

## 2. Kapitel

# Prosodische Struktur und Worterkennung bei gesprochener Sprache

*Anne Cutler*

### *1 Einführung*

Die prosodische Struktur ist eine Dimension, die der gesprochenen Sprache zugehört. Zwar kann ein im Schreiben Geübter auch bei Prosatexten beispielsweise rhythmische Effekte anzielen, die dann jedoch von der Fähigkeit des Lesers abhängen, sie mit dem „geistigen Ohr“ zu „hören“, das heißt, den geschriebenen Text im Geiste in seine gesprochene Form umzuwandeln. In diesem Kapitel wird dargestellt, dass und wie Hörer prosodische Information ausgiebig und auf verschiedenartige Weise heranziehen, um gesprochene Äußerungen zu erkennen. Weil die Prosodie jedoch eine Eigenschaft der gesprochenen Sprache ist, welche wiederum (allein schon aus Gründen der empirischen Handhabbarkeit) weit seltener zum Gegenstand psycholinguistischer Forschung wurde als schriftliche Sprache, ist die Untersuchung der Rolle der Prosodie bei der Erkennung sprachlicher Äußerungen vergleichsweise unterentwickelt. Eine jüngst erschienene umfassende Literaturübersicht über dieses Gebiet (Cutler, Dahan & van Donselaar, 1997), die die Rolle der Prosodie sowohl beim Verstehen von Syntax- und Diskursstrukturen als auch bei der Erkennung gesprochener Wörter umfaßt, weist etwa 300 Literaturangaben auf, was im Vergleich mit beispielsweise der Literatur zur visuellen Worterkennung, selbst wenn man hier nur die Arbeiten heranzieht, die auf einem einzigen experimentellen Verfahren beruhen (der lexikalischen Entscheidungsaufgabe), nur eine winzige Menge darstellt. Darüber hinaus kommen Cutler et al. zu dem Schluß, dass die Literaturlage wenig ausgewogen ist: Manche Fragestellungen wurden immer wieder zum Forschungsgegenstand, wobei sich die Untersuchungen nur in kleinen Details unterscheiden, während andere Themen völlig unbeachtet blieben. Dies trifft auch auf die Forschung über verschiedene Spra-

chen zu; die meisten Forschungsarbeiten beziehen sich, wie in allen Bereichen der Psycholinguistik, auf das Englische, aber von anderen Sprachen haben einige in der Forschung beträchtliche Aufmerksamkeit auf sich gezogen, andere blieben völlig unbeachtet. Besonders relevant ist in diesem Zusammenhang der Vergleich zwischen dem Deutschen und dem Niederländischen: Die prosodische Struktur dieser beiden Sprachen ist sehr ähnlich, und in der phonetischen Literatur wurden auch beide Sprachen ausführlich beschrieben. Doch während die psycholinguistische Literatur inzwischen eine erhebliche Anzahl an experimentellen Untersuchungen zur Verarbeitung der Prosodie im Niederländischen aufweist, gibt es auffallend wenige vergleichbare Untersuchungen im Deutschen.

Der Schwerpunkt des vorliegenden Kapitels liegt darin, wie die prosodische Struktur zur Worterkennung bei gesprochenen Äußerungen beitragen kann. „Prosodische Struktur“ bezieht sich (wie in der Phonetik und Psycholinguistik allgemein angenommen) auf die linguistische Struktur, die in den suprasegmentalen Eigenschaften von Äußerungen zum Ausdruck kommt. Man beachte, dass sich in Nachbardisziplinen auch andere Definitionen der Prosodie im aktiven Gebrauch befinden: Für Ingenieure und andere Fachleute der angewandten Sprechforschung bezieht sich der Ausdruck „Prosodie“ auf die suprasegmentalen Eigenschaften selbst - Tonhöhe, Sprechtempo, Lautstärke und zeitliche Verlaufsmuster des Sprechens; Phonologen verstehen unter Prosodie die abstrakte Organisation, die die Gruppierung und relative Auffälligkeit von phonologischen Einheiten bestimmt; in der Poetik beschreibt die Prosodie die Normen von Versmaßen. Zwar sind diese Definitionen alle miteinander verwandt - und dies in einem Maße, dass die Forscher manchmal unbeabsichtigt aneinander vorbeireden können, weil sie nicht erkennen, dass sich ihre Kollegen auf eine andere Definition beziehen; doch enthalten die Definitionen unterschiedliche Implikationen hinsichtlich dessen, was in den Einzugsbereich der Prosodieforschung fällt und was nicht. Das vorliegende Kapitel beschränkt sich diesbezüglich auf die Verarbeitung der sprachlichen Struktur, die durch die suprasegmentale Struktur einer Äußerung vermittelt wird, und dabei auch nur auf diejenigen Elemente der sprachlichen Struktur, die für die Wortidentifikation unmittelbar relevant sind.

Die Befundlage, die hier im Überblick dargeboten wird, stammt ausschließlich von experimentellen Untersuchungen zur Sprachverarbeitung aus psycholinguistischen Labors. Dabei treten natürlich auch die üblichen Probleme bei der experimentellen Kontrolle der zu untersuchenden Phänomene auf; wo eine sprachlich relevante Struktur in Wirklichkeit durch mehrere unterschiedliche Informationsquellen im Signal realisiert ist, ist es manchmal nötig, separate Untersuchungen durchzuführen, um die einzelnen Beiträge dieser Komponenten herauszuarbeiten, und Befunde, die für sich genommen nicht-sprachlicher

Natur sind oder zumindest nicht unmittelbar etwas mit suprasegmentalen Phänomenen zu tun haben, parallel zu berücksichtigen. Beispielsweise hat der lexikalische Stress (also der Wortakzent) im Englischen sowohl segmentale als auch suprasegmentale Korrelate in der Wortform; seine Verarbeitung kann, wie weiter unten verdeutlicht wird, nicht verstanden werden, wenn man nicht auch Befunde zur segmentalen Verarbeitung, zur Verarbeitung von Rhythmus und zur akustischen Verteilung suprasegmentaler Merkmale berücksichtigt. Ähnliches gilt für andere Fragestellungen in diesem Gebiet; im Grunde liegt das eben gerade an der Natur der prosodischen Information: Sie ist nämlich nicht auf die eine oder andere Weise dem Signal aufgelagert, sondern sie ist eine jedem Aspekt der gesprochenen Form innewohnende Determinante.

In ähnlicher Weise umfaßt die Rolle der Prosodie bei der Worterkennung in gesprochenen Äußerungen viele Facetten. Einerseits gibt es mehrere Wege, auf denen die prosodische Struktur auf Satzebene die Worterkennung indirekt, aber entscheidend erleichtert. Andererseits kann die prosodische Information auf Wortebene auch direkt an der lexikalischen Verarbeitung beteiligt sein. Diese beiden Themenfacetten werden im folgenden ausführlich behandelt.

## *2 Prosodie auf Satzebene*

Die prosodische Struktur auf Satzebene umfaßt den Gesamtrhythmus einer Äußerung, die verschiedenen Arten und Weisen der Gruppierung von Wörtern sowie die Muster der Hervorhebung der einzelnen Satzkonstituenten. Hörer können Informationen über syntaktische Beziehungen und die Diskursstruktur direkt und indirekt aus der Satzprosodie ableiten (Cutler et al., 1997); sie können die Satzprosodie aber auch beim Prozeß der Worterkennung heranziehen. Hervorhebungen können beispielsweise einzelne Wörter leichter oder schwerer erkennbar machen; die rhythmische Struktur kann einen Rahmen schaffen, innerhalb dessen sich die Grenzen zwischen einzelnen Wörtern feststellen lassen; und wo in der prosodischen Struktur eine Grenze oberhalb der Wortebene enkodiert ist, geht sie gleichzeitig mit einer Wortgrenze einher.

### 2.1 Rhythmus und Gruppierung

Die meisten gesprochenen Äußerungen verlaufen mehr oder weniger kontinuierlich - zwar verursachen obstruente Konsonanten (das ist die Menge aller Konsonanten außer den Liquiden [r] und [l] und den Nasalen [n] und [m], die dadurch definiert ist, dass der Luftstrom beim Austritt behindert wird) zuweilen kurze Untetbrechungen im Signal, nicht aber die Grenzen zwischen phonetischen Segmenten und die zwischen lexikalischen Elementen. Dennoch muß der

Hörer das Signal in wirksamer Weise mit Unterbrechungen versehen, indem er die einzelnen Wörter identifiziert, aus denen sich die Äußerung zusammensetzt. Der Hörer kann (für gewöhnlich) nicht im voraus wissen, was der Sprecher sagen will; er kann das Gesagte nur verstehen, indem er die Wörter erkennt, aus denen die Äußerung besteht, und dazu muß er herausfinden, welche Anteile des sprachlichen Inputs zu welchen Wörtern gehören. Der kontinuierliche Verlauf des Sprechsignals kompliziert die Sache, indem er keine expliziten Hinweise bietet, die diesen Segmentierungsprozeß unterstützen würden.

Zu den Grenzen zwischen syntaktischen Phrasen finden sich jedoch oft Entsprechungen in der prosodischen Struktur, so dass der Hörer dieser Struktur Abgrenzungsinformation entnehmen kann. Dies zeigt sich in Untersuchungen, bei denen den Hörern Äußerungen dargeboten werden, deren segmentale Information entfernt wurde (zum Beispiel durch Filterung oder spektrale Manipulation des Sprechsignals); die abgegebenen Urteile über die Lokalisation von Grenzen sind bei derartig aufbereitetem Input ziemlich exakt (Collier & 't Hart, 1975; de Rooij, 1976; de Pijper & Sanderman, 1994). Eine äußerst effiziente Grundlage für die Urteile der Hörer bilden der Tonhöhenverlauf (zum Beispiel Streeter, 1978; Beach, 1991; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992) oder Information über Länge und Dauer (zum Beispiel Lehiste, Olive & Streeter, 1976; Nooteboom, Brox & de Rooij, 1978; Scott, 1982). Wenn Hörer solche Hinweise heranziehen, um das Vorhandensein einer syntaktischen Grenze zu postulieren, dann postulieren sie damit implizit auch das Vorhandensein einer Wortgrenze. Obwohl die Ableitung von Hinweisen auf Wortgrenzen aus der Wahrnehmung syntaktischer Grenzen in Experimenten zum Worterkennen nicht explizit untersucht wurde, legen einige indirekte Belege, die im weiteren zusammengefaßt werden, die Annahme nahe, dass dieser Weg der Ableitung für die Hörer tatsächlich möglich ist.

Abgesehen von dieser Möglichkeit, Information über Wortgrenzen auf dem Wege der Kalkulation syntaktischer Grenzen zu erhalten, leiten Hörer solche Information jedoch auch dadurch ab, dass sie ihr Wissen über die rhythmische Struktur ihrer Muttersprache nutzen. Der Rhythmus ist eine grundlegende Eigenschaft der Sprache (und vieler anderer menschlicher Verhaltensweisen), aber er ist nicht in jeder Sprache gleich; deshalb unterscheidet sich die Art und Weise, in der der Rhythmus mit Wortgrenzen zusammenhängt, von Sprache zu Sprache. Im Englischen beispielsweise basiert der Rhythmus auf dem Stress. Das lässt sich leicht an Hand der Formen englischer Lyrik erkennen — sie sind generell durch die Anzahl und Position der Taktschläge definiert (Hexameter, Trochäus, etc.). Taktschläge können in englischen Versen nur auf starke Silben fallen - das sind Silben, die einen vollen (nicht reduzierten) Vokal enthalten. So liegen bei „JACK and JILL went UP the HILL“ die Schläge *au{Jack,ßll, up* und *hill*. *Went* wäre ebenfalls ein möglicher Ort für eine Hebung - „JACK and JILL,

they WENT upHILL" könnte man im selben Rhythmus sprechen. Hingegen werden *and* und *the* normalerweise als weiche Silben — mit reduzierten Vokalen — gesprochen, weshalb sie keine Betonung tragen können: „JACKson AND Jill WENT inLAND" oder „JACK and JILL climbed THE big HILL" empfindet das englische Ohr als hochgradig unangenehm.

Es gibt zahlreiche experimentelle Belege dafür, dass englische Hörer diesen betonungsbasierten Rhythmus zur Ableitung von Information über Wortgrenzen heranziehen; sie segmentieren gesprochene Sprache nämlich am Beginn starker Silben (das heißt der Silben mit vollen Vokalen, die potentiell betont werden können). Wenn sich Sprecher des Englischen beispielsweise auf eine Weise verhalten, an der die fehlerhafte Wahrnehmung von Wortgrenzen beteiligt ist, dann neigen sie am häufigsten dazu, Wortgrenzen vor starken Silben anzunehmen (zum Beispiel hören sie *by loose analogy* als *by Luce and AlUrgy*), oder sie entfernen Wortgrenzen vor schwachen Silben (zum Beispiel hören sie *how big is it?* als *how bigoted?*; Cutler & Butterfield, 1992). Auch finden es englische Hörer schwer, ein echtes Wort, das in eine ansonsten sinnlose Sprechsequenz eingebettet ist, zu entdecken, wenn die das Wort enthaltende Sprechsequenz aus zwei starken Silben besteht (zum Beispiel *mint* in [mmtef]); sie haben es jedoch leichter, ein Wort in einer Sprechsequenz zu entdecken, die eine starke und eine nachfolgende schwache Silbe umfaßt (zum Beispiel *mint'm* [mintaf]; Cutler & Norris, 1988). Cutler und Norris führen an, dass sich der genannte Unterschied deshalb ergibt, weil die Hörer [mmtef] am Beginn der zweiten starken Silbe trennen, so dass sie, um das eingebettete Wort zu erkennen, das sprachliche Material über einen Segmentierungspunkt hinweg zusammensetzen müssen, während die Lautfolge [mint 3 f] im Gegensatz dazu der Entdeckung des eingebetteten Wortes nicht im Wege steht, da sie einfach nicht getrennt wird, weil die zweite Silbe schwach ist.

Warum sollten englische Hörer den Betonungsrhythmus in dieser Weise heranziehen? Statistische Untersuchungen des englischen Wortschatzes und der Verteilungsmuster in spontaner Sprache (Cutler & Carter, 1987) haben gezeigt, dass die Strategie, das Englische am Auftakt starker Silben zu segmentieren, in der Tat eine äußerst nützliche Methode ist, um Wortanfänge zu lokalisieren — die meisten lexikalischen Wörter (Nomina, Verben, Adjektive) fangen tatsächlich mit einer starken Silbe an. Die Verteilungsmuster im niederländischen Wortschatz gleichen denen des Englischen, wobei sogar noch mehr niederländische als englische Wörter in der ersten Silbe einen vollen Vokal aufweisen (van Heuven & Hagman, 1988; Schreuder & Baayen, 1994). Experimente auf der Vorlage der beiden Untersuchungen von Cutler und Norris (1988) sowie Cutler und Butterfield (1992) ergaben im Niederländischen ähnliche Befunde (Vroomen, van Zon & de Gelder, 1996). Die Ausnutzung des Betonungsrhythmus scheint somit eine Strategie zu sein, die die Hörer deshalb einsetzen, weil

sie eine effiziente (Teil-)Lösung für das Problem bietet, das sich aus der Schwierigkeit ergibt, Wortgrenzen im kontinuierlichen Sprachfluß zu lokalisieren. (Das Deutsche besitzt, wie das Niederländische, einen sehr hohen Anteil an Wörtern mit vollem Vokal in der ersten Silbe Q. Bölte, persönliche Mitteilung], doch liegen bislang keine entsprechenden experimentellen Befunde vor.)

Eine solche Lösung bietet sich jedoch nicht für Sprecher von Sprachen ohne Betonungsrhythmus. So zeigt etwa das Französische nicht diese Kontrastierung zwischen starken und schwachen Silben, die sich in den germanischen Sprachen beobachten lässt. Experimentelle Untersuchungen zur Verarbeitung von gesprochenem Französisch deuten daraufhin, dass sich die Hörer hier auf einen Prozeß der Segmentierung des sprachlichen Inputs in Silbeneinheiten stützen können; betrachtet man die beiden Wörter *palace* und *palmier* (die beide mit denselben drei Phonemen beginnen), so bilden beim Wort *pa-lace* die ersten beiden Phoneme, beim Wort *pal-mier* die ersten drei Phoneme die erste Einheit. Eine breite Vielfalt experimenteller Aufgaben, die die prälexikalische Verarbeitung, die lexikalische Verarbeitung und die Repräsentation von Wörtern im Gedächtnis betreffen, führt zu Ergebnissen, die zeigen, wie wichtig dieses Vorgehen bei der Worterkennung im Französischen ist (Mehler, Dommergues, Frauenfelder & Segui, 1981; Segui, Frauenfelder & Mehler, 1981; Cutler, Mehler, Norris & Segui, 1986; Dupoux & Mehler, 1990; Pallier, Sebastián-Galles, Felguera, Christophe & Mehler, 1993; Kolinsky, Morais & Cluytens, 1995; Peretz, Lussier & Beland, 1996).

Obwohl es sich bei der Verwendung des stress-basierten Rhythmus im Englischen und bei der Verwendung der silbischen Segmentierung im Französischen um recht unterschiedliche Lösungen des Problems der Segmentierung kontinuierlicher Sprache zu handeln scheint, kann man sie auch als ähnlich betrachten: Wie der Stress im Englischen, so ist die Silbe im Französischen die Basis der rhythmischen Struktur. Diese Symmetrie legte die Hypothese nahe (vgl. zum Beispiel Cutler, Mehler, Norris & Segui, 1992), dass Hörer letztlich eine universell anwendbare Lösung des Segmentierungsproblems wählen könnten, indem sie zu dessen Lösung jedwede rhythmische Struktur nutzen, die ihre jeweilige Sprache kennzeichnet. Daraus folgt, dass sich bei einer Sprache, deren rhythmische Struktur auf einem anderen Konstrukt als dem des Stress oder der Silbe beruht, Belege für die Heranziehung solcher rhythmischer Strukturen bei der Segmentierung des Sprechflusses finden lassen sollten. Japanisch ist eine solche Sprache; ihr Rhythmus lässt sich an Hand einer Einheit unterhalb der Silbenebene beschreiben, der Mora. (Zum Beispiel besitzt das Wort *tanshi* drei Morae: *ta-n-sht*). Otake, Hatano, Cutler und Mehler (1993), Cutler und Otake (1994) sowie Otake, Hatano und Yoneyama (1996) führten Untersuchung zur prälexikalischen Verarbeitung bei japanischen Hörern durch und fanden tatsächlich konsistente Belege für eine mora-basierte Segmentierung.

Es ist nicht die rhythmische Struktur des Inputs selbst, die das geeignete Segmentierungsverfahren mit sich bringt; sonst könnte ja der Hörer bei jeder beliebigen Sprache die Wortgrenzen mit Erfolg heraushören. Die Erfahrung zeigt, dass dies mit Sicherheit nicht der Fall ist. Es scheint vielmehr so zu sein, dass Hörer Segmentierungsverfahren auf der Basis ihrer Erfahrung mit ihrer Muttersprache herausgebildet haben und dass sie die geeigneten Verfahren für andere Sprachen nicht beherrschen. So lassen englische Hörer bei französischem Input keine Anhaltspunkte für eine silbische Segmentierung erkennen (Cutler et al., 1986), genauso wenig wie japanische Hörer (Otake et al., 1996); in entsprechender Weise zeigen englische Hörer bei japanischem Input keine Hinweise auf eine mora-basierte Segmentierung (Otake et al., 1993; Cutler & Otake, 1994), was wiederum ebenso für französische Hörer (Otake et al., 1993) oder niederländische Hörer (Kakehi, Kato & Kashino, 1996) gilt. Darüber hinaus wenden Hörer gegebenenfalls ihre muttersprachspezifischen Verfahren auf den fremdsprachigen Input an, selbst in Fällen, in denen diese Verfahren überhaupt nicht gut funktionieren können. So wenden französische Hörer die silbische Segmentierung auf englische Wörter wie *palace* und *palpitate* an, wo es englische Hörer nicht tun würden (Cutler et al., 1986); entsprechend benutzen sie die silbische Segmentierung auch bei japanischem Input (zum Beispiel segmentieren sie *tanshi* bevorzugt als *tan-shi*; Otake et al., 1993). Japanische Hörer wiederum segmentieren englischen Input, wo es möglich ist, nach Morae (beispielsweise zeigen sie — anders als englische Hörer — eine leichtere Verarbeitung von Nasalen am Silbenende in Wörtern wie *incur*, Cutler & Otake, 1994); Entsprechendes gilt für französische und für spanischen Input (beispielsweise reagieren sie mit gleicher Geschwindigkeit auf Zielreize, die aus einer Konsonant-Vokal-Verbindung bestehen wie *w'iepa-*, gleich ob sie in einer offenen oder in einer geschlossenen Silbe stehen; Otake et al., 1996).

Abschließend muß daraufhingewiesen werden, dass die Ausnutzung der unterschiedlichen Ebenen der linguistischen Struktur bei der Segmentierung nicht durch das bloße Vorhandensein dieser Strukturtypen in einer Sprache bestimmt ist. Bei allen Konzepten, die sich bei der Beschreibung dieser von Sprache zu Sprache bestehenden Unterschiede bei der Segmentierung als relevant erwiesen haben — Stress, Silbe, Mora - handelt es sich um phonologische Konstrukte, die im Prinzip auf jede Sprache angewandt werden können. Nur über die Rolle, die die relevanten Einheiten im Rhythmus der jeweiligen Sprache spielen, erlangen sie auch ihre Rolle für die Segmentierung. Und obwohl beispielsweise das Konzept des Stress vom Konzept der Silbe abhängt (betont/unbetont sind Eigenschaften von Silben, nicht von Silben teilen), verleiht dieser Umstand der Silbe in einer vom Betonungsrhythmus getragenen Sprache nicht dieselbe Rolle, die sie in einer Sprache mit silbischem Rhythmus einnimmt. Der Befund von Cutler et al. (1986), denen es nicht gelang, bei englischen Hörern silbische Segmentierungen aufzuweisen, wurde vielfach repliziert (Bradley, Sánchez-Casas

& Garcia-Albea, 1993; Cutler, Norris & Williams, 1987; Kearns, 1994; Taft & Hambly, 1985). Für das Deutsche fand eine Untersuchung von Höhle und Schriefers (1995) Reaktionsmuster, die nur bei Wörtern mit Endbetonung und offener Anfangsilbe mit einer silbischen Segmentierung übereinstimmen (das heißt, *ku-* wurde in *Kulanz* schneller entdeckt als *kul-*). Für das Niederländische fanden Zwitterlood, Schriefers, Lahiri und van Donselaar (1993) Belege für eine silbische Segmentierung, doch eine vergleichbare Untersuchung von Vroomen und de Gelder (1994) konnte solche Effekte nicht nachweisen. Cutler (1997) gab niederländischen Hörern das leicht in Silben zerlegbare französische Sprachmaterial von Mehler et al. (1981) vor; die Probanden zeigten —wie schon die englischen Hörer aus der Untersuchung von Cutler et al. (1986) — nicht dieselben Reaktionsmuster wie französische Hörer; allerdings reagierten sie, in Übereinstimmung mit den deutschen Probanden von Höhle und Schriefers, schneller auf die silbischen Zielreize in Wörtern mit einer offenen Anfangsilbe (das heißt, *pa-* wurde in *palace* schneller entdeckt als *pal-*). Man beachte, dass der Akzent in der Rezitationsaussprache französischer Wörter auf die Endsilbe fällt, so dass der Befund in diesen beiden Fällen der Segmentierung am Beginn einer betonten Silbe entspricht, was wiederum die Standardsegmentierung in Sprachen mit stress-basiertem Rhythmus darstellt.

Der Betonungsrhythmus des Englischen, Niederländischen oder Deutschen ist selbst nicht durch die Lokalisation von Wortgrenzen bestimmt; in diesen Sprachen kann der Stress an unterschiedlichen Stellen im Wort auftreten. In manchen Sprachen ist Stress jedoch festgelegt; hier muß er immer in derselben Position innerhalb des Wortes liegen. Zu den Sprachen mit derart festgelegter Betonung gehört zum Beispiel Finnisch, wo jedes Wort auf der ersten Silbe betont wird, oder Polnisch, wo der *Stress* immer auf die vorletzte Silbe fällt. Man könnte meinen, dass eine festgelegte Wortbetonung einen ausgezeichneten Hinweis auf die Lokalisation von Wortgrenzen bietet; tatsächlich aber kann es auf fast paradoxe Weise möglich sein, dass den Hörern in Sprachen mit festliegender Betonung weniger explizite akustische Korrelate des Stress zur Verfügung stehen als in Sprachen mit variablem Stress. Das liegt daran, dass die explizite Realisierung des Wortakzents überflüssig sein kann, wenn seine Position vollständig vorhersagbar ist. Suomi, McQueen und Cutler (1997) führten ein Segmentierungsexperiment im Finnischen durch, in dem sie dieselbe Aufgabe einsetzen wie Cutler und Norris (1988), bei der man, wie oben beschrieben, Wörter in sinnlosen Lautsequenzen erkennen soll. Das Experiment war eigentlich nicht auf Stress, sondern auf Harmonie zwischen Vokalen gerichtet (die es im Finnischen erforderlich macht, dass zwei Vokale im selben Wort aus kompatiblen Vokalklassen stammen müssen). Die Hörer in der Untersuchung von Suomi et al. hörten zweisilbige Wörter (zum Beispiel *palo*) mit vorangehender oder nachfolgender Konsonant-Vokal-Umgebung (*kupalo*, *pabku*); die entstandenen deisilbigen sinnlosen Wörter wurden mit dem unmarkierten Prosodiemuster für dreisilbige Wörter ge-



sprachen, das üblicherweise als Betonung auf der ersten Silbe beschrieben wird. Das wichtigste Ergebnis dieser Untersuchung bestand darin, dass die Vokalharmonie als Segmentierungshinweis funktionierte: Wörter, denen eine disharmonische Umgebung vorangeht (also Silben, die einen Vokal aus einer Klasse enthalten, die mit den Vokalen im Wort nicht kompatibel ist), wurden schneller entdeckt als Wörter, denen ein harmonischer Kontext vorangeht. Stress war jedoch für die Interpretation eines Kontrollexperiments relevant, bei dem die Wörter aus ihrem Kontext herausgeschnitten und Hörern im Rahmen einer lexikalischen Entscheidungsaufgabe vorgegeben wurden; für die Wörter, deren vorausgegangener Kontext entfernt worden war, führte dies zum Verlust derjenigen Silbe, auf der nominell die Betonung gelegen hatte, weshalb für diese Wörter im Vergleich zu den Wörtern mit nachfolgendem Kontext eine prosodische Abnormität zu erwarten gewesen wäre. Die Reaktionen der Hörer wiesen jedoch keinen Effekt auf, der auf eine derartige Abnormität zurückzuführen wäre; wenn überhaupt, dann wurden Wörter *Wie pah*, akustisch herausgeschnitten aus *kupalo*, ein wenig schneller erkannt als Wörter wie *palo*, herausgeschnitten aus *paloku*. Suomi et al. sprachen die Vermutung aus, dass die sogenannte Initialbetonung im Finnischen eigentlich ein allmähliches Absinken der Grundfrequenz und der Amplitude über das Wort hinweg darstelle und dass für ihre korrekte Realisierung einfach nur die Beziehung zwischen aufeinanderfolgenden Silben wichtig sei; diese Beziehung bliebe durch das Entfernen vorausgehender oder nachfolgender Silben unberührt.

Französisch ist eine weitere Sprache mit einem einheitlichen Prosodiemuster, aus dem sich Informationen über einige Wortgrenzen gewinnen lassen könnten. Das Französische besitzt nicht denselben Stress wie das Englische, aber der Wortakzent fällt auf die letzte Silbe einer rhythmischen Gruppe, so dass die rechte Grenze einer rhythmischen Gruppe zugleich immer auch die rechte Grenze eines Wortes ist. Französische Hörer scheinen in der Lage zu sein, diese Regelmäßigkeit zu nutzen, um die Entdeckung einer Zielsilbe an der Grenze einer rhythmischen Gruppe im Vergleich zu einer identischen Silbe an einer anderen Position zu beschleunigen (Dahan, 1996); die rhythmische Struktur erleichtert somit indirekt die lexikalische Verarbeitung.

Die prosodische Struktur, wie sie im Rhythmus einer Sprache enthalten ist, hilft den Hörern also auf mehrfache Weise bei der effizienten lexikalischen Segmentierung. Der charakteristische Rhythmus einer Sprache ist zweifellos ein reales Phänomen; er spielt nicht nur bei der lexikalischen Segmentierung und bei anderen Formen der Verarbeitung eine Rolle, sondern auch und am offensichtlichsten bei den bevorzugten Versmaßen. Der Rhythmus liefert jedoch keine direkten Signale für die Position von Wortgrenzen, sondern unterstützt die Segmentierung eher indirekt, indem er einen Rahmen schafft, innerhalb dessen die Hörer Hypothesen über mögliche Positionen von Wortgrenzen aufstellen können, oder indem er

durch Inferenzen von der Segmentierung auf hierarchisch höhere Konstituenten die Weiterführung der lexikalischen Segmentierung ermöglicht.

In frühen Untersuchungen zum Sprechrhythmus wurde oft angenommen, dass der Rhythmus direkt kodiert sein sollte, und zwar in Form einer regelmäßigen zeitlichen Abfolge der Einheiten im Sprechsignal; dieser Forschungsstrang endete darin, dass Phonetikforscher die Annahme einer direkten Kodierung zurückwiesen (für eine Übersicht siehe Cutler, 1991). In psycholinguistischen Untersuchungen wurde auch die Möglichkeit angesprochen, dass rhythmische Regelmäßigkeiten existieren und von den Hörern genutzt werden könnten. So gaben Shields, McHugh und Martin (1974) Hörern in einer Untersuchung zur Phonementdeckung sinnlose Wörter vor, die in echte Sätze eingebettet waren, und fanden, dass die Hörer das erste Phonem des sinnlosen Wortes schneller entdeckten, wenn die erste Silbe (in der sich das Target befand) betont war, als wenn sie unbetont war. Der Effekt verschwand jedoch, wenn das sinnlose Wort in eine Reihe anderer sinnloser Wörter eingebettet war, was die Annahme nahelegt, dass die Beschleunigung nicht nur einfach auf einen akustischen Vorteil zurückzuführen war. Die Autoren kamen zu dem Schluß, dass der Zeitverlauf vokalsprachlicher Ereignisse aus der zeitlichen Redundanz im Signal vorhersehbar ist und dass die Hörer die zeitliche Struktur heranziehen können, um kommende Wortakzente vorherzusagen.

Andere Untersuchungen unterstützten diese Sichtweise, nach der Stress vorhergesagt werden kann, indem sie zeigten, dass die Störung der zeitlichen Struktur die Leistung bei vielen Wahrnehmungsaufgaben beeinträchtigt. Zum Beispiel fand Martin (1979), dass die Verlängerung wie auch die Verkürzung eines einzelnen Vokals eine wahrnehmbare kurzfristige Änderung im Tempo eines gesprochenen Satzes verursachen kann und die Reaktionszeiten bei der Phonementdeckung ansteigen lässt. Meltzer, Martin, Mills, Imhoff und Zohar (1976) fanden, dass phonemische Zielreize, die ein wenig aus ihrer Position im normalen Sprachfluß verschoben wurden, langsamer entdeckt wurden. (Dazu entfernt man einen kleinen Abschnitt des Signals direkt vor dem Zielphonem). Buxton (1983) ersetzte in einer Phonementdeckungsaufgabe das Wort, das dem Wort mit dem Zielphonem vorausgeht, durch ein anderes Wort und fand, dass diese Veränderung die Reaktionszeit erhöht, wenn die beiden Wörter eine unterschiedliche Anzahl an Silben enthalten. Alle diese Befunde stimmen mit dem Vorschlag überein, dass Hörer einen regelmäßigen Rhythmus verarbeiten, den sie bei der Vorhersage zeitlicher Muster heranziehen; wenn die Manipulation des Sprechsignals dazu führt, dass sich diese Vorhersagen als falsch erweisen, wird die Erkennung kurzfristig gestört.

Spätere Befunde stellten diese Interpretation jedoch in Frage. Mens und Povel (1986) führten im Niederländischen ein Experiment durch, das dem von Bux-

ton (1983) im Englischen durchgeführten Experiment nachgebildet war; wiederum wurde die zeitliche Modifikation dadurch erreicht, dass das Wort vor dem Zielwort durch ein Wort mit einer unterschiedlichen Silbenanzahl ersetzt wurde (beispielsweise wurde *kat* - Katze — durch *kandidaat* - Kandidat — ersetzt). Es gelang Mens und Povel nicht, die Vorhersagbarkeits Effekte, die in anderen Untersuchungen beobachtet worden waren, zu replizieren. In ähnlicher Weise konnten Pitt und Samuel (1990) den Befund von Shields et al. (1974) nur schwach replizieren; in ihrer Untersuchung zur Phonementdeckung verwendeten sie akustisch kontrollierte Zielwörter, die in einen natürlichen Satzkontext eingebettet waren. Sie fanden, dass die Vorhersagbarkeit nur dann eintrat, wenn das Wort in eine rhythmisch hochgradig regelmäßige Wortliste eingebettet war. Pitt und Samuel vermuteten, dass natürliche Satzkontexte wohl eher wenig Gelegenheit bieten, die Vorhersage hinsichtlich der Position betonter Silben in Anwendung zu bringen.

Zu einer ähnlichen Schlußfolgerung kamen auch Mehta und Cutler (1988), die in einem Phonementdeckungsexperiment Unterschiede in den Effektmustern fanden, die sich bei spontan gesprochenem Material gegenüber gelesenen Sprachmaterial beobachten ließen. Ein Unterschied bestand darin, dass beim Gelesenen, nicht aber beim spontan gesprochenen Material Zielreize, die im Satz weiter hinten stehen, schneller entdeckt wurden als weiter vorne im Satz auftretende Zielreize. Die beiden im Experiment verwendeten Materialmengen waren inhaltlich identisch; das Gelesene Material war vom Originalsprecher von einem Transkript des spontanen Gesprächs abgelesen worden. So kamen Mehta und Cutler zu dem Schluß, dass der Effekt bei der abgelesenen Sprache keine semantische Vorhersagbarkeit reflektiert, sondern eine zeitliche Regelmäßigkeit. Zusammengefasst sprechen diese Befunde jedoch nur für eine wenig bedeutende Rolle der rhythmischen Regelmäßigkeit. Um zu einer echten Vorhersagbarkeit zu gelangen, die die Hörer auf Einzelwortebene nutzen können, muß, wie Pitt und Samuel zeigten, eine durchgängige Regelmäßigkeit bestehen, und dies scheint beim spontanen Sprechen bei weitem nicht der Fall zu sein. Bei Gelesener Sprache kann ein solcher Effekt bei längeren Sätzen ins Spiel kommen, bei denen der Schlußteil leichter zu verarbeiten ist als die vorderen Satzteile.

Tyler und Warren (1987) verwendeten eine sogenannte Wortüberwachungsaufgabe, bei der die Hörer reagieren müssen, wenn sie ein zuvor spezifiziertes Zielwort entdecken. Damit untersuchten sie die Effekte der Störung der zeitlichen Struktur bedeutungsleerer (aber syntaktisch akzeptabler) Sätze als Funktion der Effekte dieser Störung auf die prosodische Gruppierung; längere Entdeckungslatenzen ließen sich dann beobachten, wenn die phonologische Phrasenbildung gestört wurde, was daraufhinweist, dass Gruppierungseffekte, wie sie sich in der prosodischen Struktur realisieren, eine größere Rolle spielen als die einfache zeitliche Vorhersagbarkeit, die sich aus der rhythmischen Regelmäßigkeit ergibt.

Die rhythmische Struktur wird bei der lexikalischen Segmentierung verwendet, jedoch nur indirekt, indem sie Hypothesen über die Position von Wortgrenzen lenkt; die rhythmische Struktur steuert die Verarbeitung gesprochener Sprache jedoch im allgemeinen nicht dadurch, dass sie schon im Vorfeld Voraussagen über die Struktur des gesprochenen Signals an sich zulässt.

Es gibt jedoch einen Weg, wie der Zeitverlauf gesprochener Ereignisse die Wahrnehmung von Wortgrenzen unterstützen kann - diese Unterstützung tritt jedoch nicht auf der Ebene der Satzprosodie, sondern auf der Ebene von Segmenten in Erscheinung. Die Zeitcharakteristik auf Segmentebene variiert mit der Position im Wort, und Hörer können sich dieser Variation in der Segmentierung bedienen. So untersuchte Quene (1992, 1993) zeitverlaufsbezogene Minimalpaare wie *naam op - na mop* im Niederländischen; er fand, dass die verlängerte Dauer eines Konsonanten am Wortende (insbesondere bei sonoren Konsonanten wie [m]) für die Hörer besonders hilfreich war. Die Gesamtdauer einer Silbe (in erster Linie jedoch die Dauer des Vokals) bildete in weiteren Untersuchungen einen zuverlässigen Anhaltspunkt: So fanden Nakatani und Schaffer (1978), dass die relative Silbendauer Hörern ermöglichte, englische Adjektiv-Nomen-Folgen wie *noisy dog* und *bold design* zu unterscheiden, wenn diese in Form iterierten Sprechens dargeboten wurden. (Dabei wird eine natürliche Äußerung durch eine Reihe von Wiederholungen jeweils derselben Silbe, etwa der Silbe *ma*, imitiert.) Rietveld (1980) berichtet über ähnliche Befunde für mehrdeutige Lautketten im Französischen (zum Beispiel *le couplet complet—le couple est complei*). Solche mehrdeutigen Lautketten werden in der natürlichen Sprache wohl nicht oft vorkommen; gleichwohl kommen zeitweise mehrdeutige Sequenzen durchaus vor. Eingebettete Wörter sind ein solcher Fall (so ist *Stau* in *Staub* eingebettet, dies wiederum in *Staupe*); Hörer achten auf die Dauer der Segmente, um zwischenzeitliche Mehrdeutigkeiten, die sich aus solchen Einbettungen ergeben können, nach Möglichkeit zu vermeiden. Christophe, Dupoux, Bertoncini und Mehler (1994) zeigten, dass Neugeborene unterscheiden können, ob eine zweisilbige Sequenz wie *mati* aus einem Wort stammt (*mathematicieri*) oder über die Wortgrenze zwischen zwei Wörtern hinweg entnommen wurde (*panorama typique*); die relative Silbendauer wies zwischen den beiden Typen von Zweisilbern bedeutsame Unterschiede auf. Es ist eindeutig, dass menschliche Hörer fein abgestimmte Kapazitäten für Unterschiede im zeitlichen Verlaufsmuster besitzen, und diese Fähigkeiten helfen bei der Segmentierung genauso wie der Satzrhythmus und die Gruppierung.

## 2.2 Hervorhebung

Wörter, die den Satzakzent tragen, werden schneller verarbeitet als andere Wörter. Somit werden bei der Phonementdeckung Zielreize, die in betonten Wörtern liegen, schneller entdeckt als Zielreize in unbetonten Wörtern (Cutler &

Foss, 1977; Mehta & Cutler, 1988); die Verifikation gehörter Wörter erfolgt schneller bei betonten als bei unbetonten Wörtern (van Donselaar & Lentz, 1994); Aussprachefehler werden bei betonten Wörtern schneller registriert als bei unbetonten Wörtern (Cole, Jakimik & Cooper, 1978; Cole & Jakimik, 1980). Betonte Wörter besitzen eine erhöhte akustische Klarheit (größere spektrale Tonschärfe: Koopmans-van Beinum & van Bergem, 1989; verlängerte Dauer: Klan, 1976; van Samen & Olive, 1990; Eefting, 1991; Dahan & Bernard, 1996), und das kann sicherlich zur Erleichterung ihrer Verarbeitung beitragen.

Gleichwohl beruht der Verarbeitungsvorteil betonter Wörter nicht allein auf akustischen Faktoren. Das zeigt der Befund von Cutler (1976), wonach die vorangehende prosodische Kontur, die zu einem betonten Wort hinführt, als solches schon eine beschleunigte Verarbeitung hervorruft. Cutler nahm Sätze in zwei prosodischen Versionen auf, in der einen war das Wort, das das Zielphonem enthält, kontrastiv betont, in der anderen lag der kontrastive Akzent an anderer Stelle. Hier ein Beispiel für das Zielphonem *Idl*: *She managed to remove the DIRTfront the rüg, but not the berry stains; She managed to remove the dirt from the RÜG, but notfrom their clothes*. Das Wort, das den Zielreiz enthält (in diesem Fall also das Wort *dirt*), wurde dann aus jeder Version herausgeschnitten und durch akustisch identische Kopien dieses Wortes ersetzt, die aus einer dritten Aufnahme desselben Satzes stammen, bei dem kein kontrastiver Akzent verwendet wurde. Das führt zu zwei Versionen eines jeden Experimentalsatzes, in denen die Wörter mit dem Zielphonem akustisch identisch waren, während sich die prosodische Kontur der vorangehenden Wörter unterschied: In einem Fall stimmte die Prosodie damit überein, dass der Satzakkzent auf der Position des Zielwortes liegt, im anderen Fall stimmte die Prosodie damit überein, dass der Akzent an irgendeiner anderen Stelle liegt. Cuder fand, dass die Probanden dennoch signifikant schneller auf das Zielphonem reagierten, wenn es sich in „betonter“ Position befand, als wenn es sich an „unbetonter“ Stelle befand. Da es keine akustischen Unterschiede zwischen den Zielwörtern selbst gab, die dieses Ergebnis erklären könnten, und da der einzige Unterschied im voranstehenden Kontext in der Prosodie lag, mußten die Hörer diese vorausgegangene prosodische Information verwendet haben, um vorherzusagen, wo der Akzent auftreten würde.

In einer späteren Untersuchung von Cutler und Darwin (1981) wurde gezeigt, dass dieser Effekt des vorhergesagten Akzents nicht beeinträchtigt wird, wenn man die Tonhöhenvariation entfernt, wenn die Sätze also in monotoner Sprechweise dargeboten wurden; auch fand der Effekt keine Beeinträchtigung, wenn die Verschlusdauer für den Zielkonsonanten — der, wie das *Idl* im obigen Beispiel, ein Verschluslaut war — manipuliert wurde. Der Effekt scheint somit nicht von irgendeiner bestimmten prosodischen Dimension abzuhängen. Man

kann weiterhin Techniken der Hybridisation des Sprachsignals einsetzen, um die zeitlichen Verlaufsmuster zwischen zwei Versionen einer Äußerung zu vertauschen, so dass sich „unmögliche“ Äußerungen ergeben (beispielsweise eine Äußerung, in der die Kontur der Grundfrequenz daraufschließen lässt, dass der Akzent auf das Wort mit dem Zielreiz zu liegen kommt, während die zeitlichen Verlaufsmuster der vorangehenden Wörter dafür sprechen, dass der Akzent woanders liegt). Bei Einsatz solcher Techniken verschwand der Effekt der Akzentvorhersage (Cutler, 1987), was daraufschließen lässt, dass die Konsistenz zwischen den einzelnen prosodischen Dimensionen wichtig ist, damit Hörer sie effizient nutzen können.

Die stabilen Effekte der Vorhersagbarkeit, die sich in diesen Akzentuntersuchungen beobachten ließen, stehen im Gegensatz zu den fragilen Vorhersagbarkeitseffekten, wenn es um die rhythmische Regelmäßigkeit geht. Interessanterweise scheinen in den Experimenten von Shields et al. (1974) und Meltzer et al. (1976) die sinnlosen Wörter, die Träger des Zielreizes waren, die wesentliche Information des Satzteils, in dem sie auftraten, transportiert zu haben, so dass der Sprecher ihnen wahrscheinlich den Satzakzent zugewiesen hätte. Deshalb ist es möglich, dass die Autoren unbeabsichtigt nicht nur die rhythmische Struktur, sondern auch den Satzakzent manipulierten, und dass die Effekte, die sie beobachteten, eher auf dem letztgenannten Faktor beruhten. Der Unterschied zwischen den Ergebnissen von Buxton (1983) und Mens und Povel (1986) könnte eine ähnliche Ursache haben. Wörter, die den Zielreiz enthalten, waren bei Buxton Nomina, und die rhythmische Manipulation wurde an einem unmittelbar vorausgehenden Adjektiv vollzogen; Nomina tragen den Satzakzent mit größerer Wahrscheinlichkeit als Adjektive, so dass es sich als wahrscheinlich erweist, dass die Manipulation der Untersuchung von Buxton die prosodische Struktur störte, die dem betonten Wort unmittelbar vorausging. Mens und Povel andererseits manipulierten Nomina und (in einigen wenigen Fällen) Verben, und das darauffolgende Wort, das den Zielreiz enthielt, war in fast allen Fällen eine Präposition oder ein Adverb; bei diesen Autoren war somit sehr wahrscheinlich ein betontes Wort von der Manipulation betroffen, während der Zielreiz sowohl in den intakten als auch in den zusammengeschnittenen Sätzen erst nach dem Satz Kern auftrat.

Die Effekte der Akzentvorhersage beruhen sicherlich nicht auf irgendeiner Beziehung des Satzakzents zur zeitlichen Struktur einer Äußerung. Stattdessen wurde die Ansicht vertreten, dass Hörer ihre Aufmerksamkeit auf betonte Wörter richten, weil diese semantisch fokussiert werden und daher die Information transportieren, die für das Erkennen der sprecherseitigen Botschaft besonders wichtig ist. Die semantische Fokussierung als solche führt bei der Phonementdeckung in genau derselben Weise zu schnelleren Reaktionen wie die prosodische Betonung; dies demonstrierten Cutler und Fodor (1979) in

einer Untersuchung, in der der semantische Fokus dadurch manipuliert wurde, dass dem Satz, in dem der Zielreiz vorkam, eine Frage voranging. Sobald sie geortet werden, erfahren fokussierte Wörter eine detailliertere semantische Verarbeitung: Bei Homophonen werden die verschiedenen Bedeutungen dann aktiviert, wenn das Wort im Fokus steht, aber nicht notwendigerweise, wenn es nicht den Fokus bildet (Blutner & Sommer, 1988), so wie auch die Wiedergabe der Oberflächenform eines Wortes wahrscheinlicher ist, wenn sich das Wort bei einem gehörten Satz im Fokus befand (Birch & Garnsey, 1995). Hörer mögen somit aktiv nach betonten Wörtern suchen, weil diese den semantisch zentralen Teil der Botschaft eines Sprechers vermitteln. Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard und Carlson (1995) demonstrierten, wie schnell der Akzent verarbeitet werden kann; in ihrer Untersuchung verfolgten sie die Augenbewegungen der Hörer, die auf einem Display eines von vier Objekten auswählen sollten. Wenn die dargestellte Objektmenge beispielsweise aus einem großen (*large*) roten Quadrat, einem großen (*large*) blauen Kreis, einem kleinen (*small*) roten Quadrat und einem kleinen (*small*) gelben Dreieck bestand, und die Probanden hörten „touch the LARGE red Square“, dann konnten sie das korrekte Objekt schon auswählen, so wie sie das kontrastiv akzentuierte Wort „large“ hörten. Anscheinend setzte sie der kontrastive Akzent in die Lage, das eine Element aus der Menge der großen Objekte auszuwählen, das sich analog zur prosodischen Kontrastierung von einem anderen Objekt abhob, dadurch dass es sich ausschließlich auf dem hervorgehobenen Attribut, nicht aber auf einem der anderen Objektattribute von einem der anderen Objekte unterschied. In der genannten Beispielanordnung kann diese implizit auf genau einen Kontrast gerichtete Interpretation des Akzents aber nur auf das große rote Quadrat im Kontrast zum kleinen roten Quadrat zutreffen; da die Anordnung keinen kleinen blauen Kreis aufweist, scheidet der große blaue Kreis als Zielobjekt schon beim Hören des akzentuierten Wortes „large“ aus.

Die Beziehung zwischen der Akzentsetzung und der semantischen Struktur wird durch eine Anzahl von Untersuchungen bekräftigt, denen zufolge Hörer solche Sätze bevorzugen oder leichter zu verarbeiten finden, in denen der Akzent auf neuer Information liegt, also auf Information, die im Diskurs nicht schon zu einem früheren Zeitpunkt vorkam (Bock & Mazzella, 1983; Terken & Nootboom, 1987; Birch & Clifton, 1995). Auch syntaktische Disambiguierung kann durch die Platzierung des Akzents geleistet werden; Read, Kraak und Boves (1980) fanden, dass in dem mehrdeutigen niederländischen Satz *wiezoent de vrouwi* der Akzent auf dem Verb dazu führt, dass Hörer die Interpretation bevorzugen, nach der die Frau das Subjekt ist und nach dem Objekt der Handlung gefragt wird. Die Erklärung für diesen Befund beziehen die Autoren jedoch auf die Beziehung zwischen Akzent und Informationsstruktur: Der Akzent auf dem Verb machte das anschließende Nomen (*vrouw*) eindeu-

tig unbetont, was impliziert, dass es als das bereits bestehende Thema des Diskurses (das Topik) betrachtet wird, was wiederum impliziert, dass es das grammatische Subjekt des Satzes bildet und das Fragewort somit das grammatische Objekt sein muß.

Die Hervorhebung von Wörtern, die durch die Satzprosodie geleistet wird, wird von Hörern somit genutzt, um Informationen über die semantischen Beziehungen innerhalb der Äußerungen abzuleiten; betonte Wörter erfahren eine effektiv bevorzugte Verarbeitung.

### 3 *Prosodie auf Wortebene*

Wörter können durch die Unterschiede in der Struktur ihrer Segmente eindeutig voneinander unterschieden werden (zum Beispiel *Bein* von *mein*, *Bahn* und *Beil*); in vielen Sprachen können sie sich auch allein an Hand suprasegmentaler Mittel unterscheiden: Im Deutschen (einer Sprache, die Stress berücksichtigt) unterscheidet sich *übersetzen* von *übersetzen*; im Japanischen bedeutet *ame* mit dem Tonhöhenakzent hoch-tief (HL von high-low) *Regen*, im Unterschied zu *ame* mit umgekehrtem Tonhöhenakzent LH, das *Bonbon* bedeutet; im Kantonesischen unterscheidet sich [si] mit hohem bleibenden Ton 1 (in der Bedeutung *Gedicht*) von [si] mit hohem ansteigenden Ton 2 (in der Bedeutung *Geschichte*) von [si] mit niedrigem bleibenden Ton 6 (in der Bedeutung *Zeit*). Die Worterkennung beim Verstehen gesprochener Sprache umfaßt somit die Verarbeitung der prosodischen Struktur, die zur Identität eines Wortes beitragen oder sie sogar allein bestimmen kann.

Der Prozeß des lexikalischen Zugriffs beim Erkennen gesprochener Wörter wird in diesem Band im Kapitel von Frauenfelder und Floccia detailliert beschrieben. In derzeitigen Modellen der Worterkennung wird angenommen, dass der einkommende sprachliche Input mehrere lexikalische Kandidaten aktiviert, die für die Erkennung miteinander konkurrieren. Sowohl übereinstimmende als auch widersprechende Information im Signal kann zum Schicksal eines Wortkandidaten beitragen: Information im Signal, die mit der gespeicherten lexikalischen Repräsentation übereinstimmt, kann die Aktivierung des zugehörigen Wortes erhöhen, während die Aktivierung reduziert wird, wenn die einkommende Information mit dem Gespeicherten nicht in Übereinstimmung zu bringen ist. Keines der aktuellen Modelle zur Erkennung gesprochener Wörter, sei es auf dem Rechner implementiert oder nicht, hat sich bis jetzt speziell der Rolle der prosodischen Information bei diesem Prozeß zugewandt. Es liegt jedoch eine Menge an relevanten experimentellen Befunden vor, die geeignet sind, Fragen der folgenden Art näher zu beleuchten: Schränkt prosodische Information die ersten Phasen der lexikalischen Aktivierung ein? Oder



spielt die prosodische Information nur eine nachgeordnete Rolle, indem sie dann ins Spiel kommt, wenn es darum geht, zwischen alternativen Wortlcandidaten auszuwählen?

### 3.1 Lexikalischer Ton

In lexikalischen Tonsprachen können sich Wörter durch ihre Tonhöhe oder den Tonhöhenverlauf der Silben unterscheiden, wie es im obigen Beispiel aus dem Kantonesischen der Fall ist. Somit wird der Ton nur durch eine einzige suprasegmentale Dimension angezeigt, nämlich durch die Grundfrequenz (FO). Die Information aus der Grundfrequenz kann selbst bei Abwesenheit segmentaler Information hoch informativ sein; so fand Ching (1985, 1988), dass sich die Identifikationsraten bei von den Lippen abgelesenen kantonesischen Wörtern stark verbesserten, wenn Information über die Grundfrequenz gegeben wurde, und zwar in Form von mit der Tonhöhe des Sprechers synchronisierten Impulsen. (Es fand sich jedoch nur eine geringe Verbesserung, wenn die Grundfrequenzinformation beim Lippenlesen englischer Wörter gegeben wurde.) Untersuchungen mit der Methode des lexikalischen Priming im Kantonesischen deuten darauf hin, dass sich die Rolle des Tons einer Silbe bei der Worterkennung analog zur Rolle des Vokals verhält (Chen & Cutler, 1997; Cutler & Chen, 1995); bei einer auditiven lexikalischen Entscheidungsaufgabe üben Überschneidungen im Ton und Überschneidungen im Vokal zwischen einem Prime-Wort und dem Target-Wort analoge Effekte aus.

Obwohl es den Anschein hat, dass ein Kontrast, der durch die Grundfrequenz hergestellt wird, auf Wahrnehmungsebene leicht zu verarbeiten sein sollte (dies ist zum Beispiel vergleichbar mit dem Kontrast zwischen zwei Noten in der Musik), finden es Hörer, die keine Erfahrungen mit einer Tonsprache haben, schwer, Töne zu unterscheiden. Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda und Keller (1996) verglichen die Gleich-ungleich-Urteile von Thai-, Kantonesisch- und Englisch-Sprechern über Töne in Thai und über musikalische Transformationen derselben Töne; Hörer, deren Muttersprache Thai oder Kantonesisch war, konnten die gesprochenen und die musikalischen Töne gleich gut unterscheiden, wohingegen englische Hörer die gespielten Töne signifikant besser unterschieden als die gesprochenen Töne. Lee, Vakoch und Wurm (1996) fanden ebenso, dass englische Hörer Schwierigkeiten hatten, Gleich-ungleich-Urteile über Tonpaare aus dem Kantonesischen oder dem Hochchinesischen (Mandarin) zu treffen; Sprecher der beiden Tonsprachen schnitten durchgehend besser ab als die englischen Hörer (wobei sie jedoch bei den Tonkontrasten ihrer eigenen Sprache bessere Leistungen erbrachten als bei der jeweils anderen Sprache).

Eine der ersten psycholinguistischen Untersuchungen des Tons bei der Worterkennung stammt von Fox und Unkefer (1985), die in einem Klassifikationsexperiment eine kontinuierliche Variation von einem Ton in Mandarin zu einem anderen einsetzten. Der Schnittpunkt, an dem die Hörer in ihrem Experiment dazu übergehen, nicht mehr den einen, sondern den anderen Ton wahrzunehmen, verschob sich in Abhängigkeit davon, ob die Konsonant-Vokal-Silbe, in der der Ton realisiert wurde, ein echtes Wort der Sprache bildete, und zwar entweder nur in Kombination mit dem einen oder nur in Kombination mit dem anderen Ton. (Zum Vergleich gab es Kontrollbedingungen, in denen entweder beide Töne oder keiner von beiden in Kombination mit der Konsonant-Vokal-Verbindung ein echtes Wort der Sprache bildeten.) Dieser lexikalische Effekt trat nur bei muttersprachlichen Hörern auf; englische Hörer zeigten keine derartige Verschiebung, und die beiden Probandengruppen unterschieden sich nicht bei den kontinuierlichen Variationen der Tonhöhe in den Kontrollbedingungen. Da die Klassifikationsaufgabe keine „on-line“-Messung darstellt (das heißt, sie setzt nicht direkt am Prozeß der Worterkennung an), erhellt der Befund von Fox und Unkefer nicht die Frage, ob der Ton schon bei der ersten Aktivierung von Wortkandidaten oder erst bei ihrer Auswahl ins Spiel kommt.

Tonale Information grenzt die Worterkennung jedoch mit geringerer Sicherheit ein als segmentale Information. In einer Untersuchung von Tsang und Hoosain (1979) hörten kantonnesische Probanden schnell dargebotene Sätze und mußten zwischen zwei Transkriptionen des Gehörten auswählen; die beiden Transkriptionen unterschieden sich nur in einem Schriftzeichen, das einen einzelnen Unterschied im Vokal einer Silbe, im Ton, oder in beiden, Vokal und Ton, darstellte. Die Genauigkeit war bei Vokalunterschieden signifikant größer als bei Tonunterschieden, und Unterschiede in beiden wurden nicht genauer auseinandergehalten als Vokalunterschiede allein. Repp und Lin (1990) ließen Hörer im Mandarin Konsonant-Vokal-Silben, die kein Wort der Sprache darstellten, nach ihrem Konsonant, Vokal oder Ton klassifizieren. Die Klassifikation des Tons erfolgte langsamer als die Klassifikation des Konsonanten oder des Vokals. Taft und Chen (1992) fanden, dass die Beurteilung von Schriftzeichen hinsichtlich ihrer Homophonie sowohl in Mandarin als auch in Kantonnesisch langsamer erfolgte, wenn sich die Aussprache der beiden Zeichen nur im Ton unterschied, verglichen mit Vokalunterschieden. In ähnlicher Weise fanden Cutler und Chen (1997), dass kantonnesische Hörer in einer auditiven lexikalischen Entscheidungsaufgabe mit signifikant höherer Wahrscheinlichkeit ein Nichtwort fälschlicherweise als echtes Wort akzeptierten, wenn sich das Nichtwort vom echten Wort nur im Ton unterschied; in einer Gleich-ungleich-Beurteilungsaufgabe waren die Reaktionen derselben Hörer langsamer und weniger genau, wenn sich zwei Silben nur im Ton unterschieden, verglichen mit dem Vorhandensein eines segmentalen Unterschieds. Bei beiden Aufgaben waren Fehler dann am wahrscheinlichsten, wenn der korrekte Ton des echten

Wortes und der falsche Ton des Nichtwortes ähnlich anfangen, oder, mit anderen Worten, wenn die Unterscheidung von der Wahrnehmung her schwer zu treffen war. Ähnliche Effekte ergeben sich bei der Wahrnehmung von Tönen in Thai, in diesem Fall durch Hörer, deren Muttersprache nicht Thai ist: Burnham, Kirkwood, Luksaneeyanawin und Pansottee (1992) fanden, dass die Schwierigkeitsreihenfolge von Tonpaaren, die englischsprachigen Hörern zur Gleich-ungleich-Beurteilung vorgegeben wurden, durch die Anfangstonhöhe der Töne gebildet wird.

Obwohl Tonkontraste in der Grundfrequenz erzeugt werden, werden sie über Vokale realisiert, weshalb sie zusammen mit der Vokalinformation verarbeitet werden. Doch Vokale an sich können sehr früh identifiziert werden; bei einer Konsonant-Vokal-Folge reicht den Hörern der Übergang vom Konsonanten zum Vokal, um den Vokal zu identifizieren (Strange, 1989). Die genannten Befunde lassen darauf schließen, dass Töne oft nicht so schnell identifiziert werden können — bei Aufgaben, bei denen die Reaktion möglichst schnell erfolgen soll, reagieren die Teilnehmer manchmal schon, bevor die tonale Information effektiv verarbeitet wurde. Obwohl also die Toninformation in Sprachen wie dem Kantonesischen für die Unterscheidung zwischen Wörtern maßgeblich ist, könnte es zutreffen, dass die segmentale Information die anfängliche lexikalische Aktivierung stärker bestimmt als die tonale Information.

### 3.2 Lexikalischer Tonhöhenakzent

Im Japanischen besitzen Wörter Muster des Tonhöhenakzents - dabei sind hohe (H) oder niedrige (L) Tonhöhen mit jeder Mora eines mehrsilbigen Wortes assoziiert. So hat beispielsweise das Wort *Tokyo* vier Morae: *to-o-kyo-o*, deren erste niedrigen Tonhöhenakzent trägt, die folgenden drei hohen. Das ganze Wort trägt somit das Muster LHHH. Wie beim lexikalischen Ton werden Kontraste des Tonhöhenakzents mit Hilfe der Variation der Grundfrequenz realisiert. Es gibt viele kurze japanische Wortpaare, die sich nur in ihrem Akzentmuster unterscheiden (so wie das oben erwähnte Wort *ame*), aber nur sehr wenige solcher Paare, an denen längere Wörter beteiligt sind. Es existiert nur eine begrenzte Anzahl an möglichen Mustern. Japanischen Hörern fällt es schwer, gemischt zusammengekopierte Wörter zu verarbeiten, die eine korrekte Segmentsequenz besitzen, aber ein unmögliches, das heißt in der Sprache nicht vorkommendes Akzentmuster aufweisen (Otake, Yoneyama, Cuder & van der Lugt, 1996).

Einige neuere Experimente führten zu der Annahme, dass japanische Hörer die Information über den Tonhöhenakzent in den frühen Phasen der Worterkennung heranziehen können, also bei der Anfangsaktivierung der Wortkandida-

ten. Cutler und Otake (1996) gaben japanischen Hörern einzelne Silben vor, die aus zweisilbigen Wörtern mit unterschiedlichem Akzentmuster ausgeschnitten waren, und ließen sie für jede Silbe angeben, in welchem der beiden Wörter sie ursprünglich gesprochen worden war. So hörten die Probanden etwa *ka* und sollten zwischen *baka* HL und *gaka* LH oder zwischen *kage* HL und *kagi* LH wählen; anders ausgedrückt, sollten sie entscheiden, ob die Silbe hohen Akzent H oder niedrigen Akzent L besaß, da die Silbe in derselben Position der beiden Auswahlwörter vorkam und der phonetische Kontext im Umfeld von *ka* abgeglichen wurde. Die Hörer führten diese Aufgabe mit großer Genauigkeit durch, wobei ihre Erfolgsquote bei Anfangsilben signifikant höher war (80 %) als bei Endsilben (68 %). Das deutet daraufhin, dass die Tonhöheninformation an frühen Stellen des Wortes am eindeutigsten realisiert wird, wo sie den Hörern bei der On-line-Erkennung gesprochener Wörter auch am nützlichsten wäre.

In einer anschließenden Untersuchung zum Wiederholungspriming fanden Cutler und Otake (1999), dass Minimalpaare im Hinblick auf den Tonhöhenakzent, so wie *ante* HL und *ame* LH, ihre Erkennung wechselseitig nicht erleichtern. Wenn also der eine Paarling dargeboten wird, aktiviert das offensichtlich nicht den anderen Paarling, was darauf schließen lässt, dass die fehlende Übereinstimmung im Tonhöhenakzent einen lexikalischen Kandidaten ausschließen kann. In einer Gating-Untersuchung (bei der das Wort, um das es geht, erraten werden muß), boten dieselben Autoren Hörern nach und nach längere Fragmente von Wörtern wie *nimotsu* HLL oder *nimono* LHH an, also Wortpaare, deren Anfangsilben (hier: *nimo*-) dieselbe segmentale Struktur besitzen, aber beim Tonhöhenakzent entgegengesetzte Werte aufweisen. Die falschen Vermutungen der Hörer, die etwa am Ende des ersten Vokals («/») geäußert wurden, betrafen überwältigend oft Wörter mit demselben Akzentmuster wie das tatsächlich gesprochene, zu erkennende Wort. Diese Ergebnisse sprechen sehr stark dafür, dass der japanische Tonhöhenakzent von den Hörern bei der Aktivierung von Wörtern genutzt wird.

Doch wird der Tonhöhenakzent genau wie der lexikalische Ton auf dem Wege der Grundfrequenz erzeugt, weshalb er nur dann zuverlässig identifiziert werden kann, wenn ein beträchtlicher Teil des Segments, das ihn trägt, schon gehört wurde. In der Gating-Untersuchung bildete der Vokal der ersten Silbe dieses notwendige Trägersegment. Waish Dickey (1996) führte ein Experiment zur Gleich-ungleich-Beurteilung durch, in dem japanische Probanden Paare zweisilbiger Wörter oder zweisilbiger Nichtwörter hörten, die entweder gleich waren oder sich in ihrem Tonhöhenakzent oder aber in ihrer Segmentstruktur unterschieden. Übereinstimmend mit den Beobachtungen von Cutler und Chen (1997) beim lexikalischen Ton, fand Waish Dickey, dass „ungleich“-Urteile bei Paaren, deren Tonhöhenakzent variiert wurde, signifikant langsamer

erfolgten als bei Wortpaaren, deren Segmentstruktur sich unterschied. Dies gilt zudem ungeachtet der Position, an der sich der segmentale Unterschied befand; selbst ein Unterschied in einem Vokal am Wortende (wo das Tonhöhenmuster des ganzen zweisilbigen Wortes jenseits allen Zweifels liegen sollte) führte zu signifikant schnelleren Reaktionen als ein Unterschied im Tonhöhenakzent.

### 3.3 Lexikalischer Stress

Die meisten experimentellen Untersuchungen zur Wortprosodie bezogen sich auf den lexikalischen Stress, und die meisten Forschungen wurden auf Englisch durchgeführt. Doch wird im folgenden Abschnitt skizziert, dass der lexikalische Stress bei der Worterkennung im Englischen nicht dieselbe Rolle spielen muß wie in anderen stress-basierten Sprachen.

Im Englischen sind Wortpaare selten, die nicht miteinander verwandt sind und sich nur in ihrem Stressmuster unterscheiden; zwar finden sich häufig Wörter mit entgegengesetztem Stress, die aus demselben Wortstamm in verschiedene Wortformklassen abgeleitet sind (zum Beispiel *import* oder *contest* in ihrer Lesart als Nomen oder als Verb), aber nur wenige solcher Paare sind lexikalisch eindeutig verschieden (beispielsweise *forearm* [*FORE*arm bedeutet *Unterarm*, *fore-ARM* bedeutet *vorbereiten*] oder *insight/incite*). Der Stress könnte im Prinzip also Minimalunterschiede zwischen Wörtern herstellen, in der Praxis kommt dies aber kaum vor.

Die Seltenheit stress-bezogener Minimalpaare gilt auch für das Deutsche, das Niederländische und andere Sprachen, die den Stress als Merkmal besitzen. Jedoch unterscheidet sich die Realisierung des Stress im Englischen ein wenig von der in anderen, eng verwandten Sprachen. Im Englischen enthalten unbetonte Silben fast immer reduzierte Vokale, und die meisten vollen Vokale weisen entweder primäre oder sekundäre Betonung auf. Dieser Zusammenhang ist in anderen germanischen Sprachen bei weitem nicht so stark. So hat beispielsweise das englische Wort *cobra* einen reduzierten Vokal — ein Schwa — in der zweiten Silbe, wo die entsprechenden deutschen und niederländischen Wörter den vollen Vokal [a] aufweisen. In ähnlicher Weise hat das englische Wort *cigar* einen Schwa in der ersten Silbe, wo beim deutschen *Zigarre* und beim niederländischen *sigaar* ein voller Vokal [i] steht. Im Deutschen und Niederländischen treten unbetonte volle Vokale weit häufiger auf als im Englischen.

Dieser sprachvergleichende Unterschied resultiert darin, dass es im Englischen weniger Wortpaare gibt, die sich suprasegmental unterscheiden lassen, bevor man sie auf Segmentebene unterscheiden kann. Man betrachte die Wörter *alibi* und *alinea* (die es sowohl im Deutschen als auch im Niederlän-

dischen gibt); beide fangen mit *ali-* an, aber in einem Wort ist die erste, im anderen Wort die zweite Silbe betont. Solche Paare kommen im Englischen praktisch nicht vor; ein Wort mit Stress auf der ersten Silbe wird dort mit größter Sicherheit in der zweiten Silbe ein Schwa aufweisen (wie im Fall des Wortes *alibi*, so wie es im Englischen existiert). Demzufolge wird bei der Worterkennung im Englischen die frühestmöglich verfügbare Information, die wegen fehlender Übereinstimmung mit der lexikalischen Repräsentation zur Ausmusterung eines Kandidaten führen kann, praktisch immer segmentale Information sein; die Verarbeitung suprasegmentaler Information trägt zur Einschränkung der Wortaktivierung wohl nur wenig Nützliches bei. (Statistische Analysen von Altmann und Carter [1989] wiesen auch tatsächlich nach, dass der Informationswert, den phonetische Segmente im Englischen übermitteln, bei Vokalen in betonten Silben am höchsten ist.)

Englische Hörer rinden die Unterscheidung zwischen vollen und reduzierten Vokalen in der Tat entscheidender als die Unterscheidung zwischen Betonungsebenen; die Montage von Vokalen mit verschiedenen Betonungsmustern führt nur dann zu unakzeptablen Resultaten, wenn die Qualität des Vokals dadurch verändert wird (Fear, Cutler & Butterfield, 1995). In der vorgenannten Untersuchung hörten die Probanden Wort-Tokens wie *audience*, deren primäre Betonung auf dem Anfangsvokal liegt, und *audition*, das eines der seltenen englischen Wörter mit einem unbetonten, aber nicht reduzierten Anfangsvokal darstellt; wurden die (in beiden Fällen nicht reduzierten, aber unterschiedlich betonten) Anfangsvokale dieser Wörter vertauscht, dann schätzten die Hörer die so entstandenen Wort-Tokens im Vergleich zu den unbearbeiteten Original-Tokens als nur unwesentlich anders ein. In einer Untersuchung zur Worterkennung unter Geräuschmaskierung fand Slowiaczek (1990), dass eine Fehlbetonung keinen signifikanten Effekt auf die Wortidentifikation hat, solange die Qualität des Vokals nicht verändert wird. Eine Veränderung der Vokalqualität stört die Worterkennung andererseits sehr wohl: So fanden Bond und Small (1983), dass bei fehlbetonten Wörtern mit Vokalveränderungen in einer Wortüberwachungsaufgabe die korrekte Betonung nicht wiederhergestellt wurde (was darauf hinweist, dass die Probanden die fehlbetonte Form wahrgenommen haben und gar keinen Zugriff auf das intendierte Wort hatten). Weiterhin fand Bond (1981), dass diejenige segmentale Verzerrung, die die Worterkennung am stärksten beeinträchtigen konnte, die Vertauschung von vollen und reduzierten Vokalen ist. Ein Experiment zur Fehlbetonung von Cutler und Clifton (1984) erbrachte den ähnlichen Befund eines weitaus stärkeren Hemmungseffekts durch eine Betonungsverschiebung, wenn die Wörter einen reduzierten Vokal besitzen (*wallet*, *saloori*) — da dies notwendigerweise auch eine Veränderung der Vokalqualität mit sich bringt - im Vergleich zu Wörtern mit zwei vollen Vokalen (*mutmeg*, *canteen*). Wortspiele, an denen eine Stressverschiebung beteiligt ist, wirken nicht (Lagerquist, 1980). Schließlich zeigten

Mattys und Samuel (1997) in einem „Migrations“-Experiment (bei dem die Erkennung von Phantomwörtern durch die Kombination von sprachlichem Material, das auf beide Ohren separat eingespielt wird, induziert wird), dass eine Fehlbetonung in einer betonten Silbe die Konstruktion des Phantomkonzepts hemmt.

Im Englischen wird die Worterkennung nicht erleichtert, wenn man das Stressmuster im voraus kennt: Weder die visuelle noch die auditive lexikalische Entscheidung wird durch die vorherige Spezifikation des Stressmusters beschleunigt (Cutler & Clifton, 1984). Weiterhin gibt es im Englischen bestimmte kanonische Korrelationen zwischen Stressmustern und Wortklassen; zum Beispiel werden zweisilbige Nomina auf der ersten, zweisilbige Verben auf der zweiten Silbe betont. Hörer kennen diese Muster und können sie bei „off-line“-Entscheidungen einsetzen, das heißt, wenn die Reaktionen nicht unter Zeitdruck erfolgen müssen. So sollten die Probanden in den Untersuchungen von Kelly und Mitarbeitern (Cassidy & Kelly, 1991; Kelly, 1988, 1992; Kelly & Bock, 1988) zweisilbige Nichtwörter so in einem Satz verwenden, als ob es Wörter seien; die Probanden behandelten Nichtwörter mit Anfangsbetonung wie Nomina und Nichtwörter mit Endbetonung wie Verben. In ähnlicher Weise sollten die Probanden aus einem Verb ad hoc ein Nomen bilden und wählten in diesem Fall ein Verb mit Anfangsbetonung; wenn sie hingegen ein Nomen ad hoc in die Rolle eines Verbs bringen sollten, wählten sie ein Nomen mit Endbetonung. Diese Art der Verwendung von Betonungsmustern beschleunigt jedoch wiederum nicht die Worterkennung: Ob ein zweisilbiges Wort mit dem kanonischen Muster übereinstimmt oder nicht, hat keinen Einfluß darauf, wie schnell die grammatische Klasse dieses Wortes beurteilt wird — *cigar* wird genauso schnell als Nomen erkannt wie *apple*, und *borrow* wird als Verb genauso schnell erkannt wie *arrive* (Cutler & Clifton, 1984).

In einer weiteren Off-line-Untersuchung von Connine, Clifton und Cutler (1987) sollten Hörer einen mehrdeutigen Konsonanten (der auf einem Kontinuum zwischen [d] und [t] variierte) entweder in *DIgress—TIgress* (wobei *tigress* ein existierendes Wort ist) oder in *diGRESS—tiGRESS* (wobei *digress* als Wort existiert) klassifizieren. Die Reaktionen der Hörer zeigten Effekte, die auf den durch die Betonung bestimmten lexikalischen Status zurückzuführen sind, so dass bei dem *Digress—Tigress-Kontinuum* öfter /t/, bei dem *diGRESS—tiGRESS-Kontinuum* öfter /d/ angegeben wurde. Eindeutig konnten die Hörer die Betonungsinformation im Signal und in ihren gespeicherten Repräsentationen dieser Wörter also nutzen, um die phonetische Ambiguität aufzulösen. Doch kann dieser off-line gewonnene Befund, wie schon die Korrelation zwischen Betonungsmustern und Wortklassen, die Rolle der Betonung im Prozeß der Wortaktivierung (das heißt on-line) nicht weiter klären.

Wenn es bei der Worterkennung primär auf die Identität der Segmente ankommt, dann sollten die wenigen stress-bezogenen Minimalpaare im Englischen — wie zum Beispiel *^era\*:-?«* — letztlich Homophone sein, so wie all die vielen anderen englischen Homophone (*match, count, etc.*). Cutler (1986) zeigte, dass dies in der Tat so ist: In einem crossmodalen Primingexperiment (bei dem die Probanden einen Satz hören und an einer bestimmten Stelle im Satz eine visuelle lexikalische Entscheidung treffen) erleichterten beide Betonungsmuster — *FOREarm* und *foreARM* — die Erkennung von Wörtern, die zu dem einen oder dem anderen Prime-Wort einen semantischen Bezug aufweisen (zum Beispiel *elbow, prepare*). L. Slowiczek (persönliche Mitteilung) fand ähnliche Priming-Effekte für Wörter, die mit beiden Lesarten von Sequenzen der Art *green house* assoziiert sind, also der Betonung als Nominalphrase (*green HOUSE* — *grünes Haus*) und der Betonung als zusammengesetztes Nomen (*GREEN house* — *Gewächshaus*). Englische Hörer scheinen somit im Prozeß *des* ersten Zugriffs auf das Lexikon nicht zwischen den beiden Wortformen zu unterscheiden, die sich nur durch ihre unterschiedliche suprasegmentale Struktur ergeben; im Prozeß der Wortaktivierung spielt der Stress keine Rolle.

Es wurde weiter oben jedoch schon angedeutet, dass diese Sachlage nur für das Englische gelten könnte. Die einzige weitere Betonungssprache, für die eine solide Menge an Experimentalbefunden existiert, ist das Niederländische, doch zumindest dort zeichnet die Befundlage ein anderes Bild. Van Heuven & Haggman (1988) analysierten ein niederländisches Korpus von 70 000 Wörtern, um den Beitrag der Betonung zur Spezifikation der Wortidentität zu ermitteln. Sie fanden, dass die Wörter - vom Wortanfang ab gezählt — im Durchschnitt nach Berücksichtigung von 80 Prozent ihrer Phoneme identifiziert werden konnten; wurde die Betonungsinformation jedoch mit einbezogen, war eine vorwärtsgerichtete Suche durchschnittlich schon nach 66 Prozent der Phoneme erfolgreich. Off-line-Experimente im Niederländischen haben Betonungseffekte auf die Wortidentifikation nachweisen können. Beispielsweise fanden van Heuven (1988) und Jongenburger (1996), dass Hörer eines von zwei niederländischen Wörtern mit segmental identischer, aber in ihrer Betonung unterschiedlicher Anfangssilbe auswählen konnten, wenn ihnen lediglich die erste Silbe dargeboten wurde (Beispiele für solche Wörter sind *ORgel* und *orKES*/"beziehungsweise Minimalpaare wie *SERvisch—servIES*). In einem Gating-Experiment ist eine Fehlbetonung der Worterkennung abträglich, wobei die falsche Betonung von Wörtern mit Endbetonung (*Plloot* statt *piLOOT*) größere Beeinträchtigung hervorruft als die falsche Betonung von Wörtern mit Initialbetonung (*viRUS* statt *Virus*; van Heuven, 1985; van Leyden & van Heuven, 1996; Koster & Cutler, 1997). In einem anderen Gating-Experiment verwendeten Jongenburger und van Heuven (1995a; siehe auch Jongenburger, 1996) betonungsbezogene Minimalpaare (wie *VOORnaam—voorNAAM*), die in einem Satzkontext dargeboten wurden. Hier fand sich interessanterweise, dass die Wortvorschläge der



Hörer nur dann eine korrekte Beurteilung der Betonung der Anfangssilbe des Zielwortes aufwiesen, wenn die gesamte erste Silbe und ein Teil des nachfolgenden Vokals bereits zur Verfügung standen; dies weist darauf hin, dass zumindest bei solchen Minimalpaaren die suprasegmentale Information keine allzu starke Einschränkung auf die Aktivierung von Wörtern ausübt. Mit dieser Annahme stimmt eine crossmodale Priming-Untersuchung im Niederländischen überein, die als direkte Replikation des Experiments von Cutler (1986) angelegt wurde; hier ließen sich keinerlei signifikante Priming-Effekte nachweisen, die auf die initialbetonten Elemente von Betonungspaaren (*VOOR-naarri*) zurückzuführen wären, und bei den Elementen mit Endbetonung (*voorNAAM*) ergaben sich uneindeutige Ergebnisse (Jongenburger & van Heuven, 1995b; Jongenburger, 1996).

Dennoch sprechen neuere Ergebnisse dafür, dass die Fehlbetonung eines holländischen Wortes ihre lexikalische Aktivierung verhindern kann. Diese Untersuchungen verwendeten eine größere Population von Wörtern, als sie durch die geringe Anzahl von Betonungs-Minimalpaaren gewährleistet werden kann. Bei der „Wortentdeckungsaufgabe“ werden eingebettete Wörter langsamer entdeckt, wenn sie innerhalb einer Lautkette auftreten, die wiederum zu einem längeren Wort fortgesetzt werden könnte; so wird das englische Wort *mess* in der Kette *doMES* wahrscheinlich deshalb langsamer erkannt als in der Kette *neMES*, weil *doMES* bei weiterer Ergänzung das Wort *domestic* bilden kann, während es kein existierendes Wort gibt, zu dem sich *neMES* erweitern ließe (McQueen, Norris & Cutler, 1994). Dieser Befund lässt sich im Niederländischen replizieren: *zee* (See) ist in *muzee* (das zu *museum* fortgesetzt werden kann) schwerer auszumachen als in *luzee*. Wenn *muzee* jedoch nicht, wie bei *museum*, auf der zweiten Silbe betont wird, sondern stattdessen auf der ersten Silbe, wenn die Probanden also *MUzee* und *LUzee* hören, dann gibt es bei den Entdeckungszeiten für *zee* keinen signifikanten Unterschied mehr (van Donselaar, Cutler & Koster, in Vorbereitung). Das lässt daraufschließen, dass in diesem Fall keine Konkurrenz mit *museum* bestand, einfach weil dieses Wort durch einen Input, dem das korrekte Betonungsmuster fehlt, nicht aktiviert wurde. So ist auch das Fragment *aLI-* ein Prime für *aLinea*, aber nicht für *Alibi*, während das Fragment *Ali-Alibi*, aber nicht *aLinea* aktiviert (van Donselaar, Koster & Cutler, in Vorbereitung); dieser Befund wurde auch für ähnliche Fragmente spanischer Wörter beobachtet, die spanischen Hörern vorgegeben wurden (zum Beispiel die ersten beiden Silben von *PRINcipe* oder *prinCIPio*; Soto, Sebastián-Galles & Cutler, in Vorbereitung).

Es wurde nicht versucht, diese letztgenannten Experimente im Niederländischen auf englisch durchzuführen. Können wir tatsächlich sicher sein, dass sich dieselben Befundmuster bei Einsatz dieser neuen Experimentalmethoden nicht auch im Englischen ergeben würden? Es ist jedoch so, dass beide Experimente

im Englischen einfach nicht repliziert werden können. Für das lexikalische Konkurrenzexperiment (*zee* in *mu<sub>z</sub>ee*) sind Wörter nötig, die mit zwei starken Silben anfangen und die ein einziges eingebettetes Wort enthalten; im Englischen gibt es keine ausreichende Anzahl solcher Wörter. Das Priming-Experiment mit Fragmenten (*alt-* in *alibi* und *aline*) benötigt ebenfalls Wortpaare, die mit zwei starken Silben anfangen; äquivalente Wörter im Englischen (wie *alibi*) enthalten in einer der relevanten Silben aber einen reduzierten Vokal. Die Tatsache, dass solche Experimente im Englischen nicht möglich sind, ist natürlich an sich schon informativ: Sie bedeutet, dass sich im Englischen für die Hörer die Gelegenheit, die Betonung in den frühen Phasen der Worterkennung heranzuziehen, selten ergibt, und dass die Wörter praktisch immer, ohne Bezug auf die Betonung, durch eine segmentale Analyse unterschieden werden können.

Welchen Schluß können wir dann bezüglich der Rolle der Betonung bei der Worterkennung ziehen? Indirekt spielt sie auf jeden Fall eine Rolle, weil betonte Silben akustisch zuverlässiger sind als unbetonte Silben. Deshalb werden betonte Silben leichter identifiziert als unbetonte Silben, wenn man sie aus ihrem Originalkontext herausschneidet (Lieberman, 1963), und Verzerrungen im Sprechsignal werden in betonten Silben mit größerer Wahrscheinlichkeit entdeckt (Cole et al., 1978; Cole & Jakimik, 1980; Browman, 1978; Bond & Garnes, 1980). In spontaner Sprache erfolgt die Entdeckung eines Zielphonems am Wortanfang ebenfalls schneller bei lexikalisch betonten Silben (Mehta & Cutler, 1988); beim spontanen Sprechen sind die akustischen Unterschiede zwischen betonten und unbetonten Silben ziemlich groß, bei abgelesenem sprachlichem Material im Labor treten diese Unterschiede nicht auf. Bei Wörtern, die in Gating-Experimenten Stück für Stück vorgegeben werden, werden betonte Silben früher erkannt als unbetonte Silben, und dies gilt für spontan gesprochenes, aber nicht für gelesenes Material (McAllister, 1991).

Das muß jedoch nicht heißen, dass der Kontrast zwischen betonten und unbetonten Silben allen Hörern auffallen würde. Französische Sprecher, in deren Sprache Wörter nicht durch Betonung unterschieden werden, zeigen große Schwierigkeiten, Betonungskontraste bei sinnlosem Material zu verarbeiten, beispielsweise wenn sie entscheiden sollen, ob der Reiz *bopelo* mit einem zuvor dargebotenen *bopelo* oder mit *boPElo* übereinstimmt (Dupoux, Pallier, Sebastian-Galles & Mehler, 1997). Dieselben Kontraste fallen Sprechern des Spanischen leicht, denn im Spanischen werden Wörter durch ihre Betonung unterschieden. Es sollte nicht vergessen werden, dass sich dieser gesamte Abschnitt mit Sprachen beschäftigte, in denen eine freie Betonung herrscht. Es gibt bis jetzt noch keine direkten Befunde über die Rolle des Stress (zum Beispiel die Effekte falscher Betonungen) in Sprachen mit festliegendem Stress, in denen

zwar Kontraste zwischen betonten und unbetonten Silben existieren, die aber nicht dazu dienen, Wörter voneinander zu unterscheiden. Indirekte Hinweise ergeben sich aus den oben genannten Befunden von Suomi, McQueen und Cutlet (1997) zur Wort-entdeckung im Finnischen - die aus dem vorangehenden einsilbigen Kontext ausgeschnittenen Wörter kann man zumindest so sehen, dass sie nicht ihren kanonischen Stress tragen, aber dies hatte keine nachteiligen Effekte auf die Wort-erkennung. Bei allem gibt es aber noch genügend Freiraum für neue Befunde aus Sprachen wie dem Finnischen oder Polnischen.

Für Sprachen, in denen Stress frei variieren kann, spricht die Befundlage jedoch dafür, dass Stress in denjenigen Sprachen, in denen er bedeutsame Information zur Wortidentifikation beiträgt, bei der anfänglichen Aktivierung lexikalischer Einträge relevant ist; das Englische gehört nicht zu diesen Sprachen. Leider erweist sich deshalb die Sprache, in der die meisten psycholinguistischen Forschungen (auf jedem Gebiet) betrieben werden, als wenig repräsentativ, was die Rolle der Wortprosodie bei der Worterkennung betrifft.

#### 4 *Schlußbemerkung*

Im Zentrum des vorliegenden Kapitels stand der Prozeß der Erkennung gesprochener Wörter und die Art und Weise, in der die prosodische Struktur diesen Aspekt des Hörens direkt beeinflusst. Natürlich gibt es viel allgemeinere indirekte Einflüsse, die man hätte berücksichtigen können. Beispielsweise spielt die Prosodie eine Rolle bei der allgemeinen Verständlichkeit; so wie Sätze mit zulässiger Syntax leichter zu verstehen sind als Sätze mit normabweichender Syntax, Sätze mit plausibler Semantik leichter als Sätze mit unplausiblem Inhalt und Sätze mit exakter phonetischer Realisation leichter als Sätze mit verzerrter phonetischer Struktur, so sind auch Sätze mit einer intakten prosodischen Struktur leichter verständlich als Sätze, deren prosodische Struktur auf die eine oder andere Weise gestört ist. Dies wurde in vielen Sprachen (darunter auch im Deutschen) nachgewiesen, und es ist keine Frage, dass die Worterkennung zu den Komponenten der Sprachverarbeitung gehört, die von solchen Manipulationen betroffen wären. Solche allgemeinen Überlegungen lagen jedoch nicht im Einzugsbereich dieses Kapitels.

Der spezifische Beitrag der prosodischen Struktur zur Worterkennung, so wurde behauptet, kommt einerseits durch Prosodie auf der Satzebene — wobei Rhythmus und Gruppierung für die Entdeckung von Wortgrenzen relevant sind und die Hervorhebung die lexikalische Verarbeitung unterstützen kann - und andererseits von der prosodischen Struktur der Wörter selbst, die, sofern sie geeignete Information enthält, für die Unterscheidung zwischen alternativen Wortkandidaten genutzt wird. In allen diesen Forschungsbereichen gibt es fast

keine experimentellen Belege für das Deutsche, obwohl Anhaltspunkte aus eng verwandten Sprachen vorliegen. Die Befundübersicht lässt die allgemeine Schlußfolgerung begründet erscheinen, dass die prosodische Struktur — wie man es auch für jede andere Ebene der linguistischen Struktur annehmen kann — von Hörern beim Prozeß der Worterkennung in dem Ausmaß herangezogen wird, in dem sie relevante und weiterführende Information bereitzustellen vermag.

### *Danksagung*

Der im vorliegenden Kapitel gegebene Überblick überschneidet sich in starkem Ausmaß mit den Abschnitten zur Worterkennung in der von Cutler, Dahan und van Donselaar (1997) zusammengestellten Literaturübersicht. Ich bin meinen Koautorinnen Delphine Dahan und Wilma van Donselaar sehr dankbar, dass sie mich bei diesem mühsamen, aber letztlich doch ertragreichen Projekt begleitet haben; das vorliegende Kapitel verdankt vieles ihrem Rat und ihrem Problemverständnis. Auch danke ich dem Verlag und den Herausgebern von *Language and Speech* für ihre Erlaubnis, die Arbeit für den genannten Artikel in diesem Zusammenhang noch einmal zu verwenden. Weiterhin gilt mein Dank Jens Bölte für seine Unterstützung beim Aufspüren von Untersuchungen zur Worterkennung im Deutschen.

### *Literatur*

- Altmann, G. T. M. & Carter, D. M. (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech and Language*, 3, 265—275.
- Beach, C. M. (1991). The Interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30, 644-663.
- Birch, S. & Clifton, C. E. (1995). Focus, accent and argument structure: Effects on language comprehension. *Language and Speech*, 38, 365-391.
- Birch, S. L. & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, 34, 232-267.
- Blutner, R. & Sommer, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language*, 27, 359-367.
- Bock, J. K. & Mazzella, J. R. (1983). International marking of given and new Information: Some consequences for comprehension. *Memory and Cognition*, 11, 64-76.
- Bond, Z. S. (1981). Listening to elliptic speech: Pay attention to stressed vowels. *Journal of Phonetics*, 9, 89-96.
- Bond, Z. S. & Garnes, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 115-132). Hillsdale, NJ: Erlbaum.
- Bond, Z. S. & Small, L. H. (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception and Psychophysia*, 34, 470—474.

- Bradley, D. C., Sánchez-Casas, R. M. & Garda-Albea, J. E. (1993). The Status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes*, 8, 197-234.
- Browman, C. P. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics*, 42.
- Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F. & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. *Proceedings of the Fourth International Conference on Spoken Language Processing*. 2514-2516). Philadelphia.
- Burnham, D., Kirkwood, K., Luksaneeyanawin, S. & Pansottee, S. (1992). Perception of Central Thai tones and Segments by Thai and Australian adults. *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium of Language and Linguistics* (pp. 546-560). Bangkok: Chulalongkorn University Press.
- Buxton, H. (1983). Temporal predictability in the perception of English speech. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 111-121). Heidelberg: Springer.
- Cassidy, K. W. & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, 30, 348-369.
- Chen, H.-C. & Cutler, A. (1997). Auditory priming in spoken and printed word recognition. In H.-C. Chen (Ed.), *The Cognitive Processing of Chinese and Related Asian Languages* (pp. 77-81). Hongkong: Chinese University Press.
- Ching, Y. C. (1985). Lipreading Cantonese with voice pitch. Paper presented to the 109th meeting, Acoustical Society of America, Austin (Abstract *Journal of the Acoustical Society of America*, 77, Supplement 1, pp. 39-40).
- Ching, Y. C. (1988). Voice pitch information for the deaf. *Proceedings of the First Asian-Pacific Regional Conference on Deafness* (pp. 340-343). Hongkong.
- Christophe, A., Dupoux, E., Bertoncini, J. & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, 95, 1570-1580.
- Cole, R. A. & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America*, 67, 965-970.
- Cole, R. A., Jakimik, J. & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, 64, 44-56.
- Collier, R. & 't Hart, J. (1975). The role of intonation in speech perception. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 107-121). Heidelberg: Springer.
- Connine, C. M., Clifton, C. E. & Cutler, A. (1987). Lexical stress effects on phonetic categorization. *Phonetica*, 44, 133-146.
- Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics*, 20, 55-60.
- CurJer, A. (1986). *Forhear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29, 201-220.

- Cutler, A. (1987). Components of prosodic effects in speech recognition. *Proceedings of the Eleventh International Congress of Phonetic Sciences* (pp. 84–87). Tallinn, Esland.
- Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord & R. Carlson (Eds.), *Music, Language, Speech and Brain* (pp. 157-166). London: MacMillan.
- Cutler, A. (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes*, 12, 839-845.
- Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.
- Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.
- Cutler, A. & Chen, H.-C. (1995). Phonological similarity effects in Cantonese word recognition. *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 106-109). Stockholm.
- Cutler, A. & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, 59, 165-179.
- Cutler, A. & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes* (pp. 183-196). Hillsdale, NJ: Erlbaum.
- Cutler, A., Dahan, D. & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141-201.
- Cutler, A. & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception and Psychophysics*, 29, 217-224.
- Cutler, A. & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, 7, 49-59.
- Cutler, A. & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, 20, 1-10.
- Cutler, A., Mehler, J., Norris, D. G. & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385–400.
- Cutler, A., Mehler, J., Norris, D. G. & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24, 381–410.
- Cutler, A. & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.
- Cutler, A., Norris, D. G. & Williams, J. N. (1987). A note on the role of phonological expectations in speech segmentation. *Journal of Memory and Language*, 26, 480–487.
- Cutler, A. & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824-844.
- Cutler, A. & Otake, T. (1996). The processing of word prosody in Japanese. *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 599–604). Adelaide.

- Cuder, A. & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, 105, 1877-1888.
- Dahan, D. (1996). The role of rhythmic groups in the segmentation of continuous French speech. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1185-1188). Philadelphia.
- Dahan, D. & Bernard, J. M. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech*, 39, 341-374.
- van Donselaar, W., Cuder, A. & Koster, M. (in preparation). *Voomaam* is not a homophone: Lexical prosody and lexical access in Dutch.
- van Donselaar, W., Koster, M. & Cuder, A. (in preparation). Lexical stress and lexical activation in Dutch.
- van Donselaar, W. & Lentz, J. (1994). The function of sentence accents and given/new Information in speech processing: Different strategies for normal-hearing and hearing-impaired listeners? *Language and Speech*, 37, 375-391.
- Dupoux, E. & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language*, 29, 316-335.
- Dupoux, E., Pallier, C., Sebastian-Galles, N. & Mehler, J. (1997). A destressing deafness in French? *Journal of Memory and Language*, 36, 406-421.
- Eefting, W. (1991). The effect of 'information value' and 'accentuation' on the duration of Dutch words, syllables and Segments. *Journal of the Acoustical Society of America*, 89, 412-424.
- Fear, B. D., Cutler, A. & Buttenneid, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, 1893-1904.
- Fox, R. A. & Unkefer, J. (1985). The effect of lexical Status on the perception of *of tone*. *Journal of Chinese Linguistics*, 13, 69-90.
- van Heuven, V. J. (1985). Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America*, 78, 21.
- van Heuven, V. J. (1988). Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing. *Proceedings of Speech '88, 7th FASE Symposium* (pp. 811-818). Edinburgh.
- van Heuven, V. J. & Hagman, P. J. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans & A Hulk (Eds.), *Linguistics in the Netherlands 1988* (pp. 59-68). Dordrecht: Foris.
- Höhle, B. & Schriefers, H. (1995). Ambisyllabizität im Deutschen: Psycholinguistische Evidenz. *Akten des 29. Linguistischen Kolloquiums*. Tübingen: Niemeyer.
- Jongenburger, W. (1996). *The role of lexical stress during spoken-word processing*. Ph. D. thesis, Leiden.
- Jongenburger, W. & van Heuven, V. J. (1995a). The role of linguistic stress in the time course of word recognition in stress-accent languages. *Proceedings of the Fourth European Conference on Speech Communication and Technology* (pp. 1695-1698). Madrid.

- Jongenburger, W. & van Heuven, V. J. (1995b). The role of lexical stress in the recognition of spoken words: prelexical or postlexical? *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 368-371). Stockholm.
- Kakehi, K., Karo, K. & Kashino, M. (1996). Phoneme/syllable perception and the temporal structure of speech. In T. Otake & A. Cutler (Eds.) *Phonological Structure and Language Processing: Cross-Linguistic Studies* (pp. 125-143). Berlin: Mouton de Gruyter.
- Kearns, R. K. (1994). *Prelexical speech processing in mono- and bilinguals*. PhD thesis, University of Cambridge.
- Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language*, 27, 343-358.
- Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99, 349-364.
- Kelly, M. H. & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 389-403.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208-1221.
- Kolinsky, R., Morais, J. & Cluytens, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language*, 34, 19-40.
- Koopmans-van Beinum, F. J. & van Bergem, D. R. (1989). The role of 'given' and 'new' in the production and perception of vowel contrasts in read text and in spontaneous speech. *Proceedings of the European Conference on Speech Communication and Technology* (pp. 113-116). Edinburgh.
- Koster, M. & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. *Proceedings of the Fifth European Conference on Speech Communication and Technology* (pp. 2167-2170). Rhodes.
- Lagerquist, L. M. (1980). Linguistic evidence from paranomasia. *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*, 185-191.
- Lee, Y.-S., Vakoch, D. A. & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25, 527-542.
- Lehiste, I., Olive, J. P. & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60, 1199-1202.
- van Leyden, K. & van Heuven, V. J. (1996). Lexical stress and spoken word recognition: Dutch vs. English. In C. Cremers & M. den Dikken (Eds.), *Linguistics in the Netherlands 1996* (pp. 159-170). Amsterdam: John Benjamins.
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172-187.
- Martin, J. G. (1979). Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America*, 65, 1286-1297.
- Mattys, S. L. & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language*, 36, 87-116.



- McAllister, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech*, 34, 1-26.
- McQueen, J. M., Norris, D. G. & Cuder, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning Memory and Cognition*, 20, 621-638.
- Mehler, J., Dommergues, J.-Y., Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20, 298-305.
- Mehta, G. & Cuder, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, 31, 135-156.
- Meiner, R. H., Martin, J. G., Mills, C. B., Imhoff, D. L. & Zohar, D. (1976). Reaction time to temporally displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 277-290.
- Mens, L. & Povel, D. (1986). Evidence against a predictive role for rhythm in speech perception. *The Quarterly Journal of Experimental Psychology*, 38A, 177-192.
- Nakatani, L. H. & Schaffer, J. A. (1978). Hearing „words“ without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, 63, 234—245.
- Nooteboom, S. G., Brokx, J. P. L. & de Rooij, J. J. (1978). Contributions of prosody to speech perception. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the Perception of Language* (pp. 75-107). Chichester: Wiley.
- Otake, T., Hatano, G., Cuder, A. & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 358—378.
- Otake, T., Hatano, G. & Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cuder (Eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies* (pp. 183-201). Berlin: Mouton de Gruyter.
- Otake, T., Yoneyama, K., Cuder, A. & van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, 100, 3831—3842.
- Pallier, C., Sebastián-Galka, N., Felguera, T., Christophe, A. & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language*, 32, 373-389.
- Peretz, I., Lussier, I. & Bcland, R. (1996). The roles of phonological and Orthographie code in word stem completion. In T. Otake & A. Cutler (Eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies* (pp. 217-226). Berlin: Mouton de Gruyter.
- de Pijper, J. R. & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, 96, 2037-2047.
- Pitt, M. A. & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 564-573.
- Quene, H. (1992). Oration cues for word segmentation in Dutch. *Journal of Phonetics*, 20, 331-350.
- Quene, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America*, 94, 2027-2035.

- Read, C. Kraak, A. & Boves, L. (1980). The Interpretation of ambiguous who-questions in Dutch: The effect of intonation. In W. Zonneveld & F. Weerman (Eds.), *Linguistics in the Netherlands 1977-1979* (pp. 389-410). Dordrecht: Foris.
- Repp, B. H. & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception. *Journal of Phonetics*, 18, 481-495.
- Rietveld, A. C. M. (1980). Word boundaries in the French language. *Language and Speech*, 23, 289-296.
- deRooy, J. J. (1976). Perception of prosodic boundaries. *IPO Annual Progress Report*, 11, 20-24.
- van Santen, J. P. H. & Olive, J. P. (1990). The analysis of contextual effects on segmental duration. *Computer Speech and Language*, 4, 359-390.
- Schreuder, R. & Baayen, R. H. (1994). Prefix Stripping re-revisited. *Journal of Memory and Language*, 33, 357-375.
- Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, 71, 996-1007.
- Sedivy, J., Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K. & Carlson, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 375-380). Hillsdale, NJ: Erlbaum.
- Segui, J., Frauenfelder, U. H. & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, 72, 47-AT7.
- Shields, T. L., McHugh, A. & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, 102, 250-255.
- Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, 33, 47-68.
- Soto, S., Sebastian-Galles, N. & Cutler, A. (submitted). Segmental and suprasegmental cues for lexical access in Spanish.
- Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America*, 85, 2135-2153.
- Streeter, L. A. (1978). Acoustic determinants of phrase boundary location. *Journal of the Acoustical Society of America*, 64, 1582-1592.
- Suomi, K., McQueen, J. M. & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, 36, 422-444.
- Taft, M. & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language Processing in Chinese* (pp. 151-172). Amsterdam: Elsevier.
- Taft, M. & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, 24, 320-335.
- Terken, J. & Nootboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, 2, 145-163.

- Tsang, K. K. & Hoosain, R. (1979). Segmental phonemes and tonal phonemes in comprehension of Cantonese. *Psychologin*, 22, 222-224.
- Tyler, L. K. & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language*, 26, 638-657.
- Vroomen, J. & de Gelder, B. (1994). Speech segmentation in Dutch: No role for the syllable. *Proceedings of the Third International Conference on Spoken Language Processing Vol 3* (pp. 1135-1138). Yokohama.
- Vroomen, J., van Zon, M. & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition*, 24, 744-755.
- Walsh Dickey, L. (1996). Limiting-domains in lexical access: Processing of lexical prosody. In M. Dickey & S. Tunstall (Eds.), *University of Massachusetts Occasional Papers in Linguistics 19: Linguistics in the Laboratory* (pp. 133-155).
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.
- Zwitserslood, P., Schriefers, H., Lahiri, A. & van Donselaar, W. (1993). The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 260-271.