

Lautmuster deutscher Spontansprache
Symbolphonetische Strukturen und signalphonetische Zeitvorgänge
Abschlußbericht im Forschungsprojekt Ko 331/22 der Deutschen Forschungsgemeinschaft

K. J. Kohler

Antragsteller

Prof. Dr. Klaus Kohler, Institut für Phonetik und digitale Sprachverarbeitung, CAU Kiel

Bearbeiter

Prof. Dr. Klaus J. Kohler,

Dr. Jonathan E. J. Rodgers: 01.10.98 - 30.09.00 (BATIIa),

Thomas Wesener (MPhil, Cambridge): 01.04.00 - 30.09.02 (BATIIa/2),

Benno Peters, MA: 01.01.01 - 30.04.03 (BATIIa/2)

Hilfskräfte: 01.10.97 - 30.06.03

Berichts- und Förderungszeitraum

01. 10. 1997 - 31. 08. 2003

5 pdf-Dokumente mit Audiobeispielen

Kohler: Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions

Kohler, Peters, Wesener: Phonetic exponents of disfluency in German spontaneous speech

Peters: Die Datenbasis *The Kiel Corpus*

Peters, Kohler, Wesener: Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache

Peters, Kohler, Wesener: Melodische Satzakkentmuster in prosodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion

Kiel, 15. Oktober 2003

1 Ziele

Durch integrierte symbol- und signalphonetische Analyse eines umfangreichen Datenkorpus spontansprachlicher Dialoge soll die Phonetik des Standarddeutschen auf Äußerungsebene, d.h. oberhalb lexikalischer Zitiertformen, erfaßt und im Vergleich von Spontan- und Lesesprache (Sätze und Texte) beschrieben werden. Die zu untersuchenden Phänomene sind

- die artikulatorische Veränderung in der Wortfolge (Artikulationsabschwächung)
- die Prosodie (Satzbetonung, Intonation, Zeitstrukturierung)
- nicht-verbale Sprech- und Gesprächssteuerung (Pausen, Atmen, Häitationen etc.).

Das Projekt verfolgt vier Ziele:

- segmentelle und prosodische Beschreibung deutscher Spontansprache in der Gegenüberstellung großer Korpora aus Spontan- und Lesesprache (phonetische Deskription)
- Zuordnung der empirischen Korpusdaten zu Kategorien und Regeln der artikulatorischen Reduktion sowie des Kieler Intonationsmodells (KIM) (phonetische Evaluation)
- Modellierung der segmentellen und prosodischen Satz- und Äußerungsphonetik durch Modellkategorien und generalisierende Regeln (phonetische Modellierung)
- Erklärung der artikulatorischen und prosodischen Regeln durch Rückführung auf Voraussetzungen in der lautsprachlichen Kommunikation (phonetische Explikation).

Die Trennung phonetischer Phänomene in eine segmentelle und eine prosodische Ebene ist eine in langer Tradition gefestigte metasprachliche Heuristik, die der Korrektur in mehrerlei Hinsicht bedarf:

1. Segmentelle Satzphonetik kann nicht ohne Berücksichtigung der prosodischen Bedingungen untersucht werden: prosodische Phrasierung, Akzentuierung und der Grad prosodischer und morphosyntaktischer Kohäsion zwischen sequentiellen sprachlichen Einheiten bestimmen die Ausprägung segmenteller Reduktion.
2. Auch außerhalb der Prosodie in der üblichen Definition finden sich Erscheinungen, die nicht sinnvollerweise segmentell, sondern nur als lange (artikulatorisch, akustisch, perzeptuell definierbare) Komponenten erfaßt werden können, die sich über Segmente legen, also suprasegmentell oder prosodisch in einem weiteren Sinn sind, z.B. Glottalisierung, Nasalisierung, Palatalisierung, Velarisierung, Labialisierung.
3. Andererseits haben segmentelle Erscheinungen an bestimmten Stellen im Sprechfluß Signalwirkung für prosodische Kategorien, wie z.B. Glottalverschlüsse zur Markierung oder Verstärkung von prosodischen Phrasengrenzen oder Assimilations- und Reduktionserscheinungen zur Anzeige des Fehlens prosodischer Phrasengrenzen trotz syntaktischer Grenzen.

2 Verarbeitungsplattform

2.1 Datenbasis

Die phonetischen Fragestellungen erfordern umfangreiche orthographisch und phonetisch annotierte Datenbasen zur Spontan- und zur Lesesprache. Dafür konnten großenteils die Signalaufzeichnungen aus den Projekten VERBMOBIL und PHONDAT herangezogen werden. Sie mußten allerdings durch Datenerhebungen in natürlicheren spontansprachlichen Szenarien als dem Knopfdruckszenario der Terminabsprache ergänzt werden (*Lindenstraße-Korpus*).

Ebenfalls verfügbar war das Verarbeitungsprogramm *xassp* zur Etikettierung der Signalrohdaten. Mit seiner Hilfe wurden, weitgehend durch geschulte Hilfskräfte, die Dialog- und Lesekorpora in

rechnerverarbeitbarer Form nach einem vorgegebenen Symbolinventar gemäß festgelegten Anwendungskonventionen transkribiert.

Es lagen folgende Daten segmentell und prosodisch etikettiert vor (siehe auch pdf-Dokument B. Peters "Die Datenbasis *The Kiel Corpus*") :

- das spontansprachliche Korpus der CD-ROMs #2-3 aus dem Verbmobil-Szenario der Terminabsprache: 77 Dialoge aus 11 kompletten Dialogsitzungen mit 22 Sprechern (13 männlich, 9 weiblich), 2,5 Stunden, 25.000 Wörter; auf dieser spontansprachlichen Datenbasis sind die Ergebnisse im DFG-Projekt 331/22-2 gegründet (siehe pdf-Dokumente)
- das spontansprachliche Korpus der CD-ROM #4 sowie 5 Einzeldialoge von CD-ROM #2, ebenfalls aus dem Verbmobil-Szenario der Terminabsprache: 40 Dialoge - 35 aus 5 kompletten Dialogsitzungen mit 10 Sprechern (5 männlich, 5 weiblich) sowie 5 Einzeldialoge von 10 Sprechern (6 männlich, 4 weiblich), 1,5 Stunden, 15.000 Wörter
- 6 überlappende Dialoge aus dem Video-Task-Szenario (12 Sprecher, 4 männlich, 8 weiblich, 1,5 Stunden, 13.500 Wörter), die zur Untersuchung des 'turn taking' eingesetzt werden können

2.2 Prosodisches Modell KIM und prosodisches Symbolisierungssystem PROLAB

Durch die Arbeiten im DFG-Schwerpunkt "Formen und Funktionen der Intonation" wurden in den 80er Jahren theoretische Grundlagen für die Untersuchung der deutschen Prosodie geschaffen und darauf aufbauend empirische Beschreibungen geliefert (Altmann 1988). In Kiel wurde ein prosodisches Gesamtmodell für Intonation und Zeitstrukturierung (Kieler Intonationsmodell = KIM; Kohler 1991, 1997) konzipiert. Entscheidend für diese Modellentwicklung ist eine prosodische Phonologie als Schnittstelle zwischen der syntaktisch-semantic-pragmatischen Funktion und der signalphonetischen Manifestation.

KIM bot die erforderliche Grundlage für die Entwicklung einer prosodischen Notation auf phonologischer Basis, parallel zum Phonemalphabet für die segmentelle Lauttranskription. Das Ergebnis ist PROLAB (Kohler 1997). Dieses Notationssystem wurde zunächst in der Bearbeitung der Lesesprache erprobt. Die aus dieser Arbeit bekannten prosodischen Kategorien wurden dann auch auf die Analyse der Spontansprache übertragen, modifiziert und erweitert. Die erfolgreiche Anwendung von PROLAB zur Etikettierung von Spontansprache ist inzwischen langfristig und umfangreich erprobt (weit über 100 Dialoge von insgesamt 5,5 Stunden Dauer, 53.500 fortlaufende Wörter). Das Transkriptionssystem und die Konventionen seiner Verwendung sind kontinuierlich überarbeitet, empirisch dokumentiert und zu einem Etikettierlehrgang für Studierende und Mitarbeiter ausgebaut worden. Die Dokumentation ist institutsintern auf den Rechnern bereitgestellt und wird zur Einarbeitung von Hilfskräften in die prosodische Etikettierung von Spontansprache eingesetzt sowie als PROLAB-Handbuch verfügbar gehalten (Peters 2001). Das Handbuch wird entsprechend den gewonnenen Erfahrungen und der empirischen und theoretischen Weiterentwicklung des Modells laufend überarbeitet.

Da vor allem in der angelsächsischen, speziell der amerikanischen, aber auch in der deutschen Prosodieforschung heute das Intonationsmodell Pierrehumberts und Beckmans (1980, 1986) sowie die darauf aufbauenden Etikettiersysteme ToBI (Beckman and Ayers 1997) und GToBI (Grice und Baumann 2000) als wissenschaftliche Bezugsrahmen weit verbreitet sind, bedarf es einer Rechtfertigung, weshalb das hier zur Diskussion stehende Projekt von einer anderen theoretischen und methodischen Grundlage ausgeht. KIM und PROLAB unterscheiden sich von (G)ToBI und dem zugrundeliegenden Modell in vier grundsätzlichen Punkten, die insbesondere für die

Untersuchung natürlicher, spontansprachlicher Interaktionen von großer Bedeutung sind. Ein Vergleich der beiden Systeme unter diesen vier Gesichtspunkten soll eine wissenschaftliche Begründung für die Bevorzugung von KIM und PROLAB gegenüber ToBI und seiner Modellgrundlage geben. Es soll damit auch dargelegt werden, welches der spezifische Kieler Beitrag zur Prosodieforschung im allgemeinen ist und in welchen Punkten ToBI und seine theoretische Grundlage von dieser Forschung profitieren kann.

(1) *Tonkonturen vs. Tonstufen*

KIM und PROLAB stehen in der europäischen Tradition der Intonationsforschung, u.a. vertreten durch die Londoner Schule der Phonetik, und basieren auf einem System holistischer Intonationskonturen und ihrer signifikanten syntagmatischen Komponenten. ToBI und seine Modellgrundlage stehen in der amerikanischen Tradition der Tonstufenanalyse, initiiert durch Pike (1945), die auf zwei distinktive Stufen - 'hoch' ('high') und 'tief' ('low') - reduziert und in drei Kategorien - 'Tonakzente' ('pitch accents'), 'Phrasenakzente' ('phrase accents') und 'Grenztöne' ('boundary tones') - eingeteilt wurde.

Die äußerungsprosodische Ebene ist im Gegensatz zur segmentellen und silbentonalen sowohl in der Produktion als auch in der Perzeption und der semantischen Zuordnung durch globale Strukturen und ihre Synchronisierung mit der lexikalischen und syntaktischen Ebene bestimmt. Das ist u.a. durch die Kieler Untersuchungen zur F₀-Gipfelverschiebung entlang der Zeitskala eines sonst gleichbleibenden Sprachsignals gezeigt worden (Kohler 1991): die zeitlich verschobenen Gipfelkonturen bilden eine Serie holistischer terminaler Intonationsmuster, deren Wahrnehmung sich mit der Synchronisierung zum Sprachsignal, nicht mit den Tonstufen ändert, und zwar kategorial an ausgezeichneten Silbenstellen mit gleichzeitig unterschiedlicher semantischer Zuordnung. Daraus resultieren drei distinktive Zeitpositionen einer Gipfelkontur im Bereich der betonten Silbe eines akzentuierten Wortes: 'früh', 'mittel', 'spät'. In ToBI werden diese drei Gipfeltypen in phrasenfinaler Position symbolisiert durch H+!H* L-L% ('früh'), H* L-L% bzw. L+H* L-L% ('mittel') und L*+H L-L% ('spät'). In dieser Kategorisierung kommt die zeitliche Progression in der Positionierung des gleichen Gipfelmusters nicht konzeptionell zum Ausdruck.

Außerdem wird durch diese Zuordnung zu unterschiedlichen Tonsequenzen nicht expliziert, daß in einer Gipfelverschiebungsserie von links nach rechts 'frühe' Gipfel perzeptorisch kategorial zu 'nicht frühen' umkippen, während 'nicht frühe' eher als eine graduelle Veränderung wahrgenommen werden. 'frühe' Gipfel bilden eine Klasse für sich gegenüber der Klasse der 'nicht frühen' Gipfel, einerseits phonetisch durch eine unterschiedliche Fokussierung auf den abfallenden bzw. den aufsteigenden Ast der F₀-Gipfelkontur in der Synchronisierung zum betonten Vokal, andererseits semantisch durch den Gegensatz 'bekannt, wissend, feststehend, geschlossen' vs. 'neu, erfahrend, erkennend, offen'. Die Bedeutung des 'späten' Gipfels unterscheidet sich von der des mittleren durch die *zusätzliche* semantische Komponente des 'Gegensatzes zum eigenen Urteil'. Die phonologische Kategorisierung in ToBI reflektiert weder die phonetische noch die semantische Klassenbildung. Die ursprüngliche Pierrehumbertsche Phonologisierung des 'frühen Gipfels' als H+L* war viel besser, da sie den Gegensatz zu den H*-Komponenten der anderen beiden Gipfelmuster, d.h. eine Fokussierung auf die Tiefen der absteigenden gegenüber den Höhen der aufsteigenden F₀-Konturen adäquater erfaßte.

GToBI (Grice und7 Baumann 2000) markiert den typischen 'frühen' Gipfel wieder als H+L*, hat daneben aber noch einen zweiten Typ des 'frühen Gipfels', der gradueller abfällt und als H+!H* symbolisiert wird. Die semantische Einordnung als 'feststehend' bleibt erhalten; es kommt im zweiten Typ eine persönliche Bewertung hinzu. Der Unterschied zwischen beiden Typen liegt in der Geschwindigkeit des F₀-Abstiegs, was in der GToBI-Kategorisierung nicht zum Ausdruck kommt, weil die Differenzierung wieder auf Tonebenen festgemacht wird und nicht die Verlaufsstrukturen erfaßt. Auch in 'mittleren' und 'späten Gipfeln' kann der Abstieg hinsichtlich

seiner zeitlichen Strukturierung ('schneller/langsamer') differenziert werden. Diese Geschwindigkeitsparameter müssen auch noch in KIM und PROLAB eingebaut werden.

Ein warnendes "Du!" (mit steigend-fallend-steigendem Tonhöhenverlauf) ist eine globale Intonationseinheit im Gegensatz z.B. zu einer steigend-fallenden in "Du." oder einer steigenden in "Du?", und diese Intonationseinheiten werden als solche produziert und perzipiert und ganzheitlich mit unterschiedlichen Bedeutungen verknüpft. ToBI hingegen spaltet diese einheitlichen Konturen in eine Sequenz von drei Toneinheiten 'hoch/tief' auf: 'Tonakzent' + 'Phrasenakzent' einer 'Intermediärphrase' ('intermediate phrase') + 'Grenzton' einer 'Intonationsphrase' ('intonation phrase'). Der 'Schleifton' auf dem warnenden "Du!" benötigt zu seiner eindeutigen Signalbestimmung vier signifikante F0-Punkte - den tiefen Anfang, das Maximum, das Minimum und den nachfolgenden höheren Endpunkt. ToBI läßt maximal zwei Tonpunkte in einem 'pitch accent' zu, ohne diese Begrenzung theoretisch zu begründen; die weitere Einschränkung der bitonalen Kombinationsmöglichkeiten wird mit dem Verweis auf die Empirie in einer Sprache plausibel gemacht (wobei das Englische eine dominierende Referenzfunktion hat, ähnlich dem Latein in der Diskussion zur Universalgrammatik), aber wiederum nicht erklärt. Der 'Schleifton' benötigt also zwei weitere Tonpunkte, die ihm durch Phrasentöne zugewiesen werden. Während "Du." und "Du?" ohne weiteres durch *einen* Phrasenton auskommen könnten, nämlich den finalen, braucht "Du!" einen zweiten, der dann auch kategorial von einem anderen Typ sein muß. Er ist dann der 'phrase accent', der mit nichtfinalen Intonationsphrasen verbunden wird. Für eine finale Intonationsphrase wird nun generell die Sequenz aus nichtfinaler und finaler Phrasentonmarkierung postuliert. Was bei "Du!" aus dem Zwang zur phonetischen Beschreibungsadäquatheit folgt, wird in den übrigen Fällen zum Systemzwang.

ToBI kann schließlich komplexe prosodische Phänomene bei obligatorischer Aufteilung auf drei Akzentkategorien überhaupt nicht mehr erfassen, wenn z.B. auf einen Schleifton ein steigender Ton an einer weiteren Satzakkzentposition innerhalb einer prosodischen Phrase folgt. Solch eine Möglichkeit gibt es beispielsweise im Deutschen: "*War er das wirklich?*" mit zwei Satzakkzenten auf "*war*" und "*wirklich*" sowie mit einem Schleifton auf "*war er das*" und anschließend einem getrennten steigenden Ton auf "*wirklich*", der unter dem Endpunkt des Schleiftons einsetzt, ohne ein wahrnehmbares syntagmatisches prosodisches Gliederungssignal für eine intervenierende Phrasengrenze (u.a. keine finale Längung). Diese intonatorische Abfolge ist auditiv und semantisch klar verschieden von der Sequenz aus fallendem Ton auf "*war er das*" + steigendem Ton auf "*wirklich*". Der Schleifton bringt Erstaunen, Nicht-glauben-wollen zum Ausdruck, während das andere Muster nur das Verb fokussiert und kontrastiert. Außerdem ist das vorgestellte Schleiftonmuster auch auditiv und semantisch verschieden von einem solchen mit intervenierender Phrasengrenze: in dem Fall wird "*wirklich*" zum Nachtrag. (Ferner unterscheiden sich alle genannten Muster von der einakzentigen Realisierung mit dem Schleifton über der gesamten Wortfolge: da ist "*wirklich*" unbetont, verliert die nachdrückliche Bedeutung "*tatsächlich, in Wirklichkeit*" und wird zur Modalpartikel.) ToBI müßte zwei *pitch accents* auf "*war*" und "*wirklich*" ansetzen, könnte aber dem ersten nur zwei Tonstufen zuordnen und damit den Schleifton nicht abbilden. Selbst das kontrafaktische Postulat einer Intermediärphrase würde das Problem nicht lösen; erst der Ansatz einer Intonationsphrase gäbe die theoretische Möglichkeit der Repräsentation des Schleiftons im theoretischen Rahmen von ToBI, aber das ist erst recht nicht phänomenadäquat.

Die besprochenen globalen Intonationskonturen sind mit betonten Wörtern in der Äußerung verbunden und werden ihrerseits in eine Hierarchie übergeordneter globaler Muster konkateniert, die hinsichtlich ihrer prosodischen Kohäsion skaliert sind zwischen kompletter Integration und völliger Separation. So können Gipfelkonturen eine Sequenz mit sukzessive tieferen F0-Gipfelpunkten ('downstep') bilden; größere prosodische Kohäsion wird durch die Integration in ein 'Hutmuster' mit Einebnung der Abstiegs-Aufstiegs-Folgen erreicht, größere Separation andererseits

durch Vergrößerung der F0-Einbuchtungen zwischen den Gipfelpunkten und durch Neueinsatz des F0-’downstep’. Derartige globale Konkatenationskonturen werden schließlich in prosodische Phrasen eingegliedert, deren Grenzen unterschiedlich starke Trennungssignale aufweisen (siehe (4)).

Ein Intonationsmodell muß die grundlegende Eigenschaft der Prosodie erfassen, eine Hierarchie globaler Strukturen (Tonhöhenverläufe, Intensitäts- und Zeitstrukturen) in syntagmatischer Abfolge mit wechselnden Graden der Kohäsion zwischen den konkatenierten Einheiten zu bilden. Die ausschließliche Reduktion der Äußerungsprosodie auf Sequenzen lokaler Töne in zwei Stufen, wie sie nach ToBI und dem zugrundeliegenden autosegmentell-metrischen Intonationsmodell vertreten wird, ist nicht phänomengerecht und deshalb von KIM und PROLAB nicht übernommen worden.

(2) *Perzeptorische Basis vs. Orientierung an F0 und abstrakter Phonologie*

In KIM und PROLAB ist das Ziel die Erstellung einer prosodischen Phonologie und die Annotation großer Datenkorpora der Lese- und Spontansprache mit einem darauf aufbauenden Etikettiersystem. Beide - phonologische Modellierung und Annotation - operieren auf perzeptorischer Basis, d.h. für die Kategorienbildung sind die für Hörer signifikant differenzierten prosodischen Muster entscheidend. In der Modellierung hat deshalb die kontrastive auditive Evaluation prosodischer Äußerungsphänomene durch Bearbeiter sowie Versuchspersonen in formalen Perzeptionsexperimenten einen zentralen Stellenwert; F0-Analysen und linguistische Systemüberlegungen haben dabei nur Hilfsfunktionen. Diese Modellbildung basiert auf experimenteller Phonologie, d.h. die experimentelle Untersuchung füllt nicht vorweg konzipierte phonologische Kategorien mit phonetischer Substanz, sondern trägt entscheidend zur phonologischen Kategorienbildung selbst bei. Experimentelle Phonologie - im Gegensatz zur Laborphonologie - bringt das Labor in die Phonologie anstatt die Phonologie ins Labor.

Das hat zur Konsequenz, daß Etikettierer ihre Annotationen in erster Linie auf auditiver Basis durchführen müssen und dabei die graphischen Ausgaben von F0-Verläufen nur unterstützend heranziehen dürfen, nicht aber als die Kategorisierung bestimmende Faktoren. Dies wird nicht nur durch die algorithmisch bedingten Signalanalysefehler und die vielfältigen mikroprosodischen Störeinflüsse auf die Äußerungsprosodien notwendig, sondern auch deshalb, weil die “allotonische” Variabilität in der F0-Ausprägung unter vielfachen kontextuellen Einflüssen noch gar nicht bekannt ist und weil mit der Notwendigkeit neuer Kategorien an spontansprachlichem Material stets gerechnet werden muß. Letzteres hat auch zur Konsequenz, daß initiale Kategorienansätze nur einen heuristischen Steuerungswert haben können, deshalb nicht verabsolutiert werden dürfen und stets für Modifikationen offen sein müssen. Die an die Daten herangetragenen phonologischen Kategorisierungen, die F0-Signalanalysen und die auditive Beurteilung konvergieren in der interaktiven Resynthese prosodisch etikettierter Äußerungen in der Weise, daß prototypische Realisierungen verschiedener distinktiver Annotationsmöglichkeiten synthetisiert und kontrastiv perzeptorisch evaluiert werden, um die am besten passende symbolische Zuordnung festzulegen.

In ToBI verführt die Verwendung der Tonstufenetiketten ‘hoch’ und ‘tief’ dazu, sie in einem signalphonetischen (statt symbolischen) Sinn zu interpretieren und mit konkreten F0-Ebenen in Verbindung zu bringen, also die F0-Kurven zur Richtschnur zu machen. In den ‘Guidelines for ToBI labelling’ (Beckman und Ayers 1997) wird in 1.4 “Marianna made the marmelade.” mit L* *? L* H-H% annotiert, in 2.3 “My name is Marianna.” mit H* H-H%. Der generelle Tonhöhenverlauf von “made the marmelade” und “my name is Marianna” ist völlig vergleichbar: ein hoher F0-Anstieg in weiblichen Stimmen ab der letzten betonten Silbe bis zum Äußerungsende, nach einem mehr oder weniger ebenen tiefen F0. Der Unterschied in der Zuordnung zu L* bzw. H* hängt ganz offenbar damit zusammen, daß im ersten Beispiel (L*) F0 am tiefsten Punkt vor dem Anstieg bei etwa 140Hz liegt, im zweiten Beispiel (H*) aber bei 170Hz. Gleichzeitig setzt aber im ersten Fall F0 bei ca. 200Hz, im zweiten Fall bei ca. 230Hz ein, d.h. der Vorlauf vor dem Anstieg ist insgesamt um ca. 30Hz nach oben verschoben. Das hängt eher mit unterschiedlichem Register zusammen und

kann auf keinen Fall die Vergabe von H* vs. L* als unterschiedliche phonologische Kategorisierung einer Intonationskontur rechtfertigen. In beiden Fällen müßte L* vergeben bzw. das globale Intonationsmuster als eine hoch ansteigende Talkontur kategorisiert werden.

Liegt in den soeben zitierten Beispielen eine inadäquate Kategorienzuordnung aufgrund einer Überkonkretisierung vor, so zeigen die folgenden Beispiele gerade die gegenteilige Problematik einer Überabstraktion, weil eine phonologische Kategorisierung von der konkreten F0-Manifestation losgelöst und in einem System generalisiert wird. In 2.8 der 'Guidelines' wird die stilisierte Rufintonation, z.B. auf "Marianna", als H* !H-L% etikettiert, wo H- L% die Bedeutung 'high phrase accent' plus 'low boundary tone' verloren hat und nur noch symbolisiert, daß die Tonhöhe auf der letzten Silbe gleichbleibend in mittlerer Höhe ist, wohin sie in einer Stufe von einer höheren Ebene auf der Akzentsilbe absteigt: die Etiketten H und L werden hier zu reinen Benennungen. GToBI hat dies erkannt (Grice und Baumann 2000) und verwendet !H-%.

(3) Akzentkodierung weitgehend durch melodische vs. ausschließlich durch tonale Merkmale

In KIM und PROLAB sind Satzakzent (d.h. Betonung von Wörtern in der Äußerung) und Intonation getrennte Kategoriensysteme, was auch bedeutet, daß Satzakzent durch andere Merkmale (Dauer, Intensität, Stimmqualität) allein kodiert werden kann oder daß er sich sowohl melodisch als auch nichtmelodisch manifestiert, gleichzeitig mit unterschiedlichen Funktionen. So hat in der Äußerung "und war irgendwie nicht ganz bei Sinnen und dann aus'm - rausgetorkelt." (1041a009, neue, überlappende Dialoge) das verbale Kompositum einen primären Satzakzent auf "raus". "getorkelt" hat üblicherweise einen sekundären Satzakzent. Der primäre Akzent ist im Korpusbeispiel mit einer Gipfelkontur verbunden. Der sekundäre Akzent kann z.B. mit einer untergeordneten Gipfelkontur gekoppelt sein. Es ist aber auch die völlige Übernahme der Kodierung durch die anderen Merkmale möglich. In dem Fall ergibt sich auch die Möglichkeit, daß der Satzakzent auf dem Verb die gleiche Stärke wie derjenige auf dem adverbialen Teil bekommt (oder noch stärker wird), gleichzeitig aber die Gipfelkonturierung und damit die melodische Akzentuierung des Kompositums erhalten bleibt. Das ist es, was im Korpusbeispiel vorliegt. Die Intonation liefert die akzentuelle Fokussierung, die anderen Merkmale gleichzeitig die emphatische Heraushebung der Semantik des Verbs. Wir haben es hier also mit einem Druckakzent zu tun. Er manifestiert sich unter anderem auch durch Dauer- und Intensitätsverstärkung silbeninitialer Konsonanten (vgl. *accent d'insistance* im Französischen).

Die Manifestation des Druckakzents kann zusätzlich durch eine vorangesetzte Pause verstärkt werden, wie z. B. in "verfluchte <Pause> Schscheiße" ohne daß dadurch eine Phrasengrenze entsteht. Das prosodische Phänomen der Pause ist hier mit nachfolgender initialer, nicht mit vorangehender, finaler Längung gekoppelt und damit nicht ein Merkmal prosodischer Phrasierung, sondern ein Merkmal des Druckakzents.

Da derartige Satzakzentuierungsmuster in der Spontansprache auf Schritt und Tritt anzutreffen sind, muß die prosodische Modellierung die entsprechenden getrennten Kategorien bereitstellen. In ToBI und dem autosegmentell-metrischen Modell sind Akzente ausschließlich durch tonale Muster kodiert. Damit können sie bestimmte Bereiche der prosodischen Empirie nicht modellieren.

(4) Multifaktorielle vs. tonale Kodierung der prosodischen Phrasengrenzen

PROLAB-Etikettierungen der Datenbasis haben für flüssige Sprache (nach anfänglicher Differenzierung zwischen zwei Typen von prosodischen Phrasengrenzen) nur noch einen Typ berücksichtigt, der als ein Bündel von Merkmalen (Pause, Atmen etc., finale Längung, F0-Einschnitte, 'F0-Reset') in wechselnder Gewichtung definiert wird. Durch die wechselnde Gewichtung ergeben sich unterschiedliche Grade der Kohäsion zwischen den prosodischen Einheiten vor und nach der prosodischen Grenze, d.h. unterschiedliche Phrasengrenztypen. Aber diese werden nicht etikettiert, sondern die Information über das Merkmalsbündel und die

Gewichtung seiner Komponenten soll nach erfolgter manueller Etikettierung automatisch über die vorhandenen Etikettierungen und die segmentelle Dauerstruktur extrahiert werden, so daß dann post festum eine phonetisch fundierte und statistisch validierte Subklassifikation des einen Phrasengrenztyps vorgenommen und in die Etikettierung eingespeist werden kann.

ToBI unterscheidet unter seinen fünf 'break indices' drei (2, 3, 4), die mit prosodischer Phrasierung zusammenhängen. 3 und 4 sind ausschließlich tonal definiert, nämlich durch 'Phrasenakzent' am Ende einer 'Intermediärphrase' und durch 'Grenzton' am Ende einer 'Intonationsphrase', obwohl sie ganz offenbar auch durch andere Merkmale signalisiert werden. Das wird durch das Ansetzen von BI 2 deutlich, wo entweder die tonalen Merkmale wie für BI 3 und 4 auf der Ebene der Tonmarkierung vorhanden sind, aber Etikettierer dennoch den Eindruck haben, daß der BI schwächer ist, oder aber daß keine tonalen Merkmale vorhanden sind, aber dennoch ein prosodischer Einschnitt wahrgenommen wird. Während also BI 3 und 4 einfach die auf der Tonmarkierungsebene bereits gelieferten Kategorisierungen übernehmen und (redundant) auf der BI-Ebene neu einordnen, ist BI 2 von vornherein durch andere Grenzmerkmale bestimmt, ohne daß diese klar definiert würden, was auch zu großer Unsicherheit und niedriger Reliabilität bei der Etikettierung führen muß. Der erstgenannte Fall unter BI 2 macht aber in der Negation deutlich, daß selbst für B3 und B4 andere Merkmale als Ton eine Rolle spielen, wenn auch nur intuitiv und nicht klar expliziert.

Zusammenfassend läßt sich sagen, daß die Untersuchungsstrategie von PROLAB der von ToBI vorzuziehen ist, da sie phänomenadäquater, klarer definiert und durch die Etikettierer valider und reliabler handhabbar ist.

3 Ergebnisse

Die Untersuchungen beschränkten sich nicht auf eine Beschreibung und Modellierung satzphonetischer Phänomene in den Bereichen der artikulatorischen Reduktion und der prosodischen Muster sowie deren Interaktion, sondern zielten auf die Erklärung der Erscheinungen durch Rückführung auf allgemeine Prinzipien der sprachlichen Kommunikation. Hierzu gehören:

- die Tendenz des Sprechers, generell artikulatorischen Aufwand zu reduzieren, ihn aber auch für bestimmte Funktionen, z.B. zur emphatischen Vertärkung gezielt zu erhöhen
- die Erfordernisse des Hörers nach Distinktivität
- die durch die Kommunikationssituation gesetzten Rahmenbedingungen für Distinktivität
- die Notwendigkeit, den Sprechfluß entsprechend den semantischen und pragmatischen Intentionen syntagmatisch zu strukturieren, zum einen durch syntaktische, zum andern durch prosodische Mittel, die nicht zu konvergieren brauchen
- die Störung der flüssigen syntagmatischen Strukturierung durch Variable des Kommunikationsablaufs: Probleme der Wortfindung und Formulierung sowie Sprecherinteraktionen im Dialogablauf führen zu Abbrüchen, Unterbrechungen, Wiederaufnahmen, Häsitationen
- die paradigmatische Differenzierung linguistischer Einheiten, z.B. wortklassenspezifische Reduktionsregeln

- die Zuordnung melodischer Muster zu semantisch-pragmatischen Kategorien gemäß einem *Frequenzcode*, der Hochtöne mit Unterordnung, Tieftöne mit Dominanz verbindet (Ohala 1984).

Hinsichtlich der Einbeziehung des Hörers und der sprachlichen Funktion in die Analyse der Äußerungsphonetik unterscheidet sich die hier vorgestellte Forschung grundlegend vom Ansatz sowohl der Artikulatorischen Phonologie (Browman and Goldstein 1992) als auch der Auto-segmentell-metrischen Phonologie und ToBI (Pierrehumbert 1980, Beckman & Ayres 1994).

3.1 Artikulationsabschwächung

In der ersten Projektphase (1997-2000) wurde die Artikulationsabschwächung untersucht. Die Ergebnisse sind in Kohler (2001a: AIPUK 35(2001)) im einzelnen präsentiert worden. Sie seien hier noch einmal zusammengefaßt.

3.1.1 Theoretische Grundlagen

- Die allgemeinen Prinzipien der sprachlichen Kommunikation bedingen statistisch variable Reduktionsmuster. Formen unterschiedlicher Reduktionsgrade lassen sich entlang artikulatorischer Skalen von kanonisch-phonologischer Form bis zu extremer Vereinfachung anordnen.
- Artikulierte Sprache kann als Abfolge von Öffnungs-Schließungsbewegungen des Ansatzrohrs analysiert werden, in die Symbolfolgen phonetischer Transkriptionen übersetzt werden müssen.
- Im Sprachproduktionsprogramm operieren Reduktionen über solchen globalen Bewegungseinheiten und ihren Abfolgen via Reduktionskoeffizienten, die auf hoher Verarbeitungsebene dem Programm den zu erzielenden Reduktionsgrad vorgeben.

3.1.2 Hypothesen

Hypothese 1 - die Bedingung des Sprechstils

Ceteris paribus sind satzphonetische Reduktionsphänomene häufiger in spontaner als in gelesener Sprache.

Hypothese 2 - die Bedingung des Satzakzents

Ceteris paribus sind satzphonetische Reduktionsphänomene häufiger in unakzentuierter Position und am häufigsten in unakzentuierten Funktionswörtern.

Hypothese 3 - die Bedingung der syntagmatischen Kohäsion

Der Grad der Kohäsion zwischen syntagmatischen Einheiten (Silben, morphosyntaktische und prosodische Strukturen) bestimmt die Manifestation von Reduktionsphänomenen an den Grenzen bzw. im Innern.

Hypothese 4 - Reduktion der räumlichen und zeitlichen Extension von Öffnungsgesten

Vokalische Öffnungen in Öffnungs-Schließungsbewegungen tendieren in unakzentuierten Funktionswörtern und unmittelbar nach betonten Silben in Inhaltswörtern zur Reduktion zu [ə] und schließlich zur Tilgung. Die Tilgung ist besonders häufig in der artikulatorischen Struktur 'orale Okklusion + Schwa + Nasal'.

Hypothese 5 - Eliminierung des apikalen Artikulators in Nasal-/Plosivproduktionen

- (a) In Silben mit [ə]-Tilgung gemäß Hypothese 4 passen sich apikale Okklusionen an labiale und dorsale an.

- (b) Die Anpassung erfolgt auch progressiv über Wortgrenzen hinweg.

Da die wortinterne Kohäsion in (a) größer ist als die wortübergreifende Kohäsion in (b), ist für die Anpassung unter (a) eine größere Auftretenshäufigkeit zu erwarten. Die syntaktische und prosodische Gruppierung und der dadurch resultierende Kohäsionsgrad über die Wortgrenze bestimmen, ob die apikale Geste ausgeführt wird oder nicht.

Hypothese 6 - keine Eliminierung des apikalen Artikulators in Frikativsequenzen

Hypothese 7 - Kontrolle der Velumbewegung für Plosive in nasaler Umgebung

- (a) In Gesten gemäß Hypothese 4 kann der Reduktionsgrad durch regressive Anpassung der Velumbewegung in Sequenzen von 'Lenisplosiv + Nasal' über die Reduktion des Öffnungsgrads und die Ortsassimilation hinaus erhöht werden. Es ist zu erwarten, daß eine solche Erhöhung weniger häufig auftritt.
- (b) Für Fortisplosive ist eine solche Reduktion wegen der sehr viel längeren zeitlichen Sicherheitszone im Plosivverschluß allenfalls in unakzentuierten Funktionswörtern zu erwarten.
- (c) In der umgekehrten Abfolge 'Nasal + Plosiv' ist eine noch geringere Auftretenswahrscheinlichkeit der Nasalisierung zu erwarten, weil der Prozeß hier den perzeptiv gewichtigeren initialen Teil der Geste betrifft. Der Prozeß ist am seltensten über Wortgrenzen hinweg.
- (d) In beidseitig nasalem Kontext können Lenisplosive entweder nasaliert oder glottalisiert werden; ersteres wird für Fortisplosive wieder selten auftreten.

Hypothese 8 - Reduktion der räumlichen und zeitlichen Extension von Schließungsgesten

- (a) Lenisierung von Obstruenten zwischen Vokoidöffnungen aufeinander folgender Gesten: Plosive und Frikative > Approximanten, stimmlos > stimmhaft infolge Dauerreduktion
- (b) Eliminierung der Okklusion für einen Nasal bei gleichzeitiger Nasalisierung des vorangehenden Vokals: dies ist im Deutschen eine extreme Reduktion und hängt stark von den Bedingungen der Hypothesen 1 - 3 ab.
- (c) Geminatenreduktion: sie ist im Deutschen eine generelle Tendenz, besonders in Klustern, vor allem in Folgen von unbetonten Silben und Funktionswörtern.

Hypothese 9 - Reduktion des Laterals

Sie tritt vor allem in gestenfinaler Position in unbetonten Silben und Funktionswörtern auf.

Hypothese 10 - Glottalisierung in wortinitialen Vokalen

Die Markierung der Wortgrenze durch das Einsetzen des intialen Vokals variiert entlang einer Intensitätsskala von starkem Glottalverschluß über schwächere Glottalisierung und Amplitudenmodulation zu völligem Schwund. Der Stärkegrad ist eine Funktion der Bedingungen in Hypothesen 1 - 3.

3.1.3 Deskription und Evaluation

- Die Daten des *Kiel Corpus* zeigen unter den Hypothesen 4 - 10 einen starken Effekt
 - des Sprechstils (Hypothese 1), mit Ausnahme der Geminatenreduktion in Hypothese 8

- der Akzentuierung (Hypothese 2)
 - der Kohäsion (Hypothese 3).
- Die progressive Reduktion von ‘Plosiv + Schwa + Nasal’ nach Hypothesen 4, 5 und 7 schlägt sich in einer gegenläufigen Häufigkeit nieder.
 - Unterschiedliches Verhalten von Lenis- und Fortisplosiven hinsichtlich der Nasalierung bzw. Glottalisierung nach Hypothese 7 ist bestätigt worden.
 - Schwächere Kohäsion über Wortgrenzen als innerhalb der Wörter führt weniger häufig zu Orts- bzw. Nasalassimilation nach Hypothesen 4 und 7.
 - Das Fehlen der apikalen Ortsassimilation in Frikativen nach Hypothese 6 ist bestätigt worden.
 - Die Regularitäten der Lenisierung, der Reduktion von Nasalen, Lateralen und Geminaten folgen den Hypothesen 8 und 9: allerdings sind die Auftretenshäufigkeiten, abgesehen von der Geminatenreduktion, recht klein, verglichen mit den Befunden gemäß den Hypothesen 4, 5, 7.
 - Die relative Häufigkeit der Schwaelision in ‘Plosiv + Schwa + Nasal’ liegt bei 90%, gleichzeitige Ortsassimilation etwas niedriger, aber Nasalassimilation ist unter 40%.
 - Die Integration von ‘Plosiv+Schwa+Nasal’ zu einer Plosiv-Nasal-Verbindung gleichen Artikulationsorts ist im heutigen Deutsch lexikalisiert. Realisierungen mit [ə] sind Verstärkung.

3.1.4 Modellierung

- Die in die Hypothesen eingegangenen und in der Korpusanalyse statistisch validierten Bedingungen für artikulatorische Reduktion in verbundener Sprache erlauben nunmehr die Vorhersage von Aussprachevarianten deutscher Wörter in dem durch das *Kiel Corpus* repräsentierten Deutsch, jeweils ausgehend von kanonischen Zitierformen.
- Diese Modellierung läßt sich in zwei Richtungen ausbauen:
 - im synchronen Sprachvergleich
 - und im historischen Lautwandel.
 Für beide gilt z.B. gleichermaßen, daß Reduktion, Nasalierung, Elision von Vokalen, Lenisierung von Obstruenten, Ortsassimilation und Nasalierung von Plosiven sowie Reduktion von Lateralen und Geminaten, jeweils in spezifischen Kontexten, immer wieder auftreten, weil sie als allgemeine Erscheinungen in der sprachlichen Kommunikation anzusetzen sind. Sie werden gesteuert durch die Tendenz des Sprechers, artikulatorischen Aufwand zu reduzieren, und kontrolliert durch das situationsspezifische Erfordernis des Hörers nach Distinktivität.

3.2 Prosodie

Artikulatorische Prozesse auf Äußerungsebene operieren innerhalb prosodischer Strukturen. Diese sind nach dem Kieler Intonationsmodell (KIM) auf drei Ebenen angesiedelt (Kohler 1991, 1997):

- syntagmatische prosodische Phrasierung zur inhaltlichen Gliederung des Gesprochenen mit unterschiedlichen Graden der Kohäsion an den Grenzen durch Bündel phonetischer Merkmale
- Prominenzabstufungen zur semantischen Fokussierung und Emphase

- den tonalen Akzenten zugeordnete paradigmatische holistische Melodiemuster zum Ausdruck semantisch-pragmatischer Funktionen.

Prosodische Phrasierung, Prominenzabstufungen sowie Melodiemuster bilden Rahmenbedingungen für die artikulatorische Gestaltung verbundener Sprache

- Assimilationen und Glottalisierungsphänomene in Abhängigkeit von der Phrasenkohäsion
- artikulatorische Reduktion bzw. Verstärkung bei Fehlen oder Vorhandensein des Akzents.

Die prosodische Strukturierung auf den drei Ebenen ist aber auch ein eigenständiges Untersuchungsfeld der Signal-Funktionsbeziehung des Gesprochenen. Dies ist der Gegenstand der zweiten Projektphase (2000-2003). Auch hier sollte die Spontansprache im Vergleich zur Lesesprache untersucht werden, da Hypothese 1 zur Artikulationsabschwächung entsprechend auch hier anzusetzen ist: Phrasierung, Akzentuierung und Melodiemuster sind unterschiedlich in beiden Sprechstilen und erlauben eine sofortige Entscheidung darüber, ob jemand abliest oder frei spricht. Da das lesesprachliche Korpus jedoch weitgehend aus Einzelsätzen und nur zu einem ganz geringen Teil aus Texten bestand und da die drei prosodischen Strukturebenen sich nur in Texten voll entfalten können, war die Vergleichbarkeit mit dem spontansprachlichen Korpus eingeschränkt. Es hätten zusätzliche gelesene Texte aufgenommen und etikettiert werden müssen, wozu jedoch die Zeit fehlte. Daher wurde in der zweiten Projektphase die Spontansprache nicht mehr im Vergleich zur Lesesprache untersucht.

3.2.1 Prosodische Phrasierung

3.2.1.1 Theoretische Grundlagen

- Die syntagmatische Strukturierung des Gesprochenen erfolgt auf zwei Ebenen: durch syntaktische Struktur und durch prosodische Phrasierung, die auf einer graduell gestuften Konvergenz-Divergenz-Skala interagieren.
 - Prosodische Phrasierungsgrenzen werden einem Hörer durch verschiedene Bündelungen signalphonetischer Merkmale angezeigt:
 - segmentelle Längung im Vergleich zur lautlichen Umgebung
 - separierende Intonationsmuster (stark fallend, stark steigend, fallend-steigend)
 - Neueinsatz der F₀-Deklinations der Satzakkente (Reset)
 - Pausen, Atmen und andere paralinguistische Phänomene (z.B. Schnalzlaut)
 - Phonationsphänomene (Verhauchung, Knarren).
 - Die unterschiedlichen Bündelungen bilden eine Hierarchie der Kohäsions-/Trennungsstärke zwischen prosodischen Phrasen.
 - Im Fall der Konvergenz zwischen syntaktischer Struktur und prosodischer Kohäsion ergibt sich für den Hörer die **flüssige prosodische Phrasierung**.
 - Durch Probleme der inhaltlichen Äußerungsplanung, Wortfindung/Formulierung und Artikulation können sich unterschiedliche Grade der Divergenz zwischen beiden Ebenen ergeben.
 - Zum einen werden syntaktisch schwache Grenzen, z.B. zwischen Artikel und Substantiv, prosodisch stärker gewichtet, als ihnen aufgrund der syntaktischen Struktur zusteht.
 - Zum anderen wird die segmentelle Längung erhöht und weiter über sprachliche Einheiten ausgedehnt: Häsilationslängung.
 - Ferner werden Häsilationspartikeln eingefügt.
 - Schließlich kann die syntaktische Struktur selbst durch Abbrüche, Wiederaufnahmen und Wiederholungen gestört werden.
- Daraus resultiert für den Hörer eine Skala zunehmend **unflüssiger prosodischer Phrasierung**.

- Prosodische Phrasierung am Ende eines Redebeitrags bzw. vor der Übernahme durch einen Dialogpartner ist phonetisch anders ausgeprägt als turnintern.

3.2.1.2 Hypothesen

Hypothese 1 - signalphonetische Ausprägung auditiv etikettierter Phrasengrenzen

Auditiv etikettierte Phrasengrenzen haben mindestens eines der signalphonetischen Merkmale.

Hypothese 2 - finale Längung

(a) Segmentelle Längung ist das Merkmal mit der größten Vorkommenshäufigkeit.

(b) Sie ist turnfinal stärker ausgeprägt als turnintern.

Hypothese 3 - Bündelung der Merkmale und Kohäsionsgrade

Bündel der signalphonetischen Merkmale kodieren in Abhängigkeit von Anzahl und Ausprägung den Grad der Separierung bzw. Kohäsion zwischen aufeinander folgenden Phrasen und bilden so die inhaltliche Gliederung und Äußerungsplanung ab.

Hypothese 4 - unflüssige prosodische Phrasierung

Phrasengrenzen bei Störungen des Sprechflusses (Häsitationslängung/syntaktische Abbrüche) unterscheiden sich in der Ausprägung der phonetischen Merkmale von flüssiger Phrasierung.

3.2.1.3 Deskription und Evaluation

- In den Daten des *Kiel Corpus* findet sich in 93.1% der turnintern etikettierten Phrasengrenzen mindestens eines der signalphonetischen Merkmale. Damit ist Hypothese 1 bestätigt.
- Finale Längung kommt turnintern häufiger vor als Reset, separierende Intonationsmuster oder Pause/Atmen. Bei Zusammenfassen der tonalen Merkmale Reset und Intonationsmuster ist die Klasse der durch F0 kodierten Merkmale jedoch die größte. Unter diesen Bedingungen ist Hypothese 2(a) nicht bestätigt.
- Finale Längung ist turnfinal stärker ausgeprägt. Damit ist Hypothese 2(b) bestätigt.
- Die signalphonetischen Merkmale treten in vielfältiger Bündelung auf. Die häufigste Bündelung ist finale Längung + Pause/Atmen. Die Korpusstatistik erlaubt eine Einteilung der turninternen flüssigen Phrasengrenzen in zwei Klassen: die tonal und/oder temporal kodierte (PG1) und die Klasse mit Pause/Atmen als zusätzlichem Merkmal. Bei kurzer Pause/Atmen (<700ms) wird PG2, bei langer Pause/Atmen ohne tonale Kodierung PG3, mit gleichzeitiger tonaler und temporaler Kodierung PG4 angesetzt. Anhand ausgewählter Beispiele lassen sich diese vier Klassen von Phrasengrenzen in eine Hierarchie abnehmender Kohäsion einordnen, die die inhaltliche Gliederung und Äußerungsplanung abbildet. Diese Untersuchung muß noch auf das gesamte Korpus ausgedehnt werden: so ist Hypothese 3 nur teilweise bestätigt.
- An Phrasengrenzen mit Zögerungslängung oder syntaktischen Abbrüchen treten weit weniger separierende Konturen als bei flüssiger Phrasierung auf: in vier Fünftel der Fälle kommt es nur zu geringen tonalen Bewegungen. Außerdem variiert die Häsitationslängung sehr stark, ist oft extremer und erstreckt sich über mehrere Silben oder sogar Wortgruppen. Phrasengrenzen an syntaktischen Abbrüchen unterscheiden sich durch (harte) Glottalisierung/Glottalverschluß von der (weichen) finalen Laryngalisierung flüssiger Phrasengrenzen.

3.2.1.4 Modellierung

Die bisherige Korpusanalyse hat teilweise die Voraussetzungen für eine umfassende Modellierung der syntagmatischen Gliederung des Gesprochenen durch prosodische Phrasierung geschaffen. Sie muß noch durch eine umfangreichere Datenevaluation hinsichtlich der Kohäsion und ihrer funktionalen Zuordnung sowie durch Analyse spontanerer Interaktionsdaten (*Lindenstraße-Korpus*) ergänzt werden, die auch die Einbeziehung des Sprecherwechsels in die prosodische Phrasierung ermöglicht. Dann wird sich eine automatische Subklassifikation der etikettierten Phrasengrenzen nach den signalphonetisch definierten Typen sowie die automatische Bestimmung des Kohäsionsgrades zwischen aufeinander folgenden Phrasen durchführen lassen, als Grundlage für die semantisch-pragmatische Interpretation sprachlicher Interaktionen.

3.2.2 Melodische Gestaltung prosodischer Phrasen

3.2.2.1 Theoretische Grundlagen

- Innerhalb der prosodischen Phrasen werden Prominenzabstufungen vorgenommen durch
 - Druckakzent
 - temporalen Akzent
 - tonalen Akzent
 einzeln sowie in wechselnder Kombination.
- Mit den tonalen Akzentstellen werden paradigmatische holistische Melodiemuster verbunden. Sie lassen sich in vier Basisgruppen einordnen:
 - (steigend-)fallende Konturen = Gipfelkonturen
 - (fallend-)steigende Konturen = Talkonturen
 - (steigend-)fallend-steigende = kombinierte Konturen
 - ebene Konturen.
- Die Gipfel- und die kombinierten Konturen können in drei, die Talkonturen in zwei Stufen mit der Ansatzrohrartikulation synchronisiert werden: früh - mittel - spät; früh - spät.
- Die Tal- und die kombinierten Konturen können mittelhoch oder hoch ansteigen.
- Gipfel- und kombinierte Konturen können leicht oder bis in den tiefen Stimmbereich fallen. Als Grenzfall einer Gipfelkontur ohne Abstieg wird die in einem Plateau auslaufende Kontur klassifiziert, wenn der Anstieg schnell ist, im Gegensatz zu einem langsamer steigenden Tal.
- Die Basismuster und ihre Modifikationen fokussieren Höhen und Tiefen in akzentuierten Vokalen in unterschiedlicher Weise. Steigende, vor allem hoch steigende betonen die Höhen und signalisieren dem Hörer Offenheit des Sprechers, also z.B. die nichts präjudizierende Bitte um eine ja-nein-Entscheidung auf eine Frage (*Frequency Code*). Am anderen Ende der melodischen Skala liegt der frühe Gipfel, der am stärksten die Tiefen fokussiert und dem Hörer Dominanz signalisiert, z.B. eine abschließende Schlußfolgerung, eine feststehende Tatsache oder unabänderliche Sachlage. Nicht-frühe Gipfel betonen unter den Gipfelkonturen stärker die Höhen und signalisieren den Beginn einer Argumentation bzw. neue Erfahrung, der späte Gipfel tut dies im Gegensatz zur Erwartung oder zum bisher Bekannten/Erfahrenen
- Damit lassen sich direkte Beziehungen zwischen dem Signal (hier dem Grundtonverlauf) und einer allgemeinen Äußerungssemantik etablieren. Die konkrete Bedeutung in der jeweiligen Kommunikationssituation entsteht durch das Zusammenwirken dieser generalisierbaren Komponente auf situativer Äußerungsebene mit der Wort- und Satzsemantik sowie dem situativen, sozialen und kulturellen Vorwissen, in das die Kommunikation eingebettet ist.

3.2.2.2 Fragestellungen und Hypothesen

Die Basismuster und ihre Modifikationen kommen in einakzentigen Phrasen vor und lassen sich nach bestimmten Regeln, die über eine umfangreiche Datenbasis statistisch erfaßt werden können, in zweiakzentigen und mehrakzentigen Phrasen verknüpfen (*Konkatenation*). Bei der Konkatenation entstehen entweder *Sequenzen* von Mustern oder *Integrationen* zu übergeordneten Mustern (z.B. Hutmuster).

Hypothese 1 - semantisch-pragmatische Funktionen in Mustersequenz und Musterintegration

In Mustersequenzen gehen die semantisch-pragmatischen Funktionen der Basiskonturen und ihrer Modifikationen ein; in übergeordneten integrierten Mustern treten neue, zusätzliche Funktionen hinzu.

Hypothese 2 - melodische Kohäsion innerhalb prosodischer Phrasen

Übergeordnete integrierte Muster haben einen höheren Grad der Kohäsion zwischen den Basismustern als einfache Mustersequenzen, z.B. Hutmuster vs. Gipfelsequenzen.

Hypothese 3 - melodische Effekte durch Äußerungsplanung und Dialogsteuerung

Die melodische Gestaltung prosodischer Phrasen wird nicht nur durch Semantik und Pragmatik gesteuert, sondern auch durch die damit interagierenden Ebenen der Äußerungsplanung und Dialogsteuerung.

3.2.2.3 Deskription und Evaluation

- Die letzte Akzentstelle in Phrasen zeigt turnfinal bzw. turnintern folgende Regularitäten:
 - In beiden Positionen sind mittlere Gipfel unter den Gipfelkonturen mit einem Anteil von ca. 50% gleich stark vertreten. Turnfinal kommen aber wesentlich mehr frühe, turnintern wesentlich mehr späte Gipfel vor. Dies ist verbunden mit der größeren Wahrscheinlichkeit einer zusammenfassenden abschließenden Schlußfolgerung am Turnende.
 - Das wird noch weiter dadurch unterstrichen, daß turnfinale Gipfel überwiegend stark fallen, wodurch in Kombination mit der frühen Gipfelposition der Tiefton zusätzlich fokussiert und damit die Funktion des Abschließens verstärkt wird. Turnintern ist der Anteil der leicht fallenden Muster sehr viel größer, und zwar zunehmend vom frühen zum mittleren und schließlich zum späten Gipfel. Der nicht-terminale Abstieg hat eine weiterweisende Funktion; das ist im Fall des frühen Gipfels am wenigsten wahrscheinlich.
 - Der Anteil hoch ansteigender Täler ist turnfinal deutlich höher als turnintern. Talkonturen stehen im Zusammenhang mit Fragen in enger Verbindung zum Turnwechsel.
 - Sowohl stark fallende Gipfelkonturen als auch hoch ansteigende Talkonturen sind also mit dem Turnwechsel gekoppelt, entweder mit einer abschließenden Aussage oder einer Frage vor der Turnübergabe. Die übrigen Konturen haben insbesondere weiterweisende Funktion im Innern eines Sprechbeitrags.

- In der Sequenz von Basismustern in mehrakzentigen Phrasen zeigt sich eine erhöhte Auftretenswahrscheinlichkeit gleicher Konturen in Folge, insbesondere zweier früher Gipfel oder zweier Täler. Damit wird die durch das Muster signalisierte Funktion über der Gesamtphrase verstärkt.

- Es finden sich in der Sequenz von Basismustern aber auch Restriktionen. So ist die Abfolge von spätem und frühem Gipfel relativ selten. Dies kann drei Gründe haben:
 1. Der späte Gipfel hat die Funktion, etwas Neues als Gegensatz zur Erwartung in die Argumentation einzuführen (einen neuen Gedanken, eine neue Erfahrung oder

Beobachtung im expliziten Kontrast zum Bekannten), d.h. er vermittelt auch die Expressivität des Sprechers, u.a. Verwunderung. Im Szenario der nüchternen Terminabsprachen ist diese emotionale Komponente weniger wahrscheinlich. Folglich wären späte Gipfel allein und erst recht in der Sequenz mit frühen seltener als die übrigen Gipfelmuster.

2. In der Sequenz bringt jedes Basismuster seine Funktion an der separat fokussierten Akzentstelle ein, d.h. der späte Gipfel vermittelt das Besondere, kontrastierend Herausgehobene, der frühe Gipfel das unumstößlich Geltende und abschließend Festzuhaltende, woraus schließlich der expressive Gegensatz von Verwunderung und Resignation resultieren kann. Damit ergeben sich semantische Unverträglichkeiten in der Abfolge. Dies gilt insbesondere dann, wenn die auf Lexik und Syntax basierende Semantik mit diesem prosodisch vermittelten Gegensatz nicht konform geht.

Wenn z.B. in "HEUTE ist schönes FRÜHLINGSwetter." der erste Akzent einen späten Gipfel mit der Funktion der Einführung des Besonderen erhält, der zweite Akzent aber einen frühen Gipfel mit der Funktion der unumstößlichen Feststellung, dann paßt das nicht gut zusammen, weil die Wortsemantik von "Frühlingswetter" eine Fortsetzung der positiven Expressivität, z.B. mit einem zweiten späten Gipfel, erwarten läßt, nicht aber ein resignierendes Akzeptieren des Unabänderlichen. Die Situation wäre anders in "HEUTE ist typisches NOVEMBERwetter."

3. Wenn die Anzahl der Silben und der periodische Anteil zwischen den Gipfeln eine bestimmte Grenze unterschreitet, kann sich ein später Gipfel vor einem frühen nicht mehr separat ausprägen, ohne daß eine segmentelle Längung eintritt, die aber ihrerseits zu einer prosodischen Phrasengrenze führen kann. Dann liegt keine phraseninterne Sequenz mehr vor, und folglich wird ihre Häufigkeit im Korpus reduziert.

Weitere Untersuchungen müssen diese Hypothesen evaluieren. Insbesondere die Daten aus dem spontansprachlicheren Szenario der *Lindenstraße* haben hierbei eine große Bedeutung.

- Andererseits können Gipfelkonturen in eine übergeordnete kohäsivere Einheit integriert werden: das Hutmuster, bestehend aus mittlerem/spätem und mittlerem/frühem Gipfel, verbunden durch ein Plateau. Im Gegensatz zur Gipfelsequenz werden die Satzakkente in der Integration nicht einzeln lokal fokussiert, sondern die Gesamtkonfiguration bildet einen breiten Fokus. Dieses globale Muster führt eine neue semantische Funktion ein, die sich den Funktionen der Basiskonturen überlagert. Wird obiges Beispiel mit spätem + frühem Gipfel in einem Hutmuster realisiert, so wird kontrastierend ein neues Thema gesetzt, über das abschließend etwas ausgesagt wird. Die Semantik des frühen Gipfels an zweiter Stelle bezieht sich auf diese Thema-Rhema-Bindung insgesamt, d.h. durch das Hutmuster ergibt sich keine semantische Unverträglichkeit; außerdem gilt im Hutmuster die phonetische Restriktion nicht: die Verbindung kommt häufig vor.
- Ebene Konturen und eben auslaufende Gipfelkonturen am Phrasenende stehen häufig in Verbindung mit Störungen der Äußerungsplanung oder der Sprechflüssigkeit.
- Damit sind alle drei Hypothesen durch die Korpusanalyse bestätigt worden.

3.2.3 Spezialfall melodischer Gestaltung: syntaktische Fragen in semantisch-pragmatischer Fragefunktion

3.2.3.1 Fragestellungen und Hypothesen

Fragestellung 1 - Intonatorische Manifestation syntaktischer Fragetypen

Es gehört zu den unumstößlichen Lehrbuchmeinungen, daß syntaktisch markierte Fragen, z.B im Deutschen oder Englischen, eine feste Zuordnung von fallenden Mustern in Wortfragen und von steigenden Mustern in Verberststellungsfragen (VE-Fragen) haben. Diese Aussage gilt es anhand eines umfangreichen spontansprachlichen Korpus statistisch zu evaluieren.

Hypothese 1 - Häufigkeitsverteilungen von Gipfel- und Talkonturen in Wort- und VE-Fragen
Beide Intonationsmuster kommen in jedem der beiden syntaktischen Fragetypen vor, aber Wortfragen haben häufiger Gipfelkonturen, Verberststellungsfragen häufiger Talkonturen.

Fragestellung 2 - Semantisch-pragmatische Funktion der Intonationsmuster

Die semantisch-pragmatische Funktion einer syntaktischen Frageform in einem Kommunikationskontext bestimmt die Wahl der Intonation. Diese Aussage gilt es durch zwei komplementäre Methoden an spontansprachlichen Daten zu evaluieren:

(a) durch Textinterpretation von Wort- bzw. Verberststellungsfragen mit jeweils fallenden bzw. steigenden Mustern

(b) durch systematische Überführung mittels Sprachsynthese von natürlich produzierten fallenden in steigende Muster und umgekehrt sowie durch anschließende Interpretation der Kontextverträglichkeit und Bedeutungsänderung.

Hypothese 2 - Sprecher- vs Hörerorientierung in Gipfel- vs Talkonturen syntaktischer Fragen

In beiden syntaktischen Frageformen orientiert sich steigende Intonation auf den Befragten und bringt Freundlichkeit, Konzilianz, Offenheit hinsichtlich der zu erwartenden Antwort zum Ausdruck, während fallende Intonation auf Fakten und den Fragenden orientiert ist und Routine, Kategorialität, eingeschränkte Erwartung hinsichtlich der Antwort ausdrückt. Diese funktionale Differenzierung der Intonationsmuster läßt sich mit dem *Frequenzcode* und dessen Assoziation von Tief- bzw. Hochton mit Unterordnung vs. Dominanz verbinden.

Fragestellung 3 - Erklärung der Häufigkeitsverteilungen

Mit den unterschiedlichen syntaktischen Fragetypen sind verschiedene semantisch-pragmatische Funktionen verbunden: Aufforderung zu spezifischer Auskunft vs. Bitte um Entscheidung zwischen einer Ja-Nein-Polarität. Diese können die Verwendung fallender oder steigender Muster präjudizieren. Nur so läßt sich die intuitiv-empirisch begründete Lehrbuchmeinung erklären.

Hypothese 3 - Syntaktisch-prosodische Default-Bindung und kommunikative Überlagerung

Aufgrund ihrer semantisch-pragmatischen Defaultbindungen legen Wortfragen Fakten- und Sprecherbezug und damit Tiefton, Verberststellungsfragen hingegen Partnerbezug und damit Hochton nahe. Der Kommunikationsablauf entscheidet darüber, wie diese Default-Bindung durch eine Abschwächung der Selbst- und eine Erhöhung der Partnerorientierung bzw. umgekehrt restrukturiert wird; entsprechend kann in Wortfragen steigende Intonation zum Ausdruck von Freundlichkeit auftreten und in Verberststellungsfragen fallende Intonation zum Ausdruck einer erwarteten Entscheidungsrichtung.

3.2.3.2 Deskription und Evaluation

- Es findet sich folgende Häufigkeitsverteilung der fallenden (f), hoch steigenden (hs), mittel steigenden (ms) und sonstigen (r) Intonationsmuster in den Wort- bzw. VE-Fragen (in %):

	f	hs	ms	r
Wort	57	10	24	9

- Damit ist Hypothese 1 bestätigt:
 - Fallende und steigende Intonationsmuster kommen in beiden Fragesyntagmen vor.
 - Wortfragen haben überwiegend fallende, VE-Fragen weit überwiegend steigende Muster.
 - Wortfragen haben einen verschwindend kleinen Anteil an hoch steigenden Mustern, während diese in VE-Fragen dominieren.
- Textinterpretation und systematische Intonationsänderung an ausgewählten Beispielen konvergieren in der Aussage, daß in beiden Syntagmen steigende Intonation Freundlichkeit und Offenheit gegenüber dem Befragten zum Ausdruck bringt, während fallende Intonation Kategorialität und das Selbst des Fragenden fokussiert. Hypothese 2 ist damit auch bestätigt.
- Die unterschiedlichen Häufigkeitsverteilungen sowie die Funktionsanalyse unterstreichen die kommunikative Festlegung einer syntaktisch-prosodischen Bindung. Es ist daher sowohl inadäquat, eine feste Bindung zwischen syntaktischer Frageform und einem bestimmten Intonationsmuster anzusetzen als auch mit der Frageform eine definierte kommunikationsunabhängige semantische Funktion zu verbinden. Die syntaktische Form impliziert bestimmte Erwartungshaltungen des Sprechers dem Partner gegenüber, aber darüber lagern sich Einstellungen des Sprechers zu sich selbst, zu den Fakten und zum Partner, die durch den Kommunikationsablauf bestimmt werden.

So stellte Fries (1964) entgegen der Lehrbuchmeinung in einer empirischen Erhebung zu Entscheidungsfragen im Englischen fest, daß sie in 61,7% fallende und nur in 38,3% steigende Intonation hatten. Für die richtige Interpretation dieses statistischen Befundes ist es jedoch wichtig, in welchem Kommunikationsrahmen die untersuchten Daten anfielen. Es handelte sich um amerikanische Fernseh/Rudiosendungen zum Thema des Beruferatens (das später durch Robert Lembke im Deutschen Fernsehen populär gemacht wurde). Darin mußte eine Jury von vier Personen abwechselnd durch Ja-Nein-Fragen den Beruf oder die besondere Tätigkeit von Kandidaten herausbekommen. Diese Art des Fragens zielt darauf, eine ganz bestimmte, vom Sprecher vorkonstruierte Antwort zu erhalten, um die Rateaufgabe zu lösen; sie will es also nicht dem Befragten überlassen, eine freie Entscheidung zu treffen. In diese Kommunikationssituation paßt die fallende, auf den Sprecher orientierte Intonation, die steigende, den Befragten bittende ist weniger wahrscheinlich. Wie Selting (1995) betont die Analyse hier die Bedeutung der Kontextualisierung, konzeptualisiert aber darüber hinaus Prosodie als konstitutives eigenständiges Element der Semantik jenseits von konkreter Aktivität.

3.3 Integration der Teilbereiche des Projekts in eine globale Äußerungsphonetik

Das Projekt hat folgende grundsätzliche Punkte für phonetische Theoriebildung herausgearbeitet.

- Wortphonetik/phonologie bedarf des satz- und äußerungsphonetischen Komplements, um relevante Aussagen über lautsprachliche Kommunikation machen zu können.
- Die arbeitsheuristische Trennung in eine segmentelle und eine prosodische Ebene muß durch die Untersuchung der Wechselwirkungen korrigiert werden:
 - Zum einen hängt die segmentelle satzphonetische Ausprägung von den prosodischen Rahmenbedingungen ab.
 - Zum andern sind segmentelle Ausprägungen Teil des multifaktoriellen Merkmalsbündels prosodischer Phrasengrenzen.

- Prosodische Phonologie kann sich nicht darin erschöpfen, distinktive Lautmuster aufzustellen, sondern muß stets auch die Frage nach den damit verbundenen (linguistischen und paralinguistischen) Funktionen untersuchen. Insbesondere die funktionale Einbettung von Sprech-beiträgen in den kommunikativen Ablauf ist essentiell für die Interpretation prosodischer Muster.
- Eine grundsätzliche Funktion der Prosodie ist die syntagmatische Strukturierung des Gesprochenen entsprechend der inhaltlichen Gliederung. Prosodie interagiert hierbei mit syntaktischen Strukturen, mit denen sie konvergiert oder konkurriert.
 - Prosodische Phrasengrenzen sind signalphonetisch multifaktoriell kodiert.
 - Unterschiedliche Merkmalsbündelungen bilden eine Hierarchie der Kohäsions-/Trennstärke zwischen prosodischen Phrasen.
 - Diese Kohäsions-/Trennungsskalierung subsumiert auch unflüssige Phrasierung infolge von Problemen der Sprechplanung.
- In den prosodischen Phrasen werden Akzentstellen durch holistische melodische Muster markiert, die in regelhaften Konkatenationstypen auftreten, z. T. übergeordnete intonatorische Einheiten (z.B. Hutmuster) bilden und in allen Fällen zusammen mit anderen formalen Mitteln (Syntax, Wortwahl, segmentelle Phonetik) die Äußerungssemantik im Kommunikationsablauf zwischen Sprecher und Hörer abbilden.
- Dabei ist das allgemeine semantisch-pragmatische Prinzip sprachlicher Interaktion, der *Frequenzcode*, wirksam, dessen Manifestation mit der prosodischen Ausprägung der jeweiligen Kommunikationsaktivität interagiert. Dieses Prinzip zeigt sich in der funktionalen Differenzierung der Gipfelsynchronisierung ebenso wie in der Prosodie von Fragetypen. Diese bilden ein spezifisches Zusammenspiel von Satztyp und Prosodie sowohl per se als auch eingebettet in den Kommunikationsablauf.

3.4 Anwendung und Folgeuntersuchungen

Eine integrierte Äußerungsphonetik, wie sie in diesem Projekt für einen breites Spektrum segmenteller und prosodischer Phänomene erarbeitet wurde, ist von großer Bedeutung für die Verbesserung der automatischen Sprachsynthese hinsichtlich Verständlichkeit und vor allem Natürlichkeit (Kohler 2003b). Folgeuntersuchungen können sich in drei Richtungen anschließen:

- (1) Erweiterung der segmentellen Äußerungsphonetik auf andere Sprachen unter sprachindividuellen, -typologischen und -universellen Gesichtspunkten (siehe Kohler 2001b, 2002)
- (2) Elaborierung des Konzepts der prosodischen Kohäsion zwischen sprachlichen Phrasierungseinheiten: hierarchische Skalierung phonetischer Merkmalsbündel zur syntagmatischen Strukturierung im Gespräch (Peters, Kohler und Wesener 2002)
- (3) verstärkte Einbeziehung des Hörers in die prosodische Analyse: Perzeptionsuntersuchungen
 - zur Interaktion von Gipfelsynchronisierung und Gipfelgestalt (Kohler 2003c, Niebuhr 2003a,b) mit Ausdehnung auf andere Sprachen (Khromovskikh 2003), Entsprechendes für Talkonturen (Landgraf 2003)
 - zur prosodischen Phrasierung sowie zu Druck- vs. Tonakzent (siehe (2), Kohler 2003c)
 - zur Funktion der Prosodie, u.a. durch semantisches Differential (Dombrowski 2003).

4 Literatur

(a) Verweise

- Altmann, H. (ed.): *Intonationsforschungen*. Tübingen: Niemeyer (1988).
- Beckman, M. E., and G. M. Ayers: *Guidelines for ToBI Labelling*. Version 3, March 1997 (http://ling.ohio-state.edu/research/phonetics/E_ToBI).
- Beckman, M. E., and J. B. Pierrehumbert: Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255-309 (1986).
- Browman, C. P. and L. Goldstein: Articulatory Phonology, *Phonetica* 49, 155 - 180 (1992).
- Grice, M., und S. Baumann: Deutsche Intonation und GToBI. *Linguistische Berichte* 181, 1-33(2000).
- Fries, C. C.: On the intonation of 'Yes-No' questions in English. In: D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott , J. L. M. Trim (eds), *In Honour of Daniel Jones*. London: Longmans, 242-254 (1964).
- IPDS: *The Kiel Corpus of Read Speech*. Volume 1. CD-ROM#1. Kiel: Institut für Phonetik und digitale Sprachverarbeitung (1994).
- IPDS: *The Kiel Corpus of Spontaneous Speech*. Volumes 1-3. CD-ROM#2-4. Kiel: Institut für Phonetik und digitale Sprachverarbeitung (1995-1997).
- IPDS: *xassp User's Manual* (Advanced Speech Signal Processor under the X Window System). In: A. P. Simpson, K. J. K. Kohler and T. Rettstadt (eds.), *The Kiel Corpus of Read/Spontaneous Speech – Acoustic database, processing tools and analysis results*. AIPUK 32, 31-115. Kiel: Institut für Phonetik und digitale Sprachverarbeitung (1997).
- Kohler, K. J. (ed): *Studies in German Intonation*. AIPUK 25 (1991).
- Kohler, K. J.: Modelling prosody in spontaneous speech. In: Y. Sagisaka, N. Campbell, N. Higuchi (ed.), *Computing Prosody*, Berlin/Heidelberg/New York/Tokyo: Springer, 187-210 (1997). URL <http://www.ipds.uni-kiel.de/forschung/kim.de.html>
- Ohala, J.: An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41, 1-16 (1984).
- Pierrehumbert, J. B.: *The Phonology and Phonetics of English Intonation*. Bloomington: IULC (1980).
- Pike, K.: *The Intonation of American English*. Ann Arbor: University of Michigan Press (1945).
- Selting, M.: *Prosodie im Gespräch*. Tübingen: Niemeyer (1995).

(b) Veröffentlichungen, Tagungsbeiträge und Manuskripte im Rahmen des Projekts

- Dombrowski, E.: Semantic features of accent contours. Effects of F0 peak position and F0 time shape. In: *Proc. 15th ICPhS, Barcelona*, 1217-1220 (2003).
- Kohler, K. J.: The phonetic manifestation of words in spontaneous speech. In: D. Duez (ed), *Proc. from the ESCA Workshop on Sound Patterns of Spontaneous Speech*, La Baume-les-Aix, 13-22 (1998).
- Kohler, K. J.: Articulatory prosodies in German reduced speech. In: *Proc. 14th ICPhS, San Francisco*, 89-92 (1999a).
- Kohler, K. J.: Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German? Poster *ICPhS99 satellite meeting on non-modal vocal-fold vibration and voice-quality*, San Francisco (1999b). URL http://www.ipds.uni-kiel.de/pub_exx/kongrbrtr/plosglot.html
- Kohler, K. J. (ed): *Phrase-level Phonetics and Phonology of German*. AIPUK 34 (1999c).
- Kohler, K. J.: Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German? *AIPUK* 34, 285-325 (1999d).
- Kohler, K. J.: Linguistic and paralinguistic functions of non-modal voice in connected speech. In: *Proc. 5th Seminar on Speech Production. Kloster Seeon*, 121-124 (2000).
- Kohler, K. J. (ed): *Sound Patterns in German Read and Spontaneous Speech - Symbolic structures and gestural dynamics*. AIPUK 35 (2001a).

- Kohler, K. J.: The investigation of connected speech processes. Theory, method, hypotheses and empirical data. *AIPUK* 35, 1-32 (2001b).
- Kohler, K. J.: Variability of opening and closing gestures in speech communication. *AIPUK* 35, 33-96 (2001c).
- Kohler, K. J. : Articulatory dynamics of vowels and consonants in speech communication. *JIPA* 31, 1-16 (2001d).
- Kohler, K. J.: Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German? In: N. Grønnum and J. Rischel (eds). *To Honour Eli Fischer Jørgensen. Festschrift on the occasion of her 90th birthday*, Vol. 31 of *Travaux du Cercle Linguistique de Copenhagen*, 174-211. Copenhagen: C. A. Reitzel (2001e).
- Kohler, K. J.: Phrase-level sound structures in French. In: W. J. Barry and M. Pützer, *Festschrift für Max Mangold zum 80. Geburtstag*, *Phonus* 6, 129-157 (2002).
- Kohler, K. J.: Domains of temporal control in speech and language. From utterance to segment. In: *Proc. 15th ICPHS, Barcelona*, 7-10 (2003a).
- Kohler, K. J.: Modelling stylistic variation of speech. Basic research and speech technology application. In: *Proc. 15th ICPHS, Barcelona*, 223-226 (2003b).
- Kohler, K. J.: Neglected categories in the modelling of prosody. Pitch timing and non-pitch accents. In: *Proc. 15th ICPHS, Barcelona*, 2925-2928 (2003c).
- Kohler, K. J. and J. Rodgers: Schwa deletion in German read and spontaneous speech. *AIPUK* 35, 97-123 (2001).
- Kohler, K.J. and A. P. Simpson: *Patterns of Speech Sounds in Unscripted Communication*. Papers from an International Symposium funded by DFG, Sankelmark. *JIPA* 31 (2001).
- Kohler, K. J., B. Peters and T. Wesener: Interruption glottalization in German spontaneous speech. *Proc. Disfluency in Spontaneous Speech, DISS01*. Edinburgh, 45-48 (2001).
- Niebuhr, O.: Perceptual study of timing variables in F0 peaks. In: *Proc. 15th ICPHS, Barcelona*, 1225-1228 (2003a).
- Peters, B.: Individuelle und geschlechtsspezifische Unterschiede in der prosodischen Gestaltung deutscher Lese- und Spontansprache. In: H. D. Schlosser (ed), *Sprache und Kultur*, Forum Angewandte Linguistik 38, 153-162. Frankfurt a. M.: Lang (2000a).
- Peters, B.: 'Video Task' oder 'Daily Soap Szenario'. Ein neues Verfahren zur kontrollierten Elizitation von Spontansprache.
URL http://www.ipds.uni-kiel.de/pub.exx/bp2001_1/Linda21.html (2000b).
- Peters, B.: Trainingsmaterialien zur prosodischen Etikettierung mit dem Kieler Intonationsmodell KIM. Manuskript, IPDS Kiel (2001).
- Peters, B.: Multiple cues for phonetic phrase boundaries in German spontaneous speech. In: *Proc. 15th ICPHS, Barcelona*, 1795-1798 (2003).
- Peters, B., K. J. Kohler und T. Wesener: Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. Manuskript, IPDS Kiel (2002).
- Peters, B., K. J. Kohler und T. Wesener: Melodische Satzakkentmuster in prosodischen Phrasen deutscher Spontansprache. Statistische Verteilung und sprachliche Funktion. Manuskript, IPDS Kiel (2003).
- Rodgers, J. E. J.: Glottalisierung initialer Vokale in Funktions- und Inhaltswörtern in Abhängigkeit von der Akzentuierung. Vortrag auf der GAL-Tagung, Frankfurt (1999a).
- Rodgers, J. E. J.: Three influences on glottalization in read and spontaneous speech. *AIPUK* 34, 177-284 (1999b).
- Rodgers, J. E. J.: Word-initial vowel-onset phenomena in read and spontaneous speech in German. Poster ICPHS99 satellite meeting on non-modal vocal-fold vibration and voice-quality, San Francisco (1999c).
- Rodgers, J. E. J.: The phonatory correlates of juncture in German. Paper at Meeting of the British Association of Academic Phoneticians (BAAP), Glasgow (2000a).
- Rodgers, J. E. J.: The phonatory correlates of juncture in German. *Proc. 5th Seminar on Speech Production, Kloster Seeon*, 289-292 (2000b).

- Wesener, T.: The phonetics of function words in German spontaneous speech. *AIPUK* 34, 327-377 (1999a).
- Wesener, T.: Production strategies in German spontaneous speech: definite and indefinite articles. *Proc 14th ICPHS, San Francisco*, 687-690 (1999b).
- Wesener, T.: Some non-sequential phenomena in German function words. *JIPA* 31, 17-27 (2000).
- Wesener, T.: Aspects of 'cliticization' and style in German function words. *AIPUK* 35, 125-153 (2001).

(c) Magisterarbeiten im Zusammenhang mit dem Projekt

- Khromovskikh, T.: *Perzeptorische Untersuchungen zur Intonation der Frage im Russischen*. Magisterarbeit Kiel (2003).
- Landgraf, K.: *Steigende Intonationskonturen im Deutschen. Experimentalphonetische Untersuchungen zur auditiven Kategorisierung*. Magisterarbeit Kiel (2003).
- Niebuhr, O.: *Perzeptorische Untersuchungen zu Zeitvariablen in Grundfrequenzgipfeln*. Magisterarbeit Kiel (2003b).
- Peters, B.: *Prototypische Intonationsmuster in deutscher Lese- und Spontansprache*. Magisterarbeit Kiel (1998) (In: *AIPUK* 34, 1-177 (1999)).