

Linguistische
Arbeiten

247

Herausgegeben von Hans Altmann, Peter Blumenthal, Herbert E. Brekle,
Hans Jürgen Heringer, Heinz Vater und Richard Wiese

Bernd Pompino-Marschall

Die Silbenprosodie

Ein elementarer Aspekt der Wahrnehmung
von Sprachrhythmus und Sprechtempo

Max Niemeyer Verlag
Tübingen 1990



CIP-Titelaufnahme der Deutschen Bibliothek

Pompino-Marschall, Bernd : Die Silbenprosodie : ein elementarer Aspekt der Wahrnehmung von Sprachrhythmus und Sprechtempo / Bernd Pompino-Marschall. – Tübingen : Niemeyer, 1990

(Linguistische Arbeiten ; 247)

NE: GT

ISBN 3-484-30247-X ISSN 0344-6727

© Max Niemeyer Verlag GmbH & Co. KG, Tübingen 1990

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Printed in Germany.

Druck: Weihert-Druck GmbH, Darmstadt

Zeit ergibt sich erst anhand der kürzeren und gedehnten Zwischenräume der Dauer, die durch [...] Bewegung und Veränderung beansprucht wird, wenn ein Zustand dem anderen weicht und nachfolgt, die nebeneinander nicht bestehen können.

AUGUSTINUS, *Civitas Dei* XI, 6.

Die vorliegende Arbeit ist eine überarbeitete Version meiner von der Philosophischen Fakultät für Sprach- und Literaturwissenschaft II der Ludwig-Maximilians-Universität München im Dezember 1988 für das Fachgebiet "Phonetik und Sprachliche Kommunikation" angenommenen Habilitationsschrift.

Ich möchte an dieser Stelle den vielen, die auf ebenso vielfältige Art und Weise zur Entstehung dieser Arbeit beigetragen haben, meinen herzlichen Dank aussprechen.

Von den vielen, die ich hier nicht alle einzeln nennen kann, gilt mein Dank insbesondere meinem Lehrer, Herrn Prof. Dr. Hans G. Tillmann, der während all der Zeit ein wohlwollend-kritisches Auge auf den Fortgang der hier referierten Untersuchungen hatte und ein offenes Ohr für immer wieder auftauchende Fragen.

Danken möchte ich auch all den Mitarbeitern und Studenten des Instituts für Phonetik und Sprachliche Kommunikation, die mit mir die hier aufgeworfenen Fragen in langen Stunden diskutiert haben, und denen ich dadurch manche Anregung für die vorliegende Arbeit verdanke.

Besonders bedanken muß ich mich bei all meinen Versuchspersonen, die die oft sehr zeitaufwendigen Experimente über sich ergehen ließen.

Gedankt sei nicht zuletzt auch der Deutschen Forschungsgemeinschaft, von der die Durchführung der experimentellen Untersuchungen gefördert wurde.

INHALT

Vorbemerkungen	1
1. Die makroskopische zeitliche Strukturierung gesprochener Sprache	6
1.1. Die Isochronie–Hypothese	6
1.2. Das Problem des Sprachrhythmus und allgemeine Theorien der Rhythmuswahrnehmung und –produktion	8
1.3. Isochronie, Kontrolle der Dauer und zeitliche Kompensation	10
2. Frühe Untersuchungen zur Lokalisation des Taktes in lautsprachlichen Äußerungen	17
3. Das P–center–Phänomen	20
4. Die Psychoakustik des ‘Ereigniszeitpunktes’	27
5. Das P–center als Phänomen der Sprachwahrnehmung wie der Sprachproduktion	33
6. Experimente zum Rhythmuspunkt in gleichmäßig produzierten Sequenzen von Einzelsilben	39
6.1. Generelle Methode	39
6.2. Darstellung der Einzelexperimente	41
7. Experimente zur perzeptuellen Relevanz gemessener Anisochronie	70
8. Experimente zum Verhältnis von akustischer Anisochronie und physiologischen Abläufen	80
8.1. Darstellung der Einzelexperimente	81

9. Experimente zum Einfluß der segmentalen Komposition der Testsilben auf die P-center-Wahrnehmung	124
9.1. Zum P-center von CV-Silben	124
9.1.1. Experimente zum Einfluß systematischer Segmentdauervariationen an quasinatürlichen /ma/-Silben	124
9.1.2. Experimente zum Einfluß systematischer Segmentdauervariation bei den /ma/-Silben pegelverlaufsgleichen nichtsprachlichen Rechtecksignalanalogen	136
9.1.3. Experimente zum Einfluß der systematischen Segmentdauervariation bei quasinatürlichen /ji/-Silben	146
9.1.4. Experimente zum Einfluß der systematische Segmentdauervariation bei pegelnormierten /ji/-Silben	152
9.1.5. Ergebnis der statistischen Auswertung der Experimente mit systematischer Segmentdauervariation in CV-Silben bzw. nichtsprachlichen CV-Analogen	158
9.2. Zum P-center von VC-Silben	168
9.2.1. Experimente zum Einfluß systematischer Segmentdauervariationen bei quasinatürlichen /am/-Silben	168
9.2.2. Experimente zur systematischen Segmentdauervariation bei den /am/-Silben pegelverlaufsgleichen Rechtecksignalanalogen	174
9.2.3. Ergebnis der statistischen Auswertung der Experimente mit systematischer Segmentdauervariation in VC-Silben bzw. nichtsprachlichen VC-Analogen	180
9.3. Ist die Position des P-centers unabhängig von der phonetischen Kategorisierung?	184
9.4. Ist die Lage des P-centers einer Silbe unabhängig von deren Umgebung?	188
 10. Ein psychoakustisches Modell der P-center-Bestimmung	 191
 Anhang zu Kapitel 10: Listings der FORTRAN-Programme zur Simulation des psychoakustischen Modells	 204

11. Mit dem P-center verbundene Effekte in der Wahrnehmung gesprochener Sprache	222
12. Zur Funktion prosodisch definierter Ereignispunkte innerhalb einer Theorie der Sprachwahrnehmung	231
12.1. Die Sprachwahrnehmung im Rahmen einer allgemeinen Theorie der lautsprachlichen Kommunikation	231
12.2. Prosodisch definierte Ereignispunkte und die zeitliche Strukturierung lautsprachlicher Äußerungen	250
Zusammenfassung und Ausblick	259
Literatur	261

VORBEMERKUNGEN

Das Ziel der vorliegenden Arbeit, die in ihrem eigentlichen Kern eine größere experimentelle Studie darstellt, die in der Hauptsache dem Bereich der perzeptiven Phonetik zuzuordnen ist, ist es, zum besseren Verständnis einiger Aspekte der sprachlichen Rhythmus- und Tempowahrnehmung beizutragen. Wie generell bei experimentellen Arbeiten dieser Art, mußte auch in dieser Studie von einem recht eingeschränkten experimentellen Paradigma ausgegangen werden. Im vorliegenden Fall waren dies Untersuchungen zur Wahrnehmung eines subjektiv gleichmäßigen Rhythmus bei sehr einfachem sprachlichen und nichtsprachlichen Material sowie zur Produktion rhythmisch gleichmäßiger Sequenzen von Einzelsilben. Diese Untersuchungen haben allerdings in ihrem Ergebnis Effekte erklärbar gemacht, die – neben anderen, in dieser Arbeit ausdrücklich ausgeklammerten – als grundlegend für die Wahrnehmung von Rhythmus und Tempo in der gesprochenen Sprache, ja darüber hinausgehend für das Funktionieren der lautsprachlichen Kommunikation allgemein betrachtet werden können.

Bei der Wahl des Titels dieser Arbeit habe ich mich an die Modellierung der phonetischen Ereignisse durch Hans G. Tillmann als dreifach prosodisch strukturierte Phänomene angelehnt: Ausgehend von einem Gedankenexperiment Scriptures (1927) beschreibt Tillmann (1980) das phonetische Ereignis als eine Interaktion dreier Prosodien (der A-, B- und C-Prosodie) im Sinne von Firth (1948): Modulieren wir eine beliebige Klangeigenschaft sehr langsam, so können wir dies wahrnehmungsmäßig als eine Änderung einer bestimmten Eigenschaft verfolgen (A-Prosodie), wie dies bei der Satzmelodie der Fall ist; erhöhen wir die Modulationsfrequenz, so wird diese Wahrnehmung durch die einer rhythmischen Strukturierung (B-Prosodie) abgelöst, wie sie in Bezug auf phonetische Ereignisse in der wohlartikulierten silbischen Abfolge gegeben ist; bis bei weiterer Erhöhung der Änderungsgeschwindigkeit diese Änderung nicht mehr als solche, sondern als eine neue Klangeigenschaft wahrnehmungsmäßig imponiert (C-Prosodie), die bezüglich des phonetischen Ereignisses mit der individuellen einzelsprachlich geregelten Sprechbewegung gegeben ist. Der mit 'Silbenprosodie' umschriebene Gegenstand der vorliegende Arbeit ist das von Tillmann als B-Prosodie bezeichnete Phänomen der zeitlichen Strukturierung der Silbenabfolge im Zusammenhang mit den Phänomenen des wahrgenommenen Sprachrhythmus und Sprechtempos in der lautsprachlichen Äußerung. Im Zusammenhang mit dem Begriff der 'Prosodie' sei an dieser Stelle vorab darauf hingewiesen, daß Phänomene der Intonation – ohne daß damit deren bedeutende Rolle für die Wahrnehmung gerade des Sprechtempos geleugnet werden soll – im Rahmen der vorliegenden

experimentellen Studien ausdrücklich ausgeklammert wurden. Auch das Problem der Silbe als phonetischer Einheit selbst (vgl. Tillmann 1964) liegt außerhalb des Themenbereiches der vorliegenden Arbeit.

Die gesprochene Sprache als ein auf den unterschiedlichsten Ebenen zeitlich strukturiertes Phänomen stand lange nicht im Brennpunkt des wissenschaftlichen Interesses.

Während es in frühen experimentalphonetischen Untersuchungen noch um die Frage nach der komplexen zeitlichen Strukturierung in den einzelnen Sprachen ging (vgl. z.B. Gregoire (1899, 1911) für das Französische, Menzerath & de Oleza (1928) für das Spanische, Weitkus (1931) für das Deutsche), schränkte sich nachfolgend der Blickwinkel ein:

So ging es in den klassischen Modellen der modernen perzeptiven Phonetik, die implizit immer Modelle der Phonem–Dekodierung waren, um zeitliche Aspekte nur insofern, als die Frage darin bestand, wie mittels einer in ihrer Geschwindigkeit stark beschränkten Maschinerie (den Artikulationsorganen) eine solch hohe Rate der Informationsübertragung (im Sinne von Phonemen pro Zeiteinheit) zu erreichen ist.

Im weiteren interessierten zeitliche Phänomene, im eingeschränkten Sinn als Dauer einzelner Einheiten verstanden, einerseits insofern dieser Dauer die Funktion eines ‘cues’, d.h. eines hinreichenden akustischen Merkmals zur Bestimmung der wahrgenommenen phonetischen Kategorie zukam, andererseits als intrinsische Dauer einzelner Sprachlaute im Rahmen der künstlichen Sprachschallsynthese (vgl. Klatt 1976). Diese Situation hat sich im Laufe des letzten Jahrzehnts grundlegend geändert: Konnte Ilse Lehiste 1970 den Forschungsstand zu den sogenannten ‘suprasegmentalen’ Phänomenen in ihrer Gesamtheit noch in einer nicht zu umfangreichen Monographie zusammenfassen, so scheint dies heute selbst für einzelne Teilgebiete nicht mehr in dieser Form möglich. Diese Änderung der Situation war einerseits – für die Erforschung der Intonation wie der makroskopischen Zeitstruktur lautsprachlicher Äußerungen – bedingt durch das im Bereich der ‘Speech Technology’ wachsende Interesse an einer möglichst natürlichen Sprachschallsynthese, andererseits – speziell für die Frage nach den Mechanismen der zeitlichen Steuerung der Sprachproduktion – durch die wachsende Einsicht im Bereich der Sprachwahrnehmungsforschung, daß die dort zu beobachtenden Prozesse nicht ohne Rückgriff auf die Sprachproduktion zu erklären sind.

In ihrer Dissertation “Timing Control in Speech Production”, die einen Meilenstein für die erneute Hinwendung zur wissenschaftlichen Untersuchung der Sprachproduktion unter einem auch die suprasegmentalen Gegebenheiten mit einschließenden Blickwinkel darstellt, kennzeichnet Carol A. Fowler (1977) die unterschiedlichen Bereiche, die eine einheitliche Theorie des Zeitverhaltens bei der Sprachproduktion zu erklären habe, wobei sie eine Unterscheidung in eher mikroskopisches bzw. makroskopisches “Timing”–Verhalten vornimmt.

Ein erster, von ihr als in den Bereich eher mikroskopischer zeitlicher Steuerung fallender Effekt ist der der Koartikulation, den sie in diesem Zusammenhang nun wieder im ursprünglichen Sinn von echter ‘Koproduktion’ (wie bei Menzerath & de Lazerda (1933), die diesen Effekt erstmals be-

schrieben) versteht, und nicht im Sinne von einer lediglich assimilativen gegenseitigen Beeinflussung benachbarter Lautsegmente.

Im makroskopischeren Bereich der zeitlichen Steuerung nennt sie

- (1) die Effekte der intersegmentellen zeitlichen Abhängigkeiten innerhalb größerer Abschnitte lautsprachlicher Äußerungen (vgl. Kozhevnikov & Chistovich 1965),
- (2) die Phänomene der medialen Verkürzung wie der finalen Längung (Lehiste 1972, Lindblom & Rapp 1973),
- (3) die mit Sprechgeschwindigkeit und Betonung sich ändernden Zeitverhältnisse sowie
- (4) das Phänomen des Sprachrhythmus.

Im folgenden soll von Untersuchungen zu Effekten, die dem vierten genannten Punkt zuzuordnen sind, ausgegangen werden. Dabei mußte allerdings aufgrund der Vielzahl und Vielfalt an einzelnen Arbeiten eine stark begrenzende Auswahl getroffen werden. Der in diesen ersten Kapiteln gegebene Überblick erhebt somit keinerlei Anspruch auf Vollständigkeit, er soll vielmehr in aller Kürze die unterschiedlichen Herangehensweisen an das Problem des Rhythmus gesprochener Sprache im Spannungsfeld zwischen Linguistik einerseits und allgemeiner Rhythmustheorie andererseits kennzeichnen. Hierbei wurden die frühen Arbeiten zum Sprachrhythmus, die sich mit diesem Problem unter dem mehr literaturwissenschaftlichen Gesichtspunkt der Metrik befaßten (vgl. z.B. Sievers 1894, Meyer 1898, Scripture 1929) bewußt ausgeklammert, da der Rhythmus der gebundenen Sprache ja einen Sonderfall bzw. eine besondere Nutzung des auch in der alltäglichen Sprache vorhandenen Rhythmus darstellt, wie dies von Allen (1972) in dem folgenden Zitat verdeutlicht wird:

[W]e can paraphrase Orwell's pigs and say "Poetry is like speech, only more so". Poets use naturally occurring phenomena to their artistic advantage; one of those phenomena is speech rhythm, which they creatively select to fit some metrical scheme.
(Allen 1972: 73)

Diese Überlegungen führten schließlich zu einer recht speziellen Fragestellung, die im zentralen experimentellen Teil dieser Arbeit behandelt wird, nämlich der Frage nach den Grundlagen des subjektiv gleichmäßigen Rhythmus bei Sequenzen betonter einsilbiger Äußerungen. Auf der Grundlage der in diesen Experimenten gewonnenen Erkenntnisse sollen abschließend nicht nur die mit diesen Effekten in engem Zusammenhang stehenden Fragen nach der Dauer- und Tempowahrnehmung in der lautsprachlichen Kommunikation erörtert werden, sondern auch derzeitige Modelle zum Zusammenhang zwischen Sprachproduktion und -perzeption – insbesondere in Bezug auf deren Behandlung der Frage nach der zeitlichen Strukturierung lautsprachlicher Äußerungen – einer kritischen Würdigung unterzogen werden.

Die vorliegende Arbeit gliedert sich in drei größere Abschnitte, wobei der erste den unterschiedlichen Aspekten makroskopischer zeitlicher Strukturierung lautsprachlicher Äußerungen gewidmet

ist. Hierzu wird im ersten Kapitel von der Unterscheidung zwischen betonungs- und silbenzählenden Sprachen und der damit verbundenen Isochronie-Hypothese ausgegangen. Die negativen Befunde bezüglich dieser letztgenannten Hypothese in ihrer strikten Form führten einige Forscher zu der Annahme, daß die Wahrnehmung unterschiedlicher sprachlicher Rhythmen weniger mit physikalischen Dauerverhältnissen zu tun hat, als vielmehr mit allgemeineren Gesetzen der Rhythmuswahrnehmung. In einem zweiten Abschnitt des ersten Kapitels wird daher auch auf die allgemeinere Rhythmustheorie im Bereich der Wahrnehmung wie der Produktion eingegangen. Andere Forscher nahmen die negativen Evidenzen für die strenge Isochroniehypothese zum Anlaß, eine abgeschwächte Version derselben zu vertreten, nämlich, daß es sich hierbei um Tendenzen auf eine Isochronie hin handele, wobei unter diesem Gesichtspunkt gerade in letzter Zeit für die unterschiedlichsten Sprachen eine Fülle von experimentellen Evidenzen erarbeitet wurde. Einzelne Untersuchungen in diesem Zusammenhang werden in einem weiteren Abschnitt dieses Kapitels besprochen. Auf der anderen Seite warfen diese ursprünglichen, generell an Dauermessungen orientierten Untersuchungen aber auch die Frage nach der genauen zeitlichen Lokalisation rhythmischer Ereignisse auf. Nachdem in einem zweiten Kapitel kurz frühe Arbeiten zur damit verbundenen Frage nach dem silbischen Takt beschrieben werden, ist das dritte Kapitel dem für diese Frage zentralen sogenannten P-center-Effekt (Morton, Marcus & Frankish 1976) gewidmet. Während sich das vierte Kapitel mit den unabhängig von der phonetischen Forschung entstandenen Modellen der psychoakustischen 'Ereigniszeitpunkt'-Bestimmung beschäftigt, wird im fünften Kapitel abschließend auf die artikulatorische Interpretation des P-center-Effekts durch Fowler (1977ff) im Rahmen einer umfassenderen Theorie des intrinsischen Timings eingegangen.

Der zweite, experimentelle Teil dieser Abhandlung beinhaltet die zusammenfassende systematische Darstellung der Ergebnisse aller im Rahmen des DFG-Projekts Ti 69/22 "Signalphonetische Korrelate wahrgenommener Isochronie (P-center)" durchgeführten Experimente, die in den letzten Jahren von mir betreut wurden. Einzelne Detailergebnisse dieser Untersuchungen wurden in der Form von kurzen Kongreß-Papieren dargestellt (Pompino-Marschall (1987), Pompino-Marschall & Tillmann (1987d), Pompino-Marschall, Tillmann & Kühnert (1987c)). Dieser Teil der Arbeit besteht aus einem ersten Abschnitt zur Produktion rhythmisch gleichmäßiger Sequenzen sowie zur Überprüfung einiger Ergebnisse anhand von Wahrnehmungsexperimenten, während der zweite Abschnitt die Untersuchung der systematischen Beeinflußbarkeit der wahrgenommenen P-center-Position bei phonetisch strikt kontrolliertem Material zum Gegenstand hat. Im zehnten Kapitel, welches das theoretische Zentrum dieser Arbeit darstellt, wird ein aus den experimentellen Befunden abgeleitetes, am Verlauf der spezifischen Lautheiten orientiertes Modell der P-center-Bestimmung vorgestellt und im Zusammenhang mit den Ergebnissen anderer experimenteller Untersuchungen diskutiert. Ein letztes Kapitel dieses experimentell orientierten Teils der Arbeit ist dem Zusammenhang des P-center-Effekts mit einzelnen Phänomenen der Wahrnehmung der Sprechgeschwindigkeit gewidmet.

Der dritte Abschnitt dient schließlich der Diskussion der zu beobachtenden Effekte im Zusammenhang mit einer umfassenderen Theorie der zeitlichen Steuerung lautsprachlicher Äußerungen in Produktion und Perzeption.

1. DIE MAKROSKOPISCHE ZEITLICHE STRUKTURIERUNG GESPROCHENER SPRACHE

1.1. Die Isochronie-Hypothese

Daß sich die Sprachen der Welt unter anderem auch nach rhythmischen Gesichtspunkten unterscheiden, ist eine oft vorgenommene Feststellung (s. z.B. Pike 1945, Abercrombie 1967). Auf wen die Unterscheidung zwischen betonungszählenden ('stress timed') und silbenzählenden ('syllable timed') – sowie morenzählenden – Sprachen zurückgeht, ist nicht mehr festzustellen. Hinter dieser Klassifikation, die z.B. zur Unterscheidung der germanischen Sprachen (betonungszählend) von den romanischen Sprachen (silbenzählend) herangezogen wurde, stand die Vorstellung, daß in diesen Sprachen unterschiedliche Einheiten – nämlich Takte bzw. betonte Silben einerseits und Einzelsilben andererseits – in gleichmäßigen Zeitabständen aufeinander folgen. Diese Vorstellung stellt die Isochronie-Hypothese in ihrer strengen Form dar.

Die erste experimentelle Untersuchung zu dieser Hypothese dürfte für das Englische als betonungszählende Sprache die Arbeit von Classe (1939) darstellen. Da er in seinen Experimenten aber keine absolut gleichen zeitlichen Abstände zwischen den zu einzelnen englischen Versen mitgeklopften Takten finden konnte, verwarf er die strenge Hypothese des Betonungszählens, die eben diese zeitlich gleichen Abstände der Taktschläge impliziert. Schon er geht davon aus, daß es sich bei diesen Spracheigenschaften um Tendenzen bei der Sprachproduktion plus Eigenschaften der Perception zeitlicher Verhältnisse handeln müsse.

Auch neuere Untersuchungen mit englischem Sprachmaterial zeigten immer deutliche Abweichungen von der geforderten akustischen Isochronie: Shen & Peterson (1962) fanden so bei gelesenen Material Abstände zwischen primär betonten Silben in dem großen Bereich von einer halben bis zu dreieinhalb Sekunden. Die genannten Autoren gingen bei der Bestimmung der Betonung allerdings von dem System von Trager & Smith (1951) aus, welches nur eine primäre Betonung pro Satz vorsieht. Dieser Ausgangspunkt führte daher zu Meßintervallen, die auch Junktoren mit einschließen, was von vornherein bezweifeln läßt, daß sie einheitlich in ihrer Dauer kontrolliert werden. Bolinger (1965) fand starke Abhängigkeiten der Inter-stress-Intervalle je nach phonologischer Komposition der Takte und deren Position innerhalb der Äußerung. Auch neuere Analysen längerer gelesener Texte (Uldall 1971, 1972, Lea 1974) zeigten mehr oder weniger starke Abweichungen von der Isochroniehypothese für das Englische. Selbst bei Untersuchungen an Limericks (O'Connor

1965) fanden sich noch Differenzen zwischen den einzelnen gemessenen Intervallen in der Größe von 88 msec. Wiewohl die bisher beschriebenen Messungen an sehr heterogenem Material vorgenommen wurden, zeigten sich Abweichungen von der Isochronie-Hypothese auch in Untersuchungen mit homogenerem Testmaterial: So untersuchte O'Connor (1968) systematisch den Einfluß der Silbenstruktur auf die Dauer einsilbiger Takte des Englischen. Das Fehlen einer Kompensation der durch Erhöhung der Segmentanzahl bedingten Längung ließ auch ihn die Isochronie im Bereich der Produktion bezweifeln. Systematisch die Silbenstruktur und die Position der Takte variierte auch Lehiste (1973a, 1975a) in englischen Sätzen, die eine Länge von vier Takten aufwiesen. Sie interpretiert ihre Ergebnisse als nur teilweise für eine Isochronietendenz sprechend.

Wie die Situation für das Englische als betonungszählende Sprache ist auch die Evidenz für eine strikte Silbenisochronie in den sogenannten silbenzählenden Sprachen eher negativ (vgl. z.B. Pointon (1980) für das Spanische, Wenk & Wioland (1982) für das Französische). Diese Ergebnisse der an Dauermessungen orientierten Untersuchungen scheinen so für viele Autoren dafür zu sprechen, daß diese Messungen wenig zu tun haben mit den intuitiven linguistischen Rhythmuskategorien.

Such judgements are certainly correct, if the isochrony sought is absolute, i.e. one showing unchanging intervals and one whose measurement points are clearly present in the acoustic signal. The former condition assumes that there are no other processes, such as prepausal lengthening, that have an effect. This seems highly unlikely. The latter assumption may also be risky, given the complex relation between acoustic signal and percept. Thus, it is possible that investigators measuring foot durations are not using the same points of reference that hearer use when judging these same durations. (Hoequist 1983a: 20)

Das Fehlen einer empirischen Unterstützung für die Isochroniehypothese in ihrer strikten Form führte die verschiedenen am Sprachrhythmus interessierten Forscher zu recht unterschiedlichen – sich allerdings keineswegs immer gegenseitig ausschließenden – neuen Herangehensweisen an dieses Problem:

- Für eine erste Richtung ist es kennzeichnend, daß sie die Relevanz der traditionellen, am akustischen Signal orientierten Segmentationskriterien für die Rhythmusbestimmung anzweifelt, und deren neue experimentelle Fragestellung daraufhin lautete, wo im zeitlichen Ablauf lautsprachlicher Kommunikation rhythmische Ereignisse zu lokalisieren sind. Untersuchungen dieser Art werden unten ab Kapitel 2 etwas ausführlicher dargestellt, da diese in direkter Verbindung mit den in den eigenen Experimenten behandelten Fragen stehen.
- Auf der anderen Seite entstanden aus dieser Situation heraus in jüngster Zeit Untersuchungen, die aufgrund des Vergleichs von Dauer-Effekten in unterschiedlichen Sprachen zu entscheiden suchten, inwieweit nun tatsächlich Tendenzen der zeitlichen Regulierung auftreten, die eine Typologie des Sprachrhythmus auf der Grundlage der zeitlichen Ausprägung von bestimmten linguistischen Einheiten rechtfertigen (auf einige ausgewählte Bei-

spiele für dieses experimentelle Herangehen wird im übernächsten Abschnitt dieses Kapitels noch einzugehen sein).

- Neben diesen Ansätzen, die als von einer abgeschwächten Isochronie-Hypothese ausgehend gekennzeichnet werden können, wird von einer anderen Gruppe von Forschern (vgl. Lehiste 1977) angenommen, daß die Vorstellung von einer Isochronie bestimmter Einheiten eher auf einer Täuschung unserer Wahrnehmung beruht, nämlich der Tolerierung gewisser Dauerabweichungen wie der wahrnehmungsmäßigen Gruppierung von Einzelereignissen, so daß die – nur scheinbar speziellen – Erscheinungen des Sprachrhythmus allein auf der Grundlage allgemeinerer Erscheinungen der Rhythmusproduktion und –wahrnehmung schon erklärbar sind.

1.2. Das Problem des Sprachrhythmus und allgemeine Theorien der Rhythmuswahrnehmung und –produktion

Da die allgemeine Rhythmustheorie – sowie die damit verbundene Theorie der Zeitwahrnehmung allgemein – ein eigenes großes, an Einzeluntersuchungen reiches Gebiet darstellt (vgl. z. Überblick z.B. Woodrow 1951, Fraisse 1963, 1978, 1982), soll im folgenden nicht auf Einzelheiten, sondern nur auf grundsätzliche Ergebnisse eingegangen werden, nämlich insofern diese für die Sprachrhythmusforschung prinzipiell relevant erscheinen.

Eine erste allgemeinere Frage, die sich auf dem Hintergrund der beobachteten Streuung der in den oben referierten Untersuchungen zur Isochronie-Hypothese gemessenen Dauern aufdrängte, war natürlich, inwieweit wir in der Lage sind, Dauerunterschiede überhaupt wahrzunehmen. Ergebnisse früherer Untersuchungen, die eine Unterschiedsschwelle in der Größenordnung von 10% der Ausgangsdauern (bei einem Dauerbereich bis zu zwei Sekunden) bei Paarvergleichen feststellten (vgl. Woodrow 1951), stellten sich als auf die Sprachrhythmusperzeption so nicht übertragbar heraus. In einer Reihe von Untersuchungen fand Lehiste (1973a, 1975a, b) in Experimenten, bei denen das kürzeste und längste Element einer Vierersequenz von den Versuchspersonen zu nennen war, daß dies einerseits bei einem Material von ein- bzw. zweisilbigen natürlichgesprochenen Rhythmusgruppen wesentlich schwieriger war (d.h., daß hier weniger korrekte Urteile abgegeben wurden) als bei nachsynthetisierten Geräuschsequenzen, die diesem Sprachmaterial in den zeitlichen Verhältnissen entsprachen. Auf der anderen Seite zeigte sich, daß für diese im Gegensatz zu Einzelvergleichen komplexere Aufgabe aber auch bei nichtsprachlichem Material höhere Unterschiedsschwellen anzusetzen sind. Für Geräuschblocksequenzen einer Ausgangsdauer von 300, 400 bzw. 500 msec fand sie so für das Erkennen des kürzesten wie des längsten Intervalls eine Unterschiedsschwelle von 30 bis hin zu 100 msec. Dabei spielt auch die Position des Zielitems innerhalb der Sequenz eine große Rolle,

wobei diese Effekte mit den allgemeineren der finalen Längung (Lehiste 1970) wie des sogenannten 'time order errors' (vgl. Woodrow 1951) zu tun haben dürften (vgl. a. Benguerel & D'Arcy 1986). Zusammengenommen mit dem Ergebnis, daß natürlichsprachliche Sequenzen noch schlechter zu beurteilen sind, kommt Lehiste (1977) zu dem Ergebnis, daß ein Großteil der in den Isochronieuntersuchungen gemessenen Dauerunterschiede für den Hörer gar nicht wahrnehmbar sind. Zusätzlich die Wahrnehmung isochroner Einheiten unterstützend dürfte auch das Phänomen des sogenannten Indifferenzintervalls (Woodrow 1951, Fraisse 1963) sein. Mit diesem letztgenannten Phänomen ist der Effekt der dauerbezogenen Überschätzung kürzerer wie der der Unterschätzung längerer Intervalle in Bezug zu diesem im Bereich von 600 bis 800 msec gelegenen Intervall verbunden.

Ein weiterer, in diesem Zusammenhang wichtiger Punkt betrifft die Frage nach den allgemeineren Grundlagen der wahrnehmungsmäßigen rhythmischen Gruppierung. War die phonetisch-linguistische Ausgangsfrage – durch die Isochroniehypothese bedingt – rein auf den Blickwinkel der Dauer einzelner Einheiten begrenzt, so ergeben sich aus der allgemeineren Betrachtung unter dem Gesichtspunkt der rhythmischen Gruppierungen auch linguistisch andere Fragen, nämlich solcherart, daß gefragt wird, inwiefern u.U. verschiedene Sprachen von diesen Prinzipien in ihrer linguistischen (phonologischen, syntaktischen) Struktur unterschiedlichen Gebrauch machen (Allen 1975, Lehiste 1977).

Hierbei erscheinen zwei Fragenkomplexe der allgemeinen Rhythmustheorie relevant: (1) die Frage nach akustischen Parametern, die die perzeptuelle Gruppierung bestimmen und (2) die Frage nach den allgemeinen Mechanismen der Steuerung rhythmischen Verhaltens.

Bezüglich der Rhythmuswahrnehmung zeigte sich in psychoakustischen Untersuchungen, daß Tonsequenzen mit einem gleichmäßigen zeitlichen Abstand zwischen 0.1 und 3 sec zwischen den einzelnen Tönen, die in regelmäßigen Abständen jeweils einen lautereren Ton aufwiesen (z.B. jeder zweite oder dritte Ton), als in trochäische bzw. daktylische Takte gegliedert wahrgenommen wurden, d.h. in fallende Takte, die mit diesem lautereren Ton beginnen, während Sequenzen, die ebenso regelmäßig längere Töne beinhalteten, wahrnehmungsmäßig in jambische und anapestische Takte, d.h. in steigende Versfüße gegliedert wurden (Woodrow 1909, 1911, 1951, Fraisse 1963). Allein diese Tatsache – im Zusammenhang mit der unterschiedlichen Ausprägung und Flexibilität des Akzents in den verschiedenen Sprachen – muß zu einer nach rhythmischen Gesichtspunkten differenzierenden Wahrnehmung unterschiedlicher Sprachen führen. Festzuhalten bleibt an dieser Stelle, daß diese Prinzipien der Rhythmuswahrnehmung natürlich unterschiedliches Gewicht erhalten, je nachdem, ob wir Äußerungen gebundener Sprache oder aber alltägliche lautsprachliche Äußerungen als Untersuchungsgegenstand vor uns haben. Grundsätzlich aber gelten sie – wie schon im Einführungskapitel ausgeführt – für alle Formen gesprochener Sprache.

Ein weiterer Effekt der Rhythmuswahrnehmung – der sicherlich auch im Zusammenhang mit der Steuerung rhythmischen Verhaltens (s.u.) gesehen werden muß – ist, daß wir auch objektiv gleichmäßigen Sequenzen (z.B. von Clicks) wahrnehmungsmäßig eine rhythmische Gliederung,

bzw. größeren rhythmischen Gruppen – ohne physikalische Grundlage – eine interne Strukturierung aufprägen. So zeigt sich sowohl im Fall von nichtsprachlichen Sequenzen bei der Wahrnehmung von ‘fallenden’ Vierergruppierungen die Tendenz zur Wahrnehmung eines Nebenakzents auf jedem zweiten Ton (Woodrow 1951) als auch eine Beschränkung auf einen einfachen Kontrast in einem 1:2-Verhältnis bei der Beurteilung linguistischer relevanter Dauern (Lehiste 1987).

In einer Reihe neuerer Untersuchungen konnten Povel und Essens (Povel 1981, 1984, Povel & Essens 1985, Essens 1986) parallel zur Wahrnehmung für die Nachproduktion vorgegebener rhythmischer Sequenzen die folgenden grundsätzlichen Charakteristika feststellen: Der wahrgenommene bzw. reproduzierte Rhythmus läßt sich als ein intern induziertes hierarchisches Modell beschreiben, das eine ganzzahlige Unterteilung eines übergeordneten Taktes in Einzelereignisse vornimmt. Gerade dieser zuletzt angesprochene Punkt der Beschränkung in den generellen Möglichkeiten der Strukturierung rhythmischen Verhaltens allgemein ist für ein Modell der Sprachproduktion unter dem Aspekt der zeitlichen Steuerung von großer Bedeutung, und wir werden hierauf am Ende dieser Abhandlung wiederum zurückkommen.

Im folgenden Abschnitt wollen wir jedoch wieder auf die als Ausgangspunkt genommene Hypothese von der Isochronie unterschiedlicher sprachlicher Einheiten für die zeitliche Steuerung in den verschiedenen Sprachen zurückkommen.

1.3. Isochronie, Kontrolle der Dauer und zeitliche Kompensation

Neben den später noch zu referierenden Ansätzen, die das grundsätzliche Verfehlen der Bestätigung der starken Isochronie-Hypothese teilweise auf Fehler in der Wahl der Meßpunkte zurückführen (s.u. unter Kapitel 2 und 5), sollen in diesem Abschnitt nun Modelle vorgestellt werden, die von einer etwas abgeschwächten Isochronie-Hypothese ausgehen. Gemeinsam ist diesen Ansätzen, daß sie die typologisch scharf formulierten Verhältnisse nur als Tendenzen auf diese Isochronien hin auffassen, d.h., daß sie nach einzelnen Effekten suchen, die auf eine solche Tendenz zur Isochronie bestimmter Einheiten hindeuten, wobei sich die typologisch unterschiedlichen Sprachen bezüglich dieser Effekte unterscheiden sollten. Als ein typisches Beispiel hierfür seien an dieser Stelle die Untersuchungen von Hoequist (1983b, c) etwas ausführlicher dargestellt, zeigen sie doch in der Komplexität ihrer Ergebnisse auch schon die grundsätzlichen Schwierigkeiten, denen sich ein Modell der zeitlichen Strukturierung lautsprachlicher Äußerungen, welches die Vielzahl der zu beobachtenden Phänomene einer einheitlichen Erklärung zuführen möchte, gegenüberstellt.

In einer ersten Studie untersuchte Hoequist (1983b) die Beeinflußbarkeit der Silbendauern im Japanischen und Spanischen (als moren- bzw. silbenzählende Sprache) durch die Faktoren (1) der pho-

nologischen Länge (im Japanischen), (2) des Unterschiedes zwischen offenen und geschlossenen Silben, (3) des Akzents und (4) der Position innerhalb eines Wortes.

Das Material bestand jeweils aus Wörtern der beiden Sprachen, die von den sieben spanischen bzw. fünf japanischen Versuchspersonen – in einem Rahmensatz eingebettet – produziert wurden. An diesem Wortmaterial wurden die Dauern einzelner Silben ausgemessen. Für das spanische Material wurden die Messungen an CV- und CVC-Silben mit stimmlosen Plosiven in initialer Position vorgenommen. Bei den japanischen Wortäußerungen wurden ebenfalls CV-Silben mit initialen stimmlosen Plosiven sowie die Silben /fu/, /ji/, /ma/, /sa/, /saa/ und /san/ zur Dauermessung herangezogen.

Bezogen auf die Hypothese einer strikten Moren-Isochronie, wie sie von Han (1962a, b) für das Japanische konstatiert wurde, waren die Dauervergleiche zwischen den Silben /saa/ und /sa/ auf der einen und /san/ und /sa/ auf der anderen maßgebend: Einmorige – leichte – Silben des Japanischen sind generell offene Kurzvokalsilben, zweimorige – starke – Silben hingegen sind entweder offene Silben mit Langvokal oder geschlossene Kurzvokalsilben.

Im Gegensatz zu den allerdings hauptsächlich an Plosiv-Vokal-Silben vorgenommenen Messungen von Han (1962a, b), bei denen sich ein Dauerverhältnis von schweren zu leichten Silben in der Größenordnung von 2:1 zeigte, ergaben die Messungen von Hoequist (1983b) einerseits zwischen der offenen schweren Silbe /saa/ und der leichten Silbe /sa/ nur ein Dauerverhältnis von 1.71:1, und zeigte sich die geschlossene schwere Silbe /san/ als zwar wenig, aber signifikant länger als ihr offener Gegenpart /saa/.

Auf die etwas stärker mit der Isochronie-Hypothese übereinstimmenden Daten seiner zweiten Untersuchung (Hoequist 1983c) wird weiter unten noch eingegangen.

Der Dauerunterschied zwischen geschlossenen und offenen Silben prägte sich für die beiden verschiedenen Sprachen unterschiedlich aus: Mit einem Verhältnis von 1.8:1 war es im Japanischen stärker als mit 1.66:1 im Spanischen.

Während die Silbenlänge unter dem 'Pitch'-Akzent im Japanischen nur gering ist, zeigte sich dieser Betonungseffekt deutlich in einem Dauerverhältnis von 1.24:1 beim dynamischen Akzent im Spanischen.

Bezüglich der finalen Länge zeigte sich mit einem Dauerverhältnis von 1.3:1 für das Spanische und 1.24:1 für das Japanische kein Unterschied zwischen den beiden Sprachen.

Zusammenfassend zeigte sich so also zwar jeweils ein Einfluß aller in beiden Sprachen beobachtbarer Variablen, aber die Rangordnung bezüglich der Stärke dieser Einflüsse war unterschiedlich: Im Japanischen zeigte sich so die größte Länge beim Übergang zwischen ein- und zweimorigen Silben, gefolgt vom Einfluß der Position der Silbe innerhalb des Testwortes auf die Silbendauer, während im Spanischen die Position der Testsilbe den größten Einfluß auf die Silbendauer hat, knapp gefolgt von dem Effekt der Länge unter Betonung.

In seiner zweiten Untersuchung (Hoequist 1983c), bei der jeweils vier Sprecher der Muttersprachen Englisch, Japanisch und Spanisch als Versuchspersonen dienten, bezogen sich die Silbendauermessungen auf nach natürlichsprachlichen Äußerungen parallel produzierte Äußerungen reiteranter /ma/-Silben wie z.B. bei

engl. *demonstration* >> /mama'mama/

Neben den in der ersten Untersuchung schon getesteten Einflüssen auf die Silbendauer wurden hierbei auch die Effekte der Anzahl der vorausgehenden und nachfolgenden Silben sowie der Effekt der Nachbarschaft zu betonten Silben (Lieberman & Prince 1977) für das Englische untersucht.

In Bezug auf die Unterscheidung zwischen zwei- und einmorigen Silben (sofern die ersteren nicht – wie insbesondere bei Diphthongen im natürlichsprachlichen Material – als zwei reiterante /ma/-Silben realisiert wurden) zeigte sich für das Japanische hier eine Übereinstimmung mit den von Han (1962a, b) festgestellten Dauerverhältnissen von 2:1.

Unter Akzent zeigte sich für die spanisch- und auch für die englischsprechenden Versuchspersonen eine Verlängerung der Silbendauer im Verhältnis von 1.5:1, während im Japanischen wiederum nur ein sehr geringer Betonungseffekt festzustellen ist.

Alle drei untersuchten Sprachen zeigten den (auf Wortposition bezogenen) Effekt der finalen Längung, aber nur für die Sprecher des Englischen zeigte sich – nach Reduktion der Daten nach Maßgabe der schon festgestellten Effekte – ein kompensatorischer Verkürzungseffekt bezüglich Silben, die direkt neben primärbetonten Silben stehen, wie auch bei wortmedialen Silben.

Bei gleicher Anzahl von Folgesilben zeigte sich ein Effekt der Anzahl der vorausgegangenen Silben sowohl bei den englischsprachigen als auch bei den spanischsprachigen Versuchspersonen. Im Englischen zeigte sich darüber hinaus ein wesentlich stärkerer Verkürzungseffekt, der durch die Anzahl der Folgesilben beeinflusst war.

Das generelle Ergebnis der gesamten Untersuchung von Hoequist (1983c) kann in Form der folgenden Tabelle I-1 wiedergegeben werden:

Tabelle 1-I:
Silbendauerbeeinflussende Faktoren(nach Hoequist 1983c)

	Japanisch	Spanisch	Englisch
Phonologische Dauer	+	0	0
Längung unter Akzent	(+)	+	+
Kompensatorische Kürzung	0	0	+
Kürzung durch Folgesilben	0	0	+
Kürzung durch Vorsilben	0	+	+

0: kein Einfluß, (+): schwacher Einfluß, +: Einfluß vorhanden

Insgesamt zeigten die beiden hier referierten Untersuchungen Effekte, die für eine bezüglich einzelner Sprachen unterschiedlich abgeschwächte Isochronie-Hypothese sprechen: So scheint das Japanische mit dem Einfluß nur der phonologischen Länge und einem sehr schwachen Akzenteinfluß der Beschreibung einer morenzählenden Sprache am nächsten zu kommen, während das Englische als betonungszählende Sprache mit der Vielzahl der zu beobachtenden Effekte die stärksten kompensatorischen Maßnahmen aufweist, wobei diese allerdings auch über die Grenzen einzelner Betonungstakte hinausreichen. Das Spanische entspricht hier am wenigsten der Beschreibung als silbenzählende Sprache, wenn man darunter versteht, daß es die Silbe ist, die als Produktionseinheit in ihrer Dauer kontrolliert wird. Für Hoequist unterscheiden sich daher die unterschiedlichen Sprachen in erster Linie nicht so sehr durch die in ihrer Dauer kontrollierten Einheiten, sondern vielmehr durch die Art der zeitlichen Steuerung, so daß man das Japanische eher als **dauerregulierende** Sprache im Gegensatz zum Englischen als **dauerkompensierender** Sprache sehen könnte.

Während die Untersuchung von Hoequist (1983c) gerade für die silbenzählende Sprache Spanisch die wenigsten Hinweise auf die Silbe als primäre Steuerungseinheit ergeben hatte (in ähnliche Richtung weisen auch die Ergebnisse von Vayra, Avesani & Fowler 1987 für das Italienische), zeigten sich in jüngsten Arbeiten zum Italienischen durchaus Effekte, wie sie mit einer abgeschwächten Version der Isochronie-Hypothese verträglich sind.

In diesem Zusammenhang sind z.B. die Untersuchungen von Bertinetto & Fowler (1987) zur Akzeptabilität von Vokalverkürzungen im Italienischen und Englischen zu erwähnen. In dieser Untersuchung wurden unbetonte Vokale in parallel konstruierten italienischen und englischen Wörtern durch wiederholte Elimination einzelner Vokalperioden verkürzt. Als Material dienten dabei u.a. die folgenden Wortpaare (modifizierte Vokale fett):

<i>superfluous</i>	<i>superfluo</i>
<i>political</i>	<i>politico</i>
<i>esoteric</i>	<i>esoterico</i>
<i>canonical</i>	<i>canonico</i>

Bei der Beurteilung der modifizierten Stimuli durch die Versuchspersonen zeigte sich, daß, wie auf der Grundlage des silbenzählenden Charakters des Italienischen zu erwarten, die italienischen Versuchspersonen tatsächlich die Dauerreduktion des unbetonten Vokals wesentlich weniger tolerierten als die englischsprechenden.

Interessant in diesem Zusammenhang erscheinen auch Daten für das Französische und Italienische über zeitlich unterschiedliche Ausprägungen der Koartikulation je nach Silbenstruktur (vgl. Giannini 1987): Dort zeigte sich im Französischen – parallel zu italienischen Daten – in VCCV-Sequenzen eine den gerundeten finalen Vokal antizipierende koartikulatorische Lippenrundung schon gegen Ende des ersten Vokals, wenn eine V-CCV-Syllabierung vorlag, aber erst nach dem ersten intervokalischen Konsonanten bei einer VC-CV-Silbenstruktur.

Einen gänzlich anderen Ausgangspunkt in dem uns hier interessierenden Zusammenhang verfolgte Klaus Kohler in einem Projekt zur Produktion und Perzeption der Sprechgeschwindigkeit und des Rhythmus im Deutschen (vgl. Kohler 1986). Sein Ausgangspunkt dabei war die Frage, inwieweit es Sprechern möglich ist, eine Dauerkompression, wie sie für eine Isochronie der betonten Silben notwendig wäre, überhaupt vorzunehmen. Die Ausgangsexperimente bestanden darin, daß entweder parallel zu einem externen Zeitsignal oder nach einem vorgegebenen auditiven Muster Logatomaterial produziert werden sollte.

In einem ersten Experiment wurde als externer Zeitgeber ein Oszillographenstrahl mit einer Ablenkzeit von vier bzw. zweieinhalb Sekunden gewählt, und die Aufgabe der Versuchsperson bestand darin, in diesem Zeitrahmen fünf ein-, zwei- bzw. dreisilbige Takte zu äußern, die aus Wiederholungen einer der folgenden Silben bestanden: *pa*, *ba*, *bla*, *pasch*, *schapp* bzw. *platsch*. Das Ergebnis erwies sich als eindeutig. Die komplexeste Silbe *platsch* konnte nicht soweit komprimiert werden, daß sie als dreisilbiger Takt – unabhängig von der vorgegebenen Geschwindigkeit – hätte geäußert werden können. Ebenso ist sie bei der schnellen Geschwindigkeit nicht als zweisilbiger Takt produzierbar. Aber auch *bla*, *pasch* und *schapp* waren bei der schnellen Geschwindigkeit nicht als dreisilbige Takte produzierbar, *bla* auch nicht in der langsameren Geschwindigkeit. Kompression war also nur bis zu einem gewissen Grad zu beobachten. Die lineare Regression schließlich zeigte einen klaren Zusammenhang zwischen der Dauer der Einzelsilbe und des Taktes unabhängig von der Silbenstruktur und der Sprechgeschwindigkeit. Die Steigung der Regressiongeraden betrug .55 für den zweisilbigen bzw. .37 für den dreisilbigen Fuß.

In einem weiteren Experiment wurde die Kompression an den Wortendungen *-e* und *-ige* der Logatome

Pinn, Pinne, Pinnige
Pahn, Pahne, Pahnige
Schripps, Schrippse, Schrippsige
Schraabs, Schraabse, Schraabsige

gemessen, wobei die Logatome in drei verschiedenen Sprechgeschwindigkeiten in dem Satzrahmen

Nie hat Karl die blöde ... durchgestellt

geäußert wurden.

Die Ergebnisse zeigten eine klare Dauerkompensation bei zweisilbigen Takten, d.h. diese waren gleich lang wie die einsilbigen, dreisilbige Takte allerdings waren länger. Diese Längung jedoch war – bezogen auf die anderen variierten Parameter Silbenkomplexität und Sprechgeschwindigkeit – linear. Ebenso zeigte sich eine als linear zu beschreibende Abhängigkeit von der Silbenkomplexität, in der Rangfolge *Pinn* < *Pahn* < *Schraabs* und *Pinn* < *Schripps* < *Schraabs*, wobei die Dauern der Endungen *-e* und *-ige* im Einklang mit einer Isochronie-Hypothese parallel zu den ansteigenden Dauern der ‘Stämme’ eine linear abnehmende Dauer aufwiesen.

In einem weiteren Experiment wurden die zwei- und dreisilbigen Logatome in dem Rahmensatz ‘*Karl wird die ... treten*’ unter vier verschiedenen Intonationsmustern realisiert, wobei die Intonation I der normalen Aussage entsprach, während Intonation II, III, IV unterschiedlichem Kontrastakzent auf dem Verb, dem Testwort bzw. auf *Karl* zuzuordnen waren. Um auch den Einfluß der Wortgrenze abschätzen zu können, wurden dieselben Logatom-Äußerungen auch in dem Satzrahmen ‘*Karl hat die ... getreten*’ (ebenfalls in den vier angegebenen Intonationsmustern) produziert.

Die gemessene Dauer der Wortendung unter all den verschiedenen experimentellen Bedingungen bezogen auf die Dauer der Länge des Taktes zeigte die folgenden linearen Tendenzen. Wie die Dauer des Taktes mit wachsender Silbenkomplexität von *Pinn* über *Pahn* bis *Schraabs* zunimmt, nimmt die der Endung komplementär dazu ab. Die unterschiedlichen Intonationsmuster andererseits haben einen gleichartigen, ebenfalls linearen Einfluß auf die gemessenen Dauern. Die gemessenen Dauern der Takte wie der Endungen sind am längsten bei der neutralen Äußerung, während sie bei Kontrastakzent auf *Karl* den niedrigsten Wert aufwiesen. Die Werte unter den beiden anderen Bedingungen lagen zwischen diesen Extremen.

Auf der Grundlage dieser Ergebnisse und weiterer hier nicht im Einzelnen zu beschreibenden Experimente gelang es Kohler (1983a, 1986), ein Modell der zeitlichen Steuerung für das Deutsche bis hin zu Einzelsegmenten – ausgehend von der Dauer des Taktes – zu entwickeln. Hierbei sind alle

Modellvariablen additive Zeitkonstanten für z.B. größere Silbenkomplexität oder Sprechgeschwindigkeitsänderungen, bzw. Parameter der linearen Abhängigkeiten z.B. von Silben- oder Taktdauer.

Insgesamt zeigen so auch diese Ergebnisse eine Übereinstimmung mit einer abgeschwächten Isochronie-Hypothese.

Nachdem wir uns im letzten Abschnitt mehr mit Effekten der Sprachproduktion auseinandergesetzt haben, wie sie in Verbindung mit einer abgeschwächten Form der Isochronie-Hypothese verträglich erscheinen, wollen wir uns im folgenden der im Zusammenhang mit der Zielsetzung der eigenen Experimente wichtigen Fragestellung zuwenden, nämlich: Welche zur Wahrnehmung punktueller zeitlicher Ereignisse führenden Eigenschaften am zeitlichen Ablauf der lautsprachlichen Äußerungen sind es, die die Grundlagen für die zeitliche Steuerung bilden? Es wird also somit im folgenden weniger darum gehen, an einzelnen, wie auch immer gearteten Einheiten Dauerphänomene zu beobachten, sondern vielmehr darum, festzustellen, welche Punkte im Ablauf lautsprachlicher Äußerungen es sind, die unserer Wahrnehmung zeitlicher Phänomene der gesprochenen Sprache zugrundeliegen.

2. FRÜHE UNTERSUCHUNGEN ZUR LOKALISATION DES TAKTES IN LAUT- SPRACHLICHEN ÄUSSERUNGEN

Ausgangspunkt der Experimente von Allen (1972) waren die oben unter 1.1 referierten negativen Resultate der experimentellen Untersuchungen zur Überprüfung der Hypothese von der Isochronie der Betonungen im Englischen. Seine Überlegung war hierbei, daß die am akustischen Signal vorgenommenen Segmentationen unter Umständen nicht die für den Rhythmus relevanten Punkte im Signal trafen. Für ihn stellten sich daher die folgenden Fragen:

- (1) What are the "events" in speech that are the beats of the rhythmic structure we hear?
 (2) Where exactly do these events occur when we speak?
 (Allen 1972: 72)

Zur Beantwortung dieser Fragen konzipierte er eine Reihe von Experimenten, in denen die Versuchspersonen (1) zu bestimmten Silben einer gehörten Äußerung auf einer Taste mitklopfen sollten, (2) ein zeitlich verschiebbares Click-Signal auf die Taktposition von Äußerungen plazieren sollten und (3) in dem sie die Positionierung eines Clicks in Bezug zur Position des Taktes beurteilen sollten. Für das Mittast-Experiment wählte er aus einer freien Unterhaltung mit seinen drei Versuchspersonen jeweils drei Äußerungen, die er von fünf Fachkollegen je zweimal transkribieren ließ. Ein paar Beispiele sind mit über den jeweiligen Silben angegebenen Betonungswerten (von 0: nie als betont markiert, bis 10: in allen Transkriptionen als betont markiert) in der folgenden Tabelle (nach Allen 1972) wiedergegeben:

Tabelle 2-I:
Beispiele der Testäußerungen von Allen (1972)

Äußerung	
1	10 0 6 0 0 10 0 1 6 0 1 10 Use the weight of the line to get more and more out.
2	10 1/2 0 0 9 0 0 3 0 10 0 Spinners are particu(lar)ly good in currents.
3	5 5 1/2 9 5 0 10 0 7 0 9 See when he reared back and fired it by a guy?

Die Versuchspersonen hatten den Takt der jeweils auf einem Textblatt markierten Silbe jeweils 50 bzw. bei kürzeren Äußerungen 100 Mal mitzutasten. Dies ergab pro Einzeläußerung zwischen 500 und 600 Tastungen pro Versuchsperson. Gemessen wurde der Zeitpunkt des Tastendrucks relativ zum Vokalbeginn der betreffenden Silbe als auch die Dauer des prävokalischen Konsonanten. Bezüglich der Sicherheit der Versuchspersonen zeigte sich klar eine Abnahme der Varianz der Taktposition bei stärkerer Betonung der Silbe. Die genaue zeitliche Position der Tastung ist stark versuchspersonenabhängig, aber bei einem Kalibrationsexperiment mit einer Folge von vier Click-Signalen im gleichbleibenden Abstand von 800 msec zeigte sich eine parallele versuchspersonenabhängige Vorverlagerung der Tastung wie gegenüber dem Vokalbeginn der mitgetasteten Silbe. Zwei der Versuchspersonen von Allen zeigten eine signifikante positive Korrelation ($r = .39$ bzw. $.43$) zwischen der zeitlichen Vorverlagerung des Tastendrucks gegenüber dem Vokalbeginn und der akustischen Dauer des silbeninitialen Konsonanten: Je länger der initiale Konsonant, desto früher drückten die Versuchspersonen die Taste.

Das zweite Experiment in Allen (1972) wurde durchgeführt, um eine genauere Positionsbestimmung für den Taktschlag zu erlangen. Hierzu sollten zwei Versuchspersonen bei einem Teil des Sprachmaterials des ersten Experiments ein zeitlich verschiebbares Click-Signal so auf dem Sprachsignal plazieren, daß es mit dem Takt einer vorbestimmten Silbe übereinstimmte. Interindividuelle Unterschiede je nachdem, auf welchem Ohr der Click und auf welchem das Sprachsignal dargeboten wurde, waren nicht eindeutig zu interpretieren. Für die Position des Taktes zeigte sich das Vorgehen dieses Experiments als besser geeignet als das Mittasten: Die Antworten der Versuchspersonen wiesen wesentlich weniger Variabilität auf und waren auch interindividuell nicht so verschieden. Der Einfluß der Dauer des initialen Konsonanten auf die Vorverlegung des Clicks gegenüber dem akustischen Vokalbeginn ist parallel zu dem im ersten Experiment zu beobachteten. Mit dem letzten Experiment, bei dem die Übereinstimmung eines auf das Sprachsignal gesetzten Clicksignals mit dem Takt einer bestimmten Silbe bestimmt werden mußte, sollten eigentlich interindividuelle Unterschiede untersucht werden. Die überwiegende Mehrheit der Versuchspersonen zeigte sich allerdings außerstande, diese Aufgabe in konsistenter Weise zu lösen, was allerdings an der grundsätzlichen Schwierigkeit der auditiven Clicklokalisierung liegen dürfte (vgl. Broadbent & Ladefoged 1959). Auch für die drei erfolgreichen Versuchspersonen zeigte sich neben der generellen Übereinstimmung der Befunde mit denen der vorausgegangenen Experimente, daß bei der Positionierung des Clicksignals in einem Bereich von 200 msec Ausdehnung der Click als mit dem Silbenschlag übereinstimmend wahrgenommen wurde.

In Bezug auf die Frage nach der Produktion rhythmischer Äußerungen ist die Arbeit von Rapp (1971) zu nennen. Sie ist eine der frühesten systematischen Untersuchungen zu den phonetischen Verhältnissen bei zu Metronom gesprochenen einfachen Äußerungen. Sie ließ ihre drei schwedischen Versuchspersonen Äußerungen der Form /a'Ca:d/, wobei /C/ für die Konsonanten bzw. Kon-