Günter Kochendörfer

# Kortikale Linguistik

Teil 3: Phonetik/Phonologie

http://www.cortical-linguistics.de
15. 1. 2007

# Gesamtinhaltsübersicht

# Teil 3

# Phonetik/Phonologie

*Wenn man die traditionell als „Phonetik“ und „Phonologie“ bezeichneten linguistischen Gegenstände von den Anforderungen der kortikalen Verarbeitung her beleuchtet, ergeben sich einige grundsätzlich neue Perspektiven. Die Unterscheidung der beiden Gebiete muss neu gefasst werden. Es ergeben sich neue Argumente für das Problem des Status phonetischer bzw. phonologischer Merkmale. Schließlich wird es möglich, Aussagen zu möglichen Verarbeitungsprozessen zu machen in Bereichen, die bisher in der Forschung im wesentlichen ausgeklammert worden sind.*

# Inhalt

# 3.1 Phonetik und Phonologie

## 3.1.1 Gängige Definitionen

Man kann in Einführungstexten die Angabe finden, dass sich die Phonetik mit den physikalischen Eigenschaften von Sprachlauten beschäftigt und die Phonologie (oder Phonemik) den Bezug zur systematischen Geltung bzw. kommunikativen Relevanz lautlicher Einheiten herstellt. Bei Hall (2000: 37) wird die Systematik in angenähert „distributionalistischer“ Auslegung als Unterscheidungskriterium verwendet:

> „... Andererseits kann man die Systematik der Laute einer Sprache untersuchen, d. h. das Vorkommen bzw. Nichtvorkommen von Lauten in bestimmten Segmentfolgen. Diese Fragestellung ist Gegenstand der Phonologie.“

Mit solchen Charakterisierungen ist implizit die Vorstellung verbunden, dass die Phonologie gegenüber der Phonetik einen zusätzlichen, von der Phonetik nicht berücksichtigten, aber für die theoretische Aufarbeitung des linguistischen Gegenstands wesentlichen Aspekt hinzufügt. Der historische Hintergrund dafür ist natürlich, dass die Phonologie (unter dieser Bezeichnung) ursprünglich als Zweig der strukturalistischen Sprachtheorie durch die Linguisten der Prager Schule (in den 20er Jahren des 20. Jahrhunderts) formiert worden ist, also strukturalistische Grundbegriffe auszuprägen waren, während die Phonetik eine vorstrukturalistische Geschichte hat und damit aus strukturalistischer Sicht mit einer einfacheren, eher beschreibenden als erklärenden Herangehensweise identifiziert worden ist.

Es ist allerdings auch vielfach festgestellt worden, dass es ein problematisches Spannungsfeld zwischen den beiden Disziplinen und entsprechende Abgrenzungsprobleme gibt. Die *phonetischen* Verschriftlichungskonventionen der IPA enthalten, wieder vom Standpunkt der Phonologie aus gesehen, ein phonologisches Gerüst.

> „Although phonetics as a science is interested in all aspects of speech, the focus of phonetic notation is on the linguistically relevant aspects. For instance, the IPA provides symbols to transcribe the distinct phonetic events corresponding to the English spelling *refuse* (['rɛfjus] meaning 'rubbish' and [rɪ'fjuz] meaning 'to decline'), but the IPA does not provide symbols to indicate information such as 'spoken rapidly by a deep, hoarse, male voice'. Whilst in practice the distinction between what is linguistically relevant and what is not may not always be clear-cut, the principle of representing only what is linguistically relevant has guided the provision of symbols in the IPA." (International Phonetic Association, 1999: 4)

Bei Ladefoged & Maddieson (1996: 2) liest man, den IPA-Formulierungen sinngemäß entsprechend:

> „But linguistic phonetics does not have to account for all the sounds that humans are capable of making, or even all of those which can be made just in the vocal tract.
>
> The primary data we will try to describe are all segments that are known to distinguish lexical items within a language."

Aus wissenschaftstheoretischer Perspektive betrachtet gilt, dass Beschreibung Beschreibungskategorien voraussetzt, die – auch dem Linguisten – nicht selbstverständlich gegeben sind. Diese Beschreibungskategorien können für die Phonetik nicht einfach aus der Physik übernommen werden, denn man würde dabei riskieren, dass die Beschreibungen linguistisch uninteressant werden. Dieses Problem wird schon von Sapir im ersten Band der Zeitschrift „Language" gesehen (Sapir, 1925).

Es ist auch klar, dass die Beschreibungskategorien für die akustischen Erscheinungen von Sprachlauten nicht unbedingt identisch sein dürfen mit dem, was ein einzelner Linguist als kompetenter Sprachteilnehmer hört, denn auch der Linguist könnte schon als einjähriges Kind verlernt haben, auf bestimmte lautliche Differenzen, die in seiner Muttersprache nicht vorkommen, zu reagieren, wie es modernere Untersuchungen zum Spracherwerb nahelegen (vgl. z. B. Kuhl, 2000). Eine mögliche Abschwächung des Problems besteht dann darin, die Palette möglicher Beschreibungskategorien dadurch zu beschränken, dass man sie auf phonologischer Basis entwickelt und dieses Verfahren additiv unter Berücksichtigung möglichst vieler Sprachen zur Ableitung eines (angenähert) universellen Inventars nutzt. Dabei fungieren als Hörer dann jeweils die kompetenten Sprecher der einzelnen Sprachen.

Wieder von den einführenden Darstellungen her gesehen, führen die Abgrenzungsschwierigkeiten teilweise zu Koppelungen der konkurrierenden Fachgebiete jeweils unter dem Dach einer der beiden Disziplinen. Es kann die Phonologie als Basis erscheinen, die Phonetik wird dann so gesehen, dass sie sich mit Realisierungen von Phonemen beschäftigt oder es wird Phonetisches als „experimentelle Phonologie“ behandelt. Man findet aber auch Phonologisches als „funktionelle Phonetik“ in die Phonetik eingegliedert, hier ist dann also eher die Phonetik die Basis.

Das Stichwort „experimentelle Phonologie“ ist in diesem Zusammenhang interessant, weil es auf die augenfällige Tatsache hinweist, dass Phonetik und Phonologie unterschiedliche Forschungsmethoden verwenden. Man kann sich leicht davon überzeugen, wenn man die beiden Blackwell-Handbücher, Goldsmith (1995 ed.) zur Phonologie und Hardcastle & Laver (1997 ed.) zur Phonetik, vergleicht. In der Phonetik dominieren Labortechniken, die heute vor allem durch den Einsatz unterschiedlicher elektronischer Messverfahren charakterisiert sind; dazu kommt (eher als Übernahme von Ergebnissen aus der Biologie und Medizin) die Anatomie. In der Phonologie werden hauptsächlich nicht-experimentelle Beobachtungen an existierenden (phonetisch analysierten?) sprachlichen Formen eingesetzt, um zu Formulierungen regelhafter Zusammenhänge zu kommen. Die Möglichkeit, einen regelhaften Zusammenhang herzustellen, dient als Kriterium für den Aufbau phonologischer Theoriegebäude.

Obwohl also in der Tat und wenigstens bis tief in die 90er Jahre des 20. Jahrhunderts die methodischen Spektren verschieden sind, werden doch erhebliche Probleme mit der klassischen Position von Trubetzkoy (1939/1967:7) gesehen, an die man sich dabei erinnert fühlt:

> „... die Sprechaktlautlehre, die mit konkreten physikalischen Erscheinungen zu tun hat, muß naturwissenschaftliche, die Sprachgebildelautlehre dagegen rein sprach- (bezw. geistes- oder sozial-)wissenschaftliche Methoden gebrauchen. Wir bezeichnen die Sprechaktlautlehre mit dem Namen P h o n e t i k, die Sprachgebildelautlehre mit dem Namen P h o n o l o g i e.“

Es ist natürlich nicht ganz leicht, zu sagen, was nun der entscheidende Unterschied zwischen geisteswissenschaftlichen und naturwissenschaftlichen Methoden ist, es sei denn, man sieht ihn in der Verwendung von Messgeräten, wie es der Vergleich der Handbücher ja auch nahelegt. Dieses Kriterium dürfte aber heute, wo das arbeitende Gehirn mit bildgebenden Verfahren wie z. B. PET (positron emission tomography) oder fMRI (functional magnetic resonance imaging) beobachtet werden kann, nicht mehr dieselbe Relevanz

haben, und auch im Bereich außerhalb dieser Verfahren, wenn man z. B. an Reaktionszeitmessungen in psycholinguistischen Experimenten denkt, gilt Entsprechendes. In der Formulierung von Westermann (2000: 41), im Zusammenhang mit anderen Argumenten:

> „Es hat sich gezeigt, dass kontrollierte Beobachtungen und experimentelle Untersuchungen auch in den psychologischen und sozialwissenschaftlichen Wissenschaftsbereichen erfolgreich zur Erkenntnisgewinnung eingesetzt werden können."

Im selben Sinne hat sich auch schon Kohler (1977: 25f.) gegen Trubetzkoys Unterscheidung von Phonetik und Phonologie gewandt. Seine Definition von Phonetik lautet:

> „Der Gegenstand der Phonetik ist das Schallereignis der sprachlichen Kommunikation in allen seinen Aspekten, d. h. die Produktion, die Transmission und die Rezeption von Sprachschall einschließlich der psychologischen und soziologischen Voraussetzungen in der Kommunikationssituation zwischen Sprecher und Hörer, wobei sowohl symbol- als auch meßphonetische Betrachtungsweisen dieses Objekt prägen." (Kohler, 1977: 25; 1995: 22)

Als symbolphonetische Betrachtungsweisen gelten dann Methoden, die sich den phonologischen Methoden annähern.

Wegen der Problematik des Bezugs auf die hauptsächlich angewandten Methoden wäre es interessant, wenn man zeigen könnte, dass nicht (nur) diese Methoden, sondern (auch) die (physikalischen, biologischen) Gegenstände selbst, mit denen sich die Disziplinen beschäftigen, verschieden sind. Das setzt natürlich voraus, dass man der Meinung ist, dass die Phonologie tatsächlich einen solchen Gegenstand hat, was durchaus nicht unproblematisch ist. Wenn man Phoneme als Klassen von Phonen gleicher (bedeutungs-) unterscheidender Funktion betrachtet, kann das zur Konsequenz haben, dass man Phoneme als (biologisch) reale Einheiten verliert und dass man nur die Phone übrigbehält, die zusätzlich durch ihre bedeutungsunterscheidende Funktion charakterisiert sind, Phoneme sind dann wissenschaftliche Abstraktionen. Denn wenn eine Klassenbildung der beschriebenen Art möglich ist, heißt das noch nicht, dass es Einheiten geben müsste (die eine Entsprechung in einer Kortexrepräsentation hätten), die den einzelnen Klassen entsprechen würden. Man kann in diesem Zusammenhang auch darauf hinweisen, dass ein lautliches Phänomen, die Idee der bedeutungsunterscheidenden Funktion von Phonemen vorausgesetzt, schon dadurch phonematisch

wird, dass dem Lexikon eine einzige entsprechende Ausdrucksseite (z. B. ein Wort) hinzugefügt wird. Soll man die Phonologie als Abstraktion über Eigenschaften eines einzelsprachlichen Lexikons verstehen? Einerseits können Laute, die für eine bestimmte Person nicht unterscheidbar sind (siehe oben zu den Konsequenzen des Spracherwerbs), natürlich banalerweise nicht als bedeutungsunterscheidend verwendet werden. Andererseits sind alle unterscheidbaren Laute potenziell als bedeutungsunterscheidende möglich („potenziell bedeutungsunterscheidend“ anders gebraucht als bei Neef, 2005), und können auch zur Weiterentwicklung eines lexikalischen Bestands genutzt werden.

Die Problematik der Beziehung von Phonetik und Phonologie wird weiter verkompliziert, wenn man die generativistische Position, z. B. in der klassischen Version von Chomsky & Halle (1968) mit einbezieht. Die Einheiten Phonem und Phon werden dort aufgelöst zugunsten einer redundanzfreien und durch die Markiertheitstheorie weiter vereinfachten „zugrundeliegenden“ Repräsentation von lexikalischen Einheiten in Merkmalsbündeln. Der Unterschied zwischen den Bereichen kann unter dieser Voraussetzung darin gesehen werden, dass im phonologischen Bereich binäre, im phonetischen Bereich kontinuierlich variierte Merkmale verwendet werden (so Chomsky & Halle, 1968: 65). Damit ist aber für Chomsky & Halle nicht gesagt, dass dieses Kriterium eine scharfe Grenze zwischen den Bereichen ergibt, oder dass man mit genau zwei verschiedenen Repräsentationen von Ausdrücken zu rechnen hätte, eben der phonetischen und der phonologischen:

> „However, a grammar consists of a long sequence of ordered rules that convert initial classificatory representations [Repräsentationen mit binären Merkmalen] into final phonetic ones, and in the intermediate stages there will be representations of a highly mixed sort.“ (Chomsky & Halle, 1968: 65 f.)

Schließlich muss auch noch beachtet werden, dass Regelabfolgen in generativen Grammatiken nicht einfach übersetzt werden dürfen in Prozessschritte im Rahmen der natürlichen Sprachverarbeitung. Es gilt auch für die Phonologie die von Chomsky immer wieder betonte Abstraktheit der linguistischen Charakterisierungen.

Der Stand der Diskussion über das Verhältnis von Phonetik und Phonologie hat etwas Verwirrendes. Es geht offenbar auch nicht nur um die praxisbezogene Abgrenzung von Arbeitsbereichen oder einen Streit um Wörter, sondern um Unklarheiten auf der Ebene von Gegenständen und Methoden. Es ist offenkundig, dass klärende Gesichtspunkte fehlen und es letztlich dann wissenschaftliche Traditionen und weniger Sachzwänge sind, die eine Dif-

ferenzierung nahelegen, oder auch, wie im Fall der klassischen generativen Phonologie, erschweren.

### 3.1.2 Gibt es Argumente für die Unterscheidung von Phonetik und Phonologie in neuronalen Modellen?

Wenn man die Aufgabe der Phonetik darin sieht, die physikalische und biologische Seite der sprachlichen Produktion und Perzeption *erschöpfend* zu behandeln, hat die Phonologie banalerweise keine Funktion in einem neuronalen Modell, das heißt, es gibt keinen Anlass, die Phonologie als eigene Instanz abzugrenzen: Alles ist biologisch bzw. physikalisch, und es ist unter diesem Aspekt nicht sinnvoll, Phoneme als *mentale* Einheiten im Gegensatz zu den *physikalischen* Phonen zu betrachten. Das bedeutet aber zunächst nur, dass das Kriterium „physikalisch“ vs. „mental“ innerhalb eines neuronalen Modells nicht zur Differenzierung dienen kann, nicht, dass damit eine Differenzierung überhaupt sinnlos wird.

Wenn man den Bereichen Phonetik und Phonologie jeweils verschiedene neuronale Strukturen einigermaßen sinnvoll zuordnen möchte, dann ist das relativ leicht für die „Endpunkte“ sprachverarbeitender Prozesse, also für das Gehör einerseits und die Artikulationsorgane andererseits, die beide in der Vergangenheit unbestritten zur Domäne der Phonetik gerechnet worden sind. Die von diesen Endpunkten ausgehenden bzw. dorthin führenden neuronalen Bahnensysteme verlaufen bis in den Kortex hinein getrennt (auch räumlich), und man wird vielleicht dazu neigen, den Bereich, für den diese Trennung gilt, generell der Phonetik zuzuschlagen, unter Hinweis darauf, dass standardmäßig entsprechend spezialisierte Phonetiken (auditiv vs. artikulatorisch), aber nicht entsprechende Phonologien unterschieden werden.

Ein grundsätzliches Problem für diese Vorstellung ergibt sich allerdings aus der Struktur der Verarbeitung im Kortex selbst. Da alle neuronalen Bahnen unidirektional sind, das heißt, Aktionspotenziale nur in einer Richtung fortleiten, und es keine Unterscheidung von statisch zu denkenden Repräsentationen und Verarbeitungsstrukturen geben kann (Voraussetzung für massive zeitliche Parallelverarbeitung im Nervensystem), müssen Perzeptions- und Produktionsbahnen auch in nicht peripherienahen Bereichen überall getrennt vorhanden sein. Wenn von Engelkamp & Rummer (1999) und anderen zwar für Produktion und Perzeption getrennte lexikalische Ausdrucksseiten, aber eine für beide Richtungen *identische* Inhaltsseite („konzeptuelles System“) angenommen wird (vgl. Teil 4, „Lexikon, Morphologie“, Abschnitt

4.2.4), kann sich diese Identität nicht auf die Identität entsprechender Bahnensysteme beziehen. Die Idee der Identität der Inhaltsseiten oder noch „höher“ liegender Verarbeitungsbereiche kann nur heißen, dass eine stärkere *Parallelität* der Bahnensysteme vorauszusetzen ist. Es stellt sich dann neu die Frage, wie weit „nach unten“ diese Parallelität reicht. Obwohl die Bahnentrennung selbstverständlich ist, muss das noch nicht heißen, dass nicht nur die Phonetik, sondern auch die Phonologie damit in zwei Abteilungen, eine artikulatorische und eine auditive Phonologie, zerfallen sollte. Die folgenden Überlegungen behandeln Argumente, die man zur Entscheidung dieser Frage verwenden kann.

Wir sind als (gesunde) Sprecher in der Lage, nicht nur Wörter unserer Muttersprache zu verstehen und zu produzieren, sondern wir können auch Pseudowörter problemlos nachsprechen. (Pseudowörter sind Wörter, deren Ausdrucksseiten den Strukturen der Muttersprache entsprechen, die aber nicht schon vor einem entsprechenden psycholinguistischen Experiment lexikalisiert sind und die auch keine Bedeutung im landläufigen Sinn haben. Ein Pseudowort für einen deutschsprachigen Sprecher/Hörer, das zusätzlich zeigt, dass auch das verwendete Silben*inventar* nicht vorhanden zu sein braucht, ist z. B. [kɔftɪrks].) Dagegen sind Wörter fremder Sprachen oft nur schwer oder gar nicht in lautlich korrekter Form zu realisieren. Ein solches Problem entsteht z. B., wenn ein Sprecher des Deutschen das schwedische [ɧɵtːɪɧʉː] (orthographisch *sjuttisju*, deutsch *siebenundsiebzig*) korrekt nachsprechen soll. Dabei ist es nicht so sehr die ungenügende auditive Wahrnehmung, sondern die dieser Wahrnehmung entsprechende Produktion, die Schwierigkeiten macht. Die Schwierigkeit liegt hier darüber hinaus nicht in einer ungewöhnlichen Silbenstruktur, sondern in der Aussprache der Laute [ɧ], [ɵ] und [ʉː]

Der Hintergrund ist, dass man offenbar in der Lage ist, gelernte Lautstrukturen der eigenen Sprache beim Wiederholen von Pseudowörtern einzusetzen. Es wird gelernte Information verwendet, die nicht schon im Lexikon, in der Menge der Lexikoneinträge steckt, sondern die *einfließt* in den Aufbau neuer lexikalischer Elemente oder neuer sprachlicher episodischer Spuren. Man könnte prinzipiell auch überlegen, ob eine direkte Übernahme von Bestandteilen vorhandener lexikalischer Ausdrucksseiten eine Rolle spielen kann, aus denen sich dann die Pseudoformen zusammensetzen ließen. Die Annahme des dafür erforderliche Extraktionsprozesses dürfte aber in einer neuronalen Architektur erhebliche Probleme mit sich bringen. Also wird man die in Frage stehende Information außerhalb des Lexikons der Ausdrucksseiten ansiedeln, ohne dass damit eine räumliche Trennung notwendig impliziert wäre.

Diese gelernte Information besteht nun aber nicht nur in den Kategorien der auditiven Wahrnehmung und den artikulatorischen Programmen je für sich genommen, sondern muss auch die *Zuordnung* von auditiven Wahrnehmungen einerseits und artikulatorischen Programmen andererseits einschließen, um auf eine Wahrnehmung hin die entsprechende Produktion zu ermöglichen. Eine solche Zuordnung setzt in einer lokalistisch organisierten neuronalen Struktur entsprechende neuronale Verbindungen voraus. Verbindungen dieser Art gibt es nicht direkt zwischen Zellen der Hörbahn und der motorischen Peripherie, also jedenfalls nicht „unterhalb" des Kortex, sondern sie müssen im Kortex selbst angesiedelt werden, abstrakt dargestellt im Schema der Abbildung 3.1.2–1.

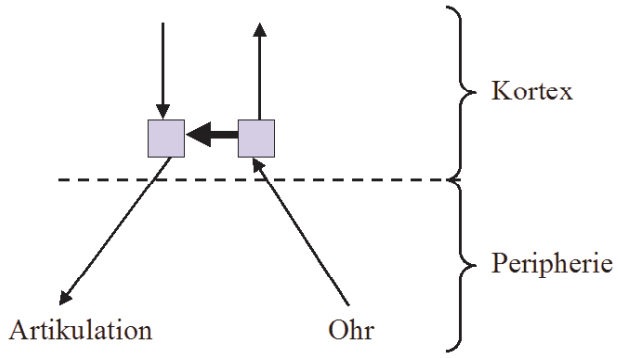



*Abbildung 3.1.2–1:* Schema zu den durch Nachsprechen von Pseudowörtern vorausgesetzten neuronalen Strukturen. Rechtecke stehen für Zellkombinationen, Pfeile sind direkt in entsprechende Bahnen zu übersetzen. Die vorauszusetzende Verbindung zwischen Perzeptions- und Produktionsstrukturen ist fett hervorgehoben.

Die in dem Schema der Abbildung 3.1.2–1 dargestellte Voraussetzung kann, wie in der folgenden Abbildung 3.1.2–2 gezeigt, direkt in eine neuronale, das heißt aus einzelnen Neuronen gebildete Struktur übersetzt werden, deren Begründung in Teil 2, „Grundlagen", Abschnitt 2.5.3 zu finden ist. Man muss dabei nicht von vornherein entscheiden, welche Art von Einheiten durch die Verbindung zwischen Produktion und Perzeption verknüpft sind. Eine einzelne Zelle bzw. ein elementarer Zellverband kann auch für eine Sequenz von Einheiten beliebiger Größe stehen. Allerdings sind, wie das Nachsprechbeispiel zeigt, lexikalische Ausdrucksseiten als Ganzes oder noch größere Gebilde ausgeschlossen (es muss nicht für jede einzelne lexikalische Ausdrucksseite einzeln gelernt werden, wie sie auszusprechen ist), sondern

es muss sich um kleinere Bauteile handeln. Wenn man Silben als Bestandteile lexikalischer Strukturen ablehnt (siehe unten 3.2.3 und Teil 4, „Lexikon, Morphologie“, Kapitel 4.5), kommen dafür Einheiten in Frage, die Phonemen entsprechen oder Komponenten davon (diese Alternative wird erst unten in Kapitel 3.2 entschieden).

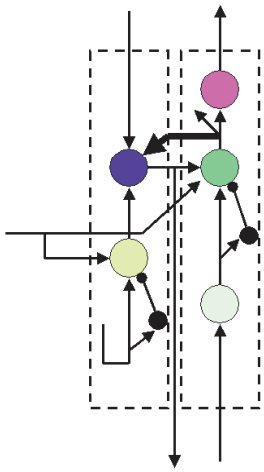


*Abbildung 3.1.2–2:* Neuronale Realisierung, wie in Teil 2, „Grundlagen“, Abschnitt 2.5.3 entwickelt. Kreise entsprechen Neuronen, die verschiedene Einfärbung weist auf etwas unterschiedliche Zellparameter, z. B. für Lernprozesse, Synapseneffektivität und Neurotransmitterwirkung, hin. Die im Schema 3.1.2–2 fett eingetragene Verbindung ist auch hier wieder fett wiedergegeben.

Die gelernten, in der beschriebenen Zuordnung bestehenden Einheiten werden beim Hören von Pseudowörtern in neue lexikalische Ausdrucksseiten oder vergleichbare episodische Spuren eingebaut. Das kann nicht so geschehen, dass ein und derselbe Baustein *X,* (oder genauer: der entsprechende Apparat) direkt über entsprechende neuronale Verbindungen als Element in alle Ausdrucksseiten, die X erfordern, eingebaut werden könnte. Ein solcher direkter Einbezug würde, wie in Teil 4, „Lexikon, Morphologie“, Abschnitt 4.3.2, gezeigt wird, zu unerwünschten Mehrdeutigkeiten im Verstehensprozess führen. Vielmehr muss *X,* um als Sequenzelement in verschiedenen lexikalischen Ausdrucksseiten dienen zu können, Instanzen ausbilden. Die Zahl der Instanzen wird größer sein, als die beim Lernen von *X* selbstverständlich entstehende Repräsentationsredundanz. Während die Instanzen entsprechend ihrer Funktion mindestens teilweise in das Lexikon bzw.

genauer: in die lexikalischen Ausdrucksseiten einbezogen sind, bleiben die Bausteine-als-Typen – jedenfalls grundsätzlich – davon getrennt. Auf diese Weise ergibt sich eine Zwischenschicht von Elementen, die

- in einer Zuordnung von perzeptionsseitigen und produktionsseitigen neuronalen Strukturen bestehen, und
- als komplette Einheiten (über einen Instanzenbildungsprozess) in lexikalische Ausdrucksseiten einbezogen werden können,
- aber trotzdem nicht dem Lexikon angehören.

Weil der „Einbau“ in lexikalische Ausdrucksseiten immer komplett, das heißt durch die komplette Zuordnung von perzeptions- und produktionsseitigen Strukturen (wenn auch nicht notwendig in einem Schritt) erfolgt, ist es nicht erforderlich, perzeptionsseitige Einheiten und produktionsseitige Einheiten mit Bezug auf die Distribution in lexikalischen Ausdrucksseiten zu unterscheiden. Die Distribution ist auf dieser Verarbeitungsebene für beide Verarbeitungsrichtungen notwendig gleich. Wenn in der Phonologie die Distribution, das heißt die Menge möglicher Umgebungen, ausschlaggebendes definierendes Kriterium für ein Phonem ist, und Phoneme in der linguistischen Tradition nicht für Produktion und Perzeption verschieden behandelt werden, sind die Bausteine, von denen hier die Rede war, gute Kandidaten für das, was man als Phoneme bezeichnet hat. Das gilt unter der oben schon angesprochenen stillschweigenden und vorläufigen Voraussetzung, dass es nicht Silben sind, die in lexikalische Sequenzen eingebunden werden und auch nicht Komponenten, die auf der Ebene phonologischer Merkmale anzusiedeln wären..

Man gewinnt damit eine Unterscheidung von Bereichen, in denen die direkte Zuordnung von Produktion und Perzeption neuronal(!) nicht stattfindet, als Gegenstand der Phonetik, und Bereichen, in denen sie stattfindet, als Gegenstand der Phonologie. Diese zunächst relativ beliebig erscheinende Differenzierung hat, wie in den folgenden Kapiteln zu zeigen sein wird, wichtige weitere Konsequenzen für anzunehmende Lernprozesse und Kodierungsformen.

Die üblichen Charakterisierungen der Fächer sind mit dieser Unterscheidung oberflächlich verträglich, und damit kann auch die Beibehaltung der Bezeichnungen legitimiert werden. Unterschiede gibt es allerdings bei Details. Komplementär verteilte Elemente, die phonetisch eine gewisse Ähnlichkeit haben, können nicht allein mit Hinweis auf die komplementäre Verteilung als Allophone in den sublexikalischen Bereich verwiesen werden. Es ist durch

Spracherwerbsprozesse nicht gewährleistet, dass in den Repräsentationen lexikalischer Ausdrucksseiten nur eine zugrundeliegende Form erscheint (also im Fall des Phonems /χ,ç/ im Deutschen die zugrundeliegende Form /ç/). Selbst wenn die Verteilung sublexikalisch verankert ist, können im Lexikon die entsprechenden Allophone auftreten (vorausgesetzt, sie sind ausreichend häufig im sprachlichen Input vertreten).

Andererseits ist es offenbar nicht so, dass bei einer Regularität in der Artikulation immer auch auf der Perzeptionsseite eine entsprechende Regularität repräsentiert sein müsste. Möglicherweise ist dafür im Deutschen das /k/ ein Beispiel, das artikulatorisch als palataler oder als velarer Plosiv erscheint, der Unterschied wird aber auditiv (wenn man versucht, beim Sprechen gegen die Verteilung zu verstoßen) nicht in allen Fällen wahrgenommen, was vielleicht auch der Grund dafür ist, dass er vielfach in Darstellungen der Phonetik des Deutschen übersehen wird. Man beachte in diesem Zusammenhang zur Erklärung das in der „Gestenphonologie“ (z. B. Browman & Goldstein, 1992) behandelte „Stehenbleiben“ artikulatorischer Einstellungen. Das bedeutet, dass man mit den für Allophone typischen Verteilungsmustern im phonetischen Bereich und mit phonetischer Begründung rechnen muss, so dass also die lexikalische Ebene dadurch nicht berührt wird.

Wenn man das letztere Argument akzeptiert, wird noch einmal deutlich, dass es nicht möglich ist, Phonetik nur artikulatorisch oder nur auditiv/akustisch zu verstehen, bzw., was häufig der Fall ist, die Phonetik aus Gründen der Anschaulichkeit hauptsächlich artikulatorisch zu betreiben. Ein gegenüber der Unterscheidung von Produktion und Perzeption neutrales Phon kann nur eine verschleiernde Fiktion sein. Für den Bereich der neuronalen Modellbildung gilt allerdings, dass die akustische, ausschließlich mit dem äußeren Schallereignis beschäftigte Phonetik, die neutral ist gegenüber der Unterscheidung von Produktion und Perzeption, auf Hilfsfunktionen beschränkt bleibt.

Insgesamt ist die Unterscheidung von phonologischen und phonetischen Bereichen in der Sprachverarbeitung und der Zusammenhang der Phonologie mit dem Lexikon der Ausdrucksseiten darzustellen, wie in dem Schema der Abbildung 3.1.2–3.

Zusätzlich bemerkenswert ist dabei, dass „phonematische Typen“ auch „phonematische Instanzen“ bilden, die nicht, oder noch nicht, in lexikalische Sequenzen einbezogen sind. Einzelheiten werden in Teil 4, „Lexikon, Morphologie“, dargestellt.

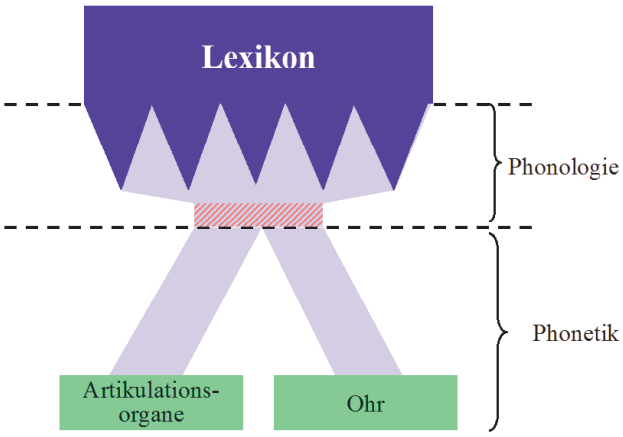



*Abbildung 3.1.2–3:* Schema zur Einordnung der Phonologie in den Sprachverarbeitungsprozess. Der Bereich der phonematischen Typen ist (rot) schraffiert. Das Lexikon verwendet nur einen Teil der gebildeten phonematischen Instanzen (durch die zackenförmige Struktur angedeutet).

Damit ist die Ausbildung von Phonemen in Lernprozessen – von den Auswirkungen des inneren Sprechens abgesehen – als unabhängig von den „Bedürfnissen" lexikalischer Ausdrucksseiten und als rein inputgetrieben anzusehen. Die definierende Funktion des Lexikons (der Menge der Lexikoneinträge) für das Phoneminventar ist schwächer, als in den üblichen phonologischen Theoriebildungen vorgesehen, vor allem, wenn dort Distribution und Bedeutungsunterscheidung als Kriterien gelten. Unter dieser Voraussetzung ist dann auch selbstverständlich, dass sich der Status einer Lautrepräsentation nicht verändert, wenn dem Lexikon eine Ausdrucksseite hinzugefügt wird, die als erste ein perzeptiv unterscheidbares lautliches Element bedeutungsunterscheidend verwendet.

Die relative Unabhängigkeit der Phonologie vom mentalen Lexikon bedeutet nicht, dass das Phoneminventar nicht mehr einzelsprachlich bestimmt ist. Der Input durch die spezifische sprachliche Umgebung legt dieses Inventar in Spracherwerbsprozessen fest, und damit ergibt sich indirekt auch ein Einfluss des Lexikons, hier als mehr oder weniger fluktuierender und im Verlauf der Sprachgeschichte herausgebildeter Besitz der Sprachgemeinschaft verstanden. Analoges gilt auch für die Phonetik. Sofern dort an der Peripherie universelle Komponenten angenommen werden, kann zusätzlich an den Vergessensprozess im Spracherwerb erinnert werden, der im vorigen Abschnitt schon einmal erwähnt worden ist, und der zu einer sprachspe-

zifischen phonetischen Ausstattung zunächst in der Perzeption und davon abhängig dann auch in der Produktion führt.

Die Unterscheidung von Phonetik und Phonologie bleibt, wenn man die vorangegangenen Überlegungen akzeptiert, auch für ein neuronales Modell, in dem es ausschließlich um biologische, also letztlich physikalische Gegenstände geht, sinnvoll und wichtig. Für die Phonologie gilt nicht nur die für Konzepte allgemein gültige Abstraktheit gegenüber der äußeren Realität, also die Unabhängigkeit von den zufälligen Eigenschaften des einzelnen Lautvorkommens, sondern auch eine zusätzliche Abstraktion bezüglich der Besonderheiten von Perzeption einerseits und Produktion andererseits. Das entspricht – mit einigen interpretatorischen Anpassungen – letztlich der schon von Sapir (1921: 55) formulierten Position:

> „These considerations as to phonetic value [innerhalb des Systems einer Einzelsprache] lead to an important conception. Back of the purely objective system of sounds that is peculiar to a language and which can be arrived at only by a painstaking phonetic analysis, there is a more restricted "inner" or "ideal" system which, while perhaps equally unconscious as a system to the naïve speaker, can far more readily than the other be brought to his consciousness as a finished pattern, a psychological mechanism. The inner sound-system, overlaid though it may be by the mechanical or the irrelevant, is a real and an immensely important principle in the life of a language."

### 3.1.3 Biologisch unmögliche Annahmen in phonetischen und phonologischen Theorien

Dieser Abschnitt dient dazu, einige weitverbreitete Ideen, hauptsächlich im Bereich der Phonologie, als für das Verständnis der neuronalen Strukturen der Sprache irrelevant (aber nicht unbedingt irrelevant unter anderen Aspekten) oder neuronal nicht möglich aus der Diskussion der folgenden Kapitel auszuklammern.

*Die „motor theory of perception"*

Die „motor theory of perception" in der in Liberman & Mattingly (1985) revidierten Form besagt, dass die phonetische Analyse sprachlicher Wahrnehmungen durch ein universelles sprachspezifisches Modul geleistet wird, das einen Analyse-durch-Synthese-Prozess auf der Basis artikulatorischer

Gesten, als motorische Programme verstanden, durchführt. Damit werden auch die lexikalischen Ausdrucksseiten als „artikulatorisch“ aufgefasst. Zur Begründung dieser Annahme wird u. a. darauf hingewiesen, dass bei Plosiv-Vokal-Abfolgen ein und derselbe Konsonant wahrgenommen wird, obwohl das tatsächlich beobachtbare akustische Muster des Übergangs (nach der durch den Plosiv verursachten Stille) zu verschiedenen Vokalen verschieden ist.

> „... we suggest what has seemed obvious since the importance of the transitions was discovered: the listener uses the systematically varying transitions as information about the coarticulation of an invariant consonant gesture with various vowels, and so perceives this gesture.“ (Liberman & Mattingly, 1985: 6)

Die Theorie enthält mehrere unterschiedlich zu bewertende Komponenten.

Am ehesten kann noch die Voraussetzung akzeptiert werden, dass sprachliche Schallereignisse anders, das heißt durch andere apparative Komponenten verarbeitet werden, als nichtsprachliche akustische Wahrnehmungen. Details sind allerdings durchaus diskussionsbedürftig und werden unten in Kapitel 3.3 noch genauer behandelt.

Analyse-durch-Synthese ist im Bereich der Phonetik/Phonologie weniger problematisch als in der Syntax, wo eine solche Annahme rasch aufgegeben worden ist. Die Synthese geht von möglichen artikulatorischen Gesten aus und führt zur Vorhersage von „Kandidaten“ für die Interpretation des auditiven neuronalen Inputs. Die Zahl der Kandidaten ist wesentlich überschaubarer als im Beispiel der Syntax, und eine zeitparallele Weiterverarbeitung kann angenommen werden. Ein entsprechender Apparat müsste angeboren sein, Lernvorgänge sind kaum vorstellbar.

Kritischer ist, dass neuronale Verbindungen in Perzeptionsrichtung(!) von der Artikulation her in das für die auditive Perzeption zuständige Modul hinein vorausgesetzt werden, da anders die für die Wahrnehmung zuständigen Konzeptknoten (in einem lokalistischen System die entsprechenden Großmutterzellen) keine als artikulatorisch beschreibbare Bedeutung haben können. Wenn man einmal von der Analyse-durch-Synthese absieht, kann man sich eine neuronale Architektur vorstellen, wie in Abbildung 3.1.3–1 schematisch dargestellt. Hören von Gesprochenem wird damit identisch mit einer vom akustischen Ereignis ausgelösten artikulatorischen Ersatzwahrnehmung und verliert den auditiven Charakter. Das sollte auch bedeuten, dass sprachliche Vorstellungen immer ausschließlich artikulatorische Vorstellungen sind, da es keinen spezifischen *auditiven* Weg in Produktionsrichtung gibt, also keine Möglichkeit auditiver sprachlicher Vorstellungen. Letzteres

widerspricht naiver Beobachtung. Wenn man aber auditive sprachliche Vorstellungen (Ersatzwahrnehmungen) annehmen möchte, müssen auch die entsprechenden auditiven Konzepte vorhanden sein. Wie ein neuronaler Synthesizer aussehen müsste, der sie aus den artikulatorischen Konzepten ableiten und einer Ersatzwahrnehmung zur Verfügung stellen würde, ist offen.

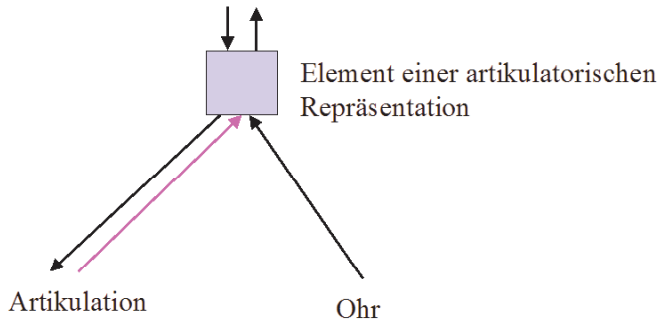



*Abbildung 3.1.3–1:* Schema zur Einbeziehung artikulatorischer Elemente in eine Perzeptionsarchitektur. Die das gezeigte Repräsentationselement als artikulatorisch auszeichnende Bahn ist rot hervorgehoben.

Die entscheidende Schwierigkeit entsteht aber wohl dadurch, dass man auch erklären muss, was Kinder in der Babbelphase des Spracherwerbs tun. Offenbar kontrollieren (gesunde) Kinder ihre sprachliche Produktion über das Gehör. Sie lernen, Sprachlaute zu produzieren, die den gehörten entsprechen, was bereits etablierte auditive Konzepte voraussetzt. Es ist nicht so, dass Kinder babbeln, um hören zu lernen, oder um zu lernen, vorhandene(?) motorische Programme einzusetzen.

Wenn von der Modularität des sprachlichen Perzeptionsapparats ausgegangen wird, kann das im Sinne von Fodors Modularitätsbegriff (Fodor, 1983) zu einer größeren Effektivität des Verarbeitungsprozesses führen. Die Details sind aber unklar. Man vgl. dazu die folgende Passage aus einer „panel discussion“ (Bellugi et al., 1991: 370):

> **„JANET WERKER:** I'd like to address this question to Al [Liberman]. Supposing there is a phonetic module and supposing also that Louis [Goldstein] and Cathe [Browman]'s theory of articulatory phonology turns out to be accurate, and supposing that all parts of it including the idea that the lexicon is specified according to articulatory parameters turn out to be accurate, and that this articulatory phonology can account for both production and perception, then how would you define the bounds of the phonetic module?

> **LIBERMAN:** I'm not sure. I guess the right answer to that is that in the end that's an empirical question. We have to work on that."

Es ist schwierig, sich Modularität vorzustellen, ohne dass Schnittstellen vorausgesetzt werden, die symbolisch kodierte Information repräsentieren.

Insgesamt ist nicht zu sehen, wie die „motor theory of perception" in einer neuronalen Struktur realisiert sein könnte. Experimentelle Befunde, die für die Theorie ins Feld geführt worden sind, müssen neu interpretiert werden.

*Regeln aus generalisierenden Beobachtungen*

Regeln können prinzipiell in einer lokalistischen Architektur durchaus als lokalisierbare Repräsentationen erscheinen. Hier verhalten sich lokalistische Architekturen also anders als verteilte, die diesbezüglich zu entschiedener Kritik Anlass gegeben haben (z. B. Pinker & Prince, 1988).

Beispiele für mögliche Regelrepräsentationen im phonologischen Bereich sind kontextbedingte (sequenzieller Kontext!) Bahnverzweigungen mit parallelen Bahnen für Produktion und Perzeption, wie in Abbildung 3.1.3–2 dargestellt.

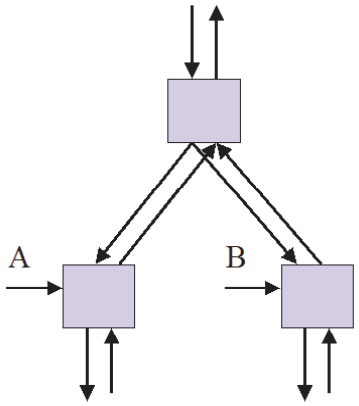



*Abbildung 3.1.3–2:* „Regel" für einen Kontextzusammenhang.

Diesem Schema entspricht die phonologische Regel für ç, aber nicht in der einfacheren Form

/ç/ ⟶ [χ] / [a, o, ɔ u, ʊ] ___

bei der ein ç in der Perzeption auch im Kontext von [a], [o], [ɔ], [u], und [ʊ] zugelassen würde. Regeln müssen auf phonologischer Ebene in beiden Ver-

arbeitungsrichtungen korrekt funktionieren. Es muss also auch der Kontext für die Beibehaltung des ç spezifiziert werden:

$$/ç/ \longrightarrow \left\{ \begin{array}{l} [\,\chi\,]/[\,\mathrm{a},\ \mathrm{o},\ \mathrm{ɔ},\ \mathrm{u},\ \mathrm{ʊ}\,] \_\_ \\ [\,ç\,]/[\,\mathrm{ø},\ \mathrm{œ},\ \mathrm{e},\ \mathrm{ɛ},\ \mathrm{y},\ \mathrm{ʏ},\ \mathrm{i},\ \mathrm{ɪ},\ \mathrm{n},\ \mathrm{l},\ \mathrm{r},\ \#\,] \_\_ \end{array} \right\}$$

Regelrepräsentationen der beschriebenen Art sind also prinzipiell möglich und können die erwarteten Leistungen erbringen. Die entscheidende zusätzlich zu klärende Frage ist dann aber, ob solche Strukturen, soweit sie nicht angeboren sind, durch Lernprozesse zustande kommen können. Mit dieser Frage beschäftigt sich Kapitel 3.5.

Man beachte, dass Regeln in der in Abbildung 3.1.3–2 angedeuteten Form in den Verarbeitungsprozess einbezogen sind. Der Prozess verwendet sie nicht, sondern sie sind Bestandteile des Prozesses (Apparats). Regeln können nicht Regelinventare bilden, die in einem passiven Speicher untergebracht sind. Das würde Adressierungsvorgänge und den Transport kodierter Daten voraussetzen, wie für symbolverarbeitende Modelle typisch. Regeln sind damit nur dann möglich, wenn ihnen buchstäblich selbst Verarbeitungskapazität zukommt und wenn sie in die jeweils aktuellen Verarbeitskomponenten integriert sind, nicht, wenn daran gedacht wird, dass sie eine anderwärts gespeicherte Repräsentation regelhaft verändern. Vorgänge der letzteren Art werden von phonologischen Prozessen in der generativen Phonologie und verwandten Vorstellungen vorausgesetzt.

Die in der Phonologie verwendete Methodik kann zur Formulierung von Regeln führen, überall, wo man eine generalisierende Beobachtung machen kann. Das sind dann teilweise Regeln, zu denen keine neuronale Entsprechung denkbar ist. Wenn z. B. festgestellt wird, dass in natürlichen Sprachen die Zahl der oralen Phoneme immer die Zahl der Nasale übersteigt, so handelt es sich dabei zwar um eine generalisierende Beobachtung, aber eine Beobachtung, die nicht einer Regel entspricht, der eine neuronale Repräsentation zugewiesen werden könnte. Dass für das Phänomen letztlich (in evolutionären Dimensionen wirksame) biologische Faktoren verantwortlich sind, soll damit nicht bestritten werden.

Es gibt auch einzelsprachliche Generalisierungen, die man als solche formulieren kann, die also in diesem Sinn „gelten“, die aber trotzdem nicht biologisch realistisch sind. Man erinnere sich an das Problem der Regeln und zugrundeliegende Strukturen in dem Stil, wie sie in der früheren generativen Phonologie behauptet worden sind. Ein bekanntes Beispiel sind die Regelformulierungen bei Wurzel (1970: 120) zur Beschreibung des deutschen Umlauts.

„(5) Umlauterzwingende Affixe, Konjunktiv und transitive Verben

$$[+\text{silb}] \longrightarrow [-\text{hint}]/ \left\{ \begin{array}{ll} \text{(a)} & \underline{\quad}\,\text{K}_\text{o}\ [+\text{UE}] \\ \text{(b)} & \left[ \begin{array}{l} \left\{ \begin{array}{l} +\text{Stark} \\ +\text{Mod} \\ +\text{R(5b)} \end{array} \right\} \\ +\text{Prät} \\ +\text{Konj} \end{array} \right] \\ \text{(c)} & \underline{\quad}\,\text{K}_\text{o}]_{\text{A,V}}\ [+\text{Trans}]]_\text{V} \\ \text{(d)} & \underline{\quad}\,\text{K}_\text{o}]_\text{A}\ \text{lx} \end{array} \right\}$$

Das Regelschema besagt: (Hintere) Vokale werden zu den (in der Höhe) entsprechenden vorderen vor Suffixen mit dem morphologischen Merkmal [+UE], im Präteritum Konjunktiv der starken Verben, Modalverben und einiger eigens zu diesem Zweck durch ein Regelmerkmal gekennzeichneten schwachen Verben, in aus Adjektiven oder Verben derivierten transitiven Verben sowie in lich-Ableitungen [Tiefenstruktur ist lx] aus Adjektiven.“

Konstruktionen dieser Art sind aufgegeben worden, weil man sich klar gemacht hat, dass die angenommenen zugrundeliegenden Formen, obwohl sie zu Generalisierungen führen, nicht lernbar sind. Die behaupteten Regeln spiegeln sprachgeschichtliche Vorgänge, die zu einer Regularität geführt haben, so dass die sprachgeschichtlichen Vorgänge noch erkennbar sind, aber ohne dass vom Sprecher/Hörer eine Regel zur Realisierung der beobachteten Regularitäten verwendet würde. Möglich ist z. B., dass die zur Beobachtung der Regularität führenden Wortformen einfach lexikalisiert und nicht durch entsprechende Regeln miteinander verknüpft sind.

Freiheit von Redundanz ist kein absoluter Wert, wenn man beachtet, dass Redundanz in sprachlichen Lernprozessen zwangsläufig entsteht und in einem störbaren lokalistischen System eine wichtige stabilisierende Funktion hat.

*Silben*

In Teil 4, „Lexikon, Morphologie“, Abschnitt 4.5.1, wird gezeigt, dass Silben als Komponenten lexikalischer Ausdrucksseiten nicht lernbar sind. Wenn

man diese Ansicht vertritt, wird eine buchstäblich verstandene „Silbenphonologie“ mit Bezug auf neuronale Verarbeitungsstrukturen fragwürdig. Wenn man einzelne Komponenten einer typischen nichtlinearen Silbenrepräsentation, wie sie im Beispiel der Abbildung 3.1.3–3 gegeben ist, auf ihre biologische Aussage hin überprüft, ergibt sich aus dem Lernproblem zunächst, dass die mit $\sigma$ bezeichnete Spitze der Hierarchie jedenfalls nicht einer Großmutterzelle entsprechen kann, deren Bedeutung eine *spezifische* Silbe ist.

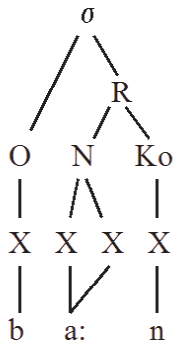



*Abbildung 3.1.3–3:* Nichtlineare Repräsentation einer Silbe (orthographisch *Bahn*). R=Reim, O=Onset, N=Nukleus, Ko=Koda. Weitere Erläuterungen im Text.

Damit ist ausgeschlossen, dass solche baumförmigen Strukturen direkt konstitutive Bestandteile lexikalischer Ausdrucksseiten sein können, was allerdings auch nicht so gedacht ist. Vielmehr steht im Hintergrund das Vorbild syntaktischer Strukturen, bei denen das Symbol S für Satz ja auch nicht für einen spezifischen Satz steht. Der Vergleich mit syntaktischen Kategorien, die durchaus für lernbar zu halten sind, hinkt aber insofern, als die syntaktischen Kategorien eine semantische Basis haben (was zu der – allerdings ungeeigneten – Annahme des semantischen Bootstrapping geführt hat), silbische Kategorien aber rein formal bestimmt sind.

Das aus Xen bestehende metrische Skelett stellt die zeitliche Struktur als Folge zeitlicher Einheiten dar und dient der Darstellung der Akzentstruktur. Akzent kann einzelnen Elementen der lexikalischen Ausdrucksseiten zugeschrieben werden und unterliegt damit nicht dem Lernproblem. Außerdem gilt, dass Länge in Lexikonrepräsentationen(!) tatsächlich in gewissem Sinne analog zu der nichtlinearen Analyse gelöst werden muss, nämlich dadurch, dass, soweit die Länge eine bedeutungsdistinktive Funktion hat, anstelle

einer einzelnen Großmuttereinheit zwei Großmuttereinheiten den entsprechenden Laut repräsentieren. Aus /ba:n/ wird also /baan/. Man vgl. dazu wieder Teil 4. Möglich ist generell die Annahme, dass Phoneme in bestimmten Silbenpositionen bestimmte, direkt in den lexikalischen Ausdrucksseiten zu verankernde Eigenschaften haben, so dass man also dem Strukturbaum der Abbildung 3.1.3–3 entsprechende Informationen, nicht die Struktur selbst, direkt im Lexikon wiederfinden würde. Diese Frage wird unten in Kapitel 3.5 noch einmal genauer behandelt.

Es mag offen bleiben, ob das Problem mit den Silbenstrukturen verschwinden würde, wenn man auf die Eigenschaft der Parallelverarbeitung im Gehirn verzichten und eine durch und durch symbolverarbeitende Gesamtstruktur voraussetzen könnte. Nur in einer symbolverarbeitenden Gesamtstruktur könnten Regeln auf Silbenstrukturen zurückgreifen.

Insgesamt gilt, dass Beschreibungen, wie sie in der nichtlinearen Phonologie gegeben werden, nicht in buchstäblich entsprechende neuronale Strukturen umsetzbar sind. Das bedeutet, dass die Erklärungen, die mit Hilfe des Silbenkonzepts für bestimmte Phänomene gegeben werden, revidiert werden müssen. Man beachte in diesem Zusammenhang, dass gezeigt werden muss, dass Erklärungen, die mit Silbengrenzen arbeiten, nicht auch unter Verwendung von Morphemgrenzen gegeben werden können. Darauf wird auch in Teil 4.5.1 hingewiesen. Hall (2000: 208) bringt Beispiele, die für eine Rolle der Silbengrenze bei der Regelung der Auslautverhärtung sprechen sollen. Die Silbengrenzen in diesen Beispielen sind aber allesamt identisch mit Morphemgrenzen (*streb+sam, Bünd+nis, bieg+sam, les+bar*). Entsprechendes gilt für eine Liste von Beispielen, die zeigen soll, dass die Silbengrenze bedeutungsdistinktiv ist, ebenda S. 235.

Es gilt zusätzlich auch hier wieder die Warnung, dass nicht alles, was als Regularität erkannt werden kann, auch einer Regel entsprechen muss bzw. als Begründung für entsprechende erforderliche Strukturen herangezogen werden darf.

*Constraints (optimality theory)*

Die „optimality theory“ erbt aus der generativen Sprachtheorie die hohe Abstraktheit aller theoretischer Aussagen.

Prince & Smolensky (2004: 5) beschreiben die Struktur einer optimalitätstheoretischen Grammatik so:

**Structure of Optimality-Theoretic Grammar**
(a) Gen ($\mathrm{In_k}$) $\longrightarrow$ {$\mathrm{Out_1}$, $\mathrm{Out_2}$, . . . }
(b) H-eval ($\mathrm{Out_i}$, $1 \leq i \leq \infty$) $\longrightarrow$ $\mathrm{Out_{real}}$

„Gen“ bezeichnet einen Generierungsprozess, der aufgrund eines Inputs (einer zugrundeliegenden Struktur) eine Menge von Outputkandidaten erzeugt, die dann durch die Bewertungsprozedur „H-eval“ in eine Reihung gebracht werden, so dass schließlich ein bester, das heißt universelle Constraints am wenigsten verletzender Output gefunden werden kann. Wichtigste Kandidaten für die Pluralform des englischen Worts *hat* von einem Input /hæt+z/ ausgehend, sind [hætz], [hæts] und [hætɪz]. Die Alternative wird dadurch zugunsten von [hæts] entschieden, dass [hætz] ein Constraint verletzt, das eine Abfolge von stimmhaftem und stimmlosem Verschlusslaut oder umgekehrt verhindert, und [hætɪz] ein Constraint, das Epenthese verhindert (Beispiel nach Hall, 2000: 325).

Charakteristisch ist, dass Prozesse beschrieben werden, die genauso wenig wie die sonst in der generativen Sprachtheorie formulierten, reale zeitliche Abfolgen, reale Outputs(!) und reale Inputs(!) spezifizieren, sondern als Prozesse in einem abstrakteren mathematischen Sinn zu verstehen sind. Beispiele: Eine unendliche Kandidatenmenge, die einer Bewertung unterliegt, kann nicht realer Bestandteil eines tatsächlich und in sehr begrenzter Zeit ablaufenden Prozesses sein (vgl. die Charakterisierung des Prozesses H-eval oben). Der Längenunterschied zweier Listen von „marks“, der zur Entscheidung über den besten Kandidaten führt, kann neuronal nicht dadurch bestimmt werden, dass so lange jeweils ein Element beider Listen entfernt wird, bis eine der Listen erschöpft ist (Prince & Smolensky, 2004: 83).

Die Annahme universeller Constraints bringt das oben schon erwähnte Bootstrapping-Problem mit sich, das, wie in Kochendörfer (2002: 5.1) gezeigt, nicht lösbar ist. Dieser Kritikpunkt wäre nur vermeidbar, wenn diese Constraints durch vollständige Ketten angeborener Bahnen mit den angeborenen Kategorien der Sinnesperipherie verknüpft gedacht werden könnten, was für Constraints wie Ons (Ons untersucht, ob Silben einen Onset haben), Dep-IO (Dep-IO verhindert Epenthesen, siehe oben) und eine größere Zahl anderer Beispiele kaum möglich sein dürfte.

Die „optimality theory“ gibt keine Hinweise darauf, wie sie im Gehirn untergebracht werden könnte, und sie ist auch gar nicht so gedacht, dass das möglich sein sollte. Es wird nicht einmal akzeptiert, dass sie realistische Berechnungsvorgänge irgendwelcher Art ermöglichen müsste.

> „It is not incumbent upon a grammar to compute, as Chomsky has emphasized repeatedly over the years. A grammar is a function that assigns structural descriptions to sentences; what matters formally is that the function is well-defined. [...] Grammatical theorists are free to contemplate any kind of formal device in pursuit of these goals ; [..]“ (Prince & Smolensky, 2004: 233)

Es ist erstaunlich, dass eine Theorie, die so wenig praktische Relevanz hat, auch im 21. Jahrhundert noch so breite, weltweite Beachtung gefunden hat.

*Buffers*

Pufferspeicher („buffers“) werden besonders in Sprachproduktionsmodellen gerne eingesetzt, um Zwischenergebnisse des Produktionsprozesses festzuhalten. Ein bekanntes Beispiel ist der „articulatory buffer“ in dem Sprachproduktionsmodell von Levelt (1989). Er wird wie folgt begründet:

> „The interface of phonological encoding and articulation involves a system that can temporarily store a certain amount of phonetic plan. [...] Sustaining a fluent, constant rate of speaking requires a storage mechanism that can buffer the phonetic plan (the speech motor program) as it develops. It can, presumably, contain a few phonological phrases. Moreover, it *must* contain a minimal amount of program in order for speech to be initiated—probably as much as a phonological word.“ (Levelt, 1989: 414)

Die „innere“ Verarbeitung geht der „äußeren“ phonetischen Realisierung von sprachlichen Äußerungen voraus, es entsteht ein Synchronisationsproblem, das durch einen Pufferspeicher gelöst werden muss. Details sind aus psycholinguistischen Experimenten abgeleitet. Die Funktion des Speichers beim Auslesen im Fall eines Äußerungsanfangs wird so beschrieben (Levelt, 1989: 421):

> „When the speaker decides to start a prepared utterance, its motor units (i.e., the phonetic plans for the phonological phrases) are retrieved from the Articulatory Buffer. The time needed to retrieve each unit depends on the total number of units in the buffer.“

Levelts Formulierungen lassen letztlich keinen Zweifel daran, dass an eine Speicherkomponente gedacht ist, die eine Anzahl von Speicherzellen enthält, die mit einem jeweils anstehenden Inhalt beschickt werden können. Die Speicherzellen selbst sind inhaltsneutral. Solche Eigenschaften sind typisch für symbolverarbeitende Modelle (vgl. Teil 2, „Grundlagen“, Abschnitt 2.2.2) und sind mit den Möglichkeiten des Kortex nicht vereinbar.

Obwohl eigentlich klar ist, dass die Annahme eines Pufferspeichers im Kortex zu grundsätzlichen Schwierigkeiten führen muss, wird in Kochendörfer (1999) zusätzlich überprüft, ob nicht ein Speicher, der begrenzte Adressierungsmöglichkeiten hat und nach dem Prinzip „first in first out“ funktioniert, im Kortex mit der Funktion eines sublexikalischen Buffers denkbar

wäre. Das Ergebnis ist auch bei der Annahme dieser Möglichkeit negativ, oder genauer: die Speicherfunktion ergibt sich letztlich als Funktion der Lexikonstruktur. In Teil 4, „Lexikon, Morphologie“, wird gezeigt, dass das ausdrucksseitige Lexikon tatsächlich eine gewisse zeitliche Flexibilität der Verarbeitung in Produktions- und Perzeptionsvorgängen gewährleistet. Es ist aber nicht so, dass ganze phonologische Wörter oder phonologische Phrasen gepuffert werden können. Bei Licht besehen ist eine Speicherung von Elementen in dieser Größenordnung auch deshalb unplausibel, weil sie ein den Beobachtungen entsprechendes rasches auditives Monitoring und rasche Reparaturen unmöglich macht.

Erklärungen von experimentellen Befunden, die die Existenz von Pufferspeichern voraussetzen, müssen auf andere Weise gegeben werden. Dabei ist u. a. an Funktionen des inneren Sprechens zu denken, die zu zeitlichen Verzögerungen des artikulatorischen Outputs führen können.

# 3.2 Die Einheit des Phonems

## 3.2.1 Vertikale Einheit

In der klassischen generativen Phonologie sind Notationen, in denen Phoneme durch einzelne Zeichen angegeben werden, nur bequeme Abkürzungen für die eigentlich gemeinten Bündel phonologischer Merkmale. Diese Auffassung gilt bis in die Gegenwart (vgl. z. B. Hall, 2000: 119). Nur die Merkmale sind, wie oben in Abschnitt 3.1.1 schon erwähnt, die eigentlich realen Bestandteile einer phonologischen Beschreibung, das Phonem als Einheit ist keine linguistisch relevante Größe. Die dabei in Frage stehende Einheit wird hier als „vertikale“ Einheit bezeichnet, im Unterschied zu einer „horizontalen“ Einheit, die als Einheit auf der Zeitachse zu verstehen ist.

Die Auflösung der vertikalen Einheit des Phonems durch den Generativismus führt dazu, dass auch die zugrundeliegenden Formen der lexikalischen Ausdrucksseiten, das heißt, die im mentalen Lexikon repräsentierten Formen, als Bündel von phonologischen Merkmalen gegeben sind. Unter dieser Annahme ist es dann möglich, die Spezifizierung der lexikalischen Segmente so weit zu reduzieren, dass nur noch kontextuell nicht vorhersagbare und nicht-universelle Merkmale festgelegt werden. Vorausgesetzt werden „interpretive conventions“ wie die folgende:

$$[u\ \text{back}] \longrightarrow \left\{ \begin{array}{l} [-\text{back}]/ \left[ \begin{array}{c} \overline{u\ \text{ant}} \\ -\text{low} \end{array} \right] \\ [+\text{back}]/ \left\{ \begin{array}{l} \left[ \overline{m\ \text{ant}} \right] \\ \left[ \overline{+\text{low}} \right] \end{array} \right\} \end{array} \right\}$$

Dabei steht $u$ für „unmarkiert“ und $m$ für „markiert“. Da angenommen wird, dass unmarkierte Merkmalsausprägungen nicht zur Komplexität ei-

nes Lexikoneintrags beitragen, werden sie nicht notiert. Das ergibt für die redundanzfreie und die entsprechende ausgefüllte Merkmalsmatrix des englischen Worts *stun* nach dem in Chomsky & Halle (1968: 415) gegebenen Beispiel die in Tabelle 3.2.1–1 dargestellten Werte.

| segment | m | m | m | m | + | + | + | + |
|---|---|---|---|---|---|---|---|---|
| consonantal | | | | | + | + | − | + |
| vocalic | | m | | | − | − | + | − |
| nasal | | | | m | − | − | − | + |
| low | | | | | − | − | − | − |
| high | | | | | − | − | + | − |
| back | | | + | | − | − | + | − |
| round | | | | | − | − | + | − |
| anterior | | | | | + | + | − | + |
| coronal | | + | | | + | + | − | + |
| continuant | | | | | + | − | + | − |
| delayed release | | | | | + | − | + | − |
| strident | | | | | + | − | − | − |

*Tabelle 3.2.1–1:* Lexikalische Repräsentation des englischen Worts *stun*. Links die redundanzfreie, markiertheitstheoretische Merkmalsmatrix (der Wert *u* für „unmarkiert“ wird nicht geschrieben), rechts die Entsprechung in binären Merkmalen.

Das Prinzip der Unterspezifizierung lexikalischer Repräsentationen wird differenziert und relativiert von Mohanan (1991), ohne dass damit die Auflösung der Einheit des Phonems grundsätzlich angegriffen würde.

Da die Redundanzfreiheit mentaler Repräsentationen generell ein fragwürdiges Konzept ist und hier mit Regelanwendungen verbunden ist, die dem Charakter nach zu einem symbolverarbeitenden Modell gehören würden, können wir diesen Aspekt erst einmal beiseite schieben. Was hier zunächst interessieren muss, ist allgemein die Möglichkeit, Merkmalsmatrizen als Repräsentationen lexikalischer Ausdrucksseiten in einer neuronalen Struktur zu sehen. Es muss gewährleistet sein, dass eine derartige Spezifikation sowohl in Produktions- als auch in Perzeptionsrichtung befriedigend funktioniert. Vom Standpunkt der Sprachperzeption her gesehen muss ein Lexikonabgleich ermöglicht werden, der zur Identifikation des auditiven Inputs und letztlich zur Zuweisung von inhaltlichen Informationen führt. Dazu müssen die Merkmalsausprägungen jeder Matrixspalte der Darstellung in Tabelle 3.2.1–1 in zeitlicher Folge in dem Sinne abgeprüft werden, dass ein Übergang zur folgenden Spalte nur dann „erlaubt“ ist, wenn alle Merkmale einer

Spalte im sprachlichen Input eine Entsprechung haben. Eine solche Überprüfung muss synchron mit dem Verlauf des Inputs erfolgen, nicht erst, wenn der Input einer lexikalischen Einheit abgeschlossen ist (vgl. dazu Teil 4, „Lexikon, Morphologie“). Die Überprüfung muss, neuronal interpretiert, in der Feststellung der Koinzidenz der Aktivität von Zellen bestehen, die die einzelnen Merkmale repräsentieren. Es werden damit zusätzliche Zellen vorausgesetzt, die in diesem Sinne als Koinzidenzdetektoren arbeiten, also z. B. in einer Anordnung, wie in Abbildung 3.2.1–1 skizziert. (Solche Schlussfolgerungen könnten natürlich auch allesamt mit dem Hinweis auf die Abstraktheit der Konzeptionen in der generativen Sprachtheorie abgelehnt werden.)

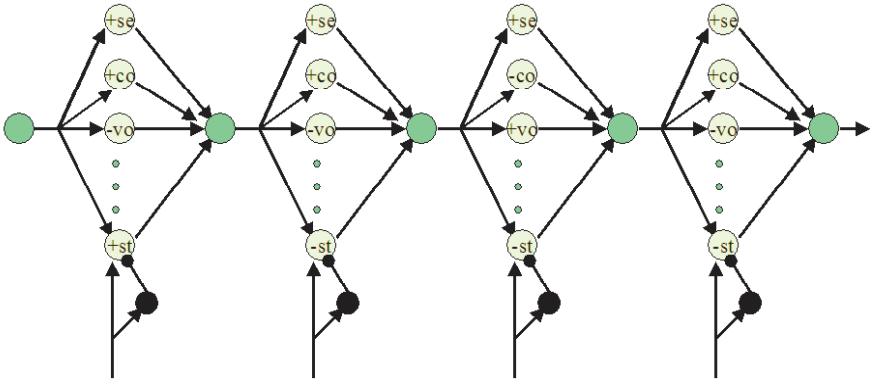



*Abbildung 3.2.1–1:* Beispiel einer neuronalen Architektur, die eine lexikalische Merkmalsmatrix realisiert. Die hemmenden Strukturen, die zu jeder einzelnen Merkmalsrepräsentation gehören, sind der Übersichtlichkeit halber nur für die untersten Merkmale angegeben.

Simulationen:
Lexikonabgleich mit der in Abbildung 3.2.1–1 wiedergegebenen Struktur:
1. Korrekter Input, der Spezifikation entsprechend.
2. Abweichender Input.
Die Symbolik der Bildschirmdarstellung ist in Teil 2, „Grundlagen“, Abschnitt 2.1.4, erklärt.

Die Simulation setzt stillschweigend voraus, dass das von Zelle zu Zelle weitergegebene Signal jeweils in einem einzelnen Impuls besteht. Ohne diese Voraussetzung kann, wie in Kochendörfer (1997: 83 ff.) gezeigt ist, der Lexikonabgleich nicht funktionieren und es sind auch die für den Aufbau des

mentalen Lexikons erforderlichen Lernprozesse nicht zu realisieren (vgl. dazu Teil 2, „Grundlagen“ 2.3.2 und 2.4.6). Ergänzende Überlegungen dazu finden sich im folgenden Abschnitt 3.2.2.

Der durch die beschriebene Konstruktion geforderte Zellaufwand ist sehr hoch. Selbst wenn man in Betracht zieht, dass eine Reduktion der Redundanz eine Erleichterung bringen würde, ist es doch so, dass, auf das Lexikon als Ganzes gesehen, wiederkehrende Merkmalskombinationen vielfach wiederholt werden müssen. Es fällt auch auf, dass die Feststellung der Vollständigkeit eines Merkmalsbündels neuronale Einheiten vorsieht, die sich wie Phonemeinheiten verhalten, mit dem Unterschied, dass sie in jeder Position, in der sie vorkommen, stets neu durch ein spezifisches Merkmalsbündel definiert werden. Da bei der Konstruktion der Abbildung 3.2.1–1 jedes einzelne Merkmal die Last der Sequenzkontrolle sozusagen mitträgt, ergibt sich – unter Beachtung der Lernprobleme, die in Teil 2, „Grundlagen“, Abschnitt 2.4.6, besprochen sind – eine zusätzliche Vermehrung der Anzahl erforderlicher Zellen. Wenn man dieses letztere Problem vermeiden möchte, kann man die Aufgabe der Sequenzkontrolle auf die die Merkmalsbündel auswertenden Zellen verlagern. Man bekommt dann eine Konstruktion, die der Abbildung 3.2.1–2 entspricht.

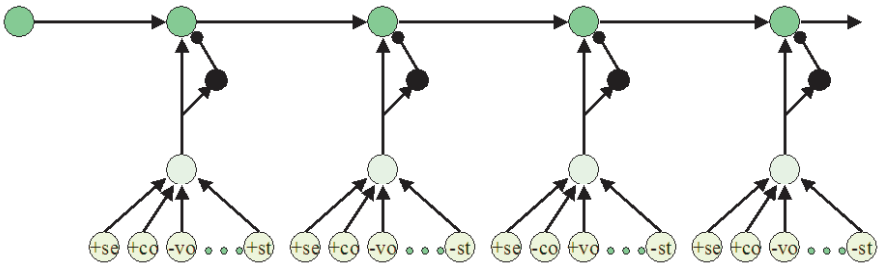



*Abbildung 3.2.1–2:* Beispiel einer neuronalen Architektur, die eine lexikalische Merkmalsmatrix realisiert. Die Sequenzkontrolle ist den geeignet erweiterten Koinzidenzdetektoren übertragen, die Merkmalseinheiten sind aus der sequenziellen Abfolge herausgenommen.

Man kann sich jetzt weiter fragen, ob wirklich anzunehmen ist, dass die mit der Sequenzkontrolle beauftragten Zellen jeweils bei der Verankerung einer lexikalischen Ausdrucksseite durch ihre entsprechenden Merkmalsbündel definiert werden und zwar so, dass so viele Repräsentationen von Merkmalsbündeln vorhanden sind, wie Segmente in den Ausdrucksseiten des

mentalen Lexikons insgesamt. Im Zusammenhang damit stehen zwei weitere Probleme:

- Kann man eine Konzeptbildung für Elemente verhindern, die wie die phonetischen bzw. phonologischen Merkmale in einem ausreichend kleinen Zeitfenster aktiviert werden?
- Wie soll man unter solchen Umständen die im vorigen Kapitel besprochenen Unterschiede zwischen dem Nachsprechen von Pseudowörtern der eigenen Sprache und dem Nachsprechen von fremdsprachlichen Wörtern erklären?

Wenn man, um diesen Schwierigkeiten Rechnung zu tragen, mit einer „Vorratsbildung“ für Einheiten rechnet, die dann als Sequenzelemente in lexikalische Ausdrucksseiten eingebaut werden und man zusätzlich beachtet, dass dieselbe merkmalsrepräsentierende Zelle Verbindungen zu verschiedenen solcher Einheiten haben kann, erhält man zwangsweise Vorstellungen, die dem klassischen Konzept des Phonems als ausdrucksseitigem Elementarbaustein entsprechen.

Ein vielleicht erwarteter Gewinn an „Einfachheit“ mentaler Repräsentationen bei einem Ersatz von Phonemen durch Merkmalsbündel, wodurch also Phonemeinheiten überflüssig werden, ist nicht zu sehen, und das würde auch in dem aus anderen Gründen zusätzlich schwierigen Fall der Annahme von redundanzfreien Merkmalsmatrizen und des Zusammenspiels mit phonologischen Regeln gelten. Man beachte, dass diese Bewertung nicht durch das Zählen von Symbolen in phonologischen Notationen zustandekommt (vgl. dazu die klassische Diskussion in der Halle-Houshoulder-Kontroverse der 60er Jahre: Householder, 1965; Chomsky & Halle, 1965; Matthews, 1968), sondern durch die Abschätzung des Aufwands an neuronalen Strukturen.

Die voranstehenden Überlegungen gelten zunächst für die Sprachperzeption. Wenn man auf die Idee der Merkmalsrepräsentation lexikalischer Ausdrucksseiten verzichtet, entfällt auch das Problem, dass es schwer fällt, entsprechende Repräsentationen zu finden, die in gewissem Sinne neutral sind gegenüber Perzeption und Produktion. Ein wahrgenommenes Pseudowort muss nicht durch einen eigenen Lernprozess mit einer für die Produktion gültigen Form verknüpft werden, und die Idee einer direkten (angeborenen?) Entsprechung von Perzeptions- und Produktionsmerkmalen ist angesichts der Vorgänge in der Lallphase des Spracherwerbs ohnehin nicht plausibel, vgl. ergänzend oben 3.1.3 zur „motor theory of perception“.

Es spricht insgesamt gesehen vieles dafür, dass Phoneme „vertikale“ Ganzheiten sind. Sie müssen durch Großmutterzellen (bzw. entsprechende kleine Zellverbände) neuronal repräsentiert werden und sind in dieser Form Bau-

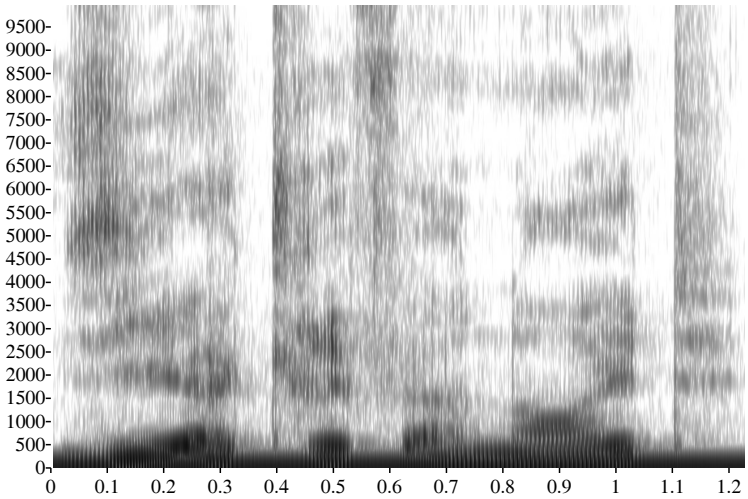
steine lexikalischer Ausdrucksseiten. Phonetische/phonologische Merkmale sind ebenfalls durch Großmutterzellen repräsentierte, bezüglich der Phoneme definitorische, näher an der Sinnesperipherie bzw. motorischen Peripherie liegende Konzepte.

## 3.2.2 Horizontale Einheit

Man hat sich in der Linguistik daran gewöhnt, von phonologischen *Segmenten* zu sprechen und betont damit die zeitliche Erstreckung der Phone/Phoneme und eine zeitliche Grenzziehung zu Vorgängern und Nachfolgern. Wenn man, wie im vorangegangenen Abschnitt vorausgesetzt, annimmt, dass eine Phonemeinheit als Signal an eine folgende Einheit gerade einen einzelnen neuronalen Impuls abgibt, und nur bei größerer Länge, also nicht notwendig, mehrere Impulse, wird die Vorstellung von der zeitlichen Erstreckung, jedenfalls auf neuronaler Ebene betrachtet, fragwürdig. Auch ein einzelner neuronaler Impuls hat natürlich eine zeitliche Dauer (von größenordnungsmäßig einer Millisekunde), diese Dauer ist aber nicht systematisch veränderlich und liegt weit unterhalb dessen, was man einem Phonem als Dauer zumessen würde. Damit schrumpft sozusagen das einzelne Element einer phonologischen Sequenz zeitlich gesehen auf einen Punkt. Diese Zeitproblematik wird hier mit dem Stichwort „horizontale Einheit“ thematisiert.

Die Vorstellung von der Dauer phonologischer Segmente wird durchaus auch gestützt durch spektrographische Analysen, die eine typische „Streifigkeit“ zeigen, obwohl immer wieder auch behauptet wird, dass das akustische Signal „kontinuierlich“, also eine Abgrenzung phonetischer Segmente grundsätzlich schwierig und vielleicht prinzipiell unangemessen sei. Dazu muss man sich klar machen, dass, solange man sich nicht in den subatomaren Bereich begibt, natürliche Ereignisse generell kontinuierlich sind und es auch für artikulatorische oder daraus entstehende akustische Phänomene nur um den Zeitbedarf für Übergänge von einem Muster (nicht „steady state“) zu einem anderen gehen kann. Wenn man in das Spektrogramm der Abbildung 3.2.2–1 **A** unter zusätzlicher auditiver Kontrolle Segmentgrenzen wie in Teilabbildung **B** einzeichnet, wird die Übergangszone jeweils durch die Strichstärke der Grenzmarkierung, die auf der Zeitskala ca. 10 ms beträgt, abgedeckt. Die Übergangszone ist gegenüber der gesamten Segmentdauer, die mehr als 60 ms beträgt, relativ kurz. Ausnahmen sind die Übergänge zwischen den Bestandteilen von Diphthongen und vielleicht auch andere Vokalübergänge, wie in dem verwendeten Beispiel von [sɪ] nach [ɛ⌣ɐ]. Ein gutes englischsprachiges Beispiel für eine entsprechende Analyse findet sich bei Kent, Dembowski & Lass (1996: 198).

**A**




**B** [ z i ɛ‿ɐ g ʁ ɪ f ɛ‿ɐ n ɔ‿ɪ t ʰ ]

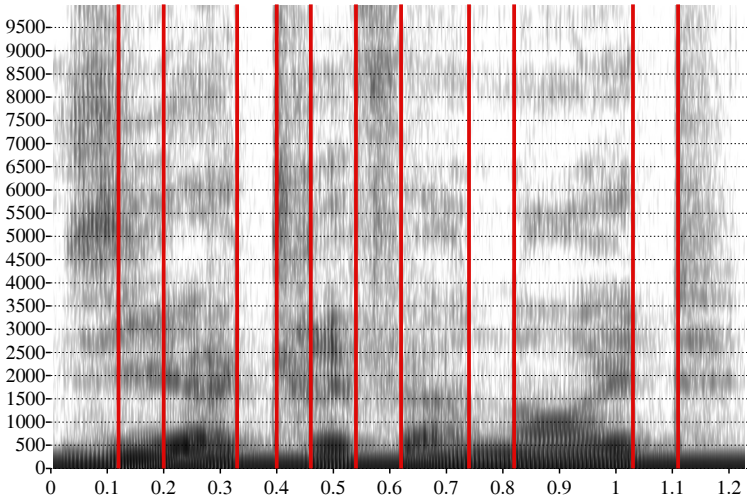



*Abbildung 3.2.2–1:* **A:** Spektrogramm des Äußerungsfragments *sie ergriff erneut* .... **B:** Dasselbe Spektrogramm mit Zuordnung der phonetischen Interpretation und Versuch der Festlegung von Segmentgrenzen (durch Abhören zusätzlich verifiziert). Aufnahme unter natürlichen Bedingungen mit Nebengeräuschen und ohne spezielle Apparaturen.

Tonbeispiel:
Ton *sie ergriff erneut* zu Abbildung 3.2.2–1.

(Alle in diesem Teil 3, „Phonetik/Phonologie“ wiedergegebenen Spektrogramme und Veränderungen an Tonbeispielen sind, wenn nicht anders vermerkt, mit dem unter http://www.praat.org erhältlichen Freeware-Programm PRAAT hergestellt.)

Es ist also durchaus möglich, phonetische Segmente als zeitlich im Bereich von einigen zehn Millisekunden erstreckte Gebilde zu sehen. Das bedeutet allerdings noch nicht, dass man auf *phonologischer* Ebene und z. B. für lexikalische Repräsentationen mit einer Abbildung dieser Zeitdauern durch eine neuronale Aktivität rechnen darf. Eine solche Abbildung könnte nur durch einen entsprechend andauernden Burst von Aktionspotenzialen geschehen. Also lautet die Frage, ob man mit solchen Bursts auf phonologischer Ebene rechnen kann.

Diese Frage kann auch schon allein mit Hinweis auf die erforderlichen lexikalischen Funktionen negativ beantwortet werden, siehe die Bemerkungen oben in Abschnitt 3.2.1 und generell Teil 4, „Lexikon, Morphologie“. Man kann aber zu einer zusätzlichen Stütze dieser Auffassung kommen, wenn man z. B. den akustischen Verlauf von stimmhaften Plosiven vor Vokalen betrachtet. Er beginnt mit einer Phase der Stille, deren Dauer wenig kürzer ist als die eines Kurzvokals. Während dieser Dauer sind alle Plosive, von einer eventuellen Stimmbeteiligung abgesehen, gleich. Es folgt ein Übergang, dessen Gestalt zusätzlich vom Folgesegment abhängig ist. Erst diese Übergangsphase legt die Qualität des Plosivs letztlich fest und muss im Verstehensprozess zur Auslösung einer spezifischen, das Phonem identifizierenden neuronalen Reaktion führen. Der diese Reaktion auslösende Plosiv ist, grob geschätzt, innerhalb 10 bis 20 ms nach Beginn des Übergangs abgeschlossen. Die im Gehirn beobachteten Impulsfrequenzen lassen aber innerhalb einer solchen Zeitspanne nicht mehr als einen einzelnen Impuls zu. Mindestens die Entscheidung für ein ganz bestimmtes Phonem im Bereich der stimmhaften Plosive wird offenbar über einen einzelnen Impuls vermittelt, die Gesamtdauer des Phonems wird nicht durch Impulse auf einer spezifischen Bahn begleitet. Einzelne Verarbeitungsschritte, die zu einer Abfolge von Impulsen in verschiedenen Zellen führen, können auf phonetischer(!) Ebene erforderlich sein, um diese Entscheidung am Ende herbeizuführen; das Phonem, als Element lexikalischer Ausdrucksseiten, muss aber als zeitliche („horizontale“) Einheit gesehen werden, die einem einzelnen neuronalen Aktionspotenzial entspricht.

Man kann diese These zusätzlich durch ein einfaches Experiment stützen, das zeigt, dass die Phase der Stille nicht nur am Äußerungsanfang, wo sie

wohl ohnehin nicht auswertbar ist, sondern auch im Innern einer Äußerung zur Identifikation eines Plosivs nicht beiträgt. Wenn man die dem Verschlusslaut [g] entsprechende Stille in der Tondatei in dem Äußerungsfragment *sie ergriff erneut* wegnimmt, ist die Veränderung auditiv kaum spürbar. In der Version ohne Stille erscheint der vorangehende Diphthong verkürzt, vielleicht sogar auf [ɛ].

Die Abbildung 3.2.2–2 zeigt das entsprechende Spektrogramm, die Tondateien ermöglichen den Nachvollzug des Vergleichs.

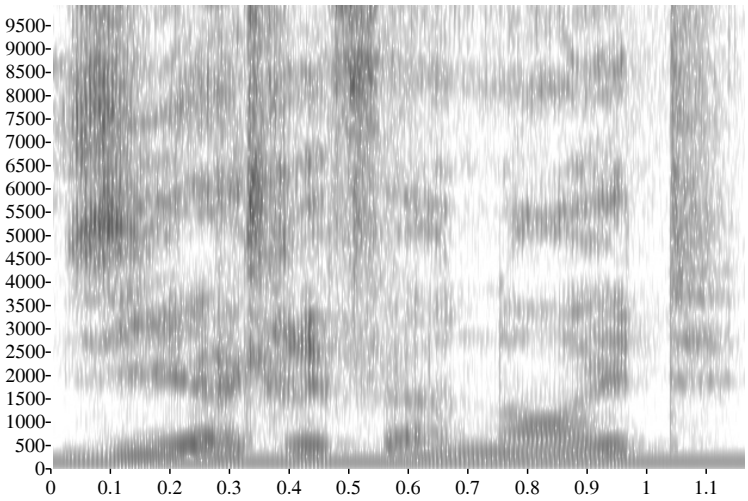



*Abbildung 3.2.2–2:* Spektrogramm des Äußerungsfragments *sie ergriff erneut* ohne die zum Plosiv [g] gehörende Stille.

Tonbeispiele:
Ton *sie ergriff erneut,* Wiederholung der vollständigen Version.

Version ohne Stille bei [g].

Man kann interpretieren, dass die Stille an der Kodierung des Verschlusslauts tatsächlich nicht notwendig beteiligt ist, sondern vielleicht nur ein Zeitraster ausfüllt. Die Wirkung einer Top-down-Aktivität zur Restituierung des fehlenden Ausschnitts kann nicht verantwortlich gemacht werden, sie würde in der zur Verfügung stehenden Zeit nicht unterkommen. Eine

andere mögliche Interpretation wäre, dass man das der Stille entsprechende phonologische Merkmal als eines unter mehreren sieht, die ausgewertet werden, und dass man annimmt, dass nur eine Auswahl, also nicht alle Merkmale zur Identifikation des Plosivs erforderlich sind. Das wäre eine prototypentheoretische Sicht, wie sie in Teil 2, „Grundlagen", Kapitel 2.4, für die lexikalische Semantik und im folgenden Abschnitt 3.2.2 ergänzend für die Phonologie diskutiert wird. Die prototypische Funktion der Kategorisierung kann aber angesichts der Variabilität möglicher Impulsbursts nur unter der Annahme einer Kodierungsform mit Einzelimpulsen funktionieren.

Insgesamt dürfte also die Schlussfolgerung kaum abweisbar sein, dass ein die Stille abbildender Impulsburst schwerlich für die lautliche Identifikation von Plosiven erforderlich sein kann. Stimmhafte Plosive haben neuronal also auf Phonemebene keine vom akustischen Signal abhängige Dauer, jedenfalls nicht, solange es sich nicht um konsonantische Längen handelt.

Ein anderes Problem, das bei dem Versuch entsteht, eine Phonemdauer in einem Impulsburst abzubilden, entsteht bei Diphthongen. Die klassische strukturalistische Diskussion unterscheidet zwischen monophonematischer und biphonematischer Wertung von Diphthongen. Es ist aber prinzipiell auch möglich, mehr als zwei Bestandteile zu sehen, und das kann dann als Argument generell gegen eine Segmentierung verwendet werden (Beispiel bei Neppert, 1999: 245: „Wieviele Laute (mit Phonemwert) stecken in einem Diphthong? Ein, zwei oder unendlich viele?"). Wenn man die Wahrnehmung und Kategorisierung von Diphthongen im Verstehensprozess betrachtet, kann man sich fragen, zu welchem Zeitpunkt ein spezifischer Diphthong als solcher identifizierbar ist. Das ist sicherlich nicht möglich, solange nur der erste Bestandteil (unter der Voraussetzung, dass man von zwei Bestandteilen ausgehen darf) verarbeitet wird, sondern erst, nachdem eine bestimmte Strecke des zweiten Bestandteils wahrgenommen worden ist. Man kann im Beispiel der Diphthonge /ai/ und /au/ im Deutschen vielleicht annehmen, dass Impulsbursts während des ersten Bestandteils mehrdeutig für /a/, /ai/ und /au/ entstehen und der zweite Bestandteil diese Alternative entsprechend einengt. Was soll aber während des Übergangs zwischen den Bestandteilen passieren und wie ist die zeitliche Erstreckung der Übergangsphase? Andererseits ist es doch so, dass zu bestimmten Zeitpunkten während der Wahrnehmung eines Diphthongs Eigenschaftskonstellationen vorliegen, die denen von Monophthongen ähneln. Diese Feststellung ist allerdings so lange nicht hilfreich, als man keinen Anhaltspunkt hat für die Festlegung der Positionen dieser Zeitpunkte. Wenn man annimmt, dass Phoneme durch einzelne Impulse auf spezialisierten Zellen kategorisiert werden, kann man ein in gewissen Grenzen variables Zeitraster gewinnen, das sich

in den Diphthong hinein fortsetzt und somit die gewünschten Zeitpunkte liefert. Die Konsequenz ist, dass der Diphthong in minimal zwei und bei entsprechender Dehnung auch in mehr Bestandteile zerlegt wird, wobei die Anzahl der Bestandteile durch entsprechende Inputeigenschaften festgelegt wird und immer überschaubar bleibt.

Die Abbildung 3.2.2–3 zeigt abschließend einen Versuch, der Äußerung von Abbildung 3.2.2–1 die den Phonemen entsprechenden Aktionspotenziale, die in den darauf spezialisierten Großmutterzellen zu erwarten sind, zuzuordnen. Ein ähnlicher Versuch findet sich in Teil 4, „Lexikon, Morphologie“, Kapitel 4.3.1. Es ist zu beachten, dass Längen zu zwei oder mehreren Aktionspotenzialen führen müssen (Längen bei Plosiven führen zunächst zu einer mehrdeutigen Reaktion, die Qualität des Plosivs wird auch hier erst durch die Übergangsphase festgelegt). Die Variabilität der Abstände zwischen den Aktionspotenzialen hält sich in den Grenzen, die für störungsfreie lexikalische Prozesse (vgl. wieder Teil 4) möglich sind.

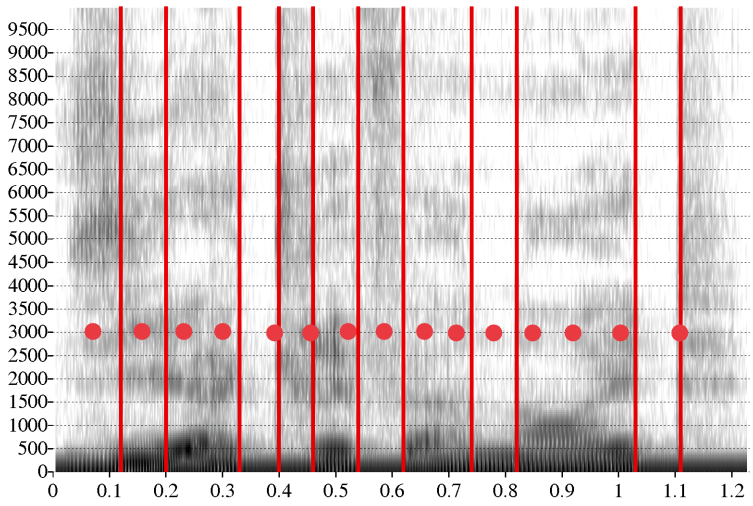



*Abbildung 3.2.2–2:* Versuch einer Zuordnung von neuronalen Impulsen (rote Punkte) auf Phonemebene zu phonetischen Segmenten in der Äußerung von Abbildung 3.2.2–3.

### 3.2.3 Natürliche Klassen, Prototypizität

Sofern Phoneme nicht universell sind, müssen sie komponentiell sein, das heißt, es ist mit Lernvorgängen zu rechnen, die angeborene Komponenten (letztlich Komponenten der Sinneswahrnehmung) zu Konzepten zusammenschließen. Das gilt natürlicherweise zunächst für die Perzeption und führt zu zwei wichtigen Konsequenzen:

- Hierarchische Lernprozesse können zu Phonemklassen führen, das heißt, es können auch phonematische Einheiten entstehen, die in herkömmlicher Sicht nicht einzelnen Phonemen, sondern Phonemklassen („natürlichen Klassen“) entsprechen.
- Wie bei allen Konzeptbildungsprozessen durch Koinzidenz muss mit Prototypizität gerechnet werden, das heißt mit einer „Unschärfe“ in einem für Prototypizität charakteristischen Sinn.

*Natürliche Klassen*

Eine gängige Definition für natürliche Klassen lautet (hier nach Hall, 2000, 122):

> „Zwei oder mehr Laute bilden eine natürliche Klasse nur dann, wenn weniger Merkmale gebraucht werden, um diese Klasse zu spezifizieren, als ein einzelner Laut hat, der zu dieser Klasse gehört.“

Wenn man annehmen möchte, dass den Phonemklassen eine neuronale Realität entspricht, hat das keine besondere Schwierigkeit: Man kann Großmuttereinheiten annehmen, die auf dieselbe Weise zustandekommen wie die Großmuttereinheiten von Einzelphonemen, nämlich durch Lernvorgänge an Zellen, die als Koinzidenzdetektoren arbeiten und also neuronale Inputs innerhalb eines Zeitfensters verknüpfen. Die Verknüpfung ist, logisch gesehen, eine UND-Verknüpfung. Wenn Phonemrepräsentationen durch ein Bündel von quasi gleichzeitig anstehenden Merkmalen gebildet werden, können auch Phonemklassenrepräsentationen durch eine Teilmenge solcher Merkmale entstehen, vorausgesetzt, dass eine solche Teilmenge ausreicht und entsprechend häufig aktiviert wird, um eine potenzielle Großmutterzelle zum Feuern zu bringen. Phonemklassen sind durch Lernvorgänge gebildete Konzepte wie Phoneme und andere Konzepte auch. Unter Beachtung dieser Möglichkeit ergibt sich, ausgehend von dem Phonem /b/ die in Tabelle 3.2.3–1 wiedergegebene Zusammenstellung möglicher (vollständig) und unmöglicher (Beispiele) Konzepte bzw. Klassen.

| Mögliche Konzepte | Merkmalskombination | Nicht mögliche Konzepte (Beispiele) | Merkmalskombination |
|---|---|---|---|
| b | plosiv+bilabial+sth | b d | plosiv+bilabial/alveolar+sth |
| b p | plosiv+bilabial | p b g k | plosiv+bilabial/velar |
| b d g | plosiv+sth | b d | plosiv+bilabial/alveolar+sth |
| b d g p t k | plosiv | usw. | |

*Tabelle 3.2.3–1:* Natürliche Klassen, an denen der Konsonant [b] beteiligt ist, nach Maßgabe möglicher Konzeptbildungsprozesse. Die definierenden Merkmale sind der Einfachheit halber in artikulatorisch motivierter Notation gegeben, anstelle der eigentlich gemeinten auditiven Eigenschaften. Im Vergleich dazu eine Auswahl von Klassenbildungen, die als nicht-natürlich gelten müssen.

Es fällt auf, dass die nicht möglichen Konzepte durch Merkmalskombinationen charakterisiert sind, die Alternativen enthalten. Die „Klasse“ /b d/ enthält einen bilabialen und einen alveolaren Laut, es ist also entweder das Merkmal *bilabial* oder das Merkmal *alveolar* innerhalb der Klasse zugelassen. Diese ODER-Verknüpfung kann nicht durch eine UND-Verknüpfung ersetzt werden. Ein Laut kann nicht gleichzeitig bilabial und alveolar sein. Damit kann ein Konzept, das die in Frage stehende Klasse repräsentieren könnte, nicht dadurch entstehen, dass gleichzeitig *bilabial* und *alveolar* als Merkmale vorliegen. Das wäre aber die Voraussetzung einer Konzeptbildung durch Koinzidenz.

Wenn man zusätzlich prüft, ob eine neuronale ODER-Verknüpfung in der gewünschten Form entstehen könnte, stößt man auf die Schwierigkeit, dass dabei eines und dasselbe der beiden Merkmale immer in kurzem zeitlichem Abstand vor dem anderen erscheinen müsste, was durch die Idee der natürlichen Klasse nicht vorausgesetzt wird (zu den entsprechenden Lernprozessen siehe Teil 2, „Grundlagen“, Abschnitt 2.4.7).

Die Beachtung neuronaler Lernprozesse ergibt damit eine klare Abgrenzung der möglichen natürlichen Klassen gegenüber Klassenbildungen, die keine neuronale Realität haben können. Es ergibt sich außerdem, allgemeiner, dass mit der Entstehung neuronaler Repräsentationen für natürliche Klassen tatsächlich zu rechnen ist und schließlich, dass natürliche Klassen prinzipiell sprachspezifisch, also nicht universell sind, mit Ausnahme von Klassen, die einzelnen phonetischen Merkmalen entsprechen.

*Prototypizität*

Kategorisierung bedeutet immer, dass nicht-invariante(!) Erscheinungen einer und derselben Kategorie zugeordnet werden. Das wird manchmal ver-

gessen. Merkmalsdetektoren reagieren grundsätzlich auf ein gewisses Reizspektrum. Die zunächst in der lexikalischen Semantik entwickelte Idee der Prototypizität betrifft aber nicht diese „einfache“ Unschärfe von Konzepten. Es wird zusätzlich beobachtet, dass das Bündel einfacherer Konzepte (semantischer Merkmale), das ein komplexeres Konzept definiert (unter der Voraussetzung, dass man die Existenz solcher Merkmale akzeptiert), als variabel erscheint. Taylor (1995: 226) sieht das auch für Phoneme, also z. B. /t/, als gegeben an:

> „Just as there are no criterial semantic features which unify all the meanings of *climb* and *over,* so there are no phonetic features which unify all the members of the /t/ phoneme and which jointly distinguish the /t/ phoneme from contrasting phoneme categories. English /t/ would normally be described as a voiceless aspirated pulmonic alveolar stop; yet some allophones of /t/ are voiced, some stop realizations are unaspirated, pulmonic air-stream mechanism is absent in the ejectives, dental and glottal articulations defeat the characterization alveolar, and not all members of /t/ are stops.“

Die hinter dieser Feststellung stehende Liste von insgesamt 14 /t/-Realisationen bei Taylor (1995: 224 f.) enthält, wie der Autor selbst feststellt, Beispiele aus verschiedenen regionalen Varietäten und verschiedenen Sprechstilen, was, wie er meint, nicht überbetont werden sollte. Natürlich entsteht aber doch die Frage, unter welchen Bedingungen überhaupt mit prototypentheoretisch interpretierbarer Variation zu rechnen ist und damit auch, welche der Beispiele in der Liste von Taylor wirklich ernst zu nehmen sind.

In Teil 2, „Grundlagen“, Kapitel 2.4, wird herausgearbeitet, wie, auf der Basis der Möglichkeiten neuronaler Strukturen, Prototypizität in der lexikalischen Semantik durch Lernprozesse entsteht. Übertragen auf die Phonologie gilt, dass Prototypizität nicht zustande kommt dadurch, dass notwendig unterschiedliche Ausprägungen eines Lauts im Input erscheinen, sondern dadurch, dass ein mehr oder weniger komplexes (ggf. durchaus auch variables) Bündel von Merkmalen dem Lernprozess zugrundeliegt und schließlich nicht alle Merkmale benötigt werden, um eine neuronale Großmuttereinheit, die die lautliche Kategorie repräsentiert, zum Feuern zu bringen. Häufiger in derselben Form erscheinende „Allophone“ sollten dagegen eigene Kategorien bilden, unabhängig davon, ob man sie traditionellerweise als Repräsentationen ein und desselben Phonems ansieht oder nicht.

Die „normale Charakterisierung“ des /t/ in dem Zitat aus Taylor (1995) dürfte wohl kaum der Realität der auditiven Wahrnehmung entsprechen, sondern ist eine linguistische Abstraktion. Wenn man sie trotzdem einmal

akzeptiert, sind z. B. [aspirated pulmonic alveolar stop], [voiceless pulmonic alveolar stop] oder [voiceless aspirated alveolar stop] mögliche prototypentheoretisch interpretierbare Varianten. Diese Varianten bilden allerdings kein Kontinuum, und die Grenzen der Kategorie /t/ sind nicht im buchstäblichen Sinne unscharf. Die einzelnen Ausprägungen sind nicht mehr oder weniger /t/-haft. Auch einzelne Merkmale sind als Komponenten von Phonemen, das heißt: auf dieser Ebene, nicht (mehr) gradiert (es gibt z. B. kein Kontinuum der Stimmhaftigkeit, vgl. dazu Taylor, 1995: 230 f.), ansonsten würde der Lernprozess, der zu phonologischen Kategorien führt, nicht funktionieren. Wie es zu entsprechenden Urteilen über bessere oder schlechtere Vertreter einer Kategorie kommt, wird in Teil 2, „Grundlagen“, Abschnitt 2.4.8, erklärt.

Wenn man einige klassische Positionen zurechtrückt, kann man also durchaus mit prototypischen Kategorien in der Phonologie rechnen.

### 3.2.4 Fazit

Die Überlegungen dieses Kapitels haben zu Ergebnissen geführt, die für das Verständnis der Phonologie von grundlegender Bedeutung sind. Es ist deshalb sinnvoll, sich zusammenfassend klar zu machen, welche Gesichtspunkte insgesamt und zusätzlich ergänzend für die vorgestellten Positionen maßgebend sind. Man vgl. dazu Teil 2, „Grundlagen“, vor allem Kapitel 2.2.

Man muss zunächst darauf hinweisen, dass es keinen Anlass gibt, die bisher in der Phonologie herrschenden Konzepte und Methoden aufzugeben, wenn man damit rechnen darf, dass Sprachverarbeitung ein symbolverarbeitender Prozess ist. Dasselbe gilt, solange man nicht behauptet, dass die Phonologie etwas mit den Leistungen des Gehirns zu tun hat oder zu tun haben sollte. Wenn man aber der Meinung ist, dass Sprache zu einem wesentlichen Teil im Gehirn stattfindet und wenn man dem Rechnung tragen möchte, wird relevant, dass der Kortex die für einen symbolverarbeitenden Prozess erforderlichen anatomischen Strukturen nicht aufweist. Es ist nicht möglich, Speichereinheiten und Prozessoren zu unterscheiden und die neuronalen Verbindungen haben nicht die Möglichkeit, kodierte Information weiterzuleiten, in einer Form, wie sie in elektronischen Systemen genutzt wird, und mit einer den äußerlich beobachtbaren Leistungen entsprechenden Geschwindigkeit. Eine Regelverwendung, wie in der generativen Phonologie vorausgesetzt, kann in einer neuronalen Architektur nicht realisiert sein. (Zur Problematik von Regeln in der Phonologie vgl. auch Abschnitt 3.1.3 und Kapitel 3.5.)

Speicherprozesse bzw. Lernprozesse bestehen in der Verstärkung vorhandener neuronaler Verbindungen und müssen dazu führen, dass „gleiche“ Umweltreize als solche erkannt werden. Was „gleich“ ist, bestimmen die durch Lernen entstandenen mentalen Konzepte. Für diese mentalen Konzepte gilt, dass sie durch Zellen repräsentiert werden, die jeweils auf ein Konzept spezialisiert sind. Eine „verteilte“ Repräsentation würde einen sog. „überwachten“ Lernprozess erfordern, der schon naiver Beobachtung widerspricht. Die Festlegung eines mentalen Konzepts (einer mentalen Kategorie) kann nun nicht dadurch geschehen, dass einfach neuronale Verbindungen, die an einem Verarbeitungsprozess beteiligt sind, verstärkt werden, sondern es muss eine Begrenzung dieses Vorgangs angenommen werden, so, dass z. B. erreicht wird, dass nur ein bestimmter, u. U. variabler Satz von Inputkomponenten in einem Wahrnehmungsprozess eine ihm zugeordnete Kategorie triggert. Aus der Notwendigkeit solcher Eigenschaften und den entsprechenden Lernbedingungen kann man ableiten, dass in Bereichen, in denen Lernvorgänge stattfinden, eine Information innerhalb eines Zeitfensters in einem einzelnen neuronalen Impuls auf einer spezifischen Bahn besteht. Wenn man annehmen muss, dass Phoneme in Lernprozesse eingebunden sind, ergibt sich die oben als „horizontale Einheit“ bezeichnete Eigenschaft von Phonemen als selbstverständliche Voraussetzung. Verzichtet man auf die Annahme der horizontalen Einheit, verliert man entsprechend die Möglichkeit, die erforderlichen Lernprozesse zu erklären.

Die horizontale Einheit des Phonems ist in mehrfacher Hinsicht Voraussetzung für Verarbeitungsprozesse oberhalb der Phonemebene. Das gilt zunächst für lexikalische Prozesse, die außer dem, was oben als „vertikale Einheit“ bezeichnet ist, auch die horizontale Einheit des Phonems voraussetzen. Man vgl. dazu allgemein Teil 4, „Lexikon, Morphologie“. Zusätzlich ist die horizontale Einheit auch die Voraussetzung für Kurzzeitgedächtnis-Phänomene und für eine Kohärenzüberwachung, die es ermöglicht, sowohl bei der Sprachperzeption als auch bei der Sprachproduktion Störungen zu erkennen und ggf. Reparaturprozesse unterschiedlicher Art auszulösen. Details werden in Teil 5, „Syntax“ beschrieben. Alle diese Funktionen sind darauf angewiesen, dass die Sprachwahrnehmung einen Impulszug liefert, der in bestimmten Grenzen relativ gleiche Abstände zwischen den einzelnen Impulsen aufweist und dass damit, sozusagen als Nebenprodukt, das Abgrenzungsproblem für Phoneme als zeitlich andauernde Segmente verschwindet. Zu welchem Zeitpunkt im Verstehen einer sprachlichen Äußerung sollte sonst Kohärenz oder Inkohärenz bescheinigt werden? Man beachte, dass Beobachtungen von Reparaturprozessen in der Produktion gesprochener Sprache darauf hindeuten, dass z. B. nicht immer das Wortende oder Silbengrenze als Zeitpunkt für eine Kohärenzkontrolle abgewartet wird. Man kann al-

so auch aus solchen Gründen der Weiterverarbeitung nicht damit rechnen, dass ein minimales phonologisches Segment durch Bursts, die aus mehreren Impulsen bestehen, begleitet wird.

Die Unterscheidung Kürze vs. Länge (z. B. Kurzvokal vs. Langvokal) verlagert sich in den Bereich lexikalischer Repräsentationen. Laute, die prinzipiell gelängt werden können (dazu gehören auch Konsonanten!), könnten phonetisch verschieden sein von Lauten, die im Spracherwerbsprozess nie als Längen erschienen sind. Die Distinktivität des Merkmals Länge wird aber erst durch das Vorhandensein entsprechender lexikalischer Ausdrucksrepräsentationen entschieden. Man überlege sich in diesem Zusammenhang, wie unter der Annahme der Dastellung lautlicher Segmente durch Impulsbursts das Merkmal Länge festgestellt werden könnte. Unter der Annahme der horizontalen Einheit von Phonemen ist das auf der lexikalischen Ebene relativ unproblematisch.

Die vertikale Einheit von Phonemen ist in dem Augenblick, wo man überhaupt Lernprozesse annimmt, die auf Koinzidenz von Signalen beruhen, unvermeidlich. Argumente, wie die oben angesprochenen Einfachheitsüberlegungen auf lexikalischer Ebene, entsprechen der linguistischen Tradition, sind aber letztlich doch eher zweitrangig. Ebenso unvermeidbar ist die Prototypizität der phonologischen Konzepte. Eine Ablehnung der vertikalen Einheit würde bedeuten, dass man prinzipiell das Lernen auf Grund von Koinzidenz als allgemeines Prinzip für den Kortex in Frage stellt. Wenn man umgekehrt mit Lernprozessen durch Koinzidenz rechnet, hat das auch zur Folge, dass man eine mehr oder weniger große Redundanz der mentalen Repräsentationen akzeptieren muss. Wenn man beachtet, dass offenbar lebenslang Neuronen im Kortex abgebaut werden, was möglicherweise zur Störung vorhandener Repräsentationen führen sollte, ist Redundanz eine notwendige und positiv zu bewertende Eigenschaft.

Die Annahme der Redundanzfreiheit sprachlicher Repräsentationen ist grundlegend für die linguistische Methodik in vielen Bereichen und gerade in der Phonologie. Die Möglichkeit oder gar Notwendigkeit redundanter Repräsentationen sollte zu einem grundsätzlichen Nachdenken über die Argumentationsstrukturen in dieser Disziplin führen.

# 3.3 Perzeption

## 3.3.1 Modellbildung als Forschungsinstrument in der auditiven Phonetik

Wenn hier von Modellen die Rede ist, sind nicht Tiermodelle gemeint (also die Verwendung von Versuchstieren zu Messzwecken), sondern symbolische Modelle nach Teil 1, „Grundlagen“, oder spezieller: Simulationen. Die Absicht dieses Abschnitts ist es, einige Gesichtspunkte zur Funktion solcher Modelle in Erinnerung zu rufen, da sie für die Bewertung des Vorgehens bei der Klärung einiger wesentlicher Fragestellungen der auditiven Phonetik von Bedeutung sind.

Der zur auditiven Phonetik rechnende Bereich der Sprachverarbeitung beginnt mit dem äußeren Ohr und endet, wenn man sich an die in Abschnitt 3.1.2 gegebenen Definitionen hält, bei der Bildung von Phonemrepräsentationen, die nicht mehr spezifisch für die Perzeption oder Produktion von Sprache sind. Wenn es um Untersuchungen zu diesem Bereich geht, sind die Möglichkeiten für Messungen am Menschen gerade in den zentraleren Teilen, die für die kortikale Linguistik interessant sind, eng begrenzt. Schon Beobachtungen, die das Innenohr betreffen, sind nur in beschränktem Umfang möglich, und die Funktionen der neuronalen Verbindungen zwischen Innenohr und Kortex, abgesehen von den anatomischen Details, die auch post mortem gewonnen werden können, bleiben, wenn man die Möglichkeit von Messungen betrachtet, weitgehend im Dunkeln. Das wirkt sich gerade im Sprachbereich besonders nachteilig aus, da Versuchstiere wie Affen und Katzen ja gerade keine Sprache besitzen und die Sprachevolution sich auch auf das Gehör erstreckt haben könnte.

Die bruchstückhaften Informationen über den Bereich der auditiven Verarbeitung bilden bestenfalls ein Puzzle mit sehr vielen fehlenden Teilen. Diese Situation ist ein typisches Anwendungsproblem für eine Form von Modell-

bildung, bei der es nicht nur um die Erzeugung von korrekten Zahlenwerten aus empirischen Untersuchungen geht, sondern die auch gültige Aussagen liefert über die Art und Weise, wie solche Zahlenwerte zustande kommen, also in unserem Fall z. B. über die zugrundeliegenden neuronalen Strukturen und deren Funktionen im Detail. Es geht auch nicht nur um eine Veranschaulichungsfunktion des Modells, und, um im Bild zu bleiben, das Puzzle kann nicht nach einer schon vorhandenen Vorlage zusammengesetzt werden, sondern die Modellentwicklung muss als ein eigener Forschungsprozess verstanden werden, der vielfach an ein Versuchs-Irrtums-Verfahren erinnert. Ziel ist es, eine funktionierende Struktur zu entwickeln (besonders wenn Simulationen eingesetzt werden). Fehlende Puzzleteile müssen so ergänzt werden, dass sich ein geschlossenes, widerspruchsfreies Ganzes ergibt.

Mit dem Mittel der Modellbildung kann also versucht werden, Informationen über nicht direkt beobachtbare und nicht messbare Details zu gewinnen. Insbesondere besteht in dem für die auditive Phonetik interessierenden Bereich die Möglichkeit, Neurophysiologie und Psychoakustik zusammenzubringen. Die von der Psychoakustik erhobenen behavioralen Daten werden durch entsprechende Verarbeitungsstrukturen verursacht, die ihrerseits neurophysiologischen Beschränkungen unterliegen. Obwohl in der Nähe der Sinnesperipherie auf neurophysiologischer Ebene oft von besonderen Bedingungen, spezifischen Zelltypen und Verschaltungen auszugehen ist, die nicht dem entsprechen, was man vom Kortex her sonst gewohnt ist, bleibt es doch möglich, einige allgemeine Grundprinzipien der neuronalen Verarbeitung auch in diesem Bereich als gültig anzunehmen. Wichtig ist auch, dass die Endpunkte der phonetischen Verarbeitung, besonders der Übergang zum lexikalischen Bereich, durch Überlegungen außerhalb der Phonetik bzw. Phonologie im engeren Sinne festgelegt sind und damit wertvolle Restriktionen für die Konstruktion von dazwischen liegenden Strukturen und Prozessen liefern.

Die konstruierten Modelle bleiben selbstverständlich hypothetisch in dem Sinne, dass sie weitergehender Diskussion unterliegen, sie haben aber nicht die mangelnde Verbindlichkeit, die, besonders in den Geisteswissenschaften, mit dem Modellbegriff häufig assoziiert wird. Brauchbare Modelle gelten, wie gute Hypothesen prinzipiell, als verletzlich und können nicht gerechtfertigt bzw. immunusiert werden mit dem Hinweis, dass es sich eben nur um Modelle handelt.

Schließlich sollte man bei aller Skepsis gegenüber der Leistung von Modellen beachten, dass sie als Hilfsmittel dienen können für die Entwicklung sinnvoller Fragestellungen für oft teure und risikobehaftete empirische Studien. Gerade im Bereich der peripherienahen auditiven Verarbeitung erinnern die

empirischen Techniken oft an das berüchtigte Suchen der Nadel im Heuhaufen – falls sie überhaupt mit Aussicht auf nennenswerten Erfolg unternommen werden können. Hier muss jeder Hinweis von besonderem Wert sein.

## 3.3.2 Neuronale Kodierung durch das Innenohr

Wahrend das Außenohr durch seine Form das Richtungshören begünstigt und im Mittelohr, u. a. durch den Mechanismus der Gehörknöchelchen, eine mechanische Aufbereitung des Schallsignals bewirkt wird, ist die Umkodierung des mechanischen Signals in eine neuronale Form eine Leistung der schneckenförmig gewundenen Kochlea. Die dort innerhalb des Corti-Organs vorhandenen vier Reihen von Haarzellen (drei Reihen „äußere“ und eine Reihe „innere“) sind für die eigentliche Sinneswahrnehmung zuständig, wobei mehr als 90% der afferenten Nervenbahnen Verbindungen mit den inneren Haarzellen haben, die in einer einzigen Reihe auf der Basilarmembran der Schnecke angeordnet sind. Die Abbildung 3.3.2–1 zeigt die Anordnung der Haarzellen und führt einige ergänzende Begriffe ein.

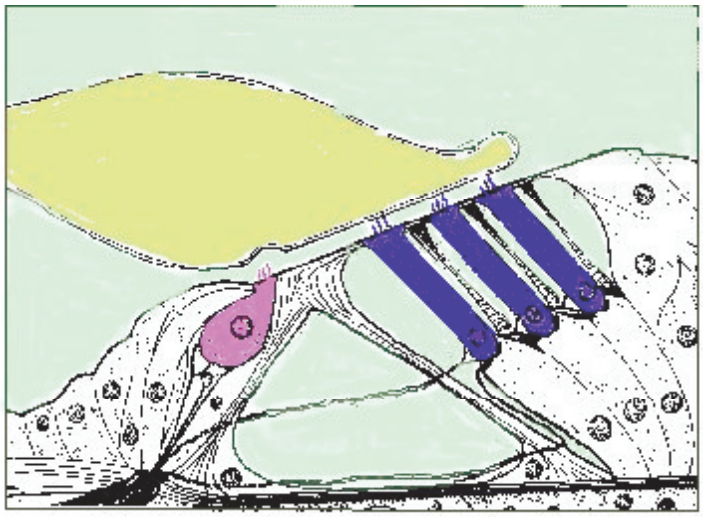


*Abbildung 3.3.2–1:* Struktur des Corti-Organs (Querschnitt). Rot: Innere Haarzelle. Blau: Äußere Haarzellen. Gelb: Tektorialmembran. Hellgrün: Lymphe.

Die bipolaren Nervenzellen des Ganglion spirale bilden Synapsen mit den inneren Haarzellen, bei Abscherung der Stereozilien (Haare) geben die Haarzellen einenTransmitter ab, der die Nervenzellen zum Feuern bringt. Da es ca. 3600 innere Haarzellen, aber ca. 30000 Fasern des Hörnerven gibt, kann man sich ausrechnen, dass pro Haarzelle ca. 8 Fasern innerviert werden. Hörnervenfasern sind regulär jeweils nur mit einer einzelnen inneren Haarzelle verknüpft.

Das Innenohr gewährleistet eine Frequenzanalyse des Schalls, deren Ergebnis in dem Erregungsmuster des Hörnerven gespiegelt ist. Die Fasern des Hörnerven sind jeweils spezialisiert auf eine bestimmte charakteristische Frequenz, auf die sie am besten (aber nicht ausschließlich) reagieren (Bestfrequenz). Wenn man die Reaktion von Hörnervenfasern auf Sinustöne unterschiedlicher Frequenz und unterschiedlicher Intensität erfasst, kann man sog. Tuning-Kurven (Abstimmkurven) erhalten, wie in Abbildung 3.3.2–2 dargestellt.

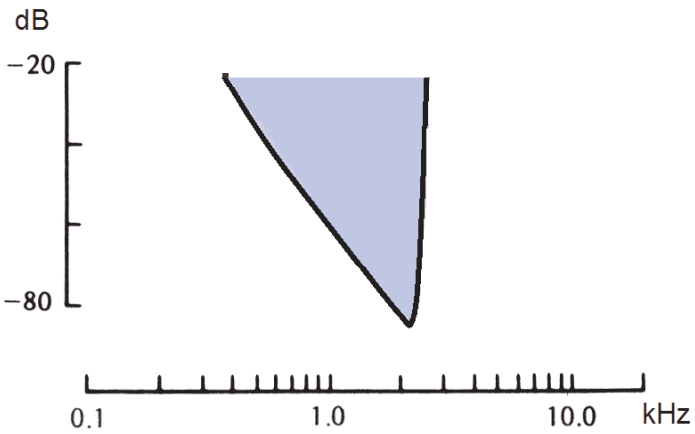



*Abbildung 3.3.2–2:* Tuning-Kurve einer Hörnervenfaser (Katze). Die Kurve gibt den Schallpegel wieder, bei dem die Hörnervenfaser gerade über die Spontanaktivität hinaus reagiert. Die farbige Fläche gibt also den Frequenzbereich wieder, auf den die Faser insgesamt spezialisiert ist. (Vereinfacht und ergänzt nach Kiang & Moxon, 1972: 725.)

Messungen an Hörnervenfasern zeigen, dass unabhängig von einem Schallinput „spontan“ Frequenzen bis 120 Aktionspotenzialen (bei 25% der Zellen unter 20 Aktionspotenzialen) pro Sekunde auftreten können (Zenner, 1994: 171). Die zeitliche Verteilung der Aktionspotenziale ist unregelmäßig.

Diese Spontanrate wird unmittelbar nach einem abrupten Abbruch des Schallsignals gedämpft. Tuningkurven der beschriebenen Art geben die Reaktion einer Hörnervenfaser abzüglich ihrer Spontanrate wieder.

Hörnervenfasern antworten also, über die Spontanrate hinaus, auch auf Sinustöne, die nicht ihrer charakteristischen Frequenz entsprechen, wenn der Schallpegel ausreichend hoch ist. Dabei gibt es eine Asymmetrie in dem Sinn, dass der erregte Bereich des Corti-Organs, in Anzahl der Haarzellen gerechnet, unterhalb der Haarzelle, deren charakteristische Frequenz dem Stimulus entspricht, breiter ist, als oberhalb. Es ist auch bemerkenswert, dass die Ausweitung des Frequenzbereichs nicht nur bei besonders lautem und nicht mehr sinnvoll wahrnehmbarem oder schädigendem Schall, sondern auch im normalen Sprachschallbereich zu beobachten ist.

Die gängige Erklärung für die Analyseleistung des Innenohrs ist die Wanderwellentheorie von Békésy (vgl. die Sammlung von Arbeiten in Békésy, 1960). Technische Details dazu sind in unserem Zusammenhang nicht von Relevanz. Die Wanderwellentheorie kann allerdings die tatsächliche Genauigkeit der Tuning-Kurven nicht erklären. Erforderlich ist nicht nur eine Verschärfung der Frequenzanalyse, sondern auch eine Verstärkung des Signals (durch eine Art „kochleären Verstärker“). Nach dem Beispiel des Sehens könnte man zur Erklärung auf das Konzept der lateralen Hemmung zurückgreifen. Die dazu erforderlichen Schaltungen sind aber in der Kochlea nicht vorhanden, und eine Verschärfung der Analyse ist, wie die Tuningkurve zeigt, schon auf dem Hörnerven und schließlich sogar durch Ableitungen aus einzelnen Haarzellen nachweisbar. Es müssen also die Umgebungsbedingungen der inneren Haarzellen zur Erklärung herangezogen werden. Da die Verstärkung und Verschärfung nur bei intakten äußeren Haarzellen möglich ist, liegt es nahe, die äußeren Haarzellen dafür verantwortlich zu machen. Es ist beobachtet worden, dass die äußeren Haarzellen sowohl unter der Innervierung über efferente Fasern als auch durch Schallreizung ihre Länge und damit die Lage der Tektorialmembran verändern können (sog. Motilität der äußeren Haarzellen). Sie haben damit eine steuernde Funktion für die Erregung der inneren Haarzellen (vgl. Zenner, 1994). Man beachte in diesem Zusammenhang die besondere Form und Lagerung der äußeren Haarzellen in Abbildung 3.3.2–1. Die technischen Details der Steuerungsfunktion sind allerdings noch nicht ausreichend geklärt.

Konform mit der Wanderwellentheorie ist die Beobachtung, dass es eine Latenzzeit für das Auftreten der neuronalen Reaktion auf einen Klick abhängig von dessen Frequenz gibt. Sie kann maximal ca. 3 ms erreichen (Stevens, 1998: 207).

Man kann bei Versuchstieren die Reaktion einer Hörnervenfaser auf Sprachschall messen und erhält dann sinnvolle, direkt mit Spektrogrammen, z. B. mit Vokalformanten in einem Breitbandspektrogramm, vergleichbare Reaktionen. Das heißt, eine mit ihrer charakteristischen Frequenz im Bereich der Frequenzen eines Vokalformanten liegende Hörnervenfaser wird den Formanten durch einen deutlich abgesetzten Impulsburst abbilden. Ein Beispiel aus Kiang & Moxon gibt die Abbildung 3.3.2–3.

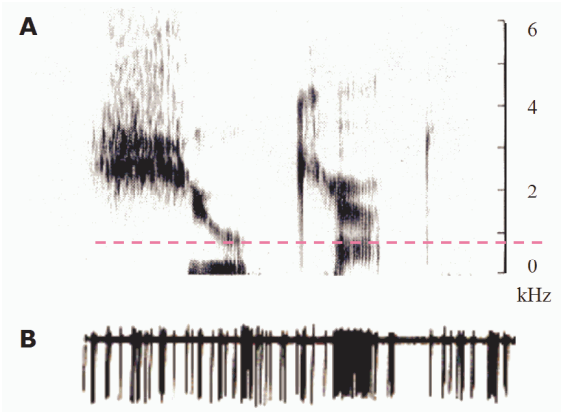



*Abbildung 3.3.2–3:* **B**: Antwort einer Hörnervenfaser mit einer charakteristischen Frequenz von 0,82 kHz auf den Input *shoo cat*, dessen Spektrogramm in **A** wiedergegeben ist. Gesamtdauer des Sprachsignals ca. 0,8 Sekunden, männlicher Sprecher. Die gestrichelte Linie im Spektrogramm deutet die charakteristische Frequenz der abgeleiteten Faser an. (Zusammengestellt und graphisch angeglichen aus den Abbildungen 6 und 7 bei Kiang & Moxon, 1972: 722 f.)

Zusätzlich zur Ortskodierung durch Auswahl bestimmter Hörnervenfasern ist auch mit der Wiedergabe der Frequenz und Stärke des Reizes innerhalb des Bursts zu rechnen. Das ergibt sich durch die Phasenkoppelung aufgrund der Funktionsweise der inneren Haarzellen. Die Phasenkoppelung gilt nach gängiger Meinung bis zu Stimuli von ca. 5 kHz, mit abnehmender Präzision. Zusätzlich ist mit einem Einfluss der Stimulusintensität zu rechnen. Außerdem sollte beachtet werden, dass schon allein aus der Tatsache, dass eine einzelne Faser nicht nur auf eine genau bestimmte Sinusfrequenz reagiert, sich eine gewisse Unregelmäßigkeit der Abfolge von Aktionspotenzialen auch bei einer solchen reizabhängigen Antwort, also nicht nur bei Spontanaktivität, ergibt.

Da die äußerste Peripherie bei allen Säugetieren ähnlich ist, sind Tierexperimente für das Verständnis der akustischen Sprachwahrnehmung in diesem äußersten Bereich durchaus relevant.

### 3.3.3 Die Hörbahn

Über die an den Hörnerven anschließenden Verarbeitungsstadien gibt es sehr wenig gesicherte Informationen, einmal abgesehen von den in Abbildung 3.3.3–1 aufgeführten groben anatomischen Gliederungen und vielen interpretationsbedürftigen Einzelbeobachtungen (vgl. Smith & Spirou, 2002).

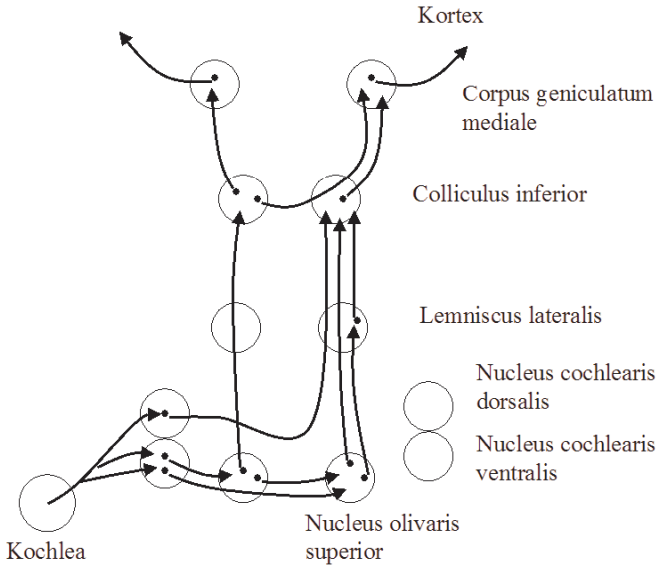



*Abbildung 3.3.3–1:* Schematische Darstellung der Hörbahn. Neuronale Verbindungen sind nur von einem Ohr ausgehend eingezeichnet. (Verändert nach Zenner, 1994: 199.)

Obwohl man auch noch im Kortex Zellen findet, die die tonotope Struktur des Ohrs spiegeln, ist es doch so, dass die Funktion der Hörbahn, auch für die Sprachverarbeitung, nicht einfach darin besteht, die vom Ohr gelieferten Informationen weiterzuleiten. Vielmehr muss man annehmen, dass wesentliche Verarbeitungsvorgänge stattfinden, die an einzelnen Zellen zu Reaktionen führen, die nicht mehr durch Sinustöne, sondern durch komplexe Schallmuster ausgelöst werden können. Schließlich muss mit Bezug auf

die Sprachverarbeitung auch gewährleistet werden, dass die Verarbeitung zu Repräsentationen führt, die koinzidenzbasierte Lernvorgänge ermöglichen. In Teil 2, „Grundlagen“, wird herausgearbeitet, dass das nur möglich ist, wenn aktivierte Informationen durch einzelne Impulse auf spezialisierten Zellen kodiert sind („Einzelimpulskodierung“).

Ausgangspunkt der Überlegungen zur Sprachverarbeitung müssen Repräsentationen sein, wie in 3.3.2–3 dargestellt. Die Frage ist jetzt, was auf dem Weg geschieht, der von solchen Repräsentationen ausgehend schließlich zu Kodierungen führt, die für die Weiterverarbeitung im Kortex geeignet sind. Man muss sich also um Anhaltspunkte zur Funktion der Hörbahn bemühen, die teilweise auch nichtsprachlich sind, um wenigstens Hinweise darauf zu bekommen, was man an Strukturen und Prozessen für die Sprachverarbeitung in diesem Bereich prinzipiell erwarten kann.

*Brainstem auditory evoked potentials*

Aus EEG-Ableitungen können BAEPs („brainstem auditory evoked potentials“) extrahiert werden, die den Verlauf einer durch einen auditiven Reiz ausgelösten neuronalen Antwort durch die einzelnen Stationen der Hörbahn direkt durch Spannungsgipfel erkennen lassen. Tabelle 3.3.3–1 ordnet den einzelnen Gipfeln jeweils die zeitliche Distanz zum Stimulus zu (Daten nach Ludman, 1998: 78).

| Ort auf der Hörbahn | BAEP-Potenzialgipfel | Zeitpunkt nach Stimulusonset |
|---|---|---|
| Hörnerv | I | 1,9ms |
| Nucleus cochlearis | II | 3,0ms |
| obere Olive | III | 4,1ms |
| Lemniscus lateralis | IV | 5,2ms |
| Colliculus inferior | V | 5,8ms 5,9ms |
| Corpus geniculatum mediale | VI | 7,6ms |
| auditorischer Kortex | VII | 9,2ms |

*Tabelle 3.3.3–1:* Elektrophysiologische Effekte der Verarbeitung auf der Hörbahn. Die einzelnen Potenzialgipfel werden üblicherweise mit römischen Zahlen angegeben.

Es ist zunächst dieses zeitliche Muster, das auch für die Sprachverarbeitung eine interessante Information darstellt. Man muss offenbar pro „Station“ mit einem Zeitbedarf von grob etwas mehr als einer Millisekunde rechnen, das entspricht, da ja doch nicht nur die zellinterne Reaktion sondern auch

die Laufzeit des Signals auf den Axonen und die Funktion des synaptischen Mechanismus einzurechnen sind, einer sehr schnellen Verarbeitung des akustischen Signals und führt zu Beschränkungen für Annahmen über die Komplexität der Verarbeitung auf der einzelnen Stufe. Wenn man diesen Zeitverlauf für die Sprachverarbeitung übernimmt, ist der Anteil der Verarbeitung auf der Hörbahn an der gesamten Verarbeitungszeit, die pro Phonem zur Verfügung steht, grob gerechnet 20%.

*Schalllokalisation*

Eine vergleichsweise intensiv untersuchte Leistung der Hörbahn ist die horizontale Ortung von Schallquellen im Raum. Das geschieht auf der Grundlage von interauraler Zeitdifferenz und interauraler Pegeldifferenz. Die Leistung der Schallortung wird für Säugetiere, also auch den Menschen, der oberen Olive zugeschrieben. Bei höheren Schallfrequenzen kann mit Pegeldifferenzen gerechnet werden, die dadurch entstehen, dass das der Schallquelle nicht zugewandte Ohr eine geringere Schallintensität registriert (Kopfschatteneffekt). Mit solchen Pegeldifferenzen ist bei niedrigeren Frequenzen nicht zu rechnen, die Schallortung ist in diesem Fall nur auf der Grundlage der Zeitdifferenz zwischen den Reaktionen beider Ohren auf die einzelnen Druckwellen möglich. Da die Zeitdifferenzen, um die es dabei geht, sehr klein sind, entsteht für die neuronale Verarbeitung ein Problem, das für das Verständnis neuronaler Prozesse von prinzipieller Bedeutung ist.

Die Auswertung der Zeitdifferenz setzt prinzipiell eine Phasenkoppelung des neuronalen Signals und Zeitverhältnisse voraus, wie sie in Abbildung 3.3.3–2 schematisch dargestellt sind.

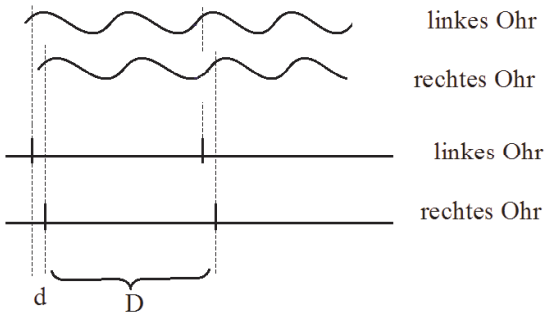



*Abbildung 3.3.3–2:* Schematische Darstellung der Phasenkoppelung und der Zeitverhältnisse bei neuronalen Signalen, die für die Raumortung von Schall niedrigerer Frequenz relevant sind. Weitere Erläuterungen im Text.

Wenn ein zuerst abgeleiteter neuronaler Impuls von einem später gebildeten unterschieden werden soll, muss prinzipiell die in Abbildung 3.3.3–2 mit $d$ bezeichnete Zeitspanne ausreichend klein sein gegenüber der mit $D$ bezeichneten.

Ein klassischer Versuch, die Schalllokalisation unter dieser Bedingung zu erklären, stammt von Jeffress (1948) (vgl. Gerstner et al., 1999). Die Grundidee ist, Signallaufzeiten auf unterschiedlich langen bzw. unterschiedlich schnellen Fasern in den Schallortungsprozess einzubeziehen. Die Schallortung kann dann durch eine Reihe von Zellen geschehen, die als Koinzidenzdetektoren arbeiten und entsprechend auf jeweils bestimmte Richtungen spezialisiert sind. Eine einzelne dieser Zellen reagiert mit einem Aktionspotenzial, wenn sie, aufgrund der unterschiedlichen Verzögerung auf den erregenden Fasern, ausreichend gleichzeitig einen Input von beiden Ohren bekommt. Die Abbildung 3.3.3–3 zeigt schematisch eine entsprechende Anordnung (ohne Aufteilung der richtungsspezifischen Neurone auf die rechte bzw. linke obere Olive).

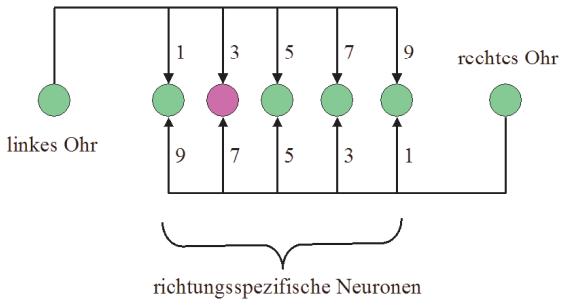



*Abbildung 3.3.3–3:* Schema zur Raumortung nach der klassischen Idee von Jeffress (1948). Die den Verbindungen zugeordneten Zahlen drücken die unterschiedlichen Laufzeiten von Aktionspotenzialen in Zeittakten (mit Bezug auf die im Folgenden verwendeten Simulationen) aus.

Das in Abbildung 3.3.3–3 gegebene Schema kann direkt in eine neuronale Struktur umgesetzt werden und liefert in der Simulation auch das erwünschte Ergebnis (Abbildung 3.3.3–4 **A**).

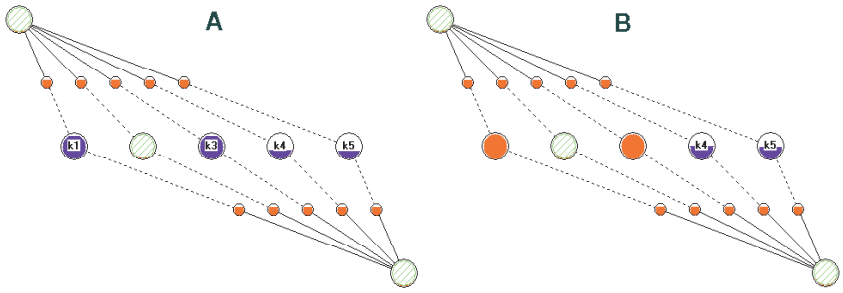



*Abbildung 3.3.3–4:* **A**: Zustand der Simulation des Schemas von Abbildung 3.3.3–3 in Zeittakt 12. Unter den Koinzidenzdetektoren hat eine einzelne Zelle ein Aktionspotenzial abgegeben und ist wie die Zellen, die den Input vom linken Ohr (oben) und vom rechten Ohr (unten) darstellen, in der Refraktärphase (grüne Schraffur). Alle anderen Zellen zeigen EPSPs unterschiedlicher Höhe (blaue Pegelstände in den Zellkreisen). Die kleinen Kreise symbolisieren die Synapseneffektivität. **B**: Derselbe Zeittakt in einer Simulation mit geringerer Abnahmerate (also längerer Dauer) des EPSP. (Aktuell feuernde Zellen sind rot gefärbt.) Die Schallortung ist unmöglich geworden.

Simulationen:
Schalllokalisation nach Jeffress (1948):
1. Funktionierende Version, mit kurzer EPSP-Dauer.
2. Version mit zu langsamer Abnahme des EPSP.

Die Schalllokalisation nach Jeffress erfordert ein feines Tuning von mehreren Parametern: Faserlänge, Effektivität der Synapsen, Dauer erregender postsynaptischer Potenziale (EPSPs). Die Abstimmung der Faserlängen ist relativ unproblematisch, bei Gerstner et al. (1999: 369) findet sich ein brauchbarer Vorschlag dazu. Schwierig wird es mit den EPSPs. Bei einer angenommenen Maximalfrequenz des Inputs von 2000 Impulsen pro Sekunde ist der große Abstand in Abbildung 3.3.3–2 0,5 ms, der kleine muss dann kleiner als 0,25 ms sein, die EPSP-Dauer dann noch einmal kleiner. Wenn die EPSP-Dauer zu lang wird, ergeben sich Störungen, wie in Abbildung 3.3.3–4 **B** und der zugehörigen Simulation demonstriert.

EPSP-Dauern von der erforderlichen Kürze sind aber neuronal unplausibel. Neuere Arbeiten zeigen nun, dass am Prozess der Raumortung auch rasche hemmende Verbindungen beteiligt sind, man vgl. dazu insgesamt Kapfer (2003) und die Überblicksdarstellung bei Yin (2002). Damit entschärft sich das Problem der zu langen EPSPs wesentlich. Das Prinzip kann am Beispiel der Abbildung 3.3.3–5 erläutert werden: Eine auf eine bestimmte Raumposition (einen bestimmten Winkel in der Horizontalen) spezialisierte Zelle

erhält nicht nur direkt einen erregenden Input von den beiden Ohren her, sondern auch mit einer beliebig kleinen Verzögerung einen hemmenden Input über zwischengeschaltete hemmende Neuronen. Die Hemmung reduziert das EPSP der integrierenden Zelle um ein geeignetes Maß und ersetzt damit die Annahme der unrealistisch kurzen EPSP-Dauer als Zelleigenschaft.

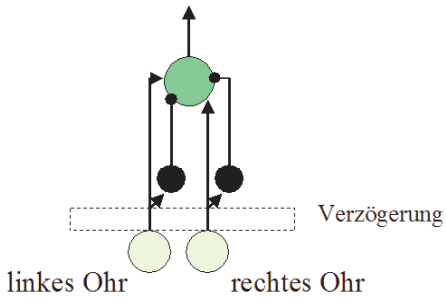



*Abbildung 3.3.3–5:* Schema zur Beteiligung der Hemmung an der Integration der von beiden Ohren her kommenden Impulse bei der Schalllokalisation.

Eine im Wesentlichen mit solchen Elementen arbeitende Simulation, aus der die Abbildung 3.3.3–6 einen Zustand zeigt, verhält sich den Erwartungen entsprechend. (Man beachte, dass letztlich das aus Gründen der Praktikabilität verwendete Zeitraster und die entsprechenden Zeitdifferenzen an der Realität gemessen noch zu grob sind. Das kann durch entsprechende Änderung der Definitionen problemlos korrigiert werden.)

Simulation:
Schalllokalisation mit Hemmung.
Der Zeittakt ist mit ca. 0,1 ms definiert.
Die ein Aktionspotenzial andeutende rote Färbung drückt in ihrer Dauer nicht die wirkliche Dauer des Aktionspotenzials aus. Refraktärphase und Aktionspotenzial zusammen sind hier 2 ms lang, ein unbeeinflusstes EPSP wäre erst nach mehreren Millisekunden auf einen vernachlässigbaren Wert abgesunken.

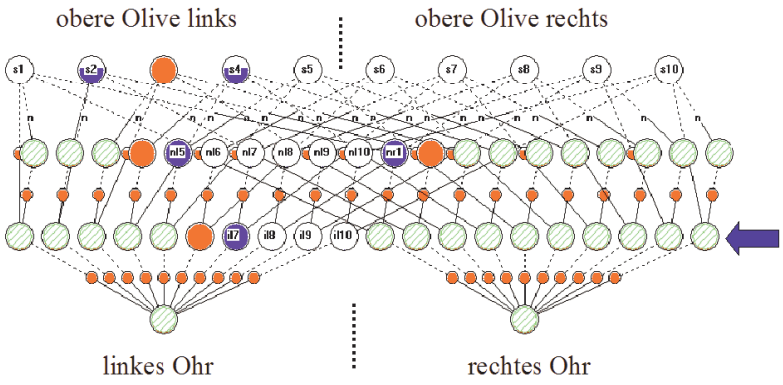



*Abbildung 3.3.3–6:* Simulation zur horizontalen Lokalisierung von Schall. Zustand in Zeittakt 16. Zur Erklärung der Symbolik vgl. die Abbildung 3.3.3–4. Hemmende Verbindungen sind mit n bezeichnet. Zur Bedeutung des Blockpfeils siehe den Text unten.

Der beschriebene Mechanismus gilt zunächst für die Auswertung von interauralen Zeitdifferenzen. Derselbe Mechanismus ist aber mit einer geringen Veränderung auch für die Auswertung von Pegeldifferenzen geeignet. Man muss dazu nur die Zellen in der in Abbildung 3.3.3–6 mit einem Blockpfeil gekennzeichneten Zellschicht mit einer zeitlichen Integrationsfunktion versehen. Wenn Pegeldifferenzen als Differenzen in der Frequenz der Aktionspotenziale erscheinen, wird sich die Differenz dadurch ausdrücken, dass die Zelle, die die höhere Rate erhält, entsprechend früher feuert. Damit ist die Pegeldifferenz in eine Zeitdifferenz umkodiert und kann nach dem für die Auswertung von Zeitdifferenzen angenommenen Schema bearbeitet werden.

Die Kodierung des Schallorts geschieht in jedem Fall durch ortsspezifische Neuronen. Eine durch dieses Modell nicht gestützte Alternative wäre die Kodierung durch eine bestimmte Frequenz der Aktionspotenziale bzw. ein bestimmtes Frequenzmuster. Man beachte aber, dass der Ort einer Schallquelle auch erinnert werden kann. Eine Frequenzkodierung dürfte damit ausgeschlossen sein. Die Beobachtung von Neuronen im Kortex von Affen, die auf unterschiedliche Orte mit unterschiedlicher Frequenz reagieren (Goldstein, 2002: 437), ist kein sicherer Hinweis auf einen Frequenzkode. Die Bemerkung bei Goldstein (2002: 439), dass „die Frage nach der neuronalen Codierung von Richtung und Abstand von Schallquellen weiterhin als noch nicht gelöst“ zu bezeichnen sei, ist zu vorsichtig.

Man beachte auch, dass die Abbildungen 3.3.3–3 bis 3.3.3–6 und die Simulationen für das linke und das rechte Ohr jeweils eine einzelne Zelle als Signalquelle vorsehen. Die Frage ist, ob man annehmen darf, dass der Input, der ja über die Haarzellen erfolgt, auf wenige oder gar nur eine einzige Zelle zusammengefasst wird. Dagegen spricht z. B., dass man Schallquellen unterschiedlicher Frequenz durchaus gleichzeitig unterschiedlichen räumlichen Positionen zuordnen kann und dass das auch zu entsprechenden Gedächtnisspuren führt. Die Konsequenz ist, dass man mehrere bis viele frequenzspezifische Anordnungen der beschriebenen Art annehmen muss.

Die Schalllokalisierung ist auch für Sprachschall eine wichtige Funktion. Sie führt zur Zuwendung zum Sprecher und damit zur Optimierung der Verständigung, vor allem unter Störbedingungen (Stichwort „Szenenanalyse“). Die neuronalen Mechanismen sind insofern interessant, als sie, wie unten noch zu zeigen sein wird, Analogien im Bereich der sprachlichen Kategorisierung auf der Hörbahn und schließlich auch im Kortex haben. Das betrifft vor allem die Funktion der Vorwärtshemmung zur Reduktion überschüssiger EPSPs.

*Verarbeitung von Sprachschall*

Die Abbildung 3.3.2–3 oben gibt die Reaktion einer einzelnen Nervenfaser des Hörnervs wieder. Bei Delgutte (1997: 511) findet sich ein Beispiel mit Ableitungen mehrerer Fasern, die in sinnvollen Zusammenhang mit dem zugehörigen Spektrogramm gebracht werden können. Wenn man Tuningkurven hinzunimmt, wird deutlich, dass es allerdings ein zusätzliches Problem gibt. Natürlich gibt es viele andere Fasern, die ebenfalls eine interessante bzw. „brauchbare“ Reaktion zeigen. Die Information wird mit zunehmender Schallintensität mehrdeutiger. Man steht also vor dem Problem, dass Sprache mit zunehmender Lautstärke schwerer verständlich sein sollte, was jeder Erfahrung widerspricht. Es ist zu beachten, dass es nicht um Lautstärken geht, die sich der Schmerzgrenze annähern, die dann Schwierigkeiten verursachen sollten, sondern durchaus um Lautstärken im normalen Sprachbereich.

Das Problem entsteht, wenn ausschließlich die Frequenz der Aktionspotenziale entsprechend spezialisierter Zellen zur Kodierung der Intensität in einem Frequenzbereich herangezogen wird. Man kann versuchen, dieses Problem dadurch zu bewältigen, dass man sich Mechanismen überlegt, die zu einer Verschärfung der Ortskodierung führen. Dafür kommen Varianten des Prinzips lateraler Hemmung in Frage, die aber offenbar nicht zu einer brauchbaren Lösung führen. Als vielversprechende Ansätze gelten solche, die nicht nur den Aktivationspegel der einzelnen frequenzspezifischen Hörner-

venfaser (Ortskodierung) auswerten, sondern sich zusätzliche Eigenschaften des Kochlea-Outputs zunutze machen. Dazu gehört vor allem die mehr oder weniger große Synchronizität der Aktionspotenziale auf verschiedenen, vor allem benachbarten Fasern, die durch die Phasenkoppelung der neuronalen Reaktion an das akustische Signal entsteht und, wie oben schon erwähnt, bei Schallen bis zu 5 kHz beobachtet wird. Es gibt verschiedene Vorschläge, die in diesem Sinne versuchen, den Effekt der Phasenkoppelung zusätzlich zur Ortskodierung auszuwerten (kurze Übersichtsdarstellungen bei Greenberg, 1996, und Delgutte, 1997).

Für die Auswertung der Synchronizität interessant ist, dass sie mit zunehmender Frequenz des Inputs abnimmt und mit zunehmender Schallintensität zunimmt. Zusätzlich ist zu beachten, dass der Bereich, in dem funktionell benachbarte Fasern als synchron betrachtet werden können, auch von der Entfernung der Haarzellen in der Kochlea abhängt. Das ergibt sich z. B. aus der oben in Abschnitt 3.3.2 schon erwähnten Beobachtung, dass die Reaktion der Kochlea frequenzabhängig um maximal 3 ms verzögert erscheint. Bei entsprechend enger Definition von „synchron“ ergeben sich entsprechend schmale Bündel synchron feuernder Hörnervenfasern.

Es ist an dieser Stelle interessant, auf einige weitere Aspekte speziell der Sprachverarbeitung hinzuweisen. Der Einbezug von Synchronizität setzt in jedem Fall die Auswertung mehrerer Fasern voraus. Die Frage ist, wie viele es sein sollen. Es ist offenbar nicht so, dass die Sprachverarbeitung auf eine besonders scharfe Analyse der Frequenzen des Sprachschalls angewiesen ist. Man beachte, dass ein Vokal nicht durch das Bündel der Mittenfrequenzen seiner Formanten synthetisiert werden kann. Gesetzt den Fall, man wäre in der Lage, die Reaktion der Hörnervenfasern durch ein entsprechendes zusätzliches Verfahren auf die Signalisierung ihrer Bestfrequenzen zu reduzieren, müßte man für die Sprachverarbeitung in einem zweiten Schritt wieder eine gewisse Unschärfe herstellen. Vielleicht ist es dann einfacher, von vornherein die Anzahl der auszuwertenden Fasern pro Analyseeinheit (was immer das sein mag) so anzunehmen, dass der Unschärfe von vornherein Rechnung getragen wird. Dabei können dann Frequenzbereiche, die nur spontane Aktivitäten oder Hintergrundrauschen durch unterschiedliche Verursacher spiegeln, aufgrund der mangelnden Synchronizität der Aktionspotenziale als quasi nicht-aktiv unterschieden werden von Frequenzbereichen (z. B. in Formantbreite), die als aktiviert zu gelten haben.

Das Problem der Auswertung von Synchronizität ähnelt dem Problem, das bei der Schalllokalisierung bewältigt werden muss, wo letztlich ja auch Synchronizität überprüft wird. Man kann sich also, wenn man eine Idee davon entwickeln möchte, am Beispiel der Schalllokalisierung orientieren. Das be-

deutet, dass man eine hemmende Funktion einführt, die den erwünschten Synchronizitätsgrad durch ein entsprechendes Zeitfenster definiert, so dass nicht extrem kurze EPSP-Dauern angenommen werden müssen, die ähnliches gewährleisten würden.

Eine mögliche Anordnung ist schematisch in Abbildung 3.3.3–7 wiedergegeben. Der Unterschied zu den Strukturen, die zur Schalllokalisierung dienen können, ist letztlich nur, dass nicht Inputs ausgewertet werden, die jeweils von beiden Ohren her kommen, und dass die resultierende Frequenz der Aktionspotenziale der integrierenden Zelle eine wesentliche Bedeutung für die Weiterverarbeitung behält.

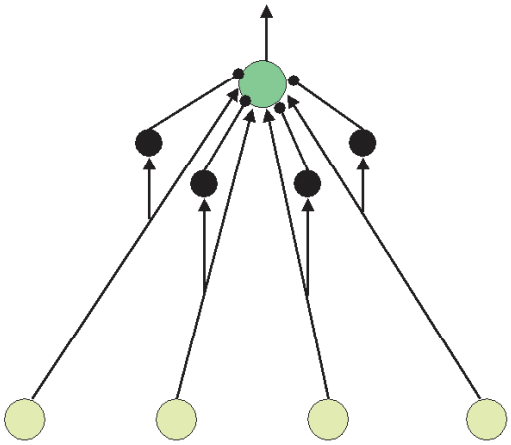


*Abbildung 3.3.3–7:* Schema einer Anordnung zur Auswertung von Synchronizität zusätzlich zur Frequenz bei prinzipiell gültiger Ortskodierung von Frequenzen auf verschiedenen Fasern der Hörbahn. Vgl. auch Abbildung 3.3.3–5.

Man kann das Grundprinzip und die Wirksamkeit in einer einfachen Simulation darstellen, die demonstriert, dass die Synchronizität tatsächlich auch unter Annahme nicht allzu exotischer Zellparameter zur Unterscheidung von Nutz- und Störschall dienen kann. In dieser Simulation sind zwei Zellbereiche (mit den Identifikationen *i1* bis *i10* und *i11* bis *i20*) gegenübergestellt, die jeweils, über alle Fasern gerechnet und an der Zahl der Aktionspotenziale gemessen, einen Input gleicher Intensität erhalten, sich aber in der Synchronizität der Impulszüge auf den einzelnen Fasern unterscheiden. (Die gleichmäßige Frequenz auf der einzelnen Faser in der Simulation ist unrealistisch, aber in diesem Zusammenhang nicht von Interesse.) Die Aktivationen der beiden Zellgruppen werden von zwei Zellen (*ome1* und *ome*

*2*) ausgewertet, die nach dem in Abbildung 3.3.3–7 dargestellten Prinzip verschaltet sind.

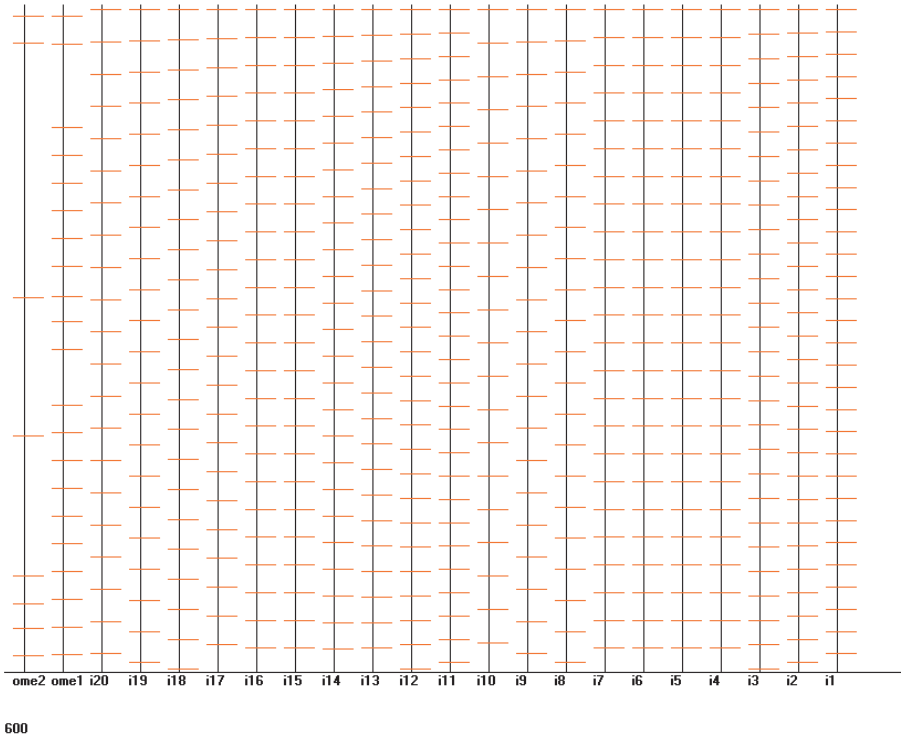



*Abbildung 3.3.3–8:* Simulationsergebnis zur Auswertung synchroner Aktivität. Die Aktionspotenziale der am unteren Rand identifizierten Zellen sind durch waagrechte (rote) Striche dargestellt, die Zeitachse verläuft von oben nach unten. Der gesamte wiedergegebene Bereich umfasst 600 Zeittakte.

Simulation:
Auswertung von Synchronizität.
Die Option „Start – Simulation bis Stop“ liefert das in Abbildung 3.3.3–8 wiedergegebene Ergebnis. Wenn man darüber hinaus mit der Leertaste einzelne Zeitzyklen berechnet, sieht man, dass die Häufung von Aktionspotenzialen bei *ome2* kurz vor Erreichen der 600-Takt-Grenze nicht andauert.

Wenn man den Zeittakt in der Simulation mit 0,1 ms veranschlagt, feuern die Einzelfasern mit einer Frequenz im Bereich von 300 bis 500 Hz. Das

Zeitfenster, innerhalb dessen Synchronizität gilt, ist 9 Zeittakte breit. Es genügen 4 synchrone Inputs, um ein Aktionspotenzial bei den integrierenden Zellen auszulösen. Unter diesen Bedingungen zählt man während 600 Zeittakten bei der Zelle *ome1*, die den stärker synchronisierten Input erhält, 21 Aktionspotenziale, bei der Zelle *ome2* 8 Aktionspotenziale. Der Kontrast kann erhöht werden, wenn man das Zeitfenster verkleinert oder eine größere Anzahl gleichzeitiger Inputs für die integrierenden Zellen verlangt.

Die Konsequenzen eines solchen Auswertungsprinzips für die Weiterverarbeitung von Sprachschall werden im folgenden Kapitel 3.3.4 diskutiert.

*Duplex perception*

„Duplex perception" meint, dass Bestandteile eines Lautspektrums unter bestimmten Umständen auch separat als nichtsprachliche akustische Ereignisse wahrgenommen werden. Diese Beobachtung der Parallelität von sprachlicher und nichtsprachlicher Wahrnehmung dient in der Tradition der „motor theory" (vgl. oben Kapitel 3.1.3) als Beleg für die Eigenständigkeit eines sprachlichen auditiven Verarbeitungsmoduls.

Wenn verschiedene Wahrnehmungskategorien prinzipiell durch verschiedene, auf eine jeweilige Kategorie spezialisierte Zellen repräsentiert sind, wird die Frage, ob man mit einem phonetischen Sprachmodul rechnen kann, das verschieden ist von Modulen mit anderer Aufgabe, zweitrangig. Unter der Voraussetzung, dass man abstrahiert von Fragen räumlicher Anordnung, evolutionsbiologischen Zwängen usw. und nur abhebt auf die zu gewährleistenden Funktionen, ist der Zusammenschluss sprachspezifischer Kategorien zu Modulen eine eher vernachlässigbare Eigenschaft. Außerdem kann man erwarten, dass ein Phänomen wie die duplex perception nicht prinzipiell sprachspezifisch sein muss, was in verschiedenen Versuchen (Überblick bei Goldinger, Pisoni & Luce, 1996: 287) auch gezeigt werden konnte.

Wenn man das Lautspektrum eines Vokals so auftrennt, dass tiefere Frequenzen auf einem Ohr, höhere Frequenzen auf dem anderen Ohr gehört werden, ist die Wahrnehmung identisch mit der eines nicht in dieser Weise zerlegten Schalls. In den folgenden Tonbeispielen ist der Langvokal [i:], ausgeschnitten aus dem Wort *bieten* so zerlegt, dass die Frequenzbestandteile unterhalb 1000 Hz dem linken und die oberhalb 1000 Hz dem rechten Ohr zugeordnet sind (Abbildung 3.3.3–9). Wenn man sich die beiden Teile separat anhört, ist der Unterschied sehr dramatisch: Der tiefere Bestandteil erscheint als Brummen, vielleicht eher als [u:], der obere als ein vielleicht entfernt [i:]-ähnliches Zwitschern. Die stereophonische Wiedergabe wird aber wieder als [i:] gehört, wobei der Laut im Raum eher rechts lokalisiert wird.

Der Lokalisierungseffekt ist erklärbar, wenn man beachtet, dass die höheren Frequenzen, die links fehlen, dort als „abschattiert“ gelten und die Pegeldifferenz zu einem entsprechenden Ergebnis führt (siehe oben zur Erklärung der Schalllokalisation bei höheren Frequenzen).

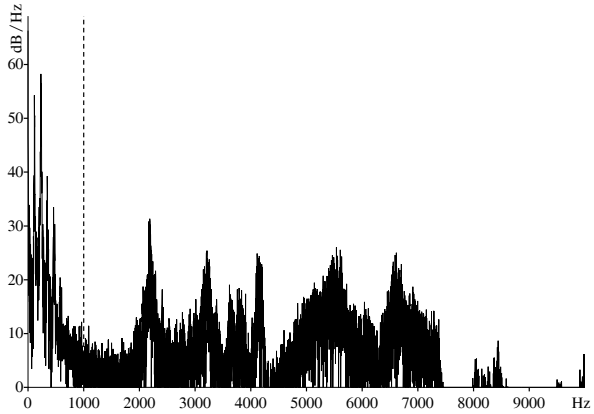



*Abbildung 3.3.3–9:* Spektrum des [i:] in den Tonbeispielen. Die in den Tonbeispielen verwendete Grenze linkes vs. rechtes Ohr ist gestrichelt eingetragen.

Tonbeispiele:
Segment [i:] in verschiedenen Versionen (Kopfhörer!):
1. Mono-Version über beide Ohren.
2. Tiefe Frequenzen links.
3. Hohe Frequenzen rechts.
4. Stereo-Version mit tiefen Frequenzen links und hohen Frequenzen rechts.

Man kann aus dieser Beobachtung zunächst folgern, dass sprachlicher Input über die beiden Ohren auf der Hörbahn (oder auch später) zu einer Einheit zusammengefasst wird. Das ist erst möglich oberhalb der oberen Olive einschließlich. Andererseits ist es nicht so, dass die beiden Bestandteile, oder auch nur das Zwitschern, zusätzlich wahrgenommen werden. Duplex perception findet also hier nicht statt. Die verwendeten Beispiele entsprechen aber in einigen Punkten nicht den klassischen Experimenten, dort werden vor allem synthetisierte Laute oder Silben verwendet. In der Version von Whalen & Lieberman (1987) sind es die Silben *ba* und *ga*, die durch drei Formanten synthetisiert und über beide Ohren gleichermaßen hörbar gemacht werden.

Es wird also hier nicht, wie in früheren Versionen, mit einem dichotischen Design experimentiert. Der Unterschied zwischen den Silben beschränkt sich auf den Übergang zwischen Konsonant und Vokal im dritten Formanten. In Isolation erscheint dieser Übergang als „chirp". Zusammen mit dem Rest der Formanten wird er nur hörbar, wenn er einen ausreichenden Pegel („duplexity threshold") hat. Die weitergehenden Schlussfolgerungen werden dann so formuliert:

> „... the results ... support the hypothesis that the phonetic mode takes precedence in processing the transitions, using them for its special linguistic purposes until, having appropriated its share, it passes on the remainder to be perceived by the nonspeech system as auditory whistles." (Whalen & Liberman, 1987: 171)

Es soll also nicht nur gezeigt werden, dass ein separates phonetisches Modul angenommen werden nuss, sondern dass es auch den Vorrang in der Verarbeitung hat und nur sozusagen überschüssige Energie an andere Prozesse weitergibt. Schließlich ergibt sich bei dieser Vorstellung auch, dass eine allgemeine nicht-sprachspezifische Schallverarbeitung, die der Sprachverarbeitung vorausgeht und ihre Ergebnisse an die Sprachverarbeitung weitergibt, nicht angenommen werden kann, sondern umgekehrt, dass die Sprachverarbeitung die Priorität hat.

> „Indeed, as the experiments reported here show, it is the phonetic module that has priority, as if its processes occurred before, not after, those that yield the standard dimensions of auditory perception." (Whalen & Liberman, 1987: 169)

Die Erklärung, metaphorisch formuliert, durch Überfließen eines Topfs ist neuronal nicht zwingend. Wenn man nicht annehmen möchte, dass Schall generell das Sprachmodul durchläuft, müsste man mit einer Unterscheidung von Schalltypen vor Erreichen dieses Moduls rechnen. Es lohnt also, den Vorgang noch etwas genauer unter die Lupe zu nehmen.

Eine Annäherung an die Verwendung synthetischer Laute kann dadurch erzielt werden, dass man eine als ausreichend angesehene Anzahl von Formanten aus natürlichem Sprachschall isoliert. Für die Kategorisierung von Vokalen sind nach üblicher Vorstellung die beiden ersten Formanten ($F_1$ und $F_2$) ausreichend. Der in den folgenden Tonbeispielen dokumentierte Versuch besteht nun darin, aus dem schon oben verwendeten Tonbeispiel des [i:] die beiden Formanten $F_1$ und $F_2$ zu isolieren (so gut es geht). Das Ergebnis ist in dem Spektrum der Abbildung 3.3.3–10 dargestellt.

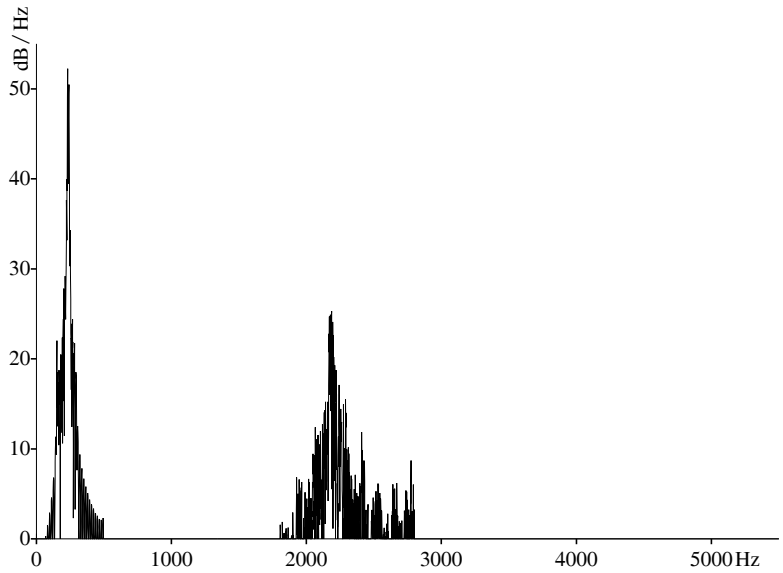



*Abbildung 3.3.3–10:* Spektrum eines durch entsprechende Filterung auf die Formanten $F_1$ und $F_2$ beschränkten [i:].

Wenn man sich die beiden Formanten getrennt über Kopfhörer anhört ($F_1$ in den Tonbeispielen links, $F_2$ rechts), entspricht das Ergebnis den Erwartungen: $F_1$ wird als Brummen, $F_2$ als Zwitschern wahrgenommen. Überraschend ist die Stereo-Version. Man hört relativ laut Brummen links und Zwitschern rechts, und nur wenig davon abgehoben einen Sprachlaut, der eher einem [y:] als einem [i:] entspricht. Das ändert sich nicht wesentlich in der Mono-Version mit $F_1$ und $F_2$ auf beiden Ohren.

Tonbeispiele:
Phon [i:], als Kombination von $F_1$ und $F_2$ (Kopfhörer!):
1. $F_1$ links.
2. $F_2$ rechts.
3. Stereo-Version mit $F_1$ links und $F_2$ rechts.
4. Mono-Version mit $F_1$ und $F_2$.

Die Formanten sind also stärker als die Kombination im Sprachlaut. Das kann doch nur heißen, dass es nicht so ist, dass zunächst der Sprachlaut bedient wird und die nichtsprachlichen Anteile mit einem Rest an Energie

versorgt werden. Da die Formanten aus einem unauffälligen Lautspektrum herausgeschnitten sind, kann man nicht annehmen, dass sie über das für das Sprachmodul erforderliche Maß hinaus verstärkt worden sind. Die Annahme des duplexity threshold, den man übersteigen muss, um duplex perzeption zu erreichen, ist also fragwürdig.

Wenn man für die auditive Verarbeitung generell und nicht nur für die Sprachverarbeitung spezialisierte Strukturen annimmt, kann man so argumentieren: Die aus dem Lautspektrum herausgeschnittenen Formanten sind nur bis zu einem gewissen Grad sprachlich, so dass sprachspezifische Verarbeitungsstrukturen nur gering ansprechen. Sie sind für sich genommen eher nichtsprachlich. Wenn Formanten über ein bestimmtes Maß hinaus verstärkt werden, reichen sie ebenfalls in den nicht-sprachlichen Bereich hinein. Wenn Frequenzausschnitte verwendet werden, die ausreichend reichhaltig sind, um als sprachlich gelten zu können, sind sie nicht oder nur in sehr geringem Maß durch nichtsprachliche Spezialisten interpretierbar, siehe das Stereo-Beispiel mit der Zweiteilung des gesamten Spektrums.

*Ergänzende Bemerkungen*

Einige sprachliche Phänomene, wie z. B. der Satzakzent, gehören vermutlich auf der Ebene der Hörbahn nicht zu den sprachspezifischen Eigenschaften, sondern werden erst im Kortex mit Ergebnissen sprachspezifischer Prozesse verknüpft.

Lautheit ist ebenfalls eine Eigenschaft, die vermutlich parallel und außerhalb des Bereichs sprachspezifischer Kategorisierungsprozesse verarbeitet wird. Daher können die zahlreichen diesbezüglichen psychoakustischen Befunde nicht ohne Probleme auf die Sprachverarbeitung bezogen werden.

Selbstverständlich gilt, dass im Kortex auditive Repräsentationen gefunden werden, die für primitiver gehalten werden, als die Ergebnisse sprachverarbeitender Prozesse. Daraus ist nicht zu schließen, dass die Sprachverarbeitung ausgehend von solchen primitiveren Repräsentationen direkt und ausschließlich im Kortex stattfindet.

### 3.3.4 Phonemdefinierende Merkmale

*Grundlagen kategorialer Wahrnehmung*

Die Wahrnehmung von Sprachlauten ist, wie psycholinguistische Experimente zur Identifikation von Lauten in Lautkontinuen zeigen, „kategorial“. Kategoriale Wahrnehmung meint, dass innerhalb eines bestimmten Variationsbereichs ein und derselbe Laut wahrgenommen wird, es gibt aber eine mehr oder weniger scharf ausgeprägte Grenze gegenüber Variationsbereichen anderer Laute. Die Abgrenzungen sind einzelsprachlich verschieden. Solche Beobachtungen passen gut zu den Annahmen der Großmutterzellentheorie und entsprechen damit den Erwartungen, die insgesamt auch für andere Bereiche der neuronalen Verarbeitung gelten.

Man könnte sich nun vorstellen, dass eine Großmutterzelle, die für einen bestimmten Sprachlaut spezifisch ist, jeweils das komplette, zu einem bestimmten Zeitpunkt anstehende Muster der Erregungen auf allen Hörnervenfasern (oder späteren Verarbeitungsstufen) in einem einfachen integrierenden Prozess auswertet, entsprechend gewichtete Effektivitäten der einzelnen Synapsen vorausgesetzt. Es gibt viele Argumente, die gegen eine solche Vorstellung sprechen. Die wichtigsten sind:

- Zu viele alternative Erregungsmuster würden zu demselben Ergebnis führen.
- Da die gültigen Muster, wie schon erwähnt, einzelsprachlich spezifisch sind, müssten entsprechende Lernprozesse gefunden werden, die die erforderliche genaue Abstimmung der Synapseneffektivität in der vorauszusetzenden Streuung zu gewährleisten hätten.
- Die Lautheit läßt sich nicht als irrelevant ausfiltern. Das Mittel der Auswertung von Synchronizität steht wegen der zeitlichen Spreizung der Reaktion der Sinneszellen über die gesamte Kochlea hinweg nicht zur Verfügung.
- Wenn mehrere unterschiedliche sprachliche Inputs gleichzeitig erfolgen, ist eine Analyse nicht möglich, da in diesem Fall ja kein Spezialist für das Erregungsmuster als Ganzes vorliegt.

Der letztere Punkt ist besonders interessant. Wenn man experimentell Formanten verschiedener Vokale mischt, hört man die einzelnen Vokale, nicht etwa ein Störgeräusch oder einen aus irgendeinem Grund dominanten Vokal. Man ist auch nicht unsicher, um welche Vokale es sich handelt. Das läßt sich zum Beispiel zeigen, wenn man zu den beiden Formanten $F_1$ und $F_2$

des [i:], aus dem Spektrum ausgeschnitten wie in Abbildung 3.3.3–10, die Formanten $F_1$ und $F_2$ eines [ɑ:] hinzufügt, so dass sie die Lücke zwischen den [i:]-Formanten ausfüllen (Abbildung 3.3.4–1).

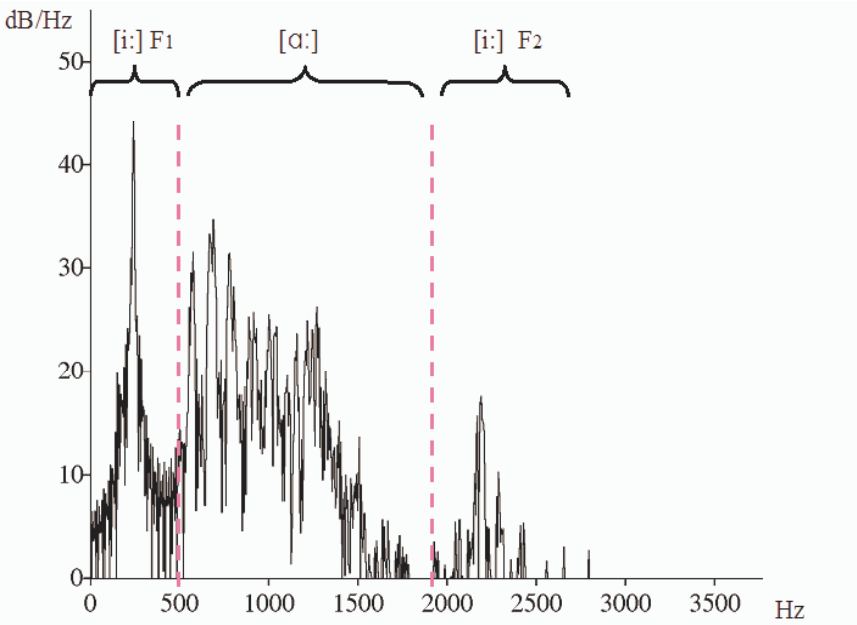



*Abbildung 3.3.4–1:* Spektrum der Formanten $F_1$ und $F_2$ eines [i:], zwischen die die Formanten $F_1$ und $F_2$ eines [ɑ:] eingefügt sind.

Der Ton zeigt, dass tatsächlich beide Vokale zu hören sind, allerdings, wie oben (Abschnitt 3.3.3) im Beispiel zur duplex perception, anstelle des [i:] wieder eher ein [y:].

Tonbeispiele:
Mischung der Formanten von [i:] und [ɑ:] (Kopfhörer!):
1. Mono-Version über beide Ohren.
2. Stereo-Version, [i:] links, [ɑ:] rechts.

Die Konsequenzen aus diesem Versuch reichen aber weiter. Man kann schließen, dass das [i:] bzw. [y:] zustandegebracht wird durch die beiden Frequenzbänder $F_1$ und $F_2$ und dass der dazwischenliegende Frequenzbereich dabei nicht beteiligt ist und verschieden ausgefüllt werden kann. Das ist

nur möglich, wenn die Formanten durch separate integrierende Zellen oder Zellverbände abgeprüft werden. Der Gesamteindruck entsteht durch Integration auf einer zweiten Stufe. Diese Beobachtung ist für das Verständnis der auditiven Phonetik von grundlegender Bedeutung, denn sie führt zu der Schlussfolgerung (zunächst für Vokale), dass phonetische Merkmale in ihrer elementarsten Form direkte Auswertungsprodukte von Formanten sein müssen und nicht etwa Auswertungsprodukte von Konstellationen aus mehreren Formanten.

Das Prinzip ist problemlos auf nicht-vokalische Laute ausdehnbar. Wenn der Unterschied zwischen [dɑ] und [gɑ] von einer Eigenschaft des dritten Formanten abhängt, kann er durch einen entsprechenden Detektor festgestellt werden. Man beachte, dass die Hörbahn Zellen enthält, die verschieden „schnell“ sind, so dass auch kurzzeitige Phänomene bearbeitet werden können. Die Abhängigkeit von den Eigenschaften eines Folgevokals kann, wenn man keine andere Lösung sieht, auch so bewältigt werden, dass man entsprechend viele verschiedene Detektoren annimmt, die alternativ aktiviert werden.

Schwieriger dürfte es sein, lautspezifische *Sequenzen* von Erregungen als Merkmale zuzulassen, wenn man solche Sequenzen als sehr kurzzeitige Ereignisse direkt von entsprechenden Erregungsmustern auf der Kochlea abgeleitet denkt. Abfolgen von Erregungsmustern, die einzeln und für sich genommen zunächst durch Detektoren festgestellt und deren Sequenz in einem zweiten Schritt abgeprüft wird, sind auswertbar, es handelt sich dann aber um einen Verrechnungsprozess, der nicht elementare Merkmale stiftet, sondern elementare Merkmale verarbeitet sozusagen zu Merkmalen zweiter Stufe.

*Das Normalisierungsproblem*

Die Formanten, deutlich vor allem bei Vokalen, haben bei Frauenstimmen, Männerstimmen und bei Kindern jeweils andere Mittenfrequenzen. Das führt auf die Vorstellung, dass die Wahnehmung von Sprachlauten eine Normalisierung voraussetzt, ehe die eigentliche Kategorisierung erfolgt. Zur Lösung dieses Normalisierungsproblems gibt es bisher keine brauchbaren Vorschläge. (Hinweise dazu bei Halberstam & Raphael, 2004.) Die Abbildung 3.3.4–2 zeigt zur Demonstration das bisher schon verwendete Spektrum eines männlichen [i:] und zum Vergleich ein weibliches Gegenstück (erwachsene Sprecherin). Man kann sehen, dass tatsächlich der zweite Formant bei der weiblichen Stimme höher liegt als bei der männlichen. Die Differenz ist allerdings, wenn man das Gesamtbild betrachtet, nicht besonders beeindruckend, so dass man aufgrund dieser Beobachtung allein schon

Zweifel haben kann, ob eine nennenswerte Normalisierung zum Ausgleich dieser Differenz wirklich erforderlich ist.

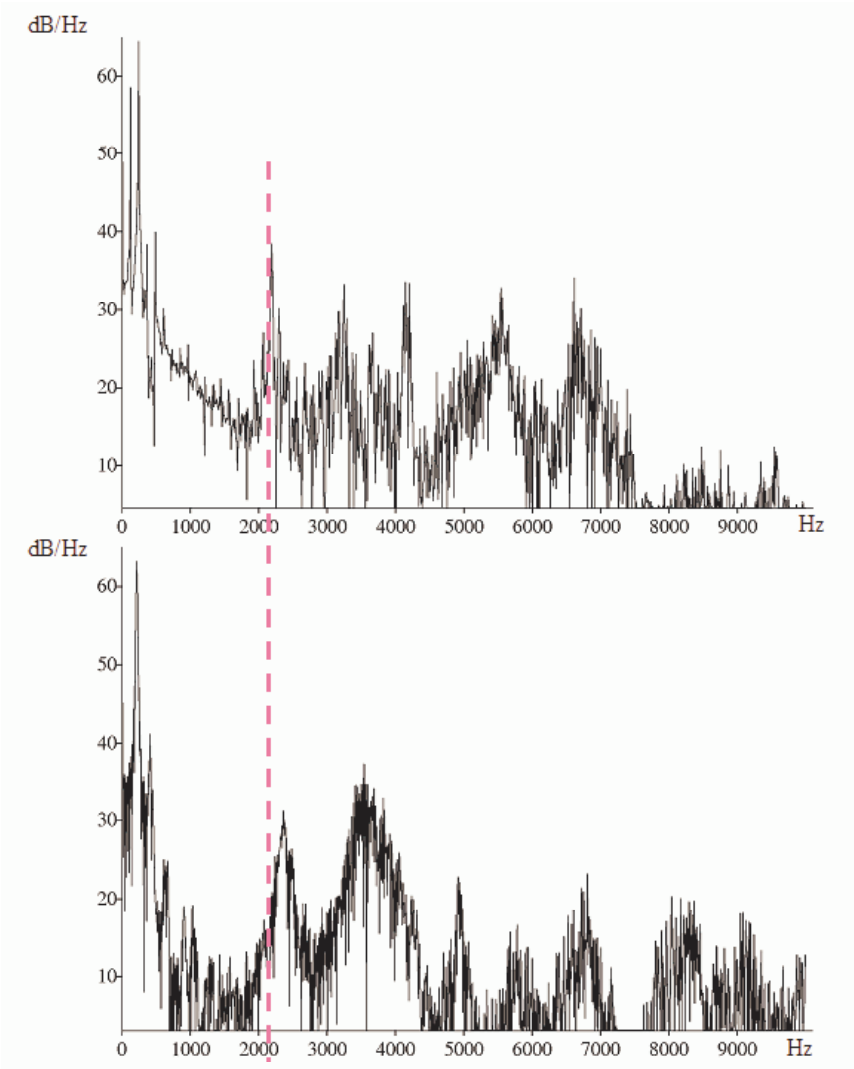



*Abbildung 3.3.4–2:* Spektrum eines [i:]. Oben: Männerstimme, unten Frauenstimme. Zur Erleichterung des Vergleichs ist die Position des zweiten Formanten der Männerstimme durch eine gestrichelte Linie angedeutet.

Ein das Normalisierungsproblem beleuchtendes interessantes Experiment besteht nun offenbar darin, die Formanten $F_1$ und $F_2$ jeweils aus Äußerungen verschiedener Sprecher auszuschneiden und zu kombinieren. Das sollte Schwierigkeiten bei der Normalisierung und damit auch bei der Kategorisierung eines Lauts ergeben. In dem folgenden Tonbeispiel sind die Frequenzen bis 999 Hz aus dem „weiblichen“ [i:] mit den Frequenzen ab 1000 Hz aus dem „männlichen“ [i:] kombiniert. Das Ergebnis ist überrraschenderweise ein klares, eher „männliches“ [i:], die erwarteten Normalisierungsschwierigkeiten treten also nicht auf. Die Normalisierung kann nicht aufgrund der Lage von $F_0$ oder $F_1$ erfolgen.

Tonbeispiele:
Segment [i:] in verschiedenen Versionen (Kopfhörer!):
1. Frauenstimme, Mono-Version über beide Ohren.
2. Tiefe Frequenzen, weiblich, links.
3. Hohe Frequenzen, männlich, rechts.
4. Zusammengesetzt, Mono-Version.
5. Stereo-Version mit tiefen Frequenzen links und hohen Frequenzen rechts.

Eine weitere Konsequenz aus diesem Experiment ist ebenfalls bemerkenswert. Der Abstand der Mittenfrequenzen von $F_1$ und $F_2$ ist nur noch 1900 Hz statt 2200 bei der reinen Frauenstimme. Es kann also doch wohl nicht die Berechnung dieses Abstands für die Lautwahrnehmung maßgebend sein. Das kann als zusätzliches Argument für die Annahme von Merkmalsdetektoren gelten, die auf Frequenzbänder einer bestimmten Breite spezialisiert sind, und zwar so, dass Abweichungen in einem bestimmten Rahmen durch einen Prototypizitätseffekt aufgefangen werden. Eine Normalisierung ist auch unter diesem Gesichtspunkt nicht unbedingt erforderlich. (Das [u:] ist im Tonbeispiel 2 deutlicher zu hören als im männlichen Gegenbeispiel oben in Abschnitt 3.3.3. Zur Begründung kann vielleicht darauf hingewiesen werden, dass der weibliche Formant weiter in den Bereich des zweiten [u:]-Formanten hinein reicht als der männliche.) Insgesamt erscheint damit das Normalisierungsproblem als mehr oder weniger offen. Eine weitergehende Klärung braucht zusätzliche Argumente und wird in Abschnitt 3.3.5 versucht.

*Ableitung einzelimpulskodierter Merkmale*

Für die Definition von Phonemkategorien, die Lernprozesse voraussetzen, ist es erforderlich, dass die definierenden Komponenten, also die entsprechenden spezialisierten Zellen, während der Dauer z. B. eines Kurzvokals in der Wahrnehmung gerade einen einzelnen Impuls abgeben. Das Ergebnis einer Merkmalsanalyse des Schallinputs liefert aber zunächst Merkmale,

die aus rascheren und auch mehr oder weniger unregelmäßigen Impulszügen bestehen, deren Frequenz bis zu einem gewissen Grad intensitätsabhängig ist (siehe oben 3.3.3 und Abbildung 3.3.3–8). Es ist ein zusätzlicher Verarbeitungsprozess anzunehmen, der den Übergang von dieser Repräsentation zu der für die Phonemkategorisierung erforderlichen Repräsentation leistet, also dafür sorgt, dass aus einem frequenzkodierten Input mit ausreichenden zeitlichen Abständen einzelne Aktionspotenziale ableitet. Das entsprechende Timing kann wegen der unterschiedlichen Sprechgeschwindigkeiten nicht ausschließlich durch einen zentralen Takt gewährleistet werden, sondern muss inputgesteuert sein.

Zusätzlich gilt, dass die ggf. anzunehmende spontane Aktivität eines Merkmalsauswerters bedeutungslos sein muss, eine bestimmte Intensitätsschwelle muss überschritten werden.

Wenn man das Prinzip der Einzelimpulskodierung so sieht, dass nach dem ersten Aktionspotenzial eine gewisse Zeitspanne verstreichen muss, ehe das zweite ausgelöst werden kann, erinnert das an die Funktion der absoluten neuronalen Refraktärphase. Man beachte, dass eine bloße Erschwerung des Potenzialaufbaus, z. B. durch einen hemmenden Input, nicht die gewünschte Leistung bringt, wenn der erregende Input mit unterschiedlicher Frequenz erfolgt. Da die „Sperrung“ des Aufbaus eines Aktionspotenzials allerdings grob vielleicht 30 bis 40 Millisekunden andauern muss, können auch Prozesse dahinter stehen, die nicht dem klassischen Konzept des Mechanismus der Refraktärphase entsprechen. Zellen, die den Anfang eines Impulszugs signalisieren, die es auf der Hörbahn nachweislich gibt (vgl. z. B. die Zusammenstellung verschiedener Zelltypen bei Trussel, 2002), können u. U. als positive Belege für die Möglichkeit eines entsprechenden Zellverhaltens herangezogen werden.

Die erforderliche Funktion kann in der Simulation mit den bisher verwendeten Mitteln demonstriert werden, wenn man tatsächlich (ersatzweise?) mit dem Parameter arbeitet, der die Dauer der Refraktärphase bestimmt. Das Problem sporadisch auftretender kurzer Impulsabstände ist dadurch zu lösen, dass man Zellparameter wählt, die eher auf das Wiederholen mehrerer Inputs abgestimmt sind, als auf die Kürze der Abstände. Das bedeutet, dass man mit relativ langen EPSP-Dauern und relativ kleiner Synapseneffektivität rechnen muss.

Der Einfachheit halber kann zur Erzeugung eines geeigneten Inputs die in Abbildung 3.3.3–8 gespiegelte Simulation wiederverwendet werden. Es müssen zwei auswertende Zellen hinzugefügt werden, die eine Verbindung zu den dort den Output bildenden Zellen *ome1* und *ome2* haben. Die Simulation ergibt auf dieser Basis für die auswertende Zelle *p1*, die *ome1* zugeordnet

ist, einen regelmäßigen Impulszug, wie er für die Repräsentation von Langvokalen bzw. anderen Dauerlauten erforderlich ist, in Zeitabständen, die den Bedürfnissen der Weiterverarbeitung, also dem Prinzip der Einzelimpulskodierung und dem dadurch vorgegebenen Impulsabstand, entsprechen.

(Es wird durch die refraktärphasenähnliche Sperrung der integrierenden Zellen nur der Mindestabstand zwischen den Aktionspotenzialen auf diesen Zellen eingestellt. Wenn die Aktionspotenziale einen zu großen Abstand haben, bricht aber die anschließende lexikalische Verarbeitung ab, es ergeben sich keine Störungen. Siehe den einzelnen Impuls in der Abbildung 3.3.4–3 bei *p2*, der durch eine vorübergehende rasche Impulsserie bei *ome2* ausgelöst wird.)

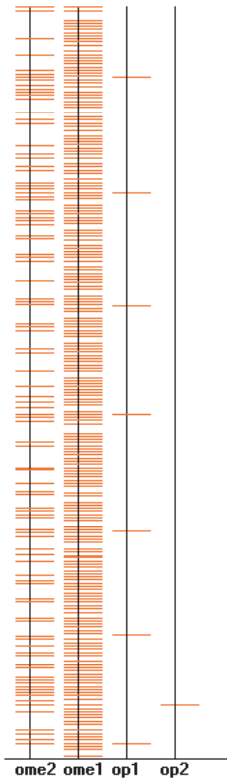



*Abbildung 3.3.4–3:* Simulation zum Übergang zur Einzelimpulskodierung. Es sind 6000 Zeittakte abgebildet. Zur Darstellung vgl. Abbildung 3.3.3–8 und die Erläuterungen dazu.

Simulation:
Übergang von Frequenzkodierung zu Einzelimpulskodierung.
Auf dem Bildschirm wird nur jeder zehnte Zeittakt abgebildet (die Aktionspotenziale allerdings vollständig), das ergibt einen Zeitraffereffekt. Die Option Start/Simulation bis Stop liefert das in Abbildung 3.3.4–2 wiedergegebene Ergebnis. Die Steuerung und Fortsetzung der Simulation mit Leertaste ist durch die Zeitrafferfunktion entsprechend verlangsamt.

*Binarität und Mehrdeutigkeit*

Unter der Voraussetzung, dass ein unmittelbar phonemdefinierendes Merkmal durch eine spezialisierte Zelle repräsentiert ist und eine funktionierende Kategorisierung (einschließlich der dahinterstehenden Lernprozesse) voraussetzt, dass diese Zelle einzelne Impulse abgibt, sind alle auditiv-phonetischen Merkmale von der Biologie her gesehen „privativ“, das heißt, sie können nicht binär sein und entweder einen positiven oder einen negativen Wert annehmen. Ein nichterregter Merkmalsspezialist, also ein Merkmalspezialist, der kein Aktionspotenzial abgibt, ist nicht ein Merkmalsspezialist mit einer negativen Ausprägung; vielmehr ist er an einer phonologischen Kategorisierung einfach nicht beteiligt. Wenn negative Merkmalsausprägungen angenommen werden, heißt das, dass das betreffende Merkmal in dieser besonderen Ausprägung vorliegt und eine spezialisierte Zelle tatsächlich erregt ist und damit zur Kategorisierung beitragen kann. Das bedeutet, dass die positive und die negative Ausprägung jeweils durch eigene Zellen repräsentiert sein müssen.

In früheren Publikationen (Kochendörfer, 1997 und 2002) wird eine neuronale Architektur vorgeschlagen, die binäre Merkmale für den Verstehensprozess nachbildet. Ausgangspunkt waren klassische Vorstellungen, nach denen es Merkmale wie ± *diffus,* ± *scharf (strident),* ± *kompakt* usw. gibt. Wenn die Idee solcher binärer Merkmale einen Sinn haben soll, müssen die Einheiten, die dann positiv oder negativ ausgeprägt erscheinen, im Wahrnehmungsprozess zunächst tatsächlich durch ein und dieselbe neuronale Einheit repräsentiert sein. Das kann nur dadurch geschehen, dass den Ausprägungen entsprechende Frequenzunterschiede der den Input kodierenden Impulszüge zugeordnet sind, also natürlicherweise für die positive Ausprägung eine höhere, für die negative Ausprägung eine niedrigere Frequenz. Auch hier gilt wieder, dass die negative Ausprägung nicht durch das Fehlen eines Inputs repräsentiert sein kann. Für die Weiterverarbeitung ist dann eine Aufspaltung erforderlich in Einheiten, die die einzelnen Ausprägungen separat repräsentieren und sich wie privative Merkmale verhalten. Das Grundprinzip ist in Abbildung 3.3.4–4 für das Merkmal *diffus* schematisch dargestellt.

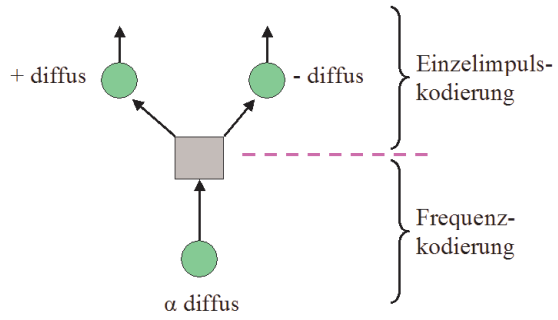



*Abbildung 3.3.4–4:* Schema zur Umkodierung von binären Merkmalen.

Wenn man binäre Merkmale und Umkodierungen der angedeuteten Art annimmt, liegt es nahe, auch den Fall vorzusehen, dass ein Input nicht klar als positive oder negative Ausprägung eines Merkmals erkannt wird, sondern gewissermaßen „dazwischen“ liegt, so dass sich also eine Mehrdeutigkeit ergibt. Mehrdeutigkeit bedeutet bei Annahme von Einzelimpulskodierung immer, dass mehrere konkurrierende Repräsentationen (Zellen) gleichzeitig aktiviert sind. Das heißt, eine mittlerere Intensität der Aktivierung eines Merkmals vor der Umkodierung muss dazu führen, dass die ihm entsprechenden privativen Merkmalseinheiten beide einen Impuls abgeben.

Diese Vorstellung von der Entstehung und Verarbeitung von Mehrdeutigkeiten im lautlichen Bereich kann ein (zusätzliches) Argument für die Annahme von binären Merkmalen sein. Das gilt natürlich nur, wenn Mehrdeutigkeit tatsächlich so zu verstehen ist. Das Merkmal *diffus* ist insofern ein kritisches Beispiel, als es das Verhältnis mehrerer Formanten beschreibt. Wenn aber, wie oben argumentiert, der einzelne Formant ein Merkmal bildet, wird die Mehrdeutigkeit direkt auf der Kochlea und den unmittelbar anschließenden Strukturen repräsentiert und kann nicht Gegenstand der mittleren Aktivität einer Formantenkombination sein. Mindestens muss also eine binäre Zwischenstufe nicht notwendig angenommen werden. Im Bereich der Vokale (des Deutschen) sind alle angenommenen binären Merkmale fragwürdig. Für den konsonantischen Bereich wird ohnehin auch in der existierenden Literatur vielfach mit privativen Merkmalen gerechnet.

Es ist nicht ausgeschlossen, dass binäre Merkmale in irgendeinem Bereich in irgendeiner Sprache eine Funktion haben könnten. Wenn das der Fall ist, müssen Strukturen zur Umkodierung, wie in Kochendörfer (1997) und (2002) beschrieben, angenommen werden. Die konsequente Modellbildung zwingt aber möglicherweise zu einer grundsätzlichen Umstrukturierung des gesamten phonetischen Merkmalsinventars im auditiven Bereich.

## 3.3.5 Lernen und Vergessen

*Allgemeine Bemerkungen*

Die äußerste Ebene der neuronalen Kodierung in jeder Sinneswahrnehmung muss auf angeborenen Strukturen beruhen, also Strukturen, die nicht Lernprozessen unterliegen. Das ist für die auditive Wahrnehmung genauso gültig wie für die visuelle Wahrnehmung und andere Sinnesmodalitäten. Strukturen und Funktionen des Innenohrs sind angeboren (vererbt) und müssen nicht durch Lernprozesse aufgebaut werden.

Die Anpassung des Verhaltens an wechselnde Umgebungsbedingungen setzt auf Verarbeitungsebenen, die an die äußerste Peripherie anschließen, Lernprozesse voraus oder jedenfalls Prozesse, die Information zu verankern in der Lage sind. Man stellt sich unter neuronalen Lernprozessen zunächst Prozesse vor, die Information durch Verstärken von Verbindungen schaffen. Information kann aber auch dadurch entstehen, dass die Auswahl unter möglichen Verarbeitungsbahnen beschränkt wird. Im Folgenden wird von „Lernen“ gesprochen, wenn neue Verbindungen funktionsfähig gemacht werden, von „Vergessen“, wenn Verbindungen funktionsunfähig werden. Dabei hat also Vergessen einen durchaus positiven, informationsschaffenden Sinn.

Die für die auditive Phonetik interessante Frage ist nun, wie die beiden in gewissem Sinn gegenläufigen Prozesse an dem Aufbau einer Sprachkompetenz im phonetischen Bereich, über die angeborenen Strukturen hinaus, beteiligt sind.

*Die Lage des zweiten Formanten*

Wenn man dem Produzenten einer sprachlichen Äußerung ein Äußerungssegment, in verschiedener Weise gefiltert (Rechteckfilter, mit Hilfe des Freeware-Programms *praat*), zu Gehör bringt und überprüft, was tatsächlich gehört wird, kommt man zu überraschenden Ergebnissen. Die im Folgenden zugrundegelegten Versuche hatten das Ziel, festzustellen, in welchem Frequenzbereich der zweite Formant des in den Beispielen der Abschnitte 3.3.3 und 3.3.4 verwendeten [i:] eines erwachsenen männlichen Sprechers beim Hören dieses Lautes durch dieselbe Person liegt.

Man wird zunächst als relativ unproblematisch annehmen, dass der zweite Formant jedenfalls nicht oberhalb 3500 Hz liegen wird. Wenn man die Frequenzen oberhalb 3500 Hz unterdrückt, ergibt der Versuch tatsächlich, dass ein [i:] gehört wird (Tonbeispiel 1).

Wenn man aber die Obergrenze des hörbaren Spektrums weiter senkt, so dass die Frequenzen oberhalb 2800 Hz unterdrückt werden, wobei also der in

der Spektralanalyse sichtbare akustische Frequenzbereich des zweiten Formanten gerade noch erhalten bleibt, wird nicht [i:] gehört, sondern eher [y:], was man vielleicht zunächst so interpretieren würde, dass höhere Formanten zur Wahrnehmung des [i:] *zusätzlich* beitragen (Tonbeispiel 2).

Diese Interpretation, die von der grundlegenden Bedeutung des im akustischen Spektrum sichtbaren zweiten Formanten ausgeht, erweist sich aber als nicht haltbar, denn die perfekte Wahrnehmung eines [i:] entsteht auch, wenn man genau diesen Formanten, das sind die Frequenzen zwischen 1800 und 2800 Hz, ausblendet (Tonbeispiel 3).

Und schließlich: Wenn man den gesamten Frequenzbereich von 1000 bis 3200 Hz unterdrückt, wird nach wie vor [i:] gehört (Tonbeispiel 4).

Wesentliche Teile des für die auditive Kategorisierung maßgebenden zweiten Formanten liegen offenbar zwischen 3200 und 3500 Hz, also weit jenseits des zweiten Formanten nach der Analyse des akustischen Signals, dessen Mittenfrequenz in einem Bereich liegt, der durchaus der in den Handbüchern angegebenen Norm für erwachsene männliche Sprecher entspricht (Abbildung 3.3.5–1).

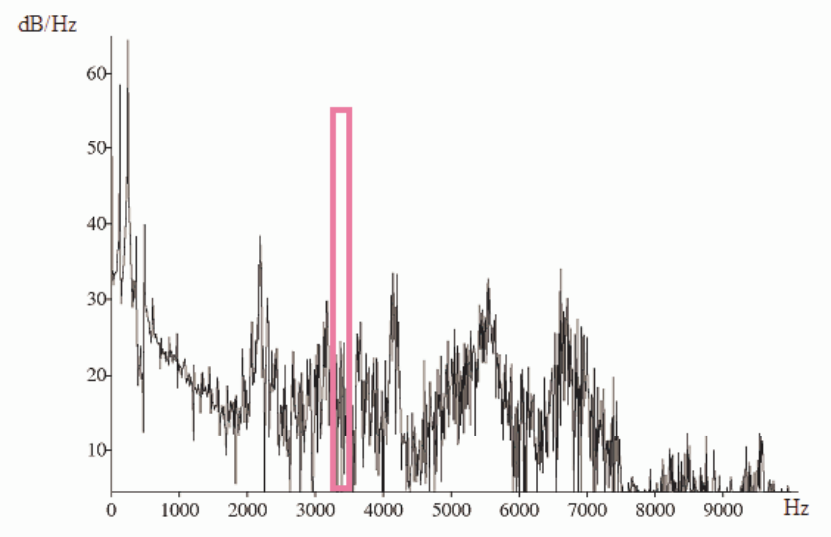



*Abbildung 3.3.5–1:* Lage des Frequenzausschnitts, der für die Wahrnehmung eines [i:] als zweiter Formant maßgebend ist.

Tonbeispiele:
Verschiedene Filterungen des Segments [i:] (Kopfhörer!):
1. Alle Frequenzen unterhalb 3500 Hz.
2. Alle Frequenzen unterhalb 2800 Hz.
3. Frequenzen 1800 bis 2800 Hz unterdrückt.
4. Frequenzen 1000 bis 3200 Hz unterdrückt.

Man beachte, dass nicht der akustische Formant $F_2$ durch einen Normalisierungsprozess auf die Position des Fensters in Abbildung 3.3.5–1 angehoben worden ist, sondern dass der akustisch dem Fenster entsprechende Frequenzbereich der Kategorisierung zugrundeliegt. Der traditionell als $F_2$ gezählte Frequenzbereich ist schlicht irrelevant. Er ist für die auditive Kategorisierung ersetzt durch einen „auditiven Formanten“ $F_2$; es bietet sich an, akustische und auditive Formanten konsequent zu unterscheiden.

Der springende Punkt ist jetzt aber, dass der auditive Formant $F_2$ für [i:] ungefähr dem akustischen Formanten $F_2$ bei Kindern entspricht.

*Phonetische Merkmale im Spracherwerb*

Phoneme können, selbst wenn man sich auf die auditive Seite beschränkt und den Zusammenhang mit der Artikulation außer Acht läßt, nicht angeboren sein. Dazu ist die Variabilität, über alle Sprachen gerechnet, zu groß. Bei den phonetischen Merkmalen, die Phoneme definieren, ist die Lage nicht so eindeutig. Wenn man akzeptiert, dass den phonetischen Merkmalen, jedenfalls auf elementarster Ebene, wie in Abschnitt 3.3.4 herausgearbeitet, Detektoren entsprechen, die ortskodierte Information unter Beachtung der Synchronizität der neuronalen Impulse auswerten, dann muss überprüft werden, ob die entsprechenden neuronalen Strukturen durch Lernprozesse entstehen können oder als angeboren gelten müssen.

Gesichtspunkte, die für Angeborenheit sprechen, sind:

- Das von den Detektoren verwendete Synchronizitätsfenster wird über die Leitungsgeschwindigkeit von Nervenfasern eingestellt, und muss angeboren sein.
- Haarzellen, deren Output (über Zwischenstufen) integriert wird, müssen (grob) benachbart sein. Entsprechende Verbindungsmuster müssen angeboren sein und können nicht aufgrund von Lernvorgängen wachsen.

Es ist u. U. möglich, an einen Lernvorgang zu denken, der schon von vornherein mit einer spezifischen Geltung angelegte Verbindungen verstärkt und damit funktionsfähig macht. Es ist aber nicht recht einzusehen, warum solche Verbindungen nicht von vornherein funktionsfähig sein sollten. Dass sie

wahrscheinlich tatsächlich zunächst funktionsfähig sind, zeigen die vielbeachteten Experimente, die demonstrieren, dass Kinder im Alter von z. B. 6 Monaten zur Diskrimination von Lauten fähig sind, die sie ein halbes Jahr später nicht mehr als verschieden erkennen. Am bekanntesten sind dazu die Arbeiten von Kuhl, z. B. Kuhl (2000), mit EKP-Technik z. B. Cheour et al. (1998). Solche Beobachtungen können als ein Hinweis darauf genommen werden, dass es einen Vergessensprozess für Merkmale gibt, der aus einem angeborenen universellen Merkmalsinventar die in der Umgebungssprache gültigen Merkmale herausfiltert. In einer etwas idealisierten Konstruktion sieht das dann so aus, wie in Abbildung 3.3.5–2 skizziert.

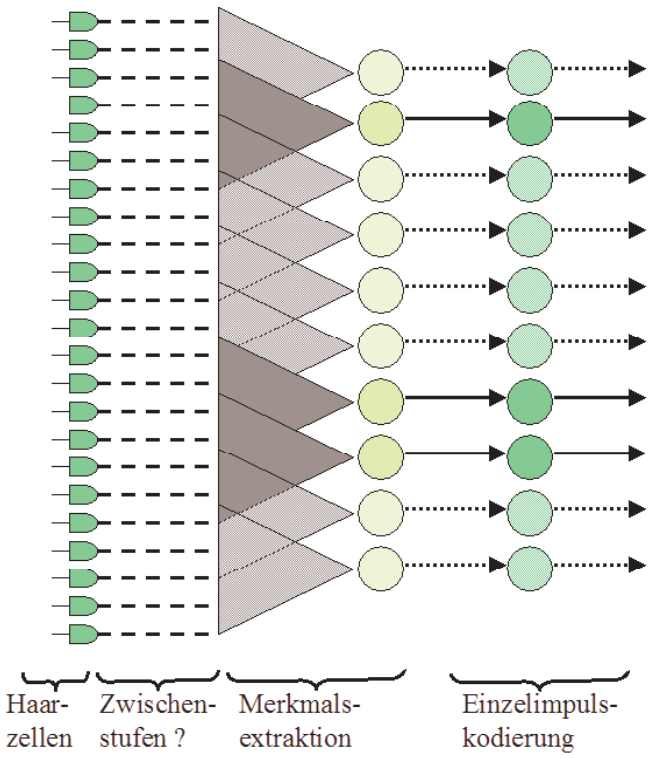



*Abbildung 3.3.5–2:* Schema zur Merkmalsextraktion. Aufgrund von Vergessensprozessen für sprachliche Stimuli unwirksam gewordene Merkmalsdetektoren sind halbtransparent und mit gepunkteten Verbindungspfeilen dargestellt.

(Man beachte in diesem Zusammenhang: Die Lokalisierung von ereigniskorrelierten Potenzialen wie die „mismach negativity“ im Kortex bedeutet nicht, dass die Ursache des Entstehens solcher Potenziale ebenfalls im Kortex anzusiedeln ist!)

Erwachsene haben die in ihrer Sprache gültigen Formanten als Kinder in einem sehr jungen Alter durch einen Vergessensprozess „gelernt“, der darin besteht, dass Detektoren für mögliche Formanten, die in der Umgebungssprache nicht vorkommen, geschwächt bzw. abgebaut werden. Beim Spracherwerb durch Vergessen gehen nicht bestimmte Frequenzen verloren, sondern Verarbeitungskategorien, die nur sprachlich vorkommen und nichtsprachlich auch nicht wahrgenommen werden können.

Es handelt sich um einen durchaus positiv zu bewertenden, die Verlässlichkeit der Kommunikation begünstigenden Vorgang. Der hauptsächlich gehörte Input ist aber wohl immer die eigene Produktion, oder jedenfalls nicht die einer Männerstimme. Erfahrungen beim späten Erwerb einer Fremdsprache deuten darauf hin, dass der Vorgang nicht oder nur schwer revidiert werden kann. Es ist also sehr naheliegend, damit zu rechnen, dass auditivphonetische Merkmale dem entsprechen, was man als Kind erworben hat.

Wenn man diesen Zusammenhang beachtet, verschwindet aber das Normalisierungsproblem. Es ist nicht zu verwundern, dass für ein in der biologischen Wirklichkeit gar nicht existierendes Problem, wie es das Normalisierungsproblem ist, so viele biologisch nie ganz befriedigende Lösungen vorgeschlagen worden sind.

*Das Problem der Lokalisierung merkmalsbezogener Prozesse auf der Hörbahn*

Die Leistung der Merkmalsdetektoren setzt voraus, dass ausreichend hohe Frequenzen vorhanden sind, die eine Auswertung der Synchronizität ermöglichen. Von dieser Voraussetzung her ist es z. B. unmöglich, den Output der Schalllokalisationsstrukturen als Input für die Merkmalsextraktion zu verwenden. Dasselbe gilt für die Strukturen, die der Wahrnehmung von Lautheit entsprechen. Wenn man beachtet, dass die Frequenz der Aktionspotenziale auf einzelnen Fasern im Verlauf der Hörbahn abnimmt, wird man die Merkmalsextraktion auf möglichst frühen Stufen der auditiven Verarbeitung annehmen.

Damit stellt sich die Frage, ob die Merkmalsextraktion schon erfolgt, ehe die Inputs von beiden Ohren „zusammengemischt“ werden, also z. B. schon auf der Ebene des Nucleus cochlearis, oder erst aufgrund des Mischungsprodukts. Leider reicht der Hinweis auf die relative „Gleichberechtigung“ der beiden Ohren nicht aus, um eine Entscheidung herbeizuführen. Von der Mo-

dellbildung her bleibt zunächst ebenfalls offen, ob die ODER-Verknüpfung, die zu dem *einen* Lexikon im Kortex führt, schon vor der Merkmalsextraktion erfolgt oder vielleicht auch erst im Anschluss an die Ableitung einzelimpulskodierter Merkmale.

In Arbeiten zur Hörbahn wird darauf hingewiesen, dass die beiden Teile des Nucleus cochlearis unterschiedliche Funktionen haben. Der ventrale Teil hat Funktionen, die mit der Schalllokalisation zu tun haben, dem dorsalen Teil werden Funktionen im Bereich der auditiven Mustererkennung zugeschrieben. Der Nucleus cochlearis dorsalis enthält offenbar auch tatsächlich Zellanordnungen, die den oben für die Merkmalsanalyse angenommenen entsprechen (Young & Davis, 2002).

Es ist eine durchaus elegante Annahme, die Merkmalsanalyse tatsächlich dem dorsalen Nucleus cochlearis zuzusprechen. Unter dieser Annahme wäre dann der nächste Verarbeitungsschritt, die Umwandlung in Einzelimpulskodierung, dem Colliculus inferior zuzuordnen, und schließlich die Zusammenführung der beidohrigen Merkmale auf eine Einheit dem Corpus geniculatum mediale.

Der Mischungsversuch in Abschnitt 3.3.4 (Abbildung 3.3.4–1) würde prinzipiell zulassen, dass schon im dorsalen Nucleus cochlearis nicht nur Merkmale gebildet werden, die einzelnen Formanten entsprechen, sondern solche, die in einem zweiten Schritt Formantenkombinationen repräsentieren. Im Extremfall würde man annehmen können, dass z. B. Vokale direkt im Nucleus cochlearis bei einer entsprechenden Formantenkombination festgestellt werden. Diese an sich schon wegen u. U. erforderlicher Lernprozesse schwierige Annahme lässt sich mit Hinweis auf die Möglichkeit weiter schwächen, Laute, deren Spektren auf beide Ohren aufgeteilt sind, zusammengesetzt zu hören, wie oben in Abschnitt 3.3.3 demonstriert worden ist. Ein analoges Experiment auf der Basis des Mischungsversuchs von Abschnitt 3.3.4 besteht darin, die Aufteilung der Schallfrequenzen so vorzunehmen, dass die Grenze *zwischen* den beiden Formanten des [ɑː] verläuft. In den folgenden Tonbeispielen liegt diese Grenze bei 800 Hz. Der obere Teil des Spektrums reicht, im Hinblick auf die auditive Wahrnehmung des zweiten [iː]-Formanten, bis 3500 Hz.

Wenn man sich die beiden Bestandteile des Spektrums separat anhört, wird ganz deutlich ein [ɑː] gehört (Fall des engen Zusammenliegens der Formanten wie bei u, ein Formant reicht), das [iː] dagegen ist weg. Die unteren Frequenzen liefern ein [ɑː] mit einem zusätzlichen Brummton, die oberen ein sehr hell eingefärbtes [ɑː], eventuell mit einem Nebengeräusch. Die Wahrnehmung beider Teile zusammen liefert neben dem etwas im Vordergrund stehenden [ɑː] auch wieder ein etwas blasseres [iː].

Tonbeispiele:
Mischung der Formanten von [iː] und [ɑː] (Kopfhörer!):
1. Frequenzen unterhalb 800 Hz.
2. Frequenzen oberhalb 800 Hz.
3. Frequenzen unterhalb 800 Hz links, oberhalb 800 Hz rechts.

Wenn prinzipiell durch den Nucleus cochlearis schon Formantenkombinationen festgestellt würden, dürfte bei Nicht-Vorhandensein eines entsprechenden Spezialisten am Ende kein Sprachlaut hörbar sein. Man kann aus den Experimenten also folgern, dass der Nucleus cochlearis offenbar Fasern aktiviert, die Merkmalen entsprechen, und dass diese Merkmale erst in einem zweiten Schritt integriert werden. Das Angebot an Merkmalen, die sich aus dem Mischungsversuch ergeben, ist nur zu den beiden Phonemen [iː] und [ɑː] kombinierbar, die dann gleichzeitig aktiviert und gehört werden.

Wenn die Leistung des Nucleus cochlearis dorsalis auf die Feststellung einzelner Merkmale beschränkt bleibt, ergibt sich als durchaus positiv zu bewertende Konsequenz, dass Lernvorgänge, die in der Verstärkung von Synapsen bestehen, erst im Anschluss an die Einzelimpulskodierung erforderlich sind, nach der oben angedeuteten Verteilung der Verarbeitungsschritte auf die einzelnen Nuclei der Hörbahn also erst im Kortex und erst, nachdem die Analyseprodukte der beiden Ohren, soweit sie gleich sind, auf dieselben Strukturen zusammengeführt sind. Die Verknüpfungen, die diese Zusammenführung leisten, können durchaus noch als angeboren gelten.

## 3.3.6 Zusammenfassende Skizze auditiv-phonetischer Prozesse

*Überblick*

Eine sicherlich etwas spekulative Zusammenstellung von phonetischen Teilprozessen der Sprachperzeption, die in Zusammenhang mit der Lautkategorisierung stehen, und deren Zuordnung zu Strukturen der Hörbahn ist in Abbildung 3.3.6–1 gegeben.

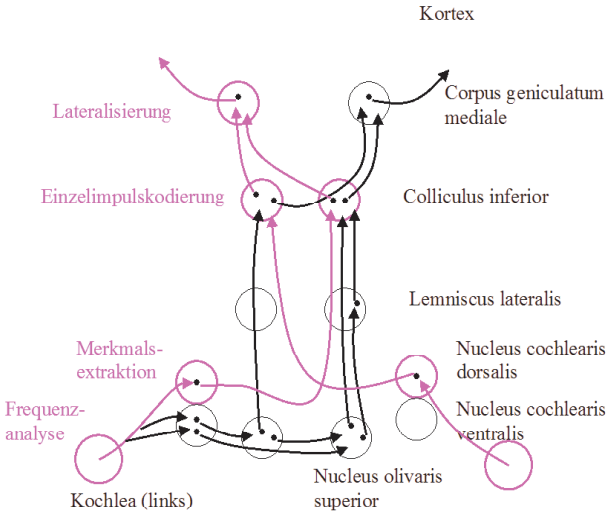



*Abbildung 3.3.6–1:* Schema der Hörbahn wie in Abbildung 3.3.3–1, mit rot eingezeichneten Strukturen, die an der sprachlichen Kategorisierung beteiligt sind. Nichtsprachliche Strukturen sind schwarz und nur für *ein* Ohr angedeutet.

Generell gilt, dass die dort eingezeichneten Bahnen nicht so gesehen werden dürfen, dass sie neutral sind gegenüber dem Input, den sie transportieren. Die gezeichneten Linien stehen für Faserbündel, in denen die einzelnen Fasern für bestimmte Bedeutungen stehen. Es kann höchstens die *Intensität* bestimmter Komponenten des Sprachschalls von Stufe zu Stufe auf der Hörbahn transportiert werden, nicht die Komponenten selbst. Das bedeutet, dass mit jedem Umschaltprozess auf der Hörbahn neue Inhalte entstehen, die jeweils in der Aktivierung inhaltsspezifischer Fasern bestehen. Die einzelnen Fasern tragen aber nur insofern Bedeutung, als sie über Umschaltprozesse mit den Sinneszellen des Innenohrs verbunden sind.

Es ist zu vermuten, dass im Spracherwerb auf dem gesamten Bereich der sprachbezogenen Hörbahn Information durch Vergessensprozesse entsteht. Strukturen, die eine Merkmalsanalyse leisten und Strukturen, die für den Übergang zur Einzelimpulskodierung und schließlich auch für die Lateralisierung erforderlich sind, sind prinzipiell angeboren (vererbt). Verarbeitende Strukturen sind immer pro Faser vorhanden, es ist nicht möglich, dass Informationen, die an verschiedene Fasern gebunden sind, gemeinsame Verarbeitungsstrukturen verwenden.

*Ohr*

Die Verarbeitung von Sprachschall beginnt mit der Frequenzanalyse und neuronalen Umkodierung im Innenohr, die zu Impulszügen auf Hörnervenfasern führt, wobei die einzelne Faser jeweils eine Bestfrequenz hat, aber nicht nur auf diese Frequenz anspricht, so dass sich gerade für den Sprachschallbereich eine Unschärfe ergibt. Die entstehenden Erregungsmuster sind über mehrere Fasern bis zu einem gewissen Grad synchron, so dass die Synchronizität zusätzlich zu dem „Ortskode“ (also der Abbildung des Schallspektrums durch die Zuordnung von Zellen entlang der Kochlea zu Bestfrequenzen) ausgewertet werden kann.

*Hörbahn*

Der Hörnerv repräsentiert noch alle sich aus dem akustischen Signal ergebenden Informationstypen. Eine erste Differenzierung wird anschließend im Nucleus cochlearis geleistet. Man beachte, dass, wie schon oben angedeutet, eine Differenzierung immer bedeutet, dass, den verschiedenen Inhalten entsprechend, verschiedene Fasern aktiviert werden. Es ist also vom Nucleus cochlearis ausgehend mit verschiedenen Verarbeitungsbahnen für die einzelnen Informationstypen zu rechnen und also auch mit der Möglichkeit, dass von dieser Verarbeitungsstufe an für verschiedene Informationstypen verschiedene Verarbeitungsschritte vorgesehen sind. Das gilt z. B. für die Schalllokalisation, für die Lautheit, wahrscheinlich auch für die Musikwahrnehmung usw.

Der Nucleus cochlearis dorsalis liefert als erstes Zwischenprodukt für die Kategorisierung von Sprachlauten die Identifikation auditiver Formanten, das heißt, es werden den auditiven Formanten entsprechende Fasern aktiviert. (Man sollte aber nicht annehmen, dass der Nukleus cochlearis dorsalis ausschließlich für sprachliche Merkmale zuständig ist.) Es findet keine Auswertung des Abstands der Formanten statt, das heißt, phonetische Merkmale beruhen nicht auf einer solchen Auswertung.

Die Bezeichnung „auditiver Formant“ weist darauf hin, dass die Leistung des Nucleus cochlearis dorsalis für die Sprachwahrnehmung nicht einfach darin besteht, die Intensitätsgipfel im akustischen Spektrum herauszufiltern, sondern dass die Filterfunktion zusätzlich an die Umstände des Spracherwerbs gebunden ist. Nur unter der Voraussetzung der Unterscheidung von akustischen und auditiven Formanten verschwindet das Normalisierungsproblem und es ist irrelevant, ob ein männlicher, weiblicher oder kindlicher Gesprächspartner gehört wird.

Kleinere Abweichungen in der akustischen Erscheinung von Sprachlauten führen nicht zu einem nennenswerten Normalisierungsproblem sondern werden dadurch irrelevant, dass grundsätzlich schon die neuronale Verarbeitung an der einzelnen Nervenzelle eine gewisse Schwankungsbreite des Inputs zulässt.

Die Annahme sprachspezifischer Strukturen (jedenfalls nach erfolgtem Spracherwerb) auf der Hörbahn ist nicht identisch mit der Annahme eines sprachspezifischen „Moduls“, wesentliche Eigenschaften des Modularitätskonzepts z. B. bei Fodor (1983) sind nicht gegeben. Man kann insbesondere nicht unterscheiden zwischen (peripheren) Prozessen, die in Modulen eingekapselt ablaufen und zentralen Prozessen, für die das nicht gilt.

Die vom Nucleus cochlearis dorsalis ermittelten auditiven Merkmale bleiben bis in den Kortex hinein getrennt, das heißt, es handelt sich um Erregungen auf entsprechenden formantenspezifischen Fasern. Die einzelne Faser, die den Nucleus cochlearis verläßt, führt, sofern sie einen für die Weiterverarbeitung positiven Wert hat, einen Impulszug, der jedenfalls eine höhere Frequenz als 100 Hz haben muss. Die Bildung von Lautkategorien im Kortex setzt aus Gründen, die in Teil 2, „Grundlagen“, Abschnitt 2.3.2, entwickelt werden, aber Einzelimpulskodierung voraus, das heißt, grob, Frequenzen jedenfalls unterhalb von 30 Hz.

Die Umsetzung in Einzelimpulskodierung muss jedem koinzidenzbasierten Lernprozess vorausgehen, die Kategorisierung von Sprachlauten kann also nicht direkt auf die Auswertung von Formantenkombinationen schon im Nucleus cochlearis folgen, solange (unbestritten) gilt, dass Sprachlaute solche Lernprozesse voraussetzen. Es ist aber davon auszugehen, dass sowohl der einfache Übergang eines frequenzkodierten in ein einzelimpulskodiertes Signal als auch die Verarbeitung von binären Merkmalen noch Vorgänge auf der Hörbahn sind. Da sie (logisch) unmittelbar auf Vorgänge im Nucleus cochlearis folgen und andere neuronale Strukturen voraussetzen als man sie dort antrifft, kann man sie dem Colliculus inferior, der direkt von dem Nucleus cochlearis dorsalis aus erreicht wird, zuordnen.

Alle Verarbeitungsvorgänge bis zu diesem Punkt erfolgen für beide Ohren getrennt und austauschbar. Anatomische Strukturen, die einen Input von beiden Ohren bekommen und damit die Grundlage für die Lateralisierung des Sprachvermögens gerade im phonetischen Bereich bilden können, sind die Nuclei des Corpus geniculatum mediale. Es ist unter normalen Umständen für die sprachliche Kategorisierung gleichgültig, über welches Ohr wir den Sprachschall wahrnehmen. Für den Übergang zur sprachdominanten Hemisphäre (bei Rechtshändern zur linken Hemisphäre) ist also ein Vorgang anzunehmen, der einem logischen ODER entspricht. Man beachte,

dass es im Hinblick auf die Weiterverarbeitung im Bereich des mentalen sprachlichen Lexikons erforderlich ist, dass von beiden Ohren (aufgrund eines im wesentlichen identischen akustischen Inputs) ausgelöste Impulse die ODER-Verknüpfung gleichzeitig (innerhalb eines ausreichend kleinen Zeitfensters, definiert z. B. durch die Refraktärphase der verarbeitenden Zellen) erreichen. Es darf in diesem Fall nicht zu Impulsdoppelungen im Kortex kommen.

Wenn die Integration von Erregungen, die von beiden Ohren her kommen, erst an dieser Stelle erfolgen kann, bedeutet das auch eine zusätzliche Bestätigung für die Annahme, dass die davor liegenden Verarbeitungsschritte sich auf phonetische Merkmale, nicht auf Laute bzw. Phoneme beziehen, denn sonst wären die Effekte der in den vorangegangenen Abschnitten durchgeführten Experimente mit der Auftrennung von Frequenzbestandteilen nicht erklärbar. Die Auswertung von Merkmalskomplexen ist dann Leistung des Kortex.

*Primärer auditiver Kortex*

Die Leistung des primären auditiven Kortex ist nicht auf die Sprachwahrnehmung beschränkt. Andererseits könnten hier möglicherweise auch schon Verarbeitungsschritte lokalisiert sein, die etwas mit der Bildung von Phonemen zu tun haben. So ist es sicherlich realistisch, mit einer hier erfolgenden größeren Aufspaltung bzw. Vervielfältigung der merkmalsspezifischen Bahnen zu rechnen. Weitere Verarbeitungsschritte sind dann eher anschließenden Kortexbereichen zuzuordnen. Der Bereich der Phonetik endet mit den durch Lernprozesse entstandenen einzelsprachlichen Phonemkategorien.

*Lautliche Vorstellungen*

Lautliche Vorstellungen sind sprachliche Produktionsprozesse, die in den phonetischen Bereich hineinreichen und eine Einwirkung auf Perzeptionsstrukturen voraussetzen. Man kann sich nicht Unterschiede von Lauten vorstellen, die man als Unterschiede nicht wahrnimmt. In Teil 2, Abschnitt 2.5.3, wird die These entwickelt, dass die für das Entstehen von Vorstellungen verantwortlichen Produktionsverbindungen von Phonemrepräsentationen ausgehen und auf Zellen führen, die zur Repräsentation von einzelimpulskodierten phonetischen (phonemdefinierenden) Merkmalen führen. Die Erregung dieser Merkmale ist gleichbedeutend mit der Ersatzwahrnehmung von Sprachschall, erzeugt über einen Vorgang, der metaphorisch gesprochen als „Rückspiegelung“ bezeichnet werden kann. Die Abbildung 3.3.5–2 stellt diesen Zusammenhang schematisch dar.

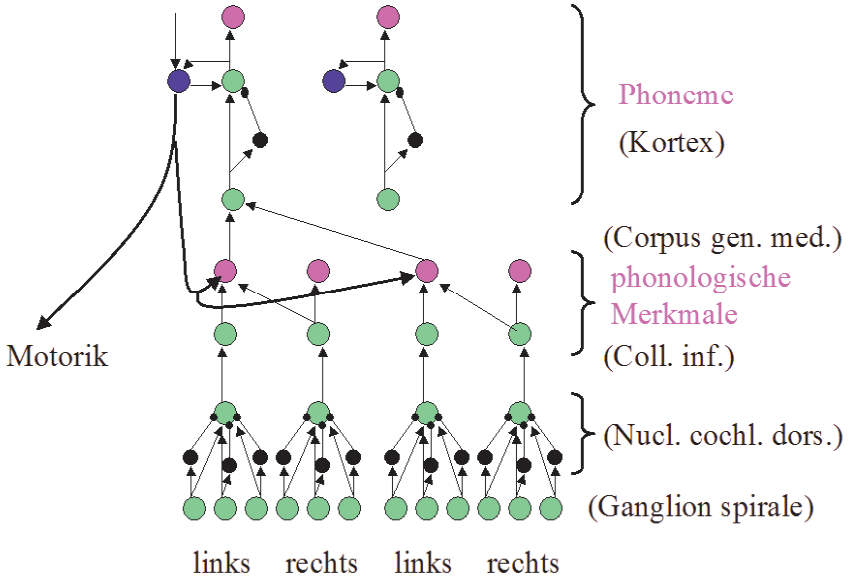



*Abbildung 3.3.5–2:* Rückspiegelungsverbindungen (fett ausgezeichnet) zur Erzeugung von auditiven Vorstellungen. Zur Begründung vgl. Teil 2, „Grundlagen", Abschnitt 2.5.3. Die Abbildung 2.5.3–8 dort verwendet Strukturen zur Auswertung binärer Merkmale, die hier weggelassen und durch die Entsprechungen für privative Merkmale ersetzt sind.

Analoge Mechanismen gelten für musikalische Vorstellungen und Vorstellungen von Geräuschen. Die Zellen, die durch die Rückspiegelungsverbindungen erregt werden, sind, ihrer Funktion entsprechend, ODER-Zellen (rote Färbung). Es ist offen, ob sie identisch sind mit den Zellen, die die Informationen von beiden Ohren verknüpfen, wie in der Abbildung 3.3.5–2 dargestellt, oder ob eine zweite Ebene von Zellen dafür zustandig ist; und es ist auch schwierig, zu entscheiden, ob an dieser Stelle Lernvorgänge erforderlich sind.

*Efferente Fasern*

Rückspiegelungsvorgänge werden durch Fasern bewirkt, die entgegen der Wahrnehmungsrichtung auf der Hörbahn verlaufen. Es scheint kein weiterer für das grundsätzliche Funktionieren erforderlicher Bedarf an solchen efferenten Fasern für die Lautkategorisierung zu bestehen. Auch die efferenten Fasern, die letztlich die äußeren Haarzellen innervieren, haben offenbar

keinen Einfluss speziell auf die Sprachverarbeitung, sondern sind von allgemeiner Bedeutung für den Hörvorgang.

*Abschließende Bemerkungen*

Es ist in diesem Kapitel 3.3 keine vollständige Erklärung von Hörphänomenen beabsichtigt. Es sollte versucht werden, sozusagen einen Unterbau zu lexikalischen Phänomenen (einschließlich der Phonemebene) zu konstruieren. Daraus ergibt sich im Wesentlichen die Beschränkung auf das Problem der Lautkategorisierung, ohne dass damit die sprachliche Relevanz anderer Erscheinungen, z. B. solcher, die zur Prosodie rechnen, in Frage gestellt wird. Lautheit, Prosodie und dergleichen werden über Strukturen bzw. Fasersysteme verarbeitet, die verschieden sind von denen der Lautkategorisierung. Psychoakustischen Beobachtungen, soweit sie nicht direkt mit Funktionen des Innenohrs in Verbindung gebracht werden können, werden deshalb ausgeklammert. Das betrifft auch die in Phonetikdarstellungen gerne behandelten Maskierungsphänomene, die nicht vollständig als Steuerungsfunktionen durch die äußeren Haarzellen verstanden werden können.

Die hier dargestellte Konzeption der Lautkategorisierung ist gegenüber klassischen Vorstellungen der auditiven Phonetik in wesentlichen Punkten verändert. Es handelt sich um Veränderungen, die es allererst ermöglichen, zu einem konsistenten Bild der Sprachwahrnehmung zu kommen. Die wichtigsten sind:

- Ablösung von der Vorherrschaft akustischer Analysen bei Versuchen, die Lautkategorisierung zu verstehen, und Lösung des Normalisierungsproblems durch Berücksichtigung von Spracherwerbsvorgängen.
- Ableitung von auditiven Merkmalen direkt aus dem Erregungsmuster der Hörbahn.
- Lernen durch Koinzidenz erst im Kortex (abgesehen von Rückspiegelungsverbindungen?).

Die verwendeten Messergebnisse sind zum größten Teil allgemein bekannt, die Gesamtkonstruktion ist ein Ergebnis konsequenter Modellbildung.

# 3.4 Produktion

## 3.4.1 Gegenstände

Die artikulatorische Phonetik, so wie sie hier verstanden wird, beschäftigt sich mit neuronalen und motorischen Prozessen, Strukturen und Repräsentationen, die von den Phonemrepräsentationen des Kortex bis zur Erregung der am Sprechen beteiligten Muskulatur führen. Der Schwerpunkt der Darstellung in diesem Kapitel 3.4 liegt weniger bei den groben anatomischen Voraussetzungen und der physikalischen Erklärung des akustischen Produkts, sondern hauptsächlich bei Problemen, die den Verlauf des Sprech*prozesses* und die erforderlichen neuronalen Voraussetzungen betreffen.

Wie bei der auditiven Phonetik gilt auch hier, dass die äußerste Peripherie relativ gut untersucht ist. Das gilt vor allem für Beobachtungen der beteiligten Muskulatur. Es wird mit zunehmender Genauigkeit und unter Einsatz elektronischer Apparate geklärt, welche Muskelgruppen welche Bewegungen zu welchen Zeitpunkten ausführen. Mit Hilfe der Elektromyographie können über eingestochene Elektroden extrazellulär Muskelaktionspotenziale abgeleitet werden. Was fehlt, ist die Klärung der funktionellen Verbindung zwischen Lexikonstrukturen bzw. dem Phoneminventar im Kortex und diesen weit außen liegenden Phänomenen. Viele Details der Innervierung der Muskulatur durch das Zentralnervensystem sind ungeklärt. In diesem Bereich sind Messungen bei menschlichen Versuchspersonen auch mit modernen Hilfsmitteln wie EEG und den verschiedenen bildgebenden Verfahren (noch) zu wenig präzise bzw. auch prinzipiell inadäquat, und mit anderen Techniken aus ethischen Gründen nur sehr begrenzt möglich. Wie bei der auditiven Phonetik auch, kann man in dieser Situation erwarten, dass ersatzweise die Technik der Modellbildung wenigstens ein Stück weit zur Klärung beiträgt.

Wenn man die Ergebnisse des vorigen Kapitels akzeptiert, ist von vornherein klar, dass es keine Eins-zu-eins-Zuordnung von auditiven und artikulatorischen phonetischen Merkmalen geben kann, obwohl man natürlich zeigen kann, welche akustischen Veränderungen durch welche artikulatorischen Einstellungen bewirkt werden. Beobachtungen, die im Verlauf des Spracherwerbs gemacht werden, ergeben außerdem, dass die artikulatorischen Eigenschaften von Sprachlauten einen Lernprozess voraussetzen, dessen linguistische Basis auditive Wahrnehmungen sind (man vergleiche dazu Teil 8, „Spracherwerb"). Die Verbindung zwischen auditiven und artikulatorischen Vorgängen bzw. Repräsentationen wird über die Phonemebene, nicht die Merkmalsebene hergestellt. Man beachte in diesem Zusammenhang, dass auditive Formanten isoliert selbstverständlich nicht nachgeahmt werden können, und Kinder im Spracherwerb nicht artikulatorische Merkmale, sondern Phoneme lernen. Das gilt natürlich besonders auch dann, wenn man annimmt, dass Bewegungen, die artikulatorischen Teilprozessen entsprechen, mindestens teilweise auch außersprachliche Funktionen haben.

Die Konstruktion eines gültigen Modells der Sprachproduktion setzt, wie bei der Sprachperzeption auch, voraus, dass die Vorgaben, die sich aus den Bedingungen der Lexikonstruktur und der daraus abgeleiteten Eigenschaften der Phonemebene ergeben, beachtet werden. Für die äußerste Peripherie gilt, dass alle im sprachlichen Rahmen und für normale sprachliche Funktionen interessanten Outputs an Muskelfunktionen gebunden sind. Einen besonderen Schwerpunkt bildet deshalb die Innervierung und Funktion der Skelettmuskulatur. Dafür werden motorische Programme oder jedenfalls Steuerungsfunktionen benötigt, deren Verhältnis zu den üblichen Merkmalsbeschreibungen in der Phonetik zu klären ist. Aus der Betonung der Prozesshaftigkeit der einschlägigen Komponenten ergeben sich charakteristische Uminterpretationen phonetischer Kategorien.

### 3.4.2 Grundlegende Annahmen über die Funktion artikulatorischer Komponenten von Phonemen

Phoneme sind Einheiten, die durch Lernprozesse etabliert werden müssen. Das bedeutet, dass Komponenten vorausgesetzt sind, die in diese Lernprozesse eingehen. Diese Komponenten sind zunächst auditiver, nicht artikulatorischer Natur und entstehen durch inputgetriebene Lernprozesse („Lernen durch Vergessen"). Die Verbindungen, die von den durch Großmutterzellen repräsentierten Phonemeinheiten zur motorischen Peripherie führen,

müssen, soweit sie nicht angeboren sind, im Spracherwerb (erst nach dem Erwerb der entsprechenden auditiven Kategorien!) durch einen Versuchs-Irrtums-Prozess aufgebaut (genauer: entsprechend verstärkt) werden. Es ist auffällig, dass Versuchs-Irrtums-Prozesse auch für andere motorische Lernvorgänge, z. B. das Greifen von Gegenständen, gelten.

Bestandteile motorischer Programme könnten prinzipiell beliebig elementar sein, so dass man annehmen könnte, dass artikulatorische phonetische Merkmale letztlich ersetzt werden müssten durch eine Vielzahl von elementaren Parametern zur Muskelsteuerung, und die motorischen Programme als Ganzes müssten ebenfalls prinzipiell keine neuronal durch einzelne Zellen repräsentierten Einheiten sein, so dass sich schematisch ein Aufbau wie in Abbildung 3.4.2–1 **A** ergibt. Die Gegenposition ist die Annahme einer hierarchischen Struktur wie in Abbildung 3.4.2–1 **B**.

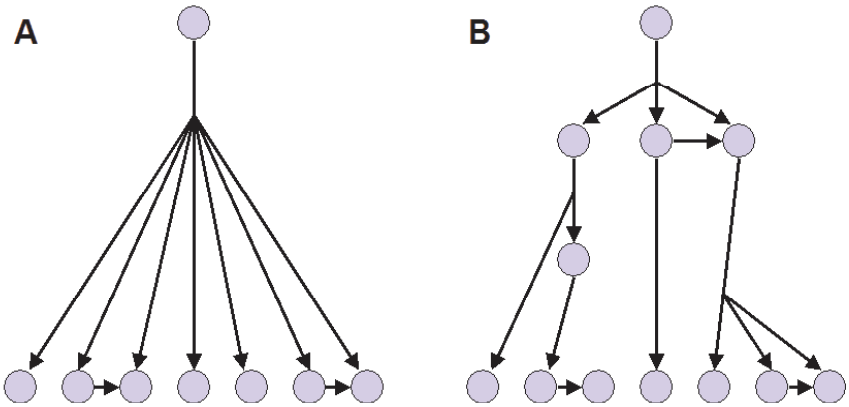



*Abbildung 3.4.2–1:* Schematische Darstellung verschiedener Möglichkeiten von Produktionsstrukturen (waagrechte Pfeile deuten zeitliche Abfolgen an). **A**: Direkte Ansteuerung von elementaren motorischen Komponenten von Phonemeinheiten aus. **B**: Hierarchische Ansteuerung von motorischen „Unterprogrammen“ durch Phonemeinheiten.

Der Vorteil von Anordnungen wie in 3.4.2–1 **B** ist vor allem die mehrfache Verwendbarkeit der „Unterprogramme“. In diesem Zusammenhang ist, wie oben schon angedeutet, darauf hinzuweisen, dass die beim Sprechen verwendeten Bewegungen wahrscheinlich durchweg auch einen nicht-linguistischen Zweck haben (Kaubewegungen, Schreien, Atmen usw.). Motorische Unterprogramme sind ganz offenbar teilweise angeboren und auch teilweise Ergebnisse von lebenswichtigen nicht-sprachlichen Lernprozessen.

Wenn man 10 Lippenmuskeln (schematisch z. B. dargestellt bei Pompino-Marschall, 2003: 57) auf jeder Gesichtshälfte unterscheiden kann, ist es schwierig, anzunehmen, dass ein spezifisch sprachlicher Lernprozess durch ein Versuchs-Irrtums-Verfahren jeden einzelnen dieser Muskeln, direkt von jeder Phonemrepräsentation ausgehend, ansteuerbar macht. Wenn der Spracherwerb auf bereits etablierte Unterprogramme zurückgreifen kann, wird er wesentlich vereinfacht. Es kann eigentlich unter Berücksichtigung aller Umstände kein Zweifel sein, dass die motorische Produktion auf hierarchischen Strukturen beruhen muss.

Die Frage ist, welchen Umfang solche von der sprachlichen Produktion verwendeten in Hierarchien eingebundene Unterprogramme haben, wie tief die Hierarchien sind (wieviele Zwischenstufen anzunehmen sind) und wie das Verhältnis zu den traditionell bzw. in den verschiedenen Varianten der Gestentheorie verwendeten artikulatorischen Merkmalen ist.

Eine Möglichkeit wäre, sich an Größenordnungen zu orientieren, die z. B. den „Zungenparametern“ entsprechen, die Hardcastle (1976: 100) so beschreibt (in deutscher Übersetzung bei Pompino-Marschall, 2003: 50):

> „1. Horizontal forward–backward movement of the tongue body
>
> 2. Vertical upwards–downwards movement of the tongue body
>
> 3. Horizontal forward–backward movement of the tip–blade
>
> 4. Vertical upwards–downwards movement of the tip–blade
>
> 5. Transverse cross-sectional configuration of the tongue body: convex–concave, in relation to the palate
>
> 6. Transverse cross-sectional configuration extending throughout the whole length of the tongue, particularly the tip and blade – degree of central grooving
>
> 7. Surface plan of the tongue dorsum – spread, tapered.“

Die Parameter (1) bis (4) beschreiben horizontale und vertikale Bewegungen der Zunge, die Parameter (5) bis (7) Veränderungen der Form. Alle Bewegungen und Veränderungen werden durch das Zusammenwirken mehrerer Muskeln erreicht. Es handelt sich also nicht mehr um Primitive nach dem Muster der Abbildung 3.4.2–1 **A**.

Auffällig ist dabei allerdings, dass nicht Artikulationen als Bewegungsziele beschrieben werden, sondern die Bewegungen selbst, die zu entsprechenden Zielen führen. Alle diese Bewegungen sind gradiert, das heißt, sie können

mehr oder weniger weiträumig, schnell usw. sein. Ein Problem dabei ist, dass Bewegungen prinzipiell nicht nur durch einen Zielzustand, sondern auch durch einen Startzustand bestimmt werden. Die erforderlichen Bewegungen ändern sich, je nachdem, welcher Zustand des Bewegungsapparats ihnen vorausgeht. Hardcastle (1976: 101) gibt deshalb einen Referenzpunkt an:

> „Movements of the tongue body are described in relation to a fixed point within the body of the tongue. The point chosen was the reference point sometimes used in X-ray measurements of lingual dimensions [...]“

Wollte man phonetische Merkmale auf diese Weise mit Bezug auf einen Startzustand definieren, müsste man alle möglichen lautlichen Umgebungen und damit alle möglichen Startzustände, die im Vollzug der sprachlichen Produktion entstehen können, einbeziehen, was sicherlich unrealistisch ist. Der von Hardcastle angenommene einheitliche Referenzpunkt könnte höchstens dann einen Wert haben, wenn zwischen den für die Lautrealisation erforderlichen Bewegungen jeweils eine neutrale Stellung der Motorik eingenommen würde, was grundsätzlich nicht der Fall ist.

Merkmalsdefinitionen müssen offenbar artikulatorische Zielzustände beschreiben, es muss in einer Merkmalsdefinition (weitgehend) offen bleiben, wie diese Zustände erreicht werden. Beschreibungen einzelner konkret realisierter artikulatorischer Bewegungen in einem konkreten Sprechvorgang können sehr verschieden aussehen, die von der phonetischen Kompetenz definierte Steuerung wird dagegen eine wesentlich geringere Variabilität aufweisen. Das ist letztlich auch das Konzept der klassischen artikulatorischen Merkmale. Es ist darüber hinaus fraglich, ob man (jedenfalls für das Deutsche) mit geregelten, das heißt *durch spezifisch sprachliche Festlegungen gesteuerten* Zustandsabfolgen unterhalb der Phonemebene (Zustandsabfolgen, die innerhalb einer Zeitspanne von größenordnungsmäßig weniger als 50 ms unterkommen), rechnen muss, und damit tatsächlich mit Repräsentationen, die die Bezeichnung „motorisches Programm“ im Sinne einer zeitlichen Abfolge motorischer Befehle verdienen. Affrikaten und Diphthonge sind keine Belege dafür, sondern müssen als Abfolgen von zwei Phonemen interpretiert werden, es handelt sich nicht um Abfolgen artikulatorischer Zustände innerhalb eines Phonems. Bei Verschlusslauten wie dem deutschen [t] z. B. in intervokalischer Position kann die Abfolge Verschluss-Verschlusslösung-Behauchung auf eine Qualität des Verschlusses zurückgeführt werden, sie muss nicht als Abfolge innerhalb eines motorischen Programms verankert sein. Die Verschlusslösung kann als Folge der Bewegungsziele des folgenden Vokals zustande kommen.

Die komponentielle Definition von Phonemen kann nicht in Zweifel gezogen werden. Ebensowenig die hierarchische Struktur der Komponenten. Die Komponenten müssen, wenn man die vorangegangenen Überlegungen akzeptiert, Bewegungsziele festlegen. Damit ist aber noch wenig gesagt über die weitergehenden Eigenschaften dieser Komponenten und über die Frage, wie die formulierten Anforderungen erfüllt werden. Man kann sich diesem Problem annähern, indem man sich die physiologischen Bedingungen etwas genauer ansieht, unter denen eine sprachliche Produktion an der äußersten Peripherie, das heißt in den beteiligten Muskeln selbst, zustande kommt.

## 3.4.3 Kodierungsprobleme

Sprachliche Produktion bedeutet Aktivierung von Skelettmuskeln.

Grundzüge des Aufbaus von Skelettmuskeln (quergestreiften Muskeln), die in unserem Zusammenhang wichtig sind, sind in Abbildung 3.4.3–1 schematisch dargestellt.

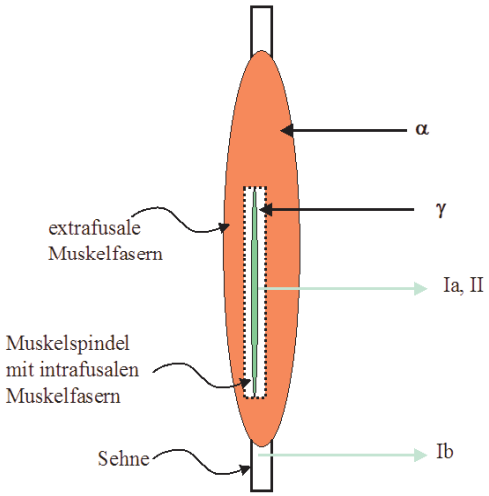



*Abbildung 3.4.3–1:* Struktur eines Skelettmuskels, schematisch. Weitere Erläuterungen im Text.

Muskeln können aktiv nur kontrahiert, nicht gedehnt werden, für die Dehnung ist die Kontraktion eines Antagonisten erforderlich. Von besonderer Bedeutung ist, dass ein Skelettmuskel in einer Kombination von efferent in-

nervierten Strukturen und von Sinnesorganen besteht, deren Zusammenwirken auch für das Verständnis der sprachlichen Produktion beachtet werden muss. Die sehr gründlich untersuchten Prinzipien der eigentlichen Umsetzung von neuronalen Impulsen in die Kontraktionskraft („elektromechanische Kopplung“) sind dagegen für uns weniger relevant.

Man unterscheidet extrafusale und intrafusale Muskelfasern.

- Extafusale Muskelfasern bilden die eigentlich krafterzeugenden Strukturen des Muskels. Die Erregung geschieht durch $\alpha$-Motoneuronen, deren Axone sog. „motorische Endplatten“ auf einzelnen Muskelfasern bilden. Das Endplattenpotenzial ist sehr hoch und löst regelmäßig Aktionspotenziale aus, die sich innerhalb der vielkernigen Muskelzelle ausbreiten und damit Muskelzuckungen hervorrufen. Es gibt nur *eine* Endplatte pro Muskelfaser, es ist keine Summation mehrerer Endplattenpotenziale erforderlich. Die Kraft der Kontraktion kann aber gesteigert werden, wenn über die Endplatte mit ausreichend hoher Wiederholrate mehrere Aktionspotenziale in zeitlicher Folge einwirken. Das ist vergleichbar mit der zeitlichen Summation bei Neuronen. Da die Dauer einer Muskelzuckung, die durch einen einzelnen erregenden Impuls ausgelöst wird, größenordnungsmäßig 100 ms betragen kann, ist diese zeitliche Summation auch für den Bereich der sprachlichen Artikulation interessant.

  Das Axon eines Motoneurons kann mehrere bis viele Muskelfasern innervieren, das Motoneuron bildet mit allen von ihm versorgten Fasern eine sog. „motorische Einheit“. Es ist außerdem nicht anzunehmen, dass eine artikulatorische Einstellung durch die Aktivität eines einzelnen Motoneurons zustande kommt. Motoneuronen für die Artikulation liegen, soweit Muskeln über Hirnnerven versorgt werden, im Hirnstamm.

- Intrafusale Muskelfasern befinden sich innerhalb der sog. Muskelspindeln (Bindegewebskapseln zwischen den extrafusalen Fasern und mit diesen verankert). Die Zahl der Muskelspindeln pro Muskel variiert je nach der Genauigkeit, mit der ein Muskel arbeitet. Die Muskelspindeln dienen nach allgemein akzeptierter Auffassung der Messung der Muskellänge. Man unterscheidet mindestens zwei Typen von intrafusalen Muskelfasern: Kernsackfasern (Nuclear-bag-Fasern, zwei pro Muskelspindel) und Kernkettenfasern (Nuclear-chain-Fasern, bis zu 10 pro Muskelspindel), die offenbar unterschiedliche Funktion haben. Details sind in der Diskussion. Die Abbildung 3.4.3–2 zeigt schematisch eine Kernkettenfaser. Beiden Typen gemeinsam ist, dass sie, efferent innerviert durch $\gamma$-Motoneuronen, an den Polen kontrahieren und über afferente Ia-Fasern Informationen über Länge und Längenänderungen

der Spindel und damit des Muskels abgeben. Die Wirksamkeit der $\gamma$-Motoneuronen weist darauf hin, dass es sich bei der Leistung der Muskelspindeln nicht nur um einen reinen Messvorgang handelt. Wenn man in Handbüchern liest, dass durch den efferenten Einfluss die Empfindlichkeit der Messung eingestellt wird, muss gefragt werden, mit welcher Funktion das geschieht.

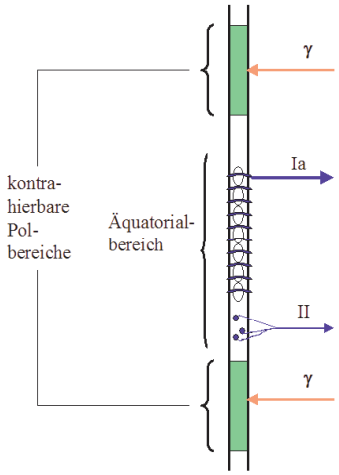



*Abbildung 3.4.3–2:* Struktur einer Kernkettenfaser, schematisch. Der Äquatorialbereich ist in Wirklichkeit nur einige 100 $\mu$m lang, die Gesamtlänge der Faser kann 12 mm erreichen.

Im Unterschied zu den Muskelspindeln haben die „Golgi-Sehnenorgane“ nur sensorische Funktion. Sie liegen am Übergang der Muskelfasern zur Sehne und liefern Informationen über den Spannungszustand des Muskels. Die Ableitung geschieht über Ib-Fasern. Synergistische Motoneuronen werden gehemmt, antagonistische erregt. Motoneuronen, die zu anderen Gelenken gehören, können sowohl erregt als auch, über zwischengeschaltete Interneuronen, gehemmt werden.

Es wird, wenn man die Literatur zur Funktion von Muskeln sichtet, rasch deutlich, dass wesentliche Details gerade der Innervierung von Muskeln nicht vollständig aufgeklärt sind, so dass man sich notgedrungen mit teilweise etwas spekulativen Bruchstücken behelfen muss. Man kann sich als Linguist z. B. ausführliche Handbuchdarstellungen wie die Beiträge von Illert in Deetjen, Speckmann & Hescheler (2005 Hg.: Kapitel 4.3 bis 4.9) ansehen und wird einigermassen enttäuscht sein, was die Besonderheiten der Muskelfunktion in der sprachlichen Produktion angeht.

Für die sprachliche Produktion erforderlich ist nicht nur die Auswahl bestimmter Muskeln (oder Muskelfasern), sondern auch die Einstellung und Beibehaltung einer bestimmten Muskellänge. Das wird besonders anschaulich am Beispiel der Vokale. Wenn man vereinfacht nur den Muskel betrachtet, der den Kiefer anhebt (Musculus masseter), gilt für ein deutsches [ɑː] eine größere Länge des Muskels, für [eː] eine mittlere Länge und für [iː] eine nicht maximale, aber doch relative Kürze. Die unterschiedlichen Längen gelten einzelsprachlich und müssen Gegenstand von Lernprozessen sein.

Der Auswahl des Muskels entspricht die Auswahl der entsprechenden $\alpha$-Motoneuronen und man könnte vielleicht zunächst annehmen, dass die Muskellänge über die Feuerfrequenz dieser Neurone eingestellt wird. Es ist aber zu beachten, dass diese Frequenz auch lastabhängig ist (von den Sehnenorganen gesteuert). Die Kraft des Muskels ist abstufbar über Rekrutierung zusätzlicher motorischer Einheiten und eben auch über die Frequenz der Aktionspotenziale der Motoneuronen. Wenn man zusätzlich durch eine spezifische Einstellung eine bestimmte Muskellänge erreichen möchte, kommen nur die intrafusalen Muskelfasern, besonders wohl die Kernkettenfasern, als wesentliche Steuerelemente in Frage. Sie werden, wie oben schon erwähnt, durch $\gamma$-Motoneuronen erregend innerviert und kontrahieren dadurch an den Polen (vgl. Abbildung 3.4.3–2). Solange der Muskel eine Länge hat, die zur Dehnung des Äquatorialbereichs der so auf einen Sollwert eingestellten intrafusalen Muskelfasern führt, geben die Ia-Afferenzen Aktionspotenziale ab, die über die entsprechenden $\alpha$-Neuronen zur Verstärkung der Muskelkontraktion und damit zu einer Entspannung der intrafusalen Fasern führen. Zusätzlich werden über Kollateralen Ia-Interneurone aktiviert, die Motoneuronen der antagonistischen Muskeln hemmen.

Das Problem der Einstellung der Muskellänge verschiebt sich dadurch auf die Frage, wie die intrafusalen Fasern durch eine entsprechende Kontraktion auf die Sollwerte eingestellt werden können. Da die Aktionspotenziale der $\gamma$-Motoneurone, wie generell im Nervensystem, nicht gradiert sind, kommt dafür nur die Frequenz innerhalb eines Impulsbursts in Frage. Anders als bei den $\alpha$-Motoneuronen ist diese Frequenz bei den $\gamma$-Motoneuronen nicht lastabhängig. Die dadurch an die Muskelspindel übermittelte Information (also entweder der Burst oder seine Wirkung) kann unverändert erhalten bleiben, solange die Einstellung der Muskellänge andauern soll. Die Feuerfrequenz der $\gamma$-Motoneuronen ist also geeignet, einen Sollwert für die Muskellänge an die Muskelspindeln zu übertragen.

Die Einstellung der Muskellängen für artikulatorische Merkmale muss pro Laut und also mit einer entsprechenden Geschwindigkeit erfolgen. In diesem Zusammenhang ist dann zusätzlich wichtig, zu beachten, dass nach den

Angaben von Matthews (1981: 209) Nuclear-chain-Fasern Feuerfrequenzen von mehr als 50 Hz brauchen, um zu kontrahieren, und dass oberhalb 150 Hz keine weitere Kontraktion erfolgt. Nuclear-bag-Fasern kontrahieren schon oberhalb 10 bis 20 Hz, die Obergrenze liegt bei 75 Hz.

Aus solchen Daten ergibt sich zwingend, dass die einzelnen Impulse, die für die Kodierung der sprachlichen Information bei lexikalischen und phonologischen Prozessen anzunehmen sind und typischerweise mit Frequenzen unterhalb 20 Hz auftreten, nicht ohne eine Umwandlung zur Steuerung der Muskellänge dienen können. Ein einzelner Impuls, der von einer Zelle produziert wird, die ein Phonem oder ein phonologisches Merkmal repräsentiert, muss auf dem Weg zur motorischen Peripherie auf spezialisierten Zellen Bursts auslösen, die die entsprechende Muskeleinstellung zur Folge haben.

Der Produktionsvorgang verläuft in dieser Hinsicht spiegelbildlich zum Perzeptionsvorgang. In der Perzeption muss frequenzkodierte Information zu einzelimpulskodierter umgewandelt werden, in der Produktion verläuft dieser Prozess umgekehrt.

Es sind sicherlich verschiedene Lösungen für einen entsprechenden Umwandlungsprozess denkbar. Da Lernprozesse einbezogen werden müssen, kann man sich vorstellen, dass im Spracherwerb eine geeignete Auswahl unter Zellen erfolgt, die durch angeborene strukturelle Eigenschaften die gewünschten Bursts produzieren. Die anzunehmenden strukturellen Eigenschaften müssen sich nicht weit von dem entfernen, was sonst für Neuronen z. B. im Kortex gilt.

Das folgende Simulationsexperiment zeigt, wie man sich den Vorgang einer Erzeugung spezifischer Bursts denken kann. Es werden fünf Neuronen dargestellt, die jeweils durch einen einzelnen Impuls aktiviert werden, und aufgrund dieses Inputs einen Burst abgeben, dessen Frequenz durch die Länge von Fasern eingestellt ist, die vom Axon auf das Neuron zurückführen (ggf. über zwischengeschaltete Neuronen) und dessen Dauer (hier als Anzahl erzeugter Impulse festgelegt, was eine entsprechende Abstimmung der Zellparameter voraussetzt, andere Festlegungen sind möglich) durch zeitliche Summation dieser Impulse auf einem hemmenden Neuron bestimmt wird. Die Abstände zwischen den Aktionspotenzialen innerhalb der Bursts betragen jeweils: 6, 7, 8, 9 und 10 Zeittakte, es werden, wenn ein Zeittakt mit einer Millisekunde veranschlagt wird, also Frequenzen von 167, 143, 125, 111 und 100 Aktionspotenzialen pro Sekunde erzeugt.

Da die Bildschirmdarstellung durch Überlagerung von Verbindungen nicht ganz optimal ist, wird die Anordnung in Abbildung 3.4.3–3 in verdeutlichter Form wiedergegeben.

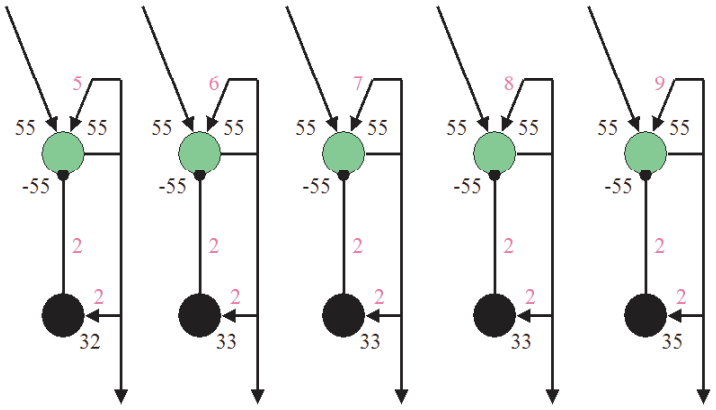



*Abbildung 3.4.3–3:* Architektur zur Erzeugung von Impulsbursts verschiedener Frequenz, aber gleicher Impulszahl. Rote Zahlen geben Leitungsverzögerungen in Zeittakten an, schwarze Zahlen die Synapseneffektivität mit Bezug auf einen angenommenen Schwellenwert von 50 und ein Ruhepotenzial von 0.

Simulation:
Erzeugung von Impulsbursts unterschiedlicher Frequenz.
Jede der erregenden Zellen erhält im Zeittakt 1 und im Zeittakt 100 einen Input durch ein einzelnes Aktionspotenzial mit überschwelliger Wirkung.
1. Bildschirm mit Darstellung der Architektur.
2. Bildschirm mit Darstellung der Impulsbursts.

Wenn man annimmt, dass die Burstfrequenzen über Faserlängen eingestellt werden, bedeutet das für den Spracherwerb (und andere motorische Lernprozesse), dass eine entsprechend große Anzahl von Zellanordnungen mit unterschiedlichen Faserlängen zu einer angeborenen Ausstattung gehört und dass die Lernprozesse in der Auswahl geeigneter Zellanordnungen bestehen. In Produktionsrichtung kann das nur als „Lernen durch Vergessen“ (vgl. oben 3.3.5) gedacht werden und muss ein Versuchs-Irrtums-Prozess sein. Das gilt jedenfalls für alle Fälle, bei denen motorische Einstellungen nicht als Ganzes angeboren sind.

Die von den Muskelspindeln ausgehenden Ia-Afferenzen, die für die Anpassung der Muskellänge an den Sollwert zuständig sind und Synapsen auf den $\alpha$-Motoneuronen bilden, genügen allein nicht, um die $\alpha$-Motoneuronen überschwellig zu erregen. Man kann also nicht davon ausgehen, dass es für die Artikulation ausreicht, die Muskellänge einzustellen und den Rest den vorhandenen Reflexschaltungen zu überlassen. Gleichzeitig muss eine ent-

sprechende efferente Erregung die $\alpha$-Motoneuronen mindestens unterschwellig depolarisieren. Sowohl die Erregung der $\alpha$- Motoneuronen als auch die der $\gamma$-Motoneuronen führen zur Auswahl der jeweils für einen Artikulationsvorgang erforderlichen Muskeln. Die Einstellung der $\alpha$-Motoneuronen muss nicht notwendig frequenzkodiert erfolgen. Die Dauer einer Muskelzuckung ist ausreichend lang, sie kann größenordnungsmäßig 100ms erreichen, so dass die von den höheren Prozessen generierten Frequenzen, ergänzt durch den lastabhängigen Einfluss der Sehnenorgane, prinzipiell ausreichen, um eine adäquate Muskelleistung zu erzielen.

Die Ansteuerung von Muskeln ist offenbar grundsätzlich (und nicht nur sprachlich) tatsächlich so, dass ein Zielzustand vorgegeben werden kann, der eine vom Startzustand her beeinflusste Reaktion zur Folge hat. Die Anpassung an den Startzustand muss nicht in die Definition des Zielzustands einbezogen werden. Das oben in 3.4.2 angesprochene Problem der Definition zustandsabhängiger artikulatorischer Bewegungen entsteht nicht.

### 3.4.4 Folgerungen für artikulatorische Merkmalsinventare

Nachdem Vorstellungen über die Natur von Lexikonrepräsentationen einerseits und über die Anforderungen aus gezielten Muskelbewegungen andererseits existieren, kann man nun versuchen, die Lücke zwischen diesen beiden Polen wenigstens spekulativ zu schließen. In diesen Bereich müssen Repräsentationen artikulatorischer Merkmale fallen.

Es ist zweifellos so, dass ein- und dasselbe artikulatorische Merkmal zur Artikulation mehrerer verschiedener Phoneme einer Sprache dienen kann. Das ist einer der wesentlichsten Anhaltspunkte für die Annahme artikulatorischer Merkmale überhaupt. Komponenten, für die das gilt, sollten in der Produktion direkt von den Phonemrepräsentationen aus erreicht werden.

Man kann nun weiter überlegen, ob die Muskellänge und damit die Einstellung der intrafusalen Muskelfasern über die $\gamma$-Motoneuronen eine solche Komponente sein könnte. Diese Einstellung impliziert allerdings an der äußersten Peripherie immer die gleichzeitige Auswahl eines Muskels bzw. einer Muskelgruppe. Die Auswahl des Muskels bzw. der Muskelgruppe durch die Aktivierung der entsprechenden $\gamma$-Motoneuronen ist aber identisch mit der Auswahl über die $\alpha$-Motoneuronen. Wenn man nicht annehmen möchte, dass die für die Einstellung *verschiedener* Muskeln erforderlichen $\gamma$-Frequenzen gleich sind, muss eine entsprechende Zuordnung erfolgen. Das bedeutet, dass $\alpha$-Aktivierung und $\gamma$-Aktivierung durch dieselbe Quelle (die

dann muskelspezifisch ist) ausgelöst werden muss. Der entsprechende Zelltyp liegt zwischen der Phonemrepräsentation und der äußersten Peripherie, womit also die Frage, ob die Muskellänge eine Komponente mit dem Status eines phonetischen Merkmals sein könnte, negativ entschieden ist.

Wenn man mit der Möglichkeit rechnet, dass das Zusammenwirken von Muskelkombinationen, nicht nur die angeborene Koordination von Agonisten und Antagonisten, aus nicht-sprachlichen Zusammenhängen in den sprachlichen Bereich übernommen werden kann, liegt es nahe, solche Kombinationen (die dann also jeweils spezifische $\gamma$-Aktivationen schon einschließen) als neuronal repräsentiert vorzusehen.

Aus diesen Überlegungen ergeben sich insgesamt hierarchische Strukturen für die Artikulation wie in Abbildung 3.4.4–1 angedeutet.

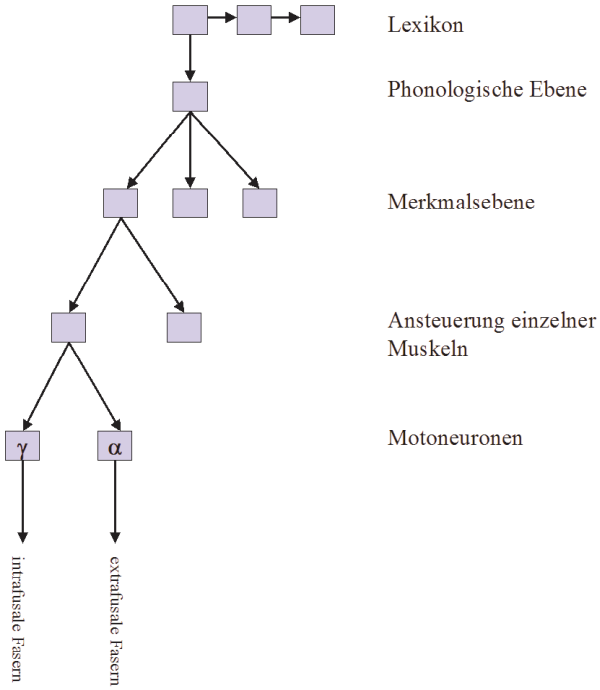



*Abbildung 3.4.4–1:* Hierarchische Struktur artikulatorischer Komponenten. Verzweigende Pfeile deuten die gleichzeitige Erregung mehrerer Komponenten (nicht Alternativen) an. Weitere Erläuterungen im Text.

In der Abbildung werden Rechtecke zur Symbolisierung von Funktionskomponenten verwendet, um anzudeuten, dass nicht notwendig an einzelne Zel-

len zu denken ist. Für alle Komponenten gilt aber das Prinzip der lokalistischen Repräsentation.

Die durch Pfeile angedeuteten neuronalen Verbindungen übermitteln bis herunter zu den Motoneuronen jeweils einzelne Impulse pro phonologischer Einheit der Lexikonrepräsentation. Impulsbursts zur Innervierung der intrafusalen Fasern werden erst auf der letzten Verarbeitungsstufe erzeugt. Wenn man das akzeptiert, muss man wohl auch voraussetzen, dass für die erforderlichen Lernprozesse entsprechende Auswahlen von Zellkombinationen schon auf der Ebene der Muskelansteuerung angenommen werden können, so dass tatsächlich, wie oben angenommen, die Muskelansteuerung immer sowohl die Auswahl des Muskels als auch seine Länge bestimmt.

Das Fazit der vorangegangenen Überlegungen ist, dass phonetische Merkmale immer Zielzustände von Bewegungen zum Gegenstand haben sollten. Diese Zielzustände sind Zustände von Muskelkombinationen, nicht von einzelnen Muskeln. Der einzelne Muskel (genauer: die einzelne motorische Einheit) wird immer angesteuert durch eine Kombination von $\alpha$-Aktivität und $\gamma$-Aktivität. Diese Kombination muss auch lokalistisch durch eine Großmuttereinheit repräsentiert sein. Wenn Bewegungsziele erreicht sind, führt eine erneute Aktivierung zur Beibehaltung einer entsprechenden Muskelstellung. Das mag auch interessant sein für das Verständnis von Beobachtungen im Rahmen von Modellen, die mit Partituren artikulatorischer Gesten arbeiten (zur Koartikulation vgl. unten 3.5.2).

Artikulatorische phonetische Merkmale sollten unter diesen Gesichtspunkten nicht einfach zur bequemen Unterscheidung von Lautklassen dienen, sondern eine neuromuskuläre Interpretation haben. Wenn man die Reihe der üblicherweise unterschiedenen Merkmale daraufhin überprüft, kann man z. B. das Merkmal „Verschlusslaut“ („Plosiv“) als in diesem Sinne nicht interpretierbar ausscheiden, da die Verschlüsse durch verschiedene jeweils spezifische Muskelkombinationen zustande kommen. Analoges gilt für „Frikativ“.

Eine Charakterisierung „bilabialer Verschluss“ ist korrekt und beschreibt eine Einheit, nicht etwa eine Kombination der Merkmale „bilabial“ und „Verschluss“. Dasselbe gilt für „bilabialer Frikativ“. Konsequenz ist, dass auch „bilabial“ als Merkmal, weil verschiedene Muskelansteuerungen impliziert sind, nicht den Kriterien entspricht, sondern eine Muskelauswahl beschreibt, ohne zugleich z. B. die Muskellänge zu bestimmen. Dieselbe Argumentation kann auf „dental“, „alveolar“, „palatal“ usw. angewandt werden. Man kann solche Merkmale nur beibehalten, wenn man akzeptiert, dass die Bezeichnungen für artikulatorische Merkmale sich auf unterschiedliche neuronale Ebenen beziehen oder die Information über die Muskellänge einschließen.

Einige andere Fälle sind weniger problematisch: Solange „stimmhaft“ als Beteiligung der Stimmlippen des Kehlkopfs in einer bestimmten Form definiert wird, ist dieses Merkmal problemlos. (Flüstern entsteht durch Blockade der entsprechenden Stimmlippenfunktion unterhalb der Merkmalsebene.) Andere analoge Fälle sind Merkmale wie „nasal“ und „lateral“.

Insgesamt ergibt sich, dass die Orientierung an den neuronalen Anforderungen zu größerer Präzision in der Verwendung von Merkmalsbezeichnungen zwingt. Das könnte zur Folge haben, dass man, je nach dem Zweck, dem solche Bezeichnungen dienen, bei der Beschreibung von artikulatorischen Eigenschaften prinzipiell mit der hierarchischen Struktur der Produktionsarchitektur rechnen muss und jeweils spezifizieren muss, auf welcher Ebene man sich mit einer Bezeichnung bewegt.

### 3.4.5 Spezielle Komponenten artikulatorischer Prozesse

*Reafferenzen*

Zu den Standards von Übersichtsdarstellungen zur phonetischen Produktion gehört der Hinweis auf die Bedeutung von Reafferenzen.

Es empfiehlt sich, die sog. „Propriozeption“ davon abzugrenzen, da sie größtenteils identisch ist mit Komponenten der Muskelsteuerung selbst, wie sie oben in Abschnitt 3.4.3 dargestellt worden sind. In einem Beitrag von K. Meßling zu dem Handbuch von Klinke, Pape & Silbernagl (2005) wird Propriozeption so definiert:

> „Unter Tiefensensibilität (Propriozeption) versteht man die Empfindungen aus den tieferen Geweben, den Muskeln, Sehnen, dem Bandapparat und den Gelenkkapseln. [...] Propriozeptoren sind die Muskelspindeln mit Afferenzen der Gruppe Ia und II und Sehnenorgane mit Ib-Afferenzen, sowie weitere Dehnungsrezeptoren mit myelinisierten Nervenfasern in Bändern und Gelenken. Die Propriozeptoren steuern Reflexe und Bewegungsprogramme, ihre Leistungen führen nur selten zu bewussten Empfindungen.“ (Klinke, Pape & Silbernagl, 2005: 635)

Als Reafferenzen in einem spezieller linguistischen Sinn sind Phänomene zu betrachten, die Produktionsvorgänge an der artikulatorischen Peripherie bzw. deren Störungen an zentralere Steuerungsinstanzen zurückspiegeln. Bekanntestes Beispiel dafür ist das Monitoring des Sprechens über das Gehör.

In Teil 2, „Grundlagen“, Kapitel 2.5 ist außerdem argumentiert worden, dass die Produktion sprachlicher Vorstellungen (das innere Sprechen) eine nicht durch das Gehör vermittelte, aber doch notwendige Form einer Rückspiegelung über sprachliche Perzeptionsbahnen erfordert.

Diese Form der Rückspiegelung beim inneren Sprechen ist deshalb hier besonders interessant, weil sie wirklich eine Reafferenz darstellt, die ein Bahnensystem verwendet, das quasi parallel zu den Produktionsbahnen verläuft. Details dazu sind schon oben in 3.3.5 und in Abbildung 3.3.5–2 dargestellt. Dort ist allerdings der Bereich der Motorik ausgeklammert. Der Rückspiegelungsvorgang, soweit dort skizziert, beschränkt sich auf Strukturen, die der Sprachperzeption zuzuordnen sind.

Damit erhebt sich hier, im Zusammenhang mit der Klärung der Motorik, die Frage, ob mit Reafferenzen zu rechnen ist, die sozusagen den motorischen Prozess parallelisieren und damit eine Rückmeldung an die zentrale Steuerung liefern. Eine solche Rückmeldung könnte z. B. eine Funktion für die Kontrolle des Ablaufs lexikalischer Phonemsequenzen haben. Das ist auch eine der Funktionen des Monitoring beim inneren Sprechen.

In Teil 2, „Grundlagen“, Kapitel 2.5 wird gezeigt, dass wegen unüberwindlicher Schwierigkeiten beim Verständnis geeigneter Lernprozesse die erforderlichen neuronalen Verknüpfungen zwischen der Merkmalsebene und der Phonemebene nicht hergestellt werden können. Wenn man sich die Bedingungen näher ansieht, die für die Steuerung lexikalischer Phonemsequenzen gelten, ergibt sich ein weiteres Argument, das aufgrund von Schwierigkeiten mit den zeitlichen Abläufen ebenfalls gegen die Annahme einer motorischen Rückspiegelung spricht. Die Steuerung lexikalischer Phonemsequenzen setzt voraus, dass ein Rückspiegelungsimpuls, der spezifisch für einen produzierten Laut ist, die Lexikonstrukturen erreicht, ehe der folgende Laut produziert wird. Im Fall eines Verschlusslauts vor Vokal kann ein solcher lautspezifischer Impuls über eine motorische Rückmeldung aber letztlich erst unmittelbar vor oder gar nach dem Start der Vokalartikulation produziert werden. Dieses Steuerungsproblem kann nur gelöst werden, wenn anstelle der motorischen Rückspiegelung mit derselben Funktion die Rückspiegelung über die Perzeptionsverbindungen erfolgt. Ein Nebeneffekt ist, dass die Motorik beim inneren Sprechen gehemmt werden muss. Das geschieht offenbar nicht immer vollständig, sondern nach Muskelgruppen abgestuft (immer Atmung, Kehlkopf, teilweise Kiefer, noch seltener Zunge), daher muss die Hemmung innerhalb der motorischen Strukturen relativ peripherienah eingreifen.

Man sollte, wenn man von Reafferenzen spricht, Phänomene, die einfach begleitende und verhaltensbestimmende Afferenzen sind (auch wenn sie sprachliche Konsequenzen haben), nicht gleichrangig mit den eigentlichen

Reafferenzen behandeln, die spezifisch sprachliche Steuerungsfunktionen haben. Ein zugegebenermaßen krasses Beispiel für einen Einfluss verhaltensbestimmender Afferenzen auf die sprachliche Produktion wäre z. B. der Einfluss von Zahnschmerzen, die bei Kältereiz verstärkt werden, auf die Bereitschaft zur ausreichenden Öffnung des Kiefers. Auch Afferenzen über die Sinnesorgane der Haut, die bewusstwerdende Empfindungen erzeugen, gehören in diese Rubrik.

Wenn man solche Beispiele ausschließt, wird die neuronale Grundlage von sprachlichen Reafferenzen durch das Schema der Abbildung 3.4.5–1 vollständig erfasst.

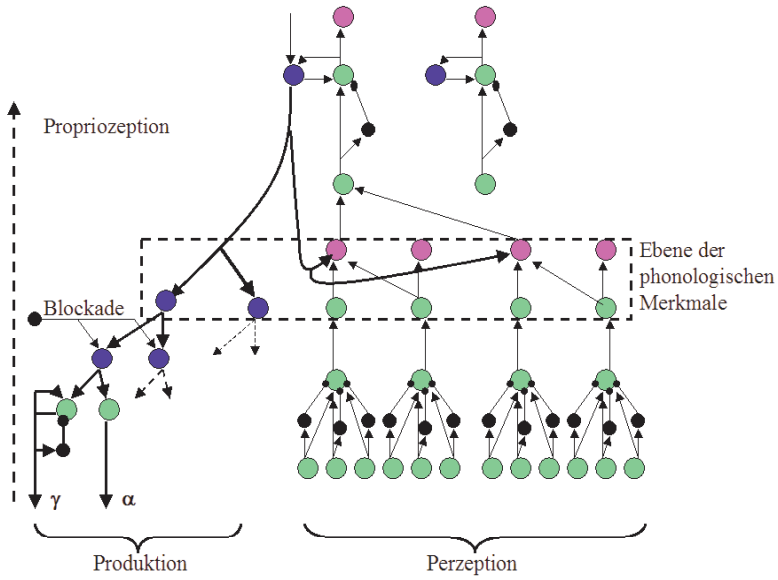



*Abbildung 3.4.5–1:* Schema zur Rückspiegelung und Propriozeption; Erweiterung des Schemas der Abbildung 3.3.5–2 für den Bereich der Produktion. Peripherienahe Reflexschaltungen sind weggelassen. Die Produktionsstrukturen sind der Übersichtlichkeit halber nicht vollständig ausgeführt und sind an der Stelle der gestrichelten Pfeile analog ergänzt zu denken.

Rückspiegelungsvorgänge mit spezifisch sprachlicher Funktion sind auf den Bereich von Strukturen beschränkt, die der Sprachperzeption zuzuordnen sind oder die mit Bezug auf die Unterscheidung von Artikulation und auditiver Wahrnehmung neutral sind. Die motorischen Steuerungsvorgänge werden nicht durch sprachspezifische, der Motorik zuzuordnende Bahnen über-

wacht. Daher entsteht das Schema der Abbildung 3.4.3 durch eine einfache Verknüpfung der Informationen aus den Abbildungen 3.3.5–2 und 3.4.4–1.

*Die Rolle der Basalganglien und des Kleinhirns*

Die Bezeichnung „Basalganglien“ bezieht sich auf eine Gruppe aus mehreren Kernen, die teilweise an der motorischen Steuerung, teilweise an der Sensorik beteiligt sind.

> „Das Striatum setzt sich als ein funktionell einheitliches System aus dem Nucleus caudatus und Putamen zusammen. Die beiden Segmente des Globus pallidus, Pars interna und Pars externa, sind an unterschiedlichen Funktionen beteiligt. Weiter gehören der Nucleus subthalamicus und die Substantia nigra, letztere mit den funktionell unterschiedlichen Kompartimenten, Pars compacta und Pars reticulata, zu den Basalganglien.“ (Schmidt, 1995 Hg.: 134; Hervorhebungen getilgt)

Das Striatum erhält als Eingangsstruktur der Gruppe Input vom zerebralen Kortex, die Ausgangskerne Globus pallidus, Pars interna und Substantia nigra, Pars reticulata haben hemmende Verbindungen mit den ventralen Kernen des Thalamus und beeinflussen auf diese Weise die Erregung des Kortex durch den Thalamus (sog. „kortiko-thalamo-kortikale Schleife“).

Bei Erkrankungen der Basalganglien ergeben sich unter anderem teilweise dramatische Bewegungsstörungen: Überschießende unkontrollierte rasche Bewegungen einerseits und Verlangsamung, erhöhte Muskelspannung und Ruhetremor andererseits. Das Standardbeispiel für eine Erkrankung, bei der es zu einer Verlangsamung der Bewegungsdurchführung kommt, ist der Morbus Parkinson. Es ist auch so, dass dabei eine Verlangsamung des Sprechens beobachtet werden kann (Bradylalie), die aber nicht im Vordergrund steht. Einiges Interesse haben auch Störungen der Prosodie (reduzierte Lautstärke, Beibehalten der gleichen Tonhöhe) gefunden (vgl. insgesamt Böhme, 2003: 360 f.).

Wenn man unterscheidet zwischen Grundfunktionen der sprachlichen Produktion und zusätzlichen Funktionen, die, wenn sie wegfallen, nicht zu grundsätzlichen Verständigungsschwierigkeiten führen, muss man schließen, dass die Basalganglien, trotz ihrer Bedeutung für die Bewegungssteuerung allgemein, offenbar keine grundsätzliche Funktion für die Realisierung lautlicher Kategorien haben. Sie sind also für eine linguistische Argumentation eher zweitrangig.

Bei Störungen des Kleinhirns, speziell des Cerebrocerebellums wird die Bewegungskoordination beeinträchtigt und die Sprache wird als „häufig eintönig und skandierend“ beschrieben (Schmidt, 1995 Hg.: 142). Auch hier gilt offenbar, dass die Realisierung lautlicher Kategorien nicht grundsätzlich gestört ist.

Prinzipiell gilt also, dass die komplizierten Steuerungsfunktionen der Basalganglien und des Kleinhirns für die sprachliche Artikulation, jedenfalls so lange man die Prosodie ausklammert und an der Realisierung von Sprachlauten als segmentalen Bestandteilen von lexikalischen Einheiten interessiert ist, praktisch irrelevant sind.

# 3.5 Phonologische Regularitäten

## 3.5.1 Zwischen Phonetik und Phonologie

Es ist unbestritten, dass auf lautlicher Ebene in natürlichen Sprachen Regularitäten beobachtet werden können, deren Beschreibung durch Regeln unter ausschließlicher Verwendung phonetischer und phonologischer Kategorien möglich ist, also ohne dass die morphologische oder die syntaktische Ebene notwendig einbezogen werden müssen. Wenn eine Beschreibung durch Angabe von Regeln möglich ist, kann daraus allerdings nicht zwingend geschlossen werden, dass solche Regeln eine neuronale Repräsentation als isolierbare Verarbeitungskomponenten im Nervensystem haben müssen. Die diesbezügliche Problematik ist teilweise schon oben in Abschnitt 3.1.3 besprochen worden. Man beachte zusätzlich, dass es eine der wichtigsten Erkenntnisse aus der Konstruktion konnektionistischer Modelle (z. B. der Modellierung des Erwerbs von Vergangenheitsformen englischer Verben in Rumelhart & McClelland, 1986b, oder des Modells TRACE in McClelland & Elman, 1986) ist, dass solche Modelle, ob sie nun als realistisch gelten können oder nicht, regelhaftes Verhalten zeigen, ohne dass Neuronengruppen im System benannt werden könnten, die ausschließlich für dieses Verhalten verantwortlich wären.

Einiges, was die Formulierung von Regeln veranlasst hat, ergibt sich aus elementaren Bedingungen der Muskelfunktionen und der Schallproduktion in der Artikulation. Ein Beispiel ist die Artikulation von [p] ohne Behauchung vor [f]. Solche Fälle werfen die Frage auf, ob man überhaupt von Regeln im phonetischen (nicht phonologischen) Bereich sprechen möchte. Es ist jedenfalls schwierig, von Regeln zu sprechen, solange es um angeborene (vererbte) unveränderliche Eigenschaften geht, die beobachtbare Konsequenzen haben. Damit wird aber die Regelbildung in einen Bereich verwiesen, in dem produktive Lernprozesse (nicht nur Lernen durch Vergessen) möglich sind. Das wiederum heißt, dass neuronale Repräsentationen von Regeln im Kortex

anzusiedeln sind, wo entsprechende strukturelle Voraussetzungen angenommen werden können.

Es entspräche auch nicht dem üblichen Regelbegriff, wenn man die phonologische Kategorienbildung durch Kombination auditiv-phonetischer Merkmale, die im Kortex stattfindet, generell als Wirkung einer Regelanwendung betrachten würde. Man könnte sich aber vorstellen, dass die Merkmale selbst, vor der Verknüpfung in ein Phonem, aber schon auf phonologischer Ebene, an bestimmte Kontexte gebunden sind. Es ergibt sich daraus die Frage, ob man mit *phonologischen* Merkmalen (die nicht Phoneme sind), die der in Kapitel 3.1.2 eingeführten Grenzziehung zwischen Phonetik und Phonologie entsprechen, überhaupt rechnen darf. Solche Merkmale müssten definitionsgemäß neutral sein gegenüber Produktion und Perzeption, sie müssten durch Lernprozesse zustande kommen, die eine Brücke zwischen den zunächst erworbenen auditiv-phonetischen Merkmalen und den entsprechenden artikulatorischen Äquivalenten schlagen müssten. Wegen der Verschiedenheit und Unmöglichkeit der 1-zu-1-Zuordnung von auditiven und artikulatorischen Merkmalen sind solche Lernprozesse aber nicht zu erwarten. Man kann also schlussfolgern, dass die Phonologie tatsächlich mit dem Phonem beginnt, also nach der Kategorienbildung durch Kombination *phonetischer* Merkmale in Perzeptionsvorgängen.

Eine (vielleicht unangenehme) Folgerung aus solchen Überlegungen ist, dass Regeln im lautlichen Bereich ausschließlich auf Phonemebene repräsentiert zu denken sind. Regularitäten können nicht an Merkmalskontexte, sondern nur an Phonemkontexte gebunden sein, auch dann, wenn sie sich schriftlich mit Bezug auf einen Merkmalskontext einfacher formulieren lassen. Man vergleiche in diesem Zusammenhang auch die Überlegungen zur vertikalen Einheit von Phonemen oben in Kapitel 3.2.1. Damit gilt auch, dass phonologische Regeln immer ersetzt werden können durch entsprechende lexikalische Verteilungen, womit also wieder die oben angesprochene Beobachtung der Konnektionisten relevant wird.

Letzteres ist auch wichtig zu beachten, wenn man sich kritisch mit den Vorstellungen der generativen Phonologie beschäftigt, insbesondere der Annahme, dass es zwei Repräsentationsebenen für lautliche Sequenzen gibt, die „zugrundeliegende Form“ und die „phonetische Form“, wobei dann die Funktion von Regeln darin besteht, die phonetische Form, ggf. über mehrere dazwischenliegende Repräsentationen (vgl. dazu auch oben Abschnitt 3.1.1), aus der zugrundeliegenden Form abzuleiten. Der „Prozess“ der Ableitung muss auf Strukturen gedacht werden, die sowohl vom Perzeptions- als auch vom Produktionsvorgang benutzbar sind und es muss sehr genau überlegt werden, ob die zugrundeliegenden Formen nicht viel einfacher durch

Repräsentationen der „phonetischen Formen“ als phonologische Formen im Lexikon ersetzt werden können. Das Hauptproblem der zugrundeliegenden Formen ist, wie schon in Abschnitt 3.1.3 angedeutet, das der Lernbarkeit, die in jedem Fall gewährleistet sein muss.

## 3.5.2 Kontextbedingte Varianten

Die Feststellung von Abhängigkeiten lautlicher Erscheinungen vom lautlichen Kontext gehört zu den Standards von Phonetik und Phonologie. Es muss im Zusammenhang mit neuronalen Bedingungen aber auf einige Punkte besonders hingewiesen werden:

- Wegen der unlösbaren Adressierungsprobleme kann es keinen phonologischen (auf Phonemebene angesiedelten) Zwischenspeicher geben, der Sequenzen von mehr als einem lautlichen Element vorhalten könnte. Komplexe Bedingungen sind nur möglich, wenn sie in der Aktivierung einer einzelnen Großmuttereinheit repräsentiert werden können. Man vgl. zum Problem von „Buffers“ den Abschnitt 3.1.3.
- Wirksamer Kontext ist, was eine Lautproduktion oder -Perzeption im Augenblick beeinflusst (vgl. Teil 2, Grundlagen, Abschnitt 2.4.6). Das setzt bei sequenziellem Kontext eine Gedächtnisleistung voraus. Ein größerer sequenzieller Abstand zu einem verursachenden Element (über unbeteiligte Sequenzelemente hinweg) ist unmöglich oder jedenfalls unwahrscheinlich (auch aufgrund von Lernproblemen!).
- Es kann nur ein vorausgehendes ein folgendes lautliches Ereignis beeinflussen, nicht umgekehrt (Kausalität!). Das erschwert, zusammen mit der Schwierigkeit von Pufferspeichern, die Vorstellung von der Abhängigkeit einer phonetischen bzw. phonologischen Variante vom Folgekontext. Es ist aber zu beachten, dass die laut werdende sprachliche Produktion auch durch inneres Sprechen vorbereitet werden kann. Die dabei gebildeten Gedächtnisspuren sind aber oberhalb der phonologischen Ebene angesiedelt.
- Regeln können eine Ersparnis an lexikalischen Repräsentationen zur Folge haben (Beispiel: Auslautverhärtung im Deutschen, siehe unten). Das Bedürfnis einer Ersparnis allein kann aber nicht die Bildung von Regeln auslösen. Speicherplatzprobleme sind für die besondere Ausformung lexikalischer Repräsentationen nicht entscheidend.
- Koartikulation ist ein phonetisches Problem. Sie entsteht dadurch, dass Muskelstellungen nicht explizit durch Änderung der zentralen Befehle

(einschließlich der Aktivierung von Antagonisten) aufgehoben werden, sondern dass die Aktivierungen einfach abklingen. Der Prozess des Abklingens kann in die Realisierung des nachfolgenden Lauts hineinreichen. Koartikulation in diesem Sinne muss also nicht durch eigens repräsentierte phonologische Regeln bewirkt werden.

In Abschnitt 3.1.3 oben ist eine neuronale Repräsentation schematisch wiedergegeben, die einer Regel für den Wechsel von [ç] und [x] im Deutschen entsprechen könnte. Eine offengelassene Frage ist die der Lernbarkeit einer solchen Struktur. Die Behandlung dieser Frage setzt voraus, dass man sich genauere Vorstellungen darüber verschafft, welche neuronalen Strukturen im Detail vorausgesetzt werden. Die schematische Darstellung wird zum Vergleich hier als Abbildung 3.5.2–1 wiederholt.

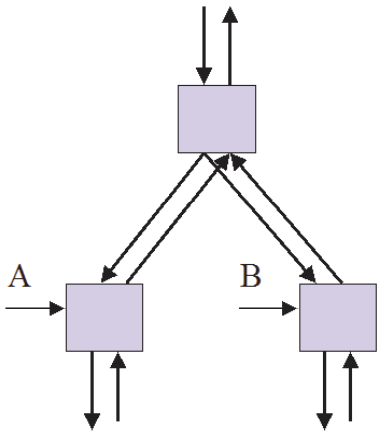



*Abbildung 3.5.2–1:* „Regel“ für eine Kontextabhängigkeit, Wiederholung des Schemas der Abbildung 3.1.3–2

Diese Regelrepräsentation ist in allen Teilen phonologisch, nicht phonetisch, da Verarbeitungsbahnen für Perzeption und Produktion dieselben Einheiten verwenden. Es gilt Einzelimpulskodierung, sodass Lernprozesse mit brauchbaren Ergebnissen jedenfalls prinzipiell ermöglicht werden.

Die Abbildung 3.5.2–2 zeigt jetzt zwei Möglichkeiten einer Präzisierung dieser Struktur, die belegen, dass doch wesentliche Probleme entstehen, wenn man sich an Voraussetzungen orientiert, die biologisch realistisch sind, sich aus verschiedenen Zusammenhängen ergeben und nicht einfach ad hoc zur Realisierung des speziellen Regeltyps angenommen werden. Die Begründung folgt anschließend.

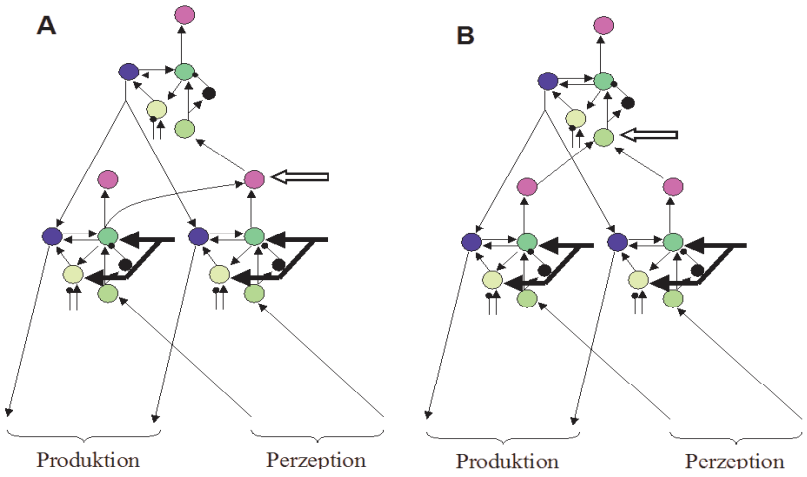



*Abbildung 3.5.2–2:* Lernbarkeitsproblem bei unterschiedlichen Strukturannahmen. Die Lernvorgänge, die an den mit Blockpfeilen gekennzeichneten Zellen erforderlich sind, können nicht erklärt werden. Weitere Erläuterungen im Text.

Die Strukturen der Abbildung 3.5.2–2 enthalten sich wiederholende Anordnungen von Zellen mit unterschiedlichen Eigenschaften. Eine dieser Anordnungen ist in Abbildung 3.5.2–3 noch einmal etwas genauer dargestellt.

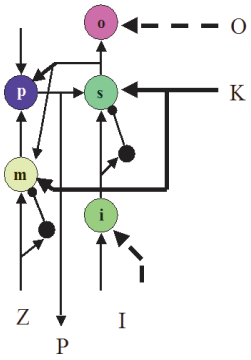



*Abbildung 3.5.2–3:* Zelltypen und typische Verschaltungen. Weitere Erläuterungen im Text.

Die in Abbildung 3.5.2–3 mit Großbuchstaben bezeichneten Verbindungen sind I=Input für die Instanzenbildung, K=Kontextverbindungen, Z=zentra-

le Produktionssteuerung, P=Verbindung zur artikulatorischen Produktion, O=ODER-Verbindung. Die Zellen *p* und *m* sind Zellen der Produktionsarchitektur. Die ausführliche Begründung der Details findet sich in Teil 2, „Grundlagen". Da die Produktionsstrukturen den Erwerb der Perzeptionsstrukturen voraussetzen, ist der letztere Vorgang zunächst für die Frage der Lernbarkeit entscheidend. Es sind die mit *i, s* und *o* bezeichneten Zellen, die an den entsprechenden Lernprozessen beteiligt sind. Die Eigenschaften dieser Zellen sind, soweit hier relevant:

**i** Die Zellen dieses Typs („instanzenbildende Zellen") bilden aus Bündeln von Eigenschaften durch Verstärken vorhandener Synapsen Kategorien der Perzeption. Eigenschaften im Bereich der Phonologie sind (z. B.) phonetische Merkmale, die als Kategorien Phoneme definieren. Die definierenden Eigenschaften müssen in einem relativ kleinen Zeitfenster (<50 ms) zur Verfügung stehen, das EPSP dieser Zellen klingt in diesem Zeitfenster ab. Die Bildung von Kategorien setzt voraus, dass das Feuern einer instanzenbildenden Zelle den Lernvorgang abschließt, also weitere Lernvorgänge nur nach dem Abbau der gebildeten Kategorie möglich sind. (Vgl. Teil 2, „Grundlagen" Kapitel 2.3.)

**s** Die Funktion dieser Zellen in ihrer spezifischen Verschaltung mit den instanzenbildenden Zellen ist die eines Schalters: In einem durch Lernen erreichten funktionsfähigen Zustand feuern diese Zellen nur, wenn sie durch einen vorangegangenen unterschwelligen Input vorbereitet sind. Das durch den vorangegangenen Input bewirkte EPSP dieser Zellen hat eine längere Dauer als das der instanzenbildenden Zellen (ca. 100ms). Diese Zellen können also z. B. zur Darstellung von Phonemabfolgen dienen (daher „sequenzenbildende" Zellen; vgl. Teil 2, „Grundlagen", Abschnitt 2.4.6, und Teil 4, „Lexikon", vor allem Kapitel 4.3). Ein brauchbarer Lernvorgang durch Verstärkung von Synapsen setzt, wie bei den instanzenbildenden Zellen auch, voraus, dass nach dem ersten Feuern der Zelle keine weitere Verbindung verstärkt werden kann.

**o** Während sich die instanzenbildenden und sequenzenbildenden Zellen in ihrem Lernverhalten prinzipiell ähneln, sind die Bedingungen bei Zellen dieses Typs dadurch verändert, dass sie ihrer Funktion nach auch feuern sollen, wenn irgendeiner der einzelnen Eingänge oder auch mehrere Eingänge erregt sind. Das heißt, Lernvorgänge müssen mehrere Eingänge mit überschwelliger Wirkung erzeugen können. Das entspricht einer ODER-Verknüpfung der Zellen. Diese „ODER-Zellen" müssen, um dieser Funktion zu genügen, von vornherein (vor einem Lernprozess) einen überschwelligen Eingang haben. Die Aktivierung dieses (oder ei-

nes anderen bereits funktionsfähigen Eingangs) öffnet ein relativ kurzes Zeitfenster, in dem durch entsprechende Aktivierung weitere Verbindungen verstärkt werden können. (Vgl. Teil 2, „Grundlagen“, Abschnitt 2.4.7.)

Wenn man von dieser Basis aus die für den Erwerb der Regel erforderlichen Lernprozesse überprüft, so macht die Gewährleistung der Kontextabhängigkeit keine Schwierigkeiten. Ein unmittelbar(!) vorausgehender Laut kann die Auswahl eines Folgelauts bestimmen. Das Vorkommen einer bestimmten Abfolge in der Perzeption kann eine entsprechende Synapse auf einer s-Zelle (wegen des zeitlichen Abstands kommt nur dieser Zelltyp in Frage) so lange verstärken, bis diese Zelle ein Aktionspotenzial abgibt. Die Spezifität für einen bestimmten Kontext wird durch das „Abschalten“ der Lernbereitschaft der s-Zelle durch dieses Aktionspotenzial gewährleistet.

Der Sinn der Regel besteht aber nicht nur darin, dass bestimmte Kontextvarianten festgelegt werden, sondern es ist auch ausgesagt, dass diese Varianten auf lexikalischer Ebene nicht mehr unterschieden sind, sondern dieselbe Repräsentation haben. Es wird also für den Perzeptionsprozess eine Struktur vorausgesetzt, bei der von den verschiedenen Varianten aus ein- und dieselbe Repräsentation erreicht wird. Das setzt eine ODER-Verknüpfung von Bahnen voraus, wie das oben in der Abbildung 3.5.2–2 **A** dargestellt ist. Die ODER-Verknüpfung wird dort (naheliegenderweise) durch eine o-Zelle geleistet. Die Lernprozesse, die für diesen Zelltyp angenommen werden müssen, bringen nun aber eine grundsätzliche Schwierigkeit mit sich: Die vorausgesetzte feste Verbindung muss von einer s-Zelle ausgehen, die zur Repräsentation einer der Varianten gehört. Von der jeweils anderen Variante ausgehend muss eine weitere Verbindung verstärkt werden. Die Verstärkung setzt voraus, dass kurz vor der Aktivierung der zu verstärkenden Synapse ein Aktionspotenzial auf der Zelle ausgelöst worden ist. Diese Bedingung kann aber nicht erfüllt werden, solange man annehmen muss, dass dafür die Wahrnehmung der alternativen Variante erforderlich ist. Wenn man an einen zentralen Auslöser des Feuerns der o-Zelle, das für den Lernvorgang erforderlich ist, denkt, verliert man die Spezifität des Lernergebnisses. Eine Struktur wie in Abbildung 3.5.2–2 **A** kann also durch Lernprozesse nicht entstehen.

Als mögliche Lösung des Problems könnte man sich vielleicht auch die Version **B** vorstellen. Die zugrundeliegende Idee ist, die Verknüpfung der Varianten erst der Repräsentation des lexikalischen Phonems zuzuschreiben. Wenn man die Existenz dieser Repräsentation nicht schon vor einem Lernprozess voraussetzen möchte, kommt nur eine Verknüpfung mit einer i-Zelle in Frage. Hier gilt aber, dass zwei überschwellige Verbindungen nicht in einem

zeitlichen Abstand entstehen können, und die zu verknüpfenden Varianten sind nicht gleichzeitig aktiviert.

Das Fazit ist, dass die Strukturen der Abbildung 3.5.2–2 und also Regelrepräsentationen, die dem Schema der Abbildung 3.5.2–1 entsprechen, nicht durch Lernprozesse entstehen können und damit für neuronale Modelle des Sprachverstehens nicht angenommen werden dürfen. Daraus folgt aber, dass auch die Produktion nicht auf solche Regeln zurückgreifen kann.

Wenn man einmal vom Problem der Lernbarkeit für eine Regel absieht, kann man auch darauf hinweisen, dass eigentlich nicht einzusehen ist, warum die Varianten nicht selbst in lexikalische Sequenzen eingebaut werden sollten. Man überlege, ob das grundsätzlich verhindert werden kann, solange man nicht eine generelle Kontrolle einführt, die dafür sorgt, dass lexikalische Strukturen möglichst redundanzfrei sind. Die Redundanzfreiheit ist aber ganz offenbar für sprachliche Repräsentationen kein Wert; andererseits kann eine Kontextabhängigkeit ein grundsätzliches Prinzip sein, das sich durch ganz allgemeine Strukturvorgaben (die sich in den Konstruktionen der Abbildungen 3.5.2–2 und 3.5.2–3 wiederspiegeln) ergibt.

Wenn man aufgrund dieses Arguments nur die Lernvorgänge beibehält, die sich auf Kontexteinflüsse beziehen und zulässt, dass die Produkte direkt als Grundlage für Phoneminstanzen der Lexikonrepräsentationen dienen, ergeben sich Strukturen, die in Kochendörfer (2002: 159) aufgrund des Vergleichs mit den Leistungen von Silbenstrukturen als „Pseudosilben“ bezeichnet worden sind. Wenn man nur vereinfacht die Kontextstrukturen für die Perzeption herauszieht, ergibt sich eine Repräsentation wie in 3.5.2–4.

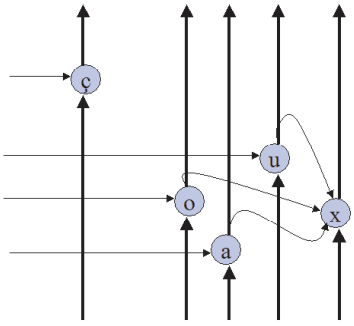



*Abbildung 3.5.2–4:* Beispiel zur Struktur von „Pseudosilben“ übernommen aus Kochendörfer (2002: 160).

Der Zusammenhang von Produktion und Perzeption setzt natürlich pro Phonem bzw. Variante, wenn man es genauer nimmt, Anordnungen wie

in der Abbildung 3.5.2–3 voraus. Die Lernvorgänge, die zu solchen Anordnungen führen, sind in Teil 2, „Grundlagen“, Abschnitt 2.5.3, angegeben.

Varianten sind jetzt in Lexikonrepräsentationen apparativ Phonemen gleichgestellt, und es erhebt sich die Frage, ob die Annahme von Regelrepräsentationen, auch in der „abgeschwächten“ Form der Pseudosilben, dann noch einen Sinn hat. Es sind zwei Gesichtspunkte zu beachten:

- Wahrscheinlich kann die Entstehung solcher Repräsentationen in Spracherwerbsprozessen nicht verhindert werden.
- Man kann erwarten, dass sich, ebenfalls mit Bezug auf Spracherwerbsprozesse, eine Steuerungsfunktion ergibt, die die Bildung von Lautinstanzen, die der Regel entsprechen, begünstigt.

Die Bildung von Regelrepräsentationen ist aber nicht Voraussetzung für eine einzelsprachlich korrekte Produktion und Perzeption auf lexikalischer Ebene.

Damit wird es jetzt auch möglich, scheinbare Abhängigkeiten von *folgendem* Kontext zu verstehen, die, wie oben angesprochen, schwer mit dem Prinzip der Kausalität zu vereinbaren sind. Ein Beispiel ist das Phonem /k/, das im Deutschen unterschiedlich artikuliert werden kann. Es ist im Wort *Kuh* velar, im Wort *Kind* palatal. Die Regelformulierung wird allerdings dadurch erschwert, dass auch Abhängigkeiten vom vorangegangenen Laut zu beobachten sind, z. B. in *dick* vs. *Dock*. Wenn man die Varianten des /k/ als lexikalisch repräsentiert ansieht, verschwindet das Problem.

### 3.5.3 Der Spezialfall der Auslautverhärtung im Deutschen

Im Unterschied zu den im vorigen Abschnitt besprochenen Fällen führt der Verzicht auf eine Regel zur Gewährleistung der Auslautverhärtung im Deutschen zu einer Aufblähung des Lexikons dadurch, dass bei flektierten Wörtern häufig endungslose Formen mit Auslautverhärtung lexikalisch verankert werden müssen.

Es ist in diesem Fall ganz interessant, einige der historischen Hintergründe zu beachten, die sehr ausführlich bei Mihm (2004) dargestellt sind. Mihm stellt fest, dass die Regel der Auslautverhärtung in der deutschen Sprache der Gegenwart auf zwei sprachnormierende Eingriffe zurückzuführen ist, einen ersten 1898 durch die Siebs-Kommission, einen zweiten durch H. de Boor und P. Diels (1957). Natürlich gibt es die Auslautverhärtung z. B.

im Mittelhochdeutschen als historischen Hintergrund, es gibt aber nicht so etwas wie eine kontinuierliche Weiterentwicklung, die konsequent zur gegenwärtigen Regel geführt hätte. Normierungen haben es aber an sich, dass sie (schon definitionsgemäß) weder die Sprachwirklichkeit korrekt wiedergeben, noch in der Sprachwirklichkeit exakt realisiert werden.

Es überrascht, dass selbst die Norm einer relativ komplexen Beschreibung bedarf:

Die entsprechenden Formulierungen in Siebs (1961; 18. Auflage) bzw. Siebs (1969; 19. Auflage) lauten:

> „Das Deutsche kennt im A u s l a u t e i n e r S i l b e o d e r e i n e s W o r t e s k e i n e s t i m m h a f t e n V e r s c h l u ß - u n d R e i b e l a u t e [...]. Tritt ein solcher Laut in den Auslaut, so v e r l i e r t e r d e n S t i m m t o n (A u s l a u t v e r h ä r t u n g): *Tage* aber *Tag* (*ta:k*), *lieben* aber *'lieplıç*, *Löwe* aber *Löwchen* (*'lø:fçən*), *weise* aber *Weisheit* (*'vaeshaet*). Das gleiche geschieht vor einem stimmlosen Konsonanten derselben Silbe: *geben* aber *gibst, gibt* (*gi:pst, gi:pt*).“ (S. 61; wörtlich, ohne die Hervorhebungen durch Sperrung, übernommen in die 19. Auflage, dort S. 84.)
>
> „Alle *b, d, g* im S i l b e n- oder W o r t a u s l a u t sind s t i m m l o s (Auslautverhärtung vgl. S. 61). Sie unterscheiden sich in nichts von stimmlosem *p, t, k,* sind also wie diese behaucht zu sprechen.“ (S. 78; in der 19. Auflage, wieder ohne Sperrung, mit kleinen Änderungen: „Wortende“ statt „Wortauslaut“, „[...] stimmlos und verhärtet[...]“, S. 104.)
>
> „Silbenschließendes *b, d, g* v o r s t i m m h a f t a n l a u t e n d e n A b l e i t u n g s s i l b e n wie: *-lich, -lein, -ling, -nis, -bar, -sam, -sal, -sel* v e r l i e r t ebenfalls d e n S t i m m t o n , ist aber w e n i g e r e n e r g i s c h (lenis) z u v e r h ä r t e n und n i c h t , wie sonst im Auslaut z u b e h a u c h e n : *lieb-lich, Feig-ling, Räd-lein, Erlaub-nis, sag-bar, bieg-sam, red-selig ('li:plıç, 'faeklıŋ, 'rɛ:tlaen, er'laopnıs, 'za:kba:r, 'bi:kza:m, 're:tze:lıç).*“ (S. 78; in der 19. Auflage wörtlich, ohne Sperrung und mit einem überflüssigen Komma, S. 105.)
>
> „In vielen Wortformen stößt silbenanlautendes, stimmhaftes *b, d, g* durch Ausfall eines folgenden Vokals mit *l, n, r* zusammen: *eb(e)nen, üb(e)ler, gold(e)ne, hand(e)le, Wand(e)rer, Wag(e)ner, reg(e)net.* Das kann bei lässigem Sprechen zu verändeter[sic!] Silbentrennung führen, indem das anlautende *b, d, g* in den Schluß der vorderen Sil-

> be hinübergezogen und dadurch stimmlos wird, so etwa: *Wag-ner ('va:k-nər* oder niederdeutsch *va:x-nər*, es *reg-net ('re:k-nət* oder niederdeutsch *'re:çnət)* u. ä. In gepflegter Sprache wird das *b, d, g* in der Regel – unter dem Einfluß verwandter Formen – zur zweiten Silbe gezogen und jedenfalls stets stimmhaft gesprochen: *ir-dne* (nach *irden*), *Bil-dner, Redner* (nach *bil-den, re-den*) und so in *Ordnung, leugne, wandle, edle, Adler, fable, schlendre, andre, Rudrer, Erobrer, weigre* usw. Ebenso in Namen wie *Rabner, Hübner, Bogner, Spindler, Friedrich, Seydlitz, Leibniz, Pegniz.* Bisweilen tritt eine entsprechende Verschiebung der Silbengrenze auch in Fremdwörtern ein: *A-blativ, Se-gment, O-blate, A-gnat*; auch dort ist das *b, d, g* stets stimmhaft zu sprechen.“ (S. 79; in der 19. Auflage mit nebensächlichen orthografischen Veränderungen S. 106.)

Die „gemäßigte Hochlautung“, die neu in die 19. Auflage einbezogen ist, unterscheidet sich von der „reinen Hochlautung“ mit Bezug auf diese Regeln nur gradmäßig (Grad der Stimmhaftigkeit und Grad der Behauchung der Verschlusslaute; vgl. in dieser Auflage S. 110).

Bei Hall (2000: 120) wird die Auslautverhärtung durch die Regel

$$[\text{-son}] \longrightarrow [\text{-sth}] \;/\; \underline{\quad} \;\#$$

bzw. silbenbezogen

$$[\text{-son}] \longrightarrow [\text{-sth}] \;/\; \underline{\quad}\,]_{\sigma}$$

erfasst.

Die Regel soll bewirken, dass alle Obstruenten am Wortende bzw. Silbenende stimmlos werden, soweit sie es nicht schon sind. Diese Regelformulierung ist zwar, an der Norm, nicht an der Sprachrealität gemessen, sicherlich korrekt, aber selbst mit dieser Einschränkung und mit der Beschränkung auf den Fall der Obstruenten unvollständig. Leider sind Vereinfachungen dieser Art in merkmalsorientierten phonologischen Beschreibungen nicht selten. Selbst die Norm ist wesentlich komplexer, von der Sprachrealität ganz abgesehen.

Wenn man einmal die Beschreibung der Norm als Wiedergabe einer Regelkomponente der sprachlichen Kompetenz akzeptiert, kann man, gerade angesichts der Komplexität und der sich daraus ergebenden Skepsis, wieder die Frage der Lernbarkeit stellen. Hier kann man nicht so einfach wie im vorigen Abschnitt auf Vorschläge zur neuronalen Realisierung zurückgreifen. Man überlege aber, wie ein Erstspracherwerbsprozess für ein einzelnes

betroffenes Wort z. B. das Wort *Bad* aussehen müsste, in einer Umgebung, in der reine Hochlautung praktiziert wird.

Das Wort *Bad* wird, entsprechende sachliche und situative Umstände vorausgesetzt, in Dialogen, in die ein Kind einbezogen wird, relativ häufig vorkommen. Es wird im Singular und kaum als Genitiv, folglich (in der Gegenwartssprache) in der Regel endungslos erscheinen, also als [bɑ:t]. Das Kind hat keinen Anlass, es anders als in dieser Form lexikalisch zu verankern, und das würde selbst dann gelten, wenn eine Regel für Auslautverhärtung bereits etabliert wäre.

Die Pluralform ist wesentlich seltener bzw. kommt in der Kommunikation mit kleineren Kindern möglicherweise gar nicht vor. Irgendwann wird das Kind aber auch die Pluralform [bɛ:dər] lernen. Man beachte jetzt, dass Auslautverhärtung wohl bedeuten sollte, dass die zugrundeliegende Form [bɑ:d] ist und im Auslaut das [d] zu [t] verhärtet wird. (Das Besondere an diesem Beispiel ist allerdings, dass im Plural auch noch der Umlaut eingeführt werden muss, was z. B. bei [tɑ:k] vs. [tɑ:gə] nicht erforderlich ist.) Das Kind müsste also den bisherigen Lexikoneintrag [bɑ:t] in [bɑ:d] verändern, obwohl die erstere Form häufig auditiv wahrgenommen wird und dadurch als lexikalische Form gestützt sein sollte. Da die zu revidierende Form unauffällig ist, wird sie auch nicht von den Bezugspersonen zurückgewiesen. Ein Vorgang des Ersatzes ist also extrem unwahrscheinlich (zu möglichen Mechanismen eines solchen Ersatzes vgl. Kochendörfer, 2002: 121 f.). Man wird letztlich nicht umhinkönnen, anzunehmen, dass die Singularform nicht aus einer durch die Pluralform oder gar das Verb *baden* veranlassten zugrundeliegenden Form abzuleiten ist, sondern dass beide Formen eigene lexikalische Repräsentationen haben (womit in diesem Fall auch das Problem des Umlauts verschwindet).

Analoges gilt für viele Singulare mit Auslautverhärtung. Ein Kuriosum ist das Wort *Löwchen*, es kommt wahrscheinlich nur in den Beschreibungen der Norm vor. Die ebenfalls in der Normbeschreibung gespiegelte Idee, das Ausbleiben der Auslautverhärtung in bestimmten Fällen durch eine Verschiebung der Silbentrennung zu erklären, bedeutet Austreiben des Teufels mit Beelzebub. Die in der Regelformulierung des Siebs genannten Beispiele sind außerdem, innerhalb der Gegenwartssprache beurteilt, uneinheitlich.

Bezüglich der erwarteten „Aufblähung“ des Lexikons ist zu beachten, dass Lexikoneinträge, die sich nur am Wortende unterscheiden, auch nur am Wortende sich verzweigende Repräsentationen haben müssen, also nicht der gesamte Eintrag wiederholt vorhanden sein *muss*. Andererseits *können* gleichlautende Lexikoneinträge auch mehrfach repräsentiert sein (erwünschte Redundanz!).

Die wesentlichen Punkte, die gegen eine neuronale Repräsentation einer Regel der Auslautverhärtung sprechen, sind, kurz zusammengefasst:

- Ein zeitlich folgendes Ereignis kann nicht auf ein vorangegangenes einwirken. Die Annahme eines phonologischen, das heißt zwischen Lexikon und Phonologie eingeschalteten Buffers, ist unrealistisch.
- Selbst wenn man die Norm als Regel akzeptiert, ist eine solche Regel außerordentlich komplex und wirkt teilweise wie eine Aufzählung von Ausnahmen (man beachte auch die „echten“ Ausnahmen wie z. B. *Nerv, Nerven, nervös*), was für eine lexikalische Repräsentation der durch die Regel abgedeckten Phänomene spricht.
- Der Rückgriff auf das Silbenkonzept ist unzulässig bzw. führt zu unrealistischen Konstruktionen. Merkmale wie Silbenende oder Wortende müssten den entsprechenden Phonemen aufgeprägt sein, um einen Einfluss auf die Artikulation zu haben. Es ist realistischer, mit einer entsprechenden phonetischen Gestalt der Phoneme zu rechnen.
- Ein vermeintlicher Gewinn an Einfachheit bzw. Kürze der Lexikonstruktur insgesamt ist bei der großen Redundanz aller sprachlicher Repräsentationen kein Argument.
- Eine mindestens vorübergehende lexikalische Repräsentation von Varianten, wie sie der Regel entsprechen, ist unvermeidlich. Lernprozesse, die von einem solchen vorübergehenden Zustand zur Verankerung der Regel führen, sind nicht erklärbar.

Das Problem der Auslautverhärtung ist ein Lehrstück für das Scheitern einer Regelformulierung an der mentalen Realität.

### 3.5.4 Regularitäten als Folge historischer Entwicklungen

Die Überlegungen der vorangegangenen Abschnitte haben gezeigt, dass es nicht ganz einfach ist, phonologische Regeln in einer neuronalen Struktur der sprachlichen Kompetenz formal so zu repräsentieren, wie es gängigen Vorstellungen entspricht. Es gilt aber nach wie vor, wie eingangs dieses Kapitels festgestellt, dass phonologische Regularitäten erkennbar sind und beschrieben werden können, und es ist auch klar, dass die Auslautverhärtung im Deutschen durchaus eine solche Regularität ist. Man muss also versuchen,

mit diesen zunächst als widersprüchlich erscheinenden Tatsachen zurechtzukommen.

Ein professioneller Sprecher kann sich die Normvorgaben des Siebs zu Eigen machen und sich bewusst der Regel der Auslautverhärtung entsprechend verhalten. Eine *bewusste* Regelanwendung ist aber nicht das, was sich aus dem natürlichen Erstspracherwerb ergibt. Regeln, die sich aus dem natürlichen Spracherwerb ergeben, sind dem Bewusstsein nicht zugänglich, man kann, metaphorisch gesprochen, nicht in die eigene Sprachkompetenz „hineinsehen". Wenn das möglich wäre, wäre das Geschäft des Linguisten wesentlich einfacher.

Man kann aber bewusst eine bestimmte, als Regel gelernte Regel anwenden. Solche Regeln müssen dabei einen anderen Status haben als die Regeln, die durch den natürlichen Spracherwerbsvorgang erzeugt werden. Wenn man eine sprachlich formulierte Regel auswendig lernt, bedeutet das den Aufbau eines episodischen Gedächtnisinhalts. Ein solcher Gedächtnisinhalt kann bewusst aufgerufen werden (Vorgang der Erinnerung) und er kann als Handlungsanweisung eingesetzt werden.

Man kann beliebige, im Augenblick gehörte akustische Signale, so weit sie den auditiven Möglichkeiten entsprechen, nachzuahmen versuchen. Das muss nicht mit Hilfe der Artikulationsorgane geschehen, man kann auch in bestimmter Weise Beifall klatschen, Klavier spielen usw. Man kann, wenn man die erforderlichen Schulkenntnisse über die Existenz von Sprachlauten hat, auch einzelne Laute mehr oder weniger erfolgreich artikulatorisch verändern. Solche Nachahmungsvorgänge können als Folgen entsprechender Handlungsanweisungen verstanden werden.

Regelformulierungen, die der Normierung oder dem Zweitspracherwerb dienen, können aber nur auf einem indirekten Weg (wenn überhaupt) zu natürlichen, den Ergebnissen des natürlichen Spracherwerbs entsprechenden Repräsentationen führen. Dieser indirekte Weg führt über die Wahrnehmung entsprechender Äußerungen, ggf. auch über den Vorgang des „inneren Sprechens".

Ein kindlicher Erstspracherwerb aufgrund von Äußerungen, deren Form einer bewusst eingesetzten Regel im Bereich der Phonologie entspricht, muss aber nicht zur isolierten Repräsentation dieser Regel führen, sondern wird mindestens zunächst ein ausdrucksseitiges Lexikon liefern, das die durch die Regel produzierten Verteilungen spiegelt. Die Regularität ist in den Äußerungen des Kindes nachweisbar, die Regel ist aber nur noch in ihren Effekten repräsentiert.

Es liegt nun nahe, bei Regularitäten, deren Regelrepräsentation innerhalb der sprachlichen Kompetenz neuronal unplausibel ist, einen historischen Prozess anzunehmen, der von mehr oder weniger bewussten, jedenfalls episodisch gesteuerten sprachlichen Veränderungen in Äußerungen ausgeht und über einen Lernprozess zu Lexikonrepräsentationen führt, die den Regularitäten entsprechen. Ursachen für das Entstehen phonologischer Varianten können sehr vielfältig, auch phonetischer Natur sein. Die Verbreitung phonologischer Regularitäten innerhalb einer Sprachgemeinschaft geschieht nach soziolinguistischen Kriterien. (Zur Theorie des historischen Wandels vgl. Croft 2000.)

# 3.6 Konsequenzen

## 3.6.1 Methoden

Wissenschaftszweige entwickeln natürlicherweise spezifische Methoden, die neben eher allgemeinen methodischen Grundprinzipien Verwendung finden. Das hat von jeher auch für die Linguistik gegolten, selbst in der Zeit der Vorherrschaft der historischen Sprachwissenschaft. Durch die Entwicklung der generativen Sprachtheorie ist der Trend zur methodischen Isolierung der Disziplin entschieden verstärkt worden, vor allem auch auf Grund der These von einem modular abgekapselten „Sprachorgan“, für das besondere Strukturen angenommen werden können. Es ist so etwas wie eine „selbsttragende“ Linguistik entstanden. Die Stützpfeiler sind einfache Beobachtungen (Typ: bestimmte Äußerungsformen kommen vor bzw. sind möglich, andere nicht). Die Theoriengebäude entstehen durch Schlussfolgerungen unter dem alles beherrschenden Prinzip der größtmöglichen Generalisierung.

Manche Linguisten lehnen Neurolinguistik und Psycholinguistik als nicht eigentlich „linguistisch“ ab, weil sie sich methodisch von „klassischen“ Bereichen der Linguistik unterscheiden. Die Biologie wird prinzipiell als irrelevant (als die Linguistik nicht betreffend) ausgeklammert. Auch die Phonetik ist von einer gewissen Ausgrenzung betroffen, mindestens solange sie ausdrücklich experimentell arbeitet.

In der Zeit nach dem Entstehen der generativen Sprachtheorie und den ersten epochemachenden Arbeiten von Noam Chomsky sind aber wesentliche Veränderungen bezüglich des Verständnisses mentaler Phänomene eingetreten, die auch die Grundannahmen des Generativismus relativieren. Man kann sich heute prinzipiell vorstellen, wie mentale Verarbeitungsprozesse (im Generativismus standardmäßig ausgeklammert) aussehen *könnten*. Das ist eine der in unserem Zusammenhang wesentlichsten Leistungen der Informatik, speziell der Forschungen im Bereich der „Künstlichen Intelligenz“. Darüber hinaus liefert die Hirnforschung biologische Informationen,

die manchmal so beschrieben werden, dass das arbeitende Gehirn direkt beobachtet werden kann. Eine selbsttragende Linguistik ist also eigentlich nicht mehr zeitgemäß. Unter diesem Gesichtspunkt kann die experimentelle Phonetik als Vorläufer einer modernen linguistischen Arbeitsweise betrachtet werden.

Die Methodenvielfalt kann zu äußerlich unvereinbaren Forschungsergebnissen führen. Die von den bildgebenden Verfahren gelieferten Ergebnisse sind z. B. nicht ohne weiteres mit den Regelsammlungen der generativen Phonologie zusammenzubringen. Als gemeinsame Basis, die eine Integration ermöglichen sollte, bietet sich der Bezug auf kleinräumige Strukturen und Prozesse im Gehirn und dem davon abhängigen peripheren Nervensystem an. Sprache findet in wesentlichen, strukturbestimmenden Teilen unbestreitbar im Gehirn statt. Das Problem dabei liegt in dem immer noch sehr beschränkten empirischen Zugang zu den entsprechenden Daten. Die im Projekt „kortikale Linguistik“ zugrundeliegende methodische Konsequenz ist die Verwendung von Techniken der Modellbildung. Die Modellbildung in der in Teil 1, „Wissenschaftstheoretische Voraussetzungen“ eingeführten Form wird zusätzlich in das Methodeninventar der Phonetik bzw. Phonologie aufgenommen. Die Konsistenzkontrolle für theoretische Konstrukte besteht in der Kontrolle des Funktionierens der entsprechenden Modelle.

Darüber hinaus sind nicht alle Möglichkeiten und Techniken, die traditionell im Bereich der Phonologie/Phonetik eingesetzt werden, von gleicher Wichtigkeit, wenn man sich speziell für den mentalen Phänomenkomplex interessiert. Einige in der klassischen Methodologie sehr hoch bewertete Verfahren werden in ihrer Bedeutung geschwächt.

Das gilt vor allem für Spektralanalysen. Spektralanalysen von Schallereignissen sind heute mit Computerhilfe sehr leicht herzustellen. Man muss aber beachten, dass sie eine grafische Widerspiegelung ausschließlich des physikalischen (also bezüglich des mentalen Geschehens externen) Signals bieten. Eigenschaften dieses externen Signals, selbst wenn sie, wie die Hervorhebung von Formanten, sehr auffällig sind, sind nicht notwendig auch sprachlich, z. B. für die mentale Analyse im Verstehensprozess, relevant. Es ist selbstverständlich nicht so, dass alles, was eine Spektralanalyse liefert, auch dann, wenn sie auf einen Frequenzbereich beschränkt wird, der den Leistungen des Innenohrs entspricht, überhaupt zentral wahrnehmbar ist. Das Verarbeitungsprodukt, wie es das Innenohr und schon die unmittelbar anschließenden Nuclei liefern, kann eine Gewichtung der physikalischen Klangeigenschaften und damit ein Muster liefern, das sich von dem grafischen Bild der Spektralanalyse mehr oder weniger weit entfernt. Spektralanalysen nehmen in Darstellungen der Phonetik einen großen Raum ein.

Das elegante technische Hilfsmittel lässt in Vergessenheit geraten, dass es Schall, und nicht Sprache zum Untersuchungsgegenstand hat. Der Übergang vom Schall zur Sprache im Forschungskontext setzt eine relativ komplexe Interpretationsleistung durch den Linguisten voraus, auf die nicht verzichtet werden kann.

Eine analoge Funktion wie die Spektralanalysen für den auditiven Bereich, haben Myogramme, Palatogramme und Röntgenaufnahmen für den Bereich der Artikulation. Auch hier ist zu beachten, dass zunächst nur die physikalische Außenseite erfassbar wird. In gewisser Hinsicht ist die Situation aber etwas günster als bei den Spektrogrammen, da die an der sprachlichen Produktion beteiligten Muskeln ja tatsächlich ein sprachlich ausgelöstes Innervationsmuster voraussetzen. Aber auch hier gilt, dass man sich vor voreiligen Schlüssen bezüglich tiefer liegender sprachlicher Prozesse und Repräsentationen hüten muss. Insbesondere ist der Einfluss nichtsprachlicher Reflexschaltungen zu beachten.

Methoden, die der Psycholinguistik nahestehen, werden zum Nachweis verschiedener Effekte, z. B. des Phonemrestaurationseffekts oder des McGurk-Phänomens, verwendet. Der Phonemrestaurationseffekt wird, da lexikalische Grundlagen eine Rolle spielen, ausführlicher in Teil 4, „Lexikon“, behandelt. Das McGurk-Phänomen (klassische Publikation McGurk & MacDonald, 1976) kann eigentlich nur beschränkt zur Erklärung phonetischer bzw. phonologischer Vorgänge herangezogen werden. Es besteht in der Beobachtung, dass die visuelle Wahrnehmung der Lippenstellung bei der Produktion eines Sprachlauts das Urteil von Versuchspersonen über den gehörten Laut beeinflusst. Man beachte, dass das Phänomen in einer Laborsituation auftritt, mit allen sich daraus ergebenden Unsicherheiten. In einer normalen Kommunikationssituation mit gesunden Gesprächspartnern spielt ein zusätzliches Lippenlesen zur Ergänzung des Hörvorgangs in den seltensten Fällen eine Rolle. Man sollte keine grundsätzlichen Schlüsse aus dem McGurk-Phänomen für die Sprachwahrnehmung ziehen.

Das klassische strukturalistische Verfahren in der Phonologie ist die Distributionsanalyse, also die Feststellung des Vorkommens eines Lauts oder einer Lautkombination in bestimmten Kontexten. Wenn man Distribution als lexikalische Eigenschaft versteht, die nicht notwendig eine phonologische Basis haben muss, sind Ergebnisse von Distributionsanalysen in ihrem Wert für Erkenntnisse über mentale phonologische Phänomene immer interpretationsbedürftig. Distributionen können, wie oben in Abschnitt 3.5.4 dargestellt, einen historischen Hintergrund spiegeln, mindestens muss ein solcher Hintergrund immer in Betracht gezogen werden.

Problematisch ist auch der Sprachvergleich zur Bestimmung von Universalien, der in der generativen Phonologie eine besondere Rolle spielt. Wenn dabei das Prinzip der größtmöglichen Generalisierung über alle Sprachen eine Rolle spielt, muss darauf hingewiesen werden, dass es dafür keine ausreichende biologische Basis gibt. Das methodische Prinzip der größtmöglichen Generalisierung, tritt mindestens für den Bereich von Phonetik, Phonologie und Lexikon völlig in den Hintergrund.

### 3.6.2 Beschreibungen

Die Folgen, die der Einsatz der Modellbildung in der Linguistik hat, werden am Beispiel der Phonetik bzw. Phonologie besonders deutlich. Man kann bei der Untersuchung der in einer bestimmten Sprache möglichen lexikalischen Ausdrucksseiten phonetische bzw. phonologische Regularitäten feststellen, die eine Beschreibung in Form einer Regel nahe legen. Eine solche Beschreibung ist, solange sie die lexikalischen Daten korrekt widerspiegelt, durchaus als korrekt zu akzeptieren. Es folgt aber daraus nicht, dass die Regel als solche mental repräsentiert, also Bestandteil eines Modells der sprachlichen Kompetenz sein muss. Der Versuch der Modellbildung zeigt u. U., wie oben in Abschnitt 3.5.2 ausgeführt, dass das nicht möglich ist.

Die Unterscheidung von Beschreibung und Modellbildung ist für eine Linguistik, die sich mit der Sprache als mentalem Besitz des Sprechers beschäftigt, von zentraler Wichtigkeit. In dieser Hinsicht sind linguistische Aussagenkomplexe (um es möglichst neutral auszudrücken) in der Vergangenheit besonders nachlässig gewesen. Das soll nicht heißen, dass nicht auch Beschreibungen, wenn man sie bewusst als solche einsetzt, eine positive Rolle in der Linguistik spielen können und es einen Wert hat, solche Beschreibungen zu entwickeln. Sie haben dann Funktionen wie oben in Abschnitt 3.5.4 behandelt. Im Fremdsprachenunterricht werden sie mit Erfolg eingesetzt. Das bedeutet, dass eine Grammatik (der Begriff im weitesten Sinn verstanden) für den Fremdsprachenunterricht nicht unbedingt die Kompetenz eines Muttersprachlers modellhaft widergeben muss. Das ist in vielen Bereichen auch gar nicht sinnvoll. Man muss sich klar machen, dass ein Spracherwerb möglich ist, der auf dem Umweg über bewusst eingesetzte „Regeln“ schließlich zu sprachlichen Repräsentationen und „natürlichen“ Prozessen führen kann, in denen diese Regeln (als repräsentierte Komponenten) dann keine Rolle mehr spielen.

Auch dann, wenn man sich ausschließlich für die mentale sprachliche Kompetenz interessiert, können linguistische Darstellungen die Form von Beschreibungen haben. Solche Beschreibungen müssen dann kompatibel sein

mit einer (aus praktischen Gründen meist fragmentarischen) Modellbildung. In Abschnitt 3.4.4 ist gezeigt, dass diese Forderung z. B. zur Veränderung des zu verwendenden phonetischen Merkmalsinventars führt. Während sich in Beschreibungen der Phonetik für den Sprachunterricht die Verwendung artikulatorischer Merkmale empfiehlt, ist in erstzunehmenden linguistischen Beschreibungen der sprachlichen Kompetenz eine strikte Trennung von artikulatorischen und auditiven Merkmalen zu fordern. Eine „Vorherrschaft“ der artikulatorischen Seite ist nicht zu rechtfertigen.

### 3.6.3 Prozesse

Den Bedingungen der neuronalen Verarbeitung, insbesondere dem Erfordernis massiver Parallelverarbeitung entsprechend, kann es keine Unterscheidung von Daten und Prozessen bzw. Prozessoren geben. Die Repräsentation eines Phonems ist identisch mit der Einrichtung, die das entsprechende Phonem erkennt bzw. produziert. Das heißt, obwohl man von Neuronen sprechen kann, die für ein bestimmtes Phonem stehen, ist doch die Gesamtheit der Phonemrepräsentation letztlich identisch mit der Gesamtheit der Verarbeitungsinstanzen vom Ohr zum Kortex und vom Kortex zu den Artikulationsorganen, die zu dem betreffenden Phonem gehören.

Das hat auch zur Konsequenz, dass Phoneme sozusagen „zitiert“, aber nicht von einer Speicherposition zu einer anderen verschoben werden können. Es kann prinzipiell keinen „Datentransport“ in diesem Sinne im Gehirn geben.

Dabei gilt selbstverständlich, dass die Reaktion einer Sinneszelle auf einen akustischen Reiz, genau so, wie die Reaktion einer einzelnen Muskelfaser, höchst mehrdeutig ist, das heißt, zu verschiedenen Phonemen gehören kann. Für die Perzeptionsseite ergibt sich damit die Frage nach den Eigenschaften des Vorgangs, der zur Auflösung dieser Mehrdeutigkeit führt. Bei Computern (Von-Neumann-Architekturen) ist eine möglichst frühe Auflösung erstrebenswert, da keine ausreichenden Möglichkeiten der Parallelverarbeitung bestehen und bei Fehlentscheidungen aufwändige Revisionsvorgänge erforderlich werden. Für das Gehirn ist umgekehrt eine möglichst späte Vereindeutigung wünschenswert. Der Normalfall für phonologische Perzeptionsprozesse dürfte sein, dass erst der lexikalische Zusammenhang oder ein noch „höherer“ Teilprozess die Mehrdeutigkeit auflöst. Die strukturalistische Idee der „räumlichen Nachbarschaft“ in Darstellungen von Phonemsystemen bezieht sich auf Möglichkeit der Mehrdeutigkeit. Auch ein Strukturalist wird nicht damit rechnen, dass das Phonemsystem im Allgemeinen tatsächlich eine eindeutige Identifizierung von Sprachlauten ermöglicht. Phonemsysteme geben ideale Verhältnisse wieder.

Man beachte, dass nicht anzunehmen ist, dass von der lexikalischen Ebene her „top-down“ eine quasi nachträglicheVerbesserung der phonologischen Eingabe erfolgt. Es gibt keine bidirektionalen Verbindungen der dafür erforderlichen Art. Ein anderer Zusammenhang (und davon zu unterscheiden) ist, dass durchaus auch in der neuronalen Sprachverarbeitung und trotz massiver Parallelverarbeitung eine Korrektur fehlgeschlagener Analyseprozesse erforderlich ist. Die entsprechenden Mechanismen sind in Teil 2, „Grundlagen“, Kapitel 2.5.4, und in Teil 5, „Syntax“ beschrieben. Sie verwenden einen Kohärenzkontrollmechanismus und die auch in der Produktion aktivierten Strukturen, zusammen mit dem Rückspiegelungsmechanismus, wie er oben in den Abschnitten 3.3.5 und 3.4.5 behandelt ist. Sie haben einen gewissen Zeitbedarf, der größer ist, als der, der für die Bearbeitung eines Sprachlauts im unproblematischen Normalfall erforderlich ist.

### 3.6.4 Abschließende Bemerkungen

Aktuelle Darstellungen der Phonetik und Phonologie enthalten Teile, die zwar von grundsätzlichem Wert sind, die aber zur modellhaften Behandlung speziell der *sprachlichen* Kompetenz nicht wesentlich beitragen. Das gilt im Bereich der Phonetik für die teilweise sehr ausführlich ausgearbeiteten Grundlagen der physikalischen Akustik. Es gehören dazu aber auch mindestens teilweise die klassischen Gegenstände der Psychoakustik, wie z. B. Lautheit, Tonheit oder das Konzept der Frequenzgruppen (critical bands).

Einige Gegenstände sind lexikalisch und somit dem Bereich der Behandlung des Lexikons zuzuordnen, das gilt z. B. für Assimilationen und Dissimilationen, Epenthesen und Tilgungen.

Die Prosodie ist, soweit suprasegmental, ein eigenständiges Phänomen, das parallel zu den phonetischen Kategorisierungsprozessen und nicht innerhalb der Bahn, die von der Lautkategorisierung zum Lexikon führt, verarbeitet wird und insofern eigenen Bedingungen unterliegt.

Über den Sinn der umfangreichen phonologischen Regelgebäude, die bis heute unter dem Einfluss der generativen Sprachtheorie konstruiert werden, muss neu nachgedacht werden. Während die strukturalistischen Beschreibungen von Distributionen als Beschreibungen von Phänomenen akzeptiert werden können, die auf lexikalischer Ebene angesiedelt sind, und ggf. eine sprachgeschichtliche Basis haben, wird es schwierig, phonologischen Regelsystemen (auf generativistischer Grundlage) eine eigenständige mentale(!) Existenz zuzuschreiben oder sie wenigstens als Beschreibungen mentaler

Phänomene zu verstehen. Soweit sie aufgrund ihrer Komplexität im Sprachunterricht keine Verwendung finden können, bleiben sie abstrakt in dem Sinne, dass sie ihre Rechtfertigung ggf. ausschließlich aus ihrer intellektuellen Eleganz beziehen. Wenn man eine solche Rechtfertigung nicht akzeptiert, kann das auf Dauer zu einer dramatischen Schrumpfung des Gegenstands Phonologie in der Linguistik führen.

Phonetik und Phonologie sind, wie andere Gebiete der Linguistik auch, an Traditionen gebunden, die Kategorien bereithalten, die gerne dort eingesetzt werden, wo es aufgrund der mangelnden Zugänglichkeit mentaler Phänomene schwierig ist, auf empirische Grundlagen zurückzugreifen. Wo es andere Möglichkeiten gibt, ist dieses gängige Verfahren nicht zu rechtfertigen, genauso, wie es eigentlich unverantwortlich ist, sich dadurch aus den entstehenden Problemen herauszumanövrieren, dass man sie einfach ignoriert.