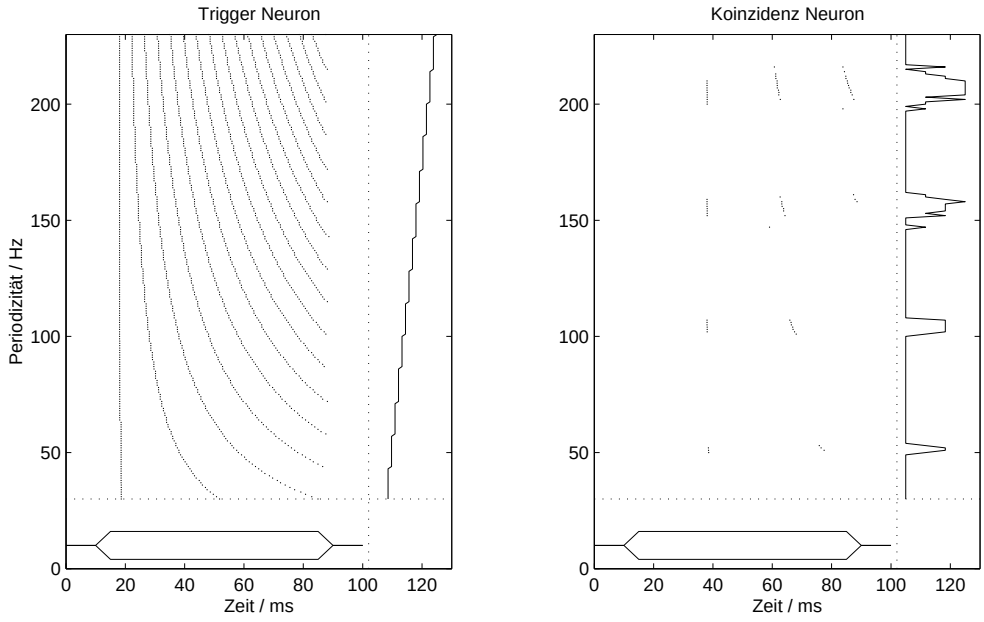

Abbildung 3.3: Bestimmung der passendsten Periodizität des Langner Schaltkreises. Die Modulationsfrequenz wurde von 30 Hz bis 230 Hz in 1 Hz Schritten variiert. Das erste Maximum bei ca. 50 Hz Periodizität zeigt die Modulationsfrequenz, auf die der Schaltkreis eingstellt ist. Die weiteren Maxima zeigen die Kammfiltereigenschaft des Langnermodells.

Plot darstellt. Man erkennt, wie die Triggerneuronantwort ab einer bestimmten Periodizität (ca. 150 Hz) zusammenbricht.

Das Verhalten der einzelnen Neuronen auf verschiedene Stimuli

Nachdem die Bedingungen an den Stimulus festgestellt wurden (50 Hz Periodizität, 1900 Hz Trägerfrequenz, 37% maximale Amplitude) der den Schaltkreis aktiviert, werden im folgenden verschiedene Stimuli als Input verwendet, und die Antwort aller Schaltkreis Neuronen dargestellt. Dies ermöglicht zum einen ein besseres Verständnis des Schaltkreises und zum anderen einen Vergleich zu Daten aus der Elektrophysiologie.

Sinus Signal

In Abb. 3.6 sind die Antworten aller Neuronen dargestellt. Entlang der Y-Achse ist die Frequenz aufgetragen, entlang der X-Achse die Zeit. Die Frequenz wurde von 500 Hz bis 5000 Hz mit einer Schrittweite von 100 Hz verändert.

Amplitudenmodulierter Stimulus

In Abb. 3.7 sind die Reaktionen der einzelnen Neuronen auf ein AM Signal dargestellt. Der Aufbau der Grafik ist wie in Abb. 3.6. Es zeigt sich, daß der Flip/Flop Schaltkreis bei einem AM Stimulus den Integrator auf die Triggerimpulse synchronisiert. Die Kammfilterantworten entstehen immer dann, wenn der Integrator über mehrere Perioden aufintegriert, also Triggerspikes ignoriert.

3.1.2 Ergebnisse bei modifizierten Verbindungsstärken

Ein wichtiger Punkt des Langnermodells ist das Verhalten des Flip/Flop Schaltkreises, wenn gleichzeitig ein Spike vom Triggerneuron das Flipneuron erreicht und ein Spike vom Integrator an dem Flopneuron ankommt. Je nach dem wie die Verbindungsparameter eingestellt sind wird in diesem

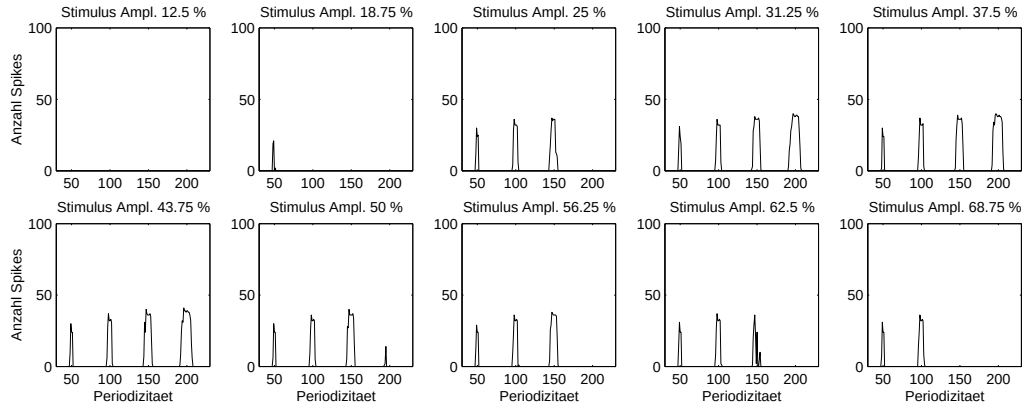

Abbildung 3.4: Die Antworten des Koinzidenzneurons auf einen AM Stimulus mit 1900 Hz, variierter Periodizität und unteschiedlichen Amplituden. Dargestellt ist für eine bessere Übersichtlichkeit nicht mehr jeder einzelne Spike, sondern nur noch die Summenantwort.

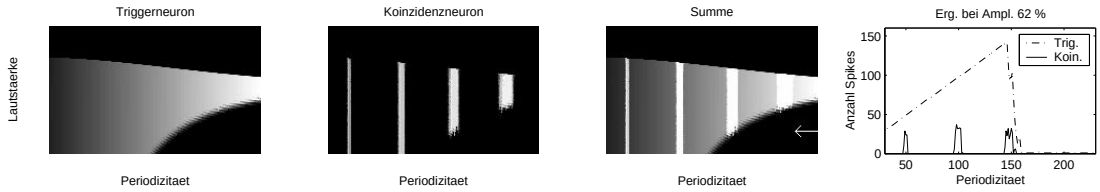

Abbildung 3.5: Genaue Analyse des Antwortverhaltens bei variierter Lautstärke: Der linke Graph stellt die Antwort des Triggerneurons durch Grauwertkodierung dar: Je heller ein Punkt, desto mehr Spikes wurden erzeugt. Auf der X-Achse ist die Periodizität aufgetragen, auf der Y-Achse die Lautstärke. Man sieht, daß das Triggerneuron nur bei mittlerer Lautstärke für alle Periodizitäten funktioniert. Im Bild rechts daneben ist entsprechend die Antwort des Koinzidenzneurons dargestellt. Die Überlagerung beider Antworten ist das dritte Bild. Ganz rechts ist zusätzlich eine Zeile der Überlagerung als Graph geplottet. Der Pfeil im dritten Graph zeigt die Zeile, die für den Plot rechts verwendet wurde.

Fall das Flipneuron oder das Flopneuron aktiv. 'Gewinnt' das Flopneuron, so wird der Integrator gehemmt, bis der Trigger den nächsten Spike generiert. Dies ist die Konfiguration, die der Schaltkreis bisher hatte. Interessant ist, was passiert, wenn das Flipneuron gewinnt. Um dies zu erreichen wird für die folgenden Ergebnisse die Verbindungsstärke so eingestellt, daß nicht das Flopneuron sondern das Flipneuron aktiv wird.

Wichtig ist die Frage deswegen, weil der Fall, daß Triggerneuron und Integratorneuron gleichzeitig Spikes erzeugen, dann eintritt, wenn der Schaltkreis auf den Stimulus antwortet, also die Periodizität des Schaltkreises mit der Periodizität des Stimulus übereinstimmt.

Die Ergebnisse sind in Abb. 3.8 dargestellt. Die obere Reihe zeigt den kompletten Bereich von 30 Hz bis 230 Hz, die untere Reihe dagegen eine Ausschnittsvergrößerung. Links sind jeweils die Ergebnisse dargestellt, wenn das Flopneuron 'gewinnt', rechts die Ergebnisse, wenn das Flipneuron sich durchsetzt. Man erkennt, daß in beiden Fällen die Summenantwort ähnlich aussieht (obere Reihe). Die Spikes dagegen weisen einen deutlichen Unterschied auf, zumindest für Periodizitäten von 50 Hz, 100 Hz und 150 Hz. Wird das Integratorneuron vom Flopneuron gehemmt (linker Fall wenn das Flopneuron sich durchsetzt), so kann erst wieder die darauffolgende Periode detektiert werden, und daher ist bei 100 Hz die zweite Spikereihe (ca. 70 ms) nicht an der gleichen Stelle, wie bei 50 Hz Periodizität (60 ms). Die Spikes bei 150 Hz liegen entsprechend bei ca. 66 ms. Das ergibt sich daraus, daß die Periode sich bei 150 Hz zu ca. 6 ms ergibt und somit das Auslassen

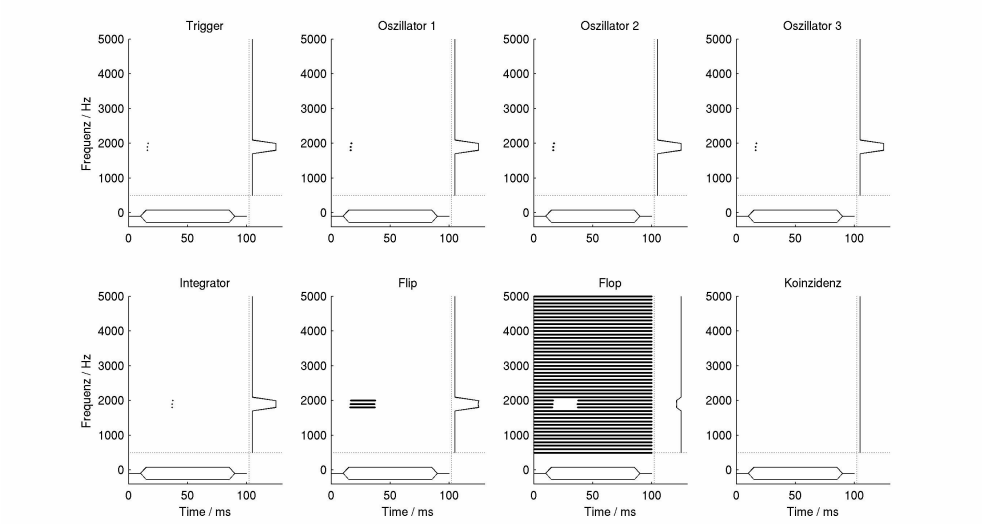

Abbildung 3.6: Die Spikes, die der ursprüngliche Langner Schaltkreis erzeugt, wenn als Stimulus Sinus-Signale mit verschiedenen Frequenzen verwendet werden (Y-Achsen). Die X-Achsen entsprechen der Zeit. Für jedes Neuron ist die Antwort in einem separatem Graph dargestellt, jeweils die zeitliche Antwort (linker Bereich in jedem Plot) und die normalisierte Summe über die Zeit (rechter Bereich). Man sieht, daß nur um die 1900 Hz überhaupt eine Änderung der Aktivität festzustellen ist, wobei das Koinzidenzneuron überhaupt keine Antwort erzeugt.

einer Periode eine Verzögerung von 6 ms erzeugt (entsprechend 10 ms bei 100 Hz).

Betrachtet man dagegen die Spikepositionen des Graphen rechts oben, so sieht man, daß die Spikes für 50 Hz, 100 Hz und 150 Hz untereinander liegen. Wenn das Flipneuron 'gewinnt', dann kann das Integratorneuron direkt nach dem Feuern wieder anfangen zu integrieren, da das Flopneuron nicht aktiv wird. Dadurch fällt die Verzögerung von einer Periode weg.

Eigentlich scheint das rechte Verhalten besser zu sein als das linke. Es werden weniger Perioden ausgelassen. Dennoch kann dieses Verhalten zu Problemen führen, wie in der unteren Abbildung durch die Ausschnittsvergrößerung klar wird. Diese zeigt einen Bereich von 45 Hz bis 55 Hz Periodizität. In dem linken Graphen sind zwei Linien von Spikes bei ca. 40 ms und 80 ms. Diese Linien sind auch im rechten Graphen. Zusätzlich sind für Periodizitäten von knapp unter 50 Hz auch noch Spikeantworten in der Mitte bei ca. 60 Hz zu sehen. Für diese Bereiche ergibt sich also für jede Periode ein Spike (50 Hz entspricht 20 ms). Obwohl dies eigentlich das Ziel eines Periodizitätsdetektors sein sollte, ist dies aber trotzdem schlecht. Dies liegt darin begründet, daß durch den Aufbau des Langner Schaltkreises eine Art Phasensprung stattfindet, wo das Flopneuron später als das Flipneuron aktiviert wird. Wenn die Periodizität etwas ansteigt, so erhöht sich auch die Spikerate des Triggerneurons und somit kommt der Input des Flipneurons früher. Das Integratorneuron dagegen benötigt eine konstante Zeit. Dadurch bekommt bei langsameren Periodizitäten (z.B. 48 Hz) das Flopneuron zuerst den Input vom Integratorneuron, und etwas später dann das Flipneuron den Input vom Triggerneuron. Das erhöht die Wahrscheinlichkeit, daß das Flipneuron sich durchsetzt. Bei 52 Hz dagegen kommt der Input vom Trigger früher als der Input vom Integrator, so daß eher das Flopneuron sich durchsetzen wird. In diesem Fall aber kann der Integrator nur jede zweite Periode aufintegrieren, und daher fallen die zusätzlichen Spikes bei 60 ms mit zunehmender Periodizität weg (vgl. Graph rechts unten). Der Schaltkreis zeigt also ein Verhalten, bei dem eine kleine Änderung des Stimulus eine Verdoppelung bzw. Halbierung der Antworten des Koinzidenzneurons erzeugt. Dieses Verhalten ist nicht erwünscht, wenn die Menge der Spikes Rückschlüsse auf den Stimulus zulassen sollen.

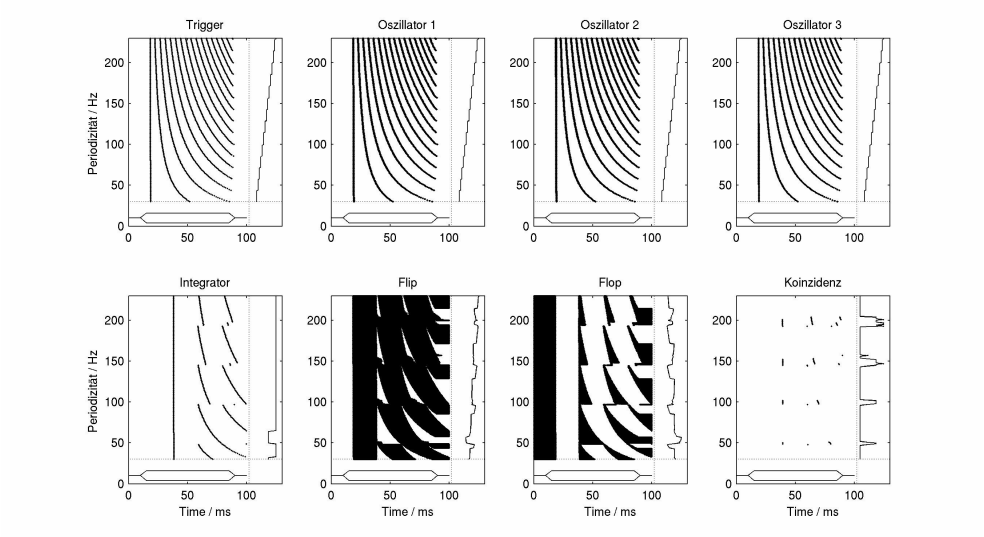

Abbildung 3.7: Darstellung wie in Abb. 3.6, als Input wurde ein AM Signal ohne Rauschen verwendet. Die Trägerfrequenz des AM Stimulus liegt bei der Bestfrequenz des Neurons.

3.1.3 Ergebnisse der Simulationen von IC Karten

Um die gewünschte Karte des ICs zu bekommen muß der einzelne Schaltkreis entsprechend oft vervielfältigt werden. Durch die systematische Veränderung der Integrationskonstante des Integratorneurons wird erreicht, daß unterschiedliche Schaltkreise auf unterschiedliche Periodizitäten antworten. Dies erzeugt die Periodizitätsachse der Karte, also eine Zeile in der Karte. Diese Zeile wird vervielfältigt und mit unterschiedlichen Ausgängen der Gammaton-Filterbank verbunden. Dadurch entsteht die Trägerfrequenzachse. Für die folgenden Versuche wird die Größe der Karte auf 50x50 festgelegt. Die Trägerfrequenz erstreckt sich von ca. 400 Hz bis 4000 Hz.

Verwendung von künstlichen Stimuli

In Abb. 3.9 sind die Ergebnisse dargestellt, wenn unterschiedliche AM Stimuli verwendet werden. Es wurden vier verschiedene Stimuli verwendet: 100 Hz/1000 Hz, 100 Hz/3000 Hz, 200 Hz/1000 Hz, 200 Hz/3000 Hz (jeweils Modulationsfrequenz/Trägerfrequenz). Man sieht, daß die Aktivitätsflecken entsprechend der Stimuliparameter auftreten. Niedrige Periodizitäten erzeugen links in der Karte Aktivitäten, hohe Periodizitäten rechts. Entsprechend niedrige Trägerfrequenzen erzeugen unten in der Karte Aktivitäten, hohe Frequenzen oben.

Zusätzlich kann man bei den beiden Stimuli mit 200 Hz Periodizität den Einfluß der Kammfilterantworten auf die Karte erkennen. Der deutliche helle Aktivitätsfleck entsteht rechts, aber kleinere, dunklere Flecken entstehen auf der gleiche Höhe weiter links. Dies sind die Neuronen, die auf 100 Hz/66 Hz/50 Hz usw. eingestellt sind. Diese werden später näher untersucht.

Interessant ist auch die unterschiedliche Breite der Antwortflecken für 100 Hz Periodizität und 200 Hz Periodizität: Die Antwort für 200 Hz ist wesentlich breiter als für 100 Hz. Das kann mit folgender Überlegung erklärt werden:

- Das Koinzidenzneuron feuert, wenn gleichzeitig Spikes vom Oszillator und vom Integrator ankommen.

- Die Spikes des Oszillators erstrecken sich über eine konstante Zeit von ca. 1 ms. Dies entspricht in etwa dem Zeitfenster, in dem der Integrator feuern muß, damit dadurch das Koinzidenzneuron feuert.

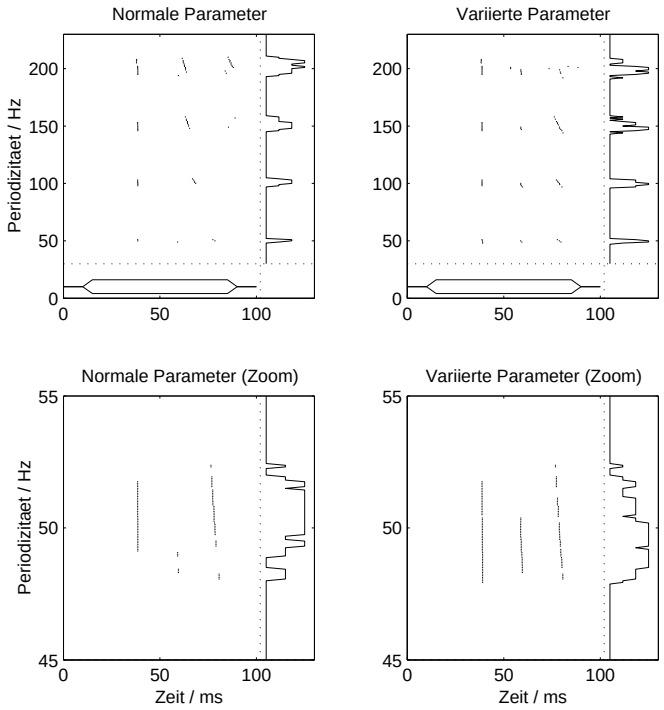

Abbildung 3.8: Ergebnisse der Untersuchung wenn die Verbindungsstärken modifiziert werden, so daß das Flip und nicht das Flopneuron aktiv wird, wenn beide gleichzeitig Input erhalten. Links 'gewinnt' das Flopneuron (wie bisher), rechts das Flipneuron (modifizierte Verbindungsstärken). In der oberen Reihe sind die Ergebnisse für 30 Hz bis 230 Hz dargestellt, in der unteren Reihe eine Ausschnittsvergrößerung der ersten Antwort. Wichtig sind die zusätzlichen Spikes im Graphen rechts unten, die zwischen den Spikes, die der bisherige Schaltkreis erzeugt hat, liegen.

  - 100 Hz Periodizität entsprechen 10 ms. Daraus ergibt sich, daß ein Schaltkreis, der auf 100 Hz eingestellt ist, in einem Bereich von 95 Hz bis 105 Hz (10.5 ms bis 9,5 ms) antwortet. Dies entspricht ca. 5 % Abweichung vom Normwert.
  - 200 Hz Periodizität entsprechen 5 ms. Entsprechend der Rechnung oben ergibt sich ein Antwortbereich eines solchen Schaltkreises von 166 Hz bis 250 Hz (6 ms bis 4 ms). Dies entspricht ca. 20 % Abweichung vom Normalwert.

- Daraus ergibt sich ein immer breiteres Tuning der einzelnen Schaltkreise entlang der Periodizitätsachse und entsprechend auch ein mit zunehmender Periodizität wachsender Fleck in der IC Karte.

Zusätzlich zeigt sich, daß die Koinzidenzneuronen immer nur jede zweite Periode feuern. Die maximale Spikeanzahl liegt bei ca. der Hälfte der möglichen Spikes. In Abb. 3.10 wurde die glei-

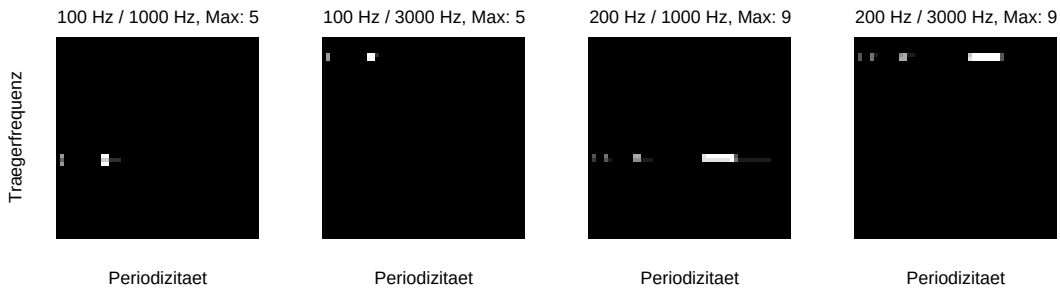

Abbildung 3.9: Ergebnisse des Langnermodells mit 50x50 Neuronen. Dargestellt ist mit Grauwerten kodiert die Anzahl der Spikes des Koinzidenzneurons. Der Stimulus war 100 ms lang, bei 100 Hz konnten also maximal 10 Spikes erzeugt werden (bei 200 Hz entsprechend 20 Spikes).

che Berechnung nochmals mit variierten Verbindungsstärken durchgeführt ('Flip gewinnt', vgl. Abschnitt 3.1.2), um den Einfluß auf die Karte darzustellen.

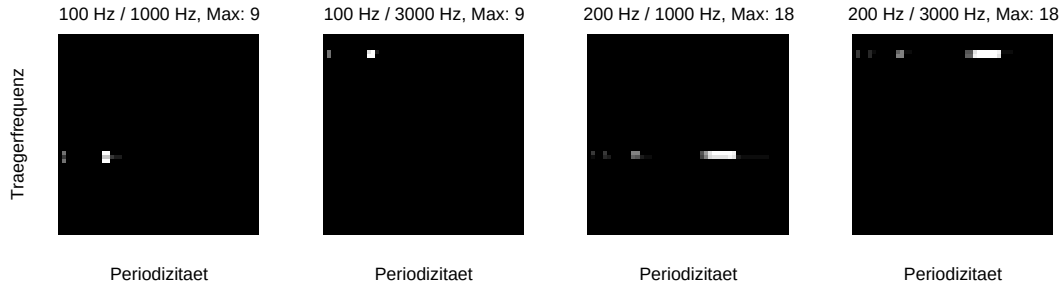

Abbildung 3.10: Die gleiche Darstellung wie in 3.9, nur wurden die Verbindungsparameter so modifiziert, daß das Flipneuron gewinnt

Man sieht, daß die maximalen Werte jetzt fast die theoretisch mögliche Anzahl Spikes erreicht. Dies erscheint eigentlich als günstigere Konfiguration. Allerdings besteht weiterhin das Problem, daß bei geringfügiger Modifikation der Periodizität des Stimulus sich die Antwortrate halbieren kann (vgl. Abschnitt 3.1.2).

Auf diese Aspekte wird in der Diskussion (vgl. Abschnitt 4.1.2 auf Seite 86) näher eingegangen.

Kammfilterantworten auf der IC Karte

Wie die Untersuchungen von einzelnen Langner Schaltkreisen zeigten, werden von ihnen Kammfilterantworten erzeugt, wenn die Periodizität des Stimulus ein vielfaches der Periodizität ist, auf das der Schaltkreis eingestellt ist. Dadurch entstehen auf der IC Karte auch Antworten an Stellen, die eigentlich nicht zum Stimulus passen. Um dies zu zeigen ist in Abb. 3.11 für vier verschiedene Periodizitäten je eine IC-Karte dargestellt. Als Stimulus wurden Klicks verwendet, das sind Impulse die sich entsprechend der jeweiligen Periodizität wiederholen. Dadurch ergeben sich vertikale Balken.

Allerdings entstehen diese Balken auf der Karte nicht bei Vielfachen sondern bei Bruchteilen der Periodizität. Wenn der Stimulus eine Periodizität von z.B. 200 Hz hat, dann ergeben sich Antworten bei 200 Hz, 100 Hz, 66 Hz, 50 Hz, ..., weil an diesen Stellen jeweils ein Schaltkreis sitzt, der auf einen Stimulus mit 200 Hz Periodizität antwortet.

Man sieht in den Plots unter den IC Karten jeweils die Anzahl der Spikes pro Periodizität, aufsummiert über die Trägerfrequenzen. Diese ist für die Kammfilterantworten deutlich geringer als die der eigentlichen Antwort, allerdings trotzdem klar erkennbar.

Verwendung von natürlichen Stimuli

In diesem Abschnitt werden die Ergebnisse eines Artikels [6] zusammengefaßt, in dem Sprachsignale und Orgeltöne mit Hilfe des Langnermodells analysiert wurden, die Abbildungen wurden direkt aus dem Artikel übernommen, und haben daher englische Bezeichnungen. Es wurden die folgenden Signale analysiert:

- Der Vokal 'a'.
- Der Vokal 'o'.
- Der Ton einer Orgelpfeife.

Zu jedem der drei Stimulustypen wurde ein Stimulus mit hoher Periodizität und mit niedrigerer Periodizität untersucht. In Abb. 3.12 sind Ausschnitte der Signale dargestellt anhand deren Periodizität erkennbar wird.

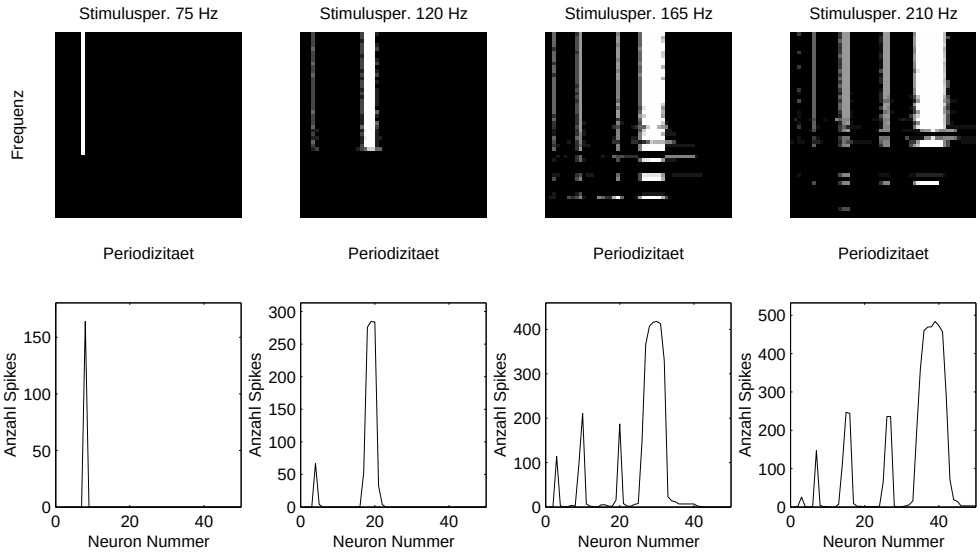

Abbildung 3.11: Die Ergebnisse zur Demonstration der Kammfilterantworten des ursprünglichen Langner Schaltkreises auf der Ebene der IC Karte. Als Stimuli wurden Klicks verwendet. Die obere Zeile zeigt die Karte, die untere die Summe der Spikes über die Frequenz.

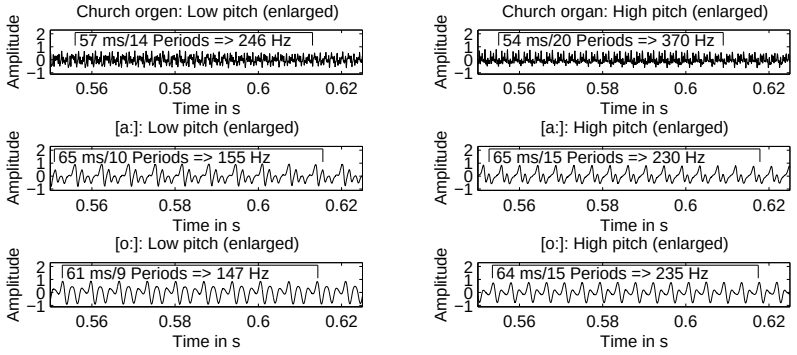

Abbildung 3.12: Ausschnitte der Originalsignale. Durch Ausmessen der Perioden wurde zusätzlich die Periodizität der Signale bestimmt und in die Plots mit eingetragen

In Abb. 3.13 sind die Ergebnisse dargestellt. Die damalige Implementation des Langnermodells hatte andere Parameter, daher deckt die Karte einen Frequenzbereich von ca. 750 Hz bis 13 kHz (Y-Achse), die Periodizität erstreckt sich von 20 Hz bis 400 Hz (X-Achse). Man erkennt deutlich die für das Langnermodell typischen Kammfilterantworten.

Ebenso deutlich zeigt sich, daß - abgesehen von den Kammfilterantworten - alle aktiven Neuronen zu einer Periodizität gehören, dadurch entstehen die vertikalen Linien. Für die Orgel sind nicht alle Trägerfrequenzbereiche aktiviert, aber für die aktivierten Bereiche gilt auch, daß sie alle bei einer Periodizität liegen. Dies bedeutet, daß die analysierten Signale in allen aktiven Trägerfrequenzbereichen die gleiche Modulationsfrequenz, also Periodizität, besitzen.

Aus der Position der vertikalen Balken kann auch die Periodizität abgeleitet werden, indem diese mit der Position von AM Signalen verglichen werden. Es ergeben sich ungefähr die Werte: 240 Hz für den tiefen Orgelton und 150 Hz für die tiefen Vokale. Für den hohen Orgelton ergibt sich 370 Hz (der Aktivitätsfleck rechts unten in der Karte) und für die hohen Vokale 230 Hz. Diese Werte sind keine exakten Werte, da ein Ton mit 231 Hz sehr ähnliche Ergebnisse erzeugt wie ein Ton mit 229 Hz. Daher kann mit dem Positionsvergleich die Periodizität nicht exakt bestimmt werden. Vergleicht man die ungefähren Werte mit der eigentlichen Periodizität des Signales (siehe Abb. 3.12), so zeigt sich, daß die durch den Positionsvergleich ermittelten Werte mit einer maximalen

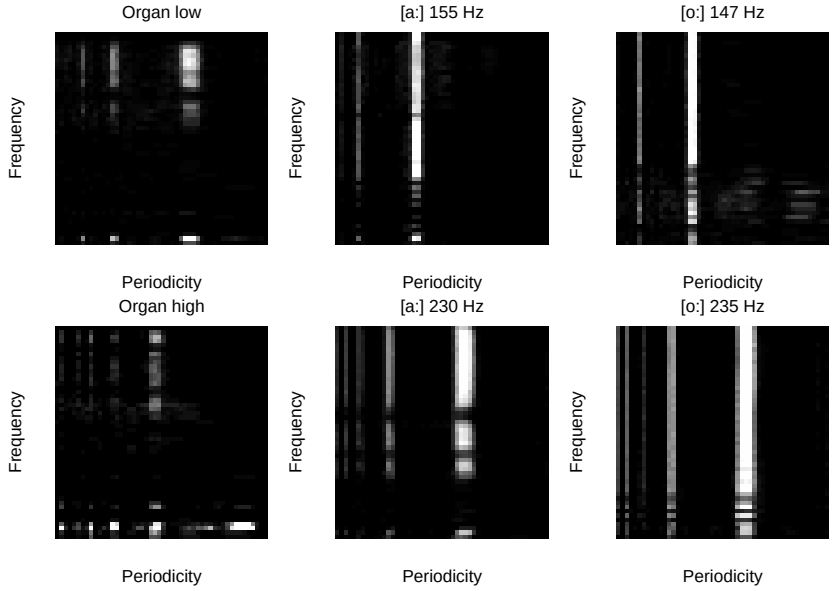

Abbildung 3.13: Ergebnisse der Untersuchungen von natürlichen Stimuli. Auf der Y-Achse sind die Trägerfrequenzen von 750 Hz bis 13 kHz aufgetragen, auf der X-Achse die Periodizität von 20 Hz bis 400 Hz.

Abweichung von 6 Hz gut übereinstimmen.

Ein anderes Verfahren, die Periodizität des Originalsignals zu bestimmen, besteht darin, daß man die Periodizität der Antwortspikes feststellt. Dazu eignet sich das folgende Verfahren, basierend auf der Korrelation:

1. Zuerst werden die Spikes einer Zeile zeitlich aufsummiert, so daß man als Ergebnis eine Funktionenmenge $Spike_f(t)$ bekommt, die zu jeder Trägerfrequenz f angibt, wieviele Neuronen im Moment t gespiket haben.

2. Von den Funktionen $Spike_f(t)$ wird dann die Autokorrelation berechnet:

$$SpAC_f(d) = \sum_{n=-\infty}^{\infty} Spike_f(d+n) * Spike_f(n)$$

3. Alle Autokorrelationen werden danach aufsummiert:

$$Erg(d) = \sum_{i=1}^{Anz_f} SpAC_f(d)$$

Die Reihenfolge des Aufsummierens, Korrelierens und wieder Aufsummierens ist sinnvoll, weil die Filterbank eine Phasenverschiebung zwischen den Frequenzkanälen erzeugt. Würden einfach alle Spikes der Karte auf einmal aufsummiert werden und danach eine Autokorrelation berechnet, so wäre das Ergebnise nicht so aussagekräftig, weil die Spikes unterschiedlicher Frequenzkanäle zwar die gleiche Periodizität haben können, zeitlich jedoch etwas versetzt sind.

Die Ergebnisse der Autokorrelationen sind in Abb. 3.14 dargestellt. Zu jedem Stimulus wurde das Maximum markiert, das neben dem zentralen Maximum das größte ist. Die für dieses Maximum geltende Verschiebung entspricht der Periodizität der von den Koinzidenzneuronen erzeugten Spikes. In der Tabelle 3.1 sind die per Hand gemessenen Periodizitäten zusammen mit den aus den Autokorrelation berechneten und den durch Positionsvergleich in der Karte bestimmten Werten eingetragen. Man sieht, daß die größte Differenz zwischen dem gemessenen und den durch Autokorrelation bestimmten Wert 4,1 Hz ist (für 'o' tief).

Weiterhin wird deutlich, daß die Positionen der vertikalen Linien in der Karte auch tatsächlich die Periodizität widerspiegeln. Die Differenz zwischen positionsbestimmter und durch Autokorrelation bestimmter Periodizität ist maximal 3,9 Hz.

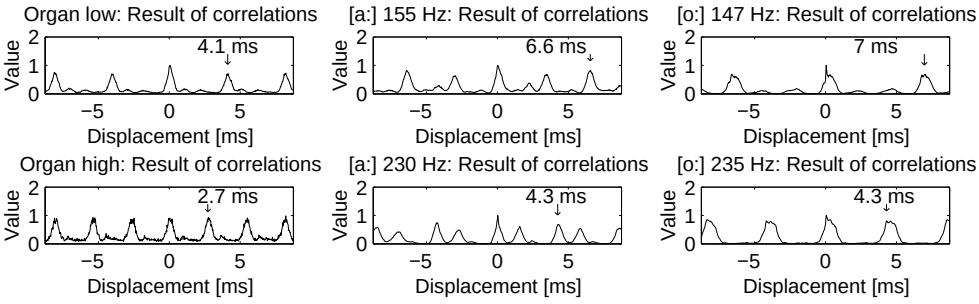

Abbildung 3.14: Ergebnis der Autokorrelationen. Für jeden Stimulus wurde zusätzlich das größte Nebenmaximum markiert und die zugehörige Zeitverschiebung angegeben.

Signal	gemessene Periodizität	durch Positionsbestimmung	aus Autokorrelation
Orgel tief	246 Hz	240 Hz	243,9 Hz
Orgel hoch	370 Hz	370 Hz	370,4 Hz
'a' tief	155 Hz	150 Hz	151,1 Hz
'a' hoch	230 Hz	230 Hz	232,6 Hz
'o' tief	147 Hz	150 Hz	142,9 Hz
'o' hoch	235 Hz	230 Hz	232,6 Hz

Tabelle 3.1: Gegenüberstellung der ermittelten Periodizitäten für die Stimuli. Die zweite Spalte ergibt sich aus der Messung per Hand, die dritte Spalte aus dem Vergleich der Positionen in der Karte und die vierte Spalte durch die Bestimmung der Maxima aus den Autokorellogramen.

Ein weiterer Aspekt, der sich aus der Analyse von natürlichen Signalen ergibt, zeigt sich in Abb. 3.15. Für diese Abbildung wurden für einen kurzen Zeitbereich die Spikes aller Neuronen der Karte aufsummiert. Man sieht, daß die Spikes nicht zufällig verteilt sind, sondern gemeinsam bevorzugt zu einem Moment auftreten, und danach für eine kurze Zeit keine Spikes auftreten. Dies zeigt, daß die aktiven Frequenzkanäle nicht nur mit der gleichen Periodizität moduliert werden, sondern diese Modulation auch in allen Frequenzkanälen phasensynchron ist. Wäre dies nicht der Fall so würden die Spikes gleichmäßig über die Zeit verteilt auftreten.

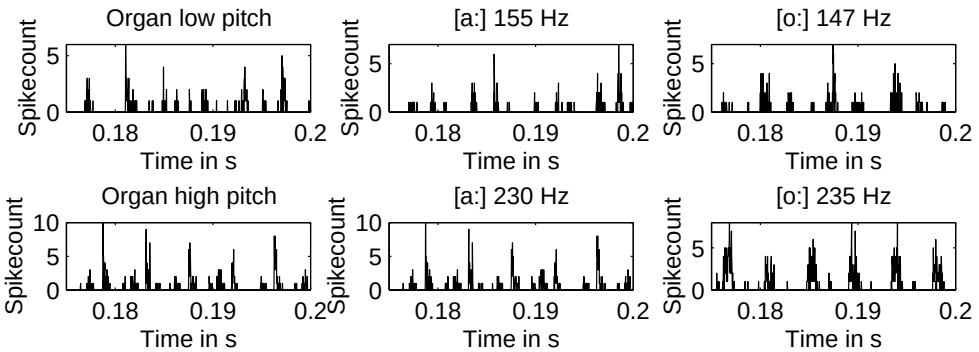

Abbildung 3.15: Darstellung der Spikes aller Neuronen für einen kleinen Zeitbereich. Man sieht, daß die Spikes gruppiert auftreten.

Einfluß der Lautstärke auf das Periodizitätstuning des Langnermodells

Ein Aspekt, der bisher nicht genauer untersucht wurde, ist, daß entsprechend dem Langnermodell die Integratorneuronen nicht einfach eine konstante Verzögerung erzeugen, wie im simulierten Modell, sondern der Input aufintegriert wird. Dadurch hängt die Verzögerung von der Amplitude des Stimulus ab, wodurch Effekte entstehen, die nicht erwünscht sind. Diese Effekte sind in Abb. 3.16 dargestellt. Als Stimulus wurden Klicks verwendet, deren Lautstärke von links oben nach rechts unten zunimmt (Bild links oben: Lautstärke 75 %, Bild rechts unten: Lautstärke 150 %).

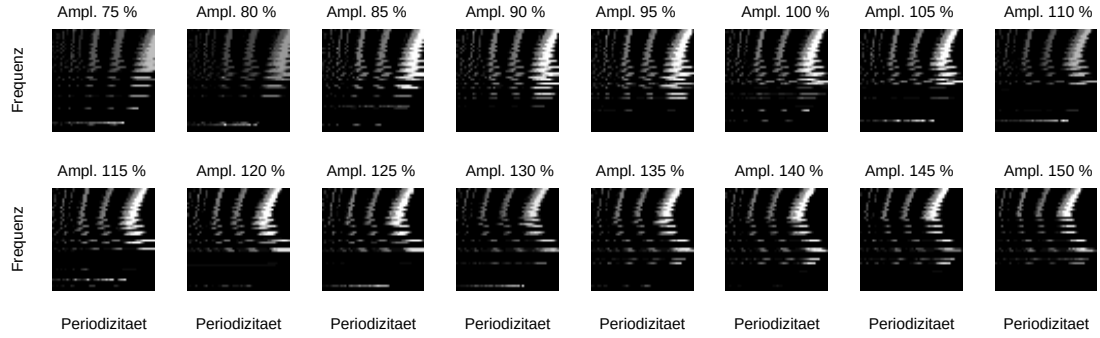

Abbildung 3.16: Darstellung des Einflusses der Lautstärke auf das Langnermodell, wenn der Integrator nicht eine feste Verzögerung erzeugt, sondern den Stimulus aufintegriert. Von Bild zu Bild wurde die Amplitude des Stimulus (Klicks) erhöht. Ansonsten ist der Stimulus identisch.

Man erkennt, daß sich die Bilder stark unterscheiden, obwohl die Periodizität immer die gleiche war, als einziges wurde die Amplitude der Klicks variiert. Dies liegt daran, daß bei der Verschaltung, die eigentlich vom Langnermodell gefordert wird, die Verzögerung der Integratoren stark vom Stimulus abhängt. Dadurch verschiebt sich das Periodizitätstuning der Schaltkreise entsprechend. In den bisherigen und nachfolgenden Simulationen wurde daher der Integrator mit einem konstanten Inputsignal verbunden. Die Konsequenzen, die sich aus dieser Änderung ergeben, werden in der Diskussion, Abschnitt 4.1.2 ab Seite 89, genauer diskutiert.

In Abb. 3.17 wurden die gleichen Berechnungen durchgeführt wie für Abb. 3.16. Allerdings wurde der Input des Integrators wie bisher konstant gehalten. Man sieht zum einen, daß die Bilder sich nicht wesentlich unterscheiden, und zum anderen sind die Antworten wieder vertikale Balken, entsprechend der Tatsache, daß alle Trägerfrequenzen die gleiche Periodizität beinhalten. Daher wird im folgenden der Integrator wieder mit konstanten Input beschaltet und nicht mit dem Stimulus Signal

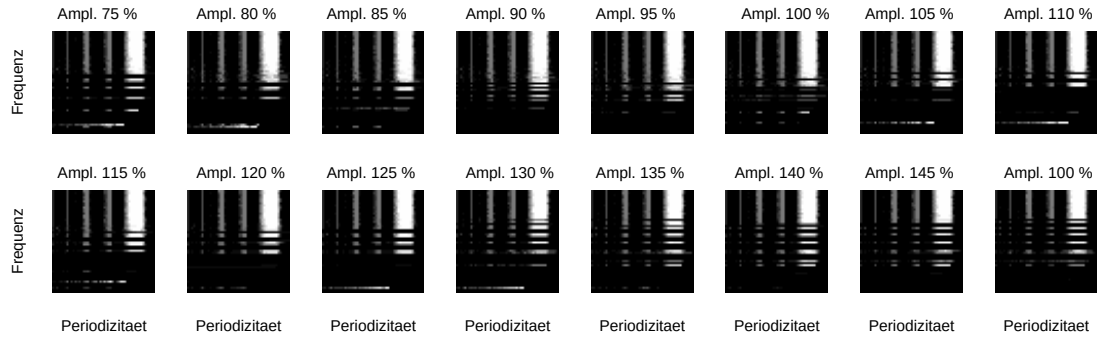

Abbildung 3.17: Dieselben Ergebnisse wie in Abb. 3.16, nur das der Integrator einen konstanten Input erhält.

3.2 Ergebnisse zum variierten Langnermodell

Analyse eines einzelnen Schaltkreises

Zuerst werden wie bei dem ursprünglichen Langnermodell wieder die Parameter Trägerfrequenz, Modulationsfrequenz und Lautstärke für den einzelnen Schaltkreis des variierten Langnermodells bestimmt.

Bestimmung der Frequenz, auf die das Triggerneuron eingestellt ist

Wie oben wird die Frequenz bestimmt, auf die das Triggerneuron eine Antwort zeigt. In Abb. 3.18 sind die Ergebnisse dargestellt. In der linken Graphik ist der gesamte Frequenzbereich abgebildet, in der rechten der Ausschnitt von 1700 bis 2100 Hz. Wie schon beim ursprünglichen Langner Schaltkreis ist der Mittelpunkt des Antwortbereichs bei 1900 Hz. Diese Frequenz wird für die folgenden Untersuchungen als Trägerfrequenz verwendet.

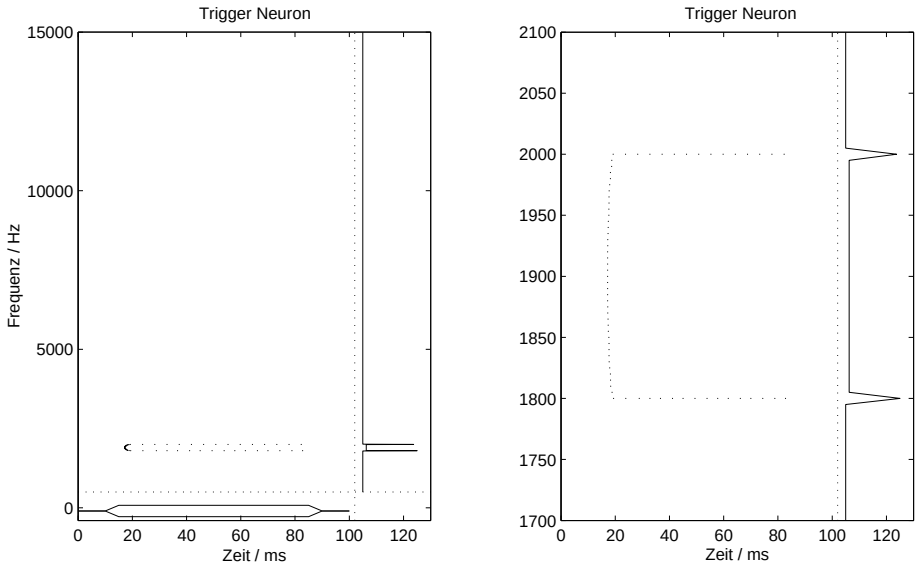

Abbildung 3.18: Antwort des einzelnen Schaltkreises des variierten Langnermodells. Der Stimulus ist ein Sinus, dessen Frequenz von 500 Hz bis 15 kHz in 5 Hz Schritten variiert wurde. Es zeigen sich Spikes in der Region um 1900 Hz (rechter Graph zeigt die Vergrößerung in diesem Bereich).

Bestimmung der Periodizität, auf die der gesamte Schaltkreis eingestellt ist

Um die Periodizität festzustellen wird ein AM Sinus mit 1900 Hz Trägerfrequenz verwendet, da der Schaltkreis auf diese Trägerfrequenz eingestellt ist. Die Periodizität wird von 30 Hz bis 230 Hz in 1 Hz Schritten variiert. Das Ergebnis ist in Abb. 3.19 dargestellt. Die Antwort liegt bei ca. 50 Hz. Im Gegensatz zum ursprünglichen Modell zeigen sich beim variierten Modell keine Kammfilterantworten, und die Spikes werden zu jeder Periodizität des Stimulus erzeugt.

Dennoch können Kammfilterantworten erzeugt werden, wie die nächsten Untersuchungen zeigen. Prinzipiell ist es so, daß jedes System, welches Spikes von einem Triggerneuron auswertet, Kammfilterantworten erzeugt, wenn das Triggerneuron nicht absolut korrekt arbeitet. Immer wenn ein Spike fehlt, wird die halbe Modulationsfrequenz gemessen, so daß ein System, welches z.B. auf 50 Hz eingestellt ist, sporadisch auch auf 100 Hz antwortet, sobald einzelne Spikes des Triggerneurons fehlen.

Bestimmung der geeignetsten Lautstärke

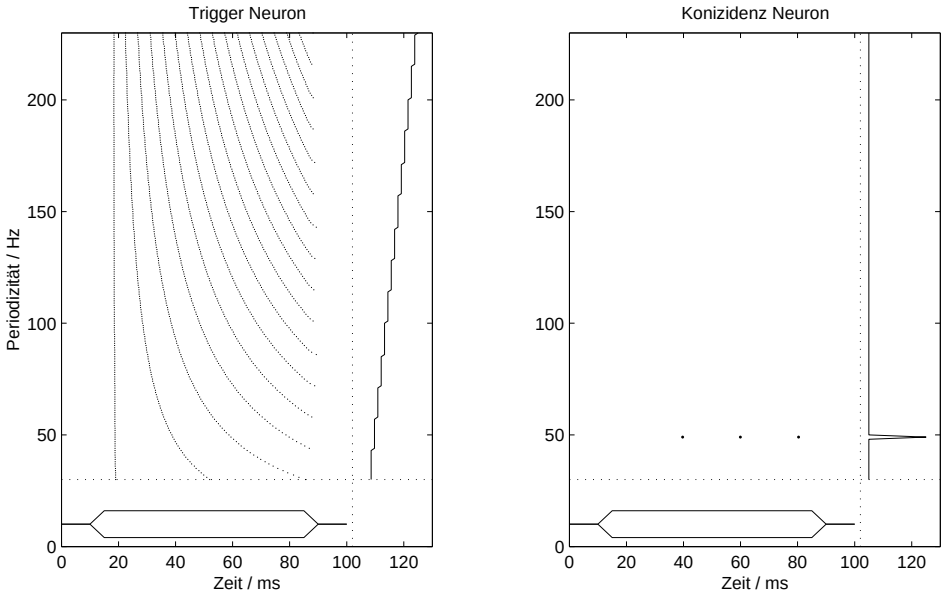

Abbildung 3.19: Antwort des einzelnen Schaltkreises (Triggerneuron/Koinzidenzneuron) des variierten Langnermodells. Als Stimulus wird ein AM Signal mit 1900 Hz Trägerfrequenz und einer Modulationsfrequenz von 30 Hz bis 230 Hz (Schrittweite 1 Hz) verwendet. Die drei Antworten des Koinzidenzneurons sind dicker dargestellt.

Nachdem die jeweils beste Trägerfrequenz und Modulationsfrequenz festgestellt wurde, wird das Verhalten des Schaltkreises bei verschiedenen Lautstärken untersucht. Dazu werden, wie für den unmodifizierten Langner Schaltkreis, die Antworten des Schaltkreises auf verschiedene Lautstärken gemessen. Die Ergebnisse sind in Abb. 3.20 dargestellt. Der Aufbau der Graphik ist wie der von Abb. 3.4 auf Seite 29. Man sieht, daß der Schaltkreis in einem Bereich von 25 % bis ca. 78 % Antworten erzeugt. Allerdings ergeben sich bei manchen Lautstärken Kammfilterantworten (z.B. bei einer Lautstärke von 62,5 %).

Dieses Verhalten wird klar, wenn - wie bei den Untersuchungen des ursprünglichen Langnerschaltkreises - die Triggerneuronantwort betrachtet wird (siehe Abb. 3.21). Der Aufbau der Abbildung ist wie der des ursprünglichen Langner Schaltkreises (Abb. 3.5 auf Seite 29). Man sieht, daß sich das Triggerneuron sehr ähnlich zum Triggerneuron des ursprünglichen Schaltkreises verhält. Die Antwort des Koinzidenzneurons dagegen sieht deutlich anders aus, da das variierte Modell so verändert wurde, daß keine Kammfilterantworten mehr erzeugt werden. Warum diese für manche Lautstärken dennoch auftreten können, wird am deutlichsten bei der Überlagerung (drittes Bild von links) klar: In dem Grenzbereich, in dem das Triggerneuron nicht mehr korrekt arbeitet, aber noch nicht komplett aufhört, Spikes zu erzeugen, kann es zum Beispiel passieren, daß das Triggerneuron nur jeden zweiten Spike erzeugt. Dies würde den restlichen Schaltkreis zu der Annahme verleiten, daß die halbe Periodizität als Stimulus anliegt, und daher erzeugt das Koinzidenzneuron entsprechend Spikes. Daher liegen die kammfilterantwortähnlichen Spikes an der Grenze, wo das Triggerneuron nicht mehr korrekt arbeitet. Dies wird auch im rechten Plot von Abb. 3.21 deutlich, die Kammfilterantwort des Koinzidenzneurons liegt genau an der Stelle, wo die Triggerneuronantworten einbrechen.

Aus diesen Gründen ist der gleiche Wert (37 %) für die Lautstärke am besten geeignet, wie für den ursprünglichen Langnerschaltkreis.

Die X-Achse ist wieder die Periodizität, die Y-Achse allerdings die Lautstärke. Diese wurde für die Untersuchung wieder von 0-1,75 verändert, allerdings mit einer Schrittweite von 0,025. Dadurch ergeben sich die 71 Zeilen der Grafik in der Darstellung. Die Anzahl Spikes sind durch Grauwerte

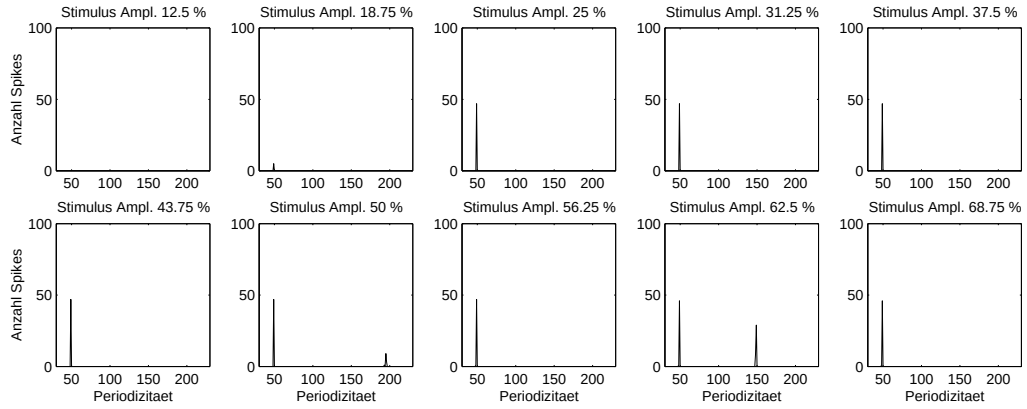

Abbildung 3.20: Ergebnis der Untersuchung um die geeignete Lautstärke festzustellen. Zu jeder Lautstärke wurde ein Graph berechnet, der auf der x-Achse die Periodizität und auf der y-Achse die Anzahl Spikes zeigt

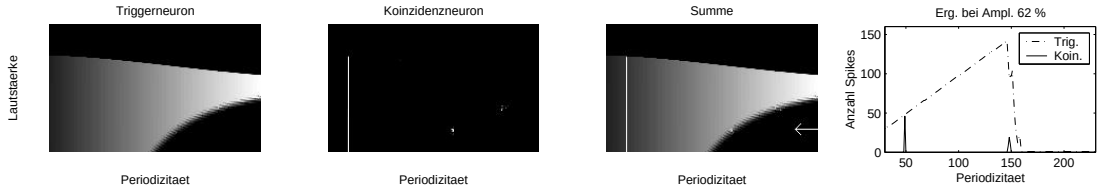

Abbildung 3.21: Darstellung der genaueren Analyse des Verhaltens des Triggerneurons in Bezug auf die Lautstärke, selber Aufbau der Graphik wie Abb. 3.5 auf Seite 29. Aufgetragen sind von links nach rechts die Periodizitäten von 30 Hz bis 230 Hz, von oben nach unten die Lautstärken von 12,5 % bis 68,75 %. Der Pfeil im dritten Graph zeigt die Zeile, die für den Plot rechts verwendet wurde.

dargestellt. Man sieht, daß bei zunehmender Periodizität (von links nach rechts) die Grauwerte heller werden, da das Triggerneuron entsprechend öfters Spikes generiert. Der Bereich, in dem das Triggerneuron über die gesamte Periodizität funktioniert, erstreckt sich allerdings nicht über die gesamte Lautstärke, sondern über eine Art dreieckigen Bereich. Bei zunehmender Lautstärke funktioniert das Triggerneuron nur noch bei niedrigen Periodizitäten. Im Grenzbereich ergeben sich Aussetzer, das Triggerneuron erzeugt nicht mehr zu jeder Periodizität einen Spike. Dadurch ergeben sich in diesem Bereich Kammfilterantworten an den Vielfachen von 49 Hz.

3.2.1 Das Verhalten der einzelnen Neuronen auf verschiedene Stimuli

Nachdem die für den Schaltkreis geeigneten Parameter festgestellt wurden, werden im folgenden die Antworten aller Neuronen auf unterschiedliche Stimuli dargestellt.

Sinus Signal

In Abb. 3.22 sind die Antworten der Neuronen auf einen reinen Sinus dargestellt. Es zeigen sich keine größeren Aktivitätsmuster, in dem Bereich von 1900 Hz ergibt sich etwas Aktivität durch die Antworten des Triggerneurons auf den Onset.

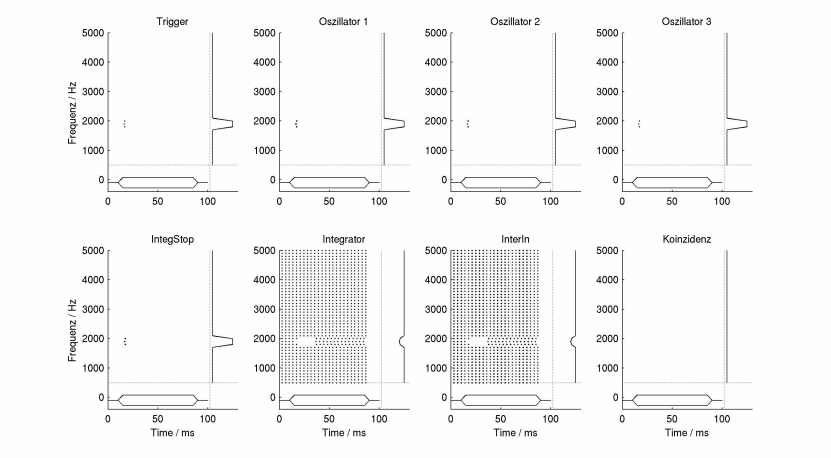

Abbildung 3.22: Die Antworten aller Neuronen des variierten Langner Schaltkreises auf Sinus Signale.

AM Signal

Die Ergebnisse, die sich aus einem AM Stimulus ergeben sind in Abb. 3.23 dargestellt. Im Gegensatz zum ursprünglichen Langner Schaltkreis zeigt sich, daß der Integrator nur Spikes erzeugt, wenn die Periodizität des Stimulus kleiner als die Periodizität ist, auf die der Schaltkreis eingestellt ist. Durch das IntegStopneuron wird der Integrator bei größeren Frequenzen zurückgesetzt, bevor er seinen Schwellwert erreicht und feuern kann. Dadurch wird dafür gesorgt, daß keine Kammfilterantworten erzeugt werden können - zumindest solange das Triggerneuron korrekt arbeitet.

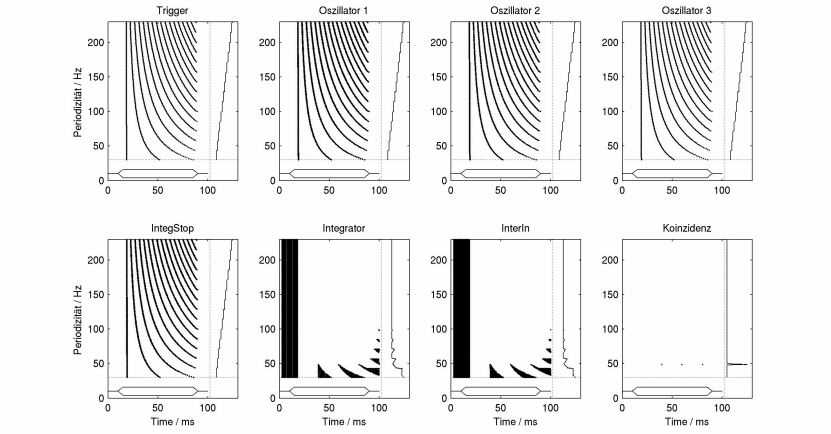

Abbildung 3.23: Die Antworten aller Neuronen des variierten Langner Schaltkreises auf AM Signale.

Weiterhin wird deutlich, daß das Koinzidenzneuron auf jede Periode antwortet. Der ursprüngliche Langner Schaltkreis erzeugt nur bei jeder zweiten Periode Spikes. Für eine möglichst genaue Repräsentation des Stimulus ist aber ein Verfahren, welches zu jeder Periode Spikes erzeugt, besser geeignet. In Abb. 3.24 ist der Bereich der Antwort des Koinzidenzneurons vergrößert dargestellt (linker Graph). Im Gegensatz zum ursprünglichen Schaltkreis, der eine Art 'Phasensprung' im Antwortverhalten erzeugt, wenn man ihn auf jede Periode einstellt, zeigt sich für den variier-

ten Schaltkreis dieses Verhalten nicht. Die erste Spikereihe bei niedrigerer Periodizität entsteht aufgrund der Rampe des Onsets.

Im rechten Graphen ist die Anzahl Spikes der Koinzidenzneuronen über die Periodizität aufgetragen. Drei verschiedene Schaltkreise wurden dazu verwendet:

- Der variierte Schaltkreis (durchgezogene Linie).
- Der ursprüngliche Schaltkreis (gestrichelte Linie).
- der ursprüngliche Schaltkreis, so verändert, daß das Flipneuron 'gewinnt' (vgl. Kapitel 3.1.2).

In diesem Graphen wird deutlich, daß der ursprüngliche Schaltkreis nur halb so oft Spikes erzeugt, wie der variierte. Das Plateau liegt bei ca. 25 Spikes für den ursprünglichen, wohingegen das Plateau des variierten Schaltkreises bei ca. 47 Spikes liegt. Der ursprüngliche Schaltkreis, bei dem das Flipneuron gewinnt hat das Plateau bei ca. 48 Spikes. Dafür zeigt sich bei ca. 52 Hz noch ein kleines Plateau bei 25 Spikes (in diesem Bereich gewinnt das Flopneuron).

Abgesehen von der Höhe des Plateaus ist die Form der Kurve sehr unterschiedlich. Während das variierte Modell eher nach dem Prinzip ganz oder gar nicht antwortet zeigt das ursprüngliche Modell eine breitere Kurve, die auch nicht so steil ansteigt bzw. abfällt. Dies gilt sowohl für den Fall, das das Flopneuron sich durchsetzt als auch das das Flipneuron sich durchsetzt. Für beide gilt auch, daß die Anzahl Spikes nicht monoton abfällt, sondern auch mehrere Nebenmaxima existieren. Im Abschnitt 4.1.2 werden diese Aspekte noch genauer diskutiert.

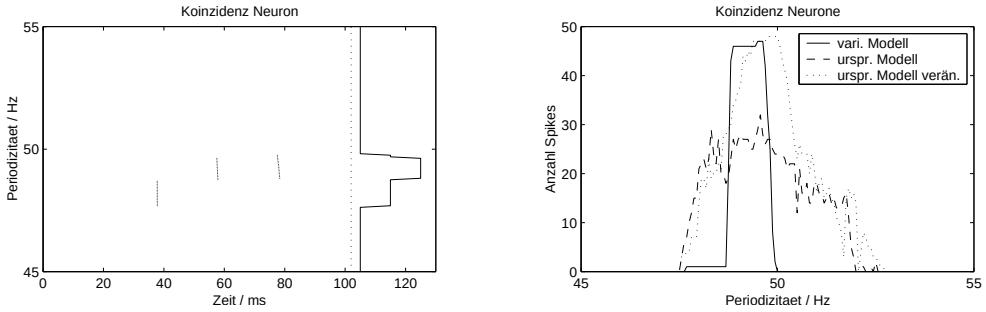

Abbildung 3.24: Linker Graph: Das Antwortverhalten des Koinzidenzneurons des variierten Schaltkreises, vergrößert für den Bereich von 45 Hz bis 55 Hz Periodizität. Man sieht, daß bei jeder Periode ein Spike erzeugt wird - die Line bei 45 ms entsteht durch die Rampe, mit der der AM-Stimulus angeschaltet wird, und kann ignoriert werden. Rechter Graph: Die Anzahl Spikes der Koinzidenzneuronen des variierten (durchgezogene Linie), des ursprünglichen (gestrichelte Linie) und des ursprünglichen mit veränderten Parametern ('Flip gewinnt', gepunktete Linie), für Periodizitäten zwischen 45 und 55 Hz. Im Gegensatz zum rechten Graph hatte der Stimulus eine Dauer von 1000 ms, so daß 50 Perioden im Zeitraum enthalten sind.

3.2.2 Ergebnisse der Simulationen von IC Karten

Wie auch für das ursprüngliche Langnermodell werden im folgendem IC Karten simuliert. Dazu wird der variierte Langner Schaltkreis 2500 mal verwendet, um eine 50x50 Karte entsprechend der Karte im Colliculus Inferior zu erzeugen. Das daraus resultierende Netzwerk wird im folgendem für verschiedene Simulationen verwendet.

Verwendung künstlicher Stimuli

Die Ergebnisse für das variierte Langnermodell für AM Stimuli sind in Abbildung 3.25 dargestellt. Die Abbildung ist entsprechend der Abbildung für das ursprüngliche Langnermodell auf Seite 31

aufgebaut, es wurden vier verschiedene Stimuli verwendet. Man sieht, daß im Gegensatz zu den Karten auf Seite 31 nur einzelne Punkte erzeugt werden, die insbesondere keine Kammfilterantworten aufweisen. Dies entspricht den Erwartungen im Vergleich zu den Ergebnissen des einzelnen variierten Schaltkreises, da das Koinzidenzneuron keine Kammfilterantworten erzeugt.

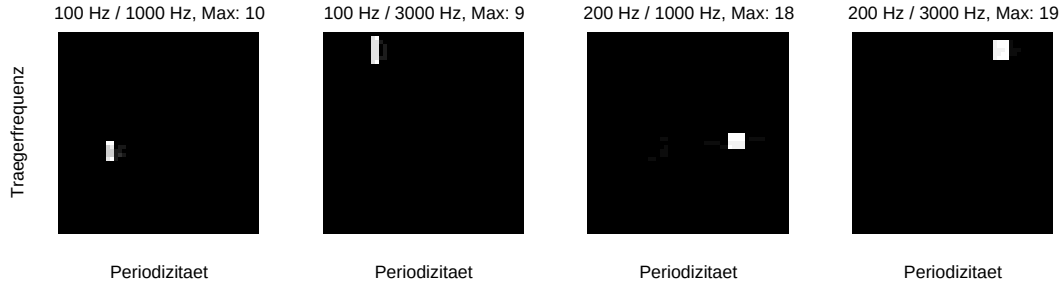

Abbildung 3.25: Die Ergebnisse der IC Karten Berechnung, wenn AM Stimuli verwendet werden. Man sieht, daß im Gegensatz zum ursprünglichen Langnermodell keine Kammfilterantworten mehr auftreten.

Anpassung der Lautstärke der Filterantworten

Für die folgenden Versuche werden gesprochene Vokale verwendet. Von jedem Vokal wurden 4 Exemplare bei unterschiedlicher Tonhöhe aufgenommen. Die Tonhöhe erstreckt sich in einem Bereich von ca. 100 Hz bis ca. 250 Hz. Die Dauer der Vokale ist 0.5 Sekunden, jeder Vokal besteht also aus 20000 Samples. Die Vokale wurden auf ihren RMS normalisiert, so daß alle Vokale die gleiche Energie haben.

Bevor die Vokale als Stimulus in das neuronale Netzwerk eingespeist werden, muß getestet werden, wie unterschiedlich deren Amplitude in den verschiedenen Frequenzkanälen ist. Wie oben schon gezeigt, reagiert das Langnermodell empfindlich auf unterschiedliche Lautstärken, so daß für realistische Ergebnisse die Lautstärke der Filterkanäle ungefähr gleich groß sein sollte. Um dies zu testen, wurden für jeden Vokal die 50 Filterkanäle $F_n(t)$ (mit $t = 0 - 0.5s$ und $n = 1..50$) berechnet, und dann diese betragsmäßig aufsummiert: $B_n = \sum_{t=0}^{0.5s} F_n(t)$. Die Funktion B_n wurde für jeden Vokal berechnet, und dann alle Funktionen aufaddiert. Das Ergebnis ist eine Funktion, die den 50 Filterkanälen jeweils einen Wert zuordnet, nämlich die Summe aller Betragssummen der Filterkanalantworten zu den Vokalen. Diese ist in Abbildung 3.26 links dargestellt.

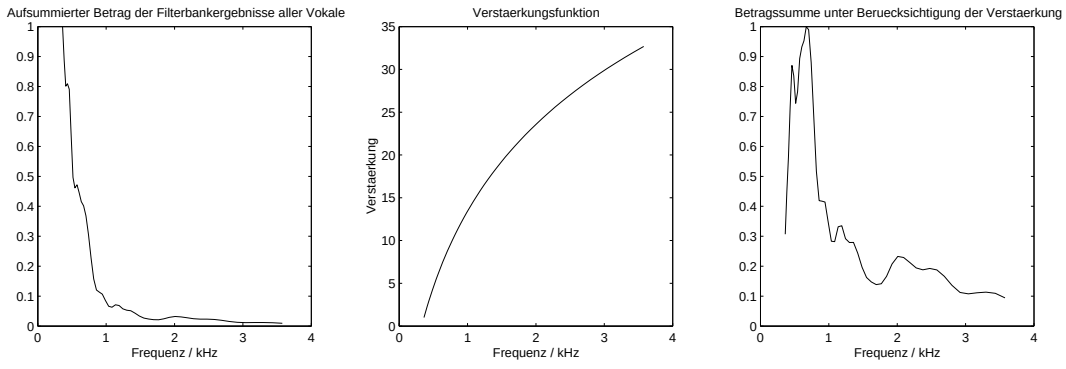

Abbildung 3.26: Links: Darstellung der Betragssumme der Filterkanäle über alle Vokale. Mitte: Die Verstärkerfunktion, die angibt, um welchen Faktor jeder Frequenzkanal verstärkt wird. Rechts: Die Betragssumme, die sich unter Anwendung der Verstärkerfunktion ergibt

Man erkennt, daß die oberen Frequenzanteile wesentlich leiser sind als die unteren Frequenzanteile. Daher wird jeder Filterkanal mit einem Verstärkungsfaktor multipliziert, der mit zunehmender Filterkanalfrequenz größer wird. Diese Verstärkungsfunktion ist in der Mitte der Abbildung 3.26 dargestellt. Wendet man diese Verstärkungsfunktion auf die Vokale an, so wird der linke Graph zum rechten Graph verändert. Rechts ist also nochmal die Summe der B_n Funktionen dargestellt, nur unter Berücksichtigung der Verstärkungsfunktion. Man sieht, daß immer noch Lautstärkeunterschiede zwischen niedrigen und hohen Frequenzen vorhanden sind. Dennoch sind die Unterschiede nicht mehr so extrem.

Vokale mit unterschiedlicher Periodizität

Mit dem leicht modifizierten Modell und den 20 Vokalen als Stimulus ergeben sich 20 IC Karten, die in Abbildung 3.27 dargestellt sind. Man erkennt, daß bei jeder der IC Karten die Aktivität auf einer Periodizitätslinie liegt. In den meisten Fällen zeigt sich der Fleck der Aktivität auch an der Stelle, wo die Formantfrequenzen [45] des Vokals liegen. Teilweise fehlt eine der beiden Formantfrequenzen. Vergleicht man die Ergebnisse z.B. aller [a:] Stimuli, dann zeigt sich, daß die Antwortmuster ähnlich zueinander sind, nur entsprechend der Periodizität auf der X-Achse verschoben.

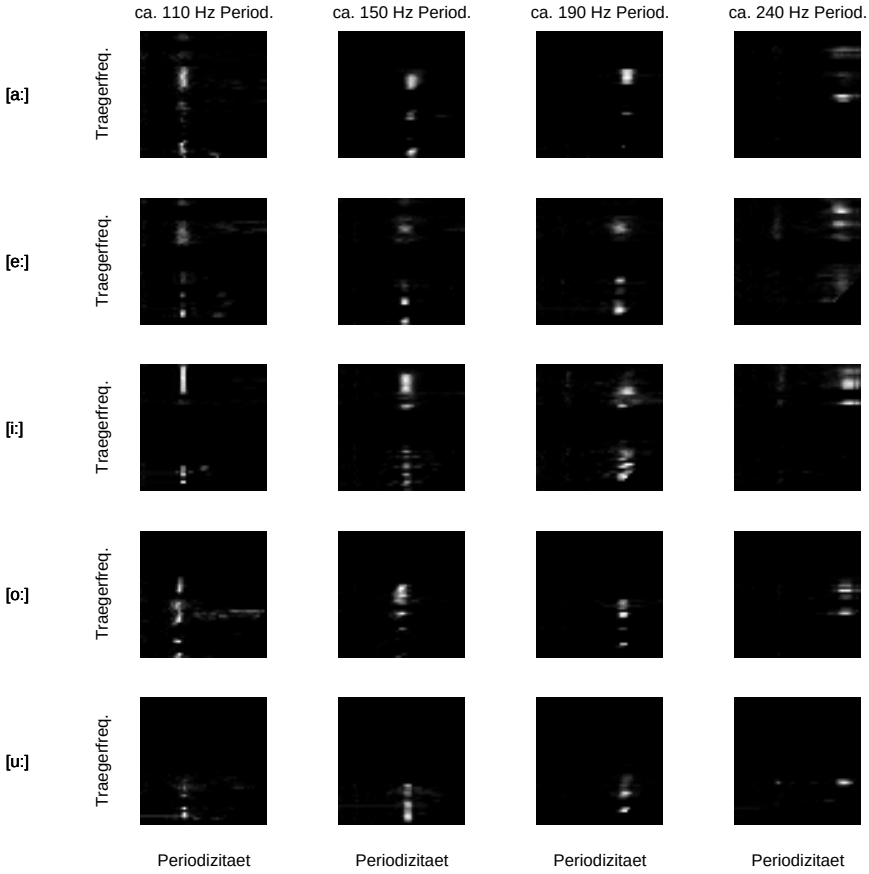

Abbildung 3.27: Die Ergebnisse, die sich mit dem variierten Modell bei Verwendung von Vokalen als Stimulus ergeben. Sowohl die Periodizitäten als auch die Formantfrequenzen sind in den Antwortmustern wieder zu erkennen.

Zur genaueren Analyse der IC Antwortkarten werden im folgenden noch weitere Berechnungen durchgeführt. In Abbildung 3.28 ist sowohl die Summe der IC Karten entlang der Frequenz als auch die Summe der IC Karten entlang der Periodizität dargestellt. Die oberen 5 Graphen zeigen zu jedem der 5 Vokale die Summe aller Ergebnisse der zugehörigen Vokale, so daß die X-Achse die Trägerfrequenz ist. Zum Beispiel wurde für den ersten Graph alle 4 IC Karten des [a:] Vokals aufsummiert, und dann die Summe über die Periodizität berechnet, also entlang der Frequenz.

Das Ergebnise ist dann zu jeder Trägerfrequenz die Anzahl Spikes, die insgesamt durch alle vier [a:] Vokale erzeugt wurden.

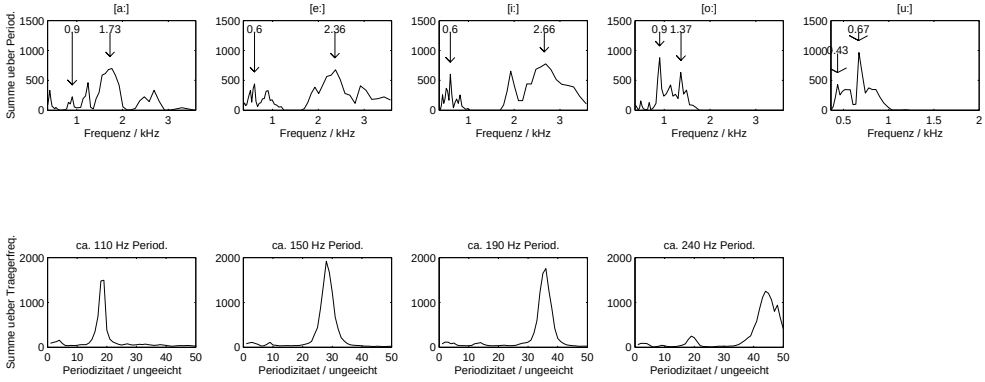

Abbildung 3.28: Ergebnis der Summation der IC-Karten aus Abb. 3.27 'nach rechts' (Zeile 1) und 'nach unten' (Zeile 2). Zusätzlich sind in der ersten Zeile die möglichen Formantfrequenzen markiert.

Entsprechend ist die zweite Zeile ähnlich aufgebaut, allerdings sind für diese vier Graphen nicht die Periodizität sondern die Frequenz summiert, und jeweils die Vokale mit der gleichen Periodizität zusammen aufaddiert. Der Graph links unten ordnet also jeder Periodizität (x-Achse) die Anzahl Spikes zu, die in allen fünf Vokalen mit Periodizität 110 Hz aufgetreten sind.

In der zweiten Zeile der Abb. 3.28 erkennt man, daß die Vokale mit gleicher Periodizität auch Aktivitätsflecken mit der gleichen x-Position auf der Karte erzeugen. Die Graphen haben ein deutliches Maxima, welches aus der Summe aller Vokale mti gleicher Periodizität entsteht. In dem Graphen für 240 Hz Periodizität zeigt sich zusätzliche eine kleine Kammfilterantwort. Anhand der x-Position der Maxima kann auf einfache Weise die Periodizität festgestellt werden - wenn die x-Achse entsprechend ausgemessen werden würde. Ein neuronales Netz in dieser Bauart hätte also die Tonhöhe des Sprechers durch Kartierung extrahiert.

Dagegen sieht man in der ersten Zeile die Formantfrequenzen der Vokale. Die angegebenen Maxima sind ungefähr an der Stelle, wo die Formantfrequenzen der Vokale liegen, das Modell erzeugt also Aktivitätsmuster, die zur Unterscheidung der Stimuli geeignet sein sollten.

Effekte durch Veränderung der Lautstärke

Um den Einfluß der Lautstärke auf die Ergebnisse der IC-Karten darzustellen ist in Abbildung 3.29 dargestellt, welche Veränderungen sich ergeben, wenn die Lautstärke des Stimulus von 67% in sechs Schritten auf 150% erhöht werden. Es ist jeweils die IC Karte für das [a:] mit Periodizität 110 Hz dargestellt.

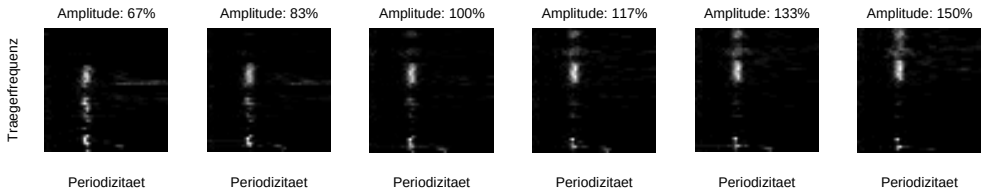

Abbildung 3.29: Einfluß der Lautstärke auf die IC Karten. Als Stimulus wurde jeweils das gleiche [a:] verwendet. Es zeigt sich, daß die Periodizität unabhängig von der Lautstärke ist, die Formantfrequenzen dagegen sich verändern.

Man sieht, daß die unterschiedliche Lautstärke keinen Einfluß auf die Periodizität hat. Dagegen verschieben sich die Aktivitätsflecken entlang der Periodizität auf der Trägerfrequenzachse. Dies ist einfach zu erklären, da jeder Langnerschaltkreis nur einen begrenzten Lautstärkebereich hat. Ist die Lautstärke zu klein ergibt sich ebenso keine Aktivität, wie wenn die Lautstärke zu groß ist. Die Auswirkungen dieses Effektes werden später in Abschnitt 3.5.1 systematisch untersucht.

Ergebnisse zur Synchronität unterschiedlicher Frequenzkanäle

Wie in Abschnitt 3.1.3 auf Seite 33 dargestellt sind die Spikes in unterschiedlichen Frequenzkanälen der Filterbank zeitlich synchronisiert. Dieser Effekt wird auch durch die variierte Version des Langnermodells extrahiert. Um dies zu zeigen, werden zwei Ergebnisse vorgestellt:

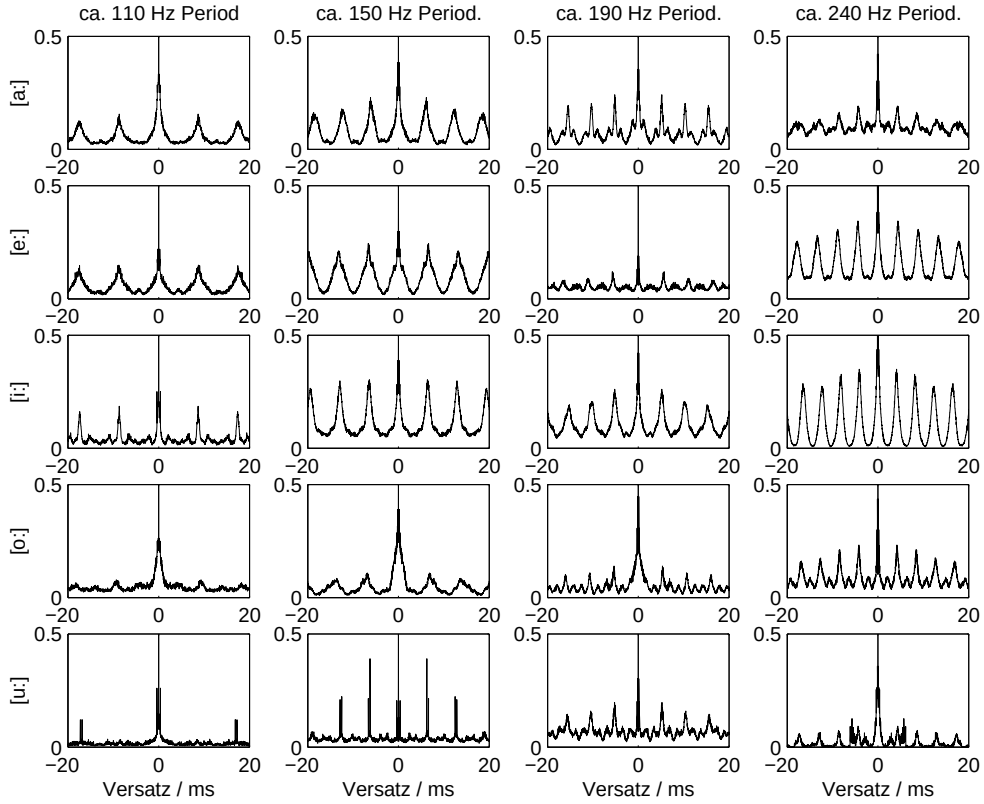

Abbildung 3.30: Ergebnisse der Autokorellogramme von $S(t)$. Die Periodizitätsangaben geben nur ungefähr die Periodizität der Vokale wieder, da diese von Vokal zu Vokal schwankt. Die Korellogramme wurden auf 1 normiert, durch die Skalierung der Y-Achse ist die obere Hälfte des zentralen Maximums außerhalb des Graphen.

- Für jeden Vokal wird die IC-Karte berechnet. Die Spikes werden räumlich aufsummiert, so daß sich eine Funktion $S(t)$ ergibt, welche zu einem Zeitpunkt t angibt, wieviele Neuronen einen Spike erzeugt haben.

  Von dieser Funktion wird die Autokorrelation berechnet. Falls die Phasenlage in den verschiedenen Frequenzkanälen nicht ähnlich ist, so ergibt sich ein flaches Autokorellogramm, da die Spikes zu jeder Phasenlage auftreten können. Sind die Phasenlagen der Frequenzkanäle dagegen sehr ähnlich, so treten die Spikes immer etwa zur gleichen Zeit auf. Dies ergibt im Autokorellogramm deutliche Nebenmaxima.

Die Ergebnisse sind in Abb. 3.30 dargestellt. Man sieht, daß bei fast jedem Vokal deutliche Nebenmaxima vorhanden sind, also die Phasenlagen der Frequenzkanäle synchron sind.

- Um deutlich zu machen, daß die Autokorellogramme ein geeignetes Maß für die Phasenlage der Frequenzkanäle sind, werden von einem Vokal ([a:]) künstlich die Phasen der Frequenzkanäle gegeneinander verschoben. Dazu wird das Zeitsignal des Vokals gefiltert, und zwar so daß die Filterbank keine Phasenverschiebung erzeugt. Dann werden die Filterkanäle zeitlich gegeneinander verschoben (um maximal eine Periode der Periodizität) und wieder aufsummiert.

  Das Ergebnis ist wieder ein [a:], der subjektive Eindruck zeigt keine wesentlichen Unterschiede zwischen dem Original [a:] und dem [a:] mit phasenverschobenen Frequenzkanälen. Für beide Vokale werden die IC Karten berechnet, die Ergebnisse sind in Abb. 3.31 dargestellt. Die obere Zeile zeigt die Ergebnisse, wenn das original [a:] verwendet wird, die untere die Ergebnisse, wenn der modifizierte Vokal verwendet wird. Man sieht, daß die Spikes (zweite Spalte von links) eher ein Blockmuster zeigen, wenn die Phasenlage nicht modifiziert wird. Dagegen ist aus der IC-Karte nicht direkt ersichtlich, das die Phasenlage modifiziert wurde. Dies ist auch zu erwarten, da die Phasenlage nur in einem Zeitsignal deutlich wird. Beide IC Karten zeigen Aktivitäten, die den Formantfrequenzen des verwendeten Vokals ([a:]) entsprechen.

  Dagegen sind die Unterschiede der Autokorellogramme sehr deutlich, die Nebenmaxima des oberen Korellogramms sind deutlich ausgeprägt, wohingegen das untere Korellogramm keine Nebenmaxima zeigt. Die Eigenschaft 'ähnliche Phasenlage' wird also durch die Autokorellogramme der zeitlichen Spikeantwort gut extrahiert.

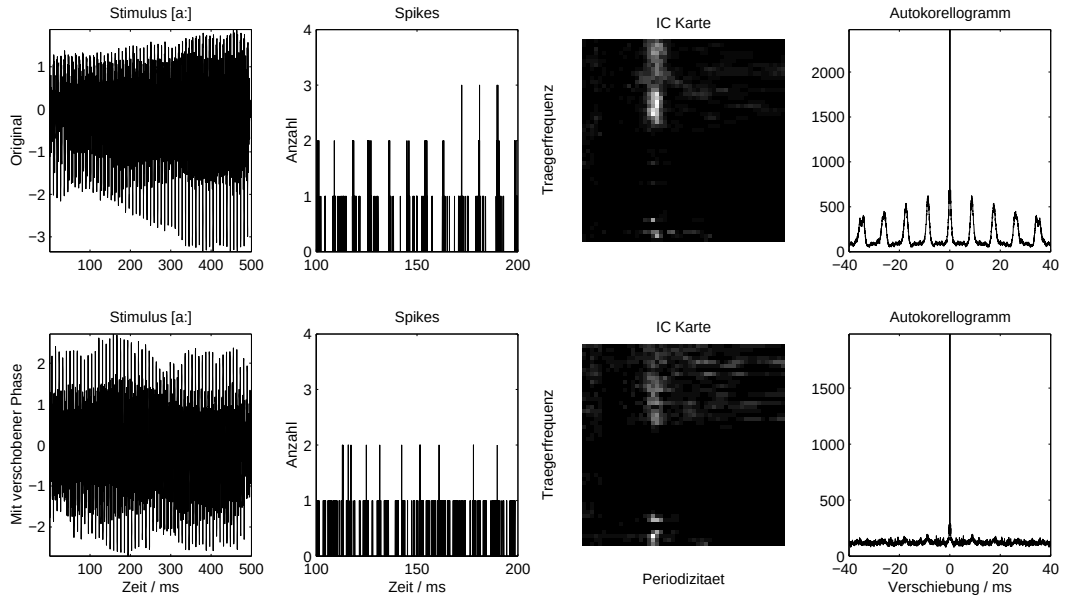

Abbildung 3.31: Vergleich des durch Phasenverschiebung der Frequenzkanäle modifizierten Vokals (untere Zeile) mit dem Originalvokal (obere Zeile). Es wurde jeweils das Zeitsignal, ein Ausschnitt der Spikefunktion $S(t)$, die IC-Summenkarte und das Autokorellogramm von $S(t)$ dargestellt.

Problematik der nichtlinearen Überlagerung

Für die einfache Trennung zweier akustischer Stimuli wäre ein Verhalten wünschenswert, welches als *lineare Überlagerung* bezeichnet wird. Damit ist gemeint, daß ein Stimulus, der aus der

Überlagerung zweier akustischer Stimuli (z.B. [a:] und [o:]) besteht, eine IC Karte, die aus der Überlagerung der IC Karten der einzelnen Stimuli besteht, erzeugt. In Abb. 3.32 ist beispielhaft das Ergebnis dargestellt, wenn man erst ein [a:], dann ein [o:] und schließlich die Überlagerung aus beiden ([a:]+[o:]) als Stimulus verwendet. Zusätzlich ist die Addition der IC Karten für das [a:] und das [o:] dargestellt. Man erkennt, daß die beiden Karten sich stark unterscheiden, also die Überlagerung nicht linear ist.

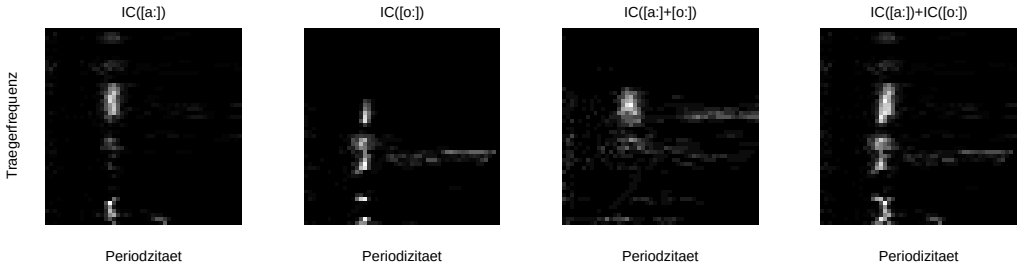

Abbildung 3.32: Darstellung des Ergebnisses der Überlagerung zweier Stimuli. Die beiden Bilder links zeigen die jeweiligen Einzelergebnisse, das zweite Bild von rechts das Ergebnis der Überlagerung vor der Simulation (Addition der Stimuli) und das rechte Bild das Ergebnis der Überlagerung nach der Simulation (Addition der IC Karten). Die beiden rechten Bilder müßten gleich aussehen, wenn die Überlagerung linear wäre.

Daß die Transformation von Stimulus zu IC Karte nur eine nicht lineare Funktion sein kann, zeigt sich durch das folgende einfache Experiment:

- Stimulus 1: Klicks mit 100 Hz.
- Stimulus 2: Klicks mit 120 Hz.
- Berechnung der vier Karten wie in 3.32.
- Berechnung eines 'InterKlickIntervall'-Histogramms für die drei Stimuli, also ein Histogramm, welches zu verschiedenen Abständen von Klicks angibt, wie oft dieser Abstand vorkommt.

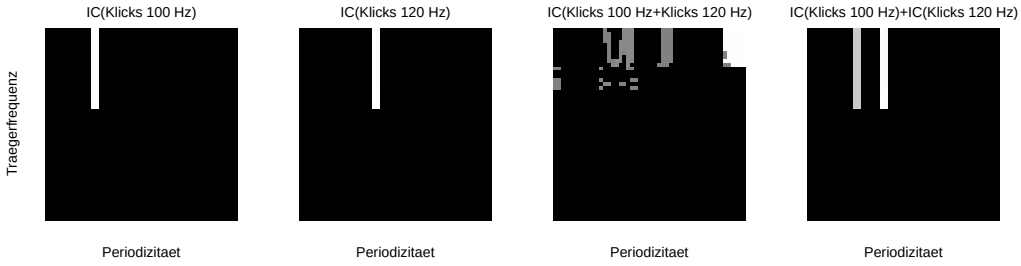

Abbildung 3.33: Darstellung wie in Abb. 3.32, nur wurden Klicks mit einer Frequenz von 100 Hz (10 ms) und 125 Hz (8 ms) verwendet.)

Bei einer linearen Überlagerung müßte das InterKlickIntervall-Histogramm der Stimulussumme wieder das selbe ergeben wie die Summe der InterKlickIntervall-Histogramme der einzelnen Stimuli. Die Histogramme der einzelnen Stimuli sind einfach, eine Frequenz von 100 Hz entspricht einem Abstand von 10 ms, eine Frequenz von 120 Hz von ca. 8 ms. Die Histogramme haben also ein Maximum bei 10 ms (100 Hz Stimulus) und 8 ms (125 Hz Stimulus).

Das Summensignal dagegen hat den ersten Klick bei 8 ms, den zweiten bei 10 ms, den dritten bei 16 ms, den vierten bei 20, ... Daraus ergeben sich Einträge im Histogramm bei 8 ms, 2 ms, 6 ms, 4 ms, ... Das bedeutet, daß das Summensignal nicht einfach die Periodizitäten von 100 Hz und 125 Hz enthält, sondern vielmehr die sich daraus ergebenen Überlagerungen, also 500 Hz, 250 Hz, 166 Hz. Diese Werte hängen jeweils von dem Verhältnis der beiden Stimulusfrequenzen ab.

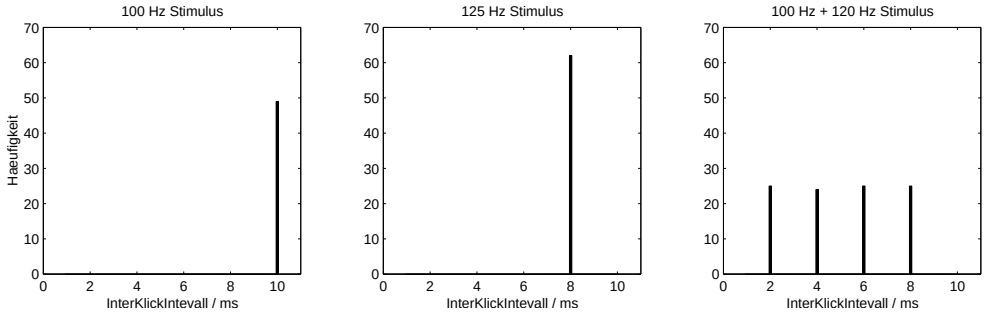

Abbildung 3.34: Darstellung der 'InterKlickIntervalle' für die Stimuli von 100 Hz, 120 Hz und der Überlagerung beider Stimuli.

Das Langnermodell kann aber nicht 'wissen', daß der Stimulus aus der Überlagerung zweier einzelner Stimuli mit 100 Hz und 125 Hz Periodizität besteht, und erzeugt als Antwort dann die sich durch die Überlagerung ergebenen Periodizitäten - in der Abbildung sieht man die bei 125 Hz, 166 Hz und 250 Hz. Zusätzlich sind noch Aktivitäten bei 100 Hz vorhanden. Das nicht lineare Verhalten ist daher nicht ein ungewollter Effekt, sondern spiegelt durchaus korrekt den Stimulus wieder. Auf diese Punkte wird noch näher in Abschnitt 4.1.4 auf Seite 91 in der Diskussion eingegangen.

3.3 Ergebnisse zur Kammfilterdiskussion

Einige biologische Messungen deuten auf die Existenz von Kammfilterantworten bei AM Stimuli hin. Allerdings scheinen diese Kammfilterantworten nicht generell vorhanden zu sein. Eine Theorie geht davon aus, daß sie nur zum Onset eines Stimulus auftreten [29]. Ausgehend von dem verwendeten Modell wird im folgenden zuerst dargestellt, unter welchen Bedingungen das variierte Modell Kammfilterantworten erzeugen kann.

3.3.1 Ergebnisse der Simulationen

Das ursprüngliche Langnermodell erzeugt prinzipiell Kammfilterantworten (vgl. zum Beispiel Abb. 3.4 auf Seite 29). Dagegen erzeugt das variierte Langnermodell im Normalfall keine Kammfilterantworten. Im folgenden werden drei unterschiedliche Methoden vorgestellt, wie dennoch Kammfilterantworten erzeugt werden können. Das Prinzip ist immer dasselbe: Sobald das Triggerneuron nicht immer korrekt arbeitet - insbesondere wenn Spikes ausgelassen werden - ergibt sich automatisch eine Kammfilterantwort des Koinzidenzneurons. Das ist ein Problem, welches jeder beliebige Schaltkreis hat, der die Antworten des Triggerneurons analysiert. In folgenden Fällen kann es für das verwendete Triggerneuron auftreten:

1. Durch unangepaßte Lautstärke:

   Wie in Abschnitt 3.2 auf Seite 38 gezeigt wurde, hängt die Antwort des Koinzidenzneurons stark von der Lautstärke des Stimulus ab. Bei entsprechender Lautstärke können auch Kammfilterantworten auftreten, wenn das Triggerneuron bei einer Periodizität, die einem vielfachen der Tuningfrequenz des Schaltkreises entspricht, aufhört korrekt zu spiken (vgl. Abb. 3.21 auf Seite 40).

2. Durch verrauschte Stimuli:

   In Abb. 3.35 ist dargestellt, wie das Koinzidenzneuron und das Triggerneuron auf verrauschte Stimuli reagieren. Als Rauschen wurde normalverteiltes Rauschen auf das AM Signal addiert, mit einem SNR von ca. 0,32. Im linken Plot ist die Antwort des Koinzidenzneurons abgebildet, im rechten Plot die des Triggerneurons. Die Länge der Stimuli war 5 Sekunden, bei 50 Hz Periodizität wären also ca. 250 Koinzidenzneuronantworten möglich. In dem Plot sieht man, daß wesentlich weniger Spikes erzeugt wurden (maximal ca. 33 Spikes). Außerdem ist der Antwortbereich wesentlich breiter im Vergleich zu den Ergebnissen mit unverrauschten Stimuli. Bei den vielfachen von 50 Hz zeigen sich auch Antworten, allerdings deutlich weniger. Dennoch ist der Effekt gut zu sehen: Wenn das Triggerneuron nicht korrekt arbeitet, dann können Kammfilterantworten auftreten.

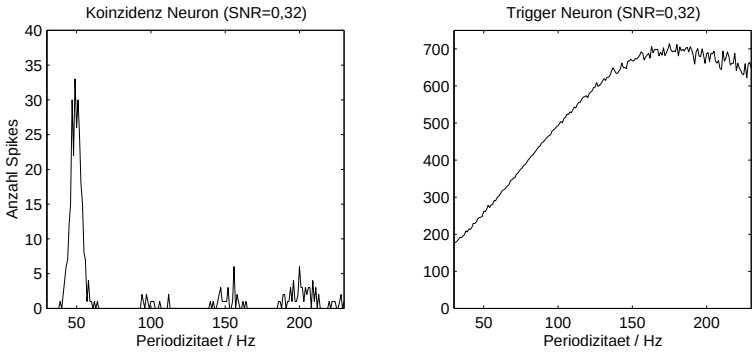

Abbildung 3.35: Die Ergebnisse, wenn das Inputsignal verrauscht wird (SNR ca. 0,32). Die Stimuluslänge war 5 Sekunden. Links das Koinzidenzneuron, rechts das Triggerneuron.

3. Durch Inhibition des Triggerneurons:

   Anstatt den Stimulus zu verrauschen kann auch das Triggerneuron durch konstante Inhibition dazu gebracht werden, nicht immer Spikes zu generieren. Dadurch fallen einfach einige Spikes weg, und es werden keine zusätzlichen falschen Spikes erzeugt. Es wurden zwei Versuche durchgeführt: wenig Inhibition und viel Inhibition. Die beiden linken Plots in Abb. 3.36 zeigen die Ergebnisse für das Koinzidenzneuron: Mit wenig Inhibition wurden kleine Kammfilterantworten erzeugt. Mit viel Inhibition dagegen zeigt sich, daß die eigentliche Antwort wegfällt, und nur noch große Kammfilterantworten erzeugt werden.

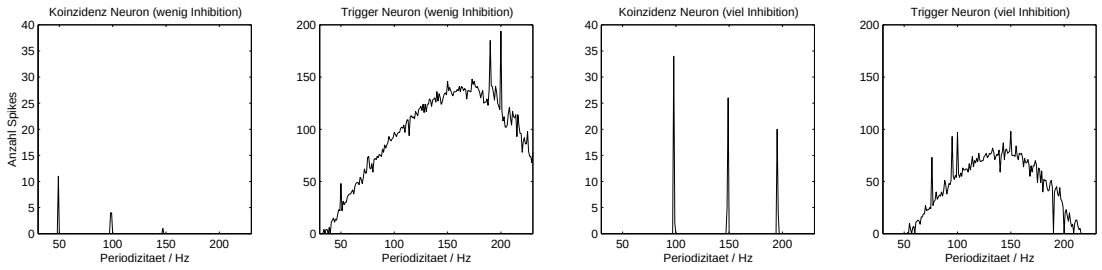

Abbildung 3.36: Die Ergebnisse, wenn das Triggerneuron mit konstanter Inhibition gehemmt wird. Die beiden linken Plots zeigen die Ergebnisse für wenig Inhibition, die beiden rechten für viel Inhibition. Die Stimuluslänge war eine Sekunde.

Es zeigt sich also, daß auch Schaltkreise, die eigentlich keine Kammfilterantworten erzeugen, in bestimmten Grenzfällen trotzdem Kammfilterantworten generieren.

Neuron	Vor Stimulus	Stimulusdauer	Nach Stimulus	Trägerfreq.	Modulationsfreq.
1	20 ms	100 ms	20 ms	2200 Hz	20-1000 Hz
2	50 ms	150 ms	50 ms	17500 Hz	50-800 Hz
3	30 ms	100 ms	20 ms	780 Hz	0-390 Hz
4	50 ms	100 ms	50 ms	1200 Hz	50-500 Hz
5	50 ms	150 ms	50 ms	7500 Hz	100-1000 Hz
6	50 ms	150 ms	50 ms	1800 Hz	25-900 Hz
7	50 ms	100 ms	50 ms	1200 Hz	50-500 Hz
8	50 ms	150 ms	50 ms	1800 Hz	25-900 Hz
9	50 ms	150 ms	50 ms	15000 Hz	0-1000 Hz

Tabelle 3.2: Die Stimulusparameter die für die Ableitungen von den Neuronen im IC verwendet wurden. Die Schrittweite der Modulationsfrequenzänderung war teilweise logarithmisch, teilweise linear. Die Spalte *Vor Stimulus* beinhaltet die Zeitdauer, in der vor dem Stimulus bereits gemessen wurde, die *Stimulusdauer* ist die Dauer des Stimulus, und entsprechend ist mit *Nach Stimulus* die Zeitdauer gemeint, die nach dem Stimulus noch weitergemessen wurde.

3.4 Analyse von Ableitungen an wachen Tieren

Im folgenden werden einige 'echte' Ableitungen von IC Neuronen analysiert und mit den Simulationsergebnissen verglichen.

3.4.1 Beschreibung der Daten

Die im folgendenden analysierten Daten stammen von Michael Ochse (Darmstadt, Forschungsgruppe von Herrn Langner) und wurden aus dem IC von wachen, also nicht narkotisierten Gerbils (Meriones unguiculatus, 3-6 Monate alt) abgeleitet. Als Stimulus wird ein AM Signal verwendet. Die Parameter der Messungen variieren von Datensatz zu Datensatz, die Dauer und die gemessenen Periodizitäten/Trägerfrequenzen sind in Tabelle 3.2 angegeben.

3.4.2 Standarddarstellung der Daten

Um einen ersten Eindruck der Ergebnissdaten zu bekommen, werden für jedes Neuron alle Spikes in ein Zeithistogramm eingetragen, unabhängig davon, welche Periodizität der Stimulus hatte. Das Ergebnis ist in Abb. 3.37 dargestellt. Die gestrichelten Linien zeigen den Onset bzw. das Ende des Stimulus an. Man erkennt deutlich die starke Onsetantwort der Neuronen auf den Stimulus.

Eine zweite Darstellung der Daten ergibt sich, wenn man anstatt der Zeit die Periodizität des Stimulus auf der X-Achse aufträgt. Dies ist in Abb. 3.38 dargestellt. Man erkennt, daß die Neuronen bei bestimmten Periodizitäten besonders viele Spikes erzeugen.

Zusätzlich zeigt sich bei manchen Neuronen (Nr. 1, 2, 5, 6 und 8), daß eine gewisse Tendenz zu Kammfilterantworten erkennbar ist. Generell ist diese allerdings nicht so stark ausgeprägt wie beim ursprünglichen Langner Schaltkreis.

3.4.3 Darstellung in Bezug auf die Latenz

In Abb. 3.39 ist noch einmal das Periodizitätsverhalten der 9 Neuronen dargestellt. Außerdem wurde die Position mit einer Linie markiert, die die geschätzte Präferenz des Neurons darstellt. Dieser Wert kann mit der Latenz (=Zeitdauer von Stimulus Onset bis zum ersten Neuronantwort bzw. zum Maximum der Spikeantwort) in einem Plot dargestellt werden (siehe Plot rechts unten). Es zeigt sich, daß sich die Latenz entlang der Periodizität systematisch verändert. Eine hohe Periodizität erzeugt eine niedrige Latenz und umgekehrt. Zum Beispiel ergibt sich bei einer Periodizität von 200 Hz eine Latenz von 6 ms.

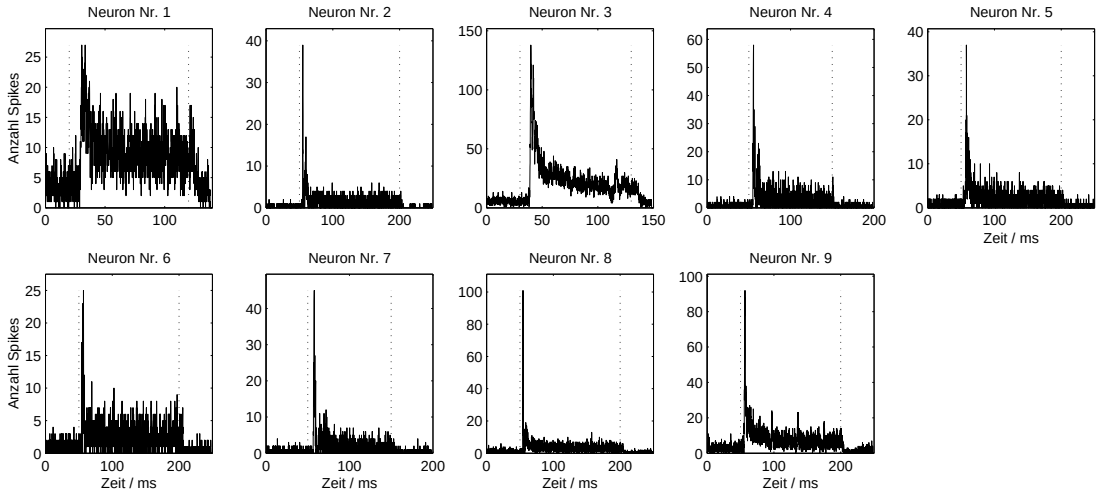

Abbildung 3.37: Die Antworten der Neuronen auf AM Reize. Die X-Achse zeigt die Zeit, die Y-Achse die Anzahl der Spikes. Je Neuron wurden unterschiedliche Periodizitäten gemessen, in dieser Darstellung ist dies nicht ersichtlich, da über die Periodizität aufsummiert wurde.

3.4.4 Darstellung der Daten über die Periodizität/Phase

Eine besonders interessante Darstellung der Daten ergibt sich durch folgende, im folgenden MoPha-Grafik genannte Darstellungstechnik:

1. Zuerst wird aus einem Zeithistogramm (vgl. Abb. 3.37) die Position der Onsetantwort festgestellt.
2. Für jede Periodizität wird nun getestet, in welcher Phasenlage die Spikes relativ zu der initialen Onsetantwort und der Periodizität des Stimulus erzeugt wurden. Dadurch ergeben sich zu jedem Spike s zwei Größen: die Periodizität des Stimulus $P(s)$ und die Phase $\phi(s)$ der Spikeposition.
3. Diese beiden Werte werden in ein 2D-Histogramm eingetragen, indem $P(s)$ als X-Wert und $\phi(s)$ als Y-Wert interpretiert werden.
4. Dieses 2D-Histogramm (MoPha Grafik) wird grauwertkodiert als Bild dargestellt.

Um die Zusammenhänge zwischen Spikepositionen, Periodizitäten und Phasen deutlicher zu machen sind in Abb. 3.40 einige künstliche Spikeantworten mit dem zugehörigen MoPha Bild dargestellt.

Nachdem das Prinzip der MoPha Grafiken deutlich gemacht wurde, werden im folgenden sowohl MoPha-Grafiken von den biologischen Daten (Abb. 3.41) als auch von den Simulationsergebnissen (Abb. 3.42) dargestellt.

Bei den MoPha-Grafiken in Abb. 3.41 zeigt sich, daß die IC Neuronen nicht einfach nur eine Periodizitätspräferenz besitzen, sondern daß deren Spikes auch eine konstante Phasenlage zum Stimulus besitzen. Um dies zu belegen, wurden die gemessenen Spikes zeitlich in vier Mengen eingeteilt:

- Spikes vor dem Stimulus.
- Spikes beim Onset.
- Spikes während der tonischen Phase (also nach dem Onset und vor dem Ende des Stimulus).
- Spikes nach dem Stimulus.

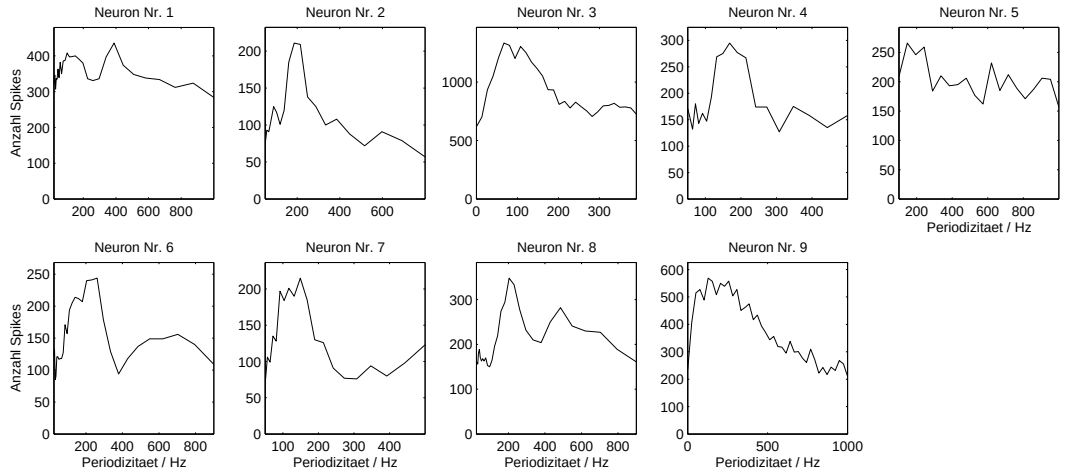

Abbildung 3.38: Die Antwort der Neuronen in Bezug auf die Periodizität (X-Achse). Es zeigt sich, daß die Neuronen auf bestimmte Periodizitäten besonders stark antworten.

Diese Einteilung wird in der untersten Zeile von Abb. 3.41 dargestellt. Die entsprechenden MoPha-Grafiken sind jeweils über den Zeitdiagrammen der Neuronen. Die linke Spalte gehört zu Neuron 1, die rechte zu Neuron 9, entsprechend ist die Zuordnung der restlichen Neuronen.

Im tonischen Bereich zeigt z.B. das 6. Neuron eine extreme Konzentration in der MoPha-Grafik in der Mitte. Das bedeutet zum einen, daß die Neuronen bei dieser Periodizität besonders stark geantwortet haben, zum anderen daß sie dies immer bei einer ähnlichen Phase des Stimulus gemacht haben. Ähnliche Konzentrationen zeigen die anderen Neuronen, teilweise allerdings nicht so stark wie beim 6. Neuron.

Aus den MoPha-Grafiken vor und nach dem Stimulus kann man erkennen, daß keine Konzentration der Spikes vorliegt. Für die Grafiken vor dem Stimulus wäre dies auch unverständlich, woher sollte das Neuron wissen, was als nächstes passiert. Dagegen sind die MoPha-Grafiken für die Zeit nach dem Stimulus durchaus interessant: Sowohl die Spontanaktivität der IC Neuronen als auch die Phasenlage der Spikes hätten durch den Stimulus direkt davor beeinflußt werden können.

Genauer wird über die MoPha-Grafiken im Kapitel Diskussion, Abschnitt 4.1.2 auf Seite 88 diskutiert.

3.4.5 Vergleich mit den simulierten Modellen

In Abbildung 3.42 sind MoPha-Bilder der Schaltkreise - sowohl ursprünglicher Langnerschaltkreis als auch variierter Schaltkreis - dargestellt. Zusätzlich zu den Standardergebnissen wurde je eine MoPha-Grafik für verrauschte Stimuli generiert, um die Antworten auf der Karte etwas größer zu machen.

Man erkennt, daß der ursprüngliche Langnerschaltkreis Kammfilterantworten erzeugt: Der Fleck wiederholt sich von links nach rechts einige Mal. Dagegen zeigt der variierte Schaltkreis nur einen Antwortfleck, so wie z.B. das MoPha-Bild von Neuron 6 in Abb. 3.41.

3.4.6 Darstellung in Bezug auf die Synchronität

Wie in Abbildung 3.38 auf Seite 53 ersichtlich wird, besitzen die Neuronen eine bestimmte Periodizität, auf die sie maximal antworten. Im folgenden wird eine ähnliche Graphik berechnet, nur daß weniger die Menge der Spikes als vielmehr die Position der Spikes berücksichtig wird. Dies wird durch folgende Funktion $S_{rel}(p)$ berechnet:

1. Berechne die Position t_{onset} des Onset der Neuronantwort

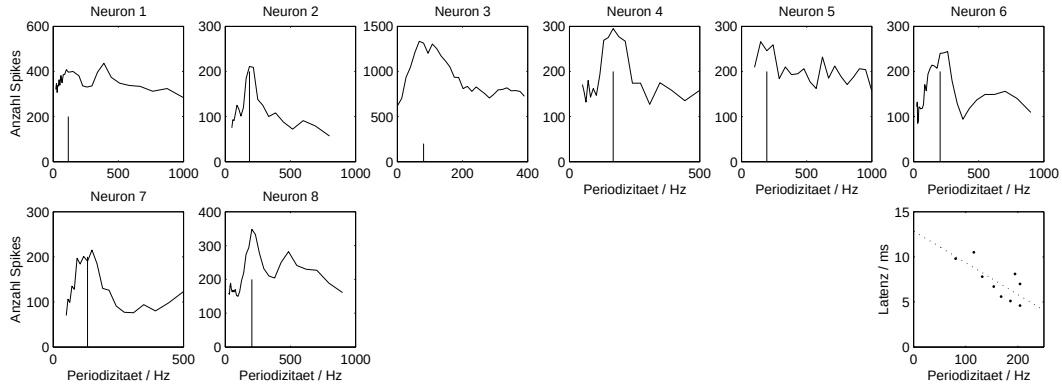

Abbildung 3.39: Darstellung des Zusammenhangs der Latenz und der Periodizität. Zuerst wurde die Präferenz der Neuronen in Bezug auf die Periodizität festgestellt, dann wurde die Latenz (Zeitdauer) zwischen dem Maximum der Neuronantwort und dem Onset berechnet. Diese Wertepaare wurden im Plot rechts unten eingetragen.

2. Für jede Stimulusperiodizität $P_{stimulus}$ wird $S_{rel}(P_{Stimulus})$ wie folgt berechnet:
   - Wähle alle Spikes aus, die frühestens 5 ms nach dem Onset aufgetreten sind.
   - Schaue für jeden dieser Spikes, ob der Spikemoment relativ zu t_{onset} ein vielfaches von $P_{stimulus}$ später aufgetreten ist. Eine Ungenauigkeit von 5% früher oder später wird dabei erlaubt.
   - Die Anzahl der Spikes, für die diese Bedingung erfüllt ist, wird durch die Gesamtanzahl der Spikes geteilt und als Ergebnis verwendet.
3. Das Ergebnis $S_{rel}(p)$ wird als Graph dargestellt.

Die Funktion S_{rel} gibt also als Ergebnis einen Wert zwischen 0 und 1 zurück, der angibt, wie häufig die Spikes synchron zur Periode aufgetreten sind. Sind die Spikes gleichmäßig über die Periode verteilt, so würde $S_{rel}(p)$ ca. 10% als Ergebnis bringen, da die erlaubte Toleranz für die Spikemomente 10% ist. Sind dagegen die Spikes immer sehr häufig an einer Stelle der Periode konzentriert, so müßte der Wert deutlich von 10% abweichen. Da anzunehmen ist, daß die Spikes dann relativ zum Onset die gleiche Position in Bezug auf die Periode haben, müßte der Wert also deutlich höher liegen.

In Abb. 3.43 sind die Ergebnisse für alle neun Neuronen dargestellt. Interessanterweise sieht man keine Funktion, die ungefähr konstant ist. Vielmehr zeigen die meisten Neuronen ein Oszillieren. Besonders deutlich ist dies bei Neuron Nr. 3 zu erkennen. Die Oszillation dieses Neurons entspricht dem vielfachen der Grundfrequenz von ca. 70 Hz.

Die Bedeutung dieser Oszillationen im Bezug auf die Kammfilterantworten wird in Abschnitt 4.1.2 auf Seite 57 genauer diskutiert.

3.5 Aus der Periodizitätskartierung ableitbare Karten

Einer der Vorteile der Implementierung von neuronalen Netzen ist, daß man sehr viele unterschiedliche Untersuchungen mit demselben Netzwerk machen kann und auch jedesmal die Aktivität aller Neuronen als Ergebnis bekommt. In der Biologie dagegen kann man zum einem von einem Neuron nicht beliebig lange ableiten, und zum anderen ist immer nur die Aktivität weniger Neuronen bekannt. Im folgenden wird dieser Vorteil ausgenutzt, um die IC Karte, durch die Periodizität und Trägerfrequenz kartiert werden, auf zwei andere Parameter zu testen, um herauszufinden, ob sich durch den Aufbau des Netzwerkes eine automatische Kartierung dieser Parameter ergibt. Getestet

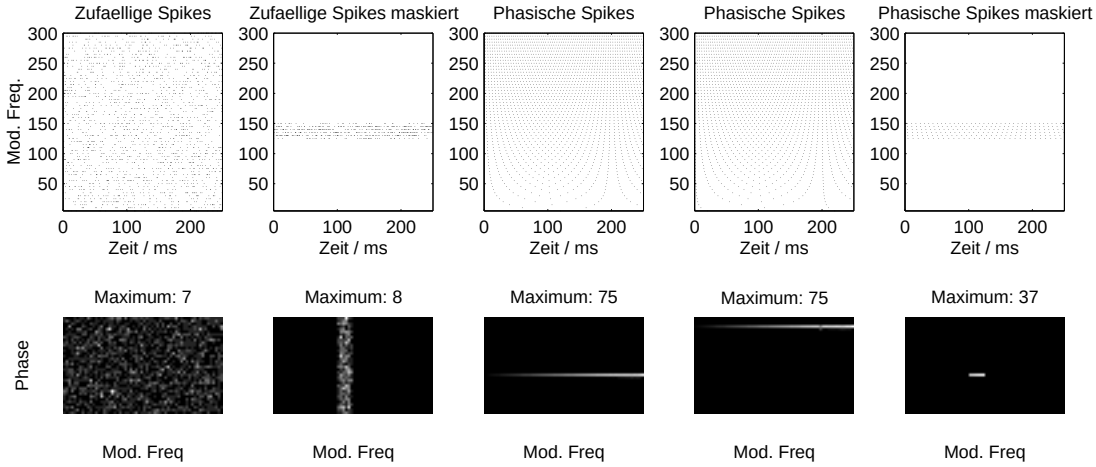

Abbildung 3.40: Darstellung von verschiedenen künstlichen erzeugten Spikeantworten als MoPha Grafik. Links: Verwendung von zufälligen Spikes, dies ergibt zufällige Punkte im MoPha-Bild. Daneben: Verwendung von zufälligen Spikes, die eingeschränkt auf einige wenige Modulationsfrequenzen sind. Dies ergibt einen senkrechten Streifen, da alle Phasen in den wenigen Modulationsfrequenzen, die eine Antwort erzeugten, enthalten sind. Dritte und vierte Grafik: Verwendung von zwei phasische Antworten mit zwei unterschiedlichen Phasen. Daraus entsteht ein waagrechter Streifen, die Position des Streifens hängt von der Phase ab. Rechts: Verwendung von Spikeantworten, in dem eine phasische Antwort auf wenige Modulationsfrequenzen eingeschränkt wurde. Dadurch entsteht ein Fleck im MoPha Bild.

werden dazu der Parameter *Lautstärke*, der Parameter *Latenz* und die Antwortbreite in Bezug auf die Periodizität.

3.5.1 Lautstärkenkarte

Für den Test, ob sich die Lautstärke systematisch auf der Karte wiederspiegelt wird ein konstanter AM Stimulus verwendet. Die Parameter des Stimulus sind 100 Hz Periodizität zu 1300 Hz Trägerfrequenz. Dieser Stimulus erzeugt in etwa einen Antwortfleck in der Mitte der Karte. Um zu prüfen, ob die Lautstärke sich systematisch auf der Karte verändert, wird diese von 0 % bis 1000 % in Schritten von 20 % variiert. Eine Auswahl der Ergebniskarten ist in Abb. 3.44 abgebildet. Man sieht, wie bei zunehmender Lautstärke zuerst der Antwortfleck breiter wird und sich dann in zwei Flecken unterteilt, die auseinanderwandern, wenn die Lautstärke weiter zunimmt.

Zusätzlich zu Abb. 3.44, in der 10 der 50 Ergebniskarten abgebildet sind, ist in Abb. 3.45 eine Zusammenfassung aller 50 Ergebnisse abgebildet. Dazu wurden aus jeder Karte erst der Punkt berechnet, wo von oben gesehen die Antwort anfängt. Zusätzlich wurde das Gleiche nochmal gemacht, nur von unten gesehen. Dadurch erhält man pro Karte zwei Punkte, die den Bereich markieren, in dem der Antwortfleck war. Die Größe des Lochs in der Mitte des Flecks wird durch dieses Verfahren allerdings nicht festgestellt. In Abb. 3.45 sind die Anfangs- und Endpositionen abgebildet, die sich daraus ergeben haben. Die X-Achse zeigt die Positionen, die Y-Achse die Lautstärke. Da immer ein Lautstärkebereich zu einer Position führen kann, kann ein Punkt auf der Linie mehreren Ergebniskarten entsprechen, daher sind wesentlich weniger Punkte als Ergebniskarten zu sehen.

Man erkennt, das bei einer geringen Lautstärke der Fleck sich an einer Stelle konzentriert, und bei zunehmender Lautstärke die Grenzen nach außen wandern. Die sich ergebene Kurve ähnelt einer Parabel. Über den Aspekt der Lautstärkenkartierung wird im Diskussionsteil, Abschnitt 4.1.5 ab Seite 92 genauer diskutiert.

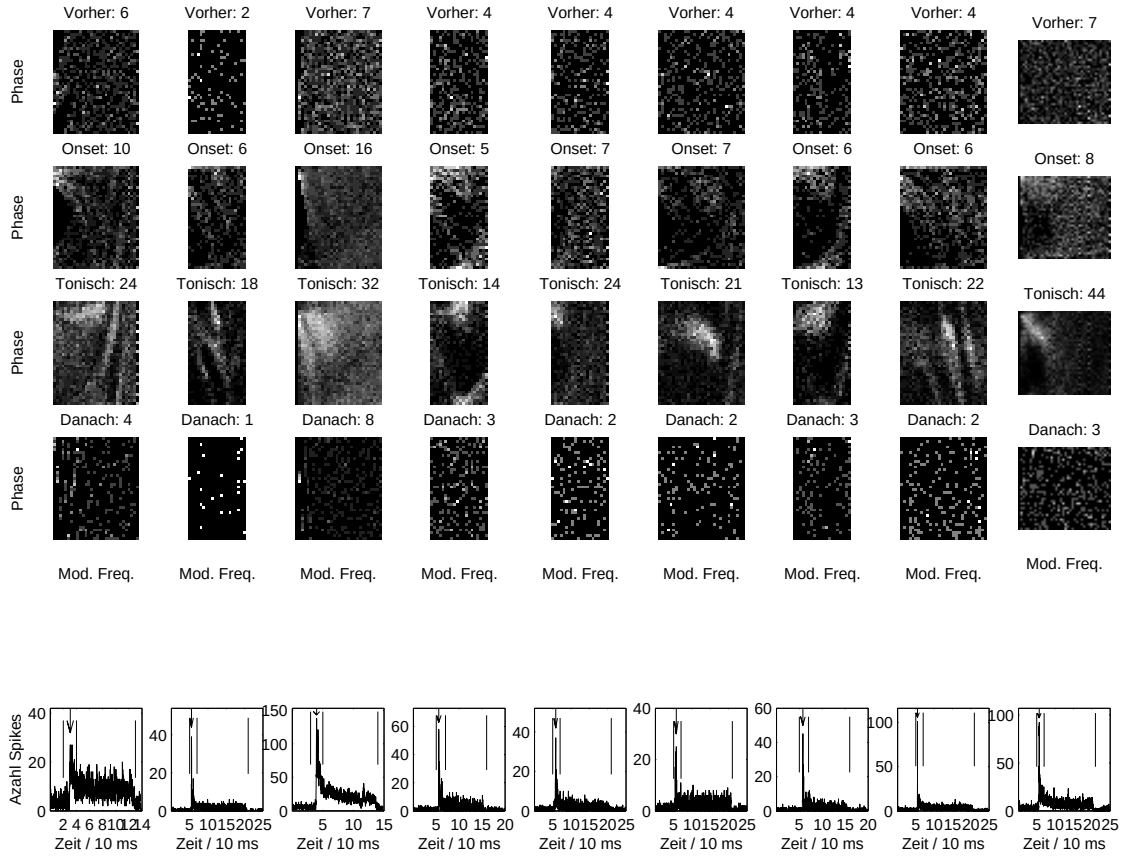

Abbildung 3.41: Einige MoPha-Grafiken, berechnet aus echten biologischen Ableitungen. Zu jedem Neuron wurde die Antwort in vier Teile unterteilt (vgl. unterste Zeile): Vor dem Ton, beim Onset, während der Ton aktiv ist und nach dem Ton. Es wurde jeweils das zugehörige MoPha-Bild berechnet. Die Zahlen über den Grafiken geben das Maximum an.

3.5.2 Latenzkarte

Zusätzlich zu der Kartierung der Amplitude des Stimulus soll geprüft werden, ob eine Latenzkartierung durch das Langnermodell erzeugt wird. Als Latenz wird die Zeit bezeichnet, die von dem Moment des Stimulus Onset ab vergeht, bis das Neuron den ersten Spike erzeugt. Um dies zu prüfen, wurden Stimulussignale mit der Dauer von 100 ms generiert, wobei die ersten 20 ms die Amplitude des Stimulus Null war. Die restlichen 80 ms waren Klicks mit einer konstanten Periodizität pro Stimulus. Es wurden Periodizitäten von 30 Hz bis 230 Hz verwendet, also 201 Stimuli generiert. Diese wurden mit dem variierten Langnermodell analysiert. Eine typische Karte, die ein solcher Stimulus erzeugt, ist links in Abb. 3.46 dargestellt. Bei der zum Stimulus passenden Periodizität entsteht ein senkrechter Balken. Dies liegt daran, daß die Klicks alle Frequenzkanäle der Gammatonfilterbank aktivieren. Allerdings erstreckt sich der Antwortbereich nicht komplett über alle Frequenzen, sondern nur bis zu einer Grenzfrequenz von ca. 850 Hz, weil die Amplitude der Filterbankantwort mit abnehmender Frequenz auch abnimmt. Ein Dirac-Impuls enthält zwar alle Frequenzen gleichmäßig, allerdings sind die Klicks nur eine Näherung von Dirac-Impulsen. Daher sind die hohen Frequenzen stärker enthalten als die niedrigen.

Wichtig für die Latenzuntersuchung ist aber weniger die Größe der Antwort als vielmehr die Zeitdauer die vom Onset des Signals bis zu den ersten Spikes vergeht. Dazu wurden für jeden Filterkanal die Neuronen betrachtet, die eine Antwort erzeugen. Die Zeitdauer, die zwischen Onset und dieser ersten Antwort vergeht, wurde protokolliert. Die Ergebnisse sind in der mittleren

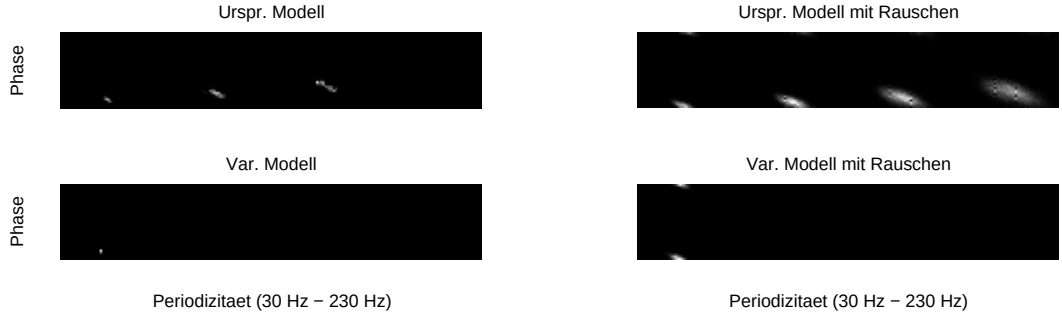

Abbildung 3.42: Ergebnisse zu den MoPha-Untersuchungen der simulierten Schaltkreise. Links sind die Ergebnisse zum ursprünglichen Schaltkreis, rechts die Ergebnisse zum variierten. Die obere Reihe wurde mit einem reinen AM-Stimulus erzeugt, die untere mit einem auf ein AM-Stimulus addierten Rauschen (SNR ca. 0.5)

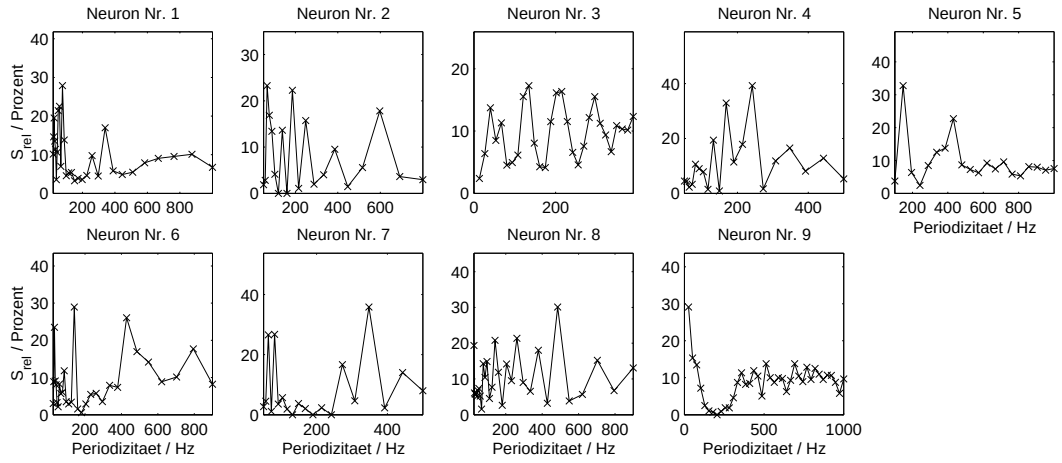

Abbildung 3.43: Darstellung der Ergebnisse der Funktion $S_{rel}(p)$ für jedes Neuron. Je größer S_{rel}, desto synchroner sind die Spikes zur Stimulusperiode aufgetreten.

Grafik in Abb. 3.46 dargestellt. Die X-Achse zeigt die Periodizität, auf die der Schaltkreis eingestellt war, die Y-Achse die Zeitdauer zwischen Onset und ersten Spike. Man erkennt deutlich eine Abhängigkeit der Latenz zur Periodizität, ein Großteil der Punkte liegt auf einer abfallenden Hyperbel. Zusätzlich gibt es weitere Hyperbeln, an denen entlang weiter Punkte liegen. Dieser Effekt ergibt sich dann, wenn das Neuron nicht auf den ersten Klick, sondern auf den zweiten oder späteren Klick eine Antwort erzeugt. Dadurch entstehen die zusätzlichen Hyperbeln. Um ein klareres Ergebnis darzustellen, wurde zu jeder Periodizität der Punkt bestimmt, der in der Mitte der nach der Latenz sortierten Punktmenge liegt (also der Median bestimmt). Das Ergebnis ist im rechten Plot zu sehen, es bleibt die Hyperbel übrig, an der die meisten Ergebnispunkte liegen.

Die Ergebnisse zeigen im Prinzip, daß eine Art Latenzkartierung auf der IC Karte vorhanden ist. Andererseits zeigt die gepunktete Linie im rechten Plot den exakt gleichen Verlauf, um wenige ms nach unten verschoben. Um diese Linie zu erzeugen, wurde der Abstand zwischen den Klicks gemessen. Dieser verändert sich natürlich mit der Periodizität, genauer: der Abstand der Klicks entspricht dem Kehrwert der Periodizität. Dadurch entsteht die Hyperbelfunktion. Bedeutet dies, daß die Latenz nicht kartiert ist, sondern durch den Stimulusaufbau erzeugt wird? Diese Frage wird in der Diskussion, Abschnitt 4.1.5 ab Seite 94 diskutiert.

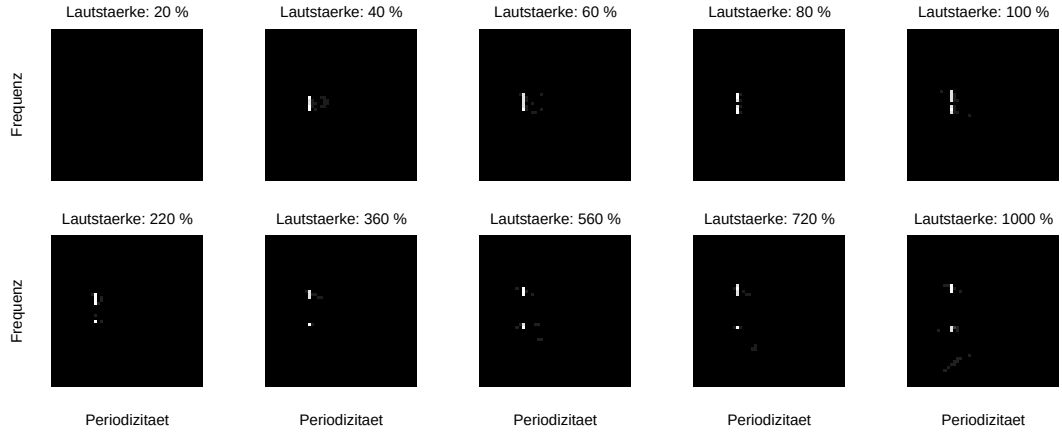

Abbildung 3.44: Ergebnis der Untersuchung zur Kartierung von Lautstärke durch das Langnermodell. Der Stimulus wird konstant bei 100 Hz/1300 Hz gehalten, die Lautstärke von 0% bis 1000% variiert. Die Dauer des Stimulus war 100 ms.

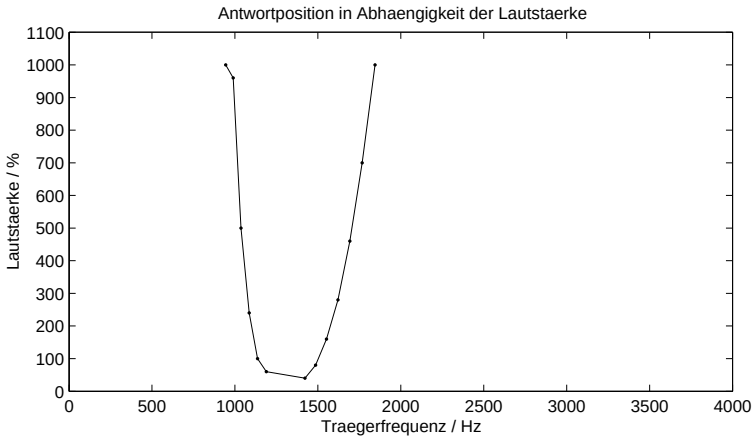

Abbildung 3.45: Zusammenfassung aller 50 Ergebnisse zur Lautstärkenuntersuchung. Auf der X-Achse sind die Positionen aufgetragen, wo die äußeren Grenzen des Fleckes in der Karte in Bezug auf die Trägerfrequenz lagen. Die Y-Achse stellt die Lautstärke dar, bei der die Karte berechnet wurde. Da mehrere unterschiedliche Lautstärken zu denselben Grenzen führen kann, repräsentiert ein Punkt mehreren Ergebniskarten

3.5.3 Antwortbreite in Bezug auf die Periodizität

Ein weiterer Parameter, der gemessen werden kann, ist die Breite der Antwort in Bezug auf die Periodizität. Damit ist gemeint, daß bei einer festen Trägerfrequenz das Modell mit verschiedenen Periodizitäten stimuliert wird und protokollliert wird, wie breit der Periodizitätsbereich ist, auf den ein einzelner Schaltkreis antwortet. Diese Messung wurde für 6 Neuronen mit einer Trägerfrequenz von ca. 2000 Hz durchgeführt. Die Neuronen liegen auf den folgenden X-Positionen in der Karte: 5, 12, 19, 26, 33, 40, 47. Die jeweilige Periodizität ergibt sich aus den Ergebniskurven. Diese sind in Abb. 3.47 abgebildet. Die X-Achse ist jeweils die Periodizität, die Y-Achse die Anzahl Spikes.

Man erkennt, daß mit zunehmender Periodizität das Antwortverhalten der Neuronen breiter wird. Während das Neuron 5 nur auf einen sehr kleinen Bereich von ca. 66 Hz bis 69 Hz feuert, erzeugt das Neuron 40 Antworten in einem Bereich von 200 Hz bis 220 Hz. Dies entspricht ungefähr der Tatsache, daß die Flecken auf der Karte immer breiter werden. Dieser Zusammenhang ist aber nicht prinzipiell so: Man könnte die Karte so konstruieren, daß für kleine Periodizitäten mehr Neuronen spendiert werden als für Große. Durch entsprechende Einstellungen würden die Flecken in der Karte für alle Periodizitäten die gleiche Breite haben, obwohl die Neuronen ein

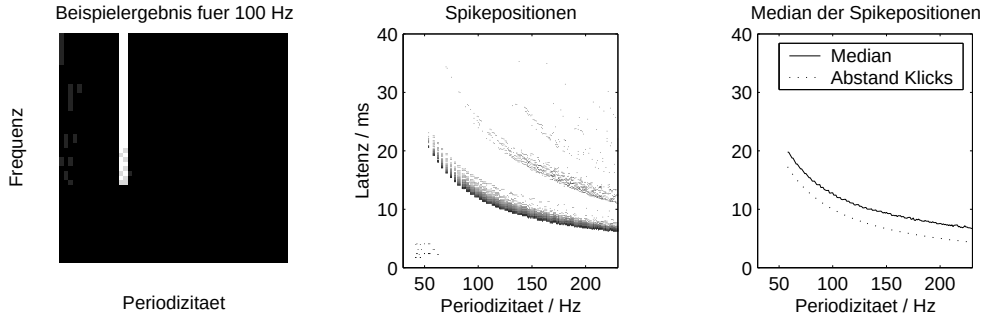

Abbildung 3.46: Die Ergebnisse zur Latenzuntersuchung. Das linke Bild zeigt beispielhaft, wie die IC-Karte aussieht, wenn ein Stimulus mit Klicks verwendet wird. Der mittleren Plot zeigt, wie die Zeitdauer zwischen Onset und erstem Spike vom Periodizitätstuning der Neuronen abhängt. Für den rechten Plot sind nur die Medianwerte zu jeder Periodizität abgebildet.

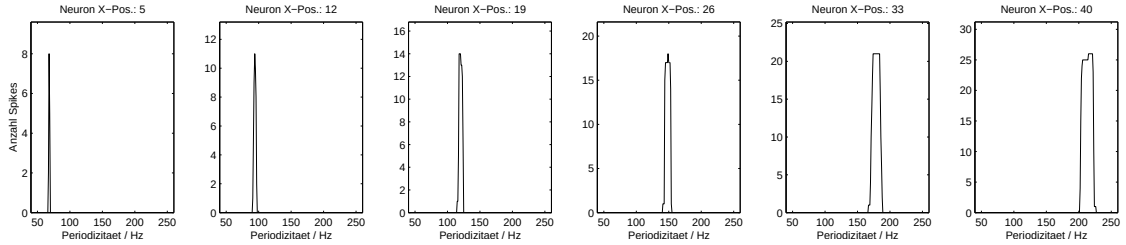

Abbildung 3.47: Darstellung des Antwortverhaltens von 6 Schaltkreisen, wenn die Periodizität variiert wird.

immer breiteres Periodizitätstuning haben. Genauer wird über diese Karte noch in der Diskussion, Abschnitt 4.1.5 auf Seite 95 diskutiert.

3.6 Ergebnisse unter Einbeziehung kortikaler Areale

3.6.1 Einführung

Im folgenden wird dargestellt, wie mit Hilfe der Informationen der Koinzidenzschicht der Stimulus klassifiziert oder separiert werden kann, indem biologisch plausible Modelle von kortikalen Arealen verwendet werden [4]. Diese werden durch spikende Assoziativspeicher modelliert. Es werden zwei Schichten von kortikalen Arealen realisiert, eine untere Schicht C, welche den primären auditorischen Cortex modelliert und ein höheres Areal D. Während in C die Muster topologisch sehr ähnlich zu der Koinzidenzschicht aussehen, werden in D reine Zufallsmuster verwendet (vgl. Abb. 3.48). Jedes Muster besteht aus einer Assembly von 50 Neuronen und repräsentiert einen Vokal bei einer festen Periodizität. Die Assemblies verändern sich während einer Simulation nicht, sondern werden beim Start eingelesen oder erzeugt. Es werden immer 20 Assemblies erzeugt, auch wenn im folgenden nur fünf oder zwei verwendet werden. Die restlichen werden durch Zufallsmuster (auch in C) aufgefüllt.

Es existieren die folgenden Verbindungen zwischen den Schichten (dabei bezeichnet K die Koinzidenzkarte, C den Assoziativspeicher des primären auditorischen Kortex und D ein höheres Areal), jedes Areal ist unterteilt in exitatorische Neuronen und inhibitorische Neuronen (die inhibitorischen Neuronen sind zusätzlich nochmals in zwei Gruppen unterteilt, werden aber im folgenden als eine zusammengefaßte inhibitorische Population betrachtet):

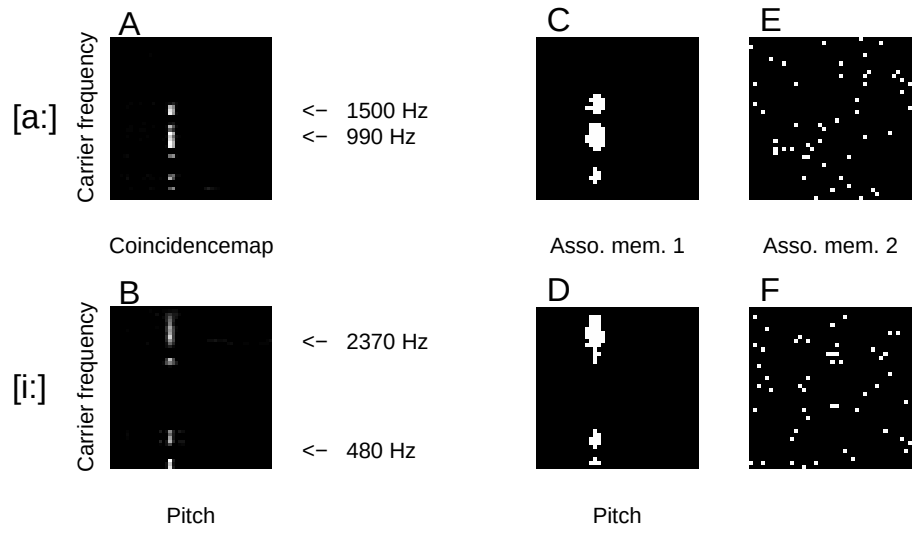

Abbildung 3.48: Erzeugung der Assemblies für die kortikalen Areale anhand zweier Vokale ([a:] und [i:]). Links sind die Ergebnisse der Koinzidenzkarte dargestellt. Die 50 aktivsten Punkte werden mit Hilfe eines Schwellwertes herausgesucht und stellen das Muster für den Assoziativspeicher C dar (mittlere Spalte). Dagegen werden in der Schicht D nur Zufallsmuster verwendet (rechte Spalte).

1. Exitatorische heteroassoziative Verbindungen:
   - $K_{ex} \rightarrow C_{ex}$: Dies ist eine topologische Verbindung, die nicht von den gespeicherten Mustern abhängt. Spikes von zum Beispiel links oben in K werden in eine Umgebung des Neurons links oben in C geleitet. Diese Umgebung ist 5×5 Neuronen groß, bzw. im Beispiel nur 3×3, weil die restlichen Neuronen außerhalb der Karte liegen. Am Rand werden die Projektionsumgebungen also einfach abgeschnitten.
   - $C_{ex} \rightarrow K_{ex}$: Diese Verbindung ist auch topologisch und hängt nicht von den gespeicherten Mustern ab. Die Umgebungsgröße ist wieder 5×5.
   - $C_{ex} \rightarrow D_{ex}$: Diese Verbindung ist rein musterabhängig. Es existieren Verbindungen genau von den Neuronen in C zu den Neuronen in D, die zum gleichen Muster gehören. Die Umgebungsgröße ist: 61×61, ein Neuron von C kann also fast in die ganze Schicht D projezieren, anhängig von seiner Position: Ein Neuron in der Mitte von C kann alle Neuronen in D erreichen, ein Neuron am Rand von C erreicht dagegen nicht alle Neuronen in D.
   - $D_{ex} \rightarrow C_{ex}$: Diese Verbindung ist wieder nur von den Mustern abhängig. Die Größe der Umgebung ist 11×11.
   - $C_{ex} \rightarrow C_{in}$: Musterabhängig, 11×11.
   - $D_{ex} \rightarrow D_{in}$: Musterabhängig, 11×11.
   - $K_{ex} \rightarrow C_{in}$: Topologisch, 5×5.
   - $C_{ex} \rightarrow K_{in}$: topologisch, 5×5.
   - $C_{ex} \rightarrow D_{in}$: Musterabhängig, 61×61.
   - $D_{ex} \rightarrow C_{in}$: Musterabhängig, 11×11.
2. Exitatorische autoassoziative Verbindungen:
   - $C_{ex} \leftrightarrow C_{ex}$: Musterabhängig, 11×11.

- $D_{ex} \leftrightarrow D_{ex}$: Musterabhängig, 61×61.

3. Inhibitorische heteroassoziative Verbindungen:
   - $K_{in} \rightarrow K_{ex}$: Topologisch, 31×31.
   - $C_{in} \rightarrow C_{ex}$: Topologisch, 11×11.
   - $D_{in} \rightarrow D_{ex}$: Topologisch, 61×61.
4. Inhibitorische autoassoziative Verbindungen:
   - $K_{in} \rightarrow K_{in}$: Topologisch, 31×31.
   - $C_{in} \rightarrow C_{in}$: Topologisch, 11×11.
   - $D_{in} \rightarrow D_{in}$: Topologisch, 61×61.

Die Verbindungsstärken werden mit X_aY_b abgekürzt, wobei $X,Y \in \{K,C,D\}$ und $a,b \in \{ex,in\}$. Zum Beispiel würde die Verbindung der exitatorischen Schicht C mit der inhibitorischen Schicht D als $C_{ex}D_{in}$ bezeichnet werden. Die gesamte Struktur der Verbindungen ist in Abb. 3.49 graphisch dargestellt.

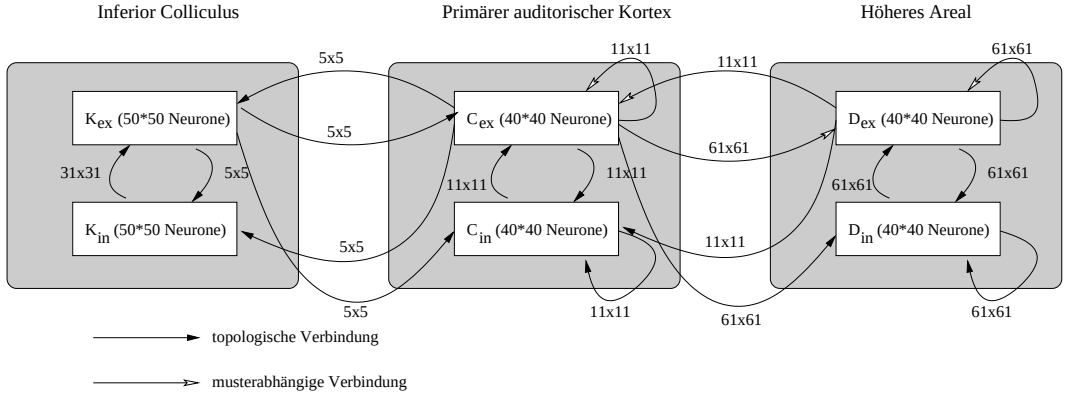

Abbildung 3.49: Das kortikale Model, zuzüglich der Koinzidenzschicht im Colliculus Inferior. Die Zahlen an den Pfeilen geben die Umgebungsgröße an, in die ein Neuron projezieren kann.

Um das Prinzip der Assoziativspeicher zu veranschaulichen, werden zuerst einige einfache Versuche mit künstlichen Stimuli durchgeführt. Anschließend wird getestet, inwieweit die Assoziativspeicher verschiedene Stimuli klassifizieren können. Danach werden die Aktivitäten dargestellt, die sich bei Verwendung von Stimuli ergeben, bei denen sich zwei Vokale überlagern.

3.6.2 Grundlagen anhand künstlicher Stimuli

Einstellung der Stärke der assoziativen Verbindungen

Zunächst werden 5 Muster eingespeichert (siehe Abb. 3.52. Für diese werden Klicks als Stimulus verwendet. Jedes der 5 Muster ensteht aus Klicks (Impulse) mit konstanter Periodizität, es wurden 100 Hz, 120 Hz, 140 Hz, 160 Hz und 180 Hz verwendet. Die Parameter der Verbindungsstärken wurden mit Standardwerten initialisiert, wobei die Rückwärtsprojektionen von D auf C und von C auf K auf 0 gesetzt wurden. Im folgenden wird demonstriert, wie Veränderungen einiger Verbindungsstärken sich auf die Ergebnisse auswirken.

Das Ergebnis der Simulationen mit den Standardwerten ist in Abb. 3.50 dargestellt. Als Stimulus wurde Klicks verwendet, die jeweils der Reihe nach 12 Periodizitäten mit 100 Hz, 12 Periodizitäten mit 120 Hz, ... enthielten. Abgebildet ist die Anzahl der Neuronen jeder Assembly, die innerhalb der letzten 4 ms gefeuert haben.

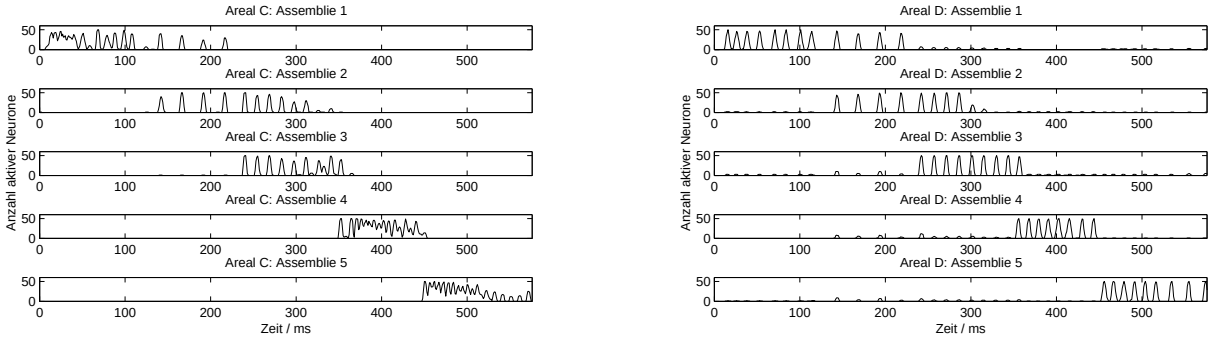

Abbildung 3.50: Ergebnisse der ersten Assoziativspeicher Simulation. Dargestellt sind die Anzahl Neuronen jeder Assembly die aktiv waren. Links die Neuronen der Schicht C, rechts die Neuronen der Schicht D. Als Stimulus wurden Klicks verwendet, deren Periodizität zuerst 100 Hz, dann 120 Hz, ... bis 180 Hz war, also genau die Stimuli, aus denen bei der Mustergenerierung die Assemblies hervorgegangen sind.

Es zeigt sich, daß, wenn die Periodizität der Klicks sich verändert, die jeweils zugehörige Assembly aktiv wird. Die einzelnen 'Aktivitätsberge' entsprechen einem einzelnen Auslesevorgang. Diese werden durch die Spikes der Neuronen aus der Koinzidenzschicht angestoßen. Man sieht, wie sich die Häufigkeit der Auslesevorgänge erhöht, wenn der Stimulus eine höhere Periodizität hat. Allerdings feuern innerhalb der 4 ms nicht immer alle 50 Neuronen. Außerdem sind längere Phasen zu sehen, in denen zwei Assemblies gleichzeitig aktiv sind. Während es für das Areal C keinen Grund gibt, warum nicht mehrere Assemblies aktiv sein sollen, sollte im Areal D nur eine Assembly aktiv sein. Dies liegt darin begründet, daß Areal C den gesamten primären auditorischen Kortex modelliert, und somit einen wesentlich größeren Bereich als Areal D, das nur einen kleinen Bereich modelliert. In diesem sollte daher nur ein Muster aktiv sein.

Um die Antworten zu verschärfen, werden für die nächste Simulation die inhibitorischen Verbindungen verstärkt. Das Ergebnis ist in Abb. 3.51 dargestellt.

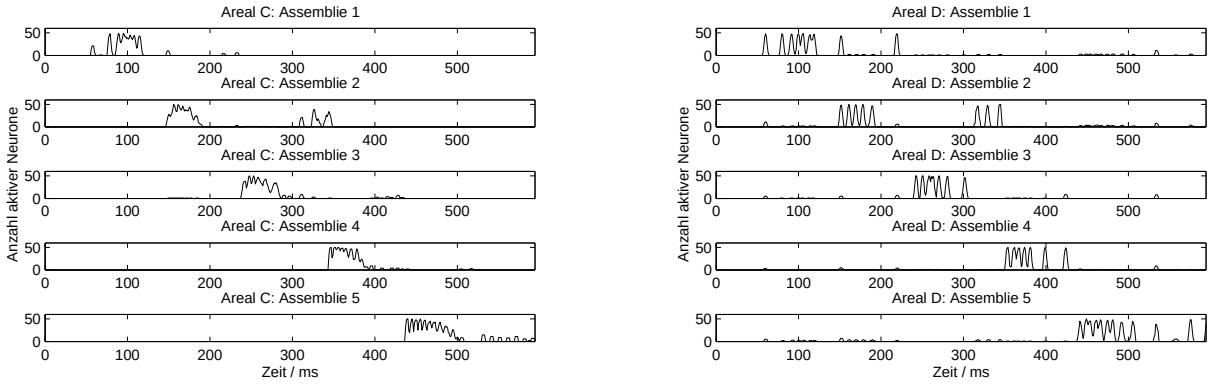

Abbildung 3.51: Ergebnis der Simulationen, wenn die Inhibition erhöht wird.

Man sieht, daß in beiden Schichten C und D nur eine Assembly gleichzeitig aktiv ist. Durch die erhöhte Inhibition setzt sich das Muster durch, dessen Neuronen zuerst aktiv wurden. Da die Muster in Areal C sehr dicht nebeneinander liegen (vgl. Abb. 3.52), kann es passieren, daß zu Beginn Neuronen feuern, die zu zwei Assemblies gehören. In diesem Fall kann eine Assembly aktiv werden, die eigentlich nicht von der K Schicht aktiviert werden sollte. Bei 300 ms sieht man, daß, obwohl in diesem Moment der Stimulus von Assembly 3 als Input anliegt, die Assembly 2 aktiv wird. Dies ist ein prinzipielles Problem, wenn die Assemblies im Assoziativspeicher keine

Zufallsmuster sind, sondern - wie in diesem Fall - fleckige Strukturen haben, und sich vielleicht teilweise überlappen. Der Assoziativspeicher irrt sich in diesem Beispiel bei 300 ms.

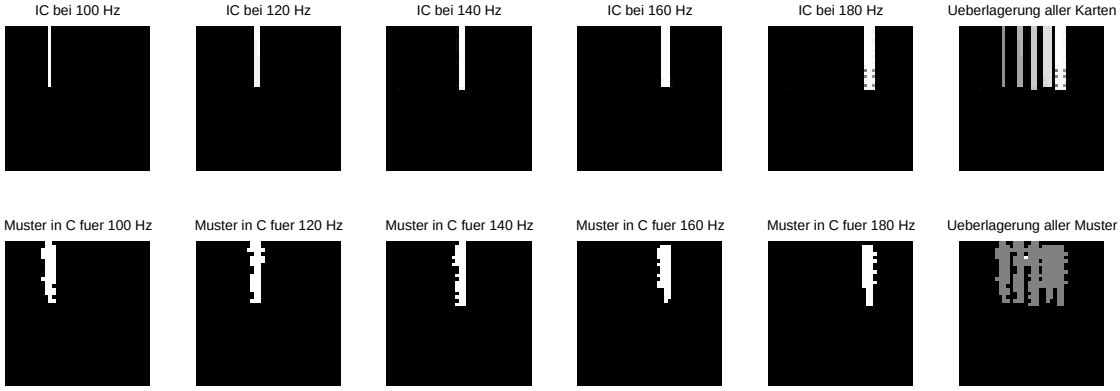

Abbildung 3.52: Obere Zeile: Die Ergebniskarten, die sich bei den jeweiligen Stimuli ergibt. Rechts ist die Überlagerung aller Karten dargestellt, an einem Punkt überlappen sich zwei Muster. Untere Zeile: Die aus den Karten generierten Muster. Rechts ist die Überlagerung aller Muster dargestellt, sie liegen dich nebeneinander.

Interessant ist in Bezug auf die Inhibition auch folgende Überlegung: In Areal D erreicht die Inhibition eines Neurons in der Mitte alle Neuronen der Schicht, und da die Muster dort Zufallsmuster sind, wird im Normalfall immer nur ein Muster aktiv werden, wenn die Inhibition entsprechend groß ist. In Areal C dagegen sind die Umgebungsgrößen kleiner gewählt, und so kann ein Muster am linken Rand auch bei großer Inhibition ein Muster am rechten Rand nicht unterdrücken. Da die Muster in C nicht zufallsverteilt sind, sondern aus den Aktivierungen der Koinzidenzkarte hervorgehen, sind sie auf einen kleineren Fleck begrenzt. Dadurch können selbst bei größerer Inhibition zwei Muster gleichzeitig aktiv werden.

Nutzen des Assoziativspeichers D

Im Gegensatz zu Assoziativspeicher C, der den kompletten primären auditorischen Kortex modelliert, wird von D nur ein kleiner Bereich eines höheren Areals modelliert [5]. Dies ermöglicht, daß die lateralen Verbindungen, also die Verbindungen zwischen zwei Neuronen innerhalb der gleichen Schicht, sich über einen wesentlich größeren Bereich erstrecken als in C. Dadurch können in D Muster, die von K und C unvollständig erkannt wurden, durch laterale Exitation komplettiert werden. Durch die Rückwärtsprojektionen können diese Zusatzinformationen wieder in Areal C und bei Bedarf auch in Areal K verwendet werden. Dieser Effekt wird mit den folgenden Simulationen demonstriert:

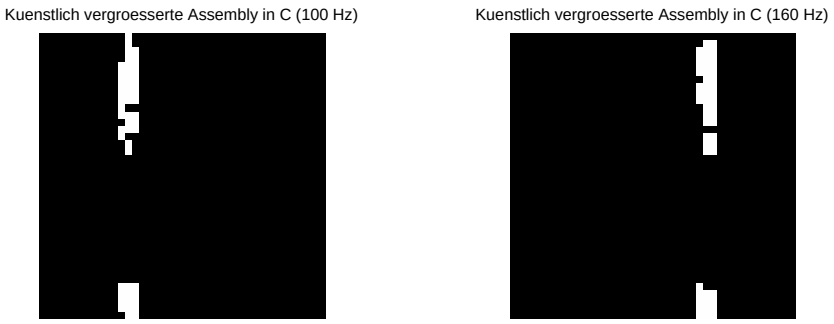

Abbildung 3.53: Die Muster zur Demonstration des Einflusses der Rückwärtsprojektionen von D nach C

Von den vorherigen Assemblies werden nur die zwei Muster gespeichert, die durch 100 Hz

und 180 Hz generiert werden. Diese werden mit einem zusätzlichen Fleck, der nicht nahe am ursprünglichen Fleck liegt, künstlich vergrößert (siehe Abb. 3.53).

- Simulation ohne Rückwärtsprojektionen von D nach C.
- Simulation mit Rückwärtsprojektionen von D nach C

Da die Spikes, die durch die Rückwärtsprojektionen erzeugt werden, entsprechend später auftreten als die Spikes, die durch Vorwärtsprojektionen aus K generiert werden, werden die beiden Simulationen sowohl mit einem Zeitfenster von 4 ms (wie bisher) und einem größerem Zeitfenster von 10 ms durchgeführt. Im letzteren Fall werden die Spikes, die durch die Projektionen von K erzeugt werden, noch mit den Spikes, die durch die Projektionen von D erzeugt werden, zusammen in der Darstellung der aktiven Spikes gezählt. Das Ergebnis ist in Abbildung 3.54 dargestellt. Der Übersichtlichkeit halber wurde nur die Assembly 1 dargestellt, die anderen Assemblies zeigten keine wesentlichen Aktivitäten.

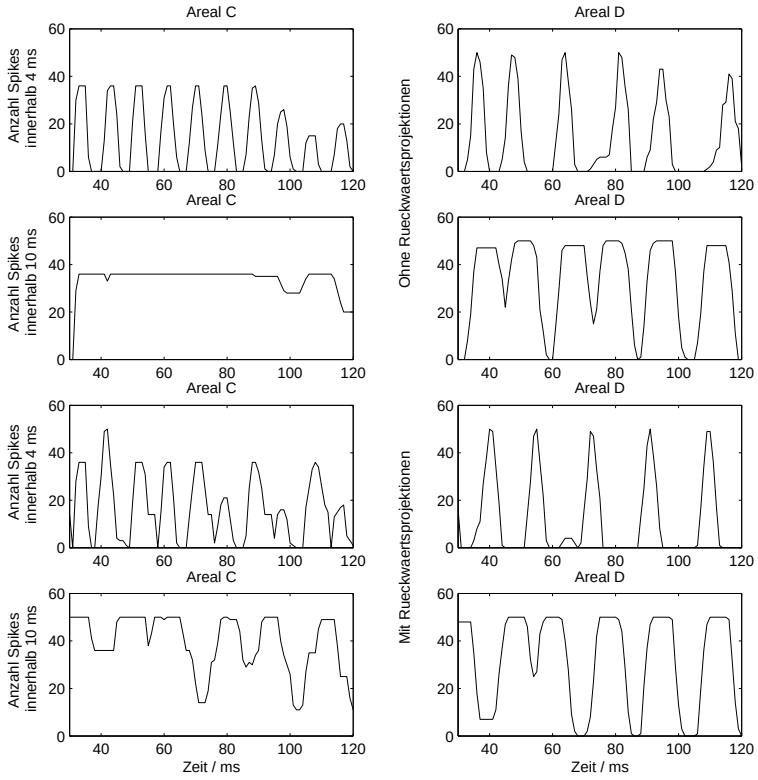

Abbildung 3.54: Darstellung des Einflusses von Rückwärtsprojektionen von D nach C. In den beiden oberen Plots sind die Rückwärtsprojektionen inaktiv, die Assembly in C kann sich trotz der autoassoziativen Rückkopplung wegen der großen Entfernung des aktiven Flecks zum nicht aktiven Fleck nicht vollständig aktivieren. In den beiden unteren Plots zeigt sich, daß die Assemblies komplett aktiviert werden, dieser Effekt ist bei dem untersten deutlicher erkennbar, weil dieser (und zum Vergleich der zweite von oben) mit einer größeren Zeitkonstante für die Overlapdetektion berechnet wurde.

Der Einfluß der Zeitkonstante für die Overlapdetektion wird an dieser Abbildung deutlich: Ist sie klein (4 ms), so sieht man die einzelnen Auslesevorgänge, aber selbst wenn alle Neuronen aktiv werden, erreicht der Wert nicht unbedingt das Maximum von 50 aktiven Neuronen innerhalb 4 ms. In der dritten Zeile sieht man auch für das Areal C bei 80 ms/95 ms/110 ms einen kleines Maximum nach dem großen Maximum. Dieses entsteht dadurch, daß die Neuronen, die bei dem Auslesevorgang bisher noch nicht gefeuert haben, erst jetzt feuern, da sie mit einer Verzögerung ca. 10 ms von den Rückwärtsprojektionen erregt werden. Da das Feuern der anderen Neuronen aber schon länger als 4 ms zurückliegt, erreicht der Overlap jetzt nicht 50 Neuron. Wird dagegen der Overlap mit 10 ms berechnet, so fallen die Spikes, die von der Vorwärtsprojektion aus K erzeugt werden, und die Spikes, die von den Rückwärtsprojektionen aus D erzeugt werden, noch in das Zeitfenster von 10 ms, und der Overlapwert erreicht 50 Neuronen. Dafür sieht man die einzelnen Auslesevorgänge schlechter.

Das Verhalten der D-Schicht ist weitgehendst unabhängig von der Rückwärtsprojektion. Auf der rechten Seite sieht man die Aktivitäten der Assembly in D, es zeigen sich keine wesentlichen Unterschiede. In beiden Fällen erreicht die Aktivität innerhalb kurzer Zeit das Maximum von 50 Neuronenn.

3.6.3 Klassifikation

Für die Klassifikation werden zunächst die schon früher verwendeten Vokale als Stimulus benutzt. Dieser Datensatz stellt zu jedem Vokal vier Exemplare mit unterschiedlicher Periodizität bereit. Von diesen werden die K-Karten berechnet und anschließend in C gespeichert. Danach werden die Vokale selbst als Stimulus verwendet und geprüft, wieviele Vokale korrekt klassifiziert werden.

Anschließend wird die Aufgabe erschwert, indem ein erweiterter Datensatz mit 18 Exemplaren je Vokal verwendet wird. Von diesen werden 4 zum Lernen für den Assoziativspeicher ausgewählt, und anschließend mit den restlichen 14 die Klassifikationsleistung geprüft.

Klassifikation von gespeicherten Mustern

Ohne Rückwärtsprojektionen

Zunächst werden die gleichen Muster zum Lernen und zum Testen verwendet. Dazu wird der bekannte Datensatz aus fünf Vokalen mit je vier Exemplaren verwendet. Für den ersten Versuch wird die Rückprojektion von D nach C auf 0 gesetzt. Jeder der 20 Vokale wird dem Modell einmal komplett (20000 Samples, entspricht 0.5 Sekunden) präsentiert und die Aktivität in C und D protokolliert. In Abb. 3.55 ist beispielhaft das Protokoll für ein [a:] mit 110 Hz abgebildet. Man erkennt, daß die erste Assembly mit Abstand am häufigsten aktiv ist.

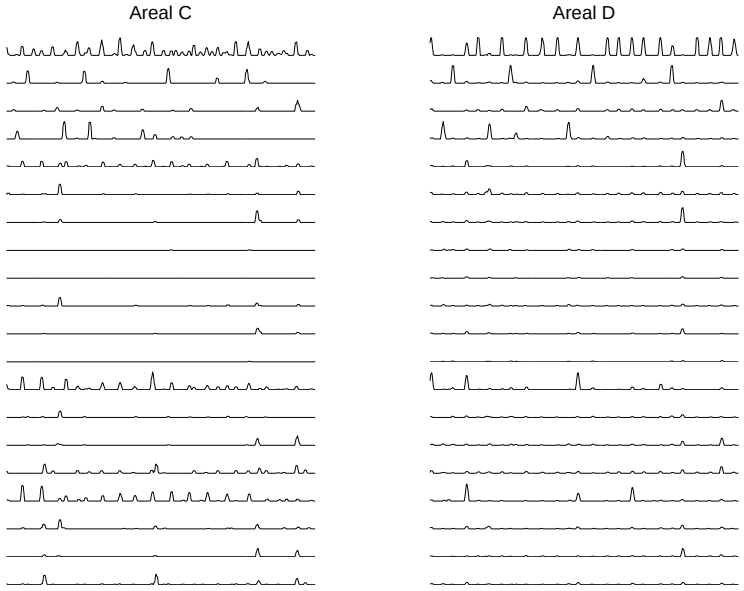

Abbildung 3.55: Beispielprotokoll der Aktivitäten in D und C, wenn der Vokal [a:] mit 110 Hz Periodizität präsentiert wird. Jeweils 4 Zeilen entsprechen den Assemblies eines Vokals, z.B. Zeile 1: [a:] bei 110 Hz, Zeile 2: [a:] bei 150 Hz. In Zeile 5-8 sind die [e:], 9-12 die [i:], 13-16 die [o:] und 17-20 die [o:] dargestellt. Die Achsenbeschriftung wurde zugunsten der Übersichtlichkeit weggelassen.

Deutlich wird auch, daß die Assembly in C oft nicht die 50 Neuronen erreicht, während in D nahezu jeder Auslesevorgang alle 50 Neuronen umfaßt. An einigen Stellen zeigt sich ein falscher Auslesevorgang, diese finden hauptsächlich in den ersten vier Assemblies statt (diese entsprechen auch ein [a:]) oder in Assembly 13 und 17 (das sind die Vokale [o:] und [u:], aber mit ähnlicher Periodizität von ca. 110 Hz).

Um nicht zu jedem Versuch die Protokolle darstellen und analysieren zu müssen, wird im folgenden ein einfacher Algorithmus angewendet, der die Aktivitäten bewertet und ausgibt, welche

Assembly am häufigsten ausgelesen wurde. Dieser kann sowohl für C als auch für D verwendet werden.

- Bestimme zu jedem Zeitpunkt t die Assembly, die die meisten aktiven Neuronen hat: Sei $A_i(t)$ die Aktivität der Assembly i zum Zeitpunkt t, dann ergibt sich die Assemblie A(t) mit den meisten aktiven Neuronen zum Zeitpunkt t zu: $A(t) = max_{i=1}^{20}(A_i(t))$
- Bestimme die Assembly W, die am häufigsten in A vorkommt, unter Berücksichtigung einer Mindestaktivität von 6 Neuronenn.

Die Assembly W ist der Gewinner dieser Simulation. Im optimalen Fall entspricht diese Assembly genau dem angelegten Vokal (also sowohl dem Vokaltyp als auch der Periodizität des Vokals). Im schlechtesten Fall sind sowohl Periodizität als auch Vokaltyp falsch. Dazwischen können die Fälle, daß nur die Periodizität oder nur der Vokaltyp falsch ist, auftreten. Für den Fall, daß sich zwei Menschen unterhalten ist die korrekte Klassifikation in Bezug auf den Vokaltyp ausreichend, wohingegen für musikalische Anwendungen die korrekte Periodizität von größerer Bedeutung sein würde.

Da jede Simulation etwas andere Ergebnisse ergibt (zufällige Initialisierungen, Rauschen) wurden die Klassifikation aller 20 Vokale zehn Mal wiederholt und jeweils der Gewinner bestimmt. Die Ergebnisse sind in Tab. 3.3 zusammengefaßt.

	Period. und Vokal korrekt	nur Period. korrekt	nur Vokal korrekt	beides falsch
Areal C	72,5 %	24 %	0.5 %	3 %
Areal D	85 %	10 %	3 %	2 %

Tabelle 3.3: Ergebnisse der 10 Wiederholungen der Klassifkation aller 20 Vokale.

Man erkennt, daß beide Assoziativespeicher Informationen aus der K Karte extrahieren. In Areal C ist von 96,5% der Vokale die Periodizität korrekt klassifiziert, in Areal D etwas weniger mit 95%. Allerdings ist die Anzahl der korrekt erkannten Vokale mit 73% in Areal C deutlich niedriger als die 88% in D. Das bedeutet, daß der Einsatz der zusätzlichen Schicht D nicht nur die Muster vervollständigt, sondern auch die Klassifikationsleistung erhöht.

	$C_{[a:]}$	$C_{[e:]}$	$C_{[i:]}$	$C_{[o:]}$	$C_{[u:]}$	$D_{[a:]}$	$D_{[e:]}$	$D_{[i:]}$	$D_{[o:]}$	$D_{[u:]}$
Input [a:]	40	0	0	0	0	40	0	0	0	0
Input [e:]	0	17	0	23	0	0	33	1	6	0
Input [i:]	0	1	32	7	0	0	0	37	1	2
Input [o:]	0	0	0	40	0	0	0	0	38	2
Input [u:]	0	1	0	22	17	2	0	0	10	28

Tabelle 3.4: Verwechslungsmatrix der Areale C und D. Jeder Vokal wurde 40 Mal präsentiert (10 Wiederholungen zu je 4 verschiedenen Periodizitäten pro Vokal). $C_{[a:]}$ steht beispielsweise für die Anzahl der Versuche, bei denen eine Assemblie mit einem [a:] gewonnen hat.

Man sieht, daß in C die Vokale [a:] und [o:] fehlerlos klassifiziert wurden, nur das [u:] macht Probleme. Dagegen ist in Areal D das [o:] nicht immer korrekt klassifiziert worden, dafür sind die Vokale [e:], [i:] und [u:] deutlich öfter korrekt klassifiziert. Interessant ist auch, daß der Vokal [e:] in C öfters dem Vokal [o:] zugesprochen wurde als dem korrekten [e:]. Diese Tendenz wurde in D deutlich abgeschwächt, von den 40 Vokalen sind jetzt 33 korrekt klassifiziert (im Gegensatz zu den 17 korrekt klassifizierten Vokalen in C).

Ein zusätzlicher Vorteil der Schicht D geht aus Abb. 3.56 hervor: Dargestellt ist für beide Schichten C und D ein Histogramm der Aktivitäten aller Assemblies. Für jeden Vokal wird zu jedem Zeitschritt eingetragen, wieviele Neuronen aktiv waren. Dies ergibt 200 Histogramme pro Areal, die Summe über diese ist jeweils in Abb. 3.56 dargestellt.

Der Vergleich der Histogramme zeigt, daß in C die Assemblies seltener mit mehr als 20 Neuronen aktiv sind als in D. Die Auswertung dieser Aktivitäten durch weitere biologische Strukturen

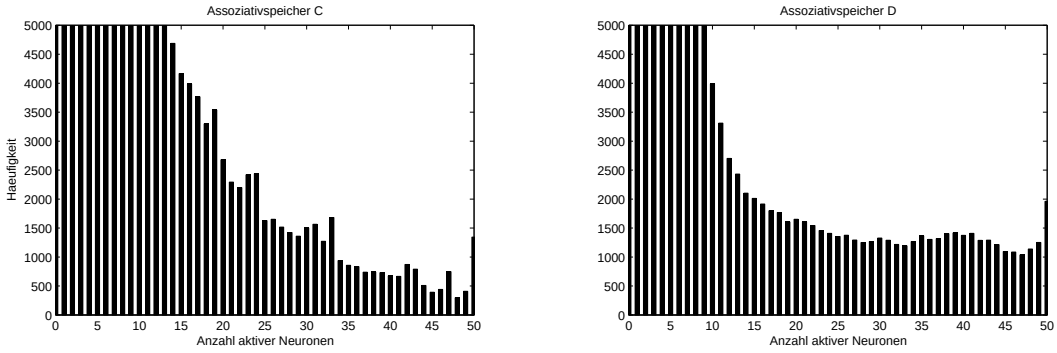

Abbildung 3.56: Darstellung der Aktivitätsverteilung der Assemblies. Die X-Achse gibt an, wieviele Neuronen einer Assembly aktiv waren, die Y-Achse, wie häufig diese Aktivität vorgekommen ist.

wäre demnach anhand von D deutlich einfacher als anhand von C. Auch aus der beispielhaften Darstellung eines Protokolls für das [a:] mit 110 Hz (Abb. 3.55) zeigt sich, daß das Areal D die Ergebnisse stärker repräsentiert als C.

Mit Rückwärtsprojektionen

Nachdem die prinzipielle Funktionsweise der Assoziativspeicher gezeigt wurde, werden die gleichen Simulationen mit aktivierten Rückwärtsprojektionen durchgeführt. Die reinen Klassifikationsleistungen sind in Tab. 3.5 dargestellt.

	Period. und Vokal korrekt	nur Period. korrekt	nur Vokal korrekt	beides falsch
Areal C	81,5 %	14,5 %	1,5 %	2,5 %
Areal D	90 %	6 %	1 %	3 %

Tabelle 3.5: Ergebnisse der 10 Wiederholungen der Klassifikation aller 20 Vokale wenn die Simulation mit aktivierten Rückwärtsprojektionen von D nach C durchgeführt werden.

Man erkennt, daß die Klassifikationsleistung in C deutlich angestiegen ist. Anstatt 72,5 % werden nun 81,5 % der Vokale komplett korrekt erkannt. Auch die Leistung von D steigt etwas an, von 85 % auf 90 %. In Tab. 3.6 sind wieder die Verwechslungsmatrizen angegeben.

	$C_{[a:]}$	$C_{[e:]}$	$C_{[i:]}$	$C_{[o:]}$	$C_{[u:]}$	$D_{[a:]}$	$D_{[e:]}$	$D_{[i:]}$	$D_{[o:]}$	$D_{[u:]}$
Input [a:]	40	0	0	0	0	40	0	0	0	0
Input [e:]	0	30	0	10	0	0	37	3	0	0
Input [i:]	0	1	38	1	0	0	0	38	0	2
Input [o:]	0	0	0	40	0	0	0	0	39	1
Input [u:]	1	0	3	18	18	0	0	3	9	28

Tabelle 3.6: Verwechslungsmatrix der Areale C und D, wenn die Rückwärtsprojektionen von D nach C aktiviert sind.

Man erkennt, daß das [e:] deutlich besser klassifiziert wird, wenn die Rückwärtsprojektionen aktiv sind (in C: 17 korrekte Vokale im Vergleich zu 30 korrekten Vokalen). Dagegen konnte das [u:] nicht wesentlich besser klassifiziert werden (in C: 17 zu 18), es wird häufig mit dem [o:] verwechselt.

Der Einfluß von D nach C wird auch in Abb. 3.57 deutlich, die Aktivitäten der Assemblies in C unterscheiden sich stark von denen in Abb. 3.55, für die keine Rückwärtsprojektionen verwendet wurden. Zum Beispiel erreicht die erste Assembly (diese entspricht dem Stimulus) jetzt fast immer 50 aktive Neuronen, wenn die Rückwärtsprojektionen aktiv sind.

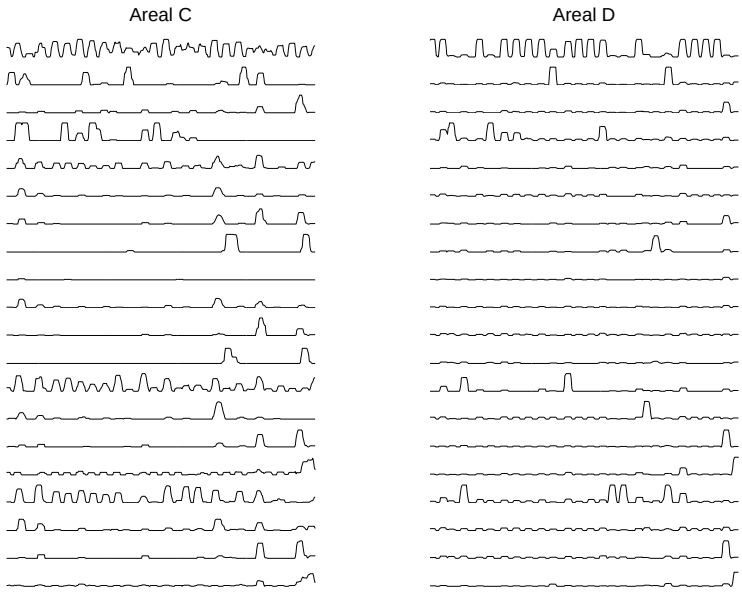

Abbildung 3.57: Beispielprotokoll der Aktivitäten in D und C, wenn der Vokal [a:] mit 110 Hz Periodizität präsentiert wird und die Rückwärtsprojektionen aktiv sind. Jeweils 4 Zeilen entsprechen den Assemblies eines Vokals

Die Tendenz, bei einem Auslesevorgang immer gleich alle Neuronen der Assembly zu aktivieren, zeigt sich auch in Abb. 3.58, die wieder die Aktivitätshistogramme der Assoziativspeicher ausgibt.

Man erkennt, daß ein großer Teil der Aktivitäten 50 Neuronen umfaßt. Bei dem Vergleich mit Abb. 3.56 muß auch die unterschiedliche Skalierung der Y-Achsen beachtet werden. In Areal C sind sogar mehr Neuronen aktiv als in Areal D, das liegt daran, daß Areal D so eingestellt ist, daß im wesentlichen nur eine Assembly aktiv sein kann, während in Areal C auch mehrere Assemblies gleichzeitig aktiv werden können - wenn sie nahe beieinander liegen und so beide den Input von K bekommen. Der Assoziativspeicher D wählt dann den Vokal aus, dessen Spikes zuerst und am häufigsten bei ihm ankommen. Durch die Rückkopplung wird anschließend dieser Vokal komplett aktiviert und die anderen unterdrückt.

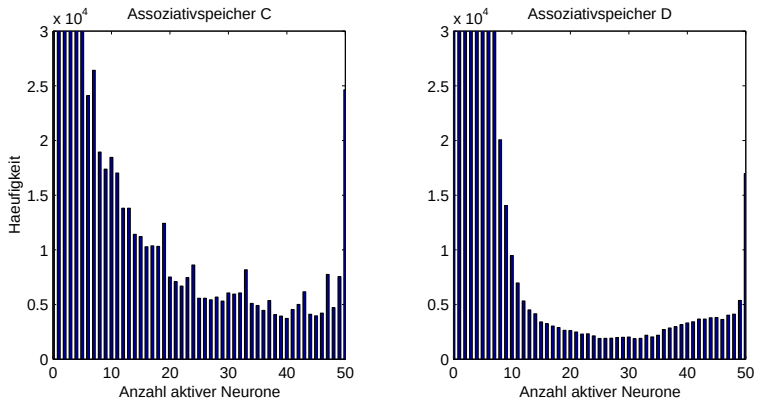

Abbildung 3.58: Darstellung der Aktivitätsverteilung der Assemblies. Die X-Achse gibt an, wieviele Neuronen einer Assembly aktiv waren, die Y-Achse, wie häufig diese Aktivität vorgekommen ist.

Klassifikation von unbekannten Mustern

Nachdem bisher nur die Vokale klassifiziert waren, die als Muster bereits eingespeichert wurden, werden im folgenden Vokale als Stimulus verwendet, die nicht gespeichert sind. Dazu wird ein neuer Datensatz verwendet, der für jeden der 5 Vokale 18 Exemplare bei verschiedenen Periodizitäten zwischen 100 Hz und 200 Hz bereitstellt. Die Vokale sind zwischen 200 ms und 500 ms lang und

	unbekannte Vokale korrekt	bekannte Vokale korrekt
Areal C	58 %	97,5 %
Areal D	57 %	97 %

Tabelle 3.7: Ergebnisse der Simulationen mit dem erweiterten Datensatz.

von demselben Sprecher gesprochen. Als Modell wird die Version mit Rückwärtsprojektionen von D nach C verwendet, da gezeigt wurde, daß diese die Klassifikationsleistung verbessern.

Zunächst werden für alle 18*5=90 Vokale die Muster berechnet, die zum Einspeichern verwendet werden. Dann werden pro Vokaltyp ([a:], [e:], ...) jeweils 4 Exemplare zufällig ausgewählt und in die Assoziativspeicher eingespeichert. In zwei Simulationsdurchläufen werden anschließend die unbekannten (14*5) und danach die bekannten Vokale (4*5) von dem so trainierten Modell klassifiziert und die Antworten jeweils protokolliert.

Um eine aussagekräftige Menge von Ergebnissen zu bekommen wird dieses Vorgehen 30 mal wiederholt. Die Ergebnisse in Tab. 3.7 sind aufgeteilt in Ergebnisse bei Verwendung bekannter Vokale (5*4*30=600 Simulationen) und Ergebnisse bei Verwendung unbekannter Vokale (5*14*30=2100 Simulationen).

Man sieht, daß die Klassifikationsleistung bei bekannten Vokalen nahezu perfekt ist, während die Klassifkationsleistung bei unbekannten Vokalen eher niedrig ist. Allerdings würde ein System, welches nur durch Raten versucht, den Vokal zu klassifizieren, im Schnitt nur 20 % Klassifikationsleistung erreichen. Interessant ist auch, daß in Areal C die Klassifikation ein bißchen besser ist als in Areal D. Für die Klassifikation mit unbekannten Vokalen sind die Verwechslungsmatrizen in Tab. 3.8 angegeben.

	$C_{[a:]}$	$C_{[e:]}$	$C_{[i:]}$	$C_{[o:]}$	$C_{[u:]}$	$D_{[a:]}$	$D_{[e:]}$	$D_{[i:]}$	$D_{[o:]}$	$D_{[u:]}$
Input [a:]	237	25	132	12	14	210	26	170	11	3
Input [e:]	73	181	76	5	85	76	191	91	5	57
Input [i:]	96	24	287	11	2	108	20	282	9	1
Input [o:]	63	7	70	182	98	74	15	62	184	85
Input [u:]	1	28	27	28	335	1	40	23	41	315

Tabelle 3.8: Verwechslungsmatrix der Areale C und D, wenn unbekannte Vokale klassifiziert werden. Pro Vokal wurden 420 Exemplare klassifiziert.

Man sieht, daß besonders [a:] und [i:] oft verwechselt wurden, wohingegen [a:] und [o:] fast nie durcheinandergebracht wurden. Das Ergebnis der Klassifikation hängt natürlich stark von den gespeicherten Mustern ab. Wenn die gespeicherten Muster gleichmäßig über die Periodizität verteilt sind, dann sind die Chancen wesentlich höher, viele unbekannte Muster zu klassifizieren, als wenn die gespeicherten Muster alle bei hohen Periodizitäten liegen und dann die Vokale mit niedrigen Periodizitäten klassifiziert werden sollen. Dementsprechend ist auch die Streuung der Klassifikationsrate sehr groß. In Abb. 3.59 sind die 30 Versuche mit der jeweiligen Trefferquote angegeben.

Man sieht, daß sowohl Klassifikationsleistungen von 100% erreicht wurden wie auch von 7%. Die 100% gehören jedesmal zum [o:]. Für diese wurden die Vokale mit den Nummern 5/10/16/18, 7/12/13/15 und 3/7/13/17 verwendet (die Periodizität steigt mit zunehmender Nummer). Die schlechteste Klassifikationsleistung beim [o:] Vokal wurde dagegen mit den Nummern 3/4/6/8 erreicht (36%). In diesem Fall wurden nur die niedrigen Periodizitäten gelernt, daher sind die [o:] Vokale mit hoher Periodizität sehr schwierig für den Assoziativspeicher.

Die schlechtesten Ergebnisse liefert der Vokal [a:] mit zwei mal 7%. Für diese beiden Durchgänge wurden die [a:] Stimuli mit den Nummern 11/12/13/15 und 4/5/6/10 zum Lernen verwendet. Der Assoziativspeicher kannte also wieder nur einen kleinen Bereich der Periodizitäten, und die [a:] mit abweichender Periodizität wurden falsch klassifiziert.

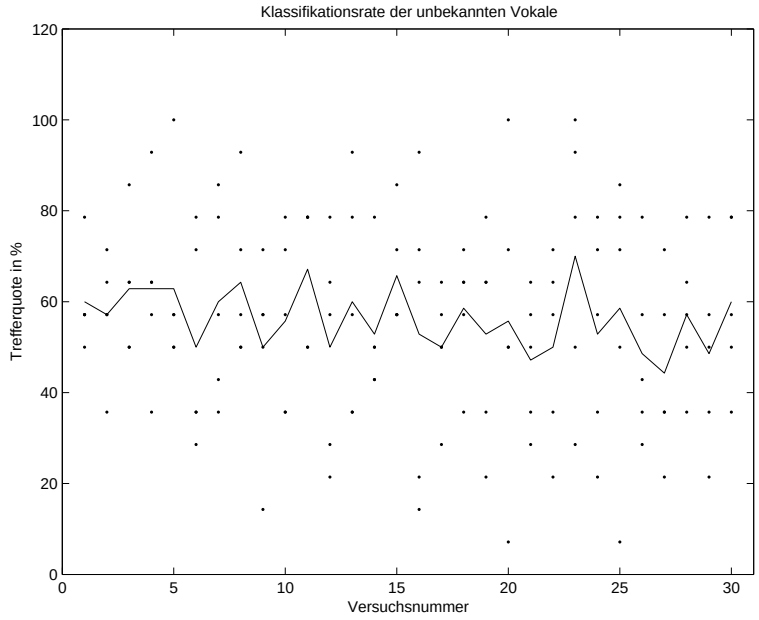

Abbildung 3.59: Darstellung der Trefferquote in Areal D bei der Klassifikation unbekannter Vokale. Jeder Punkt entspricht den 14 Exemplaren eines Vokals. Jede Spalte hat 5 Punkte, teilweise übereinanderliegend. Die durchgezogene Linie ist der Mittelwert der Klassifikationsleistung für jeden Versuch

3.6.4 Separation

Für die Separationsaufgabe wird zunächst ein Beispiel angegeben, wie der Assoziativspeicher auf zwei simultan präsentierte Vokale reagieren kann. Dazu wird ein [a:] mit mittlerer Periodizität mit einem [i:] mit niedriger Periodizität gemischt, und dem Assoziativspeicher präsentiert. Das Ergebnis ist in Abb. 3.60 dargestellt. Man erkennt, daß im wesentlichen die beiden Assemblies [a:] und [i:] mit der entsprechenden Periodizität aktiv waren. Hin und wieder wurden auch andere Assemblies aktiv, aber nie für längere Zeit. Betrachtet man die zeitlichen Verläufe der Aktivierungen genauer, so sieht man auch, daß die beiden Assemblies nicht gleichzeitig aktiv wurden, sondern immer abwechselnd, eine Zeitlang die Assembly für das [a:], dann eine Zeitlang die Assembly für das [i:] [8].

Dieses Verhalten des Assoziativspeichers kann allerdings nur erreicht werden, wenn sich die Muster der einzelnen Stimuli in der Karte linear überlagern. Daß dies im allgemeinen nicht der Fall ist, wurde in Abschnitt 3.2.2 gezeigt. Für das [a:] und das [i:] funktioniert es gut, da die Formantfrequenzen sich nur wenig überlappen. Daher überlagern sich die Muster weitgehendst linear. Im allgemeinen kann jedoch nicht von diesem Verhalten des Assoziativspeichers ausgegangen werden.

Allerdings ist das für den Cocktailparty-Effekt notwendige Verhalten auch ein etwas anderes. Ein Mensch will sich in so einer Situation nicht mehrere Sprecher gleichzeitig anhören, sondern sich auf einen bestimmten konzentrieren. Dies entspricht dem folgenden Versuch: Für jede der vier unterschiedlichen Periodizitäten, für die die Vokale vorhanden sind, wird jeweils ein Vokal verwendet: ein [a:] mit 110 Hz, ein [e:] mit 150 Hz, ein [i:] mit 190 Hz und ein [o:] mit 240 Hz. Diese vier Vokale werden zu einem Stimulus gemischt, und entsprechen in der Cocktailpartysituation vier Sprechern, die gleichzeitig reden. Eine Person, die sich jetzt auf einen von den Sprechern konzentriert, steht nahe bei diesem und schaut ihn normalerweise an. Dadurch ist dessen Stimme im Stimulus stärker repräsentiert. Außerdem versucht er die anderen Stimmen nicht wahrzunehmen. Diese beiden Strategien werden dadurch modelliert, daß zum einem die Lautstärke des Vokals, auf den sich das Modell konzentrieren soll, größer ist als die der restlichen Vokale. Zum anderen werden nur die Vokale mit der gleichen Periodizität in den Assoziativspeicher gespeichert. In die restlichen Assemblies werden Zufallsaktivitäten eingespeichert. Dies wird für jeden der vier gewählten Vokale durchgeführt. Zunächst werden also alle Vokale mit der Periodizität 110 Hz (5 Stück) und weitere 15 Zufallsmuster gespeichert. Als Stimulus wird dann $k*[a:]_{110Hz}+[e:]_{150Hz}+[i:]_{190Hz}+[o:]_{240Hz}$

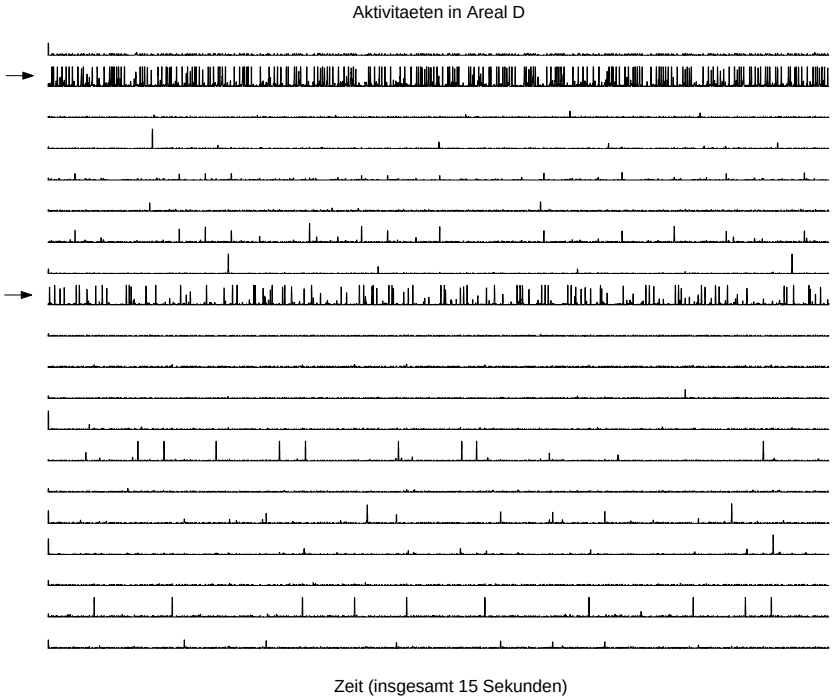

Abbildung 3.60: Darstellung der Aktivitäten der Assemblies in Areal D, wenn als Stimulus eine Mischung aus [a:] und [i:] verwendet wird. Die Pfeile zeigen die zugehörige Assembly. Die Simulation erstreckte sich über einen Zeitraum von 15 Sekunden.

verwendet (k=1, 1.5, 2, ... 4). Der Übersichtlichkeit wegen wird dieser Stimulus zukünftig mit *Stimulus [a:] mit k = ... bezeichnet.* Dieser wird 10 Mal präsentiert und die Aktivitäten in D protokolliert. In Abb. 3.61 ist beispielhaft ein Durchlauf für ein [a:] bei $k = 2$ dargestellt.

Aus der Darstellung in Abb. 3.61 wird nun die Summenaktivität berechnet, indem einfach 'von links nach rechts' für jede Assembly die Anzahl aktiver Neuronen aufsummiert wird. Dadurch ergeben sich pro Simulation 20 Werte anstatt 1200 Werte, die Auswertung und Darstellung wird vereinfacht. Dies wird auch entsprechend mit den anderen drei Vokalen durchgeführt. In Abb. 3.62 sind die Ergebnisse als Aktivitätssumme dargestellt, für jede Assembly und jeden Durchlauf wurde die Summe der Neuronen bestimmt, die in jeder Assembly aktiv waren.

In Abbildung 3.62 können unterschiedliche Ergebnisse gefunden werden: Zunächst erkennt man, daß in der ersten Spalte der Graphen (Ergebnisse für das [a:]) immer die korrekte Assembly die meiste Aktivität zeigt. Für $k = 2$ (entspricht SNR von 0.67, da alle vier Stimulivokale die gleiche Energie haben, und bei $k = 2$ das [a:] zweimal so laut wie die restlichen drei Vokale ist, also sich das Verhältnis zu 2/3 ergibt) sind die Nebenmaxima nahezu nicht mehr vorhanden, der Assoziativspeicher hat also den korrekten Vokal extrahiert. Sowohl für kleinere k als auch für größere k werden die Nebenmaxima größer, aber nie größer als das Maximum an der korrekten Position.

Zusätzlich sieht man, daß für die anderen Vokale die Separationsergebnisse nicht ganz so gut aussehen. Für das [e:] und [i:] sehen die Ergebnisse sehr ähnlich aus, für kleine k sind die Maxima nicht sehr ausgeprägt, erst ab einem SNR von 0.83 (entspricht $k = 2.5$) ergibt sich das Maximm an der korrekten Position. Für kleinere k wird das [e:] mit einem [o:] verwechselt (Maximum bei 14) wohingegen der Stimulus mit [i:] zunächst überhaupt keine Aktivitäten erzeugt.

Das Ergebnis für das [o:] sieht dagegen etwas anders aus. Das korrekte Maximum an der Position 16 bleibt etwa gleich groß. Mit zunehmendem k wird allerdings das Nebenmaximum des [a:] (an Position 4) größer, so daß gegen Ende beide Spitzen fast gleich groß sind. Auf den ersten Blick mag es verwundern, warum die Separation für zunehmendes k schlechter werden kann. Darauf wird in der Diskussion ab Seite 102 noch eingegangen.

Interessant ist auch, daß die Streuung der Ergebnisse mit 10 Wiederholungen nicht besonders groß ist. Der Median (die durchgezogene Linie) repräsentiert die Ergebnisse daher sehr gut, der

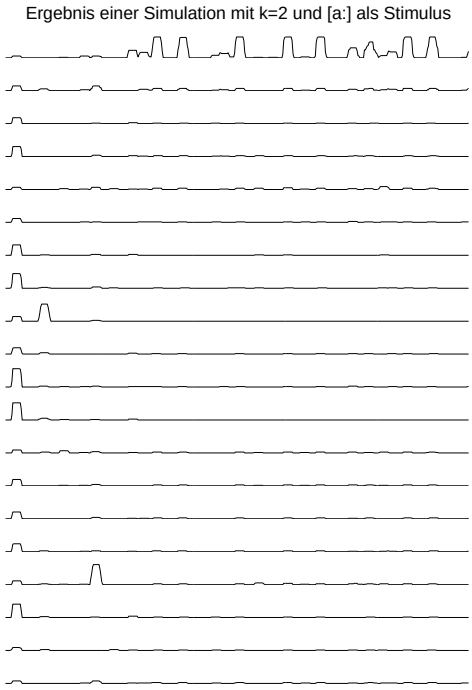

Abbildung 3.61: Ergebnis der Simulation, wenn als Stimulus $2 * [a:]_{110Hz} + [e:]_{150Hz} + [i:]_{190Hz} + [o:]_{240Hz}$ verwendet wird. Man sieht, daß die meiste Zeit die erste Assembly aktiv war, diese entspricht dem [a:] bei 110 Hz. Die geringe Aktivität der vierletzten Zeile entspricht dem [u:] mit 110 Hz Periodizität.

Assoziativspeicher sollte für ähnliche Stimuli auch ähnlich gute Ergebnisse liefern.

Zur weiteren Analyse der Ergebnisse wird aus diesen Aktivitätsgraphen für jede der 20 Assemblies die Gesamtaktivität bestimmt, indem die Anzahl der aktiven Neuronen über die zehn Wiederholungen aufsummiert wird. Für jeden Vokal und jedes k werden dann die zwei am meisten aktivierten Assemblies ausgewählt, und die Werte protokolliert. In Tab. 3.9 sind die Ergebnisse aufgeführt. Die Prozentzahlen geben an, wie stark die zweite Assembly in Bezug auf die erste Assembly aktiv war. Je kleiner die Zahl, desto sicherer war sich der Assoziativspeicher bei der Auswahl des Vokals.

k/SNR	Max. [a:]	2. [a:]	Max. [e:]	2. [e:]	Max. [i:]	2. [i:]	Max. [o:]	2. [o:]
1/0.33	1/23	13/93%	14/30	2/46%	19/12	15/81%	16/24	20/64%
1.5/0.5	1/43	17/23%	14/36	2/41%	19/16	15/69%	16/27	4/51%
2/0.67	1/45	9/19%	14/22	2/92%	11/21	19/87%	16/26	4/68%
2.5/0.83	1/43	9/28%	6/30	2/43%	11/34	19/34%	16/30	4/89%
3/1	1/43	9/35%	6/29	14/50%	11/23	15/68%	16/24	4/87%
3.5/1.17	1/38	9/53%	6/28	14/63%	11/24	15/49%	16/30	4/86%
4/1.33	1/30	9/77%	6/35	2/45%	11/27	15/28%	4/29	16/95%

Tabelle 3.9: Ergebnisse der Separationsaufgabe. Angegeben sind jeweils die zwei Assemblies mit der größten Aktivität. In der Spalte von Max. [a:] ist beispielsweise die Assemblienummer und Summenaktivität (in Tausend) der Assembly angegen, die bei dem Stimulus [a:] jeweils am stärksten aktiv war. In der nächsten Spalte 2. [a:] ist die Assemblynummer angegeben, welche die zweitgrößte Aktivität hat, die Stärke der Aktivität ist in Prozent zur maximalen Aktivität angegeben.

In der Tabelle spiegeln sich die aus den Graphen extrahierten Ergebnisse wieder. Man sieht, daß in der ersten Spalte (Max. [a:]) immer die 1. Assembly - also die von [a:] 110 Hz - am stärksten aktiv war. In der nächsten Spalte sieht man bei $k = 2$, daß das größte Nebenmaximum nur 19% des maximalen Wertes beträgt. Dies ist auch der kleinste Wert des Nebenmaximums überhaupt. Für das [e:] ergibt sich das beste Ergebnis für $k = 2.5$, mit 43%. Die 41%, die für $k = 1.5$ erreicht werden, sind zwar kleiner, aber dafür wurde nicht der korrekte Vokal erkannt. Der

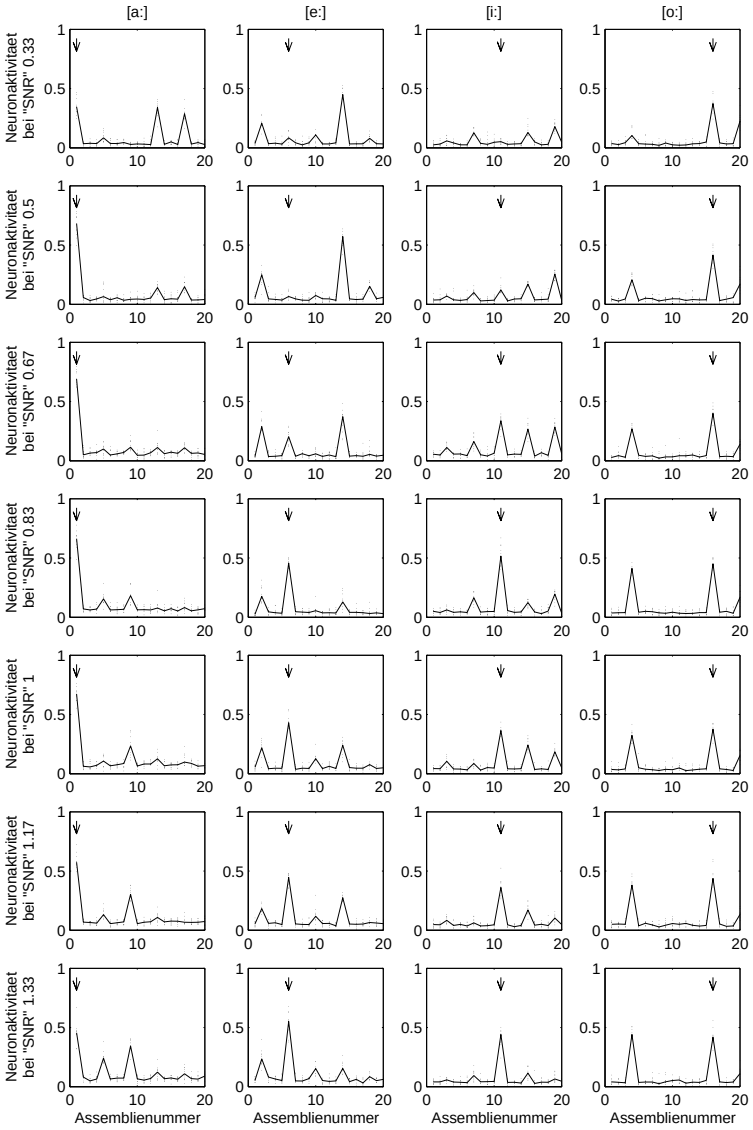

Abbildung 3.62: Ergebnis der zehn Wiederholungen des Separationsversuchs. Jeder Plot zeigt die Aktivitäten für einen Stimulus, die Punkte entsprechen dem einzelnen Versuch, die Linie dem Median der 10 Wiederholungen. Der Pfeil in jedem Graph markiert die Assemblie, die dem Vokal entspricht, dessen Periodizität im Assoziativspeicher gespeichert ist. Alle Graphen einer Spalte haben also die gleichen Muster im Assoziativspeicher gespeichert. Alle Graphen einer Zeile haben das gleiche Verhältnis von gesuchtem Vokal zu störendem Vokalen.

Assoziativspeicher hat sich 'geirrt'. Das [i:] wird bei $k = 4$ mit 28% am besten von den anderen Vokalen getrennt, wobei schon bei $k = 2.5$ 34% erreicht werden. Relativ schlecht schneidet das [o:] ab, bei $k = 1.5$ werden zwar schon 51% erreicht, allerdings wird der Wert mit zunehmenden k schlechter. Der Grund dafür wird in der Diskussion angegeben.

Zusammendfassend kann gesagt werden, daß bereits für $k = 2.5$ - also einem SNR von 0.83 - alle vier Vokale korrekt aus der Mixtur klassifiziert werden. Die Separationsaufgabe oder das Konzentrieren auf einen Sprecher wird also unter der Annahme, daß die Sprecher unterschiedliche Periodizitäten haben, bereits mit einem 'Ohr' gelöst, wenn die Zielperson etwas lauter redet als die anderen Personen.

3.7 Quellenlokalisation anhand der Spikes aus der IC Karte

Die folgenden Ergebnisse basieren auf den Ergebnissen eines in *Biological Cybernetics* veröffentlichten Artikels ([6]). Im Gegensatz zu Modellen, bei denen der Stimulus erst gefiltert wird und dann mit Kreuzkorrelationen aufbereitet wird, um die Schallquelle zu lokalisieren (zum Beispiel

[40]), wird im folgenden nicht direkt der gefilterte Stimulus kreuzkorreliert, sondern die Spikes, die sich auf der IC Karten ergeben.

3.7.1 Ergebnisse unter Verwendung des Langnermodells

Mit Hilfe der drehbaren Achse mit zwei Mikrofonen des Versuchsaufbaus wurde ein kompletter Datensatz mit 74 Stereoaufnahmen von -90 bis +90 erzeugt. Als Stimulus wurde der Vokal [a:] mit ca. 120 Hz Periodizität verwendet. Versuche mit anderen Stimuli (Sinus Signal, Rauschen und AM Sinus) zeigten, daß sich der Vokal am besten eignet. Jede Aufnahme besteht aus einem rechten Kanal und einem linken Kanal. Aus diesen zwei akustischen Signalen wurden unabhängig voneinander mit einem Langnermodell die Koinzidenzkarte berechnet, dies ergibt zwei Funktionen $S_{1/2}(f_C, f_P, t)$ die genau dann eine 1 ergeben, wenn an dem Punkt (f_C, f_P) in der Koinzidenzschicht zum Zeitmoment t ein Spike aufgetreten ist. Dabei stellt f_C die Trägerfrequenz und f_P die Periodizität dar, die Indizes $1/2$ beziehen sich auf das linke und rechte Mikrofon. Diese Funktionen wurden wie folgt weiterverarbeitet:

- Zuerst werden die Spikes entlang der Periodizitätsachse zu einer Funktion $S'_{1/2}(f_C, t) = \sum_i S_{1/2}(f_c, i, t)$ aufsummiert
- Von diesen Funktionen wurde anschließend die Kreuzkorrelation entlang der Zeit $K(f_C, dt)$ berechnet: $K(f_C, dt) = corr_t(S_1(f_C, t), S_2(f_C, t))$
- Diese Funktionen werden aufsummiert zu einer Funktion $E(\delta t) = \sum_i K(i, \delta t)$
- Von dieser Funktion wird die Position des Maximums p bestimmt.

Die Funktion $E(\delta t)$ ist eine Funktion, die zu jedem Zeitversatz δt angibt, wieviele Spikes der Funktionen $S_1/2$ um genau δt entfernt sind. Theoretisch muß diese Funktion ein Maximum an der Stelle besitzen, die dem Zeitversatz der Mikrofonaufnahmen entspricht. Aus diesem Zeitversatz läßt sich auf einfache Weise dann die Orientierung der Mikrofonachse ausrechnen. In der Praxis sind die beiden von den Mikrofonen aufgenommenen Signale nicht einfach nur zeitversetzt. Zum einen ist die Lautstärke der Signale durch die unterschiedliche Orientierung der Mikrofone und ihrer Nierencharakteristik unterschiedlich, zum anderen werden die beiden Signale noch mit Reflektionen (Echos) überlagert. Dadurch ergibt sich ein komplexes Muster, welches in der Realität dazu führt, daß selbst das menschliche Ohr nicht immer genau lokalisieren kann, wo der Ursprung der Signale ist.

Indem für jede der 74 Aufnahmen der beschriebene Algorithmus ausgeführt wird, ergibt sich eine Schar von Werten $P(w)$ für den Parameter p abhängig von der Orientierung w der Mikrofonachse bei der Aufnahme. Diese Funktion $P(w)$ ist in Abb. 3.63 dargestellt.

Man erkennt, daß die Funktion $P(w)$ sehr nahe entlang der Kreuze verläuft. Diese stellen den aus der Orientierung der Mikrofone im Moment der Schallaufnahme berechneten Zeitversatz der Mikrofonaufnahmen dar. Aus den Spikes des Langnermodells läßt sich also mit dem angegebenen Algorithmus die Orientierung der Mikrofonachse berechnen. Das bedeutet, daß nicht nur die Position des Flecks in der Karte das Signal wiederspiegelt, sondern auch genauere Informationen durch die exakte Position der Spikes repräsentiert werden. Interessant ist auch die hohe zeitliche Präzision, die sich ausgehend von den Spikes berechnen läßt. Die Spikes müssen genauer als auf eine Milisekunde erzeugt werden, sonst wären die Korrelationsergebnisse in Abb. 3.63 nicht so exakt entlang der Kreuze. Um diesen Sachverhalt besser beurteilen zu können, ist in Abb. 3.64 die Differenz der beiden Kurven dargestellt. Zusätzlich sind noch zwei Linien für -0.025 ms und +0.025 ms angegeben, diese entsprechen der Genauigkeit, die sich durch die Samplingrate von 40 kHz ergibt.

Man sieht, daß fast alle Punkte zwischen den Linien liegen. Das bedeutet, daß das Langnermodell sehr genaue Ergebnisse liefert, die in der Größenordnung der Samplerate liegt. Der Mittelwert des Betrags der Differenzpunkte liegt bei 0.012 ms, also etwa der Hälfte dessen, was der Samplingrate entsprechen würde. Dieser Aspekt wird in der Diskussion auf Seite 4.3 genauer betrachtet.

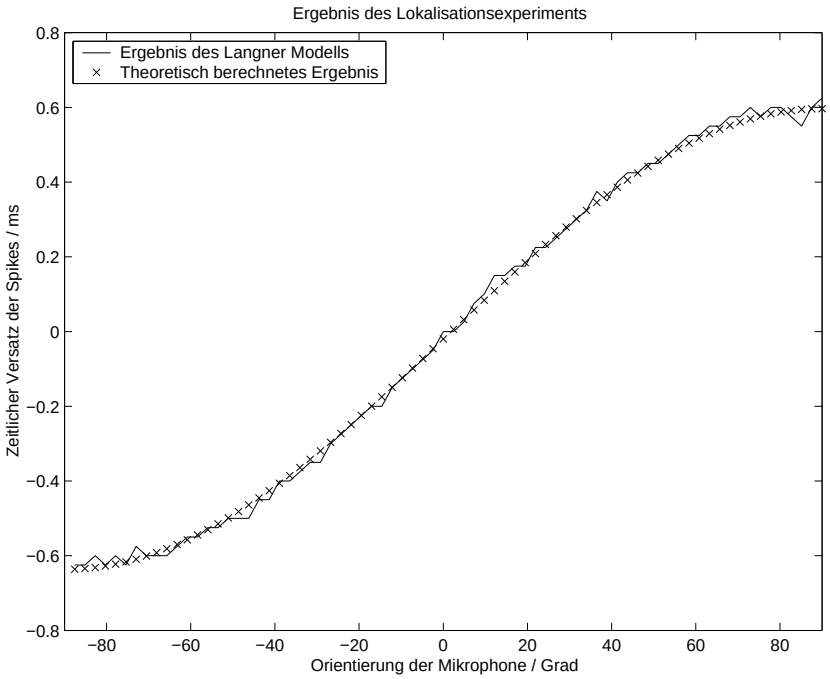

Abbildung 3.63: Darstellung der Ergebnisse des Lokalisationsexperiments. Die durchgezogene Linie stellt der von den Langnermodellen berechnete Zeitversatz der Mikrofonaufnahmen dar, die Kreuze die aus den Orientierungen berechneten Zeitversätze, also die theoretisch korrekten Werte.

3.7.2 Biologische Plausibilität

Unter der Annahme, daß das Lokalisationssystem des Menschen auf einem ähnlichen Mechanismus basiert wie der dargestellte Algorithmus, müßten die Spikes der Koinzidenzschicht an die obere Olive weitergeleitet werden, da diese einen großen Teil der Berechnungen zur Schalllokalisation durchführt [46]. Allerdings kommen die Nervenbahnen zur oberen Olive hauptsächlich aus dem Nucleus Cochlearis, und nicht aus dem Colliculus Inferior. Es sind also Rückprojektionen aus dem IC in niedere Areale notwendig, um einen Lokalisationsalgorithmus zu realisieren, der auf den Spikes des IC basiert. Die neusten Forschungsarbeiten ([32]) zeigen, daß diese Rückprojektionen entgegen früherer Zweifel doch zu existieren scheinen, zumindest sind in Arealen unterhalb des ICs (z.B.: Lemniscus Lateralis) periodotope Karten gefunden worden. Da aber dem Langnermodell entsprechend die notwendigen Informationen für eine periodotope Karte erst im IC berechnet werden, müssen diese wieder in die unteren Areale zurückgereicht werden. Somit könnten innerhalb der oberen Olive tatsächlich Informationen aus dem Colliculus Inferior verwertet werden.

Ein weiterer Hinweis ergibt sich aus einem Artikel [2], der beschreibt, wie bei psychoakustischen Versuchen festgestellt wurde, daß die Lokalisation sich verbessert, wenn bei hohen Frequenzen niedrige Periodizitäten vorhanden sind. Dies bedeutet, daß das Hörsystem anscheinend durchaus die Periodizitätsinformationen zur Lokalisierung von Schallquellen einsetzt.

Natürlich sind auch andere Mechanismen zur Lokalisation vorstellbar, die das Langnermodell nicht verwenden. Der Hauptzweck dieses Versuches ist zu zeigen, daß es prinzipiell möglich ist, die Spikes des Langnermodells für Lokalisationsaufgaben zu verwenden.

3.8 Psychophysik

Im Gegensatz zum Ansatz der Neurophsyiologen, die versuchen, die Funktion einzelner Neuronen im Gehirn zu verstehen, ist die Idee der Psychophysik, das gesamte System auf einmal zu untersuchen. Dazu werden Reize an das System angelegt und geschaut, wie das System reagiert. Im folgenden werden einige Experimente vorgestellt, die teilweise von mir und teilweise von Dritten durchgeführt wurden. Zuerst werden Ergebnisse gezeigt, die darauf hinweisen, daß die Spikesynchronität von Bedeutung für das auditorische System ist. Anschließend werden Ergebnisse darge-

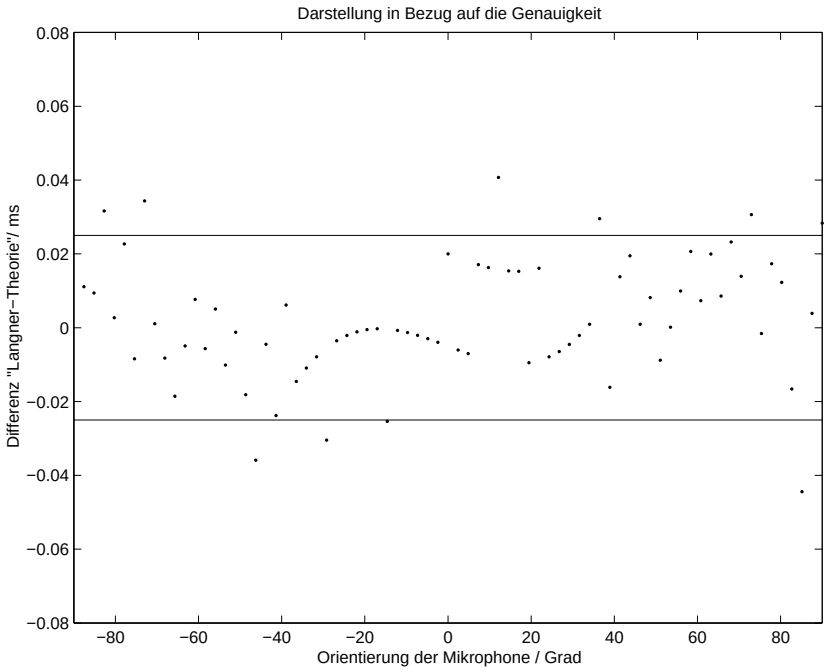

Abbildung 3.64: Darstellung zur Veranschaulichung der Genauigkeit der Lokalisationsergebnisse: Abgebildet ist die Differenz des theoretischen Ergebnisses und des experimentellen Ergebnisses. Die zwei horizontalen Linien entsprechen der Genauigkeit, die sich aus der Samplingrate von 40 kHz ergibt.

stellt, die auf eine Periodizitätskartierung entsprechend dem Langnermodell hindeuten.

3.8.1 Eigene Experimente zum Einfluß von Spike-Synchronität auf die Wahrnehmung

Das im folgenden beschriebene Experiment versucht die Fragestellung zu klären, in wie weit höhere Areale die zeitliche Feinstruktur der Spikes auswerten, oder ob durch die Ortskodierung des Langnermodells die wesentlichen Informationen ausgewertet wurden und somit der Kortex nur auf die Position der Spikes und weniger auf die genauen Zeitpunkte der Spikes achtet. Diese Fragestellung ist nicht nur für das auditorische System von Bedeutung, eine grundsätzliche Frage ist seit langem, ob Spikes mehr Bedeutung beinhalten als einfach eine Ratenkodierung, und wie diese zusätzliche Bedeutung genau aussieht.

Das Experiment läuft wie folgt ab:

1. Der Benutzer hört mit einem Kopfhörer zwei Vokale v_1 und v_2 (basierend auf einem [a:] der länge 0.5 Sekunden mit ca. 110 Hz) hintereinander, mit einer Pause von 200 ms zwischen den Vokalen.

2. Der Benutzer hat drei Felder zur Auswahl:

   - Feld 1: *Erster war lauter*
   - Feld 2: *Etwa gleichlaut*
   - Feld 3: *Zweiter war lauter*

3. Je nachdem, welches Feld angeklickt wird, reagiert der PC mit Verminderung der Lautstärke des ersten Vokals (Feld 1), keine Veränderung (Feld 2) oder Erhöhung der Lautstärke des ersten Vokals (Feld 3). Wenn dreimal hintereinander auf das Feld 2 geklickt wurde, so wird die Lautstärke des ersten Vokals auf einen zufälligen Wert zwischen 50 % und 150 % gesetzt.

Die Vokale v_1 und v_2 unterscheiden sich nur in Bezug auf die Phase in dem Frequenzbereich zwischen 100 Hz und 5000 Hz. Dazu wird v_2 wie folgt aus v_1 mit Hilfe einer zufälligen, aber

festen Verteilung w von 20000 Winkelwerten (als normierte komplexe Zahlen) zwischen $-\pi$ bis $+\pi$ erzeugt:

- Berechnung der Fouriertransformation F_1 von v_1: $F_1 = fft(v_1)$.

  Da v_1 eine Größe von 20000 Samples hat, bekommt man als Ergbenis 20000 komplexe Zahlen, die die Phasenlage und Stärke der Frequenzen von 1 Hz bis 20000 Hz angeben (da die Abtastrate 40 kHz ist).

- Verschiebung der Phasenlage der Frequenzen zwischen 100 Hz und 5000 Hz durch Multiplikation des komplexen Vektors w mit F_1: $F_2 = F_1 * w$.

- Berechnung des Vokals v_2 mit Hilfe der inversen Fouriertransformation: $v_2 = ifft(F_2)$.

  Alle Schritte zusammen können wie folgt ausgedrückt werden: $v_2 = ifft(w * fft(v_1))$

Durch diese Modifikation verändert sich die Energie des Vokals nicht. Allerdings werden die Frequenzbänder gegeneinander verschoben. Dadurch treten die Spikes aus unterschiedlichen Frequenzkanälen nicht mehr gleichzeitig auf. Wenn die kortikalen Areale nur auf die Position der Spikes im IC achten und die Zeitinformationen ignorieren, so müßten die Vokale ungefähr gleich wahrgenommen werden. Stimmt dagegen die These, daß synchrone Spikes eine verstärkte Wahrnehmung hervorrufen, dann sollte der phasenverschobene Vokal im Verhältnis zum ursprünglichen Vokal leiser wahrgenommen werden.

Für die Versuchsreihe wurde der Zufallsvektor w, der die Phasenverschiebung erzeugt, einmalig zufällig festgelegt. Dieser wurde dann für alle Versuche verwendet. Um den Einfluß der Phasenverschiebung genauer festzustellen wurde die Stärke der Phasenverschiebung in sieben Schritten von 0 % auf 100 % erhöht. Zu jeder der sieben Phasenverschiebungen wurden drei Lautstärkenverhältnisse experimentell bestimmt, indem der Stimulus v_2 erzeugt wurde und dann das oben beschriebene Verfahren sechs Mal hintereinander angewendet wurde. Von diesen sechs Werten wird dann der Mittelwert bestimmt.

Die Reihenfolge der insgesamt 21 Versuche wird zufällig gewählt, so daß sich Ermüdungseffekte nicht systematisch in den Ergebnissen niederschlagen können. Ein Versuchsdurchlauf dauert ca. 15-20 Minuten, abhängig von der Geschwindigkeit des Probanden.

Ergebnisse einer einzelnen Person

Zunächst wurde der Versuch von einer Person fünf Mal - verteilt auf mehrere Tage - durchgeführt. Dadurch ergeben sich für jede der sieben Phasenverschiebungen nicht nur drei, sondern 15 Ergebnisse. Daher können diese besser statistisch ausgewertet werden. Die Ergebnisse sind in Abbildung 3.65 dargestellt. Man sieht zu den sieben Phasenverschiebungsschritten die jeweilig subjektiv wahrgenommene Lautstärke. Die Kästen repräsentieren 50% der Datenpunkte, die restlichen Datenpunkte liegen innerhalb der beiden Begrenzungslinien, wobei bei sieben ein einzelner Datenpunkt als Ausreißer markiert ist.

Man erkennt, daß die Ergebnisse sich signifikant unterscheiden, wenn man keine Phasenverschiebung und maximale Phasenverschiebung verwendet. Interessant ist auch, daß bei keiner Phasenverschiebung der Median nicht bei 1 sondern bei ca. 0.96 liegt, obwohl die beiden präsentierten Vokale identisch sind. Insgesamt ergibt sich für den einzelnen Probanden, daß die Lautstärkewahrnehmung von 0.96 auf 0.72 fällt und somit relativ zueinander auf 75% abfällt.

Ergebnisse verschiedener Personen

Der zweite Teil besteht aus Versuchen von 8 weiteren Personen, die jede den Versuch einmal durchführten. Die Ergebnisse sind als einzelne Plots in Abb. 3.66 dargestellt. Man sieht, daß bei den meisten Personen die Lautstärke reduziert wahrgenommen wird, wenn die Phasenverschiebung zunimmt.

Bei sechs der 8 Personen ist der Trend deutlich, daß mit zunehmender Phasenverschiebung eine geringere Lautstärke wahrgenommen wird. Auffällig ist, daß bei kleiner Phasenverschiebung (Position 2) oftmals eine etwas lautere Wahrnehmung als bei keiner Phasenverschiebung zu erkennen

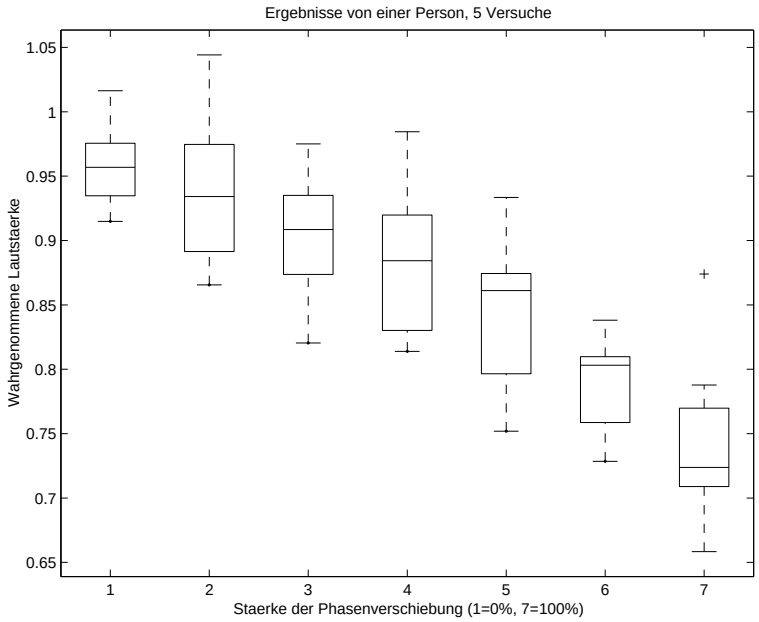

Abbildung 3.65: Die Ergebnisse, wenn eine Versuchsperson den Versuch fünf Mal durchführt. Innerhalb der Kästen liegen 50% der Datenpunkte, der Strich entspricht dem Median. Die Striche oberhalb und unterhalb der Kästen zeigen die am weitesten entfernten Punkte an, das einzelne Kreuz bei sieben markiert einen Ausreißer.

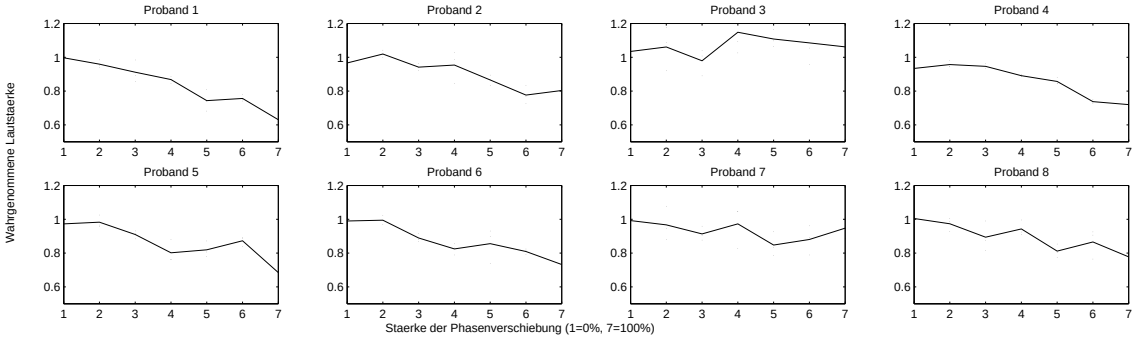

Abbildung 3.66: Ergebnisse von 8 Versuchspersonen. Jeder Graph entspricht einer Person. Auf der x-Achse ist die Stärke der Phasenverschiebung aufgetragen, auf der y-Achse die wahrgenommene Lautstärke.

ist. Dieser Effekt zeigte sich nicht bei einer einzelnen Person und mehreren Versuchen. Teilweise existieren auch Meßpunkte, bei denen die Wahrnehmung wieder lauter geworden ist. Das ist allerdings bei den meisten Personen auf die geringe Datenmenge pro Person zurückzuführen. Das hauptsächliche Resultat ist, daß die Lautstärkewahrnehmung bei maximaler Phasenveschiebung signifikant kleiner ist, als wenn keine Phasenverschiebung verwendet wird. Diese Signifikanz zeigte sich bereits bei mehreren Versuchen einer einzelnen Person. Daß der Effekt nicht von der Person abhängt, sondern bei vielen Probanden auftritt, wird durch diese zusätzlichen Ergebnisse gezeigt. Interessant ist dabei, daß bei zwei der Probanden (Nr. 3 und Nr. 7) die wahrgenommene Lautstärke auf keine erkennbare Weise systematisch von der Phasenverschiebung abhängt. Dieser Aspekt wird ab Seite 105 diskutiert.

Ergebnisse des Langnermodells

Im folgenden werden die 7 phasenverschobenen Vokale als Stimuli für Simulationen mit dem Langnermodell verwendet. Es wurde jeweils die IC Karte und das Autokorrelogramm aller Spikes berechnet. Die Ergebnisse sind in Abb. 3.67 dargestellt.

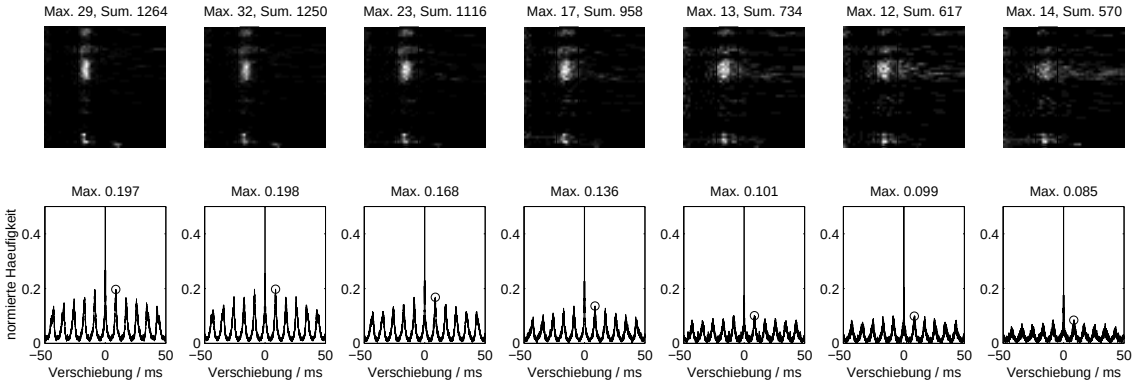

Abbildung 3.67: Ergebnisse des Langnermodells unter Verwendung von normaler Lautstärke, wenn die unterschiedlich stark phasenverschobenen Signale verwendet werden. Obere Zeile: Die IC Karten mit Maximum und Summe der Spikes im relevanten Bereich. Der Bereich der IC Karte, der für die Summe der Spikes verwendet wurde, ist durch den Kasten begrenzt. Untere Zeile: Autokorrelogramme der zeitlichen Spikefunktion. Der Kringel markiert immer die Position des Maximums an der entsprechenden Periodizitätsposition

Man sieht, daß alle extrahierten Werte kleiner werden: Sowohl die Summe der Spikes als auch die maximale Anzahl Spikes und die Synchronität nehmen mit zunehmender Phasenverschiebung ab. Um einen genaueren Eindruck von diesem Effekt zu bekommen, wird für das Langnermodell die Lautstärke von 30% bis 190% in 10% Schritten variiert. Dies ergibt 18 Werte pro Phasenverschiebung, die Ergebnisse sind in Abb. 3.68 als Boxplots dargestellt. Man sieht, daß die maximale Anzahl Spikes (erster Graph) und die Spikesumme (zweiter Graph) nicht besonders stark schwanken, wenn unterschiedliche Lautstärken verwendet werden. Dagegen unterliegt die Höhe des Maximums der Autokorrelogramme (dritter Graph) einer größeren Schwankung. Im Vergleich zu den Langnerergebnissen sind zusätzlich noch die Ergebnisse der einen Person dargestellt, die den Versuch fünf Mal durchgeführt hat.

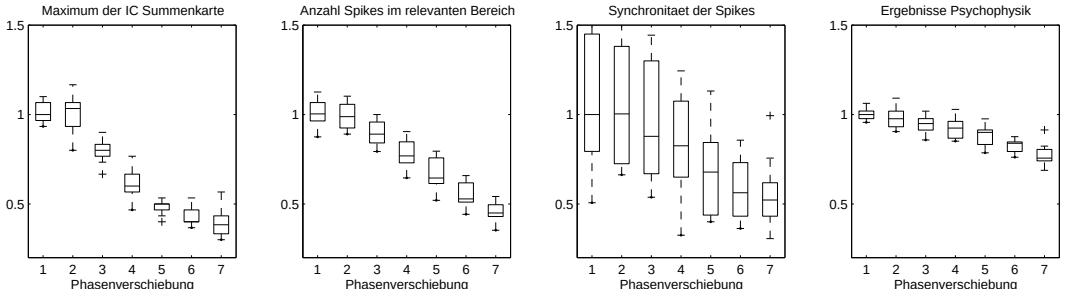

Abbildung 3.68: Darstellung der Effekte von unterschiedlicher Stimuluslautstärke auf die drei extrahierten Werte des Langnermodells (Graphen 1-3). Zum Vergleich sind zusätzlich die Ergebnisse des einzelnen Probanden dargestellt, der den Versuch fünf Mal durchgeführt hat.

Die Ergebnisse machen deutlich, daß der Effekt der reduzierten Lautstärke für das Langnermodell nicht nur auf die reduzierte Synchronität des Stimulus zurückgeführt werden kann. Vielmehr kämen auch die Anzahl der maximalen Spikes und die Summe der Spikes in Frage. Dieser Aspekt wird in der Diskussion ab Seite 105 genauer hinterfragt.

3.8.2 Weitere Experimente in Bezug auf die biologische Plausibilität des Langnermodells

Die Analyse, die durch das Langnermodell durchgeführt wird, ist ähnlich wie Analysen, die auf der Autokorrelation basieren. Man berechnet die Autokorrelation des Stimulus und kann dann die Stärke jeder Periodizität feststellen, indem man an der entsprechenden Stelle im Autokorrelogramm schaut, wie hoch der Wert ist. Da der Mensch in der Lage ist, Periodizitäten zu detektieren, auch wenn die Grundfrequenz nicht im Signal enthalten ist (der sogenannte 'missing fundamental' Effekt) - also nicht nur das Frequenzspektrum analysiert - war eine naheliegende Überlegung, ob die Neuronen vielleicht eine Autokorrelation durchführen. Diese These würde den Effekt der 'missing fundamental' erklären können.

In zwei Artikeln [21, 22] beschreibt Christian Kaernbach mehrere Psychophysikexperimente, deren Ergebnis der These entgegensprechen, daß vom auditorischen System eine Autokorrelation durchgeführt wird. Dazu wurden spezielle Stimuli verwendet, die alle Klicktrains als eine Folge von kurzen Impulsen mit unterschiedlichen Pausen dazwischen darstellen. Die Unterschiede zwischen den Stimuli sind dabei die Länge der Pausen.

1. *k*: Dieser Stimulus besteht aus Impulsen mit konstanten Abstand k.

2. *kxx*: Dieser Stimulus besteht aus einer festen Pause von k ms mit zwei zufälligen Pausen aus dem Interval $[0, 2 * k]$ ms.

   Eine kxx Stimulus mit $k = 10$ ms würde zum Beispiel wie folgt aussehen können: 0, 10, 15, 27, 37, 39, 44, 54, ... Das bedeutet, daß zu den angegebenen Zeiten Impulse vorhanden sind, ansonsten Pausen. Im Schnitt hat der Stimulus für $k = 10$ ms 100 Impulse/Sekunde.

3. *abx*: Dieser Stimulus besteht im Prinzip aus einem Klick, einer Pause von a ms, einem Klick, einer Pause von b ms einem Klick und danach wieder einer zufälligen Pause. Allerdings ist nur a zufällig gewählt, b wird danach so festgelegt, daß $a + b = 2 * k$ (k konstant). Der Wert der zufälligen Pause ist dann aus dem Intervall $[0, k]$ zu wählen.

   Ein abx Stimulus mit $k = 10$ ms würde zum Beispiel 0, 13, 20, 26, 42, 46, 53, ... sein können. Im Schnitt enthält ein abx Stimulus mit $k = 10ms$ 100 Impulse/Sekunde.

4. *x*: Dieser Stimulus ist ein einfacher Klicktrain mit Pausen aus dem Interval $[0, 2 * k]$.

   Ein x Stimulus mit $k = 10$ ms wäre zum Beispiel 0, 12, 18, 26, 45, ... Im Schnitt enthält ein x Stimulus mit $k = 10$ ms auch 100 Impulse/Sekunde

5. *kxxxx*: Dieser Stimulus besteht aus einer festen Pause von k ms mit vier zufälligen Pausen aus dem Interval $[0, 2 * k]$ ms.

   Eine kxxxx Stimulus mit $k = 10$ ms würde zum Beispiel wie folgt aussehen können: 0, 10, 15, 27, 36, 39, 49, 54, ... Im Schnitt hat der Stimulus für $k = 10$ ms 100 Impulse/Sekunde.

Die drei Stimuli sind so konstruiert, daß immer gleichviele Impulse/Sekunde erzeugt werden, wenn die k Parameter gleich gewählt werden. Weiterhin enthalten die Stimuli kxx und abx gleich viele Impulse mit dem Abstand k bzw. $2*k$. Und die Autokorrelogramme sind bei allen drei Stimuli flach bis auf die Verschiebung k bzw. $2 * k$ und die Verschiebung 0, da zwischen den periodischen Abständen der Pausen immer zufällige Abstände liegen. In Abb. 3.69 sind je ein k, kxx, abx und x Stimulus mit Intervallen von 10 ms dargestellt.

Wie man sieht zeigt sich, daß die Maxima bei 10 ms in den Korrelogrammen für abx und kxx ungefähr gleich hoch sind. Ein Algorithmus, der ähnlich zu einer Autokorrelation arbeitet, sollte also beide Stimuli ähnlich stark bewerten. Dagegen sind die Maxima des kxxxx Stimulus nur halb so hoch wie die von abx und kxx. Dieser Stimulus sollte also deutlich kleiner bewertet werden, als die beiden anderen.

Die Untersuchungen von Christian Kaernbach zeigten dagegen, daß das auditorische System des Menschen die kxx/kxxxx Stimuli wesentlich höher bewertet als den abx Stimulus. Bei Experimenten wurden dem Probanden zwei Stimuli präsentiert, einer der beiden war ein reiner Zufallsstimulus (x), der andere war entweder kxxxx oder abx. Der Proband mußte angeben, welcher

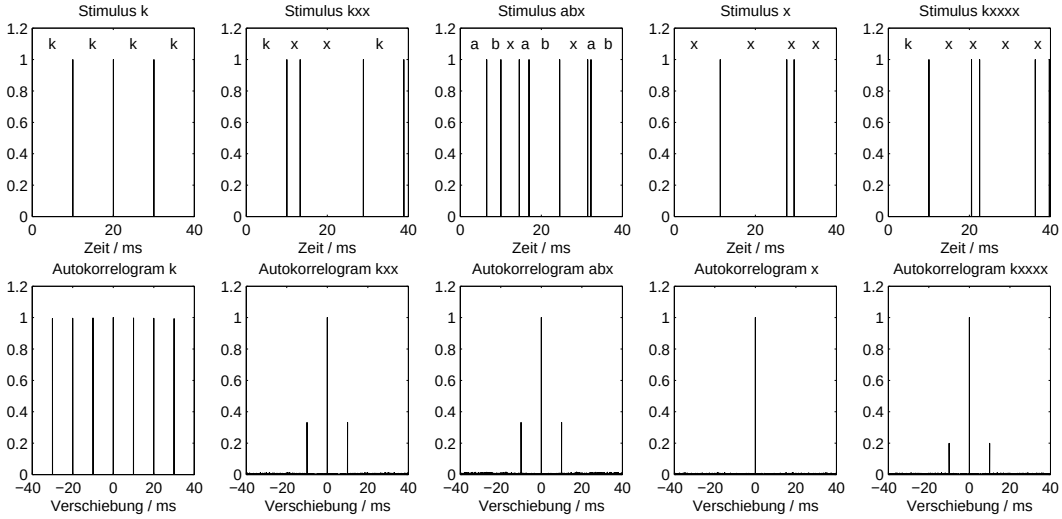

Abbildung 3.69: Obere Zeile: vier Stimuli: k, kxx, abx, x und kxxxx (Intervallbreite 10 ms, das heißt, $k = 10$ ms für k, kxx, x, kxxxx und $k = 5$ ms für abx). Untere Zeile: die zugehörigen Autokorrelationsfunktionen.

der beiden der x Stimulus war. Wurde der kxxxx Stimulus verwendet, so waren im Schnitt ca. 80 % der Antworten korrekt. Bei dem abx Stimulus dagegen wurden nur 60 % korrekt detektiert. Obwohl der kxxxx Stimulus ein deutlich kleineres Maximum in den Autokorellogrammen zeigt als der abx Stimulus, zeigt sich, daß das auditorische System den kxxxx Stimulus deutlich besser vom reinen Zufallsstimulus x unterscheiden kann als den abx Stimulus.

Dies zeigt, daß keine einfache Autokorrelation durchgeführt wird, sondern ein anderer Algorithmus. Die Schlußfolgerungen von Kaernbach sind, daß das Hörsystem sehr wohl Periodizitäten detektieren kann, aber nur wenn sie 'erster Ordnung' sind. Damit meint er solche Periodizitäten, die sich ergeben, wenn man die Abstände zwischen zwei Spikes analysiert. Dagegen werden Periodizitäten 'zweiter Ordnung', also Periodizitäten, die durch Weglassen von jedem zweiten Spike entstehen, nicht besonders stark wahrgenommen. Die Stimuli kxx/kxxxx besitzen Periodizitäten erster Ordnung, der Stimulus abx dagegen nur Periodizitäten zweiter Ordnung (es befindet sich immer genau ein Impuls innerhalb der Impulse, die die Periodizität repräsentieren). Dies zeigt sich, wenn man ein 'InterKlickIntervall'-Histogramm berechnet, also ein Histogramm über die Abstände der Impulse. In Abb. 3.70 sieht man, daß die Stimuli kxx und kxxxx einen deutliches Maximum zeigen, dagegen besitzt abx kein Maximum.

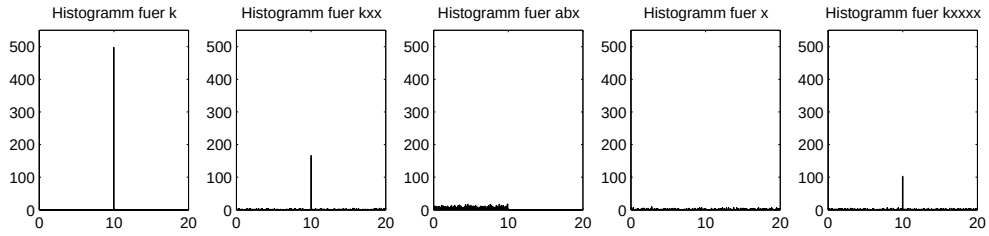

Abbildung 3.70: Analyse der Stimuli in Bezug auf ihre Periodizität 'erster Ordnung' durch Berechnung eines 'InterKlickIntervall'-Histogramm.

Die Ergebnisse aus diesen Untersuchungen werden im folgenden mit Simulationen verglichen, die sich durch Anwendung der Stimuli auf das Langnermodell ergeben. Dazu wurden Stimuli mit

der Länge 5 Sekunden (=200000 Samples) erzeugt, damit sich die Zufallseffekte herausmitteln, und die eigentlichen Ergebnisse deutlich werden.

In Abbildung 3.71 sind die Ergebnisse dargestellt. In der oberen Zeile sind jeweils die IC Karten abgebildet, in der mittleren Zeile die 5. Zeile aus den Karten und in der untersten Zeile die Summe aller Zeilen.

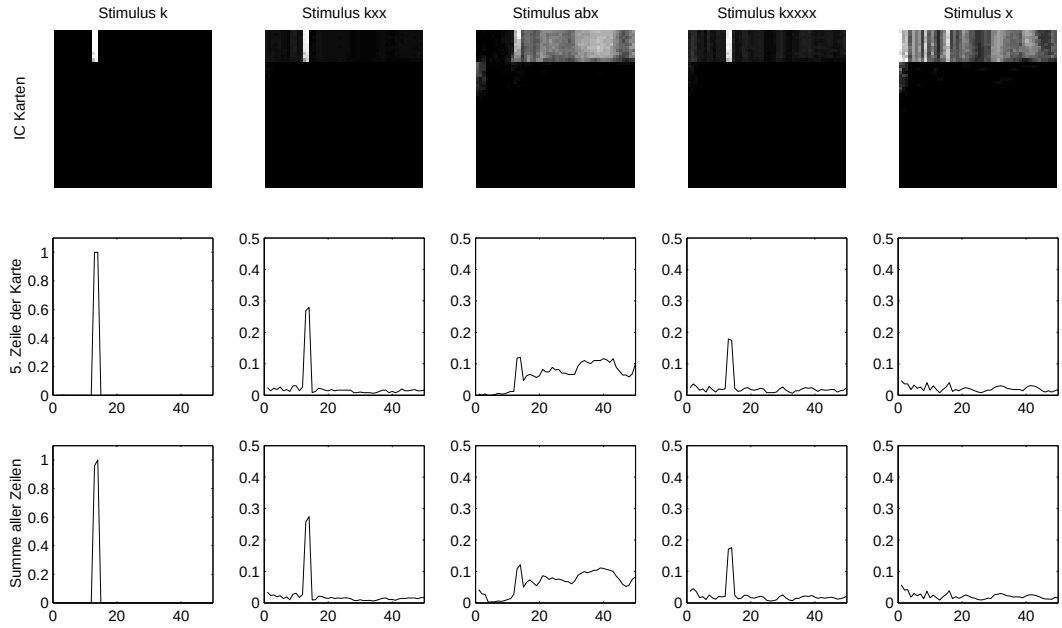

Abbildung 3.71: Darstellung der Ergebnisse, die sich durch Anwendung des variierten Langnermodells auf die obigen Stimuli ergeben. Oberste Zeile: IC Karten. Mittlere Zeile: Jeweils die 5. Zeile der IC Karten. Unterste Zeile: Summe aller Zeilen der IC Karten.

Die Ergebnisse der Simulationen zeigen, daß das Langnermodell bei dem kxx Stimulus die Periodizität deutlich stärker detektiert als bei dem kxxxx Stimulus. Das Verhältnis der Maxima des Summengraphes ist (Max_{kxxxx}/Max_{kxx}) 57 %, und entspricht somit dem Verhältnis der Maxima in den Autokorrelogrammen von 61 %. Das Entscheidende ist aber nicht so sehr das Verhältnis der kxx/kxxxx Stimuli, als vielmehr das Ergebnis unter Verwendung des abx Stimulus. Man sieht in diesem kein ausgeprägtes Maximum bei der Periodizität, sondern ein insgesamt erhöhtes Rauschniveau mit möglicherweise einem kleinen Maximum bei 100 Hz Periodizität. Vergleicht man die Verhältnisse von kxx und abx, dann ist dieses für das Autokorrelogram 100 % und für das Ergebnis der Simulationen 42 %. Absolut betrachtet zeigt sich auch, daß das Maximum des kxxxx Stimulus in den Ergebnissen deutlich größer ist (0.17), als das Maximum des abx Stimulus (0.11). Dagegen sind die Maxima der Autokorrelogramme umgekehrt, für kxxxx ergibt sich 0.19 und abx 0.33.

Insgesamt ähnelt des Ergebnis für den abx Stimulus sehr dem Ergebnis für den Zufallsstimulus x. Ein System, welches auf dem Langnermodell basiert, würde also deutliche Probleme haben, den abx Stimulus vom x Stimulus zu unterscheiden. Dagegen könnte der kxxxx Stimulus auf einfache Weise vom x Stimulus unterschieden werden.

Das bedeutet, daß das Langnermodell deutlich andere Ergebnisse im Vergleich zu der Autokorrelation liefert. Dieses Verhalten entspricht aber genau den Ergebnissen von Christian Kaernbach. Damit ermöglicht das Langnermodell eine präzise Vorhersage dieser Ergebnisse und modeliert im Gegensatz zu Verfahren, die auf Autokorrelation basieren, die Eigenschaften des Hörsystems in dieser Beziehung sehr gut. Weitere Überlegungen zu diesem Zusammenhang werden in der Diskussion im Abschnitt 4.4 auf Seite 105 dargestellt.

Interessant ist an dieser Stelle auch der Vergleich mit dem ursprünglichen Langnermodell. Die Ergebnisse, die sich mit diesem Schaltkreis ergeben, sind in Abb. 3.72 dargestellt. Man erkennt

wieder, daß die Periodizitäten der Stimuli k, kxx und kxxxx deutlich detektiert werden und der Stimulus x zu wenig Aktivität führt. Der große Unterschied ist jetzt, daß der Stimulus abx nicht mehr dem Stimulus x ähnelt, sondern eher dem Stimulus kxx mit etwas Aktivität bei höheren Periodizitäten. Der Grund dafür steckt in den Kammfilterantworten und wird auch in Abschnitt 4.4 genauer diskutiert.

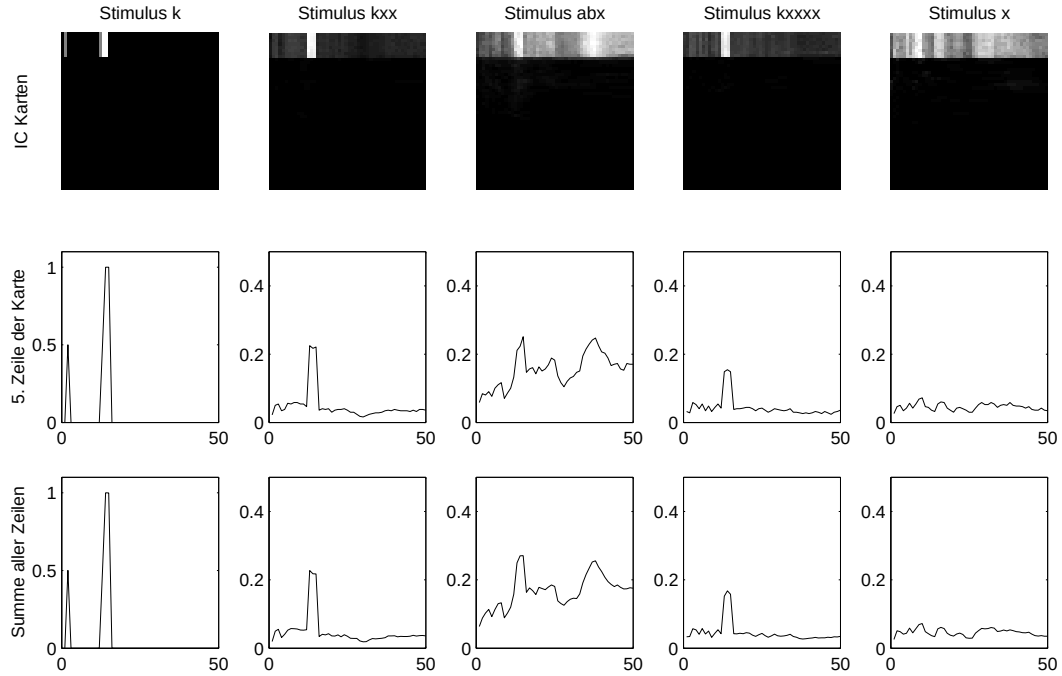

Abbildung 3.72: Ergebnisse unter Verwendung des ursprünglichen Langnermodells. Das Ergebnis für abx ähnelt jetzt dem Ergebnis von kxx

3.8.3 Periodizitätsschwankungen natürlicher Sprache

Ein wichtiger Aspekt der natürlichen Sprache - insbesondere in Bezug auf Quellentrennung - ist die Frage, wie stark die Periodizität von natürlich gesprochenen Worten schwankt. Um einen ersten Eindruck davon zu bekommen, wurde von männlichen Personen der folgendende Text vorgelesen, und die Stimme per Mikrofon aufgenommen: *a, e, i, o, u, ba, be, bi, bo, bu, baba?, bebe?, bibi?, bobo? bubu?*. Aus diesen Aufnahmen wurden per Hand die Vokale extrahiert und die Periodizität bestimmt. Dazu wurde von jeder Aufnahme das Zeitsignal vergrößert auf dem Monitor dargestellt, und mit der Maus eine Periode markiert. Das Ergebnis ist in Abb. 3.73 abgebildet. Jeder Graph repräsentiert einen Sprecher, die Punkte sind den Vokalen nach sortiert. Die ersten vier Punkte sind also die [a:] Vokale, wobei immer der vierte der fragende Vokal (der zweite Vokal von z.B. baba?) ist, und somit im Normalfall eine höhere Periodizität hat als die anderen drei. Diese erhöhte Periodizität zeigt sich allerdings nicht immer in den Vokalen. Über den Graphen ist jeweils der Median, die maximale Abweichung nach oben und die maximale Abweichung nach unten anegegeben. Die horizontalen Linien begrenzen den Bereich, der zwischen einer halben Oktave über dem Median und eine halbe Oktave unter dem Median liegt.

Die fehlenden Punkte bei manchen Graphen (z.B. Sprecher 1, das erste [a:]) ergeben sich dadurch, daß nicht jeder Vokal ausreichend klare Periodizitäten erkennen ließ. Diese wurden nicht gewertet. Man erkennt, daß bei fast allen Vokalen die Periodizität um maximal eine halbe Oktave schwankt. Interessant ist auch, daß die fragenden Vokale keine deutliche Tendenz zu höheren Periodizitäten zeigen. Auch sieht man, daß die Männerstimmen sich im Bereich von 92 Hz bis

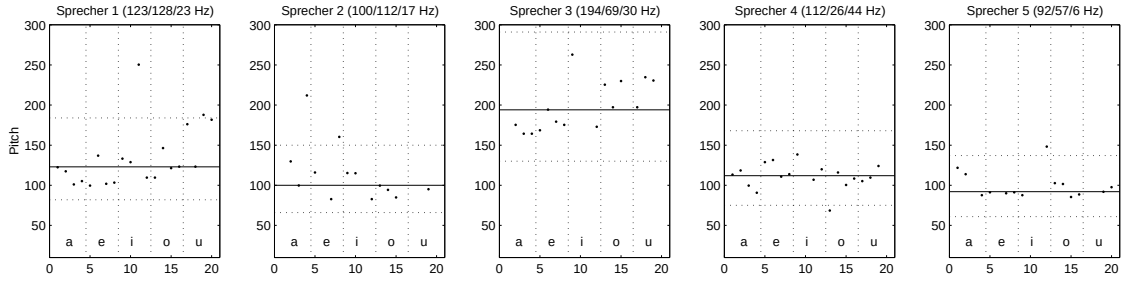

Abbildung 3.73: Darstellung der Periodizitätsschwankungen von fünf männlichen Sprechern. Die Zahlen über den Graphen geben jeweils den Median, die maximale Abweichung nach oben und die maximale Abweichung nach unten (gemessen vom Median) an.

194 Hz bewegen, wobei die 194 Hz eher eine Ausnahme darstellen, die restlichen vier Sprecher schwanken um Periodizitäten von ca. 105 Hz.

Kapitel 4

Diskussion

Im folgenden werden einige der Ergebnisse genauer diskutiert. Der Anfang der Diskussion betrachtet die Fragestellung, welche Vor- und Nachteile das ursprüngliche Langnermodell im Vergleich zum variierten Langnermodell hat. Anschließend werden die Ergebnisse mit kortikalen Arealen diskutiert und danach wird kurz auf das Lokalisationsexperiment eingegangen. Abschließend werden die Psychophysikexperimente ausführlich betrachtet.

4.1 Überlegungen zu den peripheren Arealen (Langnermodell)

Die Existenz einer Periodizitätskarte im Colliculus Inferior ist von vielen Forschern anerkannt. Der Nutzen einer solchen Karte zur Klassifikation und Separation wurde durch die Simulationen belegt. Beide Modelle sind in der Lage, die gewünschte Karte zu erzeugen, allerdings bleibt die Frage offen, welches Modell der Biologie besser entspricht oder ob die biologische Erzeugung der Karte auf komplett anderen Mechanismen beruht. Ein Modell, welches biologisch plausibel ist, muß mit den Resourcen auskommen, die in der Natur vorhanden sind und die biologischen Ergebnisse möglichst gut modellieren. Zunächst werden jeweils die verwendeten Neurontypen mit biologischen Neuronen verglichen, später die teilweise feinen Unterschiede der erzeugten Periodizitätskarte diskutiert.

4.1.1 Zuordnung der Neurontypen

Einige der Neurontypen der beiden Modelle sind identisch: Der Triggerschaltkreis, der Oszillator und das Koinzidenzneuron sind in beiden Modellen gleich aufgebaut. Für diese Neuronen gibt es gute biologische Vorbilder, das Triggerneuron entspricht Onset Zellen im Nucleus Cochlearis, der Oszillatorschaltkreis den sogenannten Chopperzellen im Nucleus Cochlearis - wobei nicht klar ist, ob diese wirklich Spikes im Kreis herum weiterleiten oder einfach nur selbständig Bursts erzeugen - und die Koinzidenzneurone im Colliculus Inferior. Die Modelle unterscheiden sich durch den Teil, der die konstante Verzögerung erzeugt. Im ursprünglichen Langnermodell werden zwei Neuronen als Flip/Flop verschaltet. Deren neuronalen Antwortmuster sind eher schwer zu finden. Das Flopneuron, welches mit einer festen Verzögerung anfängt zu spiken, wäre mit Pauser oder Buildupneuronen im Nucleus Cochlearis gleichzusetzen. Allerdings ist das Antwortverhalten des Flipneurons ungewöhnlich: Es werden erst Spikes erzeugt, doch dann, immer nach einer festen Zeit, muß es aufhören zu spiken. Dieses Verhalten entspricht keinem der bekannten Neurontypen im Nucleus Cochlearis. Auch das Integratorneuron des ursprünglichen Langnermodells ist ungewöhnlich: Es entspricht bis zu einem gewissen Grad einem Buildupneuron, da es anfangs auf den Input überhaupt nicht reagiert, und dann nach einer festen Zeit zu spiken anfängt. Während die biologischen Buildupneurone allerdings weiter Spikes erzeugen, darf das Modell Neuron des ursprünglichen Langnerschaltkreises nur genau einen - oder nur sehr wenige - Spikes erzeugen. Sonst würde das Koinzidenzneuron unabhängig vom Oszillator schon allein durch den Integrator

zum feuern angeregt. Ein Buildupneuron, welches aber nach sehr kurzer Zeit gleich wieder aufhört zu spiken, ist bisher nicht gefunden worden [1].

Dagegen sind die Neurontypen für den variierten Schaltkreis nicht übermäßig ungewöhnlich. Anstatt durch den Flip/ Flop/ Integrator-Schaltkreises wird die Verzögerung durch den IntegStop/ Integrator/ Inhib-Schaltkreis erzeugt. Diese Neuronen zeigen Antwortmuster, wie sie im Nucleus Cochlearis gefunden werden können. Das IntegStopneuron entspricht im wesentlichen einer Onsetzelle, die stark hemmend auf den Integrator wirkt. Der Integrator zeigt ein Antwortverhalten wie ein Buildupneuron, und unterscheidet sich insbesondere vom ursprünglichen Integrator dadurch, daß nicht nur ein Spike erzeugt wird, sondern nach einer Pause eine beliebig lange Zeit Spikes erzeugt werden. Dieses Muster entspricht genau einem biologischem Buildupneuron, welches erst nach einer Pause anfängt, Spikes zu erzeugen und dann bis zum Ende der Periode oder des Stimulus Spikes erzeugt. Als letzter Neurontyp wird das Inhibneuron benötigt, welches vom Integrator Input bekommt und das Koinzidenzneuron hemmt. Dieser Neurontyp könnte einfach im Colliculus Inferior vorhanden sein oder im lateralen Lemniscus liegen, und dami auch die starke Hemmung erklären, welche aus dem lateralen Lemniscus in den Colliculus Inferior projeziert wird. Beide Überlegungen wären biologisch plausibel, da sowohl Projektionen vom Nucleus Cochlearis in den lateralen Lemniscus existieren als auch inhibitorische Neuronen innerhalb des ICs.

Zusammengefaßt scheinen die Variation des Langnerschaltkreises den in der Realität vorhandenen Neuronen eher zu entsprechen als der ursprüngliche Schaltkreis. Natürlich kann man aus der Tatsache, daß man einen bestimmten Neurontyp noch nicht gefunden hat, nicht schließen, daß er nicht existiert. Dennoch erscheint ein Modell, welches nur bereits bekannte Neurontypen verwendet plausibler als ein Modell, welches nicht so ohne weiteres auf die bisher bekannten Neurontypen abgebildet werden kann.

4.1.2 Funktionelle Differenzen der beiden Modelle

Durch die Modifikationen des ursprünglichen Langner Schaltkreises ergeben sich einige Änderungen im Verhalten des Koinzidenzneurons. Die zwei wesentlichen Unterschiede sind, daß vom variierten Schaltkreis keine Kammfilter erzeugt werden und daß das Koinzidenzneuron des variierten Schaltkreises zu jeder Periode einen Spike erzeugen kann. Zunächst wird der letztere Punkt genauer betrachtet:

Unterschiedliche Konfiguration des Flip/Flop Schaltkreises

Wie auf Seite 28 gezeigt wurde, hängt das Verhalten des ursprünglichen Schaltkreises stark davon ab, ob sich das Flipneuron oder das Flopneuron durchsetzt, wenn beide gleichzeitig durch das Triggerneuron bzw. den Integrator aktiviert werden sollen. Wenn dadurch das Flopneuron aktiviert wird, so wird der Integrator solange gehemmt, bis die nächste Periode anfängt und das Flipneuron durch das Triggerneuron aktiviert wird. Setzt sich dagegen das Flipneuron gleich durch, so kann zu dieser Periode bereits wieder ein Spike des Koinzidenzneurons erzeugt werden. Im ersten Fall erzeugt der Schaltkreis nur Spikes zu jeder zweiten Periode, im zweiten Fall zu jeder Periode (vgl. Abb. 3.24 auf Seite 42). Allerdings ist es nicht so, daß das Flipneuron immer gewinnt, wenn der Schaltkreis so eingestellt ist, daß es gewinnen kann. Falls die Periode des Signals etwas kürzer ist als die Zeitverzögerung des Integrators, kann trotzdem ein Spike vom Koinzidenzneuron erzeugt, und das Flopneuron aktiviert werden. In diesem Fall erzeugt der Schaltkreis dann nur die Hälfte der möglichen Spikes. Generell ist aber ein Schaltkreis vorzuziehen, welcher nach Möglichkeit immer zu jeder Periode Spikes erzeugt, wenn sie der eingestellen Periodizität des Schaltkreises entspricht, und sonst überhaupt keine. Ein Schaltkreis, der im einen Fall doppelt soviele Spikes erzeugt als in einem anderen Fall, ist dagegen schwerer von einem höheren Areal auszuwerten. Außerdem erscheint es sinnvoll, wenn ein Schaltkreis, der zum Beispiel auf 100 Hz Periodizität eingestellt ist, auch mit 100 Hz Spikes erzeugt, und nicht mit der halben Frequenz, also 50 Hz.

In dieser Hinsicht ist der variierte Schaltkreis dem ursprünglichen funktionell überlegen, da er immer bei jeder Periode feuert, wenn die Periodizität des Stimulus mit seiner bevorzugten Periodizität übereinstimmt. In Abb. 3.24 erkennt man auch, daß der variierte Schaltkreis ein

wesentlich schärferes Tuning in Bezug auf die Periodizität zeigt. Während die Anzahl der Spikes beim ursprünglichen Modell gemächlich steigt, um dann nach dem Maximum wieder gemächlich abzufallen, zeigt die Kurve des variierten Langnermodells einen schnellen Anstieg, um dann auf dem Plateau, das der maximalen Anzahl Spikes entspricht, zu verharren. Danach fällt die Kurve wieder steil auf Null. Im Prinzip kann nicht allgemeingültig gesagt werden, daß die steile Kurve zu bevorzugen ist. Es gibt durchaus Anwendungen, bei denen eine Art gaussglocken ähnliche Repräsentation des Antwortverhaltens wünschenswert ist. Im Falle des Langnerschaltkreises ist dies allerdings nicht so. Zum einen können hier die einzelnen Langnerschaltkreise, die im IC zu einer Karte konvergieren, so eingestellt werden, daß keine Lücken entstehen, und sich jeder Schaltkreis wirklich auf seine Periodizität konzentrieren kann. Zum anderen sollen die Spikes phasensynchron zum Stimulus sein, und wenn kein scharfes Tuning der Antwortkurve existiert, so fehlen oft Spikes, wenn der Stimulus nicht genau die Mitte des Antwortbereiches trifft. Dadurch wird der Stimulus aber nur unzureichend repräsentiert.

Eine Überlegung, welche das halbierte Antwortverhalten des ursprünglichen Langnerschaltkreises als nicht so problematisch hinstellt, wäre, daß im biologischen System mehrere Schaltkreise zu einer Periodizitäts/Frequenzpaarung existieren könnten, so daß der eine Schaltkreis dann feuert, wenn der andere Pausen macht. Dies würde dazu führen, daß in der Summe bei jeder Periode ein Spike erzeugt werden würde. Allerdings würde dies einen deutlich erhöhten Aufwand bedeuten, da man für jedes Koinzidenzneuron mindestens den Verzögerungsteil benötigt. Würden mehrere Koinzidenzneurone den gleichen Integrator verwenden, so würden auch beide natürlich wieder auf die gleiche Periodizität spiken bzw. keine Spikes erzeugen. Dieser Mehraufwand ist aber unnötig, wenn der variierte Langnerschaltkreis verwendet wird, da dieser bereits sowieso zu jeder Periode Spikes erzeugt.

Zusammengefaßt läßt sich sagen, daß das Antwortverhalten des variierten Modells in Bezug auf die Anzahl der erzeugten Spikes günstiger erscheint, als das Verhalten des ursprünglichen Modells.

Überlegungen zu Kammfilterantworten

Als Kammfilterantworten werden die Spikes bezeichnet, welche der Schaltkreis nicht bei der Grundfrequenz, sondern bei einem vielfachen der Grundfrequenz erzeugt. Ein ursprünglicher Langnerschaltkreis, der auf 50 Hz eingestellt ist, antwortet auch auf 100 Hz, 150 Hz, 200 Hz, usw. (vgl. Abb. 3.7 auf Seite 31). Dagegen zeigt der variierte Langnerschaltkreis keine Kammfilterantworten, wenn ein reiner amplitudenmodulierter Sinus verwendet wird (siehe Abb. 3.19 auf Seite 39). Allerdings ist nicht klar, welches der beiden Verhalten dem biologischem Vorbild entspricht. Die Periodizitätstuningkurven der biologischen Neuronen aus Seite 53 zeigen keine größeren Kammfilteranteile, teilweise könnte man Kammfilter bei der doppelten Frequenz finden, aber keines der Neuronen zeigt Antworten bei dreifacher oder größerer Periodizität. Dies deutet eher darauf hin, daß der biologische Mechanismus keine Kammfilterantworten produziert.

Einige beispielhafte Untersuchungen von Michael Ochse zeigen, daß, wenn das Zeitfenster für die Erzeugung der Periodizitäts/Aktivitäts-Graphen systematisch verkleinert wird, die Kammfilterantworten deutlicher werden können. Die Verkleinerung des Zeitfensters erfolgt, indem der vordere Startpunkt festgehalten wird (z.b. 10 ms nach der Onset Antwort) und das hintere Ende nach vorne geschoben (z.B von 200 ms bis 40 ms nach der Onset Antwort). Dies funktioniert allerdings nicht bei allen Neuronen, sondern nur bei einer kleinen Teilmenge. Die Schlußfolgerungen von Michael Ochse ist, daß dem biologischen Mechanismus das ursprüngliche Langnermodell zugrunde liegt, und ein inhibitorischer Mechanismus aus dem lateralen Lemniscus die Koinzidenzneuronen an der Erzeugung der Kammfilterantworten hindert. Dies wäre eine Erklärungsmöglichkeit für das Verhalten des biologischen Vorbildes. Der Trick dabei ist, daß der Inhibitionsmechanismus erst mit Verspätung einsetzt, und somit am Anfang - kurz nach dem Onset die Inhibition - noch nicht wirksam ist. Daher können hier noch Kammfilter auftreten. Später allerdings wirkt die Inhibition, und es treten keine Kammfilterantworten mehr auf.

Diese Theorie stützt sich hauptsächlich auf die Beobachtung, daß bei kleinerem Zeitfenster die Kammfilterantworten stärker werden können. Leider wurde bisher noch kein quantitatives Kriterium zur Beurteilung der Kammfilterantworten entwickelt, welches auf verschiedene Szenarien

angesetzt werden könnte. Zum Beispiel könnte der Effekt der verstärkten Kammfilterantworten auch dadurch begründet werden, daß die Kammfilter bei Verwendung kleiner Zeitfenster eher auftreten als bei großen Zeitfenstern, aber unabhängig von der Position des Zeitfensters. Diese These könnte gegen die erste These geprüft werden, wenn man ein analytisches Maß für die Stärke der Kammfilterantworten hätte. Ohne ein solches Maß dagegen kann jeder die Graphen anschauen und das herauslesen, was ihm gerade geeignet erscheint.

Im Gegensatz zu den simulierten Modellen feuern die echten Neuronen im Colliculus Inferior nicht immer phasensynchron zum Stimulus. Teilweise fehlen Spikes, teilweise werden Spikes erzeugt, die nicht phasengekoppelt zum Stimulus sind. Dennoch zeigen die MoPha Bilder in Abb. 3.41 auf Seite 56 teilweise deutliche Aktivitätsflecken (am deutlichsten erkennt man das am 6. Neuron von links), die nur dadurch entstehen können, daß die Spikes synchron zum Stimulus erzeugt werden. Diese Flecken, die durch phasensynchrone, periodizitätsabhängige Spikes erzeugt werden, treten hauptsächlich für den tonischen Teil des Stimulus auf, also in der Zeit kurz nach dem Onset des Stimulus bis zum Stimulusende. Daß vor dem Stimulus keine Spikekonzentrationen zu erkennen sind, zeigt, daß die Darstellung durchaus aussagekräftig ist, da zu dieser Zeit die Neuronen zufällige Spikes erzeugen. Dagegen könnten die Spikes, die nach dem Stimulus erzeugt werden, durchaus noch phasensynchron erzeugt werden, da der Schaltkreis durch den Stimulus synchronisiert wurde. Diese vorhergehende Synchronität könnte das zufällige Feuern des Schaltkreises beinflussen. Aus den MoPha Bildern geht aber hervor, daß diese Spikes nicht synchron sind.

Für diese Spikes gibt es zwei unterschiedliche mögliche Quellen: Es handelt sich um Spontanaktivität der IC Neuronen oder sie bekommen tatsächlich Spikes von den darunter liegenden Neuronen - z.B. aus dem Nucleus Cochlearis. Im zweiten Falle würde dies bedeuten, daß die Neuronen im Nucleus Cochlearis unsynchrone Spontantaktivität aufweisen, die aber öfter gleichzeitig auftreten, so daß das Koinzidenzneuron den gleichen Input erhält als ob der Stimulus noch anliegen würde. Der erste Fall dagegen würde einfache Spontantaktivität des Koinzidenzneurons bedeuten. Wahrscheinlich treten beide Fälle auf, mit den vorhandenen Daten kann nicht zugeordnet werden, wodurch der Spike ausgelöst wurde. Wichtig ist aber die Beobachtung, daß die IC Neuronen Spontanaktivität aufweisen. Um die Antwort auf einen Stimulus korrekt zu messen, müßten die Spikes der Spontanaktivität von den gemessenen Spikes abgezogen werden. Dies ist aber nicht für jeden einzelnen Spike möglich, da nicht klar ist, ob der Spike durch den Stimulus erzeugt wurde oder durch die Spontanaktivität. Natürlich könnte man für die Berechnung von Periodizitätstuningskurven (wie in Abb.3.38 auf Seite 53) die durchschnittliche Spontanaktivität von der durchschnittlichen Stimulusaktivität abziehen. Allerdings würde sich dadurch die Kurve nur um eine Konstante nach unten verschieben. Aus der Beobachtung, daß die Spikes während der tonischen Stimuluspräsentation phasensynchron zum Stimulus erzeugt werden, kann aber ein Filter konstruiert werden, der zumindest versucht, die Spontanaktivitätsspikes von den durch den Stimulus erzeugten Spikes zu trennen. Dazu werden einfach nur die Spikes berücksichtigt, die zeitlich ungefähr in dem Moment auftreten, der aufgrund der Synchronitätsüberlegung für stimulusabhängige Spikes in Frage kommt. Würden die Neuronen nur abhängig von der Periodizität des Stimulus und nicht auch phasensynchron zum Stimulus feuern, also sich nur die Anzahl der Spikes bei einer bestimmten Periodizität erhöhen (bei gleichverteilten Spikepositionen), so würde sich die Form der Periodizitätstuningkurve nicht verändern, nur die Anzahl der berücksichtigten Spikes würde reduziert werden. Sind dagegen die Spikes immer synchron zum Stimulus - unabhängig von dessen Periodizität, so würde sich die Periodizitätstuningkurve auch nicht stärker verändern, bis auf eine evtl. leichte Verschärfung der Antworten.

Das genaue Verfahren ist in Abschnitt 3.4.6 beschrieben, in Abb. 3.43 auf Seite 57 sind die Ergebnisse dargestellt. Man erkennt, daß die Neuronen teilweise extrem starke Kammfilterantworten aufweisen. Vergleicht man die Ergebnisse mit den Periodizitätstuningkurven in Abb. 3.38, so erkennt man deutliche Unterschiede. Zum Beispiel weist Neuron 3 extreme Kammfilterantworten auf, die in der ungefilterten Periodizitätstuningkurve nicht zu erkennen waren. Anscheinend erhöht bei diesem Neuron der Stimulus die Gesamtaktivität (daher ist das Grundniveau der Tuningkurve recht hoch), aber nur für bestimmte Periodizitäten werden phasensynchrone Spikes erzeugt. Das Neuron erzeugt die Kammfilterantworten also nicht durch die Anzahl der Spikes, sondern

durch die Synchronität zum Stimulus. Diese Kammfilterantwort existiert insbesondere nicht nur am Anfang des Stimulus, sondern ist generell während des kompletten Stimulus vorhanden. Dies entspricht nicht der Theorie, daß ein Inhibitionsmechanismus nach kurzer Zeit alle Kammfilterantworten unterdrückt. Diese besondere Art von Kammfilterantworten deutet auf einen Mechanismus hin, der Spikes mit einer bestimmten Periodizität erzeugen will. Dadurch entstehen die Maxima, wenn diese Periodizität auf den Stimulus synchronisiert werden kann, bzw. die Minima, wenn diese Schaltkreisperiodizität überhaupt nicht zu der Periodizität des Stimulus paßt.

Eine wichtige Beobachtung bei Neuron 3 ist auch, daß die Kammfilterantworten nicht mit zunehmender Periodizität größer werden, sondern auf einem konstanten Niveau bleiben. Dies bedeutet, daß der Schaltkreis nicht bei jeder Periode des Stimulus Spikes erzeugen will, sondern nur bei seiner Grundperiodizität. Ein Schaltkreis, welcher zum Beispiel alle Vielfachen von 50 Hz detektieren soll, würde bei 50 Hz 50 Spikes/s erzeugen, bei 100 Hz 100 Spikes/s und so weiter. Dagegen erzeugt ein Schaltkreis, der auf 50 Hz eingestellt ist und Kammfilterantworten erzeugt, immer maximal 50 Spikes/s, da die Pause des Integrators 20 ms entspricht, und dazwischen das Koinzidenzneuron nicht feuert. Es werden also Perioden ausgelassen. Diese Beobachtung stützt die These, daß ein Modell ähnlich zum Langnermodell die Periodzitätsanalyse vornimmt.

Aus welchem Grund dieses Modell die Kammfilterantworten erzeugt, ist schwer zu beweisen. Die Simulationen zeigen, daß das variierte Langnermodell auch Kammfilterantworten erzeugen kann, wenn entweder der Stimulus verrauscht ist, die Lautstärke des Stimulus ungünstig gewählt ist oder der Triggerschaltkreis gestört wird, wie zum Beispiel durch Hemmung (vgl. Abschnitt 3.3.1 auf Seite 49). Generell unterliegt ein biologisches Neuron stärkeren Schwankungen (in Bezug auf das Neuronpotential, Laufzeiten und ähnlichem) als die simulierten Neuronen. Daher wäre es nicht überraschend, wenn das biologische Triggerneuron, welches in der Realität auch in einem großen Lautstärkebereich arbeiten muß, nicht immer hundertprozentig korrekt arbeitet, sondern auch mal Fehler macht, und einen oder wenige Spikes aussetzt. Dadurch wären Kammfilterantworten nicht der Teil eines komplexen Modells, der dem variierten Modell hinzugefügt werden muß, sondern ein notwendiges Übel der Realität, welches jedem Modell zugrunde liegt, welches auf Triggerneuronantworten basiert.

Modifikation des Integratoreingangssignals

Entsprechend dem Langnermodell soll der Integrator das akustische Signal direkt aufintegrieren. In den vorgestellten Simulationen wurde dies nur in einem Falle gemacht, um zu zeigen, welche Effekte diese Verschaltung für die Karte im Colliculus Inferior mit sich bringen. Ansonsten wurde ein konstantes Signal aufintegriert, um eine konstante Verzögerung sicherzustellen.

In Abb. 3.16 auf Seite 37 ist dargestellt, wie sich die Karte verändert, wenn die Lautstärke verändert wird. Man erkennt, daß die Aktivität auf der Karte sich abhängig von der Lautstärke verschiebt. Vergleicht man die Aktivitätsflecken von 75% Amplitude mit 150% Amplitude, so sieht man, daß ersterer eher einem schrägen Dreieck ähnelt, während letzterer bananenförmig ist. Insgesamt verschiebt sich die Aktivität der Karte nach links. Dies entspricht den folgenden theoretischen Überlegungen: Ein Integratorneuron, welches bei 100% Amplitude 10 ms benötigt, bis es feuert, braucht bei verringerter Lautstärke länger, weil das Neuronpotential die Schwelle nicht so schnell erreichen kann. Wird der Stimulus dagegen lauter, so ist die Integration schneller, und der Integrator feuert früher. Wird ein Stimulus mit 100 Hz Periodizität verwendet, so feuert dieses Neuron bei 100% Amplitude. Ist die Amplitude kleiner, so verschiebt sich das Periodizitätstuning aller Neuronen, sie benötigen jetzt länger bis zum Feuern. Das bedeutet, daß ein Neuron, welches vorher bei beispielsweise 110 Hz gefeuert hat, jetzt bei 100 Hz feuert. Wird die Amplitude dagegen größer, so benötigen die Neuronen weniger Zeit bis zum Feuern. Ein Schaltkreis, welcher bisher auf beispielsweise 90 Hz geantwortet hat, feuert jetzt bei der Stimulusperiodizität von 100 Hz. Dieses Muster - zunehmende Amplitude läßt die Aktivitäten nach links wandern, abnehmende Amplitude nach rechts - ist auch deutlich in den Ergebnissen auf Seite 37 zu erkennen.

Zum Vergleich wurde die gleiche Simulation ein zweites Mal durchgeführt, allerdings wurde diesmal der Integrator mit einem konstanten Stimulus verbunden. Vergleicht man Abb. 3.16 mit Abb. 3.17, so werden zwei Unterschiede deutlich: Zum einen verändern sich die Aktivitätsmuster

nicht - die Amplitude des Stimulus hat keinen Einfluß mehr auf das Periodizitätstuning. Zum anderen sehen die Aktivitätsflecken nicht mehr aus wie schräge Dreiecke oder krumme Bananen, sondern sind vertikale Balken. Dies entspricht dem Stimulus, der in allen Frequenzkanälen die gleiche Periodizität aufweist, und daher in der Karte ungefähr die gleiche Aktivität in allen Frequenzbändern erzeugen sollte. Wird der Stimulus dagegen mit dem Integrator verbunden, so erzeugen die - schon für jeden einzelnen Stimulus - unterschiedlich lauten Frequenzkanäle ein abweichendes Periodizitätstuning. Somit erzeugen gleiche Periodizitäten nicht mehr an den gleichen Stellen innerhalb der Karte Aktivitätsmuster, sondern jedes Frequenzband hat sein eigenes, gegen die anderen Frequenzbänder verschobenes Periodizitätstuning. Wäre diese Verschiebung konstant, könnte sie durch geeignete Initialisierung ausgeglichen werden. Da das Periodizitätstuning aber vom Stimulus abhängt, kann die Karte nicht so intialisiert werden, daß zu allen Stimuli die Aktivitätsflecken mit gleicher Periodizität in der Karte übereinanderliegen.

Unter der Annahme, daß der Mensch tatsächlich das Signal des Hörnervs mit dem Integrator aufintegriert, würde nicht nur beispielsweise die Tonhöhe eines Klaviers sich ändern, wenn die Lautstärke sich ändert. Vielmehr würde ein Ton, der in mehreren Frequenzkanälen Aktivitäten erzeugt, durch die unterschiedlichen Positionen der Aktivitäten in der Karte ungewöhnlich wahrgenommen werden, in manchen Frequenzbänder würden 100 Hz Periodizität detektiert, in anderen dagegen vielleicht 110 Hz, obwohl der Stimulus nur eine Periodizität von vielleicht 105 Hz besitzt.

Interessanterweise gibt es allerdings tatsächlich den Effekt, daß die Lautstärke eines Tones die Tonhöhe etwas beeinflussen kann [47]. Dieser ist sehr gering, aber meßbar. In der Realität könnte also eine Mischung aus beiden Möglichkeiten vorhanden sein, oder die Amplitude des Stimulus könnte die Integrationsdauer des Integratorneurons auf eine indirekte Weise beeinflussen. Dieser Effekt scheint aber deutlich kleiner zu sein, als die Ergebnisse der Simulationen in Abb. 3.16 zeigen. Für die weiteren Untersuchungen wurde daher einfach der Integrator mit einem konstanten Eingangssignal verschaltet. Dies vereinfacht auch die Aufgabe von höheren Arealen, da ein Punkt in der Karte in dieser Konfiguration genau einer Periodizität entspricht, und nicht unterschiedlichen, abhängig von der Amplitude des Stimulus.

4.1.3 Nutzen des Oszillators

Unabhängig davon, ob der ursprüngliche oder der variierte Langnerschaltkreis verwendet wird, beinhalten beide Modelle den Oszillatorschaltkreis. Für ein einfacheres Verständnis könnte man den Oszillator auch weglassen und die Spikes des Triggerneurons direkt an das Koinzidenzneuron leiten. Die einzige Funktion des Oszillators besteht in den implementierten Modellen momentan darin, den einen Spike des Triggerneurons zu vervielfachen, so daß ungefähr 3 Spikes beim Koinzidenzneuron ankommen. Man könnte meinen, daß man genausogut die Verbindungsstärke vom Oszillator zum Koinzidenzneuron verdreifachen kann, und dann dafür nur einen Spike verwenden. Würde dies funktionieren, so wäre der Oszillator tatsächlich sinnlos. Allerdings zeigt sich, daß die exakte Funktion des Langner Schaltkreises durchaus von der Verwendung des Oszillators abhängt. Zuzüglich zur Spikeanzahl spielt das Zeitverhalten eine wichtige Rolle. Ein einzelner Spike des Oszillators erfordert ein wesentlich kleineres Zeitfenster, in dem der Integratorspike eintreffen kann, um das Koinzidenzneuron zum Feuern zu bringen. Werden dagegen viele Spikes auf eine längere Zeit verteilt, so kann der Integrator Spike in einem größerem Zeitbereich eintreffen, um das Koinzidenzneuron zum feuern zu bringen. Die Anzahl der Oszillatorspikes beeinflußt also die Breite der Periodizitätstuningkurve, da ein breiteres Zeitfenster dafür sorgt, daß ein größerer Periodizitätsbereich zu Koinzidenzneuronspikes führt. Bei wenigen Spikes ist die Periodizitätstuningkurve also schmaler, bei vielen Spikes ist sie breiter.

Wie bereits gezeigt wurde, wird die Periodizitätstuningkurve auch mit zunehmender Periodizität breiter, da ein Zeitfenster von beispielsweise 1 ms bei niedrigen Periodizitäten nur kleine prozentuale Abweichungen erlaubt, bei großen Periodizitäten dagegen größere prozentuale Abweichungen. Um ein gleichmäßiges Antwortverhalten zu erreichen, könnte jetzt die Anzahl Spikes des Oszillators entsprechend angepaßt werden - bei Schaltkreisen für kleine Periodizitäten werden mehr Oszillatorspikes erzeugt, bei Schaltkreisen für größere Periodizitäten werden weniger Oszillatorspikes erzeugt. Dadurch könnte dem Verbreiterungseffekt entgegen gewirkt werden. Damit

dies beim variierten Langnerschaltkreis zuverlässig funktioniert, muß die Laufzeit der Spikes vom Integratorneuron zum Koinzidenzneuron ausreichend hoch sein. Um dies zu begründen, wird folgender Fall betrachtet: Ein Schaltkreis, der auf Periodizitäten von 10 ms eingestellt ist, soll durch entsprechend langes Feuern des Oszillatorschaltkreises auch noch Spikes erzeugen, wenn eine Periodizität von 9 ms anliegt. Dies bedeutet, daß der Oszillator ca. 1 ms lang Spikes erzeugen muß. Nehmen wir an, die Spikes vom Integrator zum Koinzidenzneuron benötigen 0,5 ms. Dann wäre der Integrator so eingestellt, daß er nach einem Spike des Triggerneurons 9,5 ms aufintegriert, bevor er einen Spike erzeugt. Der Oszillator würde jetzt nach 9 ms für 1 ms lang feuern, und das Koinzdenzneuron würde in dem Moment, wo der Spike des Integrators eintreffen würde, bereits ein angehobenes Neuronpotential haben und durch den Spike des Integrators zum Feuern gebracht werden. Das Problem ist jetzt, daß aber dieser überhaupt nicht erzeugt wird, weil der Integrator durch den Triggerimpuls bereits zurückgesetzt wurde, bevor er einen Spike erzeugen konnte. Würde dagegen der Spike des Integrators zum Koinzidenzneuron eine Laufzeit von 1,1 ms haben, so müßte das Integratorneuron bereits nach 8,9 ms feuern, um insgesamt eine Periodizität von 10 ms zu detektieren. In diesem Fall würde der Schaltkreis auch eine Periodizität von 9 ms detektieren, da der Integrator bereits nach 8,9 ms gefeuert hat, und das Zurücksetzen des Integrators durch den Triggerimpuls auf diesen Spike keinen Einfluß mehr hat. Der Oszillator würde dann 1 ms lang feuern, und somit in dem Moment, wo der Spike des Integrators eintrifft (10 ms nach dem ersten Triggerneuron Spike) immer noch feuern. Dadurch ist das Koinzidenzneuronpotential bereits angehoben, und es würde tatsächlich feuern.

Anstatt nur die Laufzeit zwischen Integrator und Koinzidenzneuron zu vergrößern, könnte man auch die Zeitdauer erhöhen, die zwischen einem Triggerimpuls und dem Zurücksetzen des Integrators vergeht. Dies hätte ebenso den gewünschten Effekt.

4.1.4 Problematik der nichtlinearen Überlagerung von Mustern

Wie auf Seite 49, Abb. 3.34 dargestellt ist, ergibt die Verwendung eines Stimulus, der eine Überlagerung zweier Signale darstellt, nicht einfach die lineare Überlagerung der einzelnen Antwortkarten, sondern ein komplexeres Muster. Eine einfache lineare Überlagerung wäre ideal für die Trennung von mehreren, gleichzeitig aktiven Schallquellen, man könnte im Prinzip direkt aus der IC Karte erkennen, welche Stimuli momentan verwendet werden. So schön dieses Verhalten auch wäre, es entspricht leider nicht dem, was das Langnermodell leistet. Dies liegt aber nicht daran, daß das Modell nicht korrekt funktioniert, es handelt sich vielmehr um das Problem, daß die Mischung der Signale die ursprünglichen Signale nicht vollständig repräsentiert. Wenn nur eine Aufnahme eines gemischten Signales vorliegt, so gibt es theoretisch unendlich viele Paare von Originalsignalen, die diese Mischung erzeugen könnten. Das Langnermodell kann also nicht wissen, welche Signale zu dem Stimulus zusammengemischt wurden. Es wird also genau die Karte erzeugt, die dem gemischten Signal entspricht, und nicht die, die der Summe der einzelnen Signale entspricht. Das Triggerneuron erzeugt natürlich im gemischten Fall auch einfach Spikes, wenn ein entsprechender Anstieg des Signals vorhanden ist. Im Prinzip sind im optimalen Fall die Spikes des Triggerneurons durchaus die lineare Überlagerung der Triggerneuronspikes der einzelnen Signale. Allerdings kann der Rest des Modells nicht unterscheiden, welche Spikes von welchem Modell erzeugt werden. Würden zum Beispiel zwei Signale mit 100 Hz und 166 Hz Periodizität zusammengemischt, so wäre der erste Spike von beiden Signalen zusammen, der zweite Spike (nach 6 ms) von dem 166 Hz Signal, der dritte Spike von dem 100 Hz Signal (nach 10 ms) und die Spikes vier und fünf von dem 166 Hz Signal (12 ms und 18 ms). Das ist dem Rest des Schaltkreises aber unbekannt, für diesen sieht es aus wie ein Stimulus, der die Spikes zu den Zeitpunkten 0 ms, 6 ms, 10 ms, 12 ms, 18 ms, ... erzeugt. Dies würde zu Aktivitäten bei 166 Hz, 250 Hz, 500 Hz, 166 Hz, ... führen - immer die Differenz zwischen zwei Spikes würde einer Antwort entsprechen. Und damit repräsentiert das Langnermodell den Stimulus durchaus korrekt, da er eben diese Periodizitäten enthält. Ohne weitere Informationen kann überhaupt keine lineare Überlagerung der einzelnen Antworten entstehen. Um diese in unserem Fall zu erzeugen, wäre es notwendig, daß der Langner Schaltkreis auch Periodizitäten mit größerer Ordnung, also unter Auslassung von Spikes, detektiert. Dies entspricht aber nicht den Ergebnissen der Psychophysik, die in Abschnitt 3.8.2 auf Seite 80 dargestellt

sind.

Dennoch können Muster unter Umständen auf der Karte ausreichend gut repräsentiert werden. Zum einen ist es möglich, daß die Periodizitäten bei unterschiedlichen Trägerfrequenzen unterschiedlich laut sind. Dann würde sich in den einen Trägerfrequenzen die Periodizität des ersten Signals durchsetzen, und in den anderen Trägerfrequenzen die Periodizität des zweiten Signals. Das Ergebnis würde dann zwar nicht mehr in allen Frequenzkanälen der linearen Überlagerung entsprechen, aber die Aktivitäten der Karte würden dennoch vernünftige Rückschlüsse auf die einzelnen Signale erlauben. Durch Mustervervollständigung und Inhibition könnten höhere Areale sich dann auf das gewünschte Signal konzentrieren.

Zum anderen könnten Aussetzer des Triggerneurons dazu führen, daß sich teilweise doch eine Überlagerung der beiden ursprünglichen Signale ergibt. Über lange Zeit betrachtet kann die Chance steigen, daß Fehler des Triggerneurons dazu führen, daß sich die Aktivitätsflecken der ursprünglichen Signale stärker zeigen, als die Aktivitäten, die durch Fehler entstehen, aber nicht in den Originalstimuli enthalten sind. In dem vorherigen Beispiel mit Periodizitäten von 6 ms und 10 ms würde zum Beispiel das Auslassen jedes zweiten Triggerneuronspikes folgendes Spikemuster ergeben (Angaben in ms): *0*, 6, *10*, 12, *18*, 20, *24*, 30, *36*, 40, *42*, 48, *50*, ... Würden jetzt nur die kursiven Momente gewertet werden, so ergäben sich Aktiviten bei 10 ms, 8 ms, 6 ms, 6 ms, 8 ms, ... Hier wurde die Aktivität zu 40 % bei 6 ms und 8 ms liegen und zu 20 % bei 10 ms. Die 8 ms sind natürlich nicht in den Originalsignalen enthalten, das Modell würde sich diese 'einbilden'. Bei einem anderem Beispiel würden dagegen die ursprünglichen Bestandteile deutlicher repräsentiert werden: Wenn 7 ms und 9 ms Periodiztät gewählt werden würde, so ergäbe sich folgendes Ergebnis: *0*, 7, *9*, 14, *18*, 21, *27*, 28, *35*, 36, *42*, 45, *49*, 56, *63*, 65, *70*, 74, *77*, 81, *84*, 90, *91*, 98, *99*, 105, *108*, 112, *117*, 119, *126*,... . Würde jeder zweite ausgelassen, so würden sich die folgenden Periodizitäten ergeben: 9, 9, 9, 8, 7, 7, 14, 7, 7, 7, 7, 8, 9, 9, 9... . Die Aktivität würde also zu 40 % bei 7 ms und 9 ms liegen und nur zu 13 % auf der 8 und zu 7% auf der 14. In diesem Falle würde durch systematische Fehler des Triggerneurons die Überlagerung ungefähr linear verlaufen. Natürlich würde ein biologisches Neuron nicht exakt jeden zweiten Impuls auslassen, aber auf lange Zeit gemittelt können sich die Effekte durchaus ähnlich positiv addieren.

Diese Überlegung ist allerdings nur begrenzt sinnvoll. Zum einem würde ein Triggerneuron, welches nur in 50% der Fälle korrekt feuert, bei ungestörten Stimuli zu viele Fehler in der Karte erzeugen. Zum anderen haben Experimente aus der Psychophysik gezeigt, daß das menschliche Hörsystem nicht zuviele Impulse ausläßt - dieser Punkt wird später (siehe Abschnitt 4.4.2 auf Seite 106 genauer diskutiert. Allerdings ist die Überlegung, daß Fehler des Triggerneurons durchaus positive Effekte haben können, prinzipiell interessant.

4.1.5 Aus dem Langnermodell ableitbare Karten im IC

Wie in Abschnitt 3.5 ab Seite 54 dargestellt wird, lassen sich anhand der Periodizitäts/Trägerfrequenzkarte möglicherweise noch andere Kartierungen ableiten. Diese Ergebnisse werden im folgenden genauer diskutiert.

Lautstärkenkartierung

Die in Abb. 3.44 auf Seite 58 dargestellen Ergebnisse, die anhand mehrerer Stimuli mit unterschiedlicher Lautstärke berechnet wurden, zeigen, daß sich ein Aktivitätsfleck in der Mitte der Karte mit zunehmender Lautstärke teilt und die beiden Teilflecken nach außen wandern. Um diesen Effekt genauer zu untersuchen, wurde die Amplitude von 0% bis 1000% in 50 Schritten variiert und jeweils die Position der beiden Flecken protokolliert. Daraus ergibt sich ein Graph, der in Abb. 3.45 auf Seite 58 abgebildet ist. Man erkennt, daß die Flecken bei kleinen Amplituden schneller auseinanderlaufen als bei großen. Woher aber kommt der Effekt, daß mit zunehmender Lautstärke zwei Flecken nach außen wandern? Die Antwort ist eigentlich ganz einfach: Der Triggerschaltkreis erzeugt Spikes, wenn der Stimulus in einem bestimmten Lautstärkebereich moduliert ist. Ist die Lautstärke zu niedrig, werden keine Spikes erzeugt, und ist sie zu hoch, auch nicht. Dieses Verhalten entspricht einem Triggerneuron, welches bei einem amplituden modulierten Signal zu jeder

Modulationsperiode einen Spike erzeugt: Nur in dem Moment, in dem die Amplitude eine feste Schwelle überschreitet, wird ein Spike erzeugt. Für einen einzelnen Schaltkreis bedeutet dies, daß bei geringer Amplitude keine Spikes erzeugt werden, dann mit zunehmender Amplitude zunächst normal gefeuert wird, bis die Amplitude zu groß wird, und dann wieder keine Spikes erzeugt werden, wenn die Amplitude nie kleiner wird als der Schwellwert.

Für die Karte bedeutet dies zunächst, daß am Anfang keine Aktivität zu erkennen ist. Dann jedoch ist die Amplitude des passenden Filterkanals so groß, daß in dieser Zeile der Karte der Triggerschaltkreis Spikes erzeugt, und der Schaltkreis mit der passenden Periode einen Aktivitätsfleck in der Karte erzeugt. Diese Aktivität hört aber auf, wenn die Amplitude noch größer wird, und der Triggerschaltkreis keine weiteren Spikes mehr erzeugt. Damit ist geklärt, warum die Aktivität in der Mitte zum Erliegen kommt. Die beiden Flecken dagegen, die um diese Position der ursprünglichen Aktivität herum entstehen, ergeben sich dadurch, daß die Filterkanäle nicht extrem scharfe Bandpaßcharakteristik haben, sondern auch für Frequenzen neben der eigentlichen Mittelfrequenz entsprechende Antworten erzeugen, nur mit verringerter Amplitude. Das bedeutet, daß ein amplitudenmodulierter Stimulus nicht nur Aktivitäten in dem Filterkanal erzeugt, dessen Mittelfrequenz der Trägerfrequenz des Stimulus entspricht, sondern auch in den benachbarten Kanälen. Diese Signale sind allerdings einiges leiser als das Signal in dem Filterkanal, der am besten zu der Trägerfrequenz paßt. Daher erzeugen die Triggerneurone der umliegenden Zeilen in der Karte keine Spikes. Wird die Lautstärke jetzt erhöht, so sind die Aktivitäten in den Filterkanälen, die nicht exakt auf die Trägerfrequenz passen, irgendwann so groß, daß deren Triggerneurone auch anfangen, Spikes zu erzeugen. Somit fangen mit zunehmender Lautstärke auch die benachbarten Zeilen der Karte an, Aktivitäten zu zeigen. Da die Filtereigenschaften so sind, daß die Signale um so stärker gedämpft werden, je weiter die Trägerfrequenz des Signals von der Mittelfrequenz des Filters entfernt ist, scheint der Fleck nach außen zu wandern. In Wirklichkeit ist es so, daß mit zunehmender Amplitude die Filterkanäle der Reihe nach zuerst die 'Betriebsamplitude' der Triggerneurone erreichen, und dann, wenn die Amplitude zu laut wird, die Triggerneurone keine Aktivitäten mehr erzeugen.

Welche Rückschlüsse läßt das Verhalten des Modells auf das biologische Vorbild zu, wenn angenommen wird, daß tatsächlich die von Langner vorgeschlagene Karte existiert? Im Gegensatz zu den Ergebnissen der Simulationen wird eine Messung normalerweise mit wenigen, oft nur einer Elektrode durchgeführt. Ein Experimentator, der eine Lautstärkenkarte suchen will, würde beispielsweise folgendermaßen vorgehen: Zuerst würde er ein beliebiges Neuron im IC aussuchen - tendenziell in der Mitte - und dann den passenden Stimulus für dieses Neuron suchen. Basierend auf dem Langnermodell sollte dazu ein amplitudenmodulierter Stimulus verwendet werden. Findet man ein Periodizitäts/Frequenz Paar, bei dem das Neuron regelmäßig feuert, so würde man jetzt den Lautstärkebereich ausmessen, bei dem es weiterhin regelmäßig feuert. Dann fährt man mit der Elektrode etwas zur Seite und wiederholt die Messung. Dabei wird nur die Lautstärke verändert, die Periodizität und die Trägerfrequenz bleiben gleich. Nun werden entlang einer Linie entsprechend viele Messungen durchgeführt und jeweils der passende Amlitudenbereich protokolliert. Am Ende des Versuches würde der Experimentator zu recht denken, daß er eine Art Lautstärkenkartierung gefunden hat, weil der Lautstärkenbereich, in dem das Neuron gut antwortet, mit zunehmender Entfernung vom ersten Neuron größer wird. Je nachdem, ob zufällig die Neuronen entlang der Periodizitätsachse oder entlang der Trägerfrequenzachse untersucht wurden, würde diese Lautstärkenkarte einen unterschiedlich starken Gradienten haben. Wenn nur wenige Versuche durchgeführt werden würden, könnte man zu dem Ergebnis kommen, daß eine zirkuläre Lautstärkenkarte vorhanden ist, bei der die kleinsten Lautstärken, bzw. die empfindlichsten Neuronen, in der Mitte liegen und die unempfindlicheren Neuronen außen.

Auch wenn diese Ergebnisse attraktiv aussehen, so wären die daraus gezogenen Schlußfolgerungen falsch - weiterhin unter der Annahme, daß die Karte des Langnermodells im IC abgebildet ist. Der Grund ist folgender: Die gefundene Karte hängt vom Stimulus ab! Der Punkt der geringsten Lautstärke, also wo die Neuronen am empfindlichsten reagieren, muß immer der gleiche sein, damit man eine sinnvolle Kartierung hat. Dies ist aber nicht der Fall: Wird die Trägerfrequenz verändert, so verschiebt sich mit ihr auch die Position der empfindlichen Neuronen, da diese eigentlich überhaupt nicht weniger oder mehr empfindlich sind, sondern nur durch die jeweiligen

Filterkanäle entsprechende Signale mit kleinerer oder größerer Amplitude bekommen. Verändert man die Trägerfrequenz, so wird ein anderer Kanal der Filterkanal mit der am besten passendsten Mittelfrequenz, und das Zentrum der Karte verschiebt sich. Dies merkt der Experimentator allerdings nicht, wenn nicht genug Versuche mit unterschiedlichen Stimuli durchgeführt werden.

Zusammengefaßt läßt sich sagen, daß aus dem Modell zwar auf den ersten Blick eine Art Lautstärkenkartierung ableitbar ist, diese These aber nicht haltbar ist, wenn sie genauer untersucht wird. Bei Untersuchungen am realen IC könnte ein Experimentator durchaus fehlgeleitet werden und eine solche Karte in die Ergebnisse der Messungen hineininterpretieren.

Latenzkartierung

Im Gegensatz zur Lautstärkenkartierung, die nur bei oberflächlicher Betrachtung aus dem Langnermodell ableitbar ist, ergibt sich die Latenzkartierung tatsächlich aus dem Langnermodell. In Abb. 3.46 auf Seite 59 ist dargestellt, wie die Latenz der Neuronen mit zunehmender Periodizität geringer wird. Dazu wird die Zeit gemessen, die zwischen dem Onset des Stimulus und der Onsetantwort der Neuronen vergeht. Diese Messung kann in dem Modell auf einfache Weise durchgeführt werden. Man sieht, daß die Latenzen ungefähr der Periodizität in ms zuzüglich einer additiven Konstante entsprechen. Ein Schaltkreis, der auf 100 Hz Periodizität antwortet, tut dies mit ca. 10 ms+2 ms=12 ms (1/(100 Hz)+Konstante) Verzögerung nach dem Onset des Stimulus. Die Konstante bleibt dabei etwa gleich groß. Diese Zeiten entstehen aus folgenden einfachen Gründen: Um eine Periodizitätsanalyse durchzuführen, muß prinzipiell eine Periode abgewartet werden - vorher kann nicht bestimmt werden, welche Periodizität anliegt. Sobald eine komplette Periode vorbei ist, kann ein Schaltkreis diese als 'vorhanden' melden. Vorher geht es prinzipiell nicht. Diese Zeit wird allerdings mit zunehmender Periodizität kleiner, da eine Periode kürzer wird, wenn die Frequenz zunimmt. Daher ist es völlig klar, daß die Latenz des Langnermodells mit der Periodizität im direkten Zusammenhang steht.

Die Konstante von ca. 2 ms entspricht der Durchlaufzeit eines Schaltkreises von Triggerimpuls erzeugen, Spike im Oszillatorschaltkreis bearbeiten und anschließend an das Koinzidenzneuron weitergeben. Diese ist unabhängig von der zu detektierenden Periodizität.

Ein einfaches Indiz für das Vorhandensein eines biologischen Periodizitätsanalysators, dessen Ergebnisse im IC ausgegeben werden, ist also die Latenz. Existiert keine Abhängigkeit zwischen Latenz und Periodizität, so wird vermutlich kein Periodizitätsanalysator verwendet. Die neun vorhandenen Neuronen konnten auf einfache Art und Weise in dieser Hinsicht ausgewertet werden, die Ergebnisse sind in Abb. 3.39 auf Seite 54 dargestellt. Im Plot rechts unten sind die jeweiligen Latenz/Periodizitätspaare als Punkte eingetragen. Die zusätzlich eingezeichnete Gerade stellt die lineare Annäherung der Punkte dar, die typischerweise durchgeführt wird. Man erkennt, daß die Latenz systematisch von der Periodizität abhängt. Allein mit den vorhandenen Punkten könnte man allerdings nicht sagen, ob ein linearer Zusammenhang oder ein anderer besteht. Greift man auf die Ergebnisse des Langnermodells zurück, so müßte man eine Hyperbel einzeichnen oder besser die Periodizität in ms ausdrücken, und dann eine lineare Approximation verwenden. Das Entscheidende ist allerdings die Tatsache, daß die Latenzen sich in der Größenordnung bewegen, die den theoretischen Überlegungen und den Ergebnissen der Simulationen entsprechen. Kleinere Abweichungen sind durch die ungenaue Messung der Periodizität auf die das Neuronen am besten antwortet, zu begründen. Wenn diese Abhängigkeit durch eine größere Anzahl von Messungen bestätigt werden könnte, so wäre die Messung der Periodizitätskartierung einfach möglich, indem man immer die Latenz mißt, und nicht für jedes Neuron verschiedene Periodizitäten hochauflösend ausprobieren muß.

Zusammengefaßt kann gesagt werden, daß die Kartierung der Latenz in der IC Karte an sich uninteressant ist, weil sie auf trivialen Grundprinzipien beruht. Im Zusammenhang mit biologischen Messungen ist sie allerdings äußerst interessant, da anhand dieser Überlegungen zum einen die Existenz der Karte einfach überprüft werden kann und zum anderen ein schnelles Verfahren zur Bestimmung des Periodizitätstunings eines Neurons im IC bereitgestellt wird.

Kartierung des Periodizitätstuning

Die Ergebnisse der Untersuchungen zur Kartierung des Periodizitätstuning sind in Abb. 3.47 auf Seite 59 dargestellt. Es ergibt sich ein Antwortverhalten, das entlang der Periodizitätsachse der IC-Karte ein immer breiteres Antwortmuster zeigt. Dieser Parameter ist wie die Latenz tatsächlich systematisch auf der Karte vorhanden und nicht wie die Lautstärke nur eine Fehlinterpretation von Messungen. Diese Kartierung resultiert aus der konstanten Breite der Oszillatorantworten (siehe oben). Würde die Anzahl Spikes, die der Oszillator erzeugt, entsprechend angepaßt, so könnte die Breite der Antwortkurven nahezu beliebig eingestellt werden.

Auch wenn diese Kartierung direkt aus einfachen Mechanismen ableitbar ist, können aus ihr interessante Fragen resultieren: Sind unterschiedlich breite Antwortkurven bei verschiedenen Periodizitäten anzustreben? Wie verhalten sich reale Neuronen? Welche Schlußfolgerungen lassen sich aus dem Vergleich von realen Neuronen und Modellneuronen ziehen? Diese Fragen werden im folgenden kurz diskutiert.

Dadurch, daß mit Hilfe des Oszillators die Breite der Neuronantworten relativ frei eingestellt werden kann, ergibt sich die Frage, welche Einstellung überhaupt anzustreben ist. Dies hängt mit der Anwendung zusammen: wenn bestimmte Bereiche des Periodizitätsspektrums von besonderem Interesse sind, so kann man für diese mehr Neuronen mit schärferem Tuning verwenden als für die nicht ganz so wichtigen Bereiche. Diese könnten mit weniger Neuronen und breiterer Neuronantwort abgedeckt werden. Für solche Zwecke wäre eine entsprechende Einstellung des Oszilllators notwendig.

Um festzustellen, wie das Tuning bei Säugetieren eingestellt ist, müßten entsprechende Messungen durchgeführt werden. Die meisten Messungen werden mit reinen Sinussignalen durchgeführt und sind daher nicht geeignet, in Bezug auf die Periodizität analysiert zu werden. Dementsprechend sind die Tuningkurven, die in der Biologie sehr häufig gemessen werden, nicht direkt vergleichbar, weil für diese die Amplitude und die Frequenz des Sinus modifiziert werden. Ein Experimentator, der diese neue Fragestellung untersuchen wollte, müßte also im IC ein Neuron suchen, dann die zugehörige Trägerfrequenz und Periodizität feststellen und dann die Periodizität variieren. Wenn genug Neuronen gemessen wurden, könnte man versuchen herauszufinden, ob sich die Antwortbreite systematisch verändert. Schaut man sich mit dieser Fragestellung die auf Seite 53 abgebildeten Periodizitätstuningkurven der biologischen Neuronen an, so erkennt man allerdings, daß diese Frage nicht eindeutig mit wenigen Neuronen beantwortbar ist. Die Kurven sind so unterschiedlich, daß nicht direkt und einfach die Breite so gemessen werden kann, daß die Daten vergleichbar sind. Tendenziell würde man allerdings nicht den Eindruck bekommen, daß die Kurven mit zunehmender Periodizität breiter werden.

Interessant wäre, wenn man bei Chopperneuronen aus dem Nucleus Cochlearis die Dauer der Antwort messen würde, und diese mit der Periodizität des zugehörigen Koinzidenzneurons im IC vergleichen würde. Aus diesem Vergleich könnte man am besten Rückschlüsse auf den realen Aufbau und die Tuningeinstellungen machen. Allerdings ist es nicht einfach, zu einem Neuron im Nucleus Cochlearis das zugehörige Neuron im Colliculus Inferior zu finden, selbst wenn die Verbindungen entsprechend des Langnermodells existieren. Ein schöner Hinweis auf die Korrektheit des Langnermodells wäre, wenn man tatsächlich einen Zusammenhang zwischen der Dauer der Oszillatorantwort und der Breite der Antwortkurven des entsprechenden Koinzidenzneurons finden könnte.

4.1.6 Abschätzung der Anzahl der notwendigen Neuronen

Eine interessante Frage ist, welche der Neuronen tatsächlich gebraucht werden und welche unnötig oft in dem Modell implementiert sind. Für einen einzelnen Langnerschaltkreis ist die Antwort relativ einfach: Wenn man auf das Zeittuning verzichten will, daß mit Hilfe des Oszillators eingestellt werden kann, so könnte man diesen einsparen. Will man dagegen die volle Funktionalität eines einzelnen Langnerschaltkreises, so wird jedes Neuron benötigt. Betrachtet man aber die große Menge von Langnerschaltkreisen, die für die Berechnung einer kompletten IC-Karte notwendig ist, so ergibt sich ein anderes Bild. Es ist zunächst nur klar, daß für jeden Punkt der Karte ein

Koinzidenzneuron benötigt wird. Die Notwendigkeit, für jedes Koinzidenzneuron einen kompletten Schaltkreis zu replizieren, ist dagegen nicht gesichert. Tatsächlich ist es so, daß einige der Neuronen nicht so häufig benötigt werden: Jedes Koinzidenzneuron bekommt Spikes vom Integrator- und Oszillatorteil des Schaltkreises. Alle Koinzidenzneuronen einer Trägerfrequenz bekommen dabei im Prinzip die gleichen Spikes von den Oszillatoren, da diese durch Triggerneuronen aktiviert werden, die alle mit demselben Filterkanal bzw. Hörnerv verbunden sind. Im Prinzip könnten anstatt beispielsweise der hier simulierten 50*50 Triggerneuronen genausogut nur 50 Triggerneuronen verwendet werden, die Spikes werden dann auf alle Schaltkreise mit der gleichen Trägerfrequenz verteilt. Entsprechend können die Oszillatorneurone eingespart werden, die die gleiche Anzahl Spikes pro Triggerspike erzeugen. Selbst wenn der Oszillator für das spezielle Zeittuning verwendet wird, könnten die Anzahl der Oszillatoren beispielsweise von 50*50 Oszillatoren auf 50*5 Oszillatoren reduziert werden. Zu jedem Frequenzkanal würden 5 Oszillatoren gehören, die unterschiedlich viele Spikes pro Triggerspike erzeugen. Falls dieses spezielle Zeittuning nicht verwendet wird, reicht ein Oszillator pro Trägerfrequenz aus. In diesem Falle werden nur 50 Triggerneuronen und 50(*3) Oszillatoren benötigt, anstatt 50*50+50*50*3=10000 Neuronen. Man könnte also 98% dieser Neuronen einsparen.

Für den Integratorteil sieht das dagegen anders aus. Ein Koinzidenzneuron, welches sowohl eine Aussage über die Periodizität als auch über die Frequenz machen soll, benötigt immer einen eigenen Integratorteil. Würden mehrere Frequenzbänder mit nur einem Integrator verbunden werden, so würde das Koinzidenzneuron nicht mehr die einzelnen Frequenzbänder repräsentieren, sondern den ganzen Bereich der Frequenzbänder, die mit dem Integrator verbunden sind. Andererseits kann ein einzelner Integrator nicht unterschiedliche Periodizitäten detektieren. Es können also keine Neuronen des Integratorteils eingespart werden, ohne die Auflösung der Karte zu reduzieren. Wenn diese allerdings reduziert werden soll, so können auch gleich die Koinzidenzneuronen eingespart werden und somit die Anzahl der Schaltkreise reduziert werden.

Der unterschiedliche Aufbau des Integratorteils der beiden Langnermodelle ergibt unterschiedliche Einsparungsmöglichkeiten. Während beim ursprünglichen Langnermodell der Integrator durch die Flip-/Flopneuronen synchronisiert wird, wird für das variierte Modell das IntegStopneuron verwendet. Für jeden Langnerschaltkreis muß in der ursprünglichen Version ein Flip/Flop Neuronpaar verwendet werden, da die unterschiedlichen Integrationszeiten der Integratorneuronen Unterschiede in dem Antwortverhalten der Flip/Flop Neuronen erfordern. Für ein Integratorneuron, der eine kurze Verzögerung erzeugt, wird das Flopneuron wesentlich früher aktiv (und hemmt den Integrator), als für ein Integratorneuron, welches eine lange Verzögerung erzeugt. Daher wird für jede Verzögerung - und somit für jedes Integratorneuron - ein Flopneuron benötigt. Andererseits hemmt das Flopneuron das Flipneuron, und damit ein Flipneuron nicht zu einem falschen Moment gehemmt wird, wird zu jedem Flopneuron ein eigenes Flipneuron benötigt. In dem ursprünglichen Langnermodell muß also der Integratorteil für jede gewünschte Periodizität komplett repliziert werden.

In dem variierten Langnermodell kann dagegen das IntegStopneuron für alle Integratoren einer Trägerfrequenz verwendet werden. Die Funktion des IntegStopneurons ist, den Integrator zurückzusetzen. Da für jede Trägerfrequenz immer entweder alle oder keine Integratorneuronen zurückgesetzt werden, kann dies auch von einem einzelnen IntegStopneuron pro Trägerfrequenz durchgeführt werden. Entsprechend sinkt die Anzahl der notwendigen Neuronen bei der verwendeten Kartengröe für den Integratorteil von 50*50*3=7500 (je Punkt in der Karte ein IntegStop Neuron, ein Integratorneuron und ein Inhibneuron) auf 50*50*2+50=5050 Neuronen.

4.1.7 Vergleich der theoretisch notwendigen Neuronen mit den vorhandenen Neuronen beim Menschen

Entsprechend der obigen Überlegungen sind für eine $n * m$ (n=Anzahl Periodzitäten, m=Anzahl Trägerfrequenzen) nicht $8 * n * m$ Neuronen notwendig, sondern es reichen bereits für das variierte Modell $5 * m + 3 * n * m$ Neuronen, oder $4 * m + 4 * n * m$ Neuronen für das ursprüngliche Modell. Eine Vorraussetzung für eine Abschätzung, wieviele Neuronen der Mensch für das Langnermodell benötigt, ist, die Größe der zu berechnenden Karten festzulegen. Untersuchungen aus

der Psychophysik haben gezeigt, daß der Mensch in der Lage ist, bei Frequenzen unterhalb 500 Hz Unterschiede von 3 Hz wahrzunehmen. Ab 1000 Hz ergibt sich etwa ein konstanter prozentualer Anteil von 0.6% der Frequenz [49]. Teilt man den vom Menschen hörbaren Bereich von 20 Hz bis 16000 Hz in einen Periodizitätsbereich von 20 Hz bis 500 Hz und einem Trägerfrequenzbereich von 500 Hz bis 16000 Hz ein, so würde dies 160 Periodizitäten ((500 Hz - 20 Hz)/3 Hz) und 580 Trägerfrequenzen ($\frac{\log(16000/500)}{\log(1.006)}$) ergeben. Wenn der Langnerschaltkreis entsprechend oft repliziert werden würde, wären 580*160*8=742 400 Neuronen notwendig. Durch die Einsparungen reduziert sich diese Zahl beim ursprünglichen Langnermodell auf 4*580+4*580*160=373 520, also ca. auf die Hälfe. Das variierte Langnermodell benötigt noch ein paar Neuronen weniger, nämlich 5*580+3*580*160=281 300 Neuronen. Im Vergleich zu den 100 000 000 000 Neuronen, die ein menschliches Gehirn ungefähr beinhaltet, sind diese Einsparungen nicht übermäßig bedeutend. Dennoch ist anzunehmen, daß im Verlaufe der Evolution sich Mechanismen durchsetzen, die möglichst wenig Resourcen verbrauchen.

Für das variierte Langnermodell kann auch die Anzahl der notwendigen Neuronen mit den in den Hörzentren des Menschen vorhanden Neuronen verglichen werden. Die Zahlen gelten - bis auf den Unterschied des Integratorteils - entsprechend für das ursprüngliche Modell. Welchen biologischen Neuronen die zusätzlichen Neuronen für den Flip/Flop Schaltkreis entsprechen ist dagegen unklar. Die Neuronen des Modells teilen sich auf in die Neuronen, die im Colliculus Inferior sitzen (Koinzidenzneuron und Inhibneuron) und den restlichen Neuronen, die im Nucleus Cochlearis sind (Triggerneuron, Oszillator, IntegStop und Integratorneuron). Exakte Daten über die Anzahl und Verschaltung der Neuronen sind nicht bekannt. Aber die ungefähren Größenordnungen können der Literatur entnommen werden. Das Volumen des Nucleus Cochlearis wird beim Menschen auf insgesamt ca. 36 mm^3 geschätzt [15], das des Colliculus Inferior auf 38 mm^3. Die Anzahl der Neuronen konnte nicht direkt in der Literatur gefunden werden, aber die Dichte der Neuronen beim Rhesusaffen liegt im Nucleus Cochlearis bei etwa 15 500 Neuronen/mm^3 und im Colliculus Inferior bei 42 000 Neuronen/mm^3 [3]. Unter der Annahme, daß die Dichte der Neuronen in den Kernen beim Rhesusaffen und beim Menschen ungefähr gleichgroß ist, können aus den entsprechenden Volumenangaben beim Menschen die Neuronanzahl bestimmt werden. Daraus ergibt sich jeweils eine Gesamtzahl der Neuronen beim Mensch von 558 000 Neuronen im Nucleus Cochlearis und 1 596 000 Neuronen im Colliculus Inferior. Insgesamt würde über 2 Millionen Neuronen zur Verfügung stehen, im Vergleich zu den knapp 300 000, die für das variierte Langnermodell notwendig sind. Allerdings ist nicht davon auszugehen, daß alle Neuronen der Kerne für das Langnermodell verwendet werden.

Betrachtet man aber die Kerne genauer, so stellt man fest, daß die meisten Neuronen des Nucleus Cochlearis im anteroventralen Bereich (ca. 14 mm^3) des Nucleus Cochlearis liegen, die vermutlich nicht für Periodizitätsdetektion, sondern für Schalllokalisation zuständig sind. Die Neuronen des Langnermodells passen besser zu Neuronen aus dem dorsalen Nucleus Cochlearis (DCN) und dem posteroventralen Bereich des Nucleus Cochlearis (PVCN). Das Triggerneuron des variierten Langnermodells könnte den Oktopus Neuronen, die im posteroventralen Teil des Nucleus Cochlearis liegen (ca. 1 mm^3, also a. 15 500 Neuronen), entsprechen. Der Oszillatorschaltkreis entspricht Chopperneuronen, die im DCN liegen. Dieser ist in drei Bereiche unterteilt, die unterschiedliche Neurontypen beinhalten. Das kleinste Volumen verbrauchen die Pyramidenzellen, die ein Volumen von 1.5 mm^3 (entspricht etwa 23 250 Neuronen) füllen. Das Antwortverhalten dieser Zellen ist unterschiedlich, sowohl Buildup-Antwortmuster (könnte dem dem Integratorneuron entsprechen) als auch Chopper-Antwortmuster (entsprechen dem Oszillatorschaltkreis) wurden gefunden. Zusätzlich zeigen sogennante 'Giant Cells', die in der Polymorphic-Schicht mit 5 mm^3 (77500 Neuronen) das größte Volumen im DCN verbrauchen, ein Buildup-Antwortmuster, welches dem Integrator Neuron entspricht. Der letzte Neurontyp, der für das Langnermodell noch benötigt wird, ist der des IntegStopneuron. Die 'Tuberculoventral' Neuronen, welche auch in der Polymorphic-Schicht liegen, könnten diesen Neuronen entsprechen, sie zeigen auch ein Onset-Antwortmuster.

Abgesehen davon, daß Neurontypen existieren, die ein dem Langnermodell entsprechendes Antwortmuster zeigen, müssen sowohl die Anzahl der Neurone ungefähr der theoretischen Anzahl entsprechen, als auch die Verbindungstypen (hemmend/erregend) passen. Von den Oktopuszel-

len, die dem Triggerneuron entsprechen sollen, existieren ca. 15 500, das Modell benötigt dagegen nur 580 Triggerneuronen. Diese Neuronen projezieren exiatatorisch in verschiedene Kerne, allerdings sind diese im Gegensatz zu dem Langnermodell nicht innerhalb des Nucleus Cochlearis, sondern bekannte Ziele sind der Colliculus Inferior und andere Hörkerne. Für die anderen Neurontypen passen sowohl die Verbindungsstruktur als auch die Anzahl der Neuronen sehr gut zu dem Langnermodell: Von den Chopperneuronen im DCN, die den Oszillatorneuronen entsprechen sollen, werden dem Modell nach 1740 Neuronen benötigt, dem Volumen nach stehenüber 20 000 Neuronen diesen Types zur Verfügung. Die exitatorischen Projektionen dieser Neuronen gehen zum Colliculus Inferior und dem lateralen Lemniscus, dem Modell nach werden Verbindungen zu den Koinzidenzneuronen benötigt, die passenderweise im Colliculus Inferior liegen sollen. Die IntegStop- und Integratorneuronen können mit den 77 500 Neuronen der Polymorphic-Schicht gleichgesetzt werden. Die Neurontypen, die ein Antwortverhalten entsprechend dem IntegStopneuron zeigen, projezieren hemmend auf den anderen Neurontyp, der ein Buildup-Antwortverhalten zeigt, und dem Integratorneuron entsprechen soll. Diese wiederum projezieren exitatorisch in den Colliculus Inferior, genau wie das variierte Langnermodell es erfordert. Die Neuronanzahl von 77 500 ist etwas geringer als die vom Modell geforderten 160*580+580=93 380 (Integratorneuronen+IntegStopneuronen). Allerdings sind die Zahlen nur Schätzungen, so daß kleine Differenzen keine Bedeutung haben.

Zusammenfassend kann gesagt werden, daß im Nucleus Cochlearis alle notwendigen Neurontypen in ausreichender Anzahl vorhanden sind, wobei bis auf die Projektionen des Triggerneurons auch die Art und das Ziel der neuronalen Projektionen mit dem variierten Langnermodell übereinstimmen. Von den geschätzten 558 000 Neuronen werden ca. 100 000 Neuronen durch das Langnermodell verbraucht, wobei diese Neuronen hauptsächlich im DCN sitzen. Die restlichen Neuronen können für andere Aufgaben verwendet werden, wie zum Beispiel die Schalllokalisation oder die Extraktion weitere Merkmale des Stimulus.

Die letzte Station im Langnermodell, das Koinzidenzneuron, wird im Colliculus Inferior vermutet. In diesem wurden tatsächlich Neuronen mit einem entsprechenden Antwortverhalten gefunden. Dieser Kern besteht aus mehreren Unterschiedlichen Bereichen, wobei die gesuchten Neuronen sich im zentralen Bereich des Colliculus Inferior befinden. Für die gewünschte Karte werden 580*160=92 800 Koinzidenzneuronen und nochmals soviele inhibitorische Interneurone benötigt. Selbst wenn beide Neurontypen im Colliculus Inferior sitzen wären damit erst ca. 12 % der Neuronen verbraucht. Falls die Interneuronen dagegen den Neuronen im lateralen Lemniscus entsprechen, die inhibitorisch in den Colliculus Inferior projezieren, so wären in diesem nur knapp 6 % der Neurone verbraucht. In beiden Fällen stehen noch sehr viele Neuronen für andere Zwecke zur Verfügung.

Der Vergleich mit den biologischen Strukturen zeigt, daß das variierte Langnermodell sehr gut zu den vorhanden Neuronen paßt, und auch die Verbindungsstruktur und die Mindestanzahl der benötigen Neuronen dem Modell entsprechen. Das wesentlich mehr Neuronen in dem Kern vorhanden sind, als für das Langnermodell notwendig sind kann mit mehreren Argumenten begründet werden:

- Es ist klar, daß zum einem die Anzahl der vorhandenen Neuronen eine relativ grobe Schätzung darstellt. Zum anderen ist die Größe der biologischen Karte auch unklar, und nur ungefähr aus den Ergebnissen von psychophysikalischen Experimenten abgeleitet. Dies bedeutet, daß alle Zahlen nur in Bezug auf die Größenordnung stimmen können, und die Tatsache, daß die Anzahl der biologischen Integratorneurone etwas geringer ist, als die von dem Modell geforderten Integratorneuronen ist daher von geringer Bedeutung.

- Das Hörsystem führt neben der Schallklassifikation mehrere weitere Aufgaben durch, von denen anzunehmen ist, daß sie teilweise auch schon im Nucleus Cochlearis wahrgenommen werden. Eine dieser Aufgaben ist die Schalllokalisation, welche bereits auf dieser Ebene viele neuronale Resourcen verbraucht. Andere Aufgaben sind aber beispielsweise die schnelle Reaktion auf gefährliche Geräusche, wie zum Beispiel das Zusammenzucken von Menschen bei überraschenden oder lauten Geräuschen. Für solche Reflexe ist weder der genaue Ort noch

der genaue Typ des Stimulus von Bedeutung, sondern einfach nur das schnelle Bereitstellen von Resourcen um der Gefahr begegnen zu können.

- Außerdem ist es klar, daß die reine Periodizitäts/Frequenzkarte nicht alle Wahrnehmungen des Menschen erklären kann. Zusätzlich zu dieser Kartierung müssen weitere Analysen stattfinden, die in der Lage sind, die Wahrnehmung eines reinen Sinustons zu erklären. Das Langnermodell zeigt bei einem solchen Stimulus bestenfalls eine Onset Antwort, oftmals aber überhaupt keine Antwort. Es ist anzunehmen, daß neben der Karte auch andere Merkmale aus dem Schallsignal extrahiert werden.

- Ein offenes Problem des Langnermodells ist auch die Lautstärke des Stimulus. Hier könnten weitere Neuronen im Nucleus Cochlearis notwendig sein, um das Langnermodell lautstärkeunabhängiger zu machen. Allerdings betrifft dies nur das Triggerneuron, wenn dieses lautstärkeunabhängig funktionieren würde, so würde der komplette Rest des variierten Langnermodells nicht verändert werden müssen. Wenn zum Beispiel anstatt ein Triggerneuron pro Periodizitäts/Frequenzpaar zehn Triggerneuronen verwendet werden, die alle auf einen unterschiedlichen Lautstärkebereich antworten, so wäre dieses erweiterte Langnermodell deutlich unabhängiger von der Lautstärke und würde nur anstatt der 580 Triggerneuronen 5800 Triggerneuronen benötigen, was bei einer Gesamtzahl von ca. 100 000 Neuronen nicht wesentlich mehr Neuronen darstellen würde.

4.1.8 Laterale Rückkopplungen innerhalb der Koinzidenzschicht

Häufig werden in neuronalen Netzen, die als Karten organisiert sind, laterale Verbindungen eingesetzt, die die Antwortmuster verschärfen können. Dazu wird die nahe Umgebung von aktiven Neuronen exitatorisch beinflußt und die weiter umliegenden Gebiete inhibitorisch. Dadurch werden eher Neuronen in der nahen Umgebung feuern, und nicht Neuronen, die weiter entfernt sind. In implementierten Versionen des Langnermodells besteht zwar die Möglichkeit, lateralen Feedback auf der Ebene der Koinzidenzschicht einzusetzen. Dagegen sprechen allerdings einige Punkte:

Wenn man laterale Verbindungen verwenden würde, so sollte man diese so gestalten, daß sie die Punkte mit gleicher Periodizität exitatorisch beeinflussen und Punkte mit gleicher Trägerfrequenz inhibitorisch. Dadurch könnte die Chance steigen, daß Koinzidenzneuronen, die der Periodizität des Sprechers entsprechen, aktiv werden. Eine typischerweise verwendete radialsymetrische Verbindungsstruktur wäre dagegen eher störend.

Prinzipiell ist das Problem aber ein anderes: die Neuronen der Koinzdenzschicht feuern schon ohne laterale Rückkopplungen genau dann, wenn es erwünscht ist. Entweder ist im entsprechenden Filterkanal die entsprechende Periodizität mit geeigneter Amplitude enthalten oder nicht. Im einen Fall feuert das Koinzidenzneuron bereits ohne Rückkopplungen, im anderen Fall feuert es sowieso nicht.

Die Idee, daß man mit lateralen Rückkopplungen die Koinzidenzneuronen zum Feuern bringt, bei denen die Amplitude des Filterkanals etwas zu klein ist, sieht auf den ersten Blick durchaus attraktiv aus. Allerdings erlaubt der Aufbau des Langnerschaltkreises nicht, daß dieses Verhalten mit Rückkopplungen innerhalb der Koinzidenzschicht erreicht wird. Diese ist bereits losgelöst von der Amplitude des Filterkanals, daher bringt eine kleine Erhöhung des Neuronpotentiales keine Vorteile. Nur das gleichzeitige Eintreffen der Oszillator- und Integratorspikes kann dieses Neuron zum Feuern bringen. Ist dagegen die Rückkopplung so stark, daß die Koinzidenzneuronen ohne weitere Spikes des Oszillators und Integrators feuern, so würde jedesmal eine ganze Periodizitätsspalte feuern, sobald ein Neuron dieser Spalte feuert, da die Rückkopplungen dann iterativ alle anderen Neuronen aktivieren würden. Dadurch würden alle Informationen verlorengehen, die in den unterschiedlichen Aktivierungen der Trägerfrequenzkanäle stecken - insbesondere auch die Informationen, welche Formantfrequenzen im Stimulus enthalten sind. Die einzige Möglichkeit, genau die Koinzidenzneuronen zum feuern zu bringen, bei denen die Amplitude des zugehörigen Filterkanals zu klein ist, wäre die folgende: Die lateralen Rückkopplungen der Koinzidenzschicht wirken nicht innerhalb der Koinzidenzschicht, sondern auf die Triggerneuronen. Bei diesen würde eine geeignete Erhöhung des Neuronpotentials die Folge haben, daß die Triggerspikes erzeugt

werden, obwohl die Amplitude des Stimulus etwas zu klein ist. Allerdings ergeben sich dadurch bereits Probleme mit den Triggerneuronen, die noch Spikes erzeugen, aber bereits am oberen Ende des Bereichs sind, in dem sie funktionieren. Durch die exitatorische Rückkopplung würden diese aufhören zu feuern. Außerdem könnte diese Verschaltung nur dann durchgeführt werden, wenn tatsächlich jedes Koinzidenzneuron ein eigenes Triggerneuron hat. Die Neuroneinsparungen, die oben vorgeschlagen wurden, wären dann nicht mehr möglich.

Zusammengefaßt kann man sagen, daß die Neuronen der Koinzidenzschicht ihre Aufgabe - nämlich zu feuern, wenn die entsprechende Periodizität im entsprechenden Filterkanal ausreichend vorhanden ist - gut erfüllen, und laterale Rückkopplungen eher Probleme mit sich bringen. Die Vervollständigung von Mustern muß sowieso nicht in der Koinzidenzschicht stattfinden, dazu sind die kortikalen Areale angeschlossen. Im Idealfall sind die Aufgaben wie folgt auf die kortikalen Areale verteilt: In der ersten Assoziativspeicherschicht werden die Muster vervollständigt, und die zweite Schicht sucht sich eines der aktiven Muster aus der ersten Schicht aus. Somit werden die Muster aus dem IC im Kortex komplettiert, und die lateralen Rückkopplungen innerhalb des ICs sind für Mustervervollständigung nicht notwendig - abgesehen davon, daß diese wie oben beschrieben auch Probleme mit sich bringen können.

4.2 Einbeziehung kortikaler Areale

Mit Hilfe von Assoziativespeichern werden die kortikalen Areale modelliert. Die verwendeten Assoziativspeicher C und D sind im Prinzip sehr ähnlich aufgebaut, repräsentieren allerdings unterschiedliche Bereiche im Gehirn. Während C den ganzen primären auditorischen Kortex repräsentiert, stellt D nur einen kleinen Ausschnitt aus einem höheren Areal dar. Daher kann in D fast jedes Neuron mit jedem anderen Neuron verbunden werden, während in C die Verbindungen nicht soweit reichen. Die Funktion von C ist dementsprechend eher die Vervollständigung von Mustern, während D aus diesen Mustern im Normalfall nur eines auswählt, auf das sich das System dann konzentriert. Die unterschiedliche Interpretation der Areale betrifft auch die Musterrepräsentation, in C sind die Muster ähnlich wie in K tonotopisch/periodotopisch, während in D reine Zufallsmuster verwendet werden. Im Prinzip ist die Verwendung von Zufallsmustern aus informationstheoretischer Sicht günstiger. Da aber aus der Biologie bekannt ist, daß der primäre auditorische Kortex tonotop/periodotop aufgebaut ist - wenn auch noch unklar ist, wie die Achsen genau angeordnet sind - wurde für C eine ähnliche Kartierung verwendet.

Im folgenden werden die einzelnen Ergebnisse, die mit Hilfe der Assoziativspeicher erzielt wurden, diskutiert.

4.2.1 Klassifikationsexperimente mit Assoziativspeicher C

Mit Hilfe des Assoziativspeichers C wurde der Einfluß der Parameter zur Einstellung der Autoassoziation und Inhibition getestet. Wie der Vergleich der Abb. 3.50 mit Abb. 3.51 ab Seite 62 zeigt, verschärfen sich die Antwortmuster, wenn die Inhibition verstärkt wird. Wenn keinerlei Inhibition vorhanden wäre, so würden alle Muster aktiv werden. Wenn zuviel Inhibition eingestellt wäre, so würden selbst die richtigen Muster kaum mehr aktiviert werden. Daher wird die Inhibition so eingestellt, daß möglichst viele Neuronen von den korrekten Mustern aktiv werden, und die anderen Muster möglichst schnell deaktiviert werden, wenn der Stimulus sich ändert. In Abb. 3.50 sieht man, daß dieser Wechsel durch die Muster nicht sehr schnell repräsentiert wird, während in Abb. 3.51 der Wechsel sehr schnell stattfindet. Anhand dieses Graphen wäre es einfacher, den Moment zu finden, in dem sich der Stimulus verändert hat.

Generell kann gesagt werden, daß die gespeicherten Stimuli jeweils sehr gut vom angelegten Stimulus aktiviert werden. Sporadisch werden auch falsche Muster aktiv, aber diese werden von den korrekten Mustern nach kurzer Zeit (50 ms) deaktiviert. Der Assoziativspeicher kann also die gespeicherten Muster sicher dem jeweiligen Stimulus zuordnen.

4.2.2 Nutzen des kortikalen Assoziativspeichers D

Der Vergleich der Aktivitätsverteilung der Assemblies (Abb. 3.54 auf Seite 64) zeigt, daß mit Rückkopplungen die Anzahl von Neuronen pro Auslesevorgang wesentlich höher ist. Meistens erreichten sie 50 Neuronen, also die komplette Assembly. Dieser Effekt ist auch klar verständlich, wenn man berücksichtigt, daß die lateralen Verbindungen innerhalb von C kurz sind, und daher ein Neuron von oben ein Neuron von unten nicht aktivieren kann. In dem gewählten Beispiel wurden die Muster künstlich um einen zusätzlichen Fleck erweitert, der außerhalb der Reichweite der restlichen Neuronen des Musters in C liegt. Nur durch die Rückkopplungen aus D kann das ganze Muster aktiviert werden: Da sich D von C an zwei wesentlichen Stellen unterscheidet, können dort die Muster komplett aktiviert werden, auch wenn von C nur unvollständige Muster angeboten werden. Zum einem ist die Verbindungsreichweite wesentlich höher, zum anderen sind Zufallsmuster gespeichert. Dieser letzte Punkt sorgt dafür, daß die Entfernung aller Neuronen, die in D zu einem Muster gehören, gering ist, insbesondere deutlich kleiner als die Verbindungsreichweite. Daher aktiviert sich die ganze Assembly in D und kann so die fehlenden Teile der Assembly in C durch Rückprojektionen aktivieren. Dies geschieht nicht sofort, sondern mit der Verzögerung, die die Spikes in beide Richtung benötigen. Daher sieht man bei kurzem Zeitfenster zwei Aktivitätsmaxima pro Auslesevorgang in C. Wird das Zeitfenster entsprechend angepaßt, so erreicht die Anzahl aktiver Neuronen 50. Die Änderung des Zeitfenster ändert aber nichts an den Antwortmustern, sondern nur an der Auswertung. Das neuronale Netzwerk hat genau die gleichen Aktivitäten wie vorher.

Ein weiterer Vorteil zeigt sich im Vergleich der Klassifikationsleistungen des Assoziativspeichers, wenn mehrere Vokale gespeichert und klassifiziert werden sollen: In Tab. 3.3 (Seite 66)sind die Ergebnisse von 20 Klassifikationen ohne Rückkopplungen aufgeführt, in Tab. 3.5 (Seite 67) die Ergebnisse mit Rückkopplungen. Während ohne Rückkopplungen in C nur 73% in Bezug auf den Vokaltyp korrekt klassifiziert wurden, sind es mit Rückkopplungen schon 83%. Interessant ist, daß sich auch die Ergebnisse in D verbesserten: von 88% (ohne Rückkopplungen) auf 91% (mit Rückkoplungen). Dies zeigt, daß beide Assoziativspeicher von den Rückkopplungen profitieren.

Betrachtet man die Klassifikationsleistungen in Bezug auf die Periodizität, so können die Assoziativspeicher in 96% der Fälle die Periodizität korrekt erkennen. Das bedeutet, daß die Periodizität besser repräsentiert wird als die Formantfrequenzen. Überspitzt könnte man sagen, daß das Langnermodell sich also eher zum Musikhören eignet als zur Analyse von Sprache. Dies ist aber nicht überraschend, weil das Modell speziell zur Extraktion der Periodizitätsinformationen entwickelt wurde. Die Eigenschaft, daß man die Formantfrequenzen auch extrahieren kann, ist dagegen ein Nebeneffekt, der sich durch die Filterbank ergibt. Die Ergebnisse für die Untersuchung der IC Karte haben auch fast immer die Periodizität gut repräsentiert, dagegen sind die Formantfrequenzen nicht immer an identischer Stelle für gleiche Vokaltypen aufgetaucht.

4.2.3 Klassifikation unbekannter Muster

Um die Aufgabe schwieriger zu gestalten, wurde für weitere Versuche ein Datensatz aus 18 Exemplaren je Vokal verwendet. Von diesen Vokalen wurden je 4 zum Lernen des Assoziativspeichers verwendet, die restlichen 14 zum Klassifizieren. Zusätzlich wurden die 4 gelernten Vokale auch klassifiziert. Die Ergebnisse sind in Tab. 3.7 auf Seite 69 aufgeführt. Man sieht, daß die Klassifikationsleistung mit den unbekannten Vokalen deutlich geringer ist (ca. 60%) als bei den vorherigen 20 Vokalen (ca. 90%). Dagegen sind die gespeicherten Muster zu 97% korrekt klassifiziert worden. Daß dieser Wert höher ist als der vorherige, wo auch 20 Vokale gespeichert und wieder ausgelesenwurden, kann daran liegen, daß die Vokale jetzt auch mit gleicher Periodizität auftreten können, und nicht jeder Vokal alle Periodizitäten belegt. Daher vereinfacht sich die Aufgabe etwas.

Eine Klassifikationsleistung von 60% ist nicht besonders gut. Allerdings sollte man dieses Ergebnis differenziert betrachten: Nehmen wir an, ein Mensch unterhält sich mit einer Person, die er gut kennt. Diese Situation entspricht dem Fall, wo die Vokale bereits im Assoziativspeicher eingespeichert sind, weil der Mensch die Stimme des Gesprächspartners bereits kennt. Er versteht ihn gut, und die Assoziativspeicher würden eine Leistung von fast 100% erreichen. Dagegen entspre-

chen die 60% Klassifikationsleistung eher dem Fall, daß sich zwei Menschen unterhalten, die sich noch nie gesehen haben, und vielleicht sogar unterschiedliche Dialekte sprechen. Beide Menschen müssen sich erst etwas in die Stimme des Gesprächspartners hineinhören, um den anderen gut zu verstehen. Dieser Vorgang dauert einige Zeit, und dann kennt man den Gesprächspartners besser und versteht ihn auch besser. Das bedeutet, daß der Mensch in diesem Fall zunächst auch nicht alles pefekt versteht. Mit Hilfe weitere kognitiver Fähigkeiten (Analyse der Grammatik, Thematische Zusammenhänge, ...) und anderen Zusatzinformationen (Lippenbewegungen, Gestiken, ...) kann man seinen Gegenüber trotzdem gut verstehen - wenn der Dialekt nicht zu unterschiedlich ist. Diese Informationen stehen dem Modell mit seinen zwei Assoziativspeichern allerdings nicht zur Verfügung. Die besonders schlechten Klassifikationsleistungen, die sich mit 7% in Abb. 3.59 auf Seite 70 gezeigt haben, sind ein weiterer Extremfall. In diesem Fall wäre ein Mensch nur unter tiefsprechenden Männern aufgewachsen und würde nun zum erstenmal einer Frau mit hoher Stimme begegnen. Vermutlich hätte er zunächst erhebliche Probleme, die Frau zu verstehen. Erst wenn er sich an die Stimme gewöhnt hat - man im Modell die entsprechenden Muster abspeichert - kann er eine ausreichende Klassifikationsleistung erreichen. In diesem Sinne würde das Modell in fast 100% der Fälle die korrekten Vokale extrahieren, wenn die Stimme bekannt ist, und somit die Muster bereits eingespeichert.

In diesem Sinne entspricht Areal D nicht dem letzten Assoziativspeicher, der alle möglichen Personen gespeichert hat, sondern eher einer Art Arbeitsspeicher, in den bei Bedarf die Muster einer Stimme geladen werden. Dann wird eine hohe Klassifikationsleistungen erreicht. Areal C müßte nicht unbedingt mit neuen Mustern geladen werden, möglicherweise würde hier ein generelles Konzept von Vokalen und Formantfrequenzen ausreichen. Um diese Effekte genauer zu untersuchen, könnte innerhalb Areal D mit schnellen hebbschen Lernen die neuen Muster eingespeichert werden. Dann sollte sich nach wenigen Präsentationen von unbekannten Mustern eine korrekte Klassifikation ergeben.

4.2.4 Separation von Signalmischungen

Im Abschnitt 3.6.4 ab Seite 70 sind die Ergebnisse dargestellt, die mit Hilfe des kompletten Systems erzielt wurden. Dazu wurde zunächst geprüft, welche Ergebnisse erzielt werden, wenn die Summe zweier Vokale, die beide im Assoziativspeicher gespeichert sind, als Stimuli verwendet wird. Das Ergebnis der Simulation zeigt, daß der Assoziativspeicher abwechselnd die Assembly des einen Vokals und die Assembly des anderen Vokals aktivert. Dieses Verhalten zeigt sich allerdings nur dann deutlich, wenn die Periodizitätskarte der Überlagerung der einzelnen Vokale entspricht. Dies ist im Normalfall nicht der Fall. Generell ist das Ergebnis bei Verwendung eines Stimulus, der aus der Summe mehrerer Signale besteht, der linearen Überlagerung umso ähnlicher, je weiter die Frequenzkomponenten der einzelnen Signale auseinander liegen. Daher zeigt sich das Oszillieren des Assoziativspeichers besonders deutlich, wenn das [a:] und das [i:] verwendet werden, da deren Formantfrequenzen sich nur wenig überlappen. Würden dagegen das [:] und das [o:] verwendet werden, so würden sich nicht so deutliche Oszillationen im Ergebnis zeigen.

Diese Oszillationen sind allerdings für die Verarbeitung von akustischen Signalen nicht so wichtig. In dem visuellen System, für das die Assoziativspeicher primär entwickelt wurden, wird die Oszillation verwendet, um komplexe Szenen in Einzelteile zu zerlegen, so daß der Assoziativspeicher eine Ellipse und ein Rechteck voneinander trennen kann. Im auditorischen Bereich dagegen will man weniger alle Komponenten des Signals einzeln getrennt haben, als vielmehr sich auf einen Sprecher konzentrieren und die anderen Geräusche dämpfen, oder, aus der Sicht des Assoziativspeichers heraus interpretiert, will man nur *ein* Signal aus der Mischung extrahieren, auf das man sich gerade *konzentriert*. In einer Cocktailpartyumgebung ist ein Mensch auch nicht bemüht, allen Sprechern gleichzeitig zu folgen, sondern hat bereits genug damit zu tun, überhaupt schon einem einzelnen Sprecher zu folgen. Diese Situation wird dementsprechend auch nicht durch den ersten Versuch zur Separation repräsentiert, sondern vielmehr durch die weiteren. Bei diesen werden dem Assoziativspeicher die Muster der Vokale einer Periodizität eingespeichert und anschließend eine Mischung präsentiert, die einen Vokal der gespeicherten Periodizität und drei weitere mit anderer Periodizität enthält. Dieser Assoziativspeicher modelliert also einen Menschen, der sich bemüht,

genau einem Sprecher zuzuhören und der nicht alle gleichzeitig verstehen will.

Die Ergebnisse des Versuches wurden unter Verwendung von vier Vokalen berechnet - jeder Vokal wurde zusätzlich in mehreren Schritten unterschiedlich laut im Vergleich zum Hintergrundgeräusch (der drei anderen Vokale) eingestellt. Aus den Ergebnissen ab Seite 70 geht hervor, daß der Assoziativspeicher durchaus in der Lage ist, die einzelnen Vokale aus der Mixtur zu extrahieren, wenn diese etwas lauter sind als die anderen. Diese Hervorhebung des einzelnen Vokals ist aber durchaus legitim, da zum einem ein Zuhörer bei einer Cocktailparty nicht in einem gleichmäßigen Abstand zu allen anderen Gästen positioniert ist, sondern vielmehr nahe bei seinem Gesprächspartner steht. Zum anderen wird er zusätzlich sein Gegenüber anschauen, und dadurch dessen akustischen Signale zusätzlich verstärkt wahrnehmen.

Die andere Vereinfachung, daß der Assoziativspeicher nur die Vokale der Periodizität gespeichert hat, die dem Zielvokal entsprechen, ist ebenso motivierbar. Dies entspricht dem kurzzeitigen 'Hineinhören' in die Stimme des Gegenübers, das jeder durchführt, der sich mit einem Menschen unterhält, den er nicht kennt. Zunächst kann man seinen Gegenüber nicht so gut verstehen, und muß sich erst an seine Stimme 'gewöhnen - im neuronalen Sinne die Muster in den Assoziativspeichern entsprechend anpassen. Hat man diesen Prozess abgeschlossen, so kann man seinen Gesprächspartner wesentlich leichter verstehen. In diesem Stadium befindet sich der Assoziativspeicher bereits, wenn er nur die Vokale mit der korrekten Periodizität gespeichert hat.

Ein interessanter Aspekt der Ergebnisse ist, daß die beste Trennung, gemessen an der Höhe des größten Nebenmaximums, nicht bei maximalem SNR auftritt. Vielmehr zeigt sich, daß die Separation am besten bei mittlerem SNR funktioniert. Dies scheint auf den ersten Blick unlogisch, da sich mit ansteigendem SNR der Stimulus einem Stimulus annähert, der nur aus dem einzelnen Vokal besteht. Allerdings stimmt dies nur dann, wenn die Lautstärke des Stimulus jeweils geeignet normiert wird. Für die durchgeführten Versuche dagegen hat die Gesamtlautstärke des Signals mit zunehmendem SNR auch zugenommen. Dadurch verschieben sich die Aktivitäten auf der Periodizitätskarte und entsprechen somit bei größerem SNR nicht mehr unbedingt dem Muster, welches im Assoziativspeicher eingspeichert wurde. Dieser Effekt würde durch die Normierung der Stimuluslautstärke behoben werden können. Alternativ dazu könnte auch versucht werden, die periphere Vorverarbeitung lautstärkeunabhängiger zu gestalten. Dies dürfte allerdings der wesentlich komplexere Ansatz sein. Die Normierung des Stimulus dagegen wurde aus folgendem Grund nicht durchgeführt: Sowohl der Ansatz, daß normiert wird, und daß nicht normiert wird, entsprechen der Cocktailparty Situation. In dem einem Fall wird das Störgeräusch leiser, in dem anderen Fall wird der Zielvokal lauter. Der erste Fall entspricht dem Verhalten der Menschen auf der Cocktailparty, bei dem die anderen Gäste sich vom gewünschten Sprecher entfernen und somit ihre Stimmen leiser werden, während die Stimme der Zielperson gleichlaut bleibt. Der zweite und verwendete Fall dagegen stellt ein Verhalten dar, das im normalen Leben die typischere Reaktion auf nicht verstandende Aussagen des Gegenübers darstellt: Der Zuhörende nähert sich mit dem Ohr dem Mund des Sprechers. Dadurch werden nicht die Hintergrundgeräusche leiser, sondern der Sprecher lauter. Dies ist das Verhalten, welches durch den Versuch simuliert wurde. In der Realität zeigt sich auch, daß ein Stimulus durchaus zu laut sein kann. Wenn sich Menschen in der Disko unterhalten, kann es durchaus sein, daß ein zu lautes Gebrüll des Gesprächspartners in das eigene Ohr keine Verbesserung, sondern eher eine Verschlechterung der Verständigung darstellt. Allerdings scheint das Verhalten des Menschen dennoch deutlich unabhängiger von der Lautstärke zu sein als das Verhalten der peripheren Verarbeitung des Modelles. Die schlechtere Wahrnehmung des Menschen tritt im Gegensatz zum Modell erst in extremen Grenzfällen auf.

4.2.5 Zusammenfassung

Die Ergebnisse zeigen, daß Assoziativspeicher zur Klassifikation erfolgreich eingesetzt werden können. In der Konfiguration, wo der Assoziativspeicher die Vokale bereits gekannt hat, konnten fast alle Vokale korrekt erkannt werden. Für unbekannte Vokale war die Erkennungsrate deutlich niedriger, dies entspricht aber dem Verhalten eines Menschen, der den Sprecher erst noch kennenlernen muß.

Die Separationsaufgabe wurde auch erfolgreich bewältigt, das [a:] konnte bereits bei niedrigem

SNR Verhältnis korrekt klassifiziert werden, und mit mittleren SNR konnten alle vier verwendeten Vokale aus dem Störgeräusch extrahiert werden.

Offen ist allerdings, wie der Mensch in der Lage ist, die Laustärke wahrzunehmen. Da das Langnermodell nicht entwickelt wurde, um die Lautstärkewahrnehmung zu erklären ist diese auch nicht direkt in der Ausgabe des Modells enthalten. Daher sind die Assoziativspeicher nicht direkt in der Lage, die Lautstärke zu verarbeiten. Dies wird anhand eines kleinen Beispiels deutlicher: Ein einfacher AM Stimulus, dessen Lautstärke so groß ist, daß er nicht mehr einen einzelnen Fleck erzeugt, sondern zwei Flecken (vgl. Abb. 3.44 auf Seite 58), sieht für den Assoziativspeicher genauso aus, wie zwei leisere AM Stimuli, die jeweils genau an den gleichen Stellen auf der Karte wie der einzelne laute Stimulus Aktivitäten erzeugen. Der Assoziativspeicher müßte um die Lautstärke verarbeiten zu können weitere Informationen bekommen.

Die Verwendung von spikenden Assoziativspeichern für die Modellierung kortikaler Areale ist demnach nicht nur auf das visuelle System beschränkt, für das die Assoziativspeicher ursprünglich entwickelt wurden. Vielmehr zeigt sich, daß sie ein allgemeineres Modell für kortikale Areale darstellen.

4.3 Lokalisations-Experiment

Die Ergebnisse des Lokalisations-Experiments (siehe Seite 75) zeigen eine hohe Übereinstimmung der experimentell gefundenen Werte mit den theoretisch berechneten Werten. Die experimentellen Werte weichen selten mehr als 25 μs ab. Diese extrem hohe Genauigkeit ist insofern interessant, als daß die Schallaufnahmen selber mit 40 kHz durchgeführt worden sind, also die Samples einen zeitlichen Abstand von 25 μs haben. Das bedeutet, daß in den meisten Fällen der Algorithmus genau den richtigen Zeitversatz berechnet hat. Würde er auch nur einen Sample daneben liegen, so würden die Ergebnisse mehr als 25 μs vom korrekten Ergebnis abweichen.

Den Effekt, daß die möglichen Positionen der Differenzen durch die Samplerate 'gequantelt' sind, also nur diskrete Werte annehmen önnen, erkennt man deutlich in der Grafik 3.64 auf Seite 76 in der Mitte. Dort ist eine Sinuswelle zu erkennen, bei der einige Punkte fehlen (zwischen 0 Grad und 20 Grad). Schaut man sich an, wo die Punkte liegen, so sieht man, daß sie über der Sinuswelle liegen, und zwar alle bis auf einen um 25 μs nach oben verschoben. Dieser eine Punkt liegt 50 μs über den Punkten der Sinuswelle. Mit einiger Phantasie kann man die Sinuswelle weiterführen, indem man Punkte um ein Vielfaches von 25 μs verschiebt. Daran erkennt man, daß durch die Diskretisierung des Schalls auch die möglichen detektierbaren Verzögerungen zwischen den Mikrophonen diskretisiert wird. In dem Graphen fällt es nicht so sehr auf, da die möglichen Punktpositionen von der Orientierung abhänge, und nicht konstant sind. Sonst würde man die diskreten Punkte ganz deutlich sehen.

Der Beobachtung entspricht auch der Mittelwert der Betragsabweichungen von ca. 12 μs. Dies ist ziemlich exakt die Hälfte der diskreten Schrittweite, und somit genau der Wert, den man von einem gut funktionierenden Algorithmus erwarten muß. Wenn der Algorithmus optimal ist, so liefert er als durchschnittliche Abweichung bei hinreichend vielen Messungen die Hälfte der diskreten Schrittweite als Ergebnis. Sind die Ergebnisse dagegen sehr schlecht, so werden die Ergebnisse deutlich größer als die Hälfte der diskreten Schrittweite.

Aus diesen Ergebnissen läßt sich ableiten, daß das Langnermodell sich sehr exakt und mit hoher Wiederholgenauigkeit auf die Phasenlage der Periodizität einstellt. Zwar sind die beiden Signale (linkes und rechtes Mikrophon) insofern identisch, als daß sie den gleichen Schall repräsentieren. Allerdings ist durch die unterschiedlichen Strecken auch die Amplitude unterschiedlich hoch. Da die Schallaufnahmen auch nicht in einer Umgebung aufgenommen wurden, die gut gegen Echos abgeschirmt ist, sind die Einflüsse von Echos auch unterschiedlich auf beide Mikrophone, da diese beispielsweise unterschiedliche entfernt zu den Wänden und den Möbeln in einem Zimmer sind. Daher unterscheiden sich die beiden Kanäle einer Aufnahme nicht nur durch den Zeitversatz, sondern auch inhaltlich. Daß dennoch die Schalllokalisation anhand der Ausgaben des Langnermodells derart exakte Ergebnisse liefert, zeigt, daß dessen Ausgabe robust gegenüber unterschiedlichen Variationen des Stimulus ist.

4.4 Psychophysik

Der Vergleich der Ergebnisse aus der Psychophysik mit den Ergebnissen der Simulationen läßt Rückschlüsse auf die Korrektheit und Grenzen der verwendeten Modelle zu. Zum einem wurde der Einfluss der relativen Phasenlage der Frequenzkomponenten in eigenen Versuchen geprüft, zum anderen konnten anhand von Psychophysikversuchen von Christian Kaernbach Indizien für das variierte Modell und Argumente gegen das ursprüngliche Langnermodell gefunden werden.

4.4.1 Ergebnisse zur Untersuchung des Einflusses der Fourier-Phase

Die Ergebnisse des Psychophysikexperimentes zum Einfluß der Phase auf die Lautstärkenwahrnehmung sind in Abschnitt 3.8.1 ab Seite 76 aufgeführt. Zusammenfassend kann gesagt werden, daß für viele Menschen Signale, die synchrone Spikes erzeugen, eine größere Lautstärkewahrnehmung hervorrufen als Signale, deren Spikesynchronität künstlich gestört wurde. Unklar ist allerdings, welcher Faktor tatsächlich für die unterschiedliche Lautstärkenwahrnehmung verantwortlich ist. Im Prinzip kommen die folgenden drei Faktoren in Frage: Summe der Spikes im betreffenden Periodizitätsbereich, maximale Anzahl Spikes in dem betreffenden Periodizitätsbereich und Synchronität der Spikes in dem betreffenden Periodizitätsbereich. Diese drei Werte sind aus den Ergebnissen der Simulationen des Langnermodells extrahiert worden (vgl. Abb. 3.67 und Abb. 3.68 ab Seite 79).

Alle drei Werte werden mit zunehmender Phasenverschiebung kleiner. Für die reduzierte Lautstärkewahrnehmung des Menschen kann allerdings die Höhe des Maximums nicht verantwortlich sein, da dieser Wert nur nach einer längeren Zeit sinnvoll bestimmt werden kann. Die Wahrnehmung vollzieht sich dagegen schon nach wenigen 10 ms. Daher ist dieser Wert eher für das Modell im allgemeinen interessant. Im Prinzip zeigt sich damit, daß die zunehmende Phasenverschiebung nicht nur die Synchronität der Spikes beeinflußt, sondern auch deren Anzahl. Wäre dies nicht so, so müßten die Spikes an jeder Stelle der Karte für sich alleine betrachtet gleich bleiben, nur im Vergleich zu anderen Positionen dürfte sich das zeitliche Verhalten ändern. Dies zeigt, daß, obwohl die Energie des Vokals durch die Phasenverschiebung nicht beeinflußt wird, die Periodizitätsinformationen sich dennoch ändern, zumindest für das Langnermodell. Interessant ist auch, daß für eine kleine Phasenverschiebung das Maximum der Spikes sogar etwas höher wird. Diesen Effekt zeigten auch drei der sechs Probanden, deren Lautstärkenwahrnehmung durch die Phasenverschiebung beeinflußt wurde.

In biologischen Systemen könnte die Anzahl der Spikes im relevanten Bereich - dieser wird so gewählt, daß ein Bereich mit konstanter Periodizität ausgewählt wird - relativ einfach ausgewertet werden. Eine Kodierung, die darauf basiert, daß ein Signal stärker ist, wenn mehr Spikes erzeugt werden, ist häufig anzutreffen. Allerdings nimmt dieser Wert zu stark ab und liegt bei maximaler Phasenverschiebung bei 45% (Psychophysik Ergebnis: 76%).

Ähnlich sieht es mit der Synchronität der Spikes aus. Dieser Parameter könnte auch einen starken Einfluß auf die höheren Areale haben, wie die Summe der Spikes. Allerdings ist der Median für maximale Phasenverschiebung mit 52% auch wesentlich niedriger als die Ergebnisse der Psychophysikexperimente.

Die Prozentwerte sind nur bedingt miteinander vergleichbar, da im biologischen System die Lautstärkewahrnehmung durch viele weitere Faktoren beinflußt wird. Zusammengefaßt kann man daher sagen, daß beide Parameter (Synchronität und maximale Anzahl Spikes) die reduzierte Lautstärkewahrnehmung erklären würden, und vermutlich sind sogar beide gleichzeitig von Bedeutung. Um dies genauer zu überprüfen, müßte man Versuche mit Stimuli durchführen, die nur einen der beiden Werte variieren, und den anderen nicht. Dazu könnten sich Klicks eignen, die so konstruiert werden könnten, das sie bei niedrigen Trägerfrequenzen eine andere Phasenlage als bei höheren Trägerfrequenzen haben.

Interessant ist auch das Ergebniss, daß bei 2 von 8 Probanden die Lautstärkenwahrnehmung nicht systematisch durch die Phasenverschiebung beinflußt wurde. Entweder sind diese beiden tatsächlich nicht in der Lage, die Phasenverschiebung durch unterschiedliche Lautstärkewahrnehmung zu detektieren, oder der Versuch an sich war unklar. Rückfragen ergaben kein spezielles Un-

verständnis zu dem Versuch. Möglicherweise besitzen beide Probanden eine reduzierte Genauigkeit bei der Lautstärkewahrnehmung, so daß der Effekt zwar vorhanden ist, aber durch die ungenaue Wahrnehmung sich nicht sichtbar in den Ergebnissen niederschlägt. Dieser Aspekt sollte weiter untersucht werden. Falls sich herausstellen würde, daß bei manchen Menschen die Lautstärkenwahrnehmung mit durchschnittlicher Genauigkeit funktioniert und die Phasenverschiebung keinen Einfluß auf die Lautstärkenwahrnehmung hat, so würde dies auf eine ungewöhnlich große Differenz in der neuronalen Verarbeitung von Signalen darstellen. Auf welcher Ebene diese Differenzen entstehen wäre ebenso interessant wie die Frage, wie diese Differenzen überhaupt entstehen konnten.

4.4.2 Korrelation vs. Langnermodell

Jedes Modell, welches das auditorische System des Menschen beschreibt, muß möglichst viele Effekte erklären können, die sich zum Beispiel aus der Psychophysik ergeben haben. Ein Beispiel für ein Psychophysikexperiment ist in Abschnitt 3.8.2 auf Seite 80 dargestellt: Um herauszufinden, in wieweit das auditorische System des Menschen zur Periodizitätswahrnehmung auf der Autokorrelation basiert, wurden spezielle Stimuli konstruiert, und mehreren Versuchspersonen präsentiert. Die Ergebnisse zeigen, daß nicht alle Periodizitäten, die sich in Autokorrelogrammen zeigen, vom Menschen wahrgenommen werden. Der Autor unterscheidet *Periodizitäten erster Ordnung* und *Periodizitäten höherer Ordnung*: Periodizitäten erster Ordnung resultieren daraus, daß in einem Stimulus mit Klicks nur der Abstand von Klick zu Klick betrachtet wird, wohingegen für Periodizitäten zweiter Ordnung auch Abstände von einem Klick zum übernächsten analysiert werden. Einer der verwendeten Stimuli hatte ausschließlich Periodizitäten zweiter Ordnung, die in Autokorrelogrammen deutliche Maxima erzeugen. Dennoch konnten die Versuchspersonen diese Periodizitäten nicht oder nur sehr schwach wahrnehmen. Dies deutet darauf hin, daß vom auditorischen System keine einfache Autokorrelation durchgeführt wird.

Interessant ist jetzt, welche Ergebnisse das variierte Langnermodell erzeugt, wenn die gleichen Stimuli verwendet werden. Diese sind in Abb. 3.71 auf Seite 82 dargestellt. Man erkennt, daß das Langnermodell deutliche Probleme hat, die Periodizität zweiter Ordnung wahrzunehmen. Für diese Stimuli ist eine mittlere Grundaktivität mit einem kleinen Maximum bei der entsprechenden Periodizität zu erkennen. Dagegen zeigen die Stimuli mit Periodizitäten erster Ordnung eine deutliche 'Wahrnehmung' der Periodizität, das Maximum ist klar zu erkennen. Auch der Stimulus (kxxxx), der in der Autokorrelation zu weniger Aktivität geführt hat als der Stimulus mit Periodizität zweiter Ordnung (abx), ergibt in den Simulationen wesentlich klarere Maxima als der abx Stimulus. Dieses Verhalten entspricht genau den Ergebnissen der Psychophysikexperimente, bei denen die Versuchspersonen den abx Stimulus nur schwer detektieren konnten, den kxxxx Stimulus dagegen deutlich besser.

Diese deutliche Übereinstimmung zwischen Experiment und Langnermodell zeigt, daß diese Modellierung in die richtige Richtung geht. Daher wird im folgenden die Frage, warum genau das Langnermodell so ähnliche Ergebnisse liefert, geklärt. Die Antwort ist an sich einfach, da ein Langnerschaltkreis so gebaut ist, daß er nur Periodizitäten erster Ordnung erkennen kann, wenn er keine Kammfilterantworten erzeugt. In diesem Fall analysiert der Schaltkreis genau den Abstand zwischen zwei hintereinanderfolgenden Triggerneuronspikes, und nicht den Abstand zwischen zwei Spikes, wenn weitere Spikes dazwischen liegen. Diese Spikes dazwischen setzen den Integrator wieder zurück, und es werden keine Spikes erzeugt. Daß der abx Stimulus dennoch ein kleines Maximum erzeugt hat, zeigt, daß das Triggerneuron nicht immer zu 100% korrekt gespiket hat, sondern auch Spikemomente ausgelassen hat. Wenn diese Spikes die zusätzlichen Spikes zwischen den 'korrekten' Periodizitätsspikes waren, so detektiert der Schaltkreis in diesem Fall die normale Periodizität. Da aber das Triggerneuron oft korrekt arbeitet, ist dieses Maximum sehr klein. Die restliche Aktivität der Karte bei höheren Periodizitäten ergibt sich genau daraus, daß die Zeiten zwischen hintereinanderfolgenden Triggerimpulsen gemessen werden, diese sind immer kürzer als die Grundperiodizität (in ms).

Anders sieht es dagegen aus, wenn das ursprüngliche Langnermodell mit Kammfilterantworten verwendet wird. Dann werden nicht nur Periodizitäten erster Ordnung detektiert, sondern im Prinzip Periodizitäten beliebiger Ordnung. Allerdings werden nicht alle Periodizitäten detektiert,

genauer betrachtet funktioniert ein Langnerschaltkreis dann wie folgt: Die Ausgangssituation ist, daß der Integrator momentan vom Flopneuron gehemmt wird. Jetzt kommt ein Impuls des Stimulus, dieser erzeugt ein Triggerneuron Spike, und dadurch wird die Hemmung des Integrators aufgehoben, weil jetzt das Flipneuron aktiv wird. Nehmen wir an, der Stimulusimpuls war einer der Klicks, die zu der Grundperiodizität gehören. Dann kommt als nächstes der 'Störspike', den aber der Schaltkreis ignoriert. Der Integrator wird im ursprünglichen Modell nicht zurückgesetzt! Beim nächsten Impuls, der jetzt wieder zur Grundperiodizität gehört, hat der Integrator das integrieren gerade beendet und feuert, so daß das Koinzidenzneuron sowohl vom Oszillator als auch vom Integrator Spikes erhält und somit auch feuert. Dieser Prozeß hängt jetzt aber davon ab, auf welchen Stimulusimpuls das Triggerneuron das erste Mal feuert. Gehört dieser nicht zur Grundperiodizität, so werden in diesem Fall die Spikes der Grundperiodizität ignoriert. Die Chance auf den ersten Impuls der Grundperiodizität zu treffen, ist aber immerhin 33%, somit ist die Antwort eines Modells mit Kammfilterantworten wesentlich stärker, als die Antwort eines Modells ohne Kammfilterantworten. Im Prinzip entsteht ungefähr die Autokorrelation im Antwortverhalten, nur daß diese jede Periodizität von jedem Stimulusimpuls aus repräsentiert, das ursprüngliche Langnermodell dagegen nur die Periodizitäten, bei denen das Triggerneuron gerade das Flopneuron deaktiviert hat. Diese Überlegungen erklären die Ergebnisse des ursprünglichen Langnerschaltkreises, dargestellt in Abb. 3.72 auf Seite 83.

Man erkennt einen deutlichen Unterschied zwischen dem abx Stimulus und dem Zufallsstimulus x. Dagegen ist die Höhe des Maximums bei der Grundperiodizität für den abx Stimulus etwa so hoch wie für den kxx Stimulus. Dies entspricht somit stärker dem Ergebnis der Autokorrelation als den biologischen Ergebnissen. Ein Modell, welches prinzipiell Kammfilter erzeugt, paßt demnach nicht so gut zu diesem Psychophysikexperiment, wie ein Modell, welches keine Kammfilterantworten erzeugt. Für ein Klassifikationssystem wäre es relativ einfach, die Ergebnisse des ursprünglichen Langnermodells zu klassifizieren, insbesondere den x und den abx Stimulus auseinanderzuhalten.

Allerdings wurde das ursprüngliche Langnermodell inzwischen um ein spezielles Inhibitionssystem erweitert. Dieses basiert darauf, daß zwar das ursprüngliche Langnermodell verwendet wird, aber jeder Schaltkreise die Schaltkreise mit der jeweils halbierten Periodizität hemmen. Abgesehen von dem großem Aufwand, den diese Hemmung bedeutet, und der Tatsache, daß diese Hemmung nicht auf alle weiteren Schaltkreise (drittel Periodizität, fünftel Periodizität, ...) wirkt, sprechen auch die angegebenen Ergebnisse gegen diese Hemmungs-These: Zwar wurde dieses Hemmungssystem nicht mit Simulationen getestet, aber man kann sich die Effekte theoretisch überlegen. Für einfache Stimuli wie einem amplitudenmodulierten Sinus oder einem Klicktrain Stimulus würde das Hemmungsystem die Kammfilterantworten reduzieren, aber nicht komplett unterdrücken (wegen den $N/3$, $N/5$, ...). In diesen Fall würden tatsächlich zu Beginn des Stimulus kurz Kammfilterantworten auftreten, um dann bald durch die neu eingeführte Hemmung unterdrückt zu werden. Für den abx Stimulus sieht das allerdings anders aus: Die Unterdrückung basiert darauf, daß der Schaltkreis mit N Hz den Schaltkreis mit $N/2$ Hz unterdrückt. Dies würde für den abx Stimulus mit 100 Hz bedeuten, daß die Spikes des Schaltkreises bei 100 Hz durch einen Schaltkreis mit 200 Hz durch Hemmung reduziert werden. Nur so könnte dieses Maximum, welches durch die Kammfilterantworten auftritt, auf ein Niveau reduziert werden, das die Ergebnisse der Psychophysikexperimente widerspiegelt. Dies wird vermutlich auch zu einem geringen Teil passieren, aber nicht so stark, wie es notwendig wäre: Damit der 200 Hz Schaltkreis feuert, muß eine entsprechende Periodizität im Stimulus enthalten sein. Der abx Stimulus enthält immer zwischen den 100 Hz Klicks einen zusätzlichen Klick. Wenn dieser genau in der Mitte liegt, so feuert der 200 Hz Schaltkreis tatsächlich und kann den 100 Hz Schaltkreis hemmen. Abhängig von der Schärfe des neuronalen Tunings kann aber schon ein Klick, der sich nicht bei 5 ms (in der Mitte), sondern bei z.B. 4 ms befindet, zur Hemmung des 100 Hz Schaltkreises führen. Dieser Klick führt zu einem Spike des 250 Hz Schaltkreises, dieser hemmt dann den 125 Hz Schaltkreis und dieser dann den 62,5 Hz Schaltkreis. Der 100 Hz Schaltkreis ist aber durch diese Hemmung nicht betroffen. Man sieht, daß es wesentlich mehr Positionen für Klicks zwischen den 100 Hz Klicks gibt, die zu keiner Hemmung führen, als Positionen, die zur Hemmung führen (2.5 ms würde auch zu Hemmung führen, da der 400 Hz Schaltkreis den 200 Hz, den 100 Hz, ... hemmen würde). Man sieht, daß die Hemmung zwar auf den ersten Blick ähnliche Effekte erzeugt wie der variierte Schaltkreis. Allerdings zeigen diese

theoretischen Überlegungen, daß bei bestimmten Stimuli der variierte Schaltkreis dem biologischen System deutlich näher kommt als der ursprüngliche Schaltkreis mit Hemmung.

4.4.3 Periodizitätsschwankungen natürlicher Sprache

Ein wichtiges Hilfsmittel für Quellentrennung im Bereich Cocktailparty-Effekt kann die Periodizität eines Sprechers sein. Ist dem Zuhörer die ungefähre Stimmlage seines Dialogpartners bekannt, so kann er mit einfachen Mittel versuchen, alle anderen Periodizitäten zu unterdrücken. Dies wäre unter Verwendung des Langnermodells einfach zu realisieren: sobald bekannt ist, welche Periodizität die Stimme des Sprechers hat, kann diese Spalte der Karte stärker beachtet, und auch die anderen Bereiche durch Inhibition unterdrückt werden. Damit dieses Vorgehen allerdings Sinn macht, muß geklärt werden, wie stark die Periodizität eines festen Sprechers schwankt. Die Ergebnisse dieser Untersuchung anhand von fünf Sprechern mit je 20 Vokalen ist in Abb. 3.73 auf Seite 84 dargestellt. Wie man erkennt schwankt die Periodizität durchaus, allerdings für fast alle Vokale innerhalb eines kleinen Bereiches von einer halben Oktave oder weniger. Das bedeutet, daß die angegebene Strategie, sich auf eine bestimmte Periodizität zu konzentrieren, durchaus Erfolg haben kann. In einer Situation, in der beispielsweise zwei Sprecher mit unterschiedlicher Periodizität auf der Karte Aktivitäten erzeugen, werden die Aktivitäten des gewünschten Sprechers in dem erwarteten Periodizitätsbereich liegen. Das Aktivitätsmuster wird nicht komplett so aussehen, wie wenn der Sprecher allein sprechen würde, aber dadurch, daß in diesem Fall bekannt ist, welcher Teil der Gesamtaktivität der Karte von Bedeutung ist, kann das Muster durch die Assoziativspeicher vervollständigt, und die uninteressanten Bereiche unterdrückt werden. Dieses Verfahren funktioniert gut, solange die Periodizitäten der Sprecher weit genug auseinanderliegen. Sind dagegen die Periodizitäten von simultan aktiven Sprechern sehr ähnlich, so müssen andere Mechanismen die Quellentrennung unterstützen. Vorstellbar sind hier eine Vielzahl von Mechanismen - die aber nicht Gegenstand dieser Arbeit sind - wie zum Beispiel binaurales Hören.

Kapitel 5

Anwendungen

Obwohl diese Forschungsarbeiten Grundlagenforschung darstellen, ergeben sich aus den Ergebnisse einige potentielle Anwendungsgebiete, die im folgenden aufgeführt werden. Die Anwendungsmöglichkeiten sind dafür in zwei Gebiete unterteilt, zum einem die Verbesserungen für technische Systeme und zum anderen der Einsatz in der Medizin für die Unterstützung von Menschen mit Hörproblemen.

5.1 Einsatz in der Technik

Viele Verbesserungen von technischen Systemen haben sich dadurch ergeben, daß die Strategien der Natur, die sich über lange Zeiträume bewährt haben, in technische Systeme umgesetzt werden. Da die Schallverarbeitung durch den Menschen jeder bisherigen technischen Lösung überlegen ist, wird im folgenden aufgezeigt, an welchen Stellen die gefundenen Ergebnisse technische Systeme verbessern können.

5.1.1 Quellen Trennung

Eine komplexe Aufgabe für schallverarbeitende Systeme ist die Trennung einer Mixtur von mehreren Schallsignalen. Mathematisch formuliert sieht das wie folgt aus:

$$R_i(t) = \sum_{n=1}^{k} a_{n,i} * S_n(t + \delta_{n,i})$$

Dabei stellt R_i die Aufnahme des i-ten Mikrofons zum Zeitpunkt t dar. Die Schallerzeuger S_n können unterschiedliche Entfernungen zu den Mikrofonen haben, daher ergeben sich unterschiedliche Laufzeiten, die durch $\delta_{n,i}$ ausgedrückt werden. Zusätzlich werden die Signale unterschiedlich stark gedämpft, so daß die Amplitude des aufgenommenen Signals $a_{n,i}$ ebenfalls von den Positionen des Mikrofon und des Schallerzeuger abhängt.

Die angegebene Gleichung stellt in einer Hinsicht eine Vereinfachung der realen Verhältnisse dar: das Auftreten von Echos wird nicht berücksichtigt. Wenn beliebige Echos berücksichtigt werden sollten, so wird aus der Linearkombination der Schallquellen eine Überlagerung von Faltungen der Schallerzeuger S_n mit einer Funktion, die die Echoverhältnisse repräsentiert.

Die Aufgabe ist in beiden Fällen, aus den aufgenommenen Signalen R_i die ursprünglichen Signale S_n zu berechnen. Für eine mathematische Lösung müßten für den einfachen Fall (ohne Echos) die Parameter $a_{n,i}$ und $\delta_{n,i}$ berechnet werden. Diese Problematik ist der Gegenstand der Forscher, die sich mit *ICA* (Independent Component Analysis) beschäftigen. Für den Fall, daß mindestens soviele Mikrofone wie Schallquellen vorhanden sind, kann die ICA eine Lösung darstellen, wobei die zeitlichen Verzögerungen und die Echos erhebliche Probleme mit sich bringen. Außerdem kann mit Hilfe der ICA nur Sprachaufnahmen analysiert werden, die bereits komplett aufgenommen

sind. Im Gegensatz zu dem biologischen Modell, welches Sample für Sample analysiert, muß also für eine Analyse mit ICA das gesamte Sprachsignal zunächst aufgenommen werden. Erst dann kann es von einem technischen System analyisiert werden. Anschließend kann entsprechend der Ergebnisse reagiert werden. Dieses System wäre nicht in der Lage einen Dialog zu führen, weil die Analyse der Sprache erst durchgeführt werden kann, wenn der Gesprächspartner bereits fertig ist. Selbst wenn die Analyse sehr schnell gehen würde könnte das System nicht zwischendurch den Gesprächspartner unterbrechen, um einen wichtigen Einwand gleich zu tätigen, weil in dem Moment das technische System überhaupt nicht weiß, was überhaupt geredet wird. Der Mensch ist dagegen in der Lage, alle diese Analysen in dem Moment durchzuführen, in dem er etwas hört. Er kann sich auch mit seinen zwei 'Mikrofonen' in komplexen akustischen Situationen, wie zum Beispiel in einer Cocktailpartysituation, zurecht finden. Die Aufgabe, sich auf einen Sprecher zu konzentrieren und die anderen Geräuschquellen zu unterdrücken, wird bis zu einem gewissen Grad ohne Probleme gelöst, und daß Menschen dazu in der Lage sind, ihren Gesprächspartner an sinnvollen Stellen zu unterbrechen, ist jeden bekannt.

Eine These, wie der Mensch dieses Problem löst, ist die Nutzung der Spikesynchronität von gleichen Schallquellen. In dieser Forschungsarbeit wurde gezeigt, daß bei den untersuchten Schallsignalen die Neuronen, die zum gleichen Stimulus gehört haben, synchron gefeuert haben. Durch diese Spikesynchronität können die Frequenzanteile wieder zu einzelnen Quellen zusammengefügt und somit das Bindungsproblem gelöst werden. Diese Synchronitäten sind schon auf der Ebene des Colliculus Inferior zu finden und können somit durch kortikale Areale direkt ausgewertet werden. Dieses Prinzip der Spikesynchronität kann helfen, die einzelnen Quellen zu trennen, bzw. sich auf eine einzelne Quelle zu konzentrieren und andere Quellen zu unterdrücken.

5.1.2 Quellenlokalisation

Das Problem der Quellenlokalisation und der Quellentrennung ist eng miteinander verknüpft. Wenn die Quellen getrennt sind, ist die Lokalisation einfach (wenn aus obiger Gleichung die δ_n Werte bekannt sind, dann kann die Position direkt berechnet werden, wenn ausreichend viele Mikrofone vorhanden sind), entsprechend vereinfacht sich die Quellentrennung, wenn die Position der Quellen bekannt sind. Schwierig ist dagegen die Lösung des einen Problems, wenn die Lösung des anderen noch unbekannt ist. Dies gilt in beide Richtungen. Der biologische Ansatz stellt sowohl für die Trennung als auch für die Lokalisation einen unabhängigen Weg zur Verfügung. Die Quellenlokalisation kann mit Hilfe der Spikes der Koinzidenzneurone durchgeführt werden. Dies wurde mit Hilfe des Langnermodells in dieser Arbeit gezeigt: Durch die Korrelation der Spikes der Koinzidenzneurone des linken Colliculus Inferior mit den jeweiligen gleichen Koinzidenzneuronen des rechten Colliculus Inferior ergeben sich Maxima in den Korrelogrammen an Positionen, die den Positionen der Schallquellen entsprechen. Diese Analyse kann unabhängig davon durchgeführt werden, ob bekannt ist, welche Neuronen zu welcher Schallquelle gehören.

5.1.3 Sprachverarbeitung

Im Gegensatz zu den bisherigen Ansätzen, die im wesentlichen auf der Analyse des Frequenzspektrums basieren, wurde in dieser Arbeit gezeigt, daß die Periodizität ein wichtiges Kriterium ist, um Sprachsignale zu verarbeiten. Zum einen ermöglicht die Analyse der Periodizität die Zuordnung der Frequenzkanäle zu den entsprechenden Sprechern, zum anderen fallen Störgeräusche und Rauschen oftmals nicht in die gleichen Periodizitätsbereiche wie das zu untersuchende Signal. Da aber Sprachsignale des gleichen Sprechers immer die gleichen Periodizitäten haben, wird durch die Aufteilung der Frequenzkanäle in verschiedene Periodizitäten das Sprachsignal automatisch von vielen Störgeräuschen getrennt. Wenn die Periodizität des gewünschten Sprechers bekannt ist (entsprechend dem 'Hineinhören', wenn man sich mit einer unbekannten Person unterhält) sind viele andere Schallsignale schon herausgefiltert, da diese eine andere Periodizität besitzen. Daher ist die Verwendung einer zweidimensionalen Repräsentation mit der Periodizität auf der einen Achse und der Trägerfrequenz auf der anderen Achese für die Analyse von Sprachsignalen und anderen natürlichen Schallsignalen besonders geeignet.

Außerdem können die Variationen der Tonhöhe, die ein Sprecher während einer längeren Zeit macht, weitere Hinweise für das Sprachverarbeitunssystem darstellen. Typischerweise wird die Tonhöhe am Ende von Fragesätzen etwas höher, diese Information könnte für Algorithmen, die auf einer höheren Ebene die Syntax und Semantik der Sprache analysieren, von großer Bedeutung sein.

5.1.4 Sprecheridentifikation

Ein weiteres Anwendungsfeld, für das die Periodizitätsinformationen wichtig ist, stellt die Sprecheridentifikation dar. Bei dieser Aufgabenstellung soll ein technisches System eine Sprachaufnahme nicht in Bezug auf den Inhalt analysieren, sondern in Bezug auf die Identität der sprechenden Person. Diese Aufgabe entspricht ungefähr dem Gegenteil der Spracherkennung. Anstatt möglichst sprecherunabhängig den textuellen Inhalt zu extrahieren will man bei der Sprecheridentifikation möglichst unabhängig von dem textuellen Inhalt den Sprecher extrahieren. Für diese Aufgabe stellt die Tonhöhe des Sprechers ein wichtiges Indiz dar, daß mit Hilfe des Langnermodells auf einfache Weise extrahiert werden kann. Dazu ist beriets das Modell ohne die kortikalen Areale ausreichend, es müßten keine Assoziativspeicher simuliert werden.

5.1.5 Vorverarbeitung für Liederkennungssystem

Ein System, welches in der Lage ist, aus einer großen Menge von Liedern das Lied herauszufinden, welches vorgesungen oder gepfiffen wird, wäre für viele Anwendungen interessant. Zum Beispiel in einem Musikladen oder für Radiosender könnte diese automatische Auswahl eine große Hilfe darstellen. Eine wichtige Strategie für ein solches System könnte die Periodizitätserkennung darstellen, da Musik aus Melodie und Rhythmus besteht, und somit das Langnermodell eine große Hilfe für die Melodieerkennung sein könnte. Für die zeitliche Struktur müßte allerdings ein zusätzlicher Mechanismus verwendet werden.

5.1.6 Realisierung der Periodizitätskarte für technische Anwendungen

Soll das Langnermodell für einen der obigen Zwecke eingesetzt werden, so muß aus Geschwindigkeitsgründen eine vereinfachte Variante implementiert werden. Im Prinzip wird ein Algorithmus benötigt, der auf einfache Art und Weise eine Periodizitäts/Trägerfrequenzkarte möglichst ähnlich wie das Langnermodell implementiert. Dieser könnte aus folgenden drei Teilen bestehen:

1. Eine Filterbank, um die Trägerfrequenzen zu extrahieren.
2. Ein triggerneuronähnlicher Mechanismus, der markante Stellen aus den Trägerfrequenzen extrahiert.
3. Ein Periodizitätsdetektor der die Periodizitäten des Triggerneuronmechanismus feststellt.

Als Filterbank kann die Gammatonfilterbank eingesetzt werden, die zwar nicht so schnell berechnet werden kann wie die Fast Fourier Transformation (FFT). Allerdings wird ein Filtersystem benötigt, das zu jeder Trägerfrequenz eine Zeitfunktion erzeugt und nicht nur einen Parameter zu einem festen Zeitfenster, wie es bei der FFT der Fall ist. Für einen Filterschritt benötigt die Gammatonfilterbank pro Kanal 12 'Multiply-Add' (mulitplizieren zweier Zahlen und aufaddieren auf eine weitere Zahl) Rechenschritte, bei 50 Kanälen und 40 kHz Abtastrate ergibt dies 24 Millionen Multiply-Add Rechenschritte pro Sekunde. Dies ist beim aktuellen Stand der Technik (2003) kein Problem - aktuelle DSPs berechnen weit über 100 Millionen Multiply-Add pro Sekunde. Der erste Schritt kann also entweder direkt vom PC aus durchgeführt werden, oder, zur Entlastung des Zentralprozessors, vorher von einem DSP ausgeführt werden. Als Ergebnis werden mehrere MByte/s erzeugt, die per USB oder einer anderen Standardschnittstelle zum PC transferiert werden könnten.

Der zweite Schritt, die Simulation eines Triggerneurons, erfordert im Prinzip eine Art Schwellwertprozess. Es muß ein Spike erzeugt werden, wenn die Amplitude des gefilterten Signals eine

Schwelle überschreitet. Allerdings darf dieser Spike nur einmal im Moment des Überschreitens erzeugt werden. Diese Operation erfordert im Prinzip nur 2 Vergleiche pro Kanal und Schritt, nämlich einen Vergleich, ob der letzte Wert noch unter der Schwelle lag und der aktuelle schon über der Schwelle liegt. In diesem Fall wird ein Ausgangsbit auf '1' gesetzt (dies repräsentiert den Spike). Da die Modulationsfrequenz und nicht die Trägerfrequenz detektiert werden soll, ist zusätzlich noch eine Art Refraktärzeit notwendig, damit die Spikes nur dann erzeugt werden, wenn die Einhüllende über den Schwellwert steigt, und nicht jedesmal, wenn die Trägerfrequenz diesen Schwellwert überschreitet. Diese Berechnungen könnten genauso vom 'Gammaton-DSP' übernommen werden. Der Vorteil wäre, daß nicht mehr 16 Bit Integer oder sogar 80 Bit Gleitkommazahlen übertragen werden müßten, sondern nur noch 1 Bit pro Kanal. Dies reduziert die Datenrate auf 2000 kBit/s - also 250 kByte/s - bei 50 Kanälen und 40 kHz Abtastrate.

Der aufwendigste Schritt ist die Analyse der Periodizität der 'Triggerspikes'. Im einfachsten Fall kann einfach die Pause zwischen zwei Spikes gemessen, und daraus die Periodizität bestimmt werden. Dieser Wert ergibt dann einen Index in eine Karte ähnlich der IC-Karte des Langnermodells: Die Y-Achse wird wieder durch die Filterbank erzeugt, die X-Achse durch die Periodizitätsanalyse. Das Aufwendige an diesem Schritt ist demnach nicht die Analyse der Periodizität, die mit wenigen Vergleichen und Subtraktionen durchgeführt werden kann. Allerdings ergibt sich eine hohe Datenrate, wenn zu jedem Zeitschritt die gesamte Karte übertragen werden soll: für jeden Schritt muß der Wert an jeder Stelle übertragen werden, dies entspricht in diesem Beispiel 50*50*40000*2 Byte/s=200 MByte/s (50 Trägerfrequenzen*50 Periodizitäten*40 kHz Abtastrate*z.B. 16 Bit Integer Werte). Wenn möglich würde sich eine Art spärliche Datenübertragung anbieten [17], so daß immer nur die wirklich aufgetauchten Spikes übertragen werden. Wenn die Karte entlang der Periodizitätsachse bis zu einer Periodizität von 250 Hz gehen würde, so würde ein solches Neuron maximal 250 Spikes pro Sekunde erzeugen. Ein Neuron bei der niedrigsten Periodizität - z.B. 50 Hz - würde maximal 50 Spikes pro Sekunde erzeugen. Über die gesamte Periodizitätsachse gemittelt ergeben sich 150 Spikes pro Sekunde. Wenn im theoretischen Extremfall alle Neuronen mit maximaler Spikerate feuern würden, so würde dies 50*50*150=375000 Spikes pro Sekunde entsprechen. Für jeden Spike müßte man die Position und den Zeitpunkt übertragen, wobei man sich in einem Echtzeitsystem den Zeitpunkt des Spikes sparen könnte, da der Übertragungsmoment dem Spikezeitpunkt entspricht. Bei einer Kartengröße von 50*50 Neuronen reichen für die Position 12 Bit, für jede Achse 6 Bit ($2^6 = 64$). Daraus ergibt sich eine Datenrate von 562500 Byte/s. Würde man den Spikezeitpunkt bei jedem Spike übertragen, so käme man bei Verwendung von 32 Bit auf 16500000 Byte pro Sekunde. Selbst dieser Wert ist wesentlich kleiner als die 200000000 Byte pro Sekunde, die bei der Übertragung der kompletten Karte zu jedem Zeitpunkt anfallen würde. Eine weiter Möglichkeit, die Datenübertragung zu optimieren ergibt sich wie folgt: durch geschickte Kodierung könnte man auch sowohl die Zeitinformationen übertragen als auch die Datenrate gering halten. Dazu überträgt man nur den 12-Bit Wert der Spikepositionen, und sendet die Zeitinformationen als eine Art 'Heart Beat'. Immer wenn die Berechnung eines neuen Simulationsschrittes beginnt, wird als erstes der aktuelle Zeitpunkt übertragen. Dann werden alle Spikes gesendet, die in diesem Simulationsschritt erzeugt werden. Wenn keine Spikes erzeugt werden, so wird einfach der nächste Simulationsschritt berechnet, und der neue Zeitpunkt gesendet. Dadurch würde in einer stillen Umgebung, wo keine Spikes erzeugt werden, eine Datenrate von $40000 * 4 = 160000$ Byte pro Sekunde übertragen werden (Abtastrate*4 Byte für die 32-Bit Zeitpunkte). Im anderem Extremfall, bei der alle Neuronen maximal schnell spiken würde dies $160000 + 562500 = 722500$ Byte pro Sekunde entsprechen. Dies wären weniger als 5% der Datenrate, die sich bei der vorherigen Kodierung ergeben hat, bei der zu jedem Spike der Zeitpunkt übertragen wurde. Allerdings hat die Übertragung der Zeitpunkte der Spikes als 32-Bit Zahl den Nachteil, daß man maximal 30 Stunden kodieren kann: mit 32 Bit kann die Simulation $2^{32}/40000/3600 = 29,8$ Stunden laufen, dann wiederholen sich die Zeitpunkte. Da aber sowieso zu jedem Simulationsschritt ein 'Heart Beat' übertragen wird, kann man diesen auch nur als festen Wert kodieren, und der Empfänger berechnet sich mit der gewünschten Auflösung den aktuellen Zeitpunkt seit Simulationsbeginn. Dadurch wird auch die Datenrate nochmals reduziert. Das endgültige Protokoll könnte dann wie folgt aussehen: Es werden immer 6 Bit Zahlen transferiert. Ist diese Zahl 63, so ist dies ein 'Heart Beat', der Empfänger erhöht seine Zeitvariable um eins. Ist dieser Wert kleiner als 50 (0-49), so ist

	Technischer Algorithmus	Biologischer Algorithmus
Simulation von 50 ms	0,2 s	21 s
Simulation von 500 ms	1,8 s	210 s

Tabelle 5.1: Laufzeiten der beiden Algorithmen im Vergleich. Es wurden jeweils Karten mit der Größe 50x50 erzeugt, einmal mit einem Stimulus von 100 ms und einmal mit einem Stimulus von 1000 ms

repräsentiert dieser und der nächste Wert einen Spike an der entsprechenden Position. Zusätzliche Informationen könnten noch durch die Zahlen 50-62 kodiert werden. Ein weiterer Vorteil ist auch, daß durch den 'Heart Beat' die Datenübertragung synchronisiert wird, falls zwischendurch Bytes verloren gehen. In diesem Fall wäre nicht klar, ob ein empfangener 6-Bit Wert das erste oder ads zweite Byte der Spikeposition darstellt. Durch den 'Heart Beat' wird die Datenübertragung zu jedem Simulationsschritt neu synchornisiert. Die maximale Datenrate ergibt sich bei dieser Kodierung zu $40000 * 1.5 + 562500 = 622500$ Byte pro Sekunde, bei einer typischen Spikerate von ca. 20 000 Spikes pro Sekunde müsten $60000 + 30000 = 90000$ Bytes pro Sekunde übertragen werden. Beide Datenraten sind sogar mit der älteren USB 1.1 Schnitstelle zu bewältigen, deren maximale Datenrate theoretisch $1,5$ MByte pro Sekunde ist.

Im Prinzip könnte dieser einfache Algorithmus eine ähnliche Karte erzeugen wie das Langnermodell. Eine Implementierung des Algorithmus wurde verwendet, um die in Abb. 5.1 dargestellten Karten zu erzeugen. Für einen besseren Vergleich wurden zusätzlich die Karten, die sich aus dem biologischen Langnermodell ergeben, mitabgebildet. Man erkennt, daß der Unterschied zwischen den Ergebnissen des schnelleren Technikalgorithmus und denen des langsameren biologischen Algorithmus nicht besonders groß sind. Da die Periodizitätseichung des technischen Algorithmus nicht exakt der des biologischen Netzwerkes entspricht, liegen die Aktivitätsflecken der linken Bilder nicht auf der selben X-Position wie die Flecken der rechten Bilder. Dies ist aber kein wesentlicher Unterschied - durch eine entsprechende Eichung der X-Achse des technischen Algorithmus wäre die Position der Flecken an der gleichen Stelle.

Eine genauere Untersuchung wäre notwendig, um zu prüfen, ob die Ergebnisse, die sich mit dem biologischen Algorithmus gezeigt haben auch mit Hilfe des technischen erreicht werden könnten. Dagegen können die Laufzeiten direkt verglichen werden. Diese sind in Tabelle 5.1 angegeben. Der technische Algorithmus zeigt eine deutlich höhere Geschwindigkeit im Vergleich zum biologischen. Die Simulationen wurden mit einem 800 MHz Athlon unter Linux durchgeführt. Durch die Verwendung eines aktuellen PCs müßte eine Echtzeitdarstellung des Langnermodells möglich sein, der technische Algorithmus ist auf einem 800 MHz Rechner um Faktor vier zu langsam. Mit einem Rechner, dessen CPU mit über 3 GHz getaktet ist, müßte die Berechnung in Echtzeit lauffähig sein. Außerdem kann der Algorithmus noch stärker optimiert werden, da die aktuelle Implementierung direkt den Logarithmus und die Exponentialfunktion einsetzt. Wenn diese durch Lookup Tabellen ersetzt werden würden, könnte eine deutliche Geschwindigkeitsteigerung erreicht werden, ca. 38 % der Zeit werden nur durch diese Berechnungen verbraucht.

5.2 Einsatz in der Medizin

Viele Erfolge der Medizintechnik lassen sich auf ein besseres Verständnis der biologischen Vorgänge zurückführen. In den letzten Jahren konnten einige Erfolge mit technischen Implantaten erzielt werden - wie zum Beispiel mit Cochlea-Implantaten - die es kranken Menschen ermöglicht haben, besser mit der Umwelt zu interagieren und damit die Lebensqualität zu steigern. Auch im visuellen Bereich existieren neue Entwicklungen, die langfristig viele blinde Menschen zumindest rudimentäre optische Reize mit Hilfe von technischen Implantaten wahrnehmen lassen. Im folgenden werden zwei Anwendungsgebiete dargestellt, für die die Ergebnisse dieser Dissertation von Bedeutung sein könnten.

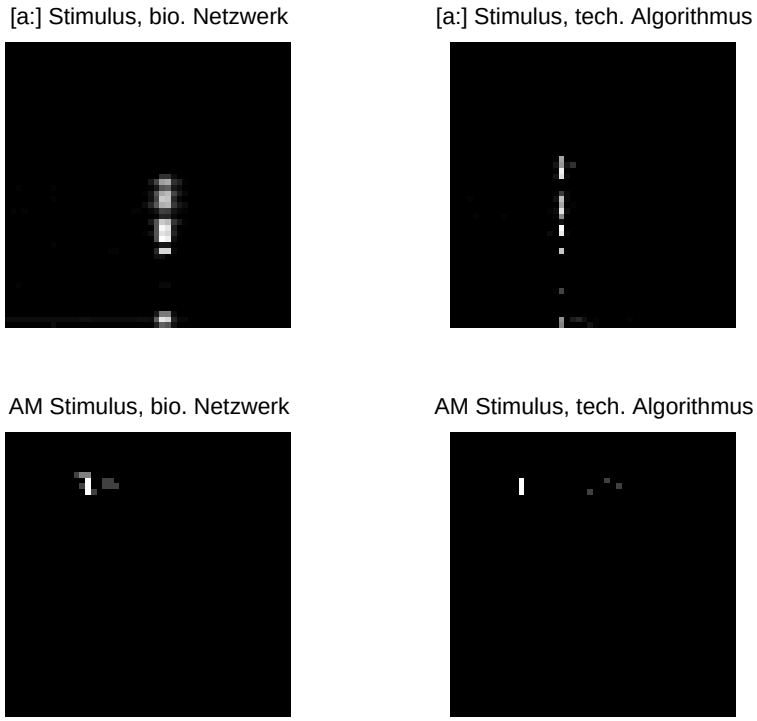

Abbildung 5.1: Vergleich des technischen Algorithmus mit dem biologisch motivierten Algorithmus zur Erzeugung von Periodizitätskarten. Dargestellt ist jeweils die Summenkarte, die sich im Colliculus Inferior ergeben würde. Links sind die Ergebnisse des biologisch motivierten Algorithmus, rechts die des technischen. Es wurden zwei verschiedene Stimuli verwendet: In der ersten Zeile der Vokal [a:] Sinus, in der zweiten Zeile ein AM Signal (100 Hz/2500 Hz).

5.2.1 Verbesserung von Cochlea-Implantaten

Die Grundvoraussetzung von Cochlea-Implantaten ist ein rudimentäres Verständnis der Funktion der Cochlea. Diese wurde in vielen Forschungsprojekten untersucht, und somit sind weite Bereiche ihrer Funktion bekannt [18]. Kurz zusammengefaßt ist die Hauptaufgabe der Cochlea die Schallreize, welche durch das Trommelfell, den Hammer, Amboß und Steigbügel aus dem Medium Luft in ein flüssiges Medium übertragen worden sind, entlang einer Frequenzachse aufzuspalten. Durch sogenannte Haarzellen wird die entstehende mechanische Wanderwelle in elektrische Reize umgesetzt. Die Wanderwelle wird abhängig von den im Schallsignal enthaltenen Frequenzen an unterschiedlichen Positionen maximale Auslenkungen erreichen. Tiefe Frequenzen erzeugen am Ende der Cochlea starke Auslenkungen der Haarzellen, hohe Frequenzen am Anfang der Cochlea. Diese Auslenkung der Haarzellen werden als elektrische Impulse weitergegeben.

Die Aufgabe eines Cochlea-Implantat ist nun diese Umsetzung des Schallsignals in elektrische Impulse. Eingesetzt werden Cochlea-Implantate bei Menschen, deren Hörsystem eigentlich funktioniert, bei denen aber aus unterschiedlichen Gründen die elektrischen Reize der Haarzellen nicht erzeugt werden. Durch das Einführen einer Elektrode können elektrische Reize erzeugt werden, die der Mensch dann als Geräusch wahrnimmt. Eine Elektronik steuert dann die Elektrode so an, daß das durch ein Mikrofon aufgenommene Schallsignal auf geeignete Weise in elektrische Reize umgewandelt wird. Dazu ist eine Filterung des Schallsignals notwendig, da an der Stelle, an der die Elektrode die elektrischen Reize erzeugt, bei einem gesunden Menschen nur ein bestimmter Frequenzbereich zu einer hohen Auslenkung der Wanderwelle führt. Neuere Implantate arbeiten daher mit mehreren Elektroden (z.B 22 Elektroden für das Implantat Nucleus 22) an verschiedenen Stellen der Cochlea um das Schallsignal möglichst genau im Frequenzraum abzubilden.

Unter Berücksichtigung der Ergebnisse, die sich für die höheren Areale, insbesonders den Colliculus Inferior, ergeben, könnte die elektronische Filterung evtl. modifiziert werden, um wichtige Inhalte des Schallsignals besser zu repräsentieren [27]. In dieser Dissertation wurde gezeigt, daß die Periodizität des Schallsignals von großer Bedeutung für die weiteren Analysen durch das Hörsystem ist. Die folgende erweiterte Vorverarbeitung für Cochlea-Implantate könnte für deren Benutzer eine Verbesserung der akustischen Wahrnehmung darstellen:

- Der Benutzer wählt einen für ihn interessanten Periodizitätsbereich.
- Das Schallsignal wird in allen Frequenzkanälen auf die enthaltenen Periodizitäten untersucht.
- Die Steuerelektronik verstärkt die Kanäle, in denen die gewählte Periodizität vorhanden ist und schwächt die anderen Kanäle ab.
- Dadurch würde das nachfolgende Hörsystem sich automatisch auf Bereiche konzentrieren, die von Interesse sein könnten.

Diese erweiterte Vorverarbeitung hätte den Vorteil, daß der Benutzer sich stärker auf gewünschte Bestandteile des Schallsignals konzentrieren könnte. Zum Beispiel ist ein Problem der üblichen Cochlea-Implantate, daß die Nutzer oft durch Umgebungslärm stärker beeinträchtigt werden als Menschen mit einem funktionierendem Hörsystem.

Alternativ könnte die manuelle Wahl von bevorzugten Periodizitätsbereichen eine große Hilfe für Cochlea-Implantatnutzer bei Cocktailpartysituationen darstellen, die auch große Probleme für die betroffenen Leute darstellen. Da die Schallsignale nur eingeschränkt von den Elektroden übertragen werden, haben die Nutzer große Probleme, sich auf einzelne Sprecher zu konzentrieren. Durch das Verstärken von der zum gewünschten Sprecher gehörenden Periodizität könnte dieses Problem abgeschwächt werden. Zumindest könnte typischerweise die Wahl von tiefen Periodizitäten die Schallreize männlicher Sprecher verstärken und die Wahl von höheren Periodizitäten entsprechend die von weiblichen.

Diese theoretischen Überlegungen könnten in der Praxis einfach geprüft werden. Die moderne Elektronik ermöglicht das einfache Aufbauen von entsprechenden Vorverarbeitungssystemen, zumindest im Labor. Sollten diese Erweiterungen Vorteile bringen, wären spezielle elektronische Chips denkbar, die diese Vorfilterung durchführen und nur geringen Platz beanspruchen.

5.2.2 Entwicklung neuer Implantate

Ein zur jetzigen Zeit futuristisch anmutendes Anwendungsgebiet für die in dieser Arbeit entwickelten Methoden wäre ein Implantat, welches nicht auf der Ebene der Cochlea arbeitet, sondern wesentlich später im Hörsystem - nämlich im Colliculus Inferior oder erst im primären auditorischen Kortex - aufsetzt. Abgesehen von der Schwierigkeit der Datenübertragung direkt von der Elektronik in das Gehirn wird durch die vorliegende Arbeit eine Möglichkeit gegeben, zumindest rudimentär die Funktionen des Hörsystems bis zum Colliculus Inferior zu ersetzen. Damit würden die Bereiche Cochlear Nucleus, Cochlea und Colliculus Inferior ersetzt werden können, falls diese durch Hirnschädigungen nicht mehr funktionieren. So wie die Cochlea-Implantate natürlich nur eine grobe Näherung der Funktion einer korrekt arbeitenden Cochlea darstellen, ist ebenso natürlich klar, daß die komplexen Funktionen des Colliculus Inferior nicht nur darin besteht, eine Periodizitäts/Trägerfrequenzkarte zu erstellen. Allerdings wurde in dieser Arbeit gezeigt, daß mit eben einer solchen Karte viele Funktionen eines Hörsystems realisiert werden können (Klassifikation, Lokalisation und Trennung von Schallquellen). Für Menschen, deren peripheres Hörsystem geschädigt ist und die daher überhaupt keine Schallwahrnehmung mehr haben, könnten diese Funktionen eine echte Verbesserung der Lebensqualität darstellen, auch wenn sie vorausichtlich nicht annähernd so gut arbeiten würden wie das biologische Original.

Für diese Anwendung könnte ein 'künstlicher Colliculus Inferior' eingesetzt werden, wie zum Beispiel das im anschließenden Kapitel vorgestellte Hardwaresystem. Voraussetzung für ein solches Implantat sind natürlich vertiefte Kenntnise in der Übertragung von elektronischen Daten in biologische Systeme.

Gegen ein IC-Implantat spricht allerdings, daß dieser chirurgisch nur schwer zu erreichen ist und die Gefahr besteht, vegetative Zentren unbeabsichtigt zu verletzen. Daher existieren Projekte [41], die sich mit Implantaten auf der Ebene des primären auditorischen Kortex beschäftigen. Eine mögliche Stimulation dieses Bereiches wäre mit dem in dieser Arbeit angegebenen Modell möglich, auch wenn für kortikale Bereiche noch wesentlich unklarer ist, welche exakte Funktion jeweils realisiert wird. Dennoch konnte anhand des verwendeten Modell gezeigt werden, wie einige Funktionen nachgebildet werden können, wie zum Beispiel das Erkennen von Vokalen oder Detektieren von Periodizitäten. Diese rudimentären Grundfähigkeiten wären für viele taube Menschen ein großer Schritt und würden eine dramatische Verbesserung für sie darstellen.

Kapitel 6

Hardware

Im folgenden werden Hardwareentwicklungen vorgestellt, die im Verlauf der Dissertation ergeben haben. Zum einem wurde für die Durchführung der Versuche ein umfangreicheres Setup benötigt, als bereits vorhanden war. Zum anderem konnte Aufgrund der Ergebnisse ein Demonstrator entwickelt werden, der die Arbeitsweise des Langnermodells dadurch erklärt, daß er die Karte, die sich im Colliculus Inferior ergibt, in Echtzeit berechnet und auf einem Monitor darstellt.

6.1 Der Versuchsaufbau

6.1.1 Einführung

Für einige Versuche, insbesondere die Lokalisationsversuche, war ein Versuchsaufbau erforderlich, der eine genaue Ausrichtung der Mikrofone zur Quelle (Lautsprecher) ermöglicht. Dazu wurde zum einen zur Stimulusgenerierung ein System entwickelt, welches akustische Stimuli auf unterschiedliche Lautsprecher verteilt. Zum anderen wurde ein vom PC steuerbarer 'Kopf' entwickelt, der durch eine PVC Stange, an der zwei Mikrofone angeschlossen waren, realisiert wurde. Dieser Aufbau wurde durch einen Artikel von Rucci [39] motiviert. Die beiden Teile werden im folgenden genauer vorgestellt.

6.1.2 Schallerzeugung

Der Versuchsaufbau besteht aus 12 Lautsprechern. Diese sind in einem Halbkreis angeordnet, so daß der Winkelabstand der Lautsprecher ca. 15 Grad ist. Der mechanische Aufbau der Lautsprecheraufhängung ermöglicht unterschiedliche Radien, von ca. 80 cm Radius bis ca. 160 cm Radius. An jedem Lautsprecher sind zwei Leuchtdioden angebracht, so daß auch optische Rückmeldungen erfolgen können.

Die Stimulussignale werden über die serielle Schnittstelle in ein hardware internes RAM (16 MByte) gespeichert und können dann per DA Wandler ausgegeben werden. Durch die Verwendung von vier unabhängigen DA-Wandlern können vier akkustische Stimuli gleichzeitig erzeugt werden. Zu jedem der 12 Lautsprecher kann jetzt ausgewählt werden, ob ein Kanal auf dem Lautsprecher ausgegeben werden soll oder nicht. In Abb. 6.1 ist der Aufbau schematisch dargestellt.

Es kann also ein Vokal von links und ein Rauschen von rechts ausgegeben werden, oder auch aus allen Lautsprechern ein Rauschen und aus manchen zusätzlich ein Vokal oder ein Sinus oder ähnliches.

Mit Hilfe der Leuchtdioden kann auf einfache Art und Weise signalisiert werden, welche Lautsprecher aktiv sind. Es kann jede Leuchtdiode beliebig und unabhängig von den anderen angesprochen werden. In Abbildung 6.2 ist der gesamte Aufbau zur Schallerzeugung nochmals dargestellt, inklusive der Position der LEDs an jedem Lautsprecher.

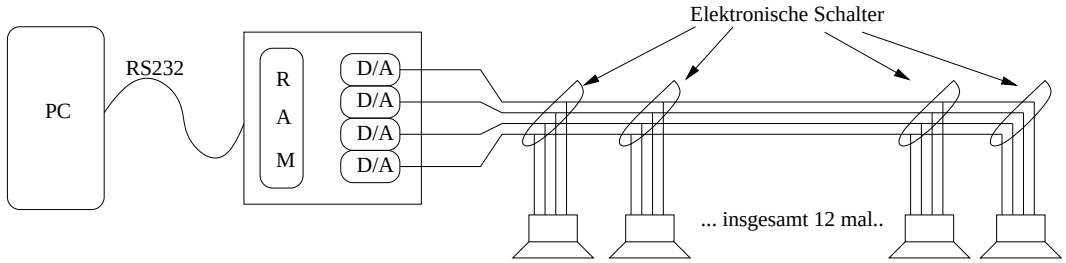

Abbildung 6.1: Schematischer Aufbau des Schallerzeugungssystems. Es stehen insgesamt 12 Lautsprecher, auf denen vier unterschiedliche Schallsignale ausgegeben werden können, zur Verfügung.

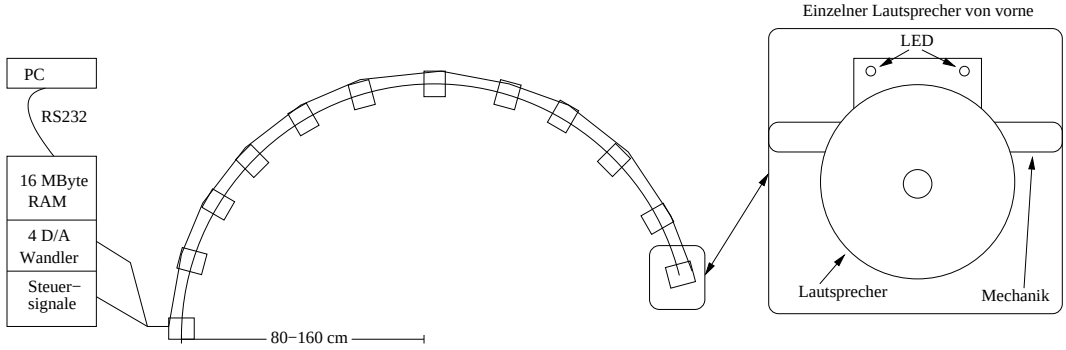

Abbildung 6.2: Komplette Darstellung des Versuchaufbaus zur Schallerzeugung.

6.1.3 Schallaufnahme

In die Mitte des Schallerzeugungssystems wird ein Drehteller mit zwei Mikrofonen und - wenn nötig - eine Videokamera gestellt. Dieser kann mit Hilfe eines Schrittmotors gedreht werden. Die Steuerung des Schrittmotors und die Auswertung des Videokamerabildes übernimmt eine Elektronik, die per serielle Schnittstelle mit dem PC verbunden ist. Die Mikrofone werden über einen Stereomikrofonvorverstärker an den Line-In-Eingang der Soundkarte des PCs angeschlossen. Für die Aufnahmen wurde eine handelsübliche Soundblaster 32 PCI verwendet, im Prinzip kann aber jede Standardsoundkarte für PCs verwendet werden.

Das Videosignal wird durch die Elektronik aufsummiert, so daß der Summenwert jeder Videozeile zur Verfügung steht. Wird die Kamera um 90 Grad gedreht montiert, so kann durch eine einfache Maximumsanalyse die Position im Videobild gefunden werden, wo eine LED leuchtet.

Während der Aufnahmen wurde der komplette Aufbau mit Schalldämmungsmatten umgeben, um die Einflüsse von Echos zu reduzieren.

Für Offlineversuche zur Schalllokalisation wird aus einem Lautsprecher ein Signal ausgegeben und die Drehteller mit den Mikrofonen nach jeder Aufnahme etwas gedreht. Dann wird wieder eine Aufnahme gemacht und dann wieder gedreht. Dadurch entsteht eine kopmlette Aufnahmesequenz des Stimulus zu jeder Raumrichtung. Diese können dann einzeln offline analysiert werden.

6.2 FPGA Implementierung des variierten Modells

Für die Realisierung des Langnermodells in Echtzeit wurde eine Hardware entwickelt, welche aus zwei Teilen besteht:

1. Signalaufbereitung:

Das zu analysierende Signal wird entweder per Mikrofoneingang oder per Line-In-Eingang an die Hardware angeschlossen. Die Signalaufbereitung filtert das Originalsignal mit mehreren analogen Bandpaßfiltern. Anschließend werden mittels einer Schwellwertoperation aus den analogen Daten digitale Daten erzeugt.

2. Signalanalyse und Darstellung der Ergebnisse:

   Die gefilterten und digitalisierten Daten werden mit einem FPGA (Field Programable Gate Array) weiterverarbeitet. Dieser realisiert eine vereinfachte Form des variierten Langnermodells. Das Ergebnis wird direkt an einem Monitor ausgegeben, ein PC ist nicht notwendig. Dargestellt wird auf dem Monitor die IC Karte in Echtzeit. Man kann dadurch die Reaktionen der Koinzidenzneuronen auf der Karte direkt mit dem Stimulus vergleichen oder auch den Einfluß verschiedener Parameter direkt auf der Karte sehen.

Diese beiden Teilsysteme werden im folgenden näher erläutert.

6.2.1 Signal Aufbereitung

Die Signal Aufbereitung führt eine Filterung mit anschließender Schwellwertuntersuchung durch. Die Abb. 6.3 stellt die einzelnen Komponenten der Aufbereitung schematisch dar. Die Schallsignale laufen von links nach rechts durch das Diagramm:

- Bei Verwendung eines Mikrofons wird das Signal erst vorverstärkt.
- Insgesamt 16 passive Filter 2. Ordnung bestehend aus je zwei Induktivitäten und zwei Kondensatoren filtern das Signal in 16 Frequenzbänder.
- Für jedes Frequenzband wird ein eigener Schwellwert eingestellt, der angibt, wie hoch die Amplitude sein muß, damit eine logische Eins (3,3 Volt) am Ausgang erzeugt wird.

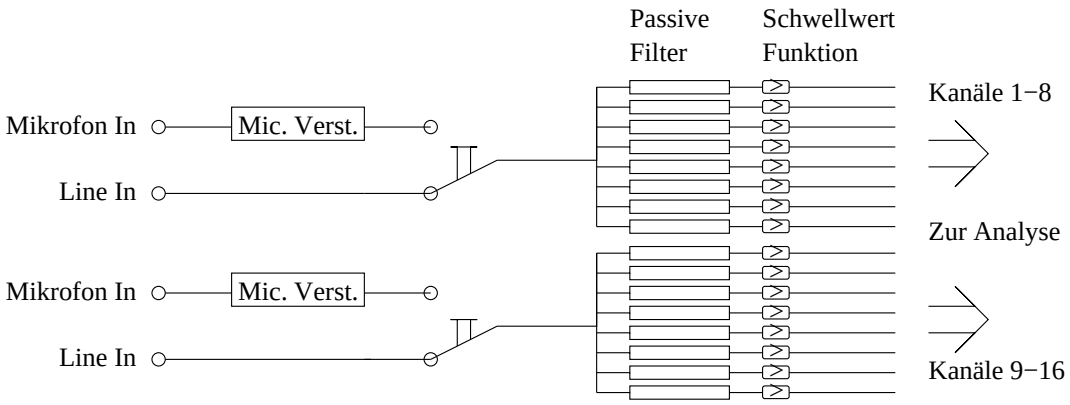

Abbildung 6.3: Schematische Darstellung der Signalaufbereitung. Aus den Schallsignalen werden gefilterte und mit einer Schwellwertfunktion digitalisierte Impulse.

Der Zustand des Ausgangs jedes Filter wird auf einer LED dargestellt, so daß man die Schwellwerte während des Betriebs auf einfache Weise einstellen kann.

Für spätere Erweiterungen besteht die Möglichkeit, zwei unabhängige Audio Signale einzuspeisen, wenn Stereosignale untersucht werden sollen. Um eine erhöhte Flexibilität zu erreichen, sind die Filter nicht auf der Vorverarbeitungsplatine untergebracht, sondern als zusätzliche Aufsteckplatine konzipiert. Dadurch können unterschiedliche Filtertopologien und Frequenzen auf verschiedenen Platinen realisiert und getestet werden. Im biologischen Sinne kann man das ungefähr mit einer auswechselbaren Cochlea vergleichen.

Die Signalaufbereitungsplatinen selbst sind auch wiederum modular aufgebaut. Jede Platine erzeugt acht Ausgangskanäle. Das Audiosignal wird an der Masterplatine angeschlossen. Diese erzeugt die Kanäle 1-8. Sind weitere Kanäle erwünscht, so können mehrere Slaveplatinen angeschlossen werden, die von der Masterplatine mit Strom und Signalen versorgt werden und jeweils wieder 8 Kanäle zur Verfügung stellen. Das verwendete System besteht aus einer Master- und einer Slaveplatine, es stehen mehrere Filterplatinen mit Frequenzen zwischen 500 Hz und 16 kHz zur Verfügung.

6.2.2 Signalanalyse und Darstellung der Ergebnisse

Die von der Signalaufbereitung vorgefilterten und per Schwellwert digitalisierten Signale werden von einem FPGA der Firma Xilinx weiterverarbeitet. Es handelt sich dabei um den Spartan II FPGA XC2S200 [20], dessen technische Daten kurz vorgestellt werden:

- 5292 Logic Cells, entspricht 200 000 Gatter
- 56 kBit embedded RAM
- 2,5 Volt Kernspannung, 3,3 Volt I/O Spannung
- bis zu 284 I/O Pins
- $>$ 100 MHz Systemtakt

Das Prinzip eines FPGAs erlaubt es, viele Aufgaben parallel durchführen zu lassen. Daher eignet sich diese Bausteinsorte deutlich besser für die Implementierung von Netzwerken als einzelne Mikroprozessoren oder DSPs, die zwar schnell rechnen, aber nur ein Programm gleichzeitig durchführen können. Der FPGA ist in der Lage, parallel viele Neurone zu simulieren. Allerdings sind die Einzelbausteine, mit denen die Rechnung realisiert wird, sehr primitiv (Gatter, Flip/Flops,...). Daher wurden die zu simulierenden Neurone stark vereinfacht, um die notwendigen Resourcen innerhalb des FPGAs niedrig zu halten.

Ausgehend vom variierten Langnermodell wurden folgende Schritte unternommen (ähnlich zum Algorithmus im Abschnitt 5.1.6 auf Seite 111):

- Reduktion des Neuronenmodells auf einen Akkumulator mit 8 Bit, zu dem der Input addiert wird (Werte zwischen -31 und +31). Dazu kommt ein Schwellwert, der mit dem Akkumulator verglichen wird. Ist der Wert des Akkumulators größer, so wird am Ausgang des Neurons eine logische Eins erzeugt, entsprechend einem Spike. Zusätzlich wird noch der Akkumulator um einen einstellbaren Wert erniedrigt.
- Reduktion des Langnerschaltkreises wie folgt:
  1. Wegfallen des Oszillatorteils
  2. Ersetzen der Integrator/Flip/Flop/Koinzidenzneurone durch einen Periodizitäts Detektor Zähler, der den zeitlichen Abstand der Triggerspikes mißt und daraus die Periodizität berechnet
  3. Realisierung des Triggerschaltkreises mit zwei Neuronen entsprechend dem oben beschrieben Neuronmodells.
- Simulation des so veränderten Schaltkreises mit Matlab.

Die Karte ergibt sich dann dadurch, daß immer, wenn ein Triggerschaltkreis einen Spike erzeugt, der aktuelle Zählerwert des zugehörigen Periodizitätsdetektorzählers ausgelesen, und als Periodizitätswert interpretiert in eine Matrix eingetragen wird. Bei jedem Spike wird der Wert an dieser Stelle um eins erhöht. Um ein Überlaufen der Matrix zu verhindern und nur die aktuellen Ergebnisse darzustellen, werden alle Matrixwerte, die nicht bereits Null sind, alle 16 ms um eins erniedrigt.

Dadurch ergibt sich eine Matrix, die auf der X-Achse die Periodizität aufgetragen hat (berechnet durch den Periodizitätsdetektorzähler) und auf der Y-Achse die Trägerfrequenz (berechnet durch die passiven Filter der Signalaufbereitung). Diese Matrix hat eine Größe von 16 Trägerfrequenzen (abhängig von den verwendeten Filterplatinen) auf 128 Periodizitätsfrequenzen (von ca. 20 Hz bis ca. 500 Hz).

Der FPGA berechnet alle notwendigen Rechenschritte innerhalb von 25 μs, dies ergibt sich durch eine eingestellte Abtastrate der I/O Pins von 40 kHz. Die Matrix wird entsprechend mit 40 kHz aktualisiert, so daß 32*128*40000 Bytes/s als Ergebnis Daten produziert werden. Dies entspricht ca. 160 MByte/s. Da die Hauptfunktion der Implementation die Demonstration des Langnermodells ist, wurde auf eine Datenübertragung zum PC verzichtet und statt dessen das Ergebnis direkt mit Hilfe eines kleinen VGA-Schaltkreis, der zusätzlich in dem FPGA realisiert wurde, als VGA Monitorsignal bereitgestellt. Wird dieses Signal auf einem Monitor ausgegeben, so wird neben einigen Zusatzinformationen die Karte als farbig kodiertes Feld ausgegeben.

Wenn mit dem Standardmodell eine vergleichbare Karte in Echtzeit erzeugt werden sollte, so würde dies folgenden Rechenaufwand bedeuten:

- 16*128 Langner Schaltkreise ergibt 2048 Schaltkreise
- Jeder der 2048 Schaltkreis wird mit 9 Neuronen realisiert, das ergibt 18432 Neuronen
- Bei einer Abtastrate von 40 kHz ergibt dies ca. 700 Millionen 'Neuronenschritte'/s. Da ein Simulationsschritt eines Neurons vieler Prozessorbefehle bedarf, müste der PC mit mindestens 7 GHz getaktet werden, wenn nur 10 Befehle pro Neuronschritt benötigt werden würden. Realistischer wäre eine Größenordnung von ca. 50 Befehlen, die aber auch nicht alle in 1 Prozessor Takt ausgeführt werden können.

Zugegebenermaßen ist das Modell, welches der FPGA simuliert, bei weiten nicht so detailreich wie das PC Modell, aber das Ergebnis ist für Demonstrationszwecke völlig ausreichend, man erkennt die Periodizität des Stimulus und zum Beispiel auch die Formantfrequenzen, wenn Vokale als Stimulus verwendet werden.

Interessant ist an dieser Stelle die Beobachtung, daß - obwohl das FPGA Modell so konstruiert ist, daß es keine intrinsischen Kammfilterantworten erzeugt - dennoch auf dem Monitor immer wieder welche zu erkennen sind. Das ist mit dem nicht zu 100 % funktionierendem Triggerschaltkreis zu erklären: Ein Aussetzer erzeugt entsprechend niedrigere Periodizitäten (vgl. Abschnitt 4.1.2 der Diskussion auf Seite 87).

Kapitel 7

Zusammenfassung

Diese Dissertation beschäftigt sich mit dem auditorischen System des Menschen. Das wesentliche Ergebnis ist ein Modell des auditorischen Systems, welches komplett im Computer implementiert und simuliert wurde. Aus diesem Modell lassen sich verschiedene neue Verfahren sowohl für technische Problemstellungen im Bereich der Musik- und Sprachverarbeitung als auch für medizinische Geräte ableiten. Im folgenden werden zunächst kurz die Komponenten des Modells und die zugehörigen Ergebnisse zusammengefaßt und anschließend die möglichen Anwendungen vorgestellt.

Das Modell ist wie das biologische Vorbild in zwei Komponenten aufgeteilt. Der erste Teil stellt die Vorverarbeitung dar, und modelliert die peripheren Bereiche und Teile des Mittelhirns bis zum Colliculus Inferior. In diesem Kern entsteht auf bisher noch unbekannte Weise eine Periodizitäts/Frequenz-Karte, so daß sich die Frequenz des Stimulus systematisch auf einer Achse im Colliculus Inferior abbildet, während die Periodizität des Stimulus orthogonal zur Frequenz variiert. Für neurophysiologische Untersuchungen an Tieren wird normalerweise ein amplitudenmodulierter Stimulus verwendet, bei dem die Modulationsfrequenz der Periodizität und die Trägerfrequenz der Frequenz entspricht. Wird die Modulationsfrequenz verändert, so wandert die Aktivität im Colliculus Inferior entlang der ersten Achse, wird dagegen die Trägerfrequenz verändert, so bewegt sich die Aktivität entlang der zweiten Achse. Während die Analyse der Trägerfrequenz bereits innerhalb der Cochlea stattfindet, indem unterschiedliche Frequenzen unterschiedliche Stellen auf der Basilarmembran aktivieren, ist der Mechanismus zur Detektion der Modulationsfrequenz bisher unbekannt. Ein Ergebnis dieser Arbeit stellt ein biologisch plausibles Modell zur Periodizitätsanalyse dar, welches auf dem Langnermodell [16] basiert.

Zunächst wurde dieses ursprüngliche Langnermodell am Computer implementiert und mit verschiedenen Stimuli getestet (siehe Abschnitt 3.1). Dadurch, daß die Neuronen im Computer simuliert werden, kann ohne Probleme jedes einzelne Neuron, aber auch das gesamte Modell auf einmal analysiert werden. Die Simulationen ergaben Ergebnisse, die nicht den biologischen Vorgaben entsprachen: das Antwortverhalten einiger für das Modell notwendiger Neuronentypen ist in der Biologie bisher nicht gefunden worden, und das Gesamtmodell erzeugt Ergebnisse, die zum einen nicht den biologischen Systemen entsprechen und zum anderen für die weitere Verarbeitung nicht sinnvoll erscheinen. Diese Punkte sind in der Diskussion, Abschnitt 4.1, genauer erklärt. Um diese Nachteile zu beheben, wurden unterschiedliche Variationen des Langnermodells neu implementiert und anhand der Simulationsergebnisse ein variiertes Langnermodell (siehe Abschnitt 2.3.5) ausgewählt. Dieses Modell erzeugt eine für kortikale Areale besser geeignete Ausgabe und entspricht der Biologie besser, da alle verwendeten Neuronen in den entsprechenden biologischen Kernen in ausreichender Zahl vorhanden sind (siehe Abschnitt 4.1.6).

Anhand des variierten Langnermodells konnten auch Kartierungen getestet werden, die im Colliculus Inferior gefunden wurden. Zum Beispiel ergibt die Verwendung des Langnermodells automatisch eine Latenzkartierung (siehe Abschnitt 3.5.2), welche auch bei biologischen Messungen gefunden wurde. Ein weiterer direkter Vergleich mit biologischen Daten wurde dadurch möglich, daß von einem Projektpartner (Michael Ochse) einige Ableitungsdaten direkt zur Verfügung gestellt wurden. Durch neu entwickelte Darstellungsmethoden konnten diese Messungen so visualisert

werden, daß die Phasensynchronizität der Messungen gut mit den Simulationsergebnissen verglichen werden konnte (siehe Abschnitt 3.4.4). In dieser Darstellung wurde auch deutlich, daß das variierte Modell besser zur Biologie paßt, als das ursprüngliche Modell.

Die zweite Komponente des Gesamtmodells sind die kortikalen Areale, welche zunächst zur Klassifikation und danach zur Separation eingesetzt wurden. Für diesen Teil des Modells existierten bisher keine konkreten Modelle wie das Langnermodell, und so wurden Assoziativspeicher aus dem visuellen Bereich für die kortikalen Areale des Modells modifiziert und verwendet (siehe Abschnitt 3.6). Mit diesen werden zwei Assoziativspeicherebenen realisiert. Die erste modelliert den primären auditorischen Kortex und die zweite ein höheres Areal. Durch diese Einteilung konnten einzelne Vokale korrekt klassifiziert und auch aus Mischungen von Vokalen separiert werden.

Der Vergleich der peripheren Areale und die guten Klassifikationsleistungen der kortikalen Areale zeigten sowohl eine hohe biologische Plausibilität des Gesamtmodells als auch eine gute Eignung für praktische Aufgaben. Mit Hilfe der ersten Hälfte des Modells kann die Periodizität eines Schallsignals extrahiert werden. Diese Information ist für viele Bereiche der Schallverarbeitung, wie zum Beispiel für die Spracherkennung, Sprecheridentifikation, Quellentrennung und auch für die Analyse von Musik wichtig (siehe Abschnitt 5.1). Zusätzlich kann durch die Analyse der zeitlichen Verzögerung zwischen zwei Mikrophonen die Ausgabe der ersten Hälfte zur Schalllokalisation verwendet werden, da die Verwendung von spikenden Neuronen die Beibehaltung der genauen zeitlichen Informationen in beiden Kanälen ermöglicht (siehe Abschnitt 3.7).

Wird dagegen das gesamte Modell verwendet, so können auch einfache Sprachsignale klassifiziert werden, auch wenn diese von anderen Geräuschen gestört werden. Dies wurde unter Verwendung von Vokalen als Sprachsignale gezeigt. Das Modell konnte sich auf einen Sprecher 'konzentrieren', und dessen Vokale korrekt klassifizieren, auch wenn zusätzliche Vokale mit einer anderen Tonhöhe im Stimulus enthalten waren (siehe Abschnitt 3.6.4).

Dadurch, daß das komplette Modell einen Großteil des menschlichen Hörsystems repräsentiert - von der Schallaufnahme bis zu den kortikalen Arealen - kann mit seiner Hilfe auf vielen Ebenen der Schallverarbeitung eine Prognose über die Funktion von neuronalen Strukturen gemacht werden. Mit diesem Wissen kann die Softwareentwicklung für neue Implantate für Menschen, deren Hörsystem geschädigt ist, unterstützt werden. Die Realisierung von Teilen des Modells in Echtzeit und ohne PC wurde in Abschnitt 6.2 gezeigt. Auch wurden einige weitere Hardwarekomponenten entwickelt, die für die durchgeführten Versuche notwendig waren (siehe Abschnitt 6.1.1).

Es zeigte sich, daß die Verwendung von Spikes zur Informationsrepräsentation wichtig ist, da in dem Modell nicht nur die Spikerate, sondern auch die Phasenlage der Spikes wichtig ist (siehe Abschnitt 3.8.2). Mit Hilfe eines psychophysikalischen Experimentes konnte gezeigt werden, daß vermutlich die Phasenlage der Spikes zueinander einen Einfluß auf die Lautstärkenwahrnehmung des Menschen hat. Der Effekt, daß gleichzeitig eintreffende Spikes eine höhere Aktivität erzeugen als gleichviele unsynchronisierte Spikes, kann nur durch die Verwendung von spikenden Neuronen, wie sie auch für das Modell benutzt werden, ausgenutzt werden. Da ähnliche Synchronitäten in Sprachsignalen vorhanden sind, kann mit ihrer Hilfe das Cocktailpartyproblem einfacher gelöst werden, da synchrone Frequenzkomponenten mit hoher Wahrscheinlichkeit zum gleichen Sprecher gehören.

Anhang

7.1 Verwendete Parameter für die Simulationen

7.1.1 Ursprüngliches Langnermodell

Die Parameter der Neuronen sind wie folgt:

	T_1	T_2	O	F	I	K
τ_x	0.5	0.5	0.25	1	20	0.25
τ_{ex}	0.25	0.25	0.25	0.25	1	0.25
τ_{in}	0.25	0.25	0.25	0.25	1	0.25
$g_{ex}max$	100	100	100	200	1000	100
$g_{in}max$	100	400	100	100	1000	100
θ	10	10	10	10	10	9
x_{ahp}	6	20	8	3	20	4

Tabelle 7.1: Die Parameter der Neuronen des ursprünglichen Langnermodells

Dabei bezeichnet T_1 die inhibitorischen Neuronen des Triggerschaltkreises, T_2 die exitatorischen Neuronen des Triggerschaltkreises, O die Oszillatorneuronen, F die Flip/Flop Neuronen, I die Integratorneuronen und K die Koinzidenzneuronen.

Die Parameter der Verbindungen sind in Tab. 7.2 angegeben. Jede Verbindung hat dabei eine Stärke und eine Verzögerung (in ms). E bezeichnet den externen graduellen Input, O_x ist das x-te Neuron des Oszillators, F_1 ist das Flip Neuron, F_2 das Flop Neuron. Die Verbindugn $E \rightarrow I$ ist die Integratorverbindung, die entsprechend der zu detektierenden Periodizität nach folgender Formel festgelegt ist:

$$[E \rightarrow I](x) = (exp((x/B) * 2) - 1)/(exp(2) - 1) * 25 + 2$$

Dabei bezeichnet x die x-Position des Neurons in der Karte und B die Breite der Karte. Der exponentielle Verlauf wurde dem linearen Verlauf vorgezogen, um die niedrigeren Periodizitäten zu begünstigen, da das Tuning der Neuronen für die hohen Periodizitäten sowieso breiter ist.

	Delay	Stärke
$E \to T_1$	0.1	8
$E \to T_2$	0.1	8.1
$T_1 \to T_2$	0.25	-50
$T_2 \to O_1$	0.025	37
$O_1 \to O_2$	0.025	10.5
$O_2 \to O_3$	0.025	10.5
$O_3 \to O_1$	0.025	10.5
$E \to I$	0.1	stetig von 2 bis 25
$T_2 \to F_1$	0.1	500
$F_1 \to F_1$	0.1	2
$F_1 \to F_2$	0.1	-9.5
$I \to F_2$	0.1	50
$F_2 \to F_2$	0.1	2
$F_2 \to F_1$	0.1	-11
$F_2 \to I$	0.1	500
$I \to K$	0.9	27
$O_2 \to K$	0.1	27

Tabelle 7.2: Parameter der Verbindungen des ursprünglichen Langnermodells, beispielsweise bezeichnet $I \to K$ eine Verbindung vom Integratorneuron zum Koinzidenzneuron. Negative Werte entsprechen Inhibition, positive Exitation.

7.1.2 Variiertes Langnermodell

Die Parameter der Neuronen des variierten Langnermodells sind in Tabelle 7.3 aufgelistet. T_1 und T_2 sind die Neuronen des Triggerschaltkreises, O die drei Oszillatorneuronen, I_S das IntegStop Neuron K_I das inhibitorische Interneuron (Koinzidenz$I$nhibitions-Neuron), I das Integratorneuron und K das Koinzidenzneuron.

	T_1	T_2	O	I_S	K_I	I	K
τ_x	0.5	0.5	0.25	1	0.25	20	0.25
τ_{ex}	0.25	0.25	0.25	0.25	0.25	1	0.25
τ_{in}	0.25	0.25	0.25	0.25	0.25	1	0.25
$g_{ex}max$	100	100	100	200	400	1000	100
$g_{in}max$	100	400	100	100	500	1000	100
θ	10	10	10	10	10	10	9
x_{ahp}	6	30	8	3	10	1	4

Tabelle 7.3: Die Parameter der Neuronen des variierten Langnermodells.

Die Parameter der Verbindungen sind in Tab. 7.4 aufgelistet. Die Verbindungsstärke des externen Inputs auf die Integratorneuronen ($E \to I$) entsprechen der Formel des ursprünglichen Langnermodells.

	Delay	Stärke
$E \to T_1$	0.1	8
$T_1 \to T_2$	0.1	-50
$E \to T_2$	0.1	8.1
$T_2 \to O_1$	0.25	50
$O_1 \to O_2$	0.025	8.5
$O_2 \to O_3$	0.025	8.5
$O_3 \to O_1$	0.025	8.5
$T_2 \to I_S$	0.2	190
$I_S \to I$	0.025	300
$E \to I$	0.1	2-25
$I \to K_I$	0.4	40
$I \to K$	0.45	30
$K_I \to K$	0.05	-31
$O_2 \to K$	0.025	8

Tabelle 7.4: Parameter der Verbindungen des variierten Langnermodells, beispielsweise bezeichnet $I \to K$ eine Verbindung vom Integratorneuron zum Koinzidenzneuron.

7.1.3 Assoziativspeicher

Die Parameter des Assoziativspeichers finden sich in der Dissertation von Andreas Knoblauch [26] wieder. Dort sind auch die notwendigen zusätzlichen Grundlagen für Assoziativspeicher vorhanden. In Tab. 7.5 sind nur die Verbindungsstärken zwischen den einzelnen Arealen aufgelistet. Die Areale X_e sind die exitatorischen Areale, X_s die inhibitorischen Areale und X_t zusätzliche inhibitorische Areale, auf die in dieser Arbeit nicht eingegangen wird. Diese sind in der Dissertation von Andreas Knoblauch genauer erklärt.

	K_e	K_s	C_e	C_s	C_t	D_e	D_s	D_t
K_e	0	0	0	0	0	0	0	0
K_s	0	0	0	0	0	0	0	0
C_e	1.5	0	2	3	2	1	0	0
C_s	1.5	0	2	3	2	2	0	0
C_t	0	0	2	3	2	0	0	0
D_e	0	0	2	0	0	3	3.5	2
D_s	0	0	1.5	0	0	3	3.5	2
D_t	0	0	0	0	0	3	3.5	2

Tabelle 7.5: Verbindungsstärken der Areale des Assoziativspeichers. Areal X_e ist das exitatorische, X_s das inhibitorische und X_t ein zweites inhibitorisches Areal. Die Verbindungen sind von oben nach unten eingetragen, beispielsweise sind die Verbindungen, die von K_e wegführen, in der ersten Spalte unter K_e.

Literaturverzeichnis

[1] A.L. Babalian, D.K. Ryugo, and E.M. Rouiller. Discharge properties of identified cochlear nucleus neurons and auditory nerve fibers in repsonse to repetitive electrical stimulation of the auditory nerve. *Exp. Brain Res.*, 153:452–460, 2003.

[2] L.R. Bernstein and C. Trahiotis. Enhancing sensitivity to interaural delays at high frequencies by using transposed stimuli. *Journal of the Acoustical Society of America*, 112:1026–1036, 2002.

[3] Samuil M. Blinkov and Ilja I. Glezer. *Das Zentralnervensytem in Zahlen und Tabellen.* VEB Gustav Fischer Verlag Jena, 1968.

[4] Borst M., Knoblauch A., Palm G. Modelling the auditory system: Preprocessing and associative memories using spiking neurons. *Neurocomputing*, accepted.

[5] Borst M., Knoblauch A., Palm G. Using spiking associative memories for auditory pattern separation. *Neurocomputing*, submitted.

[6] Borst M., Langner G., Palm G. A biologically motivated neural network for phase extraction from complex sounds. *Biological Cybernetics*, 90(2):98–104, 2004.

[7] Borst M., Palm G. Periodicity pitch detection and pattern separation using biologically motivated neural networks. In *Beiträge zur 6. Tübinger Wahrnehmungskonferenz*, page 52. Knirsch Verlag Kirchentellinsfurt, 2003.

[8] Borst M., Palm G. Periodicity pitch detection and sound separation with spiking neural networks. In *The Neurosciences from Basic Research to Therapy*, page 1046. Thieme, 2003.

[9] E. Budinger, P. Heil, and H. Scheich. Functional organization of auditory cortex in the mongolian gerbil (meriones unguiculatus). iii anatomical subdivisions and corticocortical connections. *European Journal of Neuroscience*, 12:2425–2451, 2000.

[10] E. Budinger, P. Heil, and H. Scheich. Functional organization of auditory cortex in the mongolian gerbil (meriones unguiculatus). iv. connections with anatomically characterized subcortical structures. *European Journal of Neuroscience*, 12:2452–2474, 2000.

[11] N.B. Cant. The cochlear nucleus: Neuronal types and their synaptic organiziation. *The Mammalian auditory Pathway (Springer-Verlag Berlin)*, pages 66–116, 1992.

[12] G. Ehret. The auditory cortex. *Journal of Computational Physiology*, 181:547–557, 1997.

[13] G. Ehret. *The auditory midbrain, a 'shunting yard' of acoustical information processing.* Oxford University Press, New York, 1997.

[14] G. Ehret, S.R. Hage, Egorova M., and B.A. Müller. *Auditory maps of the midbrain: the inferior colliculus.* Berlin, Springer Verlag, 2004.

[15] G. Ehret and R. Romand. *The Central Auditory System.* New York, Oxford University Press, 1997.

[16] Langner G. Temporal processing of pitch in the auditory system. *Journal of New Music Research*, 26:116–132, 1997.

[17] S.W. Golomb. Run-length encodings. *IEEE Transactions on Information Theory*, 12:399–401, 1996.

[18] H.L. Hawkins, T.A. McMullen, A.N. Popper, and R.R. Fay. *Auditory Computation.* Springer Verlag, Berlin, 1996.

[19] B. Herrnberg, S. Kempf, and G. Ehret. Basic maps in the auditory midbrain. *Biological Cybernetics*, 87:231–240, 2002.

[20] Xilinx Inc. *Spartan-II 2.5 Volt Family Field Programmable Gate Array*, volume DS001. www.xilinx.com, 2001.

[21] C. Kaernbach and C. Bering. Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *Journal of the Acoustical Society of America*, 110:1039–1048, 2001.

[22] C. Kaernbach and L. Demany. Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *Journal of the Acoustical Society of America*, 104:2298–2306, 1998.

[23] A. Knoblauch. Synchronization and pattern separation in spiking associative memory and visual cortical areas. *PhD thesis, Department of Neural Information Processing, University of Ulm, Germany*, 2003.

[24] A. Knoblauch. *Felix++ Online Dokumentation.* www.informatik.uni-ulm.de/ni/mitarbeiter/AKnoblauch/kdoc_Felix2/index.html, 2004.

[25] A. Knoblauch and G. Palm. Pattern separation and synchronization in spiking associative memories and visual areas. *Neural Networks*, 14:763–780, 2001.

[26] Andreas Knoblauch. *Synchronization and pattern separation in spiking associative memories and visual cortical areas.* Abteilung Neuroinformatik, Universität Ulm, 2003.

[27] B. Kollmeier and R. Koch. Speech enhancement based on physiological and psychoaccoustical models of modulation perception and binaural interaction. *Journal of the Acoustical Society of America*, 95:1593–1602, 1994.

[28] G Langner, M Albert, and Briede T. Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (chinchilla laniger). *Hearing Research*, 168:110–130, 2002.

[29] G. Langner, C. Simonis, S. Braun, and M. Ochse. Evidence for a pitch helix in the ventral nucleus of the lateral lemniscus in the gerbil. *Assoc.Res.Otolaryngol.*, 26:Abs. 683, 2003.

[30] Langner G., Sams M., Heil P., Schulze H. Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *J. Comp. Physiol.*, 181:665–676, 1997.

[31] Langner G., Schreiner C.E. Periodicity coding in the inferior colliculus of the cat I. Neuronal mechanisms. *Journal of Neurophysiology*, 60:1799–1822, 1988.

[32] Langner G, Simonis C, Braun S. Nachweis einer periodotopen Organisation im Lemniscus lateralis mit der 2-Deoxyglukose -Methode. *Fortschritte der Akustik, DAGA 2002, 298*, 2002.

[33] R.J. MacGregor. *Neural and Brain Modeling.* Academic Press, San Diego, 1987.

[34] B. Maess, S. Koelsch, T.C. Gunter, and A.D. Friederici. Musical syntax is processed in Broca's area: a MEG study. *Nature Neuroscience*, 4:540–544, 2001.

[35] G. Palm. On associative memories. *Biological Cybernetics*, 36:19–31, 1980.

[36] R.D. Patterson, J. Holdsworth, and M. H. Allerhand. *Auditory models as preprocessors for speech recognition*. Mouton de Gruyter, Berlin, 1995.

[37] D. Pressnitzer, I.M. Winter, and R.D. Patterson. The response of single units in the ventral cochlear nucleus of the guinea pig to damped and ramped sinusoids. *Hearing Research*, 149:155–166, 2000.

[38] W.S. Rhode and P.H. Smith. Physiological studies on neurons in the dorsal cochlear nucleus of cat. *Journal of Neurophysiology*, 56:287–307, 1986.

[39] M Rucci, G.M. Edelmann, and J. Wray. Adaption of orienting behavior: From the barn owl to a robotic system. *IEEE transactions on robotics and automation*, 15(1):96–110, 1999.

[40] M. Rucci and J. Wray. Binaural croos-correlation and auditory localization in the barn owl: a theoretical study. *Neural Networks*, 12:31–42, 1999.

[41] H. Scheich and A. Breindl. An animal model of auditory cortex prostheses. *Audiol. Neurootol.*, 7:191–193, 2002.

[42] C.E. Schreiner and G. Langner. Laminar fine structure of frequency organization in auditory midbrain. *Nature*, 388:383–386, 1997.

[43] H. Schulze, A. Hess, F.W. Ohl, and H. Scheich. Superposition of horseshoe-like periodicity and linear tonotopic maps in the auditory cortex of the mongolian gerbil. *European Journal of Neuroscience*, 15:1077–1084, 2002.

[44] Slanley M. Auditory toolbox. *Apple Computer, Inc. Advanced Technology Group*, 1993-1994.

[45] W. Strange. Evolving theories of vowel perception. *Journal of the Acoustical Society of America*, 85:2081–2087, 1989.

[46] C. Tsuchitani and D.H. Johnson. *Binaural Cues and Signal Processing in the Superior Olivary Complex*. Raven Press, Ltd., New York, 1991.

[47] J. Verschuure and A.A. van Meeteren. The effect of intensity on pitch. *Acoustica*, 32:33–44, 1975.

[48] C. von der Malsburg. Am I thinking assemblies? In G. Palm and A. Aertsen, editors, *Brain Theory*, pages 161–176. Springer-Verlag, Berlin/Heidelberg, 1986.

[49] W. Weir, J. Jesteadt and D.M. Green. Frequency discriminiation as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 61:178–184, 1977.

Lebenslauf und wissenschaftlicher Werdegang

Persönliche Angaben

Name	Marcus Borst
Geburtsdatum	14.10.1973
Geburtsort	Wilhelmshaven
Staatsangehörigkeit	Deutsch
Familienstand	verheiratet mit Ilona Borst, geb. Bihlmaier
Eltern	Dr. med. Roland Borst
	Gymnasialrätin Gertrude Borst, geb. Herrmann
Geschwister	Dr. Eva-Maria Borst
	Dipl. Ing. Alexander Borst
	Patricia Sophia Borst
	Daphne Eirene Borst
	Isis Agape Borst
	Roland Siddharta Borst

Ausbildung

1980-1984	Grundschule in Oberkochen
1984-1993	Gymnasium in Oberkochen
1993-1998	Studium der Informatik mit Nebenfach Physik an der Universität Ulm
seit 1999	Promotion in der Abteilung Neuroinformatik, Universität Ulm

Prüfungen

12.5.1993	Abitur
27.9.1995	Vordiplom in Informatik mit Nebenfach Physik
15.10.1998	Diplom in Informatik mit Nebefach Physik

Berufstätigkeit

1995-1997	GBR: Verkauf von Hard- und Software
seit 1997	OHG: Entwicklung von Hard- und Software
1995-2002	Hilfskrafttätigkeiten an der Universität Ulm
seit 1999	Wissenschaftlicher Angestellter der Abteilung Neuroinformatik der Universität Ulm