

Phonetische Intonationsmodelle und die Parametrisierung von kontrastiven Satzakkzenten im Deutschen

Rolf Siepmann
Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3
D- 80799 München

Abstract

Title: Phonetic Models of Intonation and the Parametrisation of Contrastive Sentence Accents in German.

Phonetic models of intonation postulate different properties of the f_0 contour as acoustical correlates of intonational meanings. This dissertation compares five phonetic models of intonation with respect to their analysis of contrastive sentence accents in German. The models are the Kiel intonation model KIM, the straight line approach of the Eindhoven Institute for Perception Research, the Fujisaki superposition model, the Tilt model and finally the model assumed by the recently introduced international transcription system for intonation INTSINT. For the sake of optimal comparison, the production of contrastive sentence accents was conducted in a controlled experimental setting. The accent realizations were partly checked in a perception experiment. Moreover, the contrastively accentuated utterances of the production experiment are produced out of context as so-called 'out-of-the-blue', neutral utterances from the test persons, so that a corpus of 288 contrastively accentuated and 288 corresponding neutral utterances has been created.

The accent realizations are first described acoustically using a set of 13 different parameters, which constitute the so-called MONA model. Secondly, the accent realizations are modelled in terms of the five intonation models. The different model parameters are used as predictor variables in five different Linear Discriminant Analyse (LDA) in order to compare the classification results and concomitantly the five models themselves. Finally, a LDA has been applied to the descriptive parameters of the MONA model. To sum up the results of the different LDAs, the recognition rates classifying the contrastively accentuated and the neutral utterances correctly were not very high for all six models. Therefore the descriptive power of the investigated intonation models seems to be more or less inadequate from a quantitative point of view.

Die vorliegende Arbeit ist eine leicht überarbeitete Fassung meiner Dissertation, die im Juli 2001 von der Ludwig-Maximilians-Universität in München am Institut für Phonetik und Sprachliche Kommunikation angenommen wurde.

Danksagung

Mein Dank gilt an erster Stelle Herrn Prof. Dr. Tillmann, dem Betreuer meiner Dissertation. Er verschaffte mir nicht nur gute Arbeitsbedingungen, sondern vermochte es vor allem, entstehende Knoten bei meiner Dissertation stets konstruktiv aufzulösen. Von unschätzbarem Wert war auch meine von Herrn Prof. Dr. Richter sowie gleichermaßen von Herrn Dr. Batliner er-fahrene Unterstützung. Herr Prof. Dr. Richter weckte und förderte mit großem fachlichen Überblick mein Interesse an der vorliegenden Thematik; für den weiteren Verlauf der Dissertation war die von Herrn Dr. Batliner er-fahrene Unterstützung von wesentlicher Bedeutung. Ganz besonders habe ich in diesem Zusammenhang auch von meinem bei Herrn Prof. Dr. Lieb erworbenen linguistischen Wissen profitiert.

Sehr gerne danken möchte ich auch meinen ehemaligen Kollegen am Institut für Phonetik und Sprachliche Kommunikation der LMU München für ihre vielfältige Unterstützung; zu nennen sind vor allem Daniela Oppermann und Karl Weilhammer sowie Dr. Christoph Draxler, Anja Geumann, Dr. Phil Hoole, Klaus Jaensch, Hartmut Pfitzinger, Felix Schaeffler, Daniel Sonntag, Uli Türk und Andreas Zierdt. Meinem Projektleiter in Verbmobil und SmartKom Dr. habil. Florian Schiel danke ich besonders für die verständnisvolle Ermöglichung meiner parallelen Projekt- und Dissertationsarbeit.

Christiane Hofbauer und Dirk Heckmann danke ich herzlich für ihr Korrekturlesen. Ganz spezieller Dank gilt nicht zuletzt meiner Frau Michaela für ihre vielfältige Unterstützung; ihr aktives Verständnis nicht nur für meine mathematischen Wissenslücken hat sehr zu dem Gelingen der Arbeit beigetragen. Meiner Schwester Bruni schließlich danke ich dafür, dass ihr Glauben an diese Dissertation meinen stets überflügelte.

Inhaltsverzeichnis

1. Einleitung.....	8
Theoretischer Teil.....	10
2. Aspekte der Prosodieforschung.....	10
2.1 Phänomene der Prosodie.....	10
2.2 Prosodische Transkriptionssysteme.....	12
2.3 Anwendungsgebiete der Prosodie.....	13
3. Mikro- vs. Makroprosodie.....	15
3.1 Mikroprosodische Einflüsse.....	15
3.2 Elemente der Makroprosodie.....	16
3.2.1 Globaler Tonhöhenverlauf.....	17
3.2.2 Lokaler Tonhöhenverlauf.....	18
4. Das phonologische Tonsequenzmodell.....	20
4.1 Metrische Gitter im Deutschen.....	20
4.2 Aufbau der Intonationsphrase.....	22
4.2.1 Der Nuklearakzent	23
4.2.2 Pränukleare Töne.....	23
4.3 Phonetische Abbildungsregeln.....	24
4.4 Kritik des Tonsequenzmodells.....	25
5. Funktionen von Satzakkzenten.....	27
5.1 Fokusakzente.....	28
5.2 Der Brückenakzent.....	30
5.2.1 Kontrastbildungen.....	30
5.2.2 Topikalisierungen.....	32
6. Phonetische Intonationsmodelle.....	33
6.1 Das Kieler Intonationsmodell.....	33
6.1.1 Gipfel- und Talkonturen.....	33
6.1.2 Kombination von Gipfelkonturen.....	35
6.2 Das IPO-Modell.....	36
6.2.1 Melodiekonturen des Deutschen.....	37
6.2.2 Automatische Stilisierung.....	40
6.3 Das Fujisaki-Modell.....	41
6.3.1 Der Steuermechanismus.....	41
6.3.2 Anwendung des Modells.....	42
6.3.3 Erweiterung des Steuermechanismus.....	44

6.4 Das Tilt-Modell	46
6.5 Das INTSINT-Modell.....	47
6.6 Weitere Modelle.....	49
6.7 Diskussion der Modelle.....	50
Experimenteller Teil.....	52
7. Gewinnung von Satzaktrealisierungen.....	52
7.1 Methodische Vorüberlegungen.....	52
7.2 Produktionsexperiment.....	54
7.2.1 Aufbau und Durchführung.....	55
7.2.2 Ergebnisse.....	57
7.2.2.1 Halbtonwerte.....	59
7.2.2.2 Intensitätswerte.....	65
7.2.2.3 Zeitwerte.....	69
7.2.3 Diskussion des Produktionsexperiments.....	72
7.3 Perzeptionsexperiment.....	73
7.3.1 Aufbau und Durchführung.....	73
7.3.2 Ergebnisse.....	76
7.3.2.1 Teilnehmer.....	76
7.3.2.2 Abgegebene Urteile.....	77
7.3.3 Diskussion des Perzeptionsexperiments.....	79
7.3.4 INTSINT-Synthese.....	79
8. Modellierung der Satzaktrealisierungen.	81
8.1 KIM-Modellierung.....	81
8.1.1 Vorgehen.....	81
8.1.2 Ergebnis.....	83
8.2 IPO-Modellierung.....	85
8.2.1 Vorgehen.....	85
8.2.2 Ergebnis.....	86
8.3 Fujisaki-Modellierung.....	88
8.3.1 Vorgehen.....	88
8.3.2 Ergebnis.....	89
8.4 Tilt-Modellierung.....	90
8.4.1 Vorgehen.....	90
8.4.2 Ergebnis.....	91
8.5 INTSINT-Modellierung.....	91
8.5.1 Vorgehen.....	91
8.5.2 Ergebnis.....	92
8.6 Diskussion der Modellierungen.....	94

9. Multivariate Analyse der Satzakzentmodellierungen.....	95
9.1 Unterscheidung nach Satztypen.....	97
9.2 Unterscheidung nach Sprechern	99
9.3 Diskussion der multivariaten Analysen.....	101
10. Abschließende Diskussion.....	103
Anhang A. Dialogstimuli.....	105
Anhang B. Gewichtungsfunktion.....	108
Anhang C. Verwendete Software.....	110
Anhang D. Literatur.....	111

1. EINLEITUNG

Die vorliegende Arbeit ist eine experimentelle Untersuchung von phonetischen Intonationsmodellen des Deutschen. Die Untersuchung beschränkt sich auf die Parameter der Intonationsmodelle, die die phonetischen Formen von ausschließlich kontrastiven Satzakkzenten beschreiben. Mit der Verwendung einer derart konstanten Satzakkzentfunktion ist ein optimaler Vergleich der unterschiedlichen Parametrisierungen der korrespondierenden lautlichen Formen möglich. Die Untersuchung der phonetischen Parametrisierungen berücksichtigt dabei auch den Einfluss von syntaktischen Bedingungen auf die Intonation. Damit folgt die vorliegende Arbeit prinzipiell einem Desiderat der Sprachtechnologie, phonetisch-akustische und linguistisch-syntaktische Modelle vor dem Hintergrund konkreter Anwendungen direkt aufeinander abzubilden.

Ein Satz wird üblicherweise als eine syntaktische Einheit aufgefasst, die mit einer bestimmten Bedeutung verbunden ist. Als abstrakte Einheiten der Grammatik einer Sprache stellen Sätze potentielle Äußerungen dar. Die intonatorische Form eines Satzes besteht nach traditioneller Auffassung aus den funktional relevanten lautlichen Eigenschaften des Satzes. Bei den Funktionen handelt es sich im allgemeinen um die Phrasierung, die Akzentuierung, die Kennzeichnung der Satzart und die Disambiguierung. Satzakkzente werden in dieser Arbeit der Tradition weiterhin folgend als lautliche Hervorhebungen in geäußerten Sätzen aufgefasst, die entweder direkt die propositionale Bedeutung der Sätze oder indirekt deren kontextuelle Verwendungsmöglichkeiten beeinflussen. Kontrastive Satzakkzente drücken dabei einen bestimmten Gegensatz innerhalb des Satzes oder gegenüber dem Kontext aus. Die vorliegende Arbeit behandelt kontrastive Satzakkzente, die einen kontextuellen Gegensatz ausdrücken.

Die Intonation bildet lautlich gesehen einen integralen Bestandteil der prosodischen Form einer sprachlichen Äußerung. Die prosodische Form besteht in erster Näherung aus der melodisch-rhythmischen Gestalt der Äußerung. Nach traditioneller Auffassung manifestiert sich die Intonation auditiv primär in wahrgenommenen Tonhöhenverläufen und sekundär in bestimmten Lautheits- und Dauereigenschaften sprachlicher Einheiten. In der Akustik korreliert die Tonhöhe dabei mit der Grundfrequenz (F_0) und die Lautheit mit der Intensität. Kapitel 2 gibt einen allgemeinen Überblick über verschiedene Aspekte der Prosodieforschung. In diesem Zusammenhang werden auch einzelne Formebenen der Prosodie unterschieden; Kapitel 3 behandelt die mikro- und die makroprosodische Formebene im Detail.

Die vorliegende Arbeit unterscheidet phonetische und phonologische Intonationsmodelle. Phonetische Modelle beinhalten Aussagen über die konkreten lautlichen Formen der Intonation, während phonologische Modelle sprachsystematische Aussagen über die distinktiven Eigenschaften der lautlichen Formen treffen. Eine Überführung der phonologischen in phonetische Aussagen ist mit der Formulierung von Regeln verbunden, die die abstrakten phonologischen Formen auf die konkrete phonetische Ebene abbilden. Kapitel 4 diskutiert das phonologische Tonsequenzmodell, das nicht nur für das Deutsche dominante - phonologische - Intonationsmodell.

Das Tonsequenzmodell wird in dem experimentellen Teil der vorliegenden Arbeit nicht untersucht. Das Ziel dieser Arbeit ist die Untersuchung von sprachakustisch gesehen signahohen, phonetischen Intonationsmodellen, mit denen ein phonologisches Modell nicht direkt vergleichbar ist (s.a. Kap. 4.4). Dennoch berücksichtigt der experimentelle Teil mit dem phonetischen Tilt-Modell eine Beschreibung der Intonation, die einerseits bestimmte theoretische Annahmen des

Tonsequenzmodells teilt und andererseits als dessen phonetische Spezifikation angesehen werden kann.

In Kapitel 5 werden zunächst verschiedene Funktionen von Satzakkzenten vorgestellt, die phonologisch im Rahmen des Tonsequenzmodells analysiert sind. Kapitel 6 stellt die in dieser Arbeit experimentell untersuchten phonetischen Intonationsmodelle einschließlich von zwei weiteren, in dieser Arbeit aber nicht untersuchten Modellen vor. Das erste der beiden nicht untersuchten Modelle beschreibt Grundfrequenzverläufe als Folge von bestimmten F0-Maxima, wohingegen das zweite Modell bei der entsprechenden Beschreibung eine komplexe Parametrisierungsfunktion ansetzt. Bei den untersuchten Modellen handelt es sich zum einen um die bekannten Modelle des Kieler Intonationsmodells, um die Anwendung auf das Deutsche des IPO-Modells (Modell des 'Instituut voor Perceptie Onderzoek' in Eindhoven¹) und um die entsprechende Anwendung des Fujisaki-Modells. An dieser Stelle sei bereits auf die nicht unproblematische Klassifikation des Kieler Intonationsmodells als phonetisches Intonationsmodell hingewiesen; genaueres dazu in Kapitel 6.7. Zum anderen werden in dem experimentellen Teil dieser Arbeit zwei bisher nicht am Deutschen überprüfte Modelle untersucht, das Tilt- und das INTSINT-Modell (INTSINT = International Transcription System for Intonation).

Kapitel 7 und 8 bilden den experimentellen Teil der vorliegenden Arbeit, in dem Realisierungen von kontrastiven Satzakkzenten experimentell gezielt gewonnen und im Rahmen der verschiedenen phonetischen Intonationsmodelle analysiert werden. Die Experimente elizitieren in verschiedenen Satztypen an variierenden syntaktischen Positionen stets zwei Satzakkzente, so dass intonatorisch komplexe Äußerungen entstehen, deren unterschiedliche Parametrisierung nach den genannten Intonationsmodellen phonetisch wie linguistisch gleichermaßen von besonderem Interesse ist.

In Kapitel 9 wird für jedes phonetische Intonationsmodell eine multivariate Diskriminanzanalyse durchgeführt. Den Analysen liegen die berechneten Parameter der einzelnen Intonationsmodelle für die experimentell elizitierten kontrastiven Satzakkzente sowie solche Parameter des jeweiligen Intonationsmodells zugrunde, die aus nicht gleichermaßen kontrastiv akzentuierten Varianten der experimentell gewonnenen Äußerungen berechnet wurden. Die unterschiedlichen Trennungsgüten der multivariaten Analysen der Intonationsmodelle bzgl. der kontrastiv akzentuierten und der korrespondierenden nicht kontrastiv akzentuierten Äußerungen ermöglichen einen Vergleich der in der vorliegenden Arbeit untersuchten phonetischen Intonationsmodelle.

¹) Mittlerweile das „Center for User-System Interaction.“

THEORETISCHER TEIL

2. ASPEKTE DER PROSODIEFORSCHUNG

Die Prosodie im Sinne von Sprechmelodie und -rhythmus ist ein aktuelles Forschungsgebiet, an dem im wesentlichen die Phonetik, die Linguistik, die automatische Sprachverarbeitung und die Psychologie einschließlich ihrer klinischen Bereiche beteiligt sind². Ein zusätzliches anwendungsorientiertes Forschungsgebiet bildet die Sprachdidaktik. Die empirische Prosodieforschung arbeitet mit dem klassischen Laborexperiment und seit einigen Jahren zunehmend auch mit großen, nach Möglichkeit spontansprachlichen Korpora. Mit den Korpora soll die Variabilität der Realisation von sprachlichen Einheiten repräsentativ erfasst werden. Die beiden Methoden sind grundsätzlich komplementär in dem Sinne, dass experimentell gezielt gewonnene Erkenntnisse an den genannten Korpora überprüft werden müssen. Andererseits sollten zumindest grundsätzlich Laborexperimente auch Korpusanalysen validieren (vgl. Kap. 7.1, S. 52f).

2.1 Phänomene der Prosodie

Die Prosodie manifestiert sich nach traditioneller Auffassung auditiv in dynamischen Veränderungen primär der Tonhöhe, deren Richtung, Umfang, Variationsweite und -geschwindigkeit sowie in deren äußerungs- oder sprecherbezogenem Niveau. Weitere lautliche Merkmale der Prosodie bilden Veränderungen der Lautheit und Länge von sprachlichen Einheiten. Es können auch prosodisch relevante Veränderungen der Klangfarbe und der Stimmqualität auftreten. Ein konstitutives Element der Prosodie ist zudem die Pausenstruktur einer Äußerung. Akustisch korreliert die Tonhöhe mit der Grundfrequenz (F0) und die Lautheit mit der Intensität; die beiden akustischen Parameter können physiologisch als Frequenz der Stimmbandschwingung und als Stärke des subglottalen Luftdrucks beschrieben werden.

Prosodische Formen lassen sich nach Tillmann/Mansell (1980: 108ff) auf drei interagierenden Ebenen phonetisch analysieren. Die unterste Ebene der C-Prosodie besteht aus den phonematischen Einheiten einer sprachlichen Äußerung und die mittlere Ebene der B-Prosodie aus den lautübergreifenden suprasegmentellen, silbischen Einheiten. Silben bilden in der Phonologie die kleinste prosodische Einheit, die mit distinktiven binären Merkmalen, bezogen auf eine Wortform, beispielsweise als betont oder unbetont beschrieben werden³. In der einschlägigen Literatur wird die C-Prosodie auch als Mikro- und die B-Prosodie als Makroprosodie bezeichnet (vgl. Kap. 3, S. 15ff). Die oberste Ebene der A-Prosodie ist der zumeist mehrere Silben umfassende Intonationsverlauf einer Äußerung.

Wie bereits erwähnt, hat die Intonation nach traditioneller Auffassung die Funktionen der Akzentuierung, der Disambiguierung, der Phrasierung und der Satzartdifferenzierung. Diese katego-

²) Mit der Interdisziplinarität entstehende Kommunikationsprobleme kommentieren Werner/Keller (1994: 37f) aus wohl technischer Sicht: „Writings in this field are full of terminological fuzziness, and work authored by persons with linguistic training is often totally esoteric to outsiders.“

³) Auer (1991) unternimmt den Versuch, die More als die kleinste prosodische Beschreibungseinheit für das Deutsche einzuführen.

riale linguistische Informationen sind von den eher graduellen paralinguistischen Informationen der Prosodie zu unterscheiden, die Aufschluss geben beispielsweise über den emotionalen Zustand des Sprechers und/oder über seine diskursbezogenen Einstellungen wie Hoffnungen, Wünsche oder Befürchtungen. Im Grenzbereich zwischen kategorialer und gradueller Information liegt der intonatorische Ausdruck etwa von Entschiedenheit oder Beiläufigkeit, die in der Linguistik mitunter als konnotative Bedeutungen rubriziert werden. Das in Kapitel 6.1, S. 33ff, behandelte Kieler Intonationsmodell setzt derartige Bedeutungen impressionistisch als Bedeutungen der Intonation an.

Satzakzente bilden ein zentrales Mittel der Kennzeichnung der Informationsstruktur eines Satzes, die etwa die Unterscheidung von hervorgehobener, wichtiger gegenüber nicht hervorgehobener, unwichtiger Information beinhaltet⁴. Eine weitere zentrale Funktion ist die Markierung kontrastierender Elemente des Diskurses. Die Untersuchung der Informationsstruktur steht seit geraumer Zeit, unter dem Stichwort der Schnittstelle zwischen Phonologie und Syntax, im Mittelpunkt primär der grammatisch orientierten Prosodieforschung. Konversationsanalytische Arbeiten der Prosodieforschung thematisieren dagegen die Rolle der Prosodie bei der Gesprächsorganisation⁵. Die Rolle besteht diesem Ansatz zufolge im wesentlichen in der Konstruktion von Kontextualisierungshinweisen im Diskurs. Ein umfassend untersuchtes Phänomen ist die prosodisch bestimmte Organisation des Sprecherwechsels ('turn taking').

Untersuchungen zur Prosodie müssen verschiedene Faktoren der Variabilität berücksichtigen. Die Variabilität besteht nicht nur intraindividuell bei den einzelnen Sprechern, sondern ist wenigstens noch von regionalen, sozialen und stilistischen Faktoren abhängig, wie z.B. von der stilistischen Unterscheidung freier und gelesener Reden. Auer et al. (2000) stellen ein Forschungsprojekt zur Intonation der Stadtsprachen in Berlin und Hamburg vor; auch Gibbon (1998) enthält einige Hinweise auf regionale Varianten der Intonation des Deutschen. Das Duden Aussprachewörterbuch (1990: 29f) nennt die unten folgenden allgemeinen Merkmale der deutschen Standardlautung. Die Merkmale werden an dieser Stelle genannt, um gleich zu Beginn das in dieser Arbeit untersuchte Standarddeutsch zu charakterisieren; in dem in Kapitel 7.2, S. 54ff, vorgestellten WWW-Experiment dieser Arbeit geht es dabei um die Gewinnung von regional gestreuten Hörerurteilen über kontrastive Satzakzente im Deutschen.

1. „Sie [die Standardlautung, RS] ist eine Gebrauchsnorm, die der Sprechwirklichkeit nahekommt. Sie erhebt jedoch keinen Anspruch darauf, die vielfältigen Schattierungen der gesprochenen Sprache vollständig widerzuspiegeln.
2. Sie ist überregional. Sie enthält keine typisch landschaftlichen Ausspracheformen.
3. Sie ist einheitlich. Varianten (freie Varianten und Phonemvariationen) werden ausgeschaltet oder auf ein Mindestmaß beschränkt.
4. Sie ist schriftnah, d.h. sie wird weitgehend durch das Schriftbild bestimmt.
5. Sie ist deutlich, unterscheidet die Laute einerseits stärker als die Umgangslaute, andererseits schwächer als die zu erhöhter Deutlichkeit neigende Bühnenaussprache.“

⁴) Eine genauere Definition von Satzakzenten ist wenigstens mit zwei Schwierigkeiten verbunden, die in dem vorliegenden Zusammenhang allerdings nicht weiter aufgegriffen werden sollen. So muss einerseits nach Batliner (1994) wichtige Information nicht unbedingt prosodisch markiert sein. Andererseits ist eine rein formbezogene Definition insofern problematisch, dass in einem Satz bzw. in einer Phrase mehr als eine lautliche Hervorhebung im Sinne einer Satzakzentmanifestation auftreten kann.

⁵) Selting (1992) kritisiert die grammatisch orientierten Ansätze der Prosodieforschung.

2.2 Prosodische Transkriptionssysteme

Prosodische Transkriptionen sind abhängig von den angenommenen - sprachspezifischen - Eigenschaften der Prosodie und von den Zielen der Transkription. Dabei unterscheiden sich die (segmentell-)phonetischen und die (suprasegmentell-)prosodischen Transkriptionen grundlegend voneinander. Nach den Ausführungen von Richter (1966a: 36) wird bei der phonetischen Transkription der wahrgenommene Sprachlaut auf Außerperzeptionelles, auf artikulatorische oder akustische Eigenschaften zurückgeführt. Ein Phonemsegment weist demnach trotz einer gewissen physikalischen Dauer eine punktuelle erlebnismäßige Präsenz auf. Die prosodische Transkription erfasst im Gegensatz dazu Phänomene, von denen die zeitliche Organisation ein wesentliches Merkmal ist. Nach Richter (1966a: 44) „muß [bei der prosodischen Transkription, RS] auf die Möglichkeit einer unmittelbar-evidenten hypothetischen Entschlüsselung physikalischer Substrateigenschaften verzichtet werden“, um eine Sprachmelodie „als perzeptionelles Nachrichtenobjekt fassen“ zu können.

Richter (1966b) stellt Vorschläge zur prosodischen Transkription vor, nach denen im wesentlichen die folgenden Merkmale verschriftet werden sollen:

- (a) die globale sprachmelodische Form von Äußerungen und deren suprasegmentelle Gliederung,
- (b) die Akzentverteilung,
- (c) die Schwere der Betonung von Silben,
- (d) die Betontheit von Aussprüchen und
- (e) die die Sprachmelodie von Äußerungen konstituierenden internen Tonhöhenbewegungen und Tonlagen, einschließlich bestimmter Dauerverhältnisse.

Der informelle Standard der prosodischen Transkription ist aktuell das von Silverman et al. (1992) für das amerikanische Englisch entwickelte System ToBI (Tones and Break Indices). ToBI wurde u.a. von Reyelt/Batliner (1994) an das Deutsche angepaßt. Die dem System zugrundeliegende Theorie wird in Kapitel 4, S. 20ff, vorgestellt, so dass zumindest die wesentlichen Eigenschaften dieses Verschriftungssystems daraus ableitbar sind. ToBI wird infolgedessen an dieser Stelle nicht weiter behandelt⁶. Als weitere Zeichensysteme zur prosodischen Transkription sollen die folgenden genannt werden:

- (a) LaU (Label- und Testumgebung für melodische Aspekte gesprochener Sprache) von Brindöpke/Schaffranietz (1999); LaU ist das an Spontansprache des Deutschen überprüfte System von Adrians (1991), vgl. Kap. 6.2, S. 36ff
- (b) GAT (Gesprächsanalytisches Transkriptionssystem) von Selting et al. (1998)⁷
- (c) HIAT 2 (Erweiterte Halbinterpretative Arbeitstranskription) von Ehlich/Rehbein (1979)
- (d) INTSINT (International Transcription System for Intonation) von Hirst/di Cristo (1998, 2001)⁸, vgl. Kap. 6.5, S. 47ff

⁶) Rapp (1998) entwickelt ein automatisches Verfahren zur ToBI-Transkription.; vgl. auch Kap. 6.5, S. 47f.

⁷) GAT ist nicht nur für die prosodische Transkription geeignet, sondern stellt ein umfassenderes System zur Kennzeichnung der Interaktion in natürlichen Situationskontexten dar, das auch Elemente wie Gestik und Mimik enthält.

⁸) Hirst (1999) erweitert INTSINT um die Möglichkeit, bestimmte prosodische Dauerverhältnisse zu transkribieren.

- (e) IViE (Intonational Variation in English) von Grabe /Nolan /Farrar (1998)
(f) PROLAB („Prosodic Labeling“) von Kohler (1997), vgl. Kap. 6.1, S. 33ff

In der Praxis wird im allgemeinen mit einem entsprechenden grafischen F₀-Verlauf als visuelles Hilfsmittel prosodisch transkribiert. Reinicke (1995, 1996) und Brindöpke/Schaffranietz (1999: 299ff) beschreiben zwei PC-Arbeitsumgebungen für die prosodische Transkription, die beide die Möglichkeit einer PSOLA-Synthese zur auditiven Rückmeldung der vergebenen Transkriptionszeichen beinhalten. Während die PC-Arbeitsumgebung von Reinicke (1995, 1996) nicht stilisierte F₀-Verläufe als Hilfsmittel vorsieht, bietet die Arbeitsumgebung LaU (Label- und Testumgebung) von Brindöpke/Schaffranietz (1999) die Möglichkeit, berechnete F₀-Verläufe durch lineare Interpolation zu stilisieren.

2.3 Anwendungsgebiete der Prosodie

Ein viel beachtetes Anwendungsgebiet der Prosodieforschung ist die automatische Sprachverarbeitung, die die Prosodie sowohl bei der Erkennung von gesprochener Sprache als auch bei der Sprachsynthese zunehmend implementiert. Batliner et al. (2000a) beschreiben VERBMOBIL als das weltweit erste automatische, Spontansprache verarbeitende Übersetzungssystem für Deutsch, Englisch und Japanisch, das prosodische Information in den Verarbeitungsprozess integriert.

VERBMOBIL arbeitet mit statistischen Modellen, die an einem geeignet erstellten und annotierten Korpus trainiert wurden. Das Korpus enthält auch prosodische Annotationen von bestimmten Grenzen - wie etwa syntaktische und Dialogaktgrenzen - und von verschiedenen Satzakkenten. Außerdem werden Fragen und Nicht-Fragen markiert. Der Verwendung von phonologischen statt statistischen Modellen hat nach Batliner et al. (2000c) in VERBMOBIL nicht zu positiven Ergebnissen geführt: „It is our experience that one always gets better results if one can do without such an [...] [phonological, RS] level, i.e., if one can establish a direct link between (syntactic/semantic) function and phonetic form.“ Eine neuere Entwicklung ist in diesem Zusammenhang die automatische Analyse von prosodisch kodierten Emotionen⁹⁾; vgl. dazu Batliner et al. (2000b).

In der Sprachsynthese ist die Prosodie ein wichtiger Bestandteil der Natürlichkeit der Sprachausgabe, die die Akzeptanz eines automatischen Systems erheblich erhöht. Die wichtigsten Verfahren sind nach Hess (1996) die textgesteuerte ('text to speech') und die inhaltsgesteuerte ('concept to speech') Sprachsynthese. Textgesteuerte Verfahren arbeiten mit vorgefertigten Texten und werden beispielsweise in Vorleseautomaten für Behinderte eingesetzt. Inhaltsgesteuerte Verfahren finden dagegen vorzugsweise in automatischen Dialogsystemen Verwendung, in denen die Synthese mit semantischen Repräsentationen von Dialogabschnitten arbeitet¹⁰⁾. Dabei ist eine Sprachsynthese im Vergleich mit der Spracherkennung schwieriger zu evaluieren. Lemmety (1999) diskutiert aktuelle Sprachsynthesetechniken und -systeme einschließlich deren Bewertungsmöglichkeiten. Zur Bewertung der Prosodie nennt er lediglich

⁹⁾ Emotionen äußern sich natürlich nicht nur in der Prosodie. So werden aktuell auch Systeme entwickelt, die bestimmte gestisch und/ oder mimisch kodierte Emotionen automatisch analysieren.

¹⁰⁾ Eine vornehmlich wissenschaftlichen Zwecken dienende artikulatorische Sprachsynthese entwickelt Kröger (1996) für das Deutsche.

den Test der direkten Bewertung einer Ausgabe mit Fragen etwa der Art (ebd.: 86): „Does the sentence sound like a question, statement or imperative“? Die vorliegende Arbeit untersucht in Kapitel 7, S. 52ff, in einem Kontexttest prosodische Merkmale von natürlichen und synthetischen Äußerungen dagegen mit einer indirekten Methode, die auf dem kommunikativen Wissen von Sprechern (des Deutschen) über konkrete Sprechsituationen aufbaut.

Eine weitere Anwendung der Prosodieforschung ist der Fremdsprachenunterricht und dessen Notwendigkeit, Wissen über prosodische Formen und deren kommunikative Funktionen zu vermitteln. In diesem Zusammenhang ist auch Software für den Fremdsprachenunterricht zu nennen, die dem Lernenden akustische und grafische Rückmeldungen über seine Aussprache gibt, beispielsweise in Form von automatisch erstellten grafischen F0-Verläufen seiner Äußerungen¹¹. Nach Bagshaw (1994) ist die prosodische Struktur des Lernenden stark von seiner Muttersprache und von prosodischen Stereotypen der zu erlernenden Fremdsprache geprägt. Automatische Sprachlernprogramme können dem Lernenden zunächst auf der phonetischen Ebene eine Rückmeldung über seine Aussprache geben. Eine höhere Analyseebene der Prosodie wird nach Bagshaw (ebd.: 30) mit dem Ziel verwendet, „to determine which of the detectable phonetic differences do have and do not have a semantically contrastive function in a specific language.“ Eine automatische Rückmeldung über Fehler bei den Funktionen der Intonation erfordert dann auch die automatische Analyse der Intonation des Lernenden.

¹¹) Die Verwendung von grafischen F0-Verläufen zur Schulung der Intonation von Gehörlosen behandeln Spaai/Hermes (1993); Chun (1998) und Stibbard (1996) zeigen entsprechende Möglichkeiten im Fremdsprachenunterricht auf.

3. MIKRO- VS. MAKROPROSODIE

Die Unterscheidung von prosodischen Formebenen ist eine grundlegende Voraussetzung bei der Beschreibung und Analyse von F0-Verläufen. Die Mikroprosodie besteht aus den vom Sprecher nicht kontrollierten und vom Hörer im einzelnen nicht bewußt wahrnehmbaren segmentell bedingten Variationen der prosodischen Form. Die Makroprosodie entspricht den kontrollierbaren und im einzelnen wahrnehmbaren prosodischen Formvariationen. Die Makroprosodie bildet die Grundlage der phonologischen Beschreibung der Intonation, die akustisch gesehen in bewußt wahrnehmbaren F0-Verläufen funktionale, d.h. syntaktisch, semantisch und/oder pragmatisch relevante F0-Eigenschaften postuliert¹². Den Postulaten liegt ein Ähnlichkeitskriterium zugrunde, das entweder von der qualitativ perzeptiven oder der quantitativ numerischen Gleichheit eines gemessenen F0-Verlaufs mit einem bestimmten prototypischen (idealisierten) F0-Verlauf ausgeht.

3.1 Mikroprosodische Einflüsse

Lehiste/Peterson (1961) nennen die folgenden, bis auf den ersten Punkt progressiven segmentellen Einflussfaktoren der Mikro- auf die Makroprosodie des amerikanischen Englisch:

- (a) intrinsische F0 der Vokale
- (b) Absenkung der F0 zum Silbennukleus nach stimmhaften Konsonanten, Anhebung nach stimmlosen Konsonanten
- (c) F0-Gipfel am Silbenkernanfang nach stimmlosen Konsonanten, F0-Gipfel in der Silbennukleusmitte mit langsamen Anstieg nach stimmhaften Konsonanten
- (d) F0-Perturbationen im Silbennukleus nach Verschlusslauten

Möbius/Zimmermann/Hess (1987) untersuchen im Deutschen die Faktoren a) und b) bei Plosiven und kommen wie Lehiste/Petersen (1961) zu dem Ergebnis, dass hohe Vokale insgesamt höhere intrinsische F0-Werte als tiefe Vokale haben. Eine Ausnahme bildet das unter den Werten des /E/ liegende kurze /U/. Die einzelnen Vokale wurden mit zwei Messwerten verglichen, die repräsentativ für den F0-Verlauf der Vokale sein sollen. Der erste Wert ist das arithmetische Mittel der F0-Werte aus dem zeitlich mittleren Abschnitt eines Vokals. Der zweite Wert ist der F0-Wert, der auf der Zeitachse nach zwei Dritteln des Vokals erreicht wird. Bei den Plosiven bestätigen die Meßwerte auch den zweiten Einflussfaktor, doch die Autoren folgen der Argumentation von Silverman (1986), nach der dieses Phänomen in Abhängigkeit von dem Intonationsverlauf der Testsätze zu interpretieren sei. Silverman (ebd.: 88) behauptet, „the segmental contribution to F0 is added onto the smooth intonation contour“, so dass die Intonation damit ausschlaggebend für die Richtung des nuklearen F0-Verlaufs sei.

¹²) Funktional relevante Merkmale wie die Unterscheidung von betonten und unbetonten Wortarten, die Position des Wortakzents und die Akzentuierung von morphologischen Einheiten werden in dieser Arbeit nicht behandelt. Ein Beispiel des letztgenannten Phänomens ist ein Satz wie Sie begeistert das Sonnenlicht, in dem ein morphologisch anderer Teil des Kompositums Sonnenlicht akzentuiert wird, wenn der Satz im Gegensatz zu einer Replik auf Mona begeistert das Mondlicht eine Replik auf Mona begeistert die Sonnenfinsternis ist.

Kohler/van Dommelen (1986) untersuchen in zwei Experimenten den Einfluß von /t/ und /d/ in silbenfinaler Position auf die F0 des vorhergehenden vokalischen Silbennukleus. In dem Wort leiten variieren sie die Länge der stimmlosen Verschlussphase des /t/ und die letzten 50 ms des F0-Verlaufs des Silbennukleus. Der F0-Verlauf variiert zwischen ebenem Verlauf und linearem Anstieg von 20 Hz. In den Experimenten zeigt sich, dass der steigende F0-Verlauf die Wahrnehmung des /t/ begünstigt. Damit ist bei der Analyse der F0 im Silbennukleus grundsätzlich auch die postnukleare segmentelle Struktur zu berücksichtigen.

Die meisten Methoden der F0-Modellierung verwenden auf dieser Ebene nur eine Glättungsfunktion. Dies wird damit begründet, dass mikrosegmentell bedingte F0-Variationen zumindest für die linguistischen Funktionen der Intonation irrelevant sind. Bannert/Schmidt/Smude (1994: 41) behaupten, „daß eine Fo-Kurve im Grunde, d.h. wenn alle störenden Einflüsse eliminiert werden, einen sehr gleichmäßigen Verlauf hat.“ Sie bereiten einen gemessenen F0-Verlauf in einer Reihe von manuellen Schritten auf, wobei sie u.a. Irregularitäten in einem F0-Verlauf von mehr als 10 Hz eliminieren und Lücken in großen Akzentbewegungen, die ihrer Meinung nach durch Wortakzente hervorgerufen werden, in bestimmter Art ergänzen.

Eine kaum verwendete Methode der F0-Modellierung besteht in der direkten Berücksichtigung von mikroprosodischen Einflüssen. Thorsen (1979) stellt in diesem Zusammenhang für das Dänische eine gut ausgearbeitete Methode zur Verfügung. Für das Deutsche finden sich in Heuft (1999) einige Hinweise. Heuft kommt aufgrund einer experimentellen Untersuchung zu der Ansicht, dass (ebd.: 39) „ein intuitives Wissen über die vokalintrinsischen Phänomene [...] vorhanden ist, so dass die perzeptive Irrelevanz mikroprosodischer Phänomene nicht einfach vorausgesetzt werden kann.“ Bei dem Aufbau einer Datenbank zur Erstellung von Prosodieregeln bildet sie eine Gruppe von Vokalen mit eher hoher und eine mit eher niedriger F0.

3.2 Elemente der Makroprosodie

Die Makroprosodie manifestiert sich auditiv wesentlich in dem wahrgenommenen Tonhöhenverlauf einer sprachlichen Äußerung. In der Literatur existieren in diesem Zusammenhang unterschiedliche Annahmen über die perzeptiv relevanten Bausteine von derartigen Tonhöhenverläufen. Die klassischen konträren Positionen sind in diesem Zusammenhang die Annahme von kinetischen Tönen, bestehend aus steigenden und/oder fallenden Verläufen, und die konkurrierende Annahme von zwei oder mehr Tonhöhenebenen, die in der Regel linear interpoliert werden.

In der Intonationsforschung postuliert die ursprünglich pädagogisch motivierte Britische Schule aus komplexen kinetischen Tönen bestehende Tongruppen, die holistisch interpretierte elementare Einheiten der Intonation bilden. Der Tradition der Britischen Schule folgt das in Kapitel 6.1, S. 33ff, behandelte Kieler Intonationsmodell, während die Annahme von Tonhöhenebenen charakteristisch für das in Kapitel 4, S. 20ff, vorgestellte Tonsequenzmodell ist. Die Britische Schule ist nach Fox (2000: 277f) auf Arbeiten zum Deutschen von Hermann Klinghardt, z.B. Klinghardt (1927), zurückzuführen, dessen intonatorische Einheit der Sprechakt ist. Diese nicht mit dem musikalischen Takt zu verwechselnde Einheit unterteilt Klinghardt in einen einleitenden Auftakt, in eine oder mehrere folgende Satzakkzentsilben und in den abschließenden Abtakt.

3.2.1 Globaler Tonhöhenverlauf

Tonhöhenverläufe können eine äusserungs- oder phrasenbezogene abnehmende Tendenz in ihrer Amplitude aufweisen¹³. Das damit bezeichnete Phänomen der Deklination wird akustisch beschrieben als Gefälle einer Geraden zwischen bestimmten lokalen Maxima eines F0-Verlaufs ('topline') oder analog als Gerade zwischen bestimmten lokalen Minima ('baseline'). Huber (1988: 73f) beispielsweise berechnet die Deklinationslinien für intonatorische Einheiten des gelesenen Schwedisch, indem er durch die F0-Gipfel und -Täler einer Einheit jeweils eine Regressionsgerade zieht und nach dem 'best fit'- Kriterium approximiert. Der Approximation liegt als Fehlerkriterium der Produkt-Moment-Korrelationskoeffizient von Pearson zugrunde, den Huber auf $r < 5$ setzt. Die Punkte der beiden Geraden dienen als Referenz für verschiedene 'over-' oder 'undershoots', die durch Akzentuierung, Deakzentuierung oder Prominenz verursacht sind. Die einzelnen F0-Gipfel und -Täler bestimmt Huber, indem er mindestens 10 Hz auseinanderliegende Punkte des F0-Verlaufs sucht. Initiale und finale Punkte, die ein lokales Minimum bilden, werden nach Möglichkeiten von der Berechnung der Regressionsgerade ausgeschlossen.

Deklination kann auch als der innerhalb der genannten Geraden abnehmende F0-Amplitudenverlauf bestimmt werden, wobei sich die lokalen Minima und Maxima nicht unbedingt gleich stark verringern müssen. In der Literatur treten in diesem Zusammenhang zudem die Begriffe 'downdrift' und 'downstep' auf. Hirst/di Cristo (1998: 21) definieren 'downdrift' auditiv als „iterative lowering of successive high pitches within an intonation unit separated by intermediate low points“ und 'downstep' als ein vergleichbares Absinken, allerdings ohne „intervening low tones.“ Fox (2000: 307ff) referiert verschiedene Erklärungen der Deklination, wobei er als korrelierende sprachliche Einheit der Deklination für das Deutsche den Fuß annimmt¹⁴:

- (a) Deklination ist eine Folge der Abnahme des subglottalen Luftdrucks und damit physiologisch bestimmt, d.h. sie ist vom Sprecher nicht kontrollierbar.
- (b) Deklination ist vom Sprecher kontrollierbar und wird linguistisch verwendet.
- (c) Deklination wird linguistisch verwendet, aber nur in Form einer 'once-for-all' - Entscheidung, d.h. der Sprecher trifft zu Beginn seiner Äußerung eine Entscheidung über den Deklinationsverlauf (bzw. über deren linguistische Funktion), den er dann nicht mehr kontrollieren kann.
- (d) Deklination wird linguistisch verwendet und kann innerhalb einer Äußerung neu angesetzt, also kontrolliert werden.

Gussenhoven/Rietveld (1988) untersuchen im Holländischen in verschiedenen Experimenten den wohl bekanntesten Effekt der Deklination, nach dem ein später F0-Gipfel einer Äußerung höher als ein früher Gipfel wahrgenommen wird, obwohl die F0-Werte der beiden Gipfel gleich sind. Die experimentellen Ergebnisse erklären diesen Effekt als kombinierte Folge eines einheitlichen, zeitabhängigen globalen Abfalls eines F0-Verlaufs (einer Äußerung) und eines entsprechenden lokalen Abfalls am Ende des Verlaufs. Nach Umeda (1982) ist Deklination allerdings

¹³ Blaauw (1995: 154ff) vergleicht die paragraphenbezogene Deklination in holländischer Lese- und Spontansprache.

¹⁴ Der Fuß ist ein Begriff der antiken Metrik, die den Sprechakt insbesondere mit den Begriffen des Trochäus, Daktylus und Jambus beschreibt; vgl. Maas (1999: 76ff).

ein nur in Laborsprache anzutreffendes Phänomen, das sie in ihrer Untersuchung von Spontansprache nicht beobachten kann¹⁵.

Deklinationseigenschaften spielen auch eine Rolle bei der Analyse der Variation des Tonhöhenumfangs. Ladd (1996) unterscheidet Analysen, die Tonhöhenveränderungen auf bestimmte Äußerungsabschnitte relativieren ('initialising approach'), von Analysen, die relativ zu der Stimme eines Sprechers sind ('normalising approach'). Nach Ladd (ebd.: 253) „[the initialising approach, RS] attempts to provide invariant characterizations of all tones in terms of what has preceded in the utterance.“ Bei der Modellierung der F0-Werte einer Äußerung benötigt dieser syntagmatisch arbeitende Ansatz einen geeigneten Startwert.

Der 'normalising approach' abstrahiert dagegen von Unterschieden zwischen Sprechern, so dass letztlich invariante tonale Merkmale eines idealisierten Sprechers beschrieben werden. Als geeignetes Verfahren zur Normalisierung einer sprecherspezifischen Skala schlägt Ladd die 'z-score normalisation' von Rose (1987) vor. Das Verfahren arbeitet mit den F0-Werten einer oder mehrerer Äußerungen eines Sprechers, deren Mittelwert den normalisierten Wert null erhält. Eine skalierte Normalisierung der einzelnen F0-Werte wird erzeugt, indem den einzelnen F0-Werten der Grad ihrer Standardabweichung vom Mittelwert zugewiesen wird. Ladd (ebd.: 257) plädiert für normalisierte Modelle bei der Analyse von tonalen Intonationsverläufen, weil seiner Meinung nach „the scaling of pitch targets [...] highly systematic“ ist. Er unterscheidet in diesem Zusammenhang auch zwischen dem tonalen Niveau ('overall level') und der Variationsweite eines Tonhöhenverlaufs ('span'), wobei ein höheres Niveau zumeist mit einer größeren Variationsweite verbunden ist. Die Korrelation ist der Grund für die in der Literatur häufig vorgenommenen nicht-linearen, logarithmischen Skalierung von F0-Verläufen, die beispielsweise den Vergleich von Frauen- und Männerstimmen ermöglichen soll.

Kohler (1983) untersucht den Einfluss der Sprechgeschwindigkeit auf die Variationsweite von F0-Verläufen. Er kommt zu dem Ergebnis, dass F0 mit steigender Sprechgeschwindigkeit ansteigt, wobei auch eine größere Intensität zu beobachten ist. Der Anstieg ist bei F0-Maxima größer als bei F0-Minima, und es tritt in beiden Fällen eine fallende Tendenz am Äußerungsende auf. Die obere Deklinationslinie ist steiler als die untere Linie. Mit zunehmender Sprechgeschwindigkeit nimmt auch der Abstand eines F0-Maximums zu dem folgenden Minimum ab.

3.2.2 Lokaler Tonhöhenverlauf

House (1990) entwickelt am Schwedischen ein Modell der Tonhöhenwahrnehmung mit universell linguistischem Anspruch. Nach dem Modell ist die perzeptive Unterscheidung von kinetischen und ebenen Tönen in sprachlichen Äußerungen von der spektralen Beschaffenheit des Vokals (dem Silbenkern) abhängig¹⁶. Ein steigender bzw. fallender Tonhöhenverlauf wird als ebener hoher bzw. tiefer Ton wahrgenommen, wenn in dem entsprechenden Vokal eine relativ schnelle spektrale Veränderung zu beobachten ist. Die mögliche Ableitung der kinetischen auf ebene Töne impliziert nach House (ebd.: 139), dass die nicht ableitbaren ebenen Töne gegenüber den kinetischen Tönen „more perceptually salient and more frequent“ sind.

¹⁵ Das Thema 'Labor- vs. Spontansprache' behandelt auch Kapitel 7.1, S. 52f.

¹⁶ Vergleiche in diesem Zusammenhang auch die Diskussion von Methoden der F0-Berechnung in Vary/Heute/Hess (1998: 195ff); Reetz (1996) diskutiert weitere psychophonetische Aspekte der Tonhöhenwahrnehmung, wobei er auch eine eigene zeitbasierte Methode der F0-Berechnung entwickelt.

Eine weitere Bedingung der Wahrnehmung von kinetischen Tönen ist, weiter House (1990) folgend, der tonale Beginn ca. 30-50 ms nach dem Vokalanfang, wobei der Verlauf mit der abnehmenden neuen spektralen Information des Vokals synchronisiert sein muss. Zudem darf der Vokal eine empirisch zu ermittelnde Länge nicht unterschreiten. House schlägt zunächst eine Länge von 100 ms vor, in der ein tonaler Verlauf von mindestens drei und maximal acht Halbtönen stattfindet. Ein Tonhöhenverlauf kann demnach prinzipiell aus kinetischen und aus ebenen Tönen bestehen.

Steigende und fallende Tonhöhenverläufe müssen in ihren Amplituden nicht symmetrisch sein. Nach t'Hart (1974) umfassen im Holländischen wahrnehmbare steigende F0-Verläufe mindestens 1.5 Halbtöne und wahrnehmbare fallende F0-Verläufe mindestens drei Halbtöne. Verhoeven (1994) untersucht ebenfalls im Holländischen die segmentelle Zuordnung ('alignment') von steigenden und fallenden F0-Verläufen. Zu diesem Zweck werden kurze Äußerungen akustisch derart manipuliert, dass sich der Beginn eines Verlaufs innerhalb einer akzentuierten Silbe in bestimmten Intervallen (10 bzw. 15 ms) nach rechts verschiebt. Die Aufgabe der Probanden besteht darin, die melodische Gleichheit von manipulierten und nicht manipulierten Stimuli zu beurteilen.

Die Untersuchung ergibt, dass steigende und fallende F0-Verläufe in isoliert präsentierten Stimuli gleich gut diskriminiert werden. Verhoeven (ebd.: 82) folgert¹⁷, „alignment is essentially perceived similarly in a context in which pitch movements function similarly: positional variation in isolated rises and falls is not exploited linguistically in the Dutch intonational system.“ Die Funktion der F0-Verläufe ist in diesem Fall die Markierung des Wortakzents. Die Sensitivität gegenüber der segmentellen Positionierung steigender F0-Verläufe nimmt dagegen im Vergleich mit fallenden F0-Verläufen ab, wenn die Verläufe in einem anderen prosodischen Kontext verwendet werden. Die Manipulation von steigenden F0-Verläufen des linken Pfeilers eines Hutmusters - auch Brückenakzent genannt - führt zu schlechteren Unterscheidungsraten als die Manipulation des rechten Pfeilers. Der Brückenakzent des Deutschen wird in Kapitel 5.2, S. 30 ff, vorgestellt.

¹⁷) Die englischsprachige Literatur verwendet den Terminus „Pitch“ sowohl für die wahrgenommene Tonhöhe als auch für den gemessenen F0-Verlauf.

4. Das phonologische Tonsequenzmodell

Pierrehumbert (1980) entwickelt zur Beschreibung der Intonation des amerikanischen Englisch das seitdem prominente phonologische Tonsequenzmodell. Das Modell nimmt eine Einteilung einer Äußerung in Intonationsphrasen vor, die aus Grenz- und Satzakkzenttönen aufgebaut werden. Bei den sequenziell von links nach rechts analysierten Tönen einer Intonationsphrase handelt es sich lediglich um einen Hoch- oder Tiefton im Sinne des Zielpunktes einer Tonhöhenbewegung, wobei Akzente in der Regel aus einem oder maximal zwei Tönen bestehen (mono- oder bitonale Akzente). Die einzelnen Töne werden grundsätzlich linear interpoliert, wobei Pierrehumbert annimmt (ebd. 52): “[...], a good theory of interpolation probably waits on a better understanding on the motor control for intonation”. Eine Ausnahme bei der linearen Interpolation bilden zwei hochtonige Akzente H*, bei denen ohne einen dazwischenliegenden Ton ein „dip“ (ebd.: 70) in der interpolierten tonalen Verbindung auftritt. Pierrehumbert (ebd.: 29) formuliert eine Grammatik endlicher Zustände, die wohlgeformte tonale Sequenzen von Intonationsphrasen des amerikanischen Englisch erzeugt.

Die Intonationsphrase ist eine elementare prosodische Beschreibungseinheit, die auf der Basis von nicht durch Zögern entstandenen, sog. linguistische Pausen und/oder der Koinzidenz mit bestimmten syntaktisch-semantischen Einheiten ermittelt werden¹⁸. Das Konzept der Intonationsphrase erweitern Beckman/Pierrehumbert (1986) um prosodische Untereinheiten bildende intermediäre Intonationsphrasen. Das Tonsequenzmodell setzt eine Theorie der metrischen Gitter voraus, deren Konstruktion im Deutschen zunächst Kapitel 4.1, S. 20f, beschreibt. Kapitel 4.2, S. 22f, behandelt im Anschluß die Grundlagen des tonalen Aufbaus der Intonationsphrase, die im Rahmen des Tonsequenzmodells ebenfalls für das Deutsche entwickelt wurden. Die tonalen Elemente der Intonationsphrasen sind auf einer eigenen autonomen Ebene ('tier') angesiedelt, die sich von der Ebene der metrischen Gitter unterscheidet. Die Arbeiten zum Deutschen weichen in einigen, allerdings nicht grundsätzlichen Punkten von Pierrehumbert (1980) ab.

4.1 Metrische Gitter im Deutschen

Uhmann (1991) entwickelt in ihrer Arbeit ein phonologisches Intonationsmodell von Fokusstrukturen des Deutschen, wobei sie sich zur Beschreibung von Prominenzrelationen der Theorie der metrischen Gitter von Selkirk (1984) bedient. Metrische Gitter basieren auf Text-Gitter-Regeln, die den Aufbau der metrischen Prominenz von Silben auf der Basis der Textstruktur erfassen. Die von diesen Regeln erzeugte rhythmische Struktur wird mit Wohlklangsregeln im Hinblick auf ein von Selkirk (1984) postuliertes ideales Gitter umgeformt.

Einem idealen Gitter liegt das Prinzip der rhythmischen Alternation zugrunde. Demzufolge gibt es keine zwei benachbarten prominenten Silben und können höchstens zwei nicht-prominente Silben aufeinanderfolgen. Um die rhythmische Alternation bei dem Aufbau eines metrischen Gitters zu gewährleisten, existieren Wohlklangsregeln, die die Hinzufügung, Bewegung oder Tilgung eines Schlages (x) ermöglichen. Zur Veranschaulichung der auf syntaktischen Struk-

¹⁸ Helfrich (1985) hat die Rolle von F0 bei der Bildung von Teilsatzgrenzen am Deutschen experimentell untersucht; von Heusinger (1999: 72f) stellt weitere linguistische Kriterien bei der Bestimmung von Intonationsphrasen vor, insbesondere das der traditionellen Sinneinheit.

turen operierenden Text-Gitter-Regeln diene zunächst das folgende in seiner Notation leicht modifizierte Beispiel von Uhmann (ebd.: 179):

						x	Ebene 4	
	x					x	Ebene 3	
	x		x			x	Ebene 2	
	x	x	x		x	x	Ebene 1	
	x	x	x	x	x	x	Ebene 0	
<hr/>								
	σ^*	σ		σ	σ	σ	σ^*	
(IP [AD1 [W	(Pe)	(ter)]]	[AD2 [W	(te)	(le)	(fo)	(niert)])
	AT						AT	

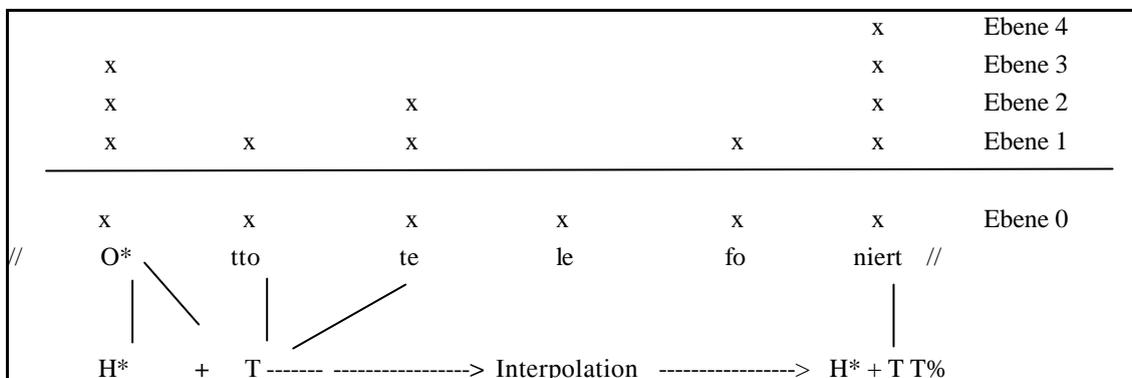
Die vorausgesetzte Silbenstruktur ist bereits syntaktisch in zwei Wörter (W) und intonatorisch in eine Intonationsphrase (IP) eingeteilt. Auf unterster Ebene erhält jede Silbe (σ) des Beispiels nach dem Silbenprinzip einen Schlag, da die Silbe einerseits die Grundeinheit rhythmischer Organisation bildet und andererseits die Assoziation des metrischen Gitters mit der Textstruktur hergestellt werden muss. Die erste Ebene erteilt allen Silben einen Schlag, die einen unreduzierten Vokal aufweisen. Sie sind potentielle Iktusträger, d.h. Träger metrischer Prominenz. Auf der zweiten Ebene wird konkret die Iktusträgerin bestimmt (σ^*), die in nativen Wörtern die erste bereits einen Schlag tragende Stammsilbe und in nicht-nativen Wörtern die letzte schwere Silbe ist. Die entstehende Verletzung der rhythmischen Alternation wird durch Schlaghinzufügung auf derselben Ebene beseitigt.

Die dritte Ebene erhält die Schläge, die auf die Zuweisung eines Akzenttons (AT) zurückgehen. Es handelt sich hierbei um Fokusakzente, die auf die entsprechende Silbe des Fokusexponenten fallen. In dem obigen Beispiel liegt enger Fokus mit telefoniert als Exponenten vor (vgl. Kap. 5.1, S. 47f), so dass der Akzentton einfach an die Iktussilbe gehen kann. Die Bildung von Akzentdomänen (AD) innerhalb von Intonationsphrasen bewirkt u.a., dass jede Domäne einen eigenen Akzentton erhält (Schlag auf Pe). Akzentdomänen umfassen die von einem Akzentton involvierten syntaktischen Konstituenten. Auf der vierten Ebene wird schließlich der letzten Silbe der Intonationsphrase nach der Regel der Endakzentstärkung immer dann ein Schlag zugeteilt, wenn die Intonationsphrase mehr als einen Akzentton erhält.

In dem diskutierten Beispiel wurden phonologisch nicht näher spezifizierte Akzenttöne mit Iktus- bzw. Akzentsilben assoziiert. Die Assoziation der aus der Intonationskontur gewonnenen Akzenttöne erfolgt innerhalb von Tondomänen. Eine Tondomäne erstreckt sich maximal von einer Akzentsilbe bis zu der finalen Silbe der entsprechenden Intonationsphrase. Bei Fokusstrukturen bilden die Fokusbereiche zudem maximale Grenzen für Tondomänen. Die mögliche Dehnung eines tropischen, d.h. nicht-gesternten Tons geht innerhalb einer Tondomäne lediglich bis zu der ersten Silbe, die auf der zweiten Ebene einen Schlag aufweist.

Die die Assoziation bestimmenden Konventionen besagen im wesentlichen, dass innerhalb einer Tondomäne sequenziell jede Silbe einen Ton erhält. Tropische Töne werden genau einer Silbe zugeordnet, wobei bei bitonalen Tönen Dehnungen oder Quetschungen auftreten können. Silben, mit denen kein tonales Merkmal assoziierbar ist und die dennoch intonatorisch relevant sind, werden unter Anlehnung an Pierrehumbert (1980) mit Interpolationsregeln erfasst. Diese

Regeln stellen die intonatorische Verbindung zwischen zwei tontragenden Einheiten her. Abschließend sei noch auf die mit der letzten Silbe einer Intonationsphrase assoziierten obligatorischen, finalen Grenztöne (T(ief)%, H(och)%) sowie auf die entsprechend mit der ersten Silbe assoziierten fakultativen Grenztöne hingewiesen. Ein weiteres Beispiel von Uhmann (ebd.: 194) verdeutlicht die Assoziation von Tönen und Silben¹⁹:



4.2 Aufbau der Intonationsphrase

In Féry (1993) findet sich eine Beschreibung des tonalen Aufbaus der Intonationsphrase bei der sog. normalen Intonation²⁰. Die normale Intonation setzt semantisch einen weiten Fokus voraus (vgl. Kap. 5.1, S. 33f). Sie besteht aus genau einem Nuklearakzent sowie aus bestimmten prä- bzw. postnuklearen Tönen.

Intermediäre Phrasen (der Intonation) werden von Féry angesetzt, wenn innerhalb einer Intonationsphrase durch Realisierung eines bitonalen Tons der Eindruck einer eigenständigen Phrase entsteht. Die Begrenzung der intermediären Phrasen erfolgt durch den Folgeton ihres letzten bitonalen Tons oder durch einen fakultativen hohen Grenztönen, wenn sie mit dem Ende einer Intonationsphrase zusammenfallen. Der Folgeton eines bitonalen Tons erfüllt im Deutschen eine doppelte Funktion. Er sorgt für die melodische Gestaltung zwischen dem Nuklearakzent und dem Grenztönen einer Intonationsphrase (IP), und er markiert die Grenze einer intermediären Phrase (iP). Die folgenden Beispiele a) und b) von Féry (ebd.: 78) sollen das Konzept der Intonations- und der intermediären Phrase verdeutlichen²¹:

L*H	H%	H*L	H*	H*L
a) [IP [iP HAT sie eine blaue Hose?]]		b) [IP [iP die KLEINE] [iP ALTMODISCHE BAHN]]		

¹⁹) Die Querstriche markieren den Anfang und das Ende einer Intonationsphrase.

²⁰) Nach Höhle (1982) ist ein normal betonter Satz mit einer in einem bestimmten Sinn größtmöglichen Anzahl von Kontexten kompatibel. Fuchs (1984) behandelt Deakzentuierung, d.h. die nicht realisierte Akzentuierung von normalerweise akzentuierten syntaktischen Konstituenten, die mit einer Bedeutungsveränderung des entsprechenden Satzes verbunden werden kann.

²¹) Großbuchstaben stehen für Akzentuierung.

4.2.1 Der Nuklearakzent

Eine Intonationsphrase besteht nach Féry (1993) im Deutschen mindestens aus einem Nuklearakzent. Der Nuklearakzent ist per definitionem der letzte realisierte Akzent einer Intonationsphrase. Er ist im allgemeinen (zumindest in der Wahrnehmung) der prominenteste Akzent der Intonationsphrase. Dem Nuklearakzent können verschiedene pränukeare Töne vorangehen sowie ein fakultativer hoher Grenzton (H%) folgen. Pränukeare Töne bestehen aus einem oder aus zwei Tönen. Als Nuklearakzente setzt Féry einen (einfachen) fallenden (H(igh)* L(ow)), einen (einfachen) steigenden (L* H), einen fallend-steigenden (H* L H%) und einen steigend-fallenden (L* H L) Akzent an²².

Die fallenden Nuklearakzente können stilisiert werden. Bei den anderen Nuklearakzentformen ist dies nach Féry nicht möglich. Im Fall von Stilisierung tritt eine Dehnung der gesternten Töne und von deren assoziierten Silben ein. Die resultierende Kontur ist mit einem Tonhöhenverlauf bei der Anrede oder dem Rufen von Personen zu vergleichen. Sie erfordert die Annahme eines weiteren Tons M(id), da der Fall eher in einen ebenen Verlauf übergeht. Die Tonhöhenfolge ist mit der gesamten phonologischen Einheit zu assoziieren, so dass kein gesternter Ton auftritt. Féry (ebd.: 101) nennt in diesem Zusammenhang folgende Beispiele c), d):

H M	H M
	/ \
c) AXEL	d) ARBEITEN!

Eine weitere Modifikation der Nuklearakzente ist die Bildung eines frühen Gipfels (vgl. Kap. 6.1.1, S. 33f). Der Gipfel eines bitonalen Tons wird dabei mit der der akzentuierten Silbe vorangehenden Silbe assoziiert. Die Verlagerung des Gipfels von der metrisch starken auf eine metrisch schwache Silbe beeinflusst nicht die Platzierung des Akzents. Die metrisch starke Silbe wird weiterhin als akzentuiert wahrgenommen. Dieses stilistische Mittel verwenden nach Féry typischerweise Fernsehreporter.

4.2.2 Pränukeare Töne

Pränukeare sind nach Féry (1993) alle mono- oder bitonalen Töne einer Intonationsphrase, die nicht final sind. Bitonale pränukeare Töne bilden eine eigene intermediäre Phrase. Sie können mit dem Nuklearakzent in eine Phrase integriert werden, wobei eine Unterdrückung der (intermediären) prosodischen Korrelate der Phrasierung, wie Pause, Dehnung der finalen Silbe oder ein größerer Umfang des tonalen Falls oder Anstiegs, stattfindet. Dabei treten Verbindungsregeln in Kraft, die im Sinne eines Ton-Sandhi zu verstehen sind. Nach diesen Regeln ist ein monotonaler pränukearer Ton als abgeleitet von einem zugrundeliegenden bitonalen Ton anzusehen. Féry (ebd.: 120f) modifiziert für das Deutsche die Verbindungsregeln (a), (b) von Gussenhoven (1984)²³:

²²) Wie bereits erwähnt, indiziert der Stern die Assoziation des Tons mit der akzentuierten Silbe.

²³) Nach Ladd (1996: 110) handelt es sich bei beiden Regeln genau genommen um „pitch accent linking rule[s]“.

- (a) $T^*T \ T^*T \ \text{--->} \ T^* + T \ T^*T$
 (b) $T^*T \ T^*T \ \text{--->} \ T^* + T^*T$

Regel (a) nimmt eine partielle und Regel (b) eine vollständige Verbindung der bitonalen Töne vor ('tonal linking'), wobei die Variable T von einem hohen oder tiefen Ton belegt werden kann. In (a) wird auf den nicht hervorgehobenen Silben zwischen dem Tonhöhenakzent T^* und dem folgenden Ton T, der auf der rechten Seite mit einer Silbe in der Nähe des zweiten Tonhöhenakzents T^* assoziiert ist, entsprechend der Belegung von T eine leicht fallende oder steigende Kontur realisiert. Das Verschwinden des Tons in (b) und die Bildung einer (idealerweise) vollständig verbundenen Kontur zwischen den beiden gesternten Tonhöhenakzenten ist mit einem deutlichen melodischen Unterschied im Vergleich zur Kontur der linken Seite von Regel (b) verbunden. Die Verbindungsregeln werden optional angewendet. Bei einem steigenden bitonalen Ton, dem ein fallender Nuklearakzent folgt, entsteht ein Brückenakzent (vgl. Kap. 5.2, S. 33ff).

4.3 Phonetische Abbildungsregeln

Möhler (1998: 82ff) stellt für das Deutsche ein Verfahren zur regelbasierten Generierung des F₀-Verlaufs einer Intonationsphrase vor, das aus den einzelnen Tönen des phonologischen Tonsequenzmodells konkrete F₀-Werte berechnet. Möhler (ebd.: 82f) betont: „Es bleibt zu überprüfen, inwieweit eine solche Intonation als natürlich empfunden wird. Sie muß allerdings in der Lage sein, all diejenige Information zu tragen, die auch in natürlichen Äußerungen in der Intonation kodiert werden kann. Ansonsten wäre eine Beschränkung auf zwei phonologische Kategorien H und L nicht gerechtfertigt.“ Das Verfahren wurde an einem von verschiedenen Sprechern erzeugten Korpus mit 22 bitonalen Akzenten, 14 aus drei Tönen bestehenden Akzenten, fünf Grenztönen und neun Realisierungen von 'downstep' entwickelt.

Das Verfahren berechnet zur Begrenzung des F₀-Bereichs einer Intonationsphrase zwei fallende Geraden, eine 'top-' und eine 'baseline'²⁴. Die Hoch- und die Tieftöne der Intonationsphrase fallen jeweils auf die entsprechende Gerade, wohingegen hohe und tiefe Grenztöne auch über die Gerade hinaus gehen können. Mit den beiden Geraden wird 'downdrift' modelliert, nach dem in einer Folge bestehend aus Tief- und Hochton der letztere Ton per default eine niedrigere Tonhöhe erreicht als ein der Folge vorhergehender Hochton. 'Upstep' eines Tons bildet den markierten Fall, der eine Äußerung semantisch modifiziert. Die einzelnen Töne werden zur Nachbildung des Intonationsverlaufs mit einem Polygonzug interpoliert, wobei die Eckpunkte des Polygonzugs in Frequenzrichtung bzgl. der 'top-' und der 'baseline' und in Zeitrichtung bzgl. der Silbenstruktur der Äußerung normiert sind.

Möhler (ebd.: 86ff) gibt Regeln für die Bestimmung der F₀-Werte der Eckpunkte der Töne an. Die bitonalen Akzente H^*L und L^*H können ein- und zweisilbig realisiert werden. Bei einer zweisilbigen Realisierung tritt der gesternte Ton nach 60% des stimmhaften Bereichs der Akzentsilbe auf und der Folgeton in der Mitte des stimmhaften Bereichs der postakzentuierten

²⁴ Die genaue Berechnung der Geraden gibt Möhler (1998) nicht an. Pierrehumbert (1980: 135) definiert die 'baseline' als eine für alle Äußerungen eines Sprechers konstante Linie, die den "hypothetical bottom of his range to which tonal values are referenced" bildet.

Silbe. Ein bitonaler Akzent wird einsilbig realisiert, wenn er auf eine phrasenfinale Silbe fällt oder in der nachfolgenden Silbe bereits ein neuer Akzent folgt. In diesem Fall beginnt der bitonale Akzent nach 30% des stimmhaften Bereichs der Akzentsilbe und endet dort nach 80%. Verbundene ('gelinkte') bitonale Akzente (vgl. Kap. 4.2, S. 22ff) werden wie ihre nicht verbundenen gesternten Varianten behandelt, wobei die Folgetöne von partiell verbundenen bitonalen Akzenten allerdings in der Mitte des stimmhaften Bereichs der postakzentuierten Silbe auftreten.

Der L*HL - Akzent wird innerhalb von zwei Silben realisiert, wobei L* nach 30% des stimmhaften Bereichs der Akzentsilbe auftritt. H tritt entsprechend nach 10% der postakzentuierten Silbe auf, wenn die Silbe mit einem stimmlosen Konsonanten beginnt. Wenn die Silbe nicht mit einem derartigen Konsonanten beginnt, dann tritt H am Ende des stimmhaften Bereichs der postakzentuierten Silbe auf. L ist stets mit dem Ende der postakzentuierten Silbe assoziiert. Der HH*L - Akzent wird nur realisiert, wenn der Akzentsilbe eine metrisch schwache Silbe vorausgeht (vgl. dazu den frühen Gipfel in Kapitel 6.1.1, S. 73f). Dabei tritt H in der Mitte des stimmhaften Bereichs der präakzentuierten Silbe, H* in der Mitte des stimmhaften Bereichs der Akzentsilbe und L nach 30% des stimmhaften Bereichs der postakzentuierten Silbe auf. H* wird um einen Faktor von im Mittel 0.7 unter der von Möhler berechneten 'topline' realisiert, wohingegen L die 'baseline' erreicht. Der Faktor wird auch zur Berechnung von 'downstep' verwendet.

Eine hohe Phrasengrenze H% bewirkt einen Anstieg des Tonhöhen- bzw. F0-Umfangs, der 30% des Umfangs oberhalb der 'topline' am Ende des stimmhaften Bereichs der phrasenfinalen Silbe endet. Wenn ein L*H - Akzent zwei Silben vor H% auftritt, dann verschmelzen L*H und H%, so dass der Hochtton des bitonalen Akzents in der vorletzten Silbe einer Intonationsphrase wiederum 30% des Umfangs über der 'topline' liegt. Eine tiefe Phrasengrenze L% erhält am Ende des stimmhaften Bereichs der phrasenfinalen Silbe einen Frequenzwert von 20% des Umfangs unterhalb der 'baseline', wenn in den letzten beiden Silben der Intonationsphrase kein H*L -, L*HL - oder HH*L - Akzent auftritt. Tritt einer von diesen Akzenten auf, dann wird L auf das Ende des stimmhaften Bereichs der phrasenfinalen Silbe verschoben, wobei der Ton ebenfalls um die genannten 20% fällt.

4.4 Kritik des Tonsequenzmodells

Das phonologische Tonsequenzmodell beschreibt invariante und semantisch funktionale abstrakte Formen der Intonation, die sprachakustisch gesehen auf konkrete F0-Verläufe abgebildet werden müssen. Die phonologische Analyse von realisierten Satzakkenten besteht somit aus der gleichzeitigen Dekodierung semantischer Funktionen und phonetischer Formen der Satzakkente. Die damit angesprochene, viel diskutierte Beziehung zwischen Phonologie und Phonetik charakterisiert Pierrehumbert (1990: 375) wie folgt: „If phonology is not related to phonetics, it models the mind of a solipsistic isolate. If phonetics is not related to phonology, it models noises and gestures to which no meaning or category structure can be assigned.“

Pierrehumbert (1980: 10) nennt als ein wesentliches Ziel ihrer prominenten Arbeit, „to investigate the rules which map these phonological representations [of stress patterns, RS] into phonetic representations.“ Pierrehumbert (1980) formuliert allerdings keine konkreten Regeln für vorgenommene phonologische Analysen, wie sie beispielsweise Möhler (1998) für das Deutsche auf-

stellt (vgl. Kap. 4.3, S. 20f). Gleichmaßen enthalten phonologische Tonsequenzanalysen des Deutschen im allgemeinen keine entsprechenden Regeln. So gehören nach Uhmann (1991: 190) beispielsweise Interpolationsregeln „nicht mehr zum Bereich der phonologischen Beschreibung des Deutschen [...], sondern [dienen] der Überführung der phonologischen Repräsentation in die phonetische Repräsentationsebene.“

Die mangelnde phonetische Spezifikation der phonologischen Analysen des Tonsequenzmodells ist der Grund für die Außerachtlassung des Modells in dem experimentellen Teil der vorliegenden Arbeit. Ladd (1996: 103) thematisiert die mangelnde phonetische Spezifikation, indem er fragt, „by what criteria do we decide that a given pitch accent consists of one or two tones? How do we determine that there is or is not a tone at any given point in a string?“ Taylor (2000: 20) argumentiert im Hinblick auf die Semantik, „there has been no evidence to show that there are strict boundaries between intonational units which signal abrupt changes in meaning.“ Die vorliegende Arbeit schliesst sich dieser Kritik grundsätzlich an; sie liefert dabei einen Beitrag zur Untersuchung der invarianten Formen bzw. Formeigenschaften von kontrastiven Satzakzenten im Deutschen. Invariante Formen sind Voraussetzung einer phonologischen Analyse der Intonation (im Deutschen).

5. Funktionen von Satzakzenten

Die sprachlichen Funktionen von Satzakzenten (im Deutschen) werden unter verschiedenen Aspekten analysiert. Ein vielbeachtetes linguistisches Konzept ist die Informationsstruktur eines Satzes (im Sinne einer potentiellen Äußerung), in der der Satzaccent einen Fokus oder ein Topik kennzeichnet. Fokus- und Topikakzente sind bestimmte Typen von Satzakzenten, die nach ihrer Funktion bestimmt werden. Die Informationsstruktur besteht dann etwa nach Jacobs (1984, 1988) aus der Fokus-Hintergrund-Gliederung eines Satzes und/oder aus der Gliederung, die den Satz in Topik und Kommentar zerlegt. Eine eher traditionelle informationsstrukturelle Beschreibungsebene bildet die Unterscheidung von neuer und alter Information des entsprechenden Diskurses.

Der Fokus eines Satzes ist der inhaltlich hervorgehobene, wichtige Teil eines Satzes, der dem nicht hervorgehobenen, unwichtigen Teil des Satzes, also dem Hintergrund, gegenübergestellt ist. Wichtige Information ist nach Jacobs (1988) nicht automatisch mit neuer Information gleichzusetzen. Die Topik-Kommentar-Gliederung eines Satzes bestimmt Jacobs (1984: 46) hingegen wie folgt: „Das Topik eines Satzes ist ein (in den allermeisten Fällen satzeinleitendes) Element, das einen Rahmen für die Interpretation des Restsatzes absteckt. Der Kommentar ist der Restsatz, der diesem Rahmen entsprechend interpretiert wird.“ Die Setzung eines Interpretationsrahmens läßt sich dabei mit der Formulierung „Was X betrifft, ...“ ermitteln²⁵. Das Topik ist mit anderen Worten die Entität, über die ein Satz etwas aussagt, und der Kommentar der Teil des Satzes, mit dem (von dem Sprecher des Satzes) etwas zu dem Topik ausgesagt wird²⁶.

Grammatisch orientierte Analysen von Satzakzenten postulieren eine Beziehung zwischen der syntaktischen Struktur, der Plazierung des Satzaccents bzw. der intonatorischen Form und der Informationsstruktur eines Satzes. Das Postulat konkurriert mit einer Auffassung, nach der die Realisierung eines Satzaccents weniger von der syntaktischen Struktur des Satzes, sondern vielmehr von den Intentionen des Sprechers in einer gegebenen Kommunikationssituation abhängt; die wohl bekannteste, pointierte Formulierung dieser Auffassung stammt von Bolinger (1972): 'Accent is predictable (if you're a mindreader)'. Monaghan (1993: 573) differenziert in diesem Zusammenhang, indem er die folgende Hierarchie von Einflussfaktoren bei der Plazierung von Satzakzenten formuliert²⁷:

Sprecherintention > Kontext > Semantik > Wortstellung

Die Hierarchie gibt der Sprecherintention die höchste Priorität, gefolgt von kontextuellen Faktoren wie etwa 'neue Information'. Der Faktor der Semantik basiert in diesem Zusammenhang auf der Unterscheidung zwischen den semantisch 'schweren' Inhalts- und den semantisch 'leichten' Funktionswörtern. Die Inhaltswörter tragen im allgemeinen den Satzaccent. Der Wortstel-

²⁵ Nach Jacobs (1984: 46) wird etwa in dem Satz Was Peter betrifft, so wird er dieses Jahr nicht kommen der Interpretationsrahmen durch die Anweisung 'Beziehe das Folgende auf Peter' gekennzeichnet.

²⁶ Eine aktuellere Diskussion dieser linguistischen Konzepte bietet von Heusinger (1999); vgl auch die verschiedenen Beiträge in Bosch/van Sandt (1999).

²⁷ Intentionalität wird in der Regel im Sinne der Sprechakttheorie von Searle (1969) aufgefaßt; an die Gricesche Konversationstheorie (vgl. z.B. Levinson (2000)) schließt Lambrecht (1994: 291) an: „To conclude, contrastiveness, unlike focus, is not a category of grammar but the result of the general cognitive process referred to as 'conversational implicatures'.“

lungsfaktor erfasst intonatorisch analysierbare Abweichungen von der sog. normalen (unmarkierten) Wortstellung, die zugleich mit einer normalen Intonation verbunden ist²⁸.

Der experimentelle Teil der vorliegenden Arbeit untersucht akustische Korrelate von Satzakkzentrealisierungen in verschiedenen einfachen Aussagesätzen mit einer markierten und einer unmarkierten Wortstellung sowie in Ellipsen. Die experimentelle Elizitierung der Satzakkzentrealisierungen ist durch den Kontext gesteuert, wobei Satzakkzentrealisierungen lediglich auf Inhaltswörtern intendiert sind; vgl. die in Anhang A, S. 105ff, aufgelisteten Stimuli.

5.1 Fokusakzente

Der Fokus kann in der syntaktischen Struktur eines Satzes unterschiedlich komplexe Einheiten einschließlich des ganzen Satzes umfassen. Bei einzelnen Wörtern handelt es sich nach Uhmann (1991) um einen engen, bei komplexen Konstituenten um einen weiten Fokus. Die verschiedenen Konstituenten sind der Fokusbereich, in dem das akzentuierte Wort den Fokusexponenten bildet. In komplex fokussierten Konstituenten erfolgt die genaue Bestimmung der den Satzakkzent tragenden Fokussilbe durch Fokusprojektionsregeln. Bei vorausgesetztem weitem Fokus ermitteln diese Regeln den Fokusexponenten, dessen Iktussilbe den Fokusakzent erhält. Ein Fokusakzent besteht aus mindestens einem Akzentton. Bei einem engen Fokus ist einfach der jeweiligen Iktussilbe des entsprechenden Wortes ein Akzent zuzuweisen. Eine genauere, weiter in die Syntax des Deutschen führende Beschreibung der Fokusprojektionsregeln unterbleibt an dieser Stelle²⁹.

Nach Féry (1993) ist die Realisierung eines Akzents innerhalb des Fokus eines Satzes mit der Bildung von phonologischen Fokusdomänen (FD) verbunden, wobei jedem Akzent eine eigene Domäne zugeordnet wird. Integrierte Fokusdomänen fassen zwei oder mehr fokussierte Konstituenten eines Satzes zu einer phonologischen Einheit zusammen. Bei der Bildung von isolierten Fokusdomänen werden die fokussierten Konstituenten in einzelne Domänen unterteilt. Die Wahl zwischen integrierten und isolierten Fokusdomänen kann von verschiedenen kulturellen, stilistischen, syntaktischen oder semantischen Bedingungen abhängen. Ein Beispiel einer kulturellen Bedingung ist nach Féry (ebd: 31) in dem folgenden Beispiel gegeben:

- (a) Alain ist nicht da. Ich glaube, dass er [F[FD im GARTEN] [FD SCHAUKELT]].
- (b) Alain ist nicht da. Ich glaube, dass er [F[FD im GARTEN spielt]].

Die isolierten Fokusdomänen in (a) basieren nach Féry im Gegensatz zu der integrierten Fokusdomäne in (b) auf der Tatsache, dass nach Féry Schaukeln eine weniger normale und vor-

²⁸) Delin/Zacharski (1997) implementieren die Faktoren in das Sprachsynthesystem BRIDGE, das Dialoge in der Domäne 'Routenplanung' simuliert. Die Implementierung zeigt, dass informationsstrukturelle Informationen deutlich zu der Verbesserung der Generierung des Tonhöhenumfangs ('pitch range') und der Satzakkzentplatzierung bei der Sprachausgabe beitragen.

²⁹) Jacobs (1983) untersucht den Zusammenhang zwischen dem Auftreten von Gradpartikeln und Akzentuierung. Mayer (1997) rekurriert auf Schubiger (1965) und sieht in dem Gradpartikelinventar des Deutschen eine Begründung der von Féry (1993) vertretenen Ansicht, dass das Intonationssystem des Deutschen ärmer als das des Englischen sei. Nach Mayer (ebd.: 19/20) verfügt das Deutsche „über ein reicheres Inventar an Gradpartikeln [...], so daß differenzierte Bedeutungsunterschiede, die in anderen Sprachen bevorzugt mit intonatorischen Mitteln ausgedrückt werden, im Deutschen auch mit anderen linguistischen Mitteln ausgedrückt werden.“

hersehbare Aktivität eines achtjährigen Jungen ist. Spielen im Gegensatz dazu eine normale und vorhersehbare Aktivität³⁰.

In vollständig fokussierten Sätzen, d.h. die als Ganze einen Fokus bilden, ist die Art der Fokusdomänenbildung in vielen Fällen optional. In manchen Sätzen gibt es jedoch syntaktische und semantische Bedingungen, die diese Option einschränken. So trägt in einfachen Sätzen bestehend aus einem Subjekt und einem Prädikat im allgemeinen das Prädikat den Nuklearakzent. Das Subjekt ist in diesem Fall ebenfalls akzentuierbar, ohne dass sich die Fokusstruktur ändert. Eine Realisierung des Nuklearakzents erfolgt auf dem Subjekt, wenn durch das Verb des Prädikats der Hörer auf die Existenz eines Individuums hingewiesen wird oder wenn das Verb mehr oder weniger semantisch leer ist (Féry (1993: 32/33)):

(c) (Hörst du das?) [F[FD Ein HUND bellt]].

(d) (Was machst du denn für ein Gesicht?) [F[FD Meine BRIEFtasche ist weg].

Das Subjekt kann ebenfalls akzentuiert sein, wenn es sich um ein 'ergatives Subjekt' handelt. In Passivsätzen sind die Oberflächensubjekte im Sinne der Generativen Grammatik stets ergativ und somit Träger des Nuklearakzents. Zudem treten 'ergative Subjekte' bei Verben auf, die ihr Perfekt mit dem Hilfsverb „sein“ bilden.

In Verbalphrasen mit einer Ergänzung wird diese akzentuiert, wobei wiederum auch die Akzentuierung des Subjekts möglich ist. In die Verbbedeutung zu inkorporierende Elemente (s.u. flachhämmern), die vom Verb getrennt auftreten, sind nicht akzentuiert (e). Ferner können freie Angaben eine eigene Fokusdomäne bilden (f). Dies gilt auch bei bestimmten Lesarten von Prädikaten (g) und bei einigen Attributen (h) (Féry (ebd.: 36/38)):

(e) [F Ernst hämmert [FD die HUNDEMARKE flach]].

(f) [F Wir sind [FD drei STUNDEN] [FD GEFAHREN]].

(g) Weil fast [FD alle FLÜCHTLINGE] [FD in dieser STADT umgekommen sind].

(h) Es ist [drei GRAMM] [GOLD].

Die Lesart eines Satzes kann in bestimmten Satztypen durch Platzierung des Nuklearakzents auf dem Subjekt determiniert werden (Féry (ebd.: 42))³¹:

(i) HUNDE müssen getragen werden/ HUNDE werden getragen

(j) HUNDE müssen GETRAGEN werden/ HUNDE werden GETRAGEN

Die Sätze in (i) sind als vollständig fokussiert aufzufassen. Sie erhalten eine 'globale', thetische Interpretation, in der die Sätze ein ganzes Ereignis ausdrücken. Die Sätze in (j) hingegen paraphrasiert Féry (ebd.: 43) durch: "should any dogs be contextually present, then you must carry them". Der Akzent auf dem Subjekt bewirkt die Bildung einer Hintergrundkonstituente, die man vergleichbar zu fokussierten Konstituenten auch als fokussiert bezeichnen kann. In dieser Art können alle nicht finalen Akzente interpretiert werden. Die weiterhin mögliche alleinige Akzentuierung des Prädikats bestimmt eine Lesart, die die tatsächliche Präsenz von Hunden voraussetzt.

³⁰) An dieser Stelle sei darauf hingewiesen, dass Féry keine experimentellen Belege für ihre phonologischen Analysen liefert.

³¹) Damit handelt es natürlich nicht mehr um einen Nuklearakzent.

5.2 Der Brückenakzent

Den tonalen Aufbau des Brückenakzents bezeichnet Sappok (1989: 431) „als Lokalisationsveränderung der Kontur im Sinne eines zeitlich erweiterten Dynamikbereichs“. Die Lokalisationsveränderung ist nach Sappok (ebd.) mit den beiden folgenden, für die Analyse des Brückenakzents zentralen Fragen verbunden³²:

- (a) „Handelt es sich um zwei Betonungen innerhalb eines Satzes?“
- (b) „Handelt es sich um die Genese einer von der ursprünglichen Satzbetonung unabhängigen Hervorhebung, wodurch die Position des ursprünglichen Satzakzents neutralisiert wird?“

5.2.1 Kontrastbildungen

Wunderlich (1991) beschreibt den Brückenakzent als Kombination von zwei einzelnen, auf unterschiedlichen einfachen Konstituenten einer syntaktischen Einheit auftretenden Akzenten. Die beiden Konstituenten gehören zu einer Intonationsphrase, die nicht mit der syntaktischen Struktur koinzidieren muss. Die analytische Trennung der beiden Akzente beruht auf einem semantischen Konzept des Kontrasts, das von einer Gegenüberstellung möglicher Alternativen in der Kommunikationssituation ausgeht³³. Bei der Kontrastierung der Bedeutungen der entsprechend akzentuierten Ausdrücke können die komplementären Bedeutungselemente entweder direkt in derselben syntaktischen Einheit oder in dem Kontext, in dem eine solche Einheit verwendet wird, auftreten. Der entscheidende Punkt der Analyse des Brückenakzents von Wunderlich ist, dass die beiden Akzente eine intonatorische Einheit bilden, der als ganzer eine Bedeutung im Sinne einer Kontrastbildung zugeordnet wird. Nach Wunderlich (1991: 242f) können Kontraste in ganz unterschiedlichen Konstruktionen auftreten:

- (1) (Weißt du schon das Neueste?) Der BUNdeskanzler ist zuRÜCKgetreten? (kontextlose Äußerung)
- (2) ErLAUBT ist, was geFÄLLT. (Sprichwörter)
- (3) Peter KOMMT, Anna GEHT. (nebeneinanderliegende Phrasen)
- (4) Willst du TEE oder KAFfee? (Entscheidungsfragen)
- (5) DREIzehn, VIERzehn, FÜNFzehn, SECHzehn. (Listen)
- (6a) (zwei mal DREI) plus VIER. (unterschiedliche syntaktische Klammerungen)
- (6b) ZWEI mal (DREI plus VIER).
- (7) WANN (**g**laubst du (dass Peter **g**esagt hat (dass Käte **k**ommt))). (unterschiedliche syntaktische Extraktionen (Fettdruck für alternative Akzentuierungen))

Der Sprecher von (1) wählt mit dem Subjekt ein bestimmtes Topik und anschließend zu diesem ein Prädikat aus einer Menge von möglichen Alternativen aus. Die Akzentuierung der entsprechenden Ausdrücke bewirkt deren Kontrastierung in ihrer spezifischen thematischen Konstruktion von Argument und Prädikat. Es ist nach Wunderlich ein allgemeines Merkmal der sprachlichen Kommunikation, mögliche Alternativen sukzessive durch die Wahl bestimmter Aus-

³² Man beachte, dass bei der Bestimmung des Brückenakzents im Gegensatz zu den Fokusakzenten der Formaspekt favorisiert wird.

³³ In Kapitel 6.1.2 werden Bedeutungen des Brückenakzents vorgestellt, die das Kieler Intonationsmodell annimmt.

drücke auszuschließen. Die einzelnen Ausdrücke werden dann durch das Mittel der Kontrastierung markiert. In diesem Sinn ist auch (2) zu interpretieren, wobei Sprichwörter eine vergleichbare kontextuelle Ungebundenheit wie kontextlos geäußerte Sätze ('out of the blue') haben.

In (3) kontrastieren bestimmte Personen und deren Aktivitäten. In (4) sind es die beiden Alternativen der Entscheidungsfrage und in (5) kontrastieren die nicht-letzten Listenelemente mit dem letzten Element der Liste. Die Klammerungen von (6a, b) lassen sich interpretieren als Kontrastierung von Elementen einer zweielementigen Liste, wobei das erste in (6a), bzw. das letzte in (6b) Element wiederum als eine Liste aufgefaßt werden kann. Die Klammerungen können zusätzlich durch intonatorische Mittel wie linguistische Pausen und Dehnungen am Ende einer Intonationsphrase indiziert werden. Zudem zeigt die tonale Kontur auf dem ersten Numeral an, ob eine links- oder eine rechtsverzweigende Struktur folgt. Bei einer linksverzweigenden Struktur wird auf dem ersten Numeral ein fakultativer monotonaler Hochton realisiert und bei einer rechtsverzweigenden Struktur ist dem Numeral (und der folgenden Silbe) eine Folge von zwei Hochtönen zugewiesen. Die extrahierte Phrase in (7) kontrastiert mit dem Bereich ihrer Spur (im Sinne der Generativen Grammatik). Die an diesen Beispielen erläuterte Konzeption von Kontrast umfaßt auch die Konzeption von Fokus. Der Fokus im Sinne von neuer Information eines Satzes läßt sich auffassen als Kontrast zu anderen möglichen neuen Informationen des Kontextes.

Wunderlich (1991: 246) zerlegt eine (potentielle) Äußerung in Kontrastphrasen (K) der aufgezeigten Art, von denen der Brückenakzent tonal abgeleitet wird. Die beiden kontextsensitiven Ersetzungsregeln i), ii) erzeugen gemeinsam den Brückenakzent in dem Sinne, dass die Zuweisung der Tonfolge 'HHL' an Fokussilben (s*) das tonale Muster eines entsprechend zu äussernden Satzes symbolisiert. Wenn Regel a) zwei benachbarten K1-Phrasen jeweils zwei Hochtöne zuweist, dann muss nach einer Tonsandhi-Regel bei der zweiten K1-Phrase eine neue Intonationsphrase mit einem Tieftone (L) als initialen Grenzton gebildet werden. Bei mehreren kontrastierenden Elementen, wie z.B. in Listen oder in Entscheidungsfragen mit mehr als zwei Alternativen, erzeugt Regel i) zunächst eine Folge von linken Brückenpfeilern und Regel ii) komplettiert abschließend einen Brückenakzent:

i) s* --> s* /__]K1	ii) s* --> s* /H__
HH	L

5.2.2 Topikalisierungen

Jacobs (1997) führt den Begriff der I-Topikalisierung für intonatorisch markierte Topik-Kommentar-Strukturen ein. Topikalisierte Sätze sind potentielle Träger des Brückenakzents. Jacobs (ebd.: 92) nennt folgendes Beispiel einer prototypischen I-Topikalisierung:

(A1) Kann man denn alle Romane von Grass empfehlen?

(B1) Na ja, √ALle kann man sicher \NICHT empfehlen (, aber sein ERSter ist zweifellos ein MEISterwerk).

Die I-Topikalisierung kann nur in Sätzen auftreten, die als Assertion oder als Anweisung interpretiert werden können. In Fragesätzen ist I-Topikalisierung nur möglich, wenn es sich um eine Echofrage im Kontext einer assertiven I-Topikalisierung handelt. Lexikalisch können die beiden Akzente einer I-Topikalisierung auf beliebigen Elementen auftreten. Die Einschränkung besteht darin, dass zu dem akzentuierten Element eine inhaltliche Alternative bestehen muß. Aus diesem Grund enthalten die beiden folgenden Sätze nicht eine mögliche Wurzelkontur, d.h. in a) ist keine Alternative zu also denkbar und in b) darf der von dem Verb regierte syntaktische Status der Präposition nicht verletzt werden³⁴:

In dem nicht geklammerten B-Teilsatz liegt auf alle ein tonaler Verlauf der Form L*H vor, den Jacobs als Wurzelkontur bezeichnet. In (A1), (B1) ist die Wurzelkontur mit einem Häkchen transkribiert. Der Schrägstrich steht dann für einen fallenden tonalen Verlauf (Grossbuchstaben kennzeichnen wie üblich einen Akzent). Die beiden interessierenden Akzente sind rhythmisch in etwa gleich stark. In dem geklammerten Teilsatz wird nach Jacobs (ebd.: 92) „zu jedem hervorgehobenen Satzteil eine inhaltliche Alternative genannt, nämlich hier 'sein erster' als Alternative zu 'alle' und die Affirmation als Alternative zur Negation (auf dem Umweg über die Folgerung aus 'ein Meisterwerk').“ Die zweite Hervorhebung markiert im Gegensatz zur ersten neue Information, wie sich aus dem A-Satz ergibt. Der erste Teilsatz ist so zu verstehen, dass nicht jeder Roman von Grass zu empfehlen ist.

a) *√ALso hat er ihn \NICHT verrissen.

b) *√VIEle Kritiker haben \NACH Gründen für Grass' Versagen gesucht.

Eine Ausnahme der Forderung nach möglichen Alternativen besteht in verbalen Phraseologismen, wie der folgende Dialog A2-B2 von Jacobs (ebd.: 96) belegen soll:

(A2) Wie wird Grass wohl auf die schlechten Kritiken reagieren?

(B2) Nun, die √HAAre wird er sich \NICHT gerade raufen (, aber ein bisschen ÄRgern wird er sich schon).

³⁴) Das Sternzeichen kennzeichnet an dieser Stelle ungrammatische Sätze.

6. Phonetische Intonationsmodelle

Die im folgenden behandelten phonetischen Intonationsmodelle (des Deutschen) beschreiben intonatorisch relevante Eigenschaften der F₀-Verläufe von sprachlichen Äußerungen. Die intonatorisch relevanten Eigenschaften werden nach funktionalen, perzeptiven und/oder artikulatorischen Kriterien ermittelt. Kapitel 6.1 bis 6.3 behandeln etablierte phonetische Modelle der Intonation des Deutschen; Kapitel 6.4 bis 6.5 haben dagegen Modelle zum Gegenstand, von denen bisher keine systematische Anwendung auf das Deutsche vorliegt. Die Modelle der Kapitel 6.1 bis 6.5 werden in dem experimentellen Teil der vorliegenden Arbeit auf konkretes Material angewendet³⁵. Kapitel 6.6 stellt einige weitere Modelle vor, die in dem experimentellen Teil der vorliegenden Arbeit allerdings unberücksichtigt bleiben. Die in dem experimentellen Teil untersuchten Modelle werden am Ende dieses Kapitels diskutiert.

6.1 Das Kieler Intonationsmodell

Das in Kohler (1991) für das Deutsche entwickelte Konturmodell KIM ('Kieler Intonationsmodell') beschreibt die Intonation in F₀-Verläufen als Gipfel- und Talkonturen. Die Gipfel- und Talkonturen bilden einen integralen Bestandteil des gesamten F₀-Verlaufs einer Äußerung, wobei die zeitliche Positionierung der Gipfel und Täler im Silbenkern phonologisch distinktiv ist. Damit ist KIM ein phonologisches Intonationsmodell.

Die phonetische Klassifizierung des Modells beruht in dem vorliegenden Zusammenhang auf der akustisch-phonetischen Spezifikation der phonologischen Eigenschaften von KIM, d.h. auf der zeitlichen Positionierung der F₀-Gipfel und -Täler bzgl. des Silbenkerns. Im folgenden werden nur die Intonationsmerkmale von KIM vorgestellt; andere Merkmale wie die Einordnung des Modells in ein Grammatikmodell oder Betonungsmerkmale bleiben unberücksichtigt. KIM ist, wie bereits erwähnt, der Britischen Schule zuzuordnen, wobei Kohler explizit auf Halliday (1967) rekurriert. Kohler arbeitet allerdings nicht wie Halliday mit distinktiven Tonmustern bestehend aus Prä nukleus und Nukleus, sondern eben mit Gipfel- und Talkonturen.

6.1.1 Gipfel- und Talkonturen

Eine Gipfelkontur kann nach KIM mit einem einfachen Abfall enden oder anschließend noch einen Anstieg aufweisen. Bei einem frühen Gipfel erreicht die F₀ des Vokals schnell ihren tiefsten Punkt, so dass der Eindruck einer eher tiefen Tonhöhe für den ganzen Vokal entsteht. Ein mittlerer Gipfel bedingt einen tiefen F₀-Verlauf zu Beginn des Vokals, dem dann der Anstieg des Gipfels folgt. Dadurch entsteht der Eindruck einer eher hohen Tonhöhe für den ganzen Vokal. Die beiden Gipfel bilden somit zwei phonologische Kategorien. Der späte Gipfel bewirkt die Dehnung des vorherigen niedrigen Verlaufs und hat im Vergleich zu den anderen Gipfeln den größeren Anstieg. Er ermöglicht nicht die kategoriale Abgrenzung gegenüber dem mittleren Gipfel, sondern unterscheidet sich von ihm nur graduell.

³⁵ Das Fujisaki-Modell wird ohne die Erweiterung von Mixdorff (1998) verwendet; vgl. Kap. 6.3.3, S. 44f.

Ein früher Gipfel ohne anschließenden Anstieg signalisiert die Entschiedenheit, die ein Sprecher mit seiner Aussage verbindet. Ein langsamerer Abfall schwächt die Entschiedenheit ab. Der frühe Gipfel kann auch kontextuell bekannte Information anzeigen. Ein mittlerer oder später Gipfel ohne anschließenden Anstieg signalisiert dagegen die Bereitschaft des Sprechers zur Diskussion seiner Aussage oder tritt bei neuer Information auf. Die Gipfel können auch einen Kontrast oder eine Emphase indizieren. Ein langsamerer Abfall bewirkt wiederum eine graduelle Abschwächung. Wenn dem Abfall eines Gipfels ein weiterer Anstieg folgt, dann kann damit die Bereitschaft des Sprechers zu weiteren Ausführungen oder die Aufforderung des Hörers dazu angezeigt werden.

Die Realisierung eines Tales in einer frühen im Gegensatz zu einer mittleren oder späten Position ist nicht mit einem kategorialen Wechsel verbunden. In einer tendenziell steigenden Kontur eines Vokals bewirkt die Verschiebung des Tiefpunktes eines Tales von einem frühen zu einem späteren Zeitpunkt nicht die kategoriale Wahrnehmung einer fallenden Tonhöhe. Die Verschiebung von einem Tal ist mit einer semantischen Differenzierung verbunden, die die von dem Sprecher ausgedrückten Emotionen oder Interessen betrifft. Ein frühes Tal drückt eine gewisse Beiläufigkeit oder ein Desinteresse aus. Ein spätes Tal tritt nach Kohler etwa in Fragen an Kinder auf, wobei die ausgedrückten Emotionen etwas unnatürlich wirken können. Die Höhe des Anstiegs nach einem Tal drückt wiederum im positiven Sinne die Offenheit des Sprechers für die Meinung des Hörers aus. Somit wird ein Tal mit folgendem Anstieg im allgemeinen in Entscheidungsfragen verwendet, bei denen der Hörer mit seiner Antwort keiner besonderen Obligation unterliegt. Einige der Konnotationen von Gipfeln und Tälern werden an dem folgenden Beispiel aufgezeigt (ebd.: 311):

Wie HEIßT du?

Ein früher Gipfel bei dem Vokal der akzentuierten Silbe indiziert eine Antwortobligation. Ein mittlerer Gipfel ist mit einer Forderung nach weiterer Information verbunden, ohne jedoch den Hörer direkt aufzufordern. Mit einem späten Gipfel wird nach Kohler dagegen ein Kontrast der Position und der Ansichten des Sprechers in dem gegebenen Kontext ausgedrückt. Folgt diesen Gipfeln ein weiterer Anstieg, so ist dies zum einen eine Aufforderung an den Hörer fortzufahren. Zum anderen unterstreicht der Sprecher damit seine Autorität gegenüber dem Hörer.

Ein frühes tiefes Tal signalisiert Faktizität, aber auch eine Beiläufigkeit oder Indifferenz gegenüber dem Hörer. Gleichzeitig ordnet sich der Sprecher damit dem Hörer unter. Die Indifferenz liegt bei einem hohem, langsam steigenden Tal nicht vor. Ein spätes tiefes Tal steht wieder für die Gleichgültigkeit des Sprechers gegenüber dem Hörer. Es ist aber auch Ausdruck der Emotionalität des Sprechers. Bei einem langsamen hohen Anstieg von diesem Tal wird wiederum Aufmerksamkeit gegenüber dem Hörer angezeigt.

6.1.2 Kombination von Gipfelkonturen

Die Kombination zweier Gipfel ohne intermediären fallenden oder steigenden F₀-Verlauf bildet einen Brückenakzent, wobei nach Kohler (ebd.: 322f) folgende Formen möglich sind:

- (a) mittlerer oder später erster Gipfel plus mittlerer zweiter Gipfel
- (b) mittlerer oder später erster Gipfel plus früher zweiter Gipfel

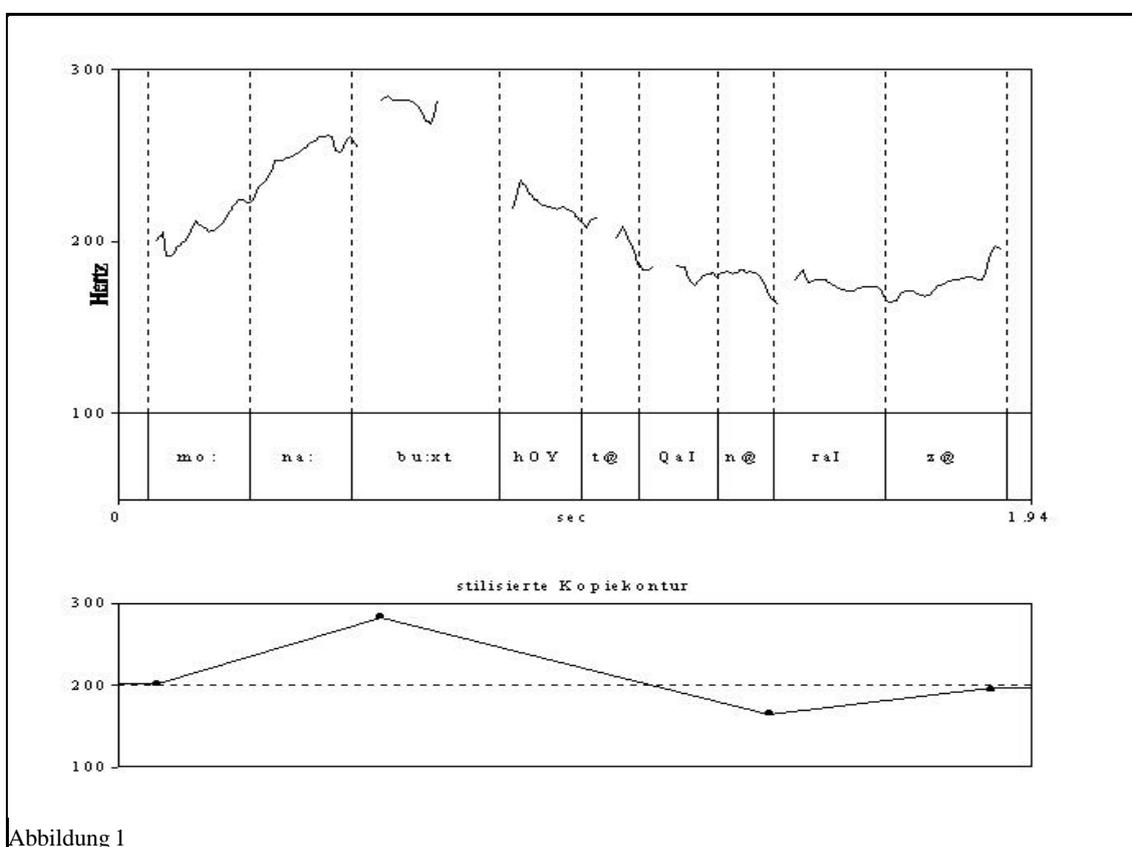
Die semantischen Funktionen von Brückenakzenten diskutiert Kohler (ebd.: 327) an dem folgenden Beispiel:

- (A) Was machen Phonetiker? Was machen Linguisten?
- (B) Phonetiker untersuchen Laute; Linguisten untersuchen Syntax.

Die betonten Silben von Phonétiker und Linguísten werden mit mittleren Gipfeln und die betonten Silben von Láute und Syntax mit frühen Gipfeln realisiert. Ein Brückenakzent entsteht auf dem ersten Teil der Antwort. Die beiden Gipfel des zweiten Teils sind hingegen deutlich voneinander getrennt. Der Brückenakzent zeigt an, dass das in der Frage vorgegebene Thema vom Sprecher akzeptiert und mit der Antwort allgemeines oder bekanntes Wissen wiedergegeben wird. Die separaten Gipfel des zweiten Teils der Antwort indizieren einen Kontrast zwischen Linguisten und Phonetikern. Wenn der letzte Gipfel in mittlerer Position realisiert werden würde, dann hätte das einen Bedeutungswechsel von der Behauptung einer allgemein bekannten Tatsache zu der Behauptung einer neuen Tatsache zur Folge. Die Verschiebung der mittleren Gipfel der Äußerung in eine späte Position bewirkt eine leichte Opposition zu der Frage in dem Sinne, dass (B) denkt, es gehe darum, was Phonetiker eigentlich machen.

6.2 Das IPO-Modell

't Hart/Collier/Cohen (1990) beschreiben für das Holländische die Methode der 'straight line approximation' des IPO-Modells (Modell des 'Instituut voor Perceptie Onderzoek' in Eindhoven³⁶), die Adrians (1991) auf das Deutsche anwendet. Die Methode ermittelt die perzeptiv relevanten Abschnitte eines F0-Verlaufs, indem sie auf einer logarithmischen Skala zwischen den Wendepunkten des Verlaufs linear interpoliert. Der linear interpolierte Verlauf geht in eine Sprachsynthese ein, die von der originalen Äußerung eine Variante mit dem linear interpolierten F0-Verlauf erzeugt. Original und synthetische Variante werden nach ihrer hörbaren Gleichheit bzw. Ungleichheit im phonetischen Sinne beurteilt. Zwei Geraden werden durch Löschen des gemeinsamen Wendepunkts zu einer neuen Geraden zusammengefasst, wenn dadurch kein hörbarer Unterschied zu dem Original entsteht. Die damit entstehenden sog. stilisierten Kopiekonturen sollen ein Minimum an Geraden, also ein Maximum an Abstraktion aufweisen; vgl. Abbildung 1 zur Illustration des Modells.



Die Methode der 'straight line approximation' beruht auf der Annahme von perzeptiv irrelevanten F0-Bewegungen. Eine Abstraktion von diesen Bewegungen besagt nicht, dass die Kopiekonturen auch die funktionalen Eigenschaften von Intonationsverläufen analysieren. Adrians kategorisiert und standardisiert allerdings die ermittelten Kopiekonturen, so dass von diesen Konturen ein melodisches Modell der deutschen Intonation (für vorgelesenen Text) entwickelt wer-

³⁶) s.a. Fußnote 1, S. 9.

den kann³⁷. In einem Perzeptionstest wird die 'perzeptive Akzeptabilität' des Modells nachgewiesen. Die F0-Verläufe des Modells nennt Adrians standardisierte Tonhöhenbewegungen. Das melodische Modell enthält auch standardisierte Tonhöhenbewegungen für akzentuierte Silben; diese Verläufe können funktional interpretiert werden.

6.2.1 Melodiekonturen des Deutschen

Das melodische Modell enthält eine standardisierte Deklinationslinie - die Basisdeklinationslinie - für (i) äßerungsfinale und eine für (ii) nicht-äßerungsfinale Intonationsverläufe. Die beiden Deklinationslinien berechnen sich nach den folgenden Formeln, wobei D für die Deklinationsgeschwindigkeit in HT (Halbtönen) pro Sekunde und t für die Dauer einer Äßerung steht (ebd.: 60, 75):

(i) für $t \geq 5 \text{ sec}$, $D = -8.5/t$

(ii) für $t \geq 5 \text{ sec}$, $D = -6.1/t$

Von der Basisdeklinationslinie werden drei weitere, übereinander liegende Deklinationslinien in einem Abstand von 7.5 HT, 2.5 HT und noch einmal 2.5 HT berechnet. Insgesamt erfassen die Deklinationslinien also einen Bereich von 12.5 HT. Auf diesen Bereich werden die standardisierten Tonhöhenbewegungen von Adrians aufgeteilt, die nach ihrer Position, ihrem Frequenzumfang in HT und ihrer Dauer in Millisekunden (ms) spezifiziert sind. Als Positionen werden Vokalanfang (VA) und Ende des stimmhaften Bereichs der entsprechenden Silbe (EST) als Referenzpunkte angesetzt (Beginn der Bewegung). So kann etwa eine steigende Tonhöhenbewegung -210 ms, -60 ms, -30 ms vor dem Vokalanfang oder genau auf dem Vokalanfang bei der Position 0 beginnen. Tabelle 1 enthält die elf standardisierten Tonhöhenbewegungen des melodischen Modells, wobei die Tonhöhenbewegungen (1) - (6) und (8) bei akzentuierten Silben vorkommen. (11) tritt bei einer prosodischen Grenze auf, wobei die Bewegung nach der Grenze fortgesetzt wird.

³⁷) Brindöpke/Schaffranietz (1999) untersuchen die von Adrians (1991) entwickelten Standardkonturen an spontansprachlichen Material und bestätigen deren Adäquatheit.

Nr.	Referenzpunkt	Position	Dauer	Umfang	Form
(1)	VA	-210 ms	180 ms	7.5 HT	Steigung
(2)	VA	-60 ms	180 ms	7.5 HT	Steigung
(3)	VA	-30 ms	60 ms	2.5 HT	Steigung
(4)	VA	0 ms	180 ms	7.5 HT	Steigung
(5)	VA	0 ms	180 ms	7.5 HT	Fall
(6)	VA	60 ms	240 ms	10 HT	Fall
(7)	VA	120 ms	variabel	7.5 HT	Fall
(8)	VA	150 ms	180 ms	7.5 HT	Steigung
(9)	EST	- 180 ms	180 ms	7.5 HT	Steigung
(10a)	EST	- 300 ms	300 ms	12.5 HT	Steigung
(10b)	EST	- 120 ms	120 ms	5 HT	Steigung
(11)	EST	0 ms	variabel	7.5 HT	Fall

Tabelle 1. Standardisierte Tonhöhenbewegungen

Die Abfolge der einzelnen Tonhöhenbewegungen in einer Kontur wird von Adrians mit Regeln ihrer möglichen Kombinationen erzeugt. Die Regeln gelten für einzelne Intonationsblöcke, in denen verschiedene Konturtypen aufgebaut werden. Insgesamt gibt es fünf Intonationsblöcke (I), die wiederum in komplexen Intonationsverläufen kombiniert auftreten können. Abbildung 2 gibt die Kombinationsmöglichkeiten der standardisierten Tonhöhenbewegungen nach Adrians (ebd.: 68) in Diagrammform wieder, wobei jeder Intonationsblock (a) – (e) eine Kontur erzeugt.

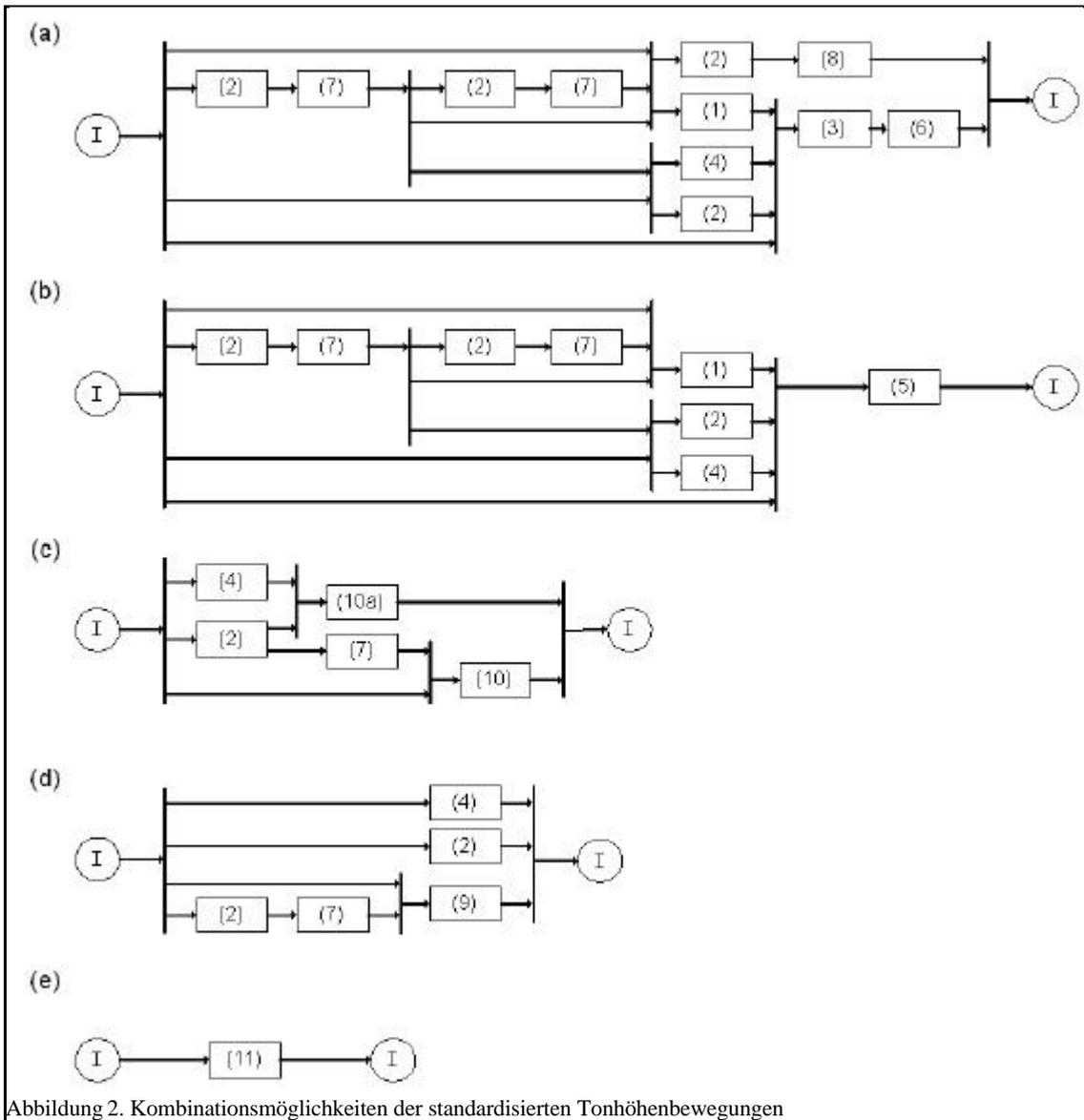


Abbildung 2. Kombinationsmöglichkeiten der standardisierten Tonhöhenbewegungen

6.2.2 Automatische Stilisierung

d'Allessandro/Mertens (1995) entwickeln ein Verfahren der automatischen Stilisierung von F0-Verläufen, das auf Annahmen des IPO-Modells aufbaut. Das Verfahren zerlegt einen F0-Verlauf nach perceptiven Kriterien in eine Folge von tonalen Segmenten. In Kapitel 8.2, S. 85f, wird das Verfahren zur Analyse der standardisierten Tonhöhenbewegungen von Adrians (1991) verwendet. Nach d'Allessandro/Mertens (1995: 262ff) werden in dem Verfahren die Schritte a) und b) für jeden Silbenabschnitt eines F0-Verlaufs durchgeführt:

- a) Anwendung der Funktion $p(t)$, die den F0-Verlauf vor dem Hintergrund der Annahme gewichtet, dass der finale Abschnitt eines tonalen Verlaufs größeren Einfluss auf die Wahrnehmung des Verlaufs hat als der initiale Abschnitt; $p(t)$ ist die zum Zeitpunkt t wahrgenommene Tonhöhe und $\alpha = 22$ ist empirisch gesetzt (die Funktion arbeitet in einem Zeitfenster von 140 ms).

$$p(t) = \frac{\left(\int_0^t e^{-\alpha(t-\tau)} f(\tau) d\tau \right)}{\left(\int_0^t e^{-\alpha(t-\tau)} d\tau \right)}$$

- b) Bildung von tonalen Segmenten, d.h. Einteilung des F0-Verlaufs in einfache ebene, steigende und fallende Tonhöhenverläufe sowie in komplexe Verläufe mit einem Richtungswechsel, also etwa steigend-zu-fallend. Zu diesem Zweck wird der F0-Verlauf logarithmisch skaliert und anschließend auf der Basis der folgenden Parameter in tonale Segmente eingeteilt:

i) ein F0-Verlauf korrespondiert nach Hart/Collier/Cohen (1990: 32f) mit einem wahrgenommenen steigenden oder fallenden Tonhöhenverlauf, wenn sich der F0-Verlauf bei einer Länge von $t = 50$ ms um $HT = 64$ Halbtöne pro Sekunde ändert; HT ist bei einer Verdoppelung von t jeweils durch vier zu dividieren. Wenn ein F0-Verlauf nicht diesem Wert entspricht, dann korrespondiert der F0-Verlauf mit einem ebenen Tonhöhenverlauf.

ii) ein in i) gebildetes tonales Segment enthält einen Richtungswechsel, wenn eine Differenz in unterschiedlichen Gefällen des tonalen Segments von mindestens 12 HT besteht. Als Wendepunkt dient dabei der F0-Wert, der am weitesten von einer zwischen dem ersten und dem letzten F0-Wert des tonalen Segments gezogenen Gerade entfernt ist.

6.3 Das Fujisaki-Modell

Fujisaki (1988) entwickelt sein Superpositionsmodell ursprünglich für die Analyse und Synthese von F0-Verläufen des Japanischen. Das Modell versteht sich als Produktionsmodell, das die entsprechenden Prozesse beim Sprechen nachbildet³⁸. In dem vorliegenden Zusammenhang interessieren ausschließlich die akustischen Merkmale des Modells. Allgemein lässt sich das Modell als ein mathematisch formuliertes, hierarchisch aufgebautes Intonationsmodell bestimmen, das Intonationsverläufe als Resultat der additiven Überlagerung von untereinander autonomen Komponenten beschreibt. Die analytische Trennung einzelner Komponenten ermöglicht dabei deren selektive Verbindung mit bestimmten segmentalen oder linguistischen Einflussgrößen.

6.3.1 Der Steuermechanismus

Das Modell von Fujisaki ist als lineares System formuliert. Das System berechnet aus zwei Eingangssignalen in additiver Verknüpfung ein Ausgangssignal, dessen fortgesetzte Konkatenation den F0-Verlauf berechnet. Als Eingangssignale erhält das Modell Phrasierungskommandos in Form von Impulsfunktionen und Akzentkommandos in Form von Rechteckfunktionen. Die Eingangssignale werden von der Phrasenkomponente mit den Mechanismen zur Steuerung der Phrasierung und von der Akzentkomponente mit den Akzentsteuermechanismen verarbeitet. Die Ausgangssignale dieser beiden Mechanismen addiert das System zu dem kleinsten asymptotischen Wert (Basiswert F_{\min}) des zu analysierenden bzw. zu erzeugenden F0-Verlaufs. Möbius (1993) wendet das Fujisaki-Modell auf das Deutsche an und übernimmt die folgende Modellierungsfunktion (ebd.: 69):

$$\ln F_0(t) = \ln F_{\min} + \sum_{i=1}^I A_{pi} \mathbf{6} G_{pi}(t - T_{0i}) + \sum_{j=1}^J A_{aj} \mathbf{6} G_{aj}(t - T_{1j}) - G_{aj}(t - T_{2j})$$

mit

$$G_{pi}(t) = \alpha_i^2 t \mathbf{6} \exp(-\alpha_i t), \text{ für } t > 0$$

$$G_{pi}(t) = 0, \text{ für } t \leq 0$$

und

$$G_{aj}(t) = 1 - (1 + \beta_j t) \mathbf{6} \exp(-\beta_j t), \text{ für } t > 0$$

$$G_{aj}(t) = 0, \text{ für } t \leq 0$$

wobei

- I = Anzahl der Phrasenkommandos
- J = Anzahl der Akzentkommandos
- $G_{pi}(t)$ = i-ter Phrasensteuermechanismus
- α_i = Dämpfungsfaktor von $G_{pi}(t)$

³⁸ Eine direkte physiologische Untersuchung der holländischen Prosodie liefert etwa Strik (1994).

A_{pi}	=	Amplitude des i-ten Phrasenkommandos
T_{0i}	=	Zeitpunkt des i-ten Phrasenkommandos
$G_{aj}(t)$	=	j-ter Akzentsteuermechanismus
β_i	=	Dämpfungsfaktor von $G_{aj}(t)$
A_{aj}	=	Amplitude des j-ten Kommandos
T_{1j}	=	Zeitpunkt des Beginns des j-ten Akzentkommandos
T_{2j}	=	Zeitpunkt des Endes des j-ten Akzentkommandos

Der modellierte F0-Verlauf $\ln F_0(t)$ ergibt sich in Abhängigkeit vom Zeitsignal additiv aus den Verläufen der Phrasen- und der Akzentkomponente. Die erste Komponente besteht aus dem Phrasensteuermechanismus $G_{pi}(t)$ und die zweite aus dem Akzentsteuermechanismus $G_{aj}(t)$. Die Parameter des Systems sind der über den gesamten F0-Verlauf konstant gehaltene Basiswert $\ln F_{\min}$ sowie die Parameter des Phrasensteuermechanismus T_{0i} , A_{pi} , α_i und die des Akzentsteuermechanismus T_{1j} , T_{2j} , A_{aj} und β_j . Die Parameter der Phrasenkomponente bleiben für die Dauer einer intonatorischen Phrase konstant und die der Akzentkomponente für die Dauer einer Akzentgruppe. Die Dämpfungsfaktoren legen den Grad der Abschwächung der Amplituden von Phrasen- und Akzentkomponente während des Schwingungsvorgangs fest.

6.3.2 Anwendung des Modells

Der Phrasensteuermechanismus des Fujisaki-Modells erfasst den eher langsamen globalen Intonationsverlauf einer Äußerung, in dem sich in der Regel syntaktische Struktur und Satzmodus manifestieren. Nach Möbius (1993: 65) wird der „globale Anteil der Intonation vom Sprecher unter Bedingungen der Vorausplanung und Vorausschau produziert [...], und zwar nicht nur hinsichtlich des äußerungsinitialen [F0-Wertes] [...], sondern auch bezüglich der Phrasierung und der Intervalle zwischen den betonten Silben.“ Für die eher schnellen lokalen Bewegungen der Akzentkomponente übernimmt Möbius das Konzept der Akzentgruppe ('prosodic stress group') von Thorsen (1988). Eine Akzentgruppe erstreckt sich von einer satzbetonten Silbe bis zur nächsten, wobei dazwischenliegende unbetonte Silben Bestandteil der Akzentgruppe sind. Die Modellparameter des Fujisaki-Modells bestimmt Möbius mit einem Software-Programm, wobei ihm als Sprachmaterial Aussage- und Fragesätze des Deutschen vorliegen, die aus ein bis zwei Phrasen bestehen. In Syntheseexperimenten werden die ermittelten Parameter perzeptiv überprüft.

Das Auftreten einer inneren Phrasengrenze unterbricht den Verlauf einer Akzentgruppe vor der Grenze. Eine Phrasengrenze zwischen zwei Akzentgruppen ist an dieser Stelle nach Möbius von einer steigend-fallend-steigenden Bewegung begleitet. Die Unterbrechung innerhalb einer Akzentgruppe bedingt in der Regel einen relativ steilen Anstieg zu der direkt vor der Grenze liegenden posttonischen Silbe. Zu der ersten Silbe nach der Phrasengrenze ist ein unregelmäßiges Absinken des Intonationsverlaufs zu beobachten. Grundsätzlich bewirkt die Phrasierung bei Grenzen innerhalb einer Äußerung im Gegensatz zu äußerungsfinalen Grenzen eine deutlich ausgeprägtere Dehnung der lautlichen Segmente vor der Grenze.

Bei der Modellierung seiner Testsätze fügt Möbius den modellierten Phrasen ein konstantes finales Kommando hinzu, das in Aussagesätzen und in Fragesätzen mit Fragepronomen einen

Amplitudenwert von -0.1 hat. Damit soll das für diese Satztypen beobachtete typische finale Absinken des Intonationsverlaufs nachgebildet werden. In Entscheidungs- und Echofragesätzen wird entsprechend ein konstanter Wert von 0.2 angenommen. Die Phrasenkommandos beginnen grundsätzlich 323 ms vor Äußerungsbeginn, vor einer Phrasengrenze oder vor Äußerungsende. Der Dämpfungsfaktor α erhält für alle Äußerungen einen konstanten Wert von 3.1 s^{-1} (Sekunden); der Dämpfungsfaktor β der Akzentkomponente ist gleichermaßen konstant auf 16.0 s^{-1} gesetzt. Bei der Modellierung der einzelnen Akzentgruppen wird genau ein Akzentkommando pro Gruppe angesetzt. Die Phrasenamplituden berechnen sich aus der Teilformel i) und die Akzentamplituden aus der Teilformel ii) von $\ln F_0(t)$:

$$\text{i) } H_p(t) = \ln F_{\min} + \sum_{(i=1)}^I A_{pi} \cdot 6G_{pi}(t - T_{0i})$$

$$\text{ii) } H_a(t) = A_a \cdot 6[G_a(t - T_1) - G_a(t - T_2)]$$

Der von Sprecher, Anzahl der Phrasen und Satzmodus abhängige Basiswert $\ln F_{\min}$ beträgt bei Möbius in einphrasigen Aussagesätzen im Mittel für Sprecher 75 Hz und für Sprecherinnen 145 Hz. In zweiphrasigen Aussagesätzen erhöht er sich jeweils um 15 Prozent. In Fragesätzen konnte Möbius wegen einer zu schmalen Datenbasis keine statistisch abgesicherten Werte ermitteln. Die Werte der Phrasenamplitude weisen bei nicht-äußerungsfinalen Phrasenkommandos eine sprecherspezifische Ausprägung auf. Zudem ist der Wert dieser Amplituden von der Akzentverteilung in der Äußerung und vom Satzmodus abhängig. In einphrasigen Aussagesätzen mit einer initialen akzentuierten und einer finalen nicht-akzentuierten Silbe sind die Werte der nicht-äußerungsfinalen Phrasenamplituden bei allen Versuchspersonen generell um etwa ein Viertel höher als bei anderen Akzentverteilungen. Der Wert der Phrasenamplitude ist nach Möbius insofern ein direktes Maß für die Deklination.

Bei den zweiphrasigen Aussagesätzen sind die Phrasenamplitudenwerte von der ersten und zweiten intonatorischen Phrase beide relativ hoch, wenn der Hauptsatz mit der zweiten Phrase zusammenfällt. Im Fall der Koinzidenz des Hauptsatzes mit der ersten Phrase haben beide Amplituden relativ niedrige Werte. Damit scheint sich die These eines 'resettings' der Deklinationslinie am Anfang einer neuen intonatorischen Phrase zu bestätigen. Die zweiphrasigen Aussagesätze haben eine sprecherspezifisch unterschiedlich ausgeprägte Markierung der Phrasengrenzen. An dieser Stelle sind die Akzentamplituden der den Phrasengrenzen unmittelbar vorausgehenden Akzentgruppen etwa um ein Viertel größer als in anderen Positionen. Es läßt sich nach Möbius (ebd.: 165) zudem in dieser Position ein 'phrase-final lengthening' bei den Akzentgruppen beobachten.

Die verschiedenen Arten von Fragesätzen differieren deutlich in der Ausprägung der (einzig) Phrasenamplitude, die eine spezifische Deklinationslinie als globale Markierung erkennen läßt. Fragesätze mit Fragepronomen haben die größere Phrasenamplitude (0.35) und damit steilste Deklination. Echofragesätze (0.14) besitzen einen relativ flachen Phrasenverlauf und Entscheidungsfragesätze (0.27) einen dagegen vergleichsweise höheren Verlauf. Bei Fragesätzen mit Fragepronomen ist wie bei Aussagesätzen ein finales Absinken des Phrasenverlaufs zu beobachten. Entscheidungs- und Echofragen weisen einen Anstieg auf. Die Ähnlichkeit von

Fragesätzen mit Fragepronomen und Aussagesätzen wird auch in den finalen Akzentkommandos der Amplituden deutlich, deren Mittelwert für beide Satztypen in etwa gleich ist. In Entscheidungs- und Echofragen sind die finalen Akzentkommandos entsprechend ihres finalen Anstiegs vergleichsweise höher. Demnach besteht eine Korrelation von finaler Kontur und globaler Phrasenamplitude.

Die statistische Analyse der Parameter der Akzentkomponente belegt die generelle Tendenz, dass kürzere Akzentgruppen höhere Amplituden als längere Akzentgruppen besitzen. Zudem ist der Amplitudenwert einer nicht-finalen Akzentgruppe höher, wenn sie sich weiter rechts in einer Äußerung befindet. Mit der Unterscheidung von Substantiven und anderen Wortarten in einer Äußerung ergibt sich eine geringe Differenz bei den Mittelwerten der Akzentamplituden. Der Mittelwert liegt bei Substantiven um etwa zehn Prozent höher als bei anderen Wortarten, die eine Absenkung von fünf Prozent verzeichnen.

Die Dauer der einzelnen Akzentkommandos korreliert grundsätzlich mit der Dauer der entsprechenden Akzentgruppe. Es besteht kein deutlicher Dauerunterschied zwischen den finalen und nicht-finalen Akzentgruppen der Testsätze, so dass nach Möbius (ebd.: 133) ein 'final lengthening' bei finalen Akzentgruppen nicht bestätigt werden kann. Die Dauer der Akzentkommandos ist allerdings in ein- und zweiphrasigen Aussagesätzen und in Fragesätzen mit Fragepronomen bei finalen Akzentgruppen um ein Viertel kürzer als bei den anderen Akzentgruppen. In Entscheidungsfragesätzen ist die Dauer um etwa ein Drittel länger. Der Anschaltzeitpunkt der Akzentkommandos stimmt bei finalen Akzentgruppen weitestgehend mit dem Beginn einer Akzentgruppe überein. Bei nicht-finalen Akzentgruppen liegt er um etwa zehn Prozent der Akzentgruppendauer nach dem Beginn der Akzentgruppe.

6.3.3 Erweiterung des Steuermechanismus

Mixdorff (1998) greift bei seiner Bestimmung von Fujisaki-Modellparametern auf den Begriff des Tonbruchs ('tone switch') von Isaenko/Schädlich (1966) zurück. Ein Akzentkommando beginnt bzw. endet mit einem Tonbruch. Die Parameter α , β setzt auch Mixdorff konstant auf $\alpha = 2.0/\text{sec}$ und $\beta = 20.0/\text{sec}$; in kurzen Äußerungen gilt $\alpha = 3.0/\text{sec}$. Der Wert für F_{\min} ist für einen Sprecher konstant und besteht aus dem Mittelwert der F0-Minima der Aussagesätze des Sprechers.

Mixdorff (ebd.: 66f) ermittelt nach linguistischen Kriterien in (nachbearbeiteten) F0-Verläufen zeitliche Korrelate von Wortakzenten und von möglichen Phrasengrenzen. Die initialen Werte der Phrasenkomponenten werden entlang den unbetonten Silben bestimmt. Akzentkommandos bestimmt Mixdorff nach dem auditiven Eindruck und nach den Wechseln in dem F0-Verlauf, die er als Tonbruch analysiert. Wenn in einem F0-Verlauf kein Tonbruch festzustellen ist, dann erhält das Akzentkommando den Wert null. Die manuell ermittelten Parameter werden mit einem automatischen Verfahren optimiert und in einer Resynthese von ihm auditiv überprüft³⁹.

Mixdorff untersucht u.a. die Modellierung nach dem Fujisaki-Modell von Satzart und Fokus im Deutschen. In den von ihm durchgeführten Experimenten verwendet er Stimuli mit engem und mit weitem Fokus sowie Frage- und Aussagesstimuli. Neben den Fokusakzent treten in den Sti-

³⁹) Mixdorff (2000) stellt ein automatisches Verfahren zur vollständigen Berechnung von Fujisaki-Parametern vor.

multi auch Nebenakzente auf. Die Untersuchung zeigt, dass ein enger Fokus grundsätzlich mit einem größeren Akzentkommando und mit einem reduzierteren Kommando für Nebenakzente modelliert werden muss als eine vergleichbare Struktur mit einem weiten Fokus. Die Akzentkommandos von engen Foki sind zudem länger als die von weiten Foki. Bei engen Foki wird zudem das Phrasenkommando reduziert, wobei dieser Effekt wiederum in Aussagen stärker als in Fragen ist. In Aussagestimuli auftretende Deklinationseffekte modelliert er, indem er ein Phrasenkommando 250 ms vor der Äußerung ansetzt.

Mixdorff stellt in seiner Untersuchung tonale Verläufe fest, in denen der mit einem Satzakzent verbundene Anstieg langsam weiter bis zu dem Beginn des finalen Anstiegs der Frage steigt. Ein derartiger Verlauf kann mit dem Fujisaki-Modell nicht beschrieben werden, so dass Mixdorff eine 'slow rise' – Komponente $G_s(t)$ in das Modell einführt. Die dem Modell additiv hinzugefügte Komponente soll die Bedingungen i) und ii) erfüllen (ebd.: 98f):

- i) Kompensation des fallenden Verlaufs in der Phrasenkomponente, so dass ein in etwa linearer Anstieg in der Phrase entsteht.
- ii) Langsamer initialer Anstieg bei der Phrasenkomponente, so dass ein konkaver globaler Verlauf der Phrase entsteht.

Der in Kapitel 6.3.1, S. 41, angegebenen Funktion $\ln F_0(t)$ ist demnach der folgende Term hinzuzufügen, wobei T_3 und T_4 wiederum Anfangs- und Endzeitpunkt des Kommandos bezeichnen⁴⁰:

$$\sum_{k=1}^K A_{Sk} [G_s(t-T_{3_k}) - G_s(t-T_{4_k})],$$

wobei

$$G_s(t) = 1 - (1 + \delta t) \exp(-\delta t), \text{ für } t \geq 0$$

$$G_s(t) = 0, \text{ für } t < 0$$

⁴⁰) Die von Mixdorff (ebd.: 48) verwendete Funktion $\ln F_0(t)$ enthält in ihrer Akzentkomponente die zusätzliche Konstante $\gamma = 0.9$, die sicherstellt, dass die Akzentkomponente ihr Maximum in endlicher Zeit erreicht.

6.4 Das Tilt-Modell

Taylor (2000) stellt das am britischen Englisch entwickelte Intonationsmodell Tilt zur Analyse und Synthese von F0-Verläufen vor, wobei in der vorliegenden Arbeit nur die Analyse des Modells interessiert. Nach Taylor (ebd.: 22) handelt es sich um ein phonetisches Modell, „because its purpose is to describe observable linguistic sound phenomema.“

Das Modell geht von der grundlegenden Einheit des intonatorischen Ereignisses ('intonational event') aus, bei dem es sich um einen tonalen Akzent ('pitch accent') oder um einen Grenzton handeln kann. Die beiden Ereignisse sind konstitutiv für Intonationsphrasen und werden mit dem Beginn der entsprechenden Silbenkerne einer Äußerung synchronisiert. Ein Ereignis ist stets auch mit einer F0-Bewegung verbunden, bei der es sich nach dem Modell nur um einen Anstieg, einen Fall oder um einen Anstieg mit einem anschließendem Fall handeln kann. Bei einer Kombination von zwei Bewegungen kann eine Bewegung partiell auftreten, also etwa ein partieller Anstieg mit einem folgenden – vollständigen – Fall.

Von einer Verbindung zwischen zwei Ereignissen wird die Amplitude des F0-Verlaufs und die Dauer berechnet. Bei den Ereignissen mit den F0-Verläufen Anstieg ('rise') und Fall ('fall') werden Tilt-Parameter nach den Formeln i) und ii) berechnet, wobei „A“ für Amplitude und „D“ für Dauer steht.

$$\text{i) } \mathit{tilt}_{amp} = \frac{(|A_{rise}| - |A_{fall}|)}{(|A_{rise}| + |A_{fall}|)} \quad \text{ii) } \mathit{tilt}_{dur} = \frac{(D_{rise} - D_{fall})}{(D_{rise} + D_{fall})}$$

Bei einem Fall besteht die Amplitude aus dem F0-Maximum des Beginns der Bewegung und bei einem Anstieg entsprechend aus dem Maximum an dem Ende der Bewegung. Die Dauer wird gemessen als der F0-Abstand zwischen dem Maximum und dem Wert an dem Ende der Bewegung, so dass ein Fall immer einen negativen Wert hat. Tritt in einem Ereignis nur eine der beiden möglichen Bewegungen auf, dann erhält die Amplitude und die Dauer der anderen den Wert null. Unter der empirischen Annahme, dass i) und ii) stark korrelieren, berechnet Taylor in iii) einen Tilt-Parameter, der die Form eines Ereignisses charakterisiert:

$$\text{iii) } \mathit{tilt} = \frac{(|A_{rise}| - |A_{fall}|)}{(2 * (|A_{rise}| + |A_{fall}|))} + \frac{(D_{rise} - D_{fall})}{(2 * (D_{rise} + D_{fall}))}$$

Die Parameter Amplitude und Dauer eines Ereignisses werden in ihrer Quantität mit den Formeln iv) und v) kodiert:

$$\text{iv) } A_{event} = |A_{rise}| + |A_{fall}| \quad \text{v) } D_{event} = D_{rise} + D_{fall}$$

Nach Taylor zeigt die Amplitude die Prominenz eines Ereignisses an, während mit der Dauer keine intonatorische Information in dem Sinne verbunden ist, dass die Dauer epiphänomenologisch mit der Amplitude variiert. Die Beschränkung des Modells auf die beiden F0-Bewegungen Anstieg und Fall sowie die Annahme von intonatorischen Ereignissen entspricht Taylor (ebd.:

22f) folgend einigen Grundannahmen des Tonsequenzmodells; vgl. Kapitel 4, S. 20ff, sowie die Besprechung der phonetischen Intonationsmodelle in Kapitel 6.7, S.50f.

6.5 Das INTSINT-Modell

Hirst /Di Cristo/Espesser (2001) entwickeln das Verfahren MOMEL ('modelling melody') zur automatischen Modellierung von F0-Verläufen. Das Verfahren berechnet Zielpunkte ('pitch targets') eines F0-Verlaufs, die potentielle Korrelate von tonalen Wendepunkten darstellen. Das Verfahren ist in ein Softwaresystem integriert ('Mes Signaux package', vgl. Anhang C, S. 110), das die visuelle und über eine Resynthese auditive Überprüfung der modellierten F0-Verläufe ermöglicht. In einer Fußnote konkretisieren Hirst /Di Cristo/Espesser (2001: 18) den Status der Zielpunkte: „These 'target points' are essentially the 'turning points' described by Gårding [...], except that, unlike turning points, target points do not necessarily imply a change of direction: downstepped and upstepped targets can thus be generated as described below.”

MOMEL arbeitet einen gemessenen F0-Verlauf in vier übergeordneten Schritten ab. Die einzelnen Schritte werden an dieser Stelle genauer beschrieben, weil MOMEL das einzige in der vorliegenden Arbeit behandelte Modellierungsverfahren ist, das keine symbolischen Informationen verwendet, so dass die einzelnen Berechnungsschritte natürlich besonders interessant sind. Der erste Schritt besteht in der Anwendung eines Glättungsverfahrens, in dem alle F0-Werte gelöscht werden, die in ihrem Betrag mehr als 5% von ihren beiden Nachbarwerten abweichen. In dem zweiten Schritt werden Kandidaten $\langle t, h \rangle$ für potentielle Zielpunkte bestimmt, in dem iterativ für jeden F0-Wert x des F0-Verlaufs die Teilschritte (a) bis (c) ausgeführt werden:

- (a) Es wird ein Zeitfenster $A = 300$ ms über x gelegt, wobei sich x in der Mitte des Fensters befindet. Anschließend werden alle F0-Werte innerhalb des Fensters gelöscht, die außerhalb eines F0-Bereichs B liegen; B hat als Minimum 50 Hz, und das Maximum errechnet sich aus dem Mittelwert der oberen 5% des gemessenen F0-Verlaufs multipliziert mit 1.3.
- (b) Es werden alle F0-Werte innerhalb des Fensters gelöscht, die von einer, ebenfalls innerhalb des Fensters berechneten quadratischen Regressionslinie mehr als 5% abweichen. (a) und (b) werden solange ausgeführt, bis kein F0-Wert mehr gelöscht werden kann.
- (c) Es wird für jedes verbleibenes x ein $\langle t, h \rangle$ aus der Regressionsgleichung $y'' = a + bx + cx^2$ bestimmt, wobei $t = -b/2c$ und $h = a + bt + ct^2$ ist.

Wenn ein F0-Wert in dem zweiten Schritt nicht gelöscht wird, dann ist ihm also als Ergebnis von (a) – (c) ein $\langle t, h \rangle$ zugeordnet. Die Folge der Kandidaten $\langle t, h \rangle$ wird dann im dritten Schritt partitioniert. Innerhalb eines - wandernden - Zeitfensters B von typischerweise 200 ms wird dabei der Durchschnittsabstand $dt(x)$ der Zeitwerte der ersten Hälfte von den Zeitwerten der zweiten Hälfte berechnet. Analog wird der Durchschnittsabstand $dh(x)$ der Frequenzwerte der ersten Fensterhälfte von denen der zweiten Fensterhälfte berechnet. Mit diesen Werten erfolgt eine Berechnung des kombinierten und gewichteten Abstandes $d(x)$:

$$d(x) = \frac{(dt(x)*wd + dh(x)*wh)}{(wd + wh)}, \text{ wobei}$$

$$wd = \frac{1}{(\text{mean } dt(x))} \quad \text{und} \quad wh = \frac{1}{(\text{mean } dh(x))}$$

Die Grenzen der Partition werden an der Stelle gesetzt, an der $d(x)$ ein lokales Maximum bildet und größer als der Durchschnitt aller kombinierten Abstände ist, d.h. $d(x)$ muss Bedingungen i) – iii) erfüllen:

- i) $d(x) > d(x-1)$
- ii) $d(x) > d(x+1)$
- iii) $d(x) > \text{mean}(d(x))$

In dem letzten Verarbeitungsschritt von MOMEL werden die Kandidaten gelöscht, deren $dt(x)$ oder $dh(x)$ größer als eine Standardabweichung von den $dt(x)$ und $dh(x)$ der anderen Fensterhälfte abweicht. Schließlich wird in jeder Partition der Durchschnittswert der verbliebenen Zielwerte als finale Schätzung von $\langle t, h \rangle$ für diese Partition berechnet.

Das 'Mes Signaux package' ermöglicht im Anschluss an MOMEL die automatische Stilisierung der berechneten Zielpunkte nach dem bereits in Kapitel 2.2, S. 12f, erwähnten Transkriptionssystem INTSINT (International Transcription System for Intonation). INTSINT verwendet die tonalen Symbole T(op), M(id), B(ottom), H(igher), S(ame), L(ower), U(pstepped) und D(ownstepped). T, M und B sind absolute Symbole, die sich auf den Tonhöhenumfang eines Sprechers bzw. auf den globalen Verlauf einer Äußerung beziehen. H, S, L, U, D sind relative Töne, die sich auf ihre tonale Umgebung beziehen. Die relativen Töne werden unterschieden in iterative Töne (U, D) und in nicht-iterative Töne (H, S, L). Der automatischen Stilisierung liegen folgende Schritte zugrunde:

- (a) T bzw. H ist der höchste bzw. tiefste Zielpunkt
- (b) M ist der erste Ton einer Äußerung und nach einer Pause von 250 ms, sofern dort noch kein anderer Ton auftritt
- (c) Alle anderen Zielpunkte sind relative Töne⁴¹. Zielpunkte direkt vor einer Pause (> 250 ms) erhalten S, H oder L.
- (d) Es wird für absolute Töne der Mittelwert und für relative Töne ein linearer Regressionswert bestimmt.
- (e) H und L können in T, U, B oder D umbenannt werden, wenn der statistische Wert des umbenannten Tons näher an dem tatsächlichen Wert liegt als der statistische Wert des nicht umbenannten Tons.
- (f) (d) und (e) werden solange wiederholt, bis keine Töne mehr umbenannt werden können.

Die in Kapitel 8.5 durchgeführte Anwendung von MOMEL und der anschließenden INTSINT-Stilisierung auf Äußerungen des Deutschen ist überprüft und manuell korrigiert worden.

⁴¹⁾ Ein Grenzwert zur Unterscheidung des Tons S von den anderen Tönen wird nicht angegeben.

6.6 Weitere Modelle

Es sollen abschließend zwei weitere Verfahren der phonetischen Intonationsmodellierung genannt werden, die allerdings in dem experimentellen Teil der vorliegenden Arbeit nicht berücksichtigt werden. Heuft (1999: 22f) beschreibt F0-Verläufe mit einer maximumbasierten Methode, die einen F0-Verlauf als Folge von Maxima auffasst. Ein F0-Maximum wird mit den folgenden Parametern beschrieben, die von Heuft in einem halbautomatischen Verfahren ermittelt:

- (a) 'delay'. Positiver oder negativer Abstand des F0-Maximums zum Beginn des betonten Vokals.
- (b) Amplitude. Höhe des F0-Maximums relativ zu einer 'top-' und zu einer 'baseline'; die Linien sind konstant für einen Sprecher und entsprechen dem F0-Maximum bzw. -Minimum, das der jeweilige Sprecher in dem von Heuft untersuchten Korpus realisiert hat.
- (c) Anstieg. Steilheit des Anstiegs vor dem Maximum; der Anstieg wird mit einer \cos^2 - Kurve der Form $\beta_{ii} = \pi / 2 * 1/ \text{ms}$ angenähert.
- (d) Fall. Steilheit des Abfalls nach dem Maximum; der Annäherung erfolgt ebenfalls mit der \cos^2 - Kurve.

Rapp (1998: 40f) verwendet im Zusammenhang mit der Entwicklung eines Systems zur automatischen prosodischen Transkription von Korpora des Deutschen eine Parametrisierungsfunktion für F0-Verläufe von Akzenten. Die Funktion $f(t)$ arbeitet in einem Zeitfenster mit der Länge von zwei Silben, also Akzent- und postakzentuierte Silbe:

$$f(t) = \alpha \tanh(\beta(t - \gamma)) + \delta e^{-(\epsilon(t - \zeta))^2} + \eta$$

Die einzelnen Parameter erhalten nach Rapp (ebd.: 41) die folgende phonetische Interpretation:

„ α korreliert mit dem tonalen Unterschied zwischen der akzentuierten und der postakzentuierten Silbe; α beschreibt die Abweichung beider Silben von dem gemeinsamen Grundniveau η

β korreliert mit der Steilheit eines Anstiegs oder Abfalls

γ korreliert mit der zeitlichen Alignierung eines Anstiegs oder Abfalls

δ korreliert mit der Höhe eines Gipfels/Tals

ϵ korreliert mit der Steilheit eines Gipfel/Tals

ζ korreliert mit der zeitlichen Alignierung eines Gipfels/Tals

η korreliert mit dem generellen F₀-Niveau”

Die Anwendung der Funktion erfordert eine manuelle, heuristisch initiale Schätzung der Parameter, die dann mit einem Optimierungsalgorithmus an den F0-Verlauf einer konkreten Akzentrealisierung angepasst werden können. Als Gütekriterium der Parameteroptimierung verwendet Rapp die Summe der Fehlerquadrate zu den gemessenen F0-Werten.

6.7 Diskussion der Modelle

Die vorgestellten phonetischen Intonationsmodelle bieten einen ganz unterschiedlichen Zugang zu dem F0-Verlauf einer Äußerung und damit zu den realisierten kontrastiven Satzakzenten. Der unterschiedliche Zugang läßt sich im einzelnen wie in (a) bis (e) beschrieben charakterisieren.

(a) Holistische F0-Verläufe in einem gesamten F0-Verlauf.

Das Kieler Intonationsmodell KIM analysiert Satzakzente als F0-Gipfel- und -Talkonturen. Die Konturen bilden einen integralen Bestandteil des gesamten F0-Verlaufs einer Intonationsphrase und werden als Ganze mit verschiedenen intonatorischen Funktionen verbunden. Die Position der F0-Gipfel und -Täler ist distinktiv, so dass KIM so gesehen ein phonologisches Intonationsmodell ist. Die in dieser Arbeit vorgenommene Klassifikation von KIM als phonetisches Modell basiert hingegen auf dessen konkreten Aussagen über F0-Verläufe, d.h. über die Position beispielsweise eines (relativen) F0-Maximums bzgl. des akzentuierten Silbenkerns. An den von KIM angenommenen Bedeutungen ist deren mangelnde Operationalisierung zu kritisieren.

(b) Linear stilisierte, separate F0-Verläufe, die von links nach rechts zu analysieren sind.

Das IPO-Modell besteht aus zwei Komponenten, der linearen Approximation von Abschnitten eines F0-Verlaufs und der Standardisierung der approximierten F0-Abschnitte. Die Approximation erfolgt ausschließlich nach perceptiven Kriterien, bei denen funktionale Eigenschaften der Intonation keine Rolle spielen. Die Approximation kann mit dem Verfahren von d'Allessandro/Mertens (1995) zur Berechnung von tonalen Segmenten automatisiert und damit auch objektiviert werden. Das Verfahren basiert auf Werten des IPO-Modells über die Wahrnehmung von Tonhöhenverläufen. Bei der Berechnung der Standardkonturen ermittelt das IPO-Modell dann auch Konturen, die beispielsweise bei Satzakzenten auftreten. Grundsätzlich spielt der funktionale Aspekt in dem IPO-Modell aber eine untergeordnete Rolle.

(c) Überlagerte F0-Verläufe, die sich aus verschiedenen Komponenten zusammensetzen.

Das Fujisaki-Modell fasst einen F0-Verlauf als additive Überlagerung einer Akzent- und einer Phrasenkomponente (und eines bestimmten minimalen F0-Wertes) auf, die das Ergebnis der mathematischen Verknüpfung von zwei unterschiedlichen Funktionen ist. In einem konkreten F0-Verlauf ermittelt das Modell die Parameter von Satzakzenten durch die Ausrichtung der Akzentkommandos auf der Zeitachse und durch deren Amplitude. Der Amplitudenwert eines Akzentkommandos ist ein abstrakter Wert, der nicht direkt auf Eigenschaften eines F0-Verlaufs abgebildet werden kann. Eine Abbildung ist nur über die Berechnung des gesamten Steuermechanismus des Fujisaki-Modells möglich, wobei die Ermittlung von dem selektiven Einfluss eines Satzakzents auf einen modellierten F0-Verlauf die Konstanzhaltung der Phrasenkomponente (und des minimalen F0-Wertes) erfordert.

(d) Intonatorische Ereignisse und deren Verbindung in F0-Verläufen.

Das Tilt-Modell analysiert F0-Verläufe in einem automatischen 'pattern matching' - Verfahren als Steigung, Fall, Steigung plus partiellen Fall, Steigung plus vollständiger Fall oder als partielle Steigung plus vollständigen Fall. Die Verläufe ohne partielle Abschnitte lassen sich als Manifestationen der mono- bzw. bitonalen, funktional relevanten Elemente des Tonsequenzmodells inter-

pretieren. Eine weitere Gemeinsamkeit des Tilt- mit dem Tonsequenzmodells ist die Aufteilung eines Intonationsverlaufs in intonatorische Ereignisse und deren Verbindung. In dem Tonsequenzmodell bestehen die intonatorischen Ereignisse aus Akzent- und Grenztönen, die linear interpoliert werden. Im Gegensatz zu den Akzent- und Grenztönen des Tonsequenzmodells haben die Tilt-Parameter graduellen, kontinuierlichen Charakter.

(e) Globale und lokale Elemente in F0-Verläufen.

Das INTSINT-Modell analysiert F0-Verläufe in globale Komponenten, die absoluten Töne, und in lokale Komponenten, die relativen Töne. Die absoluten Töne werden unter Betrachtung der gesamten Intonationsphrase vergeben, wohingegen die Vergabe der relativen Töne von links nach rechts in Abhängigkeit von der direkten Umgebung erfolgt. Die Zuweisung der einzelnen Töne basiert auf der Analyse von Zielpunkten in F0-Verläufen.

EXPERIMENTELLER TEIL

7. GEWINNUNG VON SATZAKZENTREALISIERUNGEN

Der experimentelle Teil der vorliegenden Arbeit beschreibt die Gewinnung von kontrastiven Satzaktzentrealisierungen im Deutschen. Die Gewinnung besteht aus einem Produktions- und einem komplementären Perzeptionsexperiment. In dem letztgenannten Experiment werden die produzierten Äußerungen im Hinblick auf realisierte Satzaktzente auditiv beurteilt. Das Produktionsexperiment ist ein klassisches Laborexperiment, wohingegen in dem Perzeptionsexperiment die Möglichkeiten des Internets genutzt werden, um eine möglichst große und breit gestreute Anzahl von Probanden zu erreichen.

7.1 Methodische Vorüberlegungen

Satzaktzentrealisierungen des Deutschen werden in dem experimentellen Teil dieser Arbeit nicht in spontan geäußerten Sätzen und auch nicht auf der Basis eines umfangreichen akustischen Korpus untersucht. Spontansprache ('non-scripted speech') ist die normale alltägliche Form des konkreten sprachlichen Verhaltens der Teilnehmer einer Sprachgemeinschaft und hat demnach eine bestimmte Priorität gegenüber anderen sprachlichen Verhaltensformen wie dem Vorlesen von Texten⁴². Es handelt sich bei Spontansprache natürlich nicht um ein homogenes Phänomen, sondern um ein weiter zu differenzierendes sprachliches Verhalten, das mindestens mit der Vielschichtigkeit sozialer Welten einhergeht. Nicht zuletzt aus diesem Grund ist etwa das bereits erwähnte, Spontansprache verarbeitende Übersetzungssystem VERBMOBIL für einen bestimmten Bereich entwickelt worden, der Domäne 'Reiseplanung zwischen Geschäftspartnern'⁴³.

Das in dieser Arbeit untersuchte Sprachmaterial besteht aus in einem Sprachlabor aufgenommenen Äußerungen des Deutschen, die unter den üblichen experimentellen Bedingungen der Kontrollierbarkeit, Reproduzierbarkeit und Variierbarkeit des Experimentsablaufs elizitiert wurden. Die experimentelle Methode wurde gewählt, um eine möglichst optimale Vergleichbarkeit der lautlichen Formen der elizitierten kontrastiven Satzaktzente herzustellen. Mit dem experimentellen Aufbau ist die Kontrolle über die Funktionen der Satzaktzente gegeben, so dass die Ausprägung der Akzentformen nicht auf die Variation der genannten Funktionen zurückgeführt werden kann. Damit entsteht die direkte Vergleichbarkeit der phonetischen Korrelate von kontrastiven Satzaktzenten. Wenn (kontrastive) Satzaktzente dagegen in spontansprachlichem Material

⁴²) In diesem Zusammenhang ist erneut auf die Unterscheidung von Standarddeutsch und dialektaler Ausprägung des Deutschen als eine weitere der vielen bestehenden Beschreibungsebenen hinzuweisen; vgl. dazu die in Kapitel 2.1, S. 11, aufgeführte Charakterisierung der deutschen Standardlautung des Duden Aussprachewörterbuchs (1990).

⁴³) Wahlster (2000) gibt einen Überblick über die Architektur des Systems, und Burger et al. (2000) beschreiben das Korpus, an dem VERBMOBIL statistisch trainiert wurde.

phonetisch untersucht werden sollen, dann setzt dies die Analyse ihrer Funktion(en) voraus⁴⁴. Eine derartige Analyse ist letztlich wiederum nur experimentell überprüfbar.

Werner (2000) vergleicht verschiedene Intonationsmodelle auf der Basis des ersten Teils des Kieler PHONDAT-Korpus, der aus transkribierter Lesesprache besteht (vgl. Kohler (1994)). Für seinen Vergleich entwickelt er eine Meta-Repräsentation für F₀-Verläufe, auf die er die verschiedenen Merkmale der von ihm untersuchten Intonationsmodelle abbildet. Die untersuchten Intonationsmodelle sind das IPO-Modell, das Fujisaki-Modell und eine Anwendung des Tonsequenzmodells auf das Deutsche.

Werner versucht die empirische Relevanz der einzelnen Modelle zu ermitteln, indem er eine Cluster-Analyse mit den von Kohonen (1995) entwickelten 'self-organizing maps' durchführt. Er kommt u.a. zu dem Ergebnis (ebd.: 107), „dass keines der untersuchten Modelle fähig war, das Korpusmaterial befriedigend zu klassifizieren“. Bei den von ihm klassifizierten F₀-Verläufen trifft er keine funktionale Unterscheidung, so dass er die mangelnde Klassifikationsfähigkeit der Modelle nur pauschal feststellen kann. Die vorliegende Arbeit knüpft an die Untersuchung an und stellt das IPO- und das Fujisaki-Modell sowie drei weitere phonetische Modelle anhand von semantisch konstantem Material gegenüber.

Labor- wie auch Lesesprache können Phänomene aufweisen, die in der Spontansprache bzw. in einer ihrer Ausprägungen gar nicht oder nur marginal vorkommen. Somit muss die Existenz der nicht-spontansprachlichen Phänomene natürlich auch in anderen Sprachausprägungen untersucht werden, wenn Aussagen beispielsweise über 'das Standarddeutsche' getroffen werden sollen. Dabei bieten sich prosodisch annotierte, umfangreiche Korpora an, die zumeist nach den Annahmen des Tonsequenzmodells verschriftet sind, also mit (einer Variante von) ToBI⁴⁵. Ein ToBI-Label kann mit verschiedenen Funktionen von Satzakkzenten verbunden sein, so dass die in der vorliegenden Arbeit angestrebte semantische Invarianz bei einem derartigen Korpus nicht gegeben ist.

Das Korpus des Produktions- und Perzeptionsexperiments dieser Arbeit besteht aus unterschiedlichen syntaktischen Konstruktionen, d.h. aus Sätzen mit normaler und mit markierter, topikalisierte Wortstellung und aus Ellipsen. Ellipsen sind in Spontansprache häufig auftretende Konstruktionen, deren syntaktische Akzentuierung empirisch bisher wenig untersucht worden ist. Die Verwendung des normalen und des markierten Wortstellungstyps ermöglicht dagegen die Untersuchung des Einflusses der Wortstellung auf die Satzakkzentformen. Eine weitere Besonderheit der Testsätze ist die systematische Elizitierung von zwei kontrastiven Satzakkzenten, die zudem an ganz unterschiedlichen syntaktischen Positionen auftreten. Die Intonation der Testsätze erhält dadurch eine hohe Funktionalität, so dass sich die Sätze besonders für eine Analyse nach unterschiedlichen phonetischen Intonationsmodellen eignen.

In diesem Zusammenhang ist zu beachten, dass die beiden kontrastiven Satzakkzente in einer Reihe von Fällen entweder als zwei einzelne Akzente oder als ein Brückenakzent interpretiert werden können (vgl. Kap. 5.2, S. 30ff). Die erste Interpretation setzt die Annahme von zwei

⁴⁴) Beckman (1997) stellt verschiedene Methoden der Gewinnung von Spontansprache dar; in diesem Zusammenhang sei auf ein Analyseproblem verwiesen, das van Donzel (1994: 15) wie folgt formuliert: „In the literature focus is generally detected on the basis of intonation. This definition leads to circularity, since possible acoustic features of focus are already included in the definition itself.“

⁴⁵) Das Korpus des Forschungsprojekts (und gleichnamigen Übersetzungssystem) VERBMOBIL ist die bisher wohl umfangreichste Sammlung von (nicht nur) prosodisch annotierten Daten des Deutschen; unter <http://www.phonetik.uni-muenchen.de/Bas/BasProsodie.html> finden sich genauere Informationen über das Korpus.

Formen und zwei Funktionen voraus, wohingegen bei einem Brückenakzent nur eine Form und eine Funktion zu postulieren ist. Formal liegt ein Brückenakzent vor, wenn - im Sinne der ersten Interpretation - bei der ersten Akzentsilbe ein steigender Tonhöhenverlauf auftritt, dem ein hoher tonaler Verlauf bis zu der zweiten Akzentsilbe folgt. Bei der zweiten Akzentsilbe ist ein fallender Tonhöhenverlauf zu verzeichnen.

7.2 Produktionsexperiment

Der experimentelle Teil dieser Arbeit verwendet das in Lieb (1983) zur semantischen Untersuchung von Satzakzenten vorgestellte Dialogschema. Das Schema wird allerdings nicht zur Untersuchung der Semantik, sondern der lautlichen Formen von Satzakzenten eingesetzt. Es werden somit Dialoge zwischen zwei Sprechern A, B der folgenden Art verwendet:

- A. Klaus bucht in den nächsten beiden Tagen eine Reise, und Mona auch. Ich glaube, Klaus bucht morgen eine Reise.
- B1. Mona bucht heute eine Reise.
- B2. Mona heute.
- B3. Heute bucht Mona eine Reise.

Die drei Dialoge A-B1, A-B2 und A-B3 enthalten jeweils zwei Kontrastpaare, nämlich Klaus und Mona sowie morgen und heute. Die beiden kontrastierenden Elemente in den B-Sätzen (Mona, heute) sind die intendierten Träger von - kontrastiven - Satzakzenten, die durch die A-Sätze elizitiert werden sollen. B1 ist Aussagesatz mit normaler Wortstellung, B2 eine Ellipse und B3 ein Aussagesatz mit topikalisierte Wortstellung. Das in dem Experiment verwendete Korpus ist in Anhang A, S. 105f, vollständig aufgeführt, wobei dort auch die verschiedenen syntaktischen Akzentpositionen eingetragen sind.

Die A-Sätze des Korpus wurden von einem professionellen Sprecher in der Sprachkabine des Instituts für Phonetik und Sprachliche Kommunikation (IPSK) der LMU München gesprochen und mit 16 Bit/16 kHz digital aufgezeichnet. Dem Sprecher lagen die Dialoge des Korpus der Form A-B1 vor, mit denen er sich vor dem Sprechen vertraut gemacht hat. Er wurde angewiesen, die Sätze A-B1 als einen möglichst gut zusammen passenden Dialog zwischen zwei Personen zu lesen. Dem Sprecher wurde der gesamte Dialog vorgelegt, damit er die A-Sätze mit einer Intonation produziert, die die Akzentpositionen in den B-Sätzen bereits vorgibt. Realisierte A-Sätze mit Versprechern, Zögerungen u.ä. wurden dabei wiederholt.

Die von dem Sprecher realisierten A-Sätze wurden bei der Elizitierung von kontrastiven Satzakzenten in den B-Sätzen verwendet. Die genaue Verwendungsweise ist Gegenstand des folgenden Kapitels. Dabei wurden die B-Sätze von den Probanden zu Beginn des Experiments auch einmal kontextlos, also ohne A-Satz vorgelesen ('out of the blue'). Die kontextlos geäußerten und die im Kontext eines A-Satzes elizitierten B-Sätze bilden das eigentliche Sprachmaterial der vorliegenden Arbeit.

7.2.1 Aufbau und Durchführung

Die Probanden des Produktionsexperiments befanden sich in der Sprachkabine des IPSK. In der Kabine wurde auf einem Tisch ein Computerbildschirm aufgebaut, vor dem die Probanden Platz nahmen. Den Probanden wurde mitgeteilt, dass sie alle Informationen über den Experimentablauf über den Bildschirm erhalten. Damit standen allen Probanden des Produktionsexperiments identische Informationen zur Verfügung. Das Richtmikrophon der Kabine wurde individuell angesteuert, um eine optimale Aufnahme zu gewährleisten. Dabei wurden die Äußerungen der B-Sätze ebenfalls mit 16 Bit/16kHz digital aufgezeichnet.

Die einzelnen A-Sätze des professionellen Sprechers wurden den Probanden über Lautsprecher vorgespielt. Auf dem Computerbildschirm lief dabei ein eigens entwickeltes Visual-Basic-Programm ab, das von einem Laptop außerhalb der Kabine von dem Versuchsleiter gesteuert wurde. Der Bildschirminhalt des Laptops entsprach dem Inhalt des Bildschirms der Sprachkabine. Der Versuchsleiter konnte den Beginn der einzelnen Phasen des Experiments und die Präsentation der Stimuli steuern. Das Experiment begann mit der folgenden, auf dem Computerbildschirm ausgegebenen Instruktion, wobei der Proband dem Versuchsleiter signalisierte, dass er die Instruktion gelesen hatte:

Das folgende Experiment besteht aus zwei Phasen. In der ersten Phase werden Ihnen jeweils drei Sätze präsentiert, die Sie bitte einfach nur ganz normal vorlesen. Es werden Ihnen insgesamt vierundzwanzig Satzblöcke präsentiert, die alle die folgende Form haben:

"Mona bucht heute eine Reise."

"Mona heute."

"Heute bucht Mona eine Reise."

Die aktuelle Anzahl der bereits gelesenen Satzblöcke wird in der linken oberen Ecke des Bildschirms angezeigt. Jeder Satzblock erscheint für eine bestimmte Zeit auf dem Bildschirm, die in einem Fenster unter den Sätzen angezeigt wird.

Können wir beginnen?

Das Vorlesen von einzelnen kontextlos präsentierten B-Sätzen wurde wiederholt, wenn bei diesen Sätzen Versprecher u.ä. auftraten. Der Versuchsleiter passte die Präsentationsgeschwindigkeit der Tripel während des Experimentablaufs individuell an die Sprechgeschwindigkeit der einzelnen Sprecher an. Mit der Anpassung sollte gewährleistet werden, dass der Proband die Sätze zügig und ohne übermäßige Stilisierung vorlas. Die zweite Phase des Experiments, in der ebenfalls die Präsentationsdauer der B-Sätze angepasst wurde, begann mit der folgenden Instruktion:

In diesem Teil des Experiments geht es um das natürliche Lesen einfacher Dialoge mit verteilten Rollen. Jeder Dialog des Experiments besteht aus zwei Teilen, wobei Ihnen der erste Teil über Lautsprecher vorgespielt wird. Der dazugehörige zweite Teil erscheint anschließend als Text auf dem Bildschirm. Der zweite Teil besteht nur aus einem Satz bzw. aus der Kurzform eines Satzes.

Die Kurzform eines Satzes ist z.B. 'Mona bucht', der etwa der Satz 'Mona bucht heute eine Reise' zugrundeliegt. Der zugrundeliegende Satz einer Kurzform wird stets durch ihren Kontext deutlich, den über Lautsprecher vorgespielten ersten Teil des Dialogs.

In diesem Teil des Experiments ist es nun Ihre Aufgabe, sich zunächst den ersten Teil eines jeden Dialogs genau anzuhören. Lesen Sie dann bitte den dazugehörigen zweiten Teil als Reaktion auf den ersten Satz vor. Lesen Sie den zweiten Teil bitte so vor, dass er Ihrer Meinung nach optimal dazu passt, wie der erste Teil vorgelesen wurde. Dabei können Sie einen Satz auch ruhig erneut vorlesen, wenn Sie mit einer Variante nicht zufrieden sind. Beachten Sie dabei die Ihnen verbleibende Zeit. Insgesamt werden Ihnen 72 Dialoge vorgespielt.

Können wir beginnen?

Abbildung 3 ist ein (einfarbig wiedergegebener) Screenshot des Programms aus der zweiten Phase des Experiments:

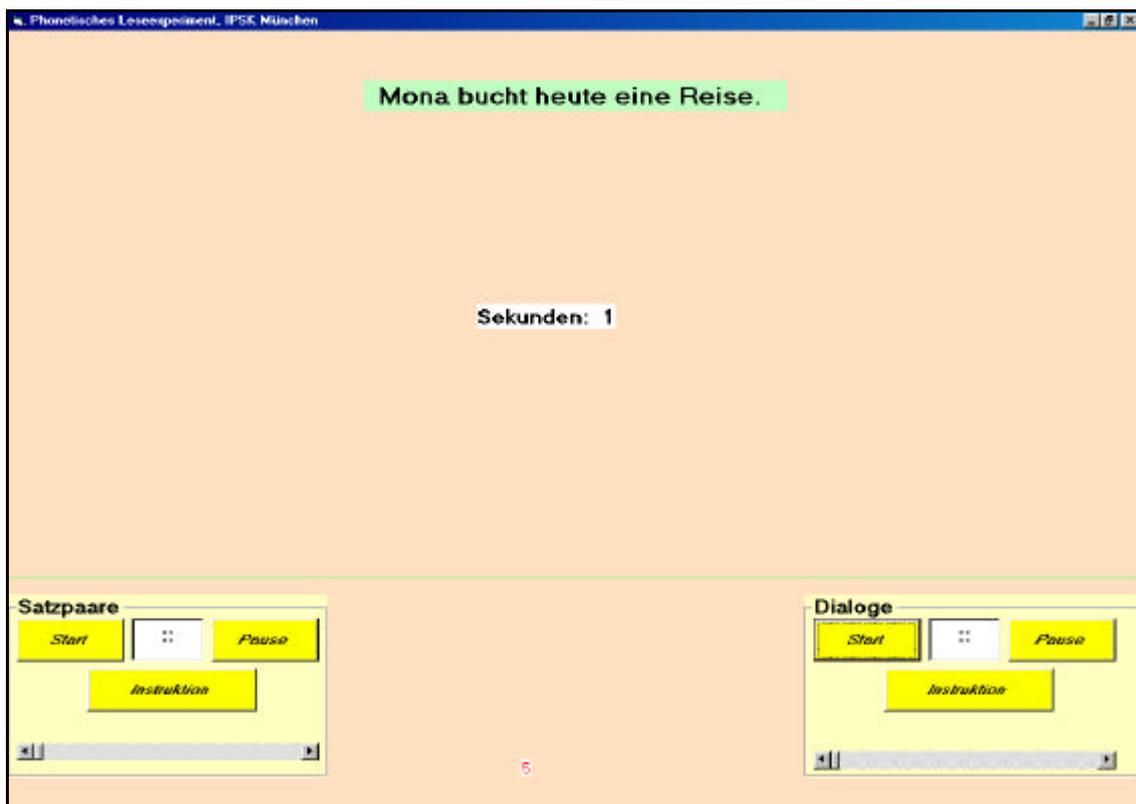


Abbildung 3: Elizitierung von Satzakzenten

Es wurden vier Probanden aufgenommen, die als Sprecher des Standarddeutschen bezeichnet werden können; Tabelle 2 beschreibt die Sprecher kurz:

Kürzel	Geschlecht	Geburtsort	Wohnort	Alter
1	m	Lindau	Augsburg	31
2	w	München	München	28
3	m	Göttingen	Hamburg	22
4	w	Göttingen	Hannover	19

Tabelle 2: Sprecher des Produktionsexperiments

Insgesamt liegt somit ein Korpus mit 576 Äußerungen des Deutschen vor. Jeder Proband produzierte dabei 144 Äußerungen, also 24 B1-Sätze, 24 B2-Sätze und 24 B3-Sätze sowie die entsprechenden 72 neutralen, kontextlos geäußerten B-Sätze ('out of the blue'). Die 288 kontextlos geäußerten Sätze heißen im folgenden neutrale Äußerungen, die verbleibenden 288 nicht kontextlos geäußerten Sätze hingegen akzentuierte Äußerungen⁴⁶. In diesem Sinne sind im folgenden auch Bezeichnungen wie etwa „neutrale Topikalisierung“ oder „akzentuierte Topikalisierung“ zu verstehen. Die 576 Äußerungen bilden insgesamt das Korpus MONA.

7.2.2 Ergebnisse

Die Aufnahmen der 576 Äußerungen des MONA-Korpus wurden manuell in Laute, Silben und Wortformen eingeteilt. Die unter Verwendung der SAMPA-Notation⁴⁷ durchgeführte Lautsegmentierung folgte den Konventionen von Geumann/Oppermann/Schaeffler (1997) und die Silbeneinteilung den Regeln des Duden Aussprachewörterbuchs (1990: 50f). Mit der in Praat (vgl. Anhang C, S. 110) implementierten Autokorrelationsmethode von Boersma (1993) erfolgte eine F0-Analyse der einzelnen Äußerungen. Anschließend wurden bestehende Oktavsprünge in den berechneten F0-Verläufen eliminiert. Kapitel 8, S. 81ff, analysiert diese F0-Verläufe nach unterschiedlichen phonetischen Intonationsmodellen.

Für die folgenden univariaten Analysen des MONA-Korpus sowie für die in Kapitel 9 vorgenommene multivariate Analyse wurden die F0-Verläufe der einzelnen Äußerungen um 10 Hz geglättet. Als Glättungsverfahren diente die in Praat implementierte Funktion 'Smooth'. Die 288 akzentuierten Äußerungen wurden für die univariaten Analysen in fünf Abschnitte eingeteilt:

- i) Der Vorlauf vom Beginn der Äußerung bis zum Beginn des ersten Akzentwortes, d.h. des Wortes mit der ersten - experimentell intendierten - kontrastiven Satzakkzentrealisierung.
- ii) Das erste Akzentwort.
- iii) Die Verbindung vom Ende des ersten Akzentwortes bis zum Anfang des zweiten Akzentwortes.
- iv) Das zweite Akzentwort.
- v) Der Nachlauf vom Ende des zweiten Akzentwortes bis zum Ende der Äußerung.

⁴⁶) Die Satzakkzentrealisierung in neutralen Äußerungen wird damit nicht behandelt.

⁴⁷) Speech Assessment Methods Phonetic Alphabet, <http://www.phon.ucl.ac.uk/home/sampa/home.htm>

Die Abschnitte i) bis v) wurden analog bei den 288 neutralen Äußerungen gebildet, so dass die Wörter des Vorlaufs, das erste 'Akzentwort' usw. der neutralen Äußerung den Wörtern des Vorlaufs, dem ersten Akzentwort usw. der korrespondierenden akzentuierten Äußerung entsprechen⁴⁸. Das Ansetzen des Wortes und nicht der Silbe als sprachliche Bezugseinheit von Satzakkenten erfolgt aus Gründen der Übersichtlichkeit, um grundsätzlich die Anzahl von fünf zu analysierenden Äußerungsabschnitten im folgenden nicht zu erhöhen. Die Modellierungen und multivariaten Analysen der Kapitel 8 und 9 gehen dagegen von der Silbe als sprachliche Bezugseinheit aus, den Annahmen der in der vorliegenden Arbeit untersuchten phonetischen Intonationsmodelle folgend.

Die vorgenommene Einteilung der neutralen Äußerungen in mit den akzentuierten Äußerungen korrespondierende Abschnitte dient der Untersuchung der akustischen Realisierung von kontrastiven Satzakkenten. In diesem Zusammenhang ist zunächst davon auszugehen, dass sich bei dem Vergleich der beiden Äußerungsvarianten die untersuchten akustischen Parameter bei den Akzentwörtern grundsätzlich stärker unterscheiden als bei den anderen Äußerungsabschnitten⁴⁹. So ist anzunehmen, dass etwa der Vergleich der Parameter des ersten Akzentwortes der akzentuierten Äußerung mit denen des korrespondierenden Akzentworts der neutralen Äußerung größere Unterschiede aufzeigt als etwa ein entsprechender Vergleich der beiden Vorläufe. Die folgenden Analysen zeigen, dass diese Annahme nur bedingt zutrifft.

Die univariate Analyse der einzelnen Äußerungsabschnitte verwendet in Halbtöne (ht) transformierte F₀-Werte, in Dezibel (db) gemessene Intensitätswerte sowie in Sekunden (sec) gemessene Zeitwerte. Die Verwendung dieser drei akustischen Parameter entspricht den Parametern aus Batliner (1991), dessen Analysen von F₀, Intensität und Länge bei Fokusakkenten u.a. eine sprecherspezifische Verwendung der drei akustischen Parameter ergab. Die folgende Auswertung arbeitet im einzelnen mit den im folgenden genannten Parametern, wobei an dieser Stelle eine Unterscheidung nach Satztypen angesetzt wird (vgl. Kap. 9.2, S. 99f):

ht und db - Mittelwert (mean), -Standardabweichung (std), -Maximum (max) und -Minimum (min)

Dauer von Vorlauf, erstes Akzentwort, Verbindung, zweites Akzentwort, Nachlauf

Zeitpunkt von ht(max), ht(min), db(max) und db(min) bzgl. des ersten und zweiten Akzentwortes

Die Mittelwerte, die Standardabweichungen, die maximalen und die minimalen Werte sind auf die einzelnen Sprecher relativiert. Von jedem berechneten Wert einer Äußerung wurde ein für den einzelnen Sprecher berechneter Mittelwert subtrahiert, also etwa von einem Halbtonmaximum eines Akzentwortes einer bestimmten Äußerung der Mittelwert aller Halbtonmaxima des jeweiligen Sprechers. Die Dauern sind dagegen auf die Dauer der einzelnen Äußerungen relativiert, d.h. es wurde etwa die Dauer eines ersten Akzentwortes durch die Dauer der gesamten Äußerung geteilt. Der Relativierung der einzelnen Zeitpunkte liegen die entsprechenden sprachlichen Einheiten Vorlauf, erstes Akzentwort usw. zugrunde. Demnach wurde beispielsweise ein

⁴⁸) Vorlauf, Verbindung und/oder Nachlauf bestehen nicht in allen Äußerungen, vgl. Anhang A.

⁴⁹) 'Sich stärker unterscheiden' ist keine objektivierte Bewertung der untersuchten akustischen Parameter und von daher rein deskriptiv zu verstehen; eine derartige Bewertung dieser wie der Parameter der kontrastiven Satzakkente der phonetischen Intonationsmodelle erfolgt in Kapitel 9 bei ihrer multivariaten Analyse.

relativer Zeitpunkt t_{rel} von einem Halbtonmaximum zum Zeitpunkt t eines Akzentwortes, welches bei t_1 beginnt und bei t_2 endet, folgendermassen berechnet:

$$t_{rel} = (t - t_1) / (t_2 - t_1)$$

Die univariaten Analysen unterscheiden im folgenden nach den drei in dem Produktionsexperiment verwendeten Satztypen (ST), der normalen Wortstellung (ST1), der Ellipse (ST2) und der Topikalisierung (ST3). Wie oben beschrieben, werden die fünf Äußerungsabschnitte Vorlauf (Vor), erstes Akzentwort (AW1), Verbindung (Ver), zweites Akzentwort (AW2) und Nachlauf (Nach) der beiden Äußerungsvarianten gegenübergestellt. Die ht- bzw. db-Mittelwerte gelten dabei als Index für die globalen Tonhöhen- bzw. Lautheitsverläufe, wohingegen die entsprechenden Standardabweichungen als Index für deren Variation der Verläufe aufgefaßt werden. Die verschiedenen Zeitwerte geben weiteren Aufschluß über die Verläufe. In diesem Zusammenhang ist zu beachten, dass im Gegensatz zu den beiden anderen Satztypen bei der Ellipse Vorlauf, Verbindung und Nachlauf, sofern überhaupt vorhanden, mit dem ersten bzw. dem zweiten Akzentwort stets eine syntaktische Konstituente unterhalb des Satzes bilden.

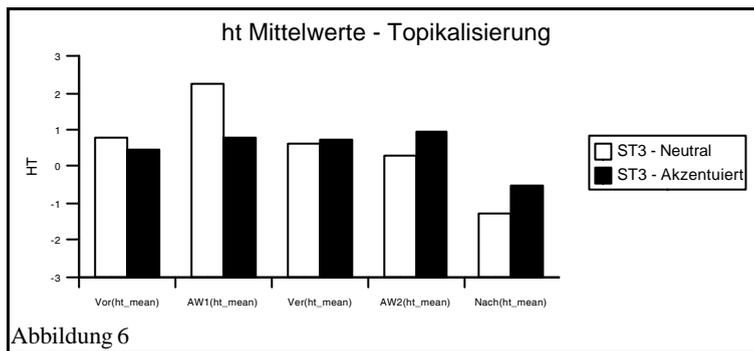
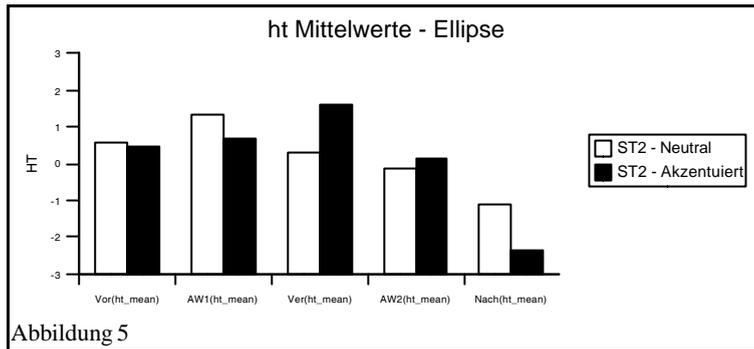
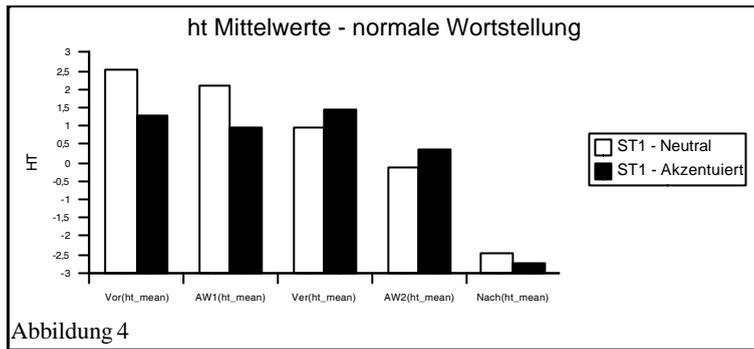
7.2.2.1 Halbtonwerte

Die berechneten F0-Werte x der Äußerungen wurden nach einer in Praat (vgl. Anhang C, S. 110) implementierten Formel relativ zu 100 Hz in Halbtöne transformiert:

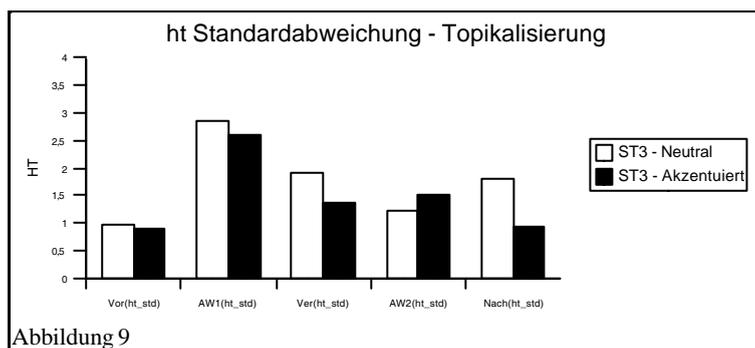
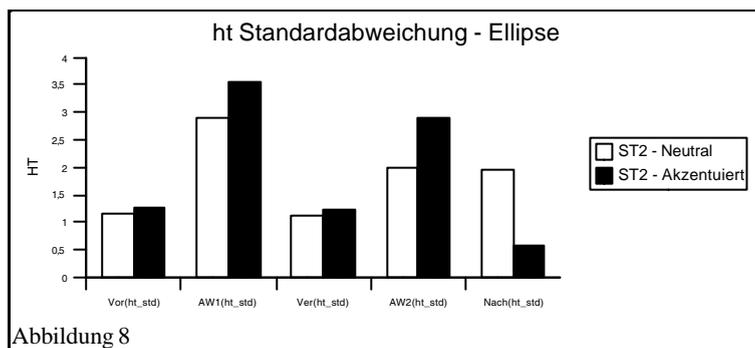
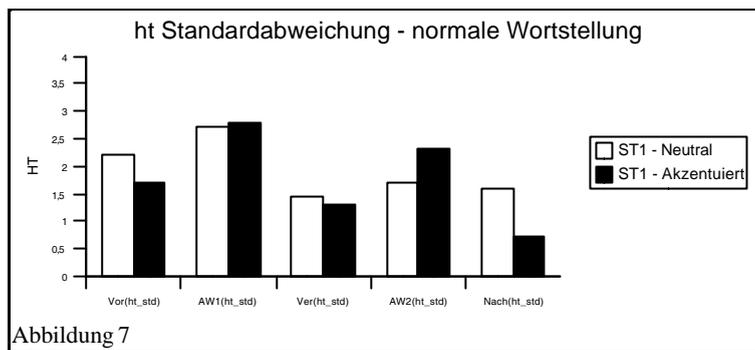
$$ht = 12 * \ln(x/100) / \ln 2$$

Die folgenden Abbildungen 4 bis 6 enthalten die gemittelten Halbtonwerte der fünf Äußerungsabschnitte je eines Satztyps unterschieden nach neutralen und akzentuierten Äußerungen. Die neutralen Äußerungen haben dabei ab dem ersten Akzentwort⁵⁰ generell fallende Halbtonmittelwerte. Hingegen steigen bei den akzentuierten Sätzen die Halbtonmittelwerte tendenziell bis zur Verbindung (neutrale Wortstellung, Ellipse) bzw. bis zu dem zweiten Akzentwort (Topikalisierung) an. Die gemittelten Halbtonwerte der neutralen Äußerungen sind bei dem ersten Akzentwort deutlich größer als die entsprechenden Halbtonmittelwerte der akzentuierten Äußerungen, bei dem zweiten Akzentwort dagegen etwas schwächer. Am deutlichsten ist dies bei der Topikalisierung in Abbildung 6 zu beobachten. Hier ist der Halbtonmittelwert der neutralen Sätze bei dem ersten Akzentwort ausgezeichnet hoch und fällt im Anschluss. Bei den akzentuierten Sätzen ist der Mittelwert des ersten Akzentwortes deutlich kleiner, innerhalb der Äußerung allerdings fast so groß wie der Halbtonmittelwert des zweiten Akzentwortes.

⁵⁰) Wie bereits erwähnt, beruht die Bezeichnung „Akzentwort“ bei den neutralen Äußerungen allein auf einer positionellen Übereinstimmung mit dem entsprechenden Akzentwort der akzentuierten Äußerungen; s.a. die Hinweise auf S. 57 unten.



Abbildungen 7 bis 9 enthalten die entsprechenden Standardabweichungen der Halbtonwerte. Die beiden ersten Satztypen weisen die größte Variation bei den Akzentwörtern auf, wobei die Differenz zwischen den neutralen und den akzentuierten Äußerungen bei dem zweiten Akzentwort ausgeprägter ist. Eine vergleichbar stärkere Ausprägung ist auch bei den Vor- und Nachläufen der akzentuierten Äußerungen des ersten und zweiten Satztyps zu verzeichnen. Die topikalisierten Äußerungen unterscheiden sich vor allem bei dem zweiten Akzentwort, welches bei der akzentuierten Variante eine größere Variation als bei der neutralen Variante hat.



In den Abbildungen 10 und 11 werden die gemittelten Maximal- und Minimalwerte der einzelnen Äußerungsabschnitte von neutralen und akzentuierten Äußerungen je Satztyp verglichen. Die beiden Abbildungen stellen die Halbtonmaxima bzw. -minima für die normale Wortstellung dar. Die Halbtonmaxima der einzelnen Abschnitte neutraler Äußerungen haben eine fallende Tendenz, wohingegen die Maxima der akzentuierten Äußerungen bis zu dem zweiten Akzentwort eine gleichbleibende bis leicht steigende Tendenz aufweisen und beim Nachlauf zudem deutlich kleinere Werte aufweisen. Bei dem ersten Akzentwort sind die Halbtonmaxima neutraler Äußerungen größer, bei dem zweiten Akzentwort deutlich kleiner als bei akzentuierten Äußerungen. Betrachtet man die Halbtonminima neutraler Äußerungen, so weisen diese ähnlich wie die Maxima ab dem ersten Akzentwort eine fallende Tendenz auf. Im Gegensatz dazu variieren die Halbtonminima der akzentuierten Sätze zwischen den Äußerungsbestandteilen stärker.

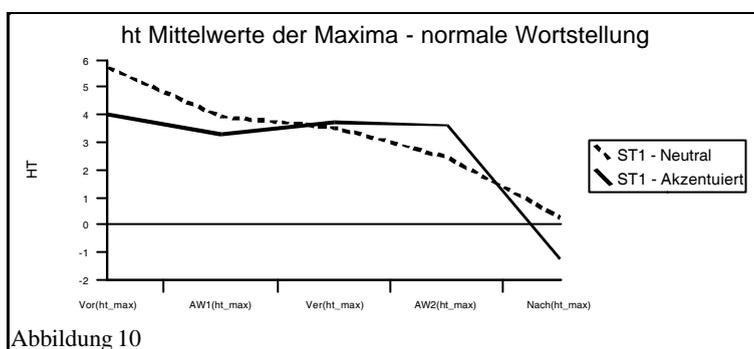


Abbildung 10

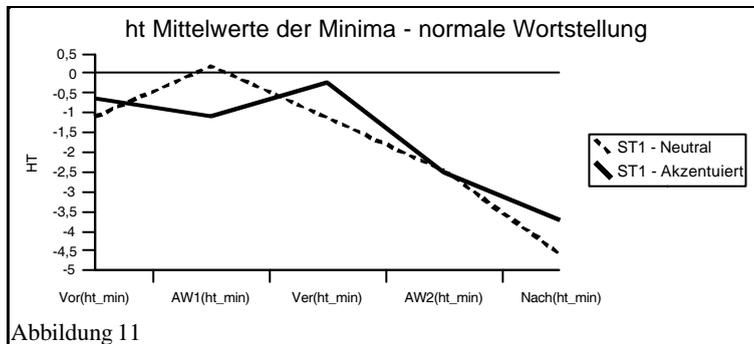


Abbildung 11

Abbildung 12 enthält die Maxima des Satztyps Ellipse. Das zweite Akzentwort ist bei der akzentuierten Äußerung besonders stark ausgeprägt, und es tritt nicht ein Abfall wie bei der neutralen Verbindung auf. Vielmehr steigen die Maxima der akzentuierten im Gegensatz zu der neutralen Variante kontinuierlich, wobei zu beachten ist, dass sich die Verbindung der Ellipse im Hinblick auf die Anzahl der Wörter wie auf deren syntaktischen Status stark von der der anderen Satztypen unterscheidet. Die Ausprägung der Minima in Abbildung 13 zeigt für neutrale Äußerungen eine ähnliche Entwicklung wie in Abbildung 11, Halbtonminima - normale Wortstellung. Bei den akzentuierten Varianten nehmen die Halbtonminima ein erstes - lokales - Minimum bei Akzentwort 1 und ein zweites bei Akzentwort 2 ein.

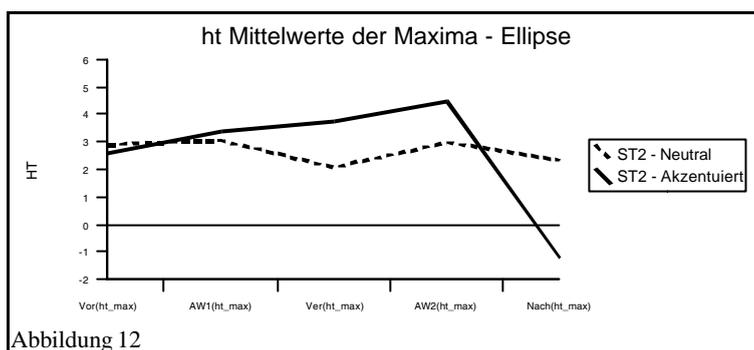


Abbildung 12

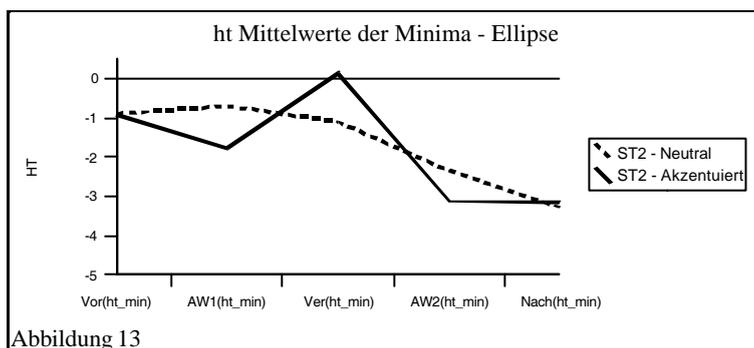


Abbildung 13

Die Maxima in Abbildung 14 haben bei beiden Varianten einen vergleichbaren Verlauf, außer dass die akzentuierten Varianten in dem Verlauf von der Verbindung bis dem zweiten Akzentwort schwächer fallen. Die Halbtonminima der Topikalisierung in Abbildung 15 zeigen für Vorlauf und erstes Akzentwort vergleichbare Einwicklungen. Die neutrale Variante zeigt anschließend eine fast gleichbleibende Tendenz, wohingegen die Halbtonminima bei der Verbindung ein lokales Maximum mit anschließend leicht abfallendem Trend aufweist.

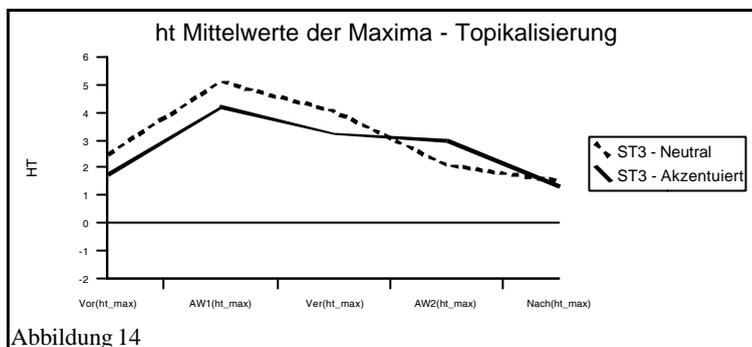


Abbildung 14

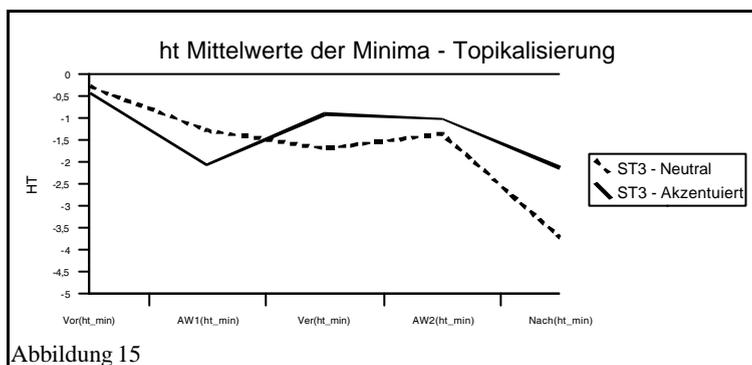
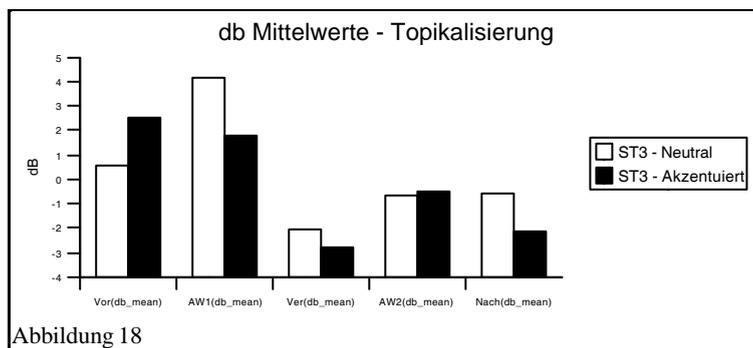
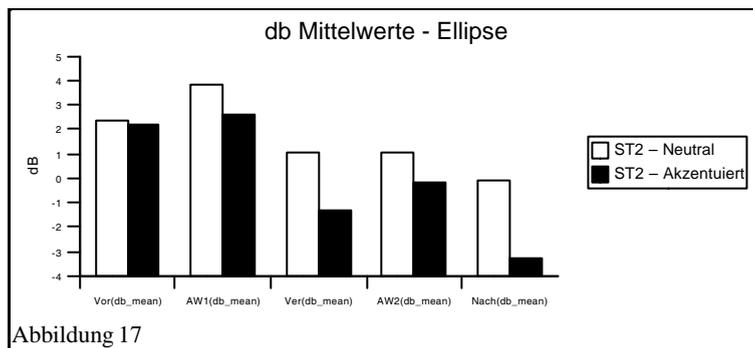
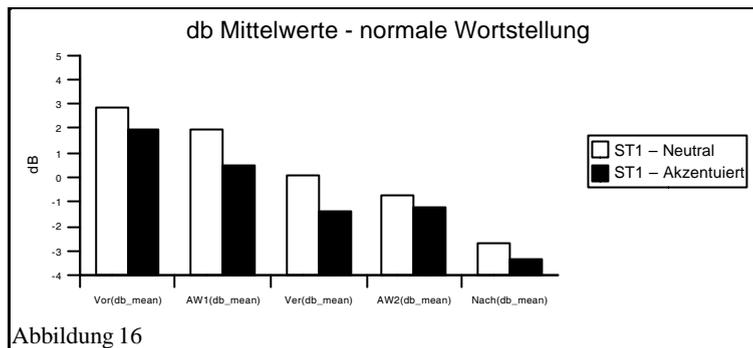


Abbildung 15

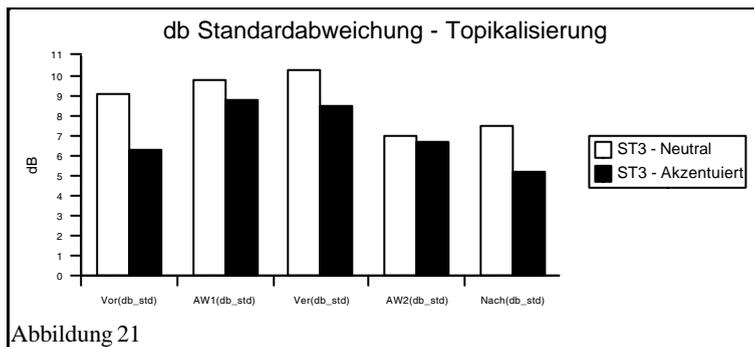
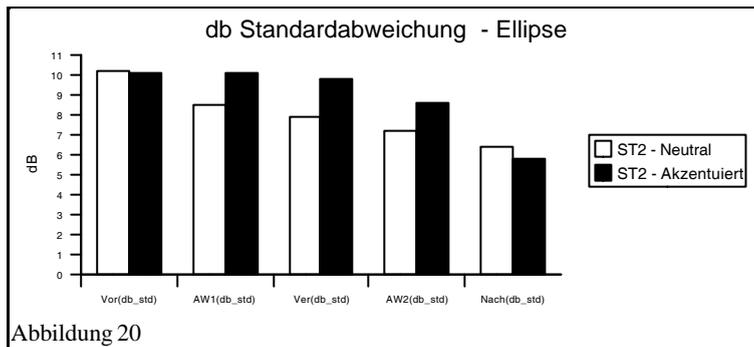
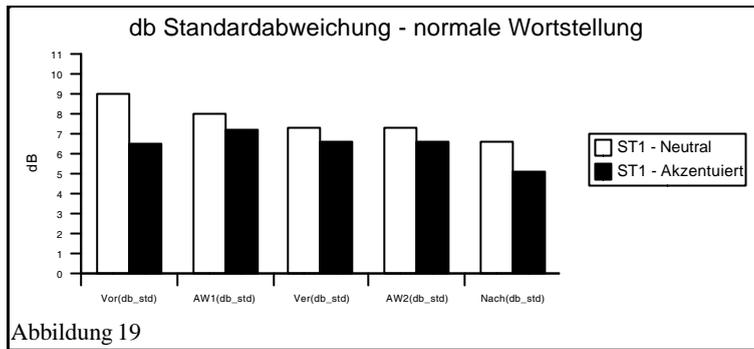
Zusammengefasst über alle Satztypen liegen die Halbtonmaxima neutraler Äußerungen bei dem ersten Akzentwort über den entsprechenden Werten der akzentuierten Äußerungen; bei dem zweiten Akzentwort verhält sich diese Ausprägung genau umgekehrt. Bei den Halbtonminima liegt der Wert für akzentuierte Sätze bei dem ersten Akzentwort jeweils unter dem Wert für neutrale Sätze. Dieser Wert kann als lokales Minimum angesehen werden, d.h. es folgt ein Anstieg auf den Wert bei der Verbindung. Der weitere Verlauf ist abhängig von dem jeweiligen Satztyp.

7.2.2.2 Intensitätswerte

Abbildungen 16 bis 18 geben die gemittelten Intensitätswerte wieder. Insgesamt läßt sich bei allen drei Satztypen sowohl bei den neutralen als auch bei den akzentuierten Äußerungen eine fallende Tendenz beobachten. Die neutralen Varianten wurden im Mittel lauter gesprochen. Die akzentuierten Varianten weisen bei dem zweiten Akzentwort eine positive Ausprägung auf, d.h. sie werden im Vergleich zu Verbindung und Nachlauf der Äußerung durchschnittlich mit größerer Intensität realisiert. Die neutralen Varianten der Sätze haben diese Ausprägung nicht. Dagegen sind die Werte des ersten Akzentwortes der Ellipse und der Topikalisierung bei der neutralen im Vergleich mit der akzentuierten Variante größer.



Abbildungen 19 bis 21 illustrieren die relative Konstanz der Variationsweite der beiden Äußerungsvarianten bei der normalen Wortstellung und der Ellipse. Bei beiden Satztypen haben die Werte eine fallende Tendenz. Die akzentuierten Varianten haben bei der Ellipse zudem die insgesamt größere Standardabweichung. Die Topikalisierungen hingegen variieren ein wenig stärker, wobei die neutralen Varianten hohe Werte bei der Verbindung und die akzentuierten Varianten ein vergleichbar hohe Werte bei dem ersten Akzentwort aufweisen. Im Anschluss daran fallen bei beiden topikalisierten Äußerungsvarianten die Werte der Standardabweichung.



Die Abbildungen 22 und 23 stellen die dB Maxima und Minima der einzelnen Satztypen dar. Die dB Maxima und Minima der beiden Äußerungsvarianten besitzen bei der normalen Wortstellung eine relativ parallele fallende Tendenz. Die dB Maxima der akzentuierten Äußerung weichen dabei lediglich bei dem zweiten Akzentwort ab, das etwas lauter gesprochen wird. Eine weitere auffällige Veränderung des Intensitätsbereichs zeigt der Wert des dB Minimums bei dem Vorlauf der neutralen Variante.

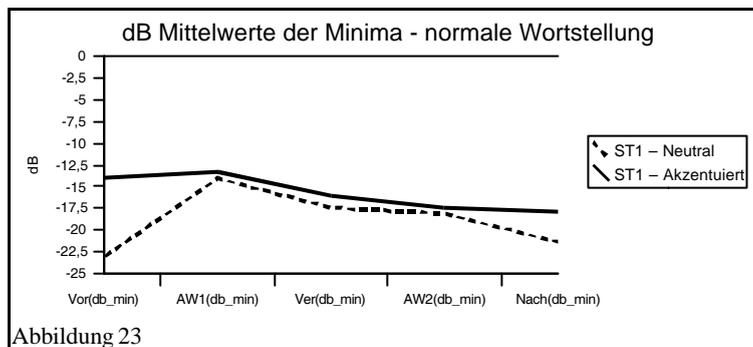
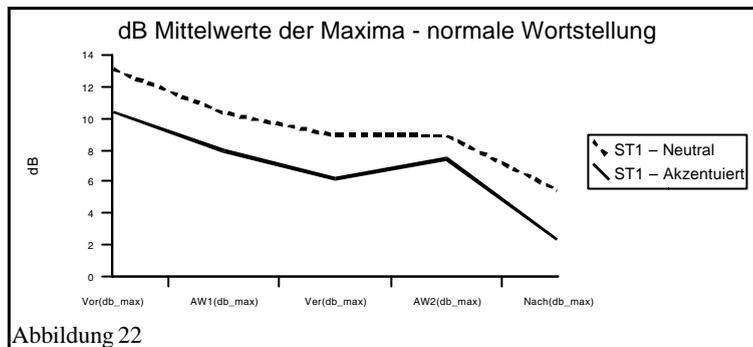


Abbildung 24 zeigt, dass die dB Maxima der Ellipse in beiden Varianten eine vergleichbare Homogenität wie die Maxima des ersten Satztyps aufweisen, wobei lediglich der Nachlauf der akzentuierten Varianten eine größere Veränderung des Intensitätsbereichs zeigt. Die dB Minima der neutralen Variante der Ellipse in Abbildung 25 besitzen ähnliche Werte wie die dB Minima der normalen Wortstellung aus Abbildung 23. Auch hier findet sich die größte positive Veränderung des Intensitätsbereichs beim Vorlauf, anschließend fallen die Werte kontinuierlich. Im Vergleich dazu zeigen die Werte der akzentuierten Varianten einen insgesamt konstanteren Verlauf.

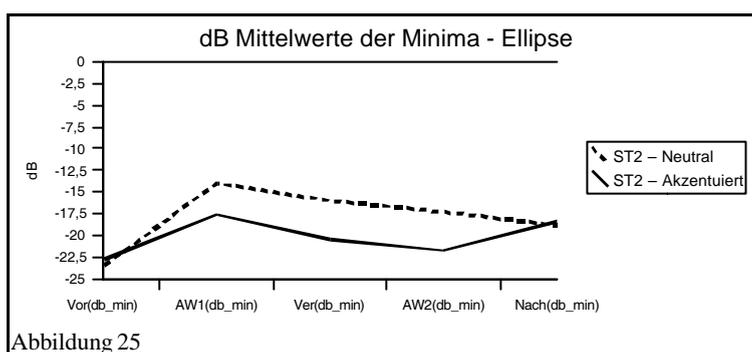
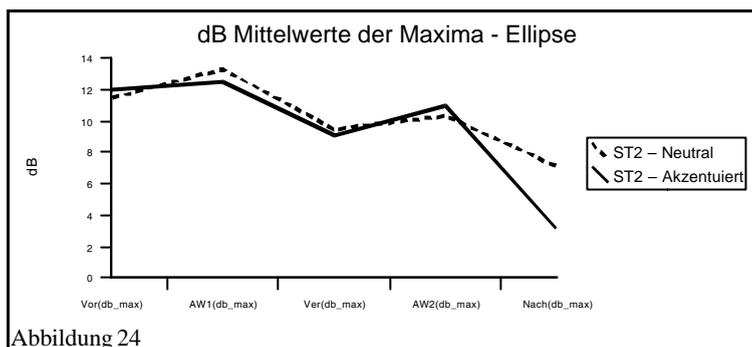


Abbildung 26 stellt die unterschiedliche Ausprägung der Maxima bei dem dritten Satztyp dar. Die neutralen Äußerungen haben einen Gipfel bei dem ersten Akzentwort mit anschließend fallender Tendenz. Die akzentuierten Äußerungen weisen zwar beim ersten Akzentwort eine schwächer ausgeprägte Gipfelkontur auf, realisieren jedoch auch bei dem zweiten Akzentwort einen Gipfel, der mit einem starken Abfall beim Nachlauf der Äußerung verbunden ist. Abbildung 27 zeigt – mit Ausnahme des Vorlaufs – einen relativ parallelen Verlauf der dB Minima der beiden Äußerungsvarianten. Beide Varianten zeigen bei der Verbindung ein lokales Minimum, bei dem zweiten Akzentwort ein lokales Maximum der dB Minima, die im Nachlauf wieder kleiner werden. Ähnlich wie bei den beiden anderen Satztypen ist bei den neutralen Äußerungen zwischen Vorlauf und ersten Akzentwort eine wesentlich stärkere Änderung der dB Minima zu beobachten als bei den neutralen Äußerungen.

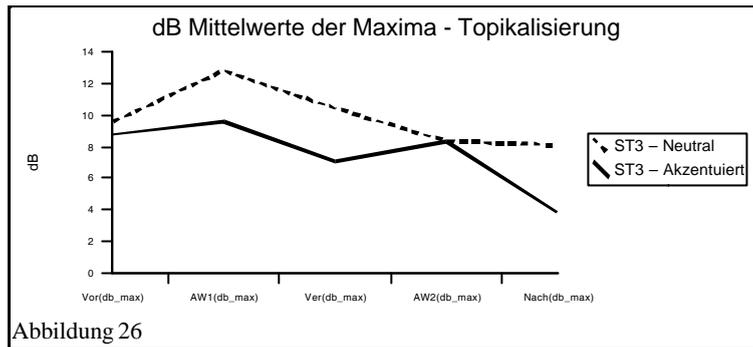


Abbildung 26

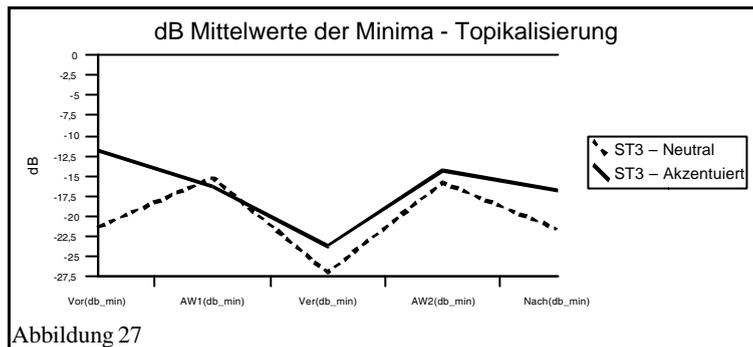


Abbildung 27

Die Abbildungen 22 bis 27 zeigen, dass sich die neutralen bzw. akzentuierten Varianten hinsichtlich ihrer dB Maxima bzw. Minima nicht wesentlich unterscheiden. Es gibt nur ein lokales Maximum der dB-Maxima bei dem zweiten Akzentwort, also eine erhöhte Intensität im Vergleich zu Verbindung und Nachlauf. Diese besondere Markierung des zweiten Akzentwortes findet sich auch bei den dB Mittelwerten (vgl. Abbildung 16 bis 18).

7.2.2.3 Zeitwerte

Die Abbildungen 28 – 32 stellen das Verhältnis von einzelnen Zeitwerten der akzentuierten im Vergleich mit den neutralen Äußerungen dar. Das Verhältnis ist auf eins normiert, d.h. die entsprechenden Zeitwerte unterscheiden sich bei beiden Äußerungsvarianten in diesem Fall nicht. Abbildung 28 enthält für alle Satztypen die Dauerwerte der einzelnen Äußerungsabschnitte in Sekunden, wobei ein positiver Wert von beispielsweise 1.05 besagt, dass die akzentuierte Äußerung an dieser Stelle um 5% länger dauert als die neutrale Äußerung. Ein negativer Wert kennzeichnet eine entsprechend kürzere Dauer der akzentuierten Äußerung. Die Abbildungen 29 - 32 sind analog mit der Ausnahme konzipiert, dass in Abbildung 28 die Wortakzentsilben des ersten und zweiten Akzentwortes als zusätzliche Äußerungsabschnitte aufgeführt sind (AS1, AS2)⁵¹.

Abbildung 28 zeigt, dass die Dauern der Akzentsilben wie die der Akzentwörter der akzentuierten Äußerungen mit zwei Ausnahmen bei allen Satztypen länger als die der neutralen Äußerungen ist. Die erste Ausnahme besteht in der kürzeren Dauer der ersten Akzentsilbe der Ellipsen und die zweite in der Gleichheit der relativen Dauern des zweiten Akzentwortes der Ellipsen. Bei dem Vor- und Nachlauf sind mit Ausnahme der Topikalisierungen die neutralen Äußerungen von größerer Dauer.

⁵¹) Die Dauern der Silben sind nach der Formel auf S. 59 auf das entsprechende Akzentwort relativiert.

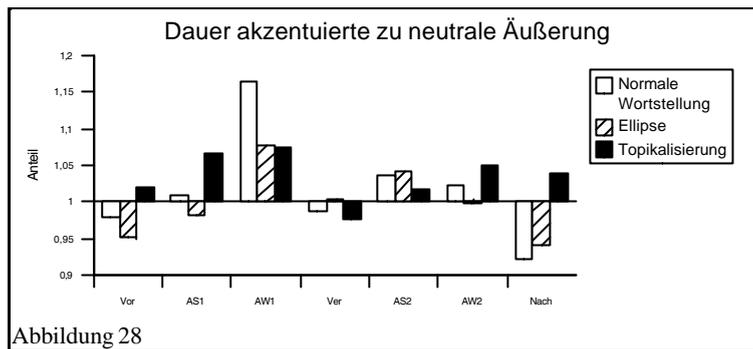


Abbildung 29 zeigt, dass die Maxima der Halbtonwerte an den relativen Positionen des ersten und des zweiten Akzentworts komplementär verteilt sind. Im Vergleich zu den neutralen tritt bei den akzentuierten Äußerungen das Halbtonmaximum im ersten Akzentwort generell später ein, wohingegen es im zweiten Akzentwort früher realisiert ist. Das Halbtonmaximum in der Verbindung zwischen den beiden Akzentwörtern tritt in allen Satztypen ebenfalls später auf. Unterschiede im Zeitpunkt ergeben sich zwischen den verschiedenen Satztypen nur bei dem Vor- und insbesondere bei dem Nachlauf.

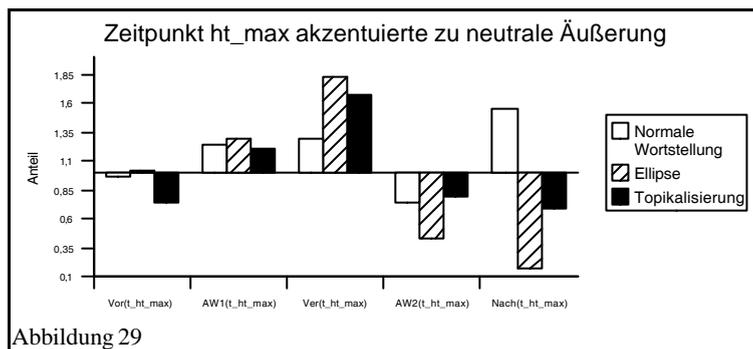
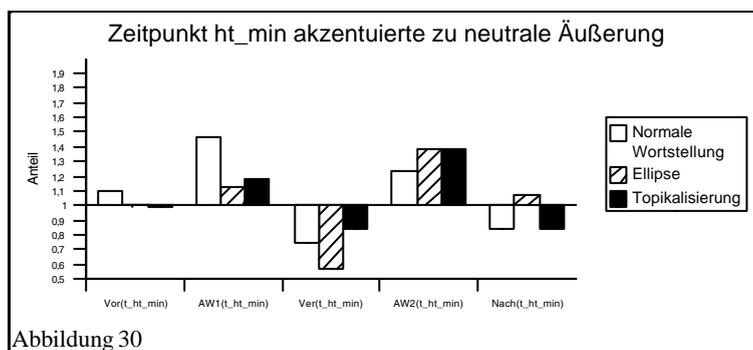


Abbildung 30 ist zu entnehmen, dass die Ausprägung der relativen Zeitpunkte der Halbtonminima bei dem ersten und zweiten Akzentwort sowie bei der Verbindung bzgl. der drei Satztypen relativ homogen ist, wenn man von der quantitativen Ausprägung im einzelnen absieht. Bei den Akzentwörtern tritt das Halbtonminimum der akzentuierten Äußerungen relativ später auf, in der Verbindung früher. Auch hier sind Unterschiede zwischen den drei Satztypen nur bei dem Vorlauf und vor allem beim Nachlauf zu erkennen.



Die relativen Zeitpunkte der Intensitätsmaxima aus Abbildung 31 verhalten sich ähnlich wie die die Zeitpunkte der Halbtonmaxima (vgl. Abbildung 29). Auch hier treten die Intensitätsmaxima von akzentuierten Äußerungen bei dem ersten Akzentwort und bei der Verbindung im Vergleich zu den neutralen Äußerungen später auf. Ausser bei den Topikalisierungen tritt das Intensitätsmaximum bei dem zweiten Akzentwort früher auf. Im Nachlauf hingegen werden bei allen Satztypen die Intensitätsmaxima deutlich früher bei akzentuierten als bei neutralen Äußerungen erreicht.

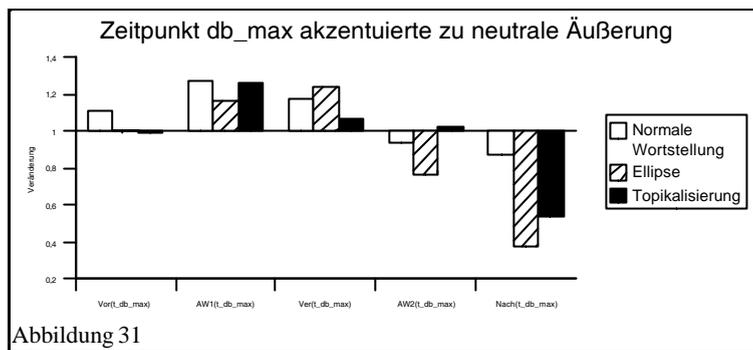
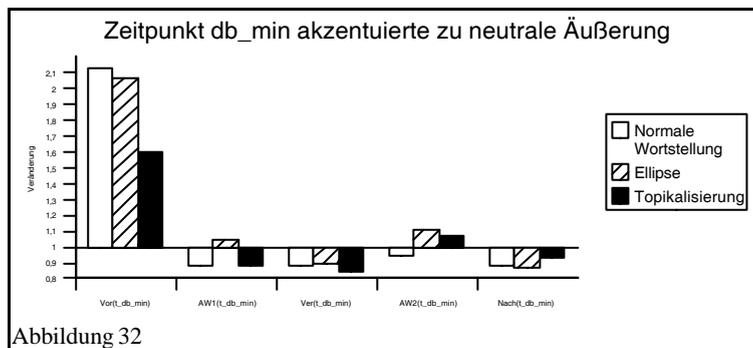


Abbildung 32 schließlich zeigt signifikante Unterschiede bei dem Zeitpunkt der Intensitätsminima des Vorlaufs. Diese treten bei akzentuierten Äußerungen für alle Satztypen sehr viel später ein. Homogen verhalten sich auch die Zeitpunkte der Intensitätsminima in den Abschnitten Verbindung und Nachlauf, die bei akzentuierten Äußerungen etwas früher auftreten. Leichte Unterschiede zwischen den Satztypen finden sich bei den beiden Akzentwörtern, wobei hier die Intensitätsminima jedoch maximal 10% früher bzw. später eintreten, was unterhalb der Verschiebungen anderer Äußerungsbereiche liegt.



7.2.3 Diskussion des Produktionsexperiments

Die Beschreibungen der in Kapitel 7.2.2 vorgestellten akustischen Parameter lassen zusammenfassend betrachtet unterschiedliche akustische Formen der akzentuierten und der neutralen Äußerungsvarianten erkennen. Die neutralen Varianten weisen demnach tendenziell einen ausgeprägten Anstieg der Halbton- und Intensitätswerte bis zu dem ersten Akzentwort auf, die anschließend bis zum Äußerungsende durchgehend fallen. Die akzentuierten Varianten hingegen bewegen sich zunächst insgesamt in einem kleineren absoluten Wertebereich. Sie haben ebenfalls einen, im Vergleich mit den neutralen Varianten allerdings schwächeren Anstieg bei dem ersten Akzentwort, dem gleichermaßen ein Abfall bis zum Äußerungsende folgt. Im Gegensatz zu den neutralen ist bei den akzentuierten Varianten noch ein lokales Maximum der Halbton- und Intensitätswerte bei dem zweiten Akzentwort zu erkennen. An dieser Stelle korrelieren Halbton- und Intensitätswerte der akzentuierten Varianten zudem besonders deutlich. Unterschiedliche akustische Formen finden sich darüber hinaus bei den relativen Dauerwerten der Akzentsilben, die bei den akzentuierten Varianten grundsätzlich größer als bei den neutralen Varianten sind.

Die unterschiedlichen akustischen Formen der akzentuierten und der neutralen Varianten entsprechen nicht der Erwartung des Produktionsexperiments, dass sich die beiden Varianten lokal bei den Akzentwörtern bzw. bei den entsprechenden Akzentsilben eher stark und bei den anderen Äußerungsteilen, also dem Vorlauf, der Verbindung und dem Nachlauf, eher schwach unterscheiden. Demnach ist vielmehr von einer auf die gesamte Äußerung bezogene, d.h. von einer global zu analysierenden Veränderung der einzelnen akustischen Parameter auszugehen, die die akzentuierten von den neutralen Äußerungen trennt. Komplexe Intonationsmuster wie der in Kapitel 5.2, S. 30ff, vorgestellte Brückenakzent bieten dabei einen Ausgangspunkt für weitere Untersuchungen.

An dieser Stelle ist allerdings darauf hinzuweisen, dass Sprecher (des Deutschen) komplexe wichtige Informationen im Diskurs wie der untersuchte Ausdruck von gleich zwei Kontrasten in dem oben beschriebenen Sinne im normalen, alltäglichen Sprachgebrauch eher nicht mit intonatorischen Mitteln übertragen. Die relativ geringe unterschiedliche Ausprägung der Standardabweichung der akzentuierten und der neutralen Äußerungen zeigt in diesem Zusammenhang, dass sich die beiden Äußerungsvarianten nicht besonders stark unterscheiden. Daraus ist zu folgern, dass zumindest bei den Sprechern des Produktionsexperiments, bei denen es sich absichtlich nicht um professionelle, geschulte Sprecher handelt, der intonatorische Ausdruck von zwei Kontrasten - innerhalb einer syntaktisch relativ einfachen Äußerung - vermutlich nicht Bestandteil ihrer alltäglichen Sprachperformanz ist. Die Diskussion des im folgenden beschriebenen Perzeptionsexperiments greift diesen Gedanken wieder auf (vgl. Kap. 7.3.3, S. 79).

7.3 Perzeptionsexperiment

Die im Produktionsexperiment gewonnenen akzentuierten Äußerungen wurden in einem Perzeptionsexperiment untersucht, in dem Probanden die Realisierung der Satzakkente indirekt bestätigen sollten. Das Perzeptionsexperiment untersuchte nur einen Teil der akzentuierten Äußerungen, weil sonst der Aufwand an Versuchspersonen zu groß gewesen wäre. Damit ist genau genommen nur bei diesen Äußerungen von realisierten Satzakkenten auszugehen, wobei der Schluss von den untersuchten Äußerungen auf das gesamte Korpus allerdings plausibel ist.

Das Perzeptionsexperiment verwendete ebenfalls das Dialogschema von Lieb (1983). In diesem Fall wurde in den A-B-Dialogen nur der nicht eingeleitete Nebensatz des ursprünglichen A-Satzes aus dem Produktionsexperiment verwendet.

A'. Klaus bucht morgen eine Reise.

B. Mona bucht heute eine Reise.

Der A'-Satz wurde eingesetzt, damit den Probanden kürzere und einfachere Stimuli präsentiert werden konnten. Außerdem wurde das Experiment im Internet durchgeführt, so dass wegen der Übertragungszeiten kleinere Sounddateien wünschenswert waren⁵².

7.3.1 Aufbau und Durchführung

Jeder Proband füllte zu Beginn des Experiments einen Fragebogen aus, in dem die folgenden Daten erhoben wurden:

Muttersprache

Muttersprache der Mutter

Muttersprache des Vaters

Bundesland/Ort (ggf. mit genaueren geographischen Angaben) der Grundschule

Bundesland/Ort (ggf. mit genaueren geographischen Angaben), der am stärksten den Dialekt geprägt hat

Beruf

Alter

Geschlecht

Der Ablauf des Experiments wurde von einem eigenen Perl-Script protokolliert, das Zeitpunkte, Reihenfolge und Häufigkeit der aufgerufenen Sounddateien der einzelnen Dialoge erfasste. Damit konnten zumindest nicht oder nur teilweise aufgerufene Dialoge, die von den Probanden aber dennoch eine Bewertung erhielten, von der Auswertung des Experiments ausgeschlossen werden. Nach der Erfassung der Personendaten erhielten die Probanden auf der WWW-Seite die folgende Instruktion:

⁵²) Für die Teilnahme an dem Experiment erhielten die Probanden eine Telefonkarte im Wert von 6 DM, die die entstehenden Telefongebühren decken sollte.

In diesem Experiment zur Betonung in der deutschen Sprache hören Sie verschiedene Satzpaare und beurteilen deren Zusammengehörigkeit. Jedes Satzpaar sollte einen kleinen Dialog bilden, wobei der zweite Satz eines jeden Dialogs allein aufgrund seiner Betonung gut oder schlecht in einen Dialog passen kann. Das folgende Beispiel verdeutlicht diese Rolle der Satzbetonung; in dem Beispiel würde der Satz des Sprechers B gut in den Dialog passen, wenn "Auto" betont wird, und er würde schlecht in den Dialog passen, wenn Sprecher B etwa "neues" betont:

Sprecher A: "Peter hat ein neues Fahrrad." Sprecher B: "Er hat ein neues Auto."

In dem Experiment sollten Sie sich jedes Satzpaar stets als Teil eines ganz alltäglichen Gesprächs zwischen zwei Personen vorstellen, die sich gerade über eine Reisebuchung Ihrer beiden Bekannten Klaus und Mona unterhalten. Hören Sie sich jedes Satzpaar genau an, und prüfen Sie, ob die Sätze auch wirklich einen Dialog zwischen den beiden Sprechern bilden. Bedenken Sie, dass ein Dialog mindestens voraussetzt, dass der zweite Sprecher den Satz des ersten Sprechers richtig verstanden hat. Erst dann kann der zweite Sprecher überhaupt richtig auf den ersten Satz reagieren. Um dies beurteilen zu können, müssen Sie sich natürlich auch den Satz des ersten Sprechers ganz genau anhören.

Wenn Sie beim Abhören des zweiten Satzes den Eindruck haben, dass der zweite Sprecher nicht richtig auf den Satz des ersten Sprechers reagiert, dann klicken Sie bitte unten auf schlecht. In diesem Fall handelt es sich um ein Satzpaar, das keinen passenden Dialog bildet. Derartige Satzpaare wurden in dem vorliegenden Experiment künstlich zusammengestellt, indem der zweite Satz aus einem anderen, vergleichbaren Dialog einfach in dieses künstliche Satzpaar kopiert wurde.

Wenn Sie hingegen den Eindruck haben, dass die beiden Sätze einen Dialog bilden, dann wählen Sie unten bitte gut oder einigermaßen. Gut besagt, dass der zweite Sprecher a) den Satz des ersten Sprechers richtig verstanden hat und b) Ihrer Meinung nach in dieser Gesprächssituation wirklich gut passend auf den Satz des ersten Sprechers reagiert. Einigermaßen besagt, dass der zweite Sprecher a) den Satz des ersten Sprechers ebenfalls richtig verstanden hat, aber b) Ihrer Meinung nach in dieser Gesprächssituation nur einigermaßen passend auf den Satz des ersten Sprechers reagiert. Wenn Sie bei einem Satzpaar aus irgendwelchen Gründen überhaupt nicht beurteilen können, ob es sich nun um einen Dialog handelt oder nicht, dann wählen Sie bitte keine Bewertung.

Beginnen Sie mit den folgenden Satzpaaren, indem Sie einfach die einzelnen Sätze anklicken. Bitte bedenken Sie, dass eine ungestörte Ausgabe der Sprachdateien Ihre Bewertung wesentlich erleichtern wird - überprüfen Sie vor der Bewertung auch, ob Ihre Soundausgabe problemlos funktioniert.

Die Skalierung der Antwortmöglichkeiten folgt einer Idee von Richter (1967). Abbildung 33 zeigt einen (einfarbig wiedergegebenen) Screenshot des Experimentablaufs im Internet.

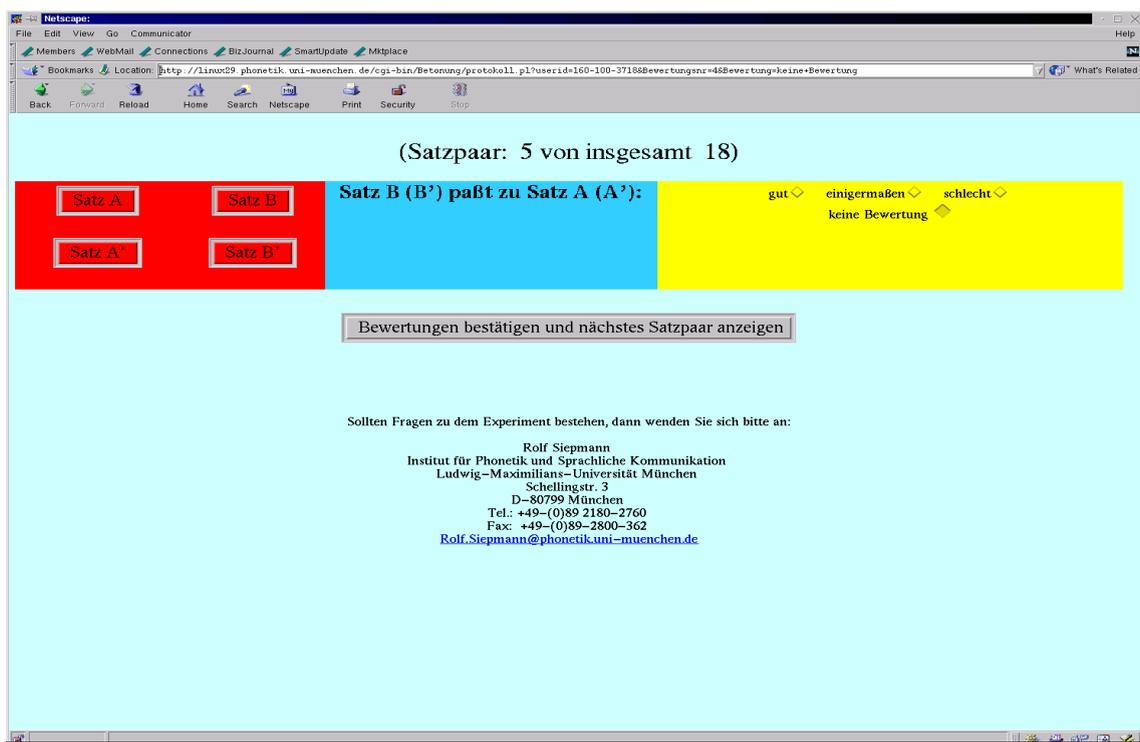


Abbildung 33

Die A'- und B'-Sätze der WWW-Seite sind mit 8 Bit/16 kHz digitalisiert, so dass Versionen der A- und B-Sätze mit kleinerer Dateigröße entstanden. Es wurden alle drei Satztypen der Dialoge 01, 02, 03, 05 und 06 verwendet (vgl. Anhang A, S. 105ff). Zusätzlich kamen drei Kontrolldialoge mit nicht zusammen passenden Satzpaaren zum Einsatz, so dass insgesamt 18 WWW-Dialoge entstanden.

Die Erhebung von Daten im Internet ist auf den Kreis der Internet-User eingeschränkt und insofern natürlich nicht repräsentativ. Dennoch ermöglicht das Internet relativ leicht, eine geographische Streuung bei den Probanden zu erreichen. Infolgedessen wurden direkte Aktionen der Rekrutierung von Sprechern durchgeführt, die viele sprachliche Regionen erfassen sollten. Die Aktionen bestanden in dem elektronischen Anschreiben von verschiedenen im Internet vertretenen Institutionen wie etwa Universitätsinstituten, Schulen und Heimatvereinen (Auswahl aufs Gradewohl). Zudem wurde das WWW-Experiment bei verschiedenen Suchmaschinen im Internet angemeldet.

7.3.2 Ergebnisse

An dem WWW-Perzeptionsexperiment nahmen insgesamt 74 Probanden teil, wobei die B-Sätze der vier Sprecher des Produktionsexperiments sukzessive zum Einsatz kamen. Der erste Sprecher wurde dabei von 19, der zweite von 16, der dritte von 17 und der vierte Sprecher von 22 Probanden des WWW-Experiments beurteilt.

7.3.2.1 Teilnehmer

Die Probanden des WWW-Experiments stammten aus verschiedenen Regionen, die nach der Karte in Abbildung 34 von Burger/Schiel (1995) regional eingeteilt wurden. Die Dialekteinstufung der Probanden basiert auf der Frage 'Bundesland/Ort, der am stärksten den Dialekt geprägt hat' (s. Fragebogen in Kap. 7.3.1, S. 73). Tabelle 3 gibt das Ergebnis der Einteilung wieder, wobei drei Probanden dort nicht auftauchen. Zwei Probanden waren aus Russland und einer hatte den Fragebogen nicht ausgefüllt. Sie wurden dennoch bei der Auswertung berücksichtigt, weil ihre Bewertungen in etwa mit den Bewertungen der anderen Teilnehmer übereinstimmen. Es konnte somit angenommen werden, dass die drei Probanden über gute Kenntnisse des Deutschen verfügen.

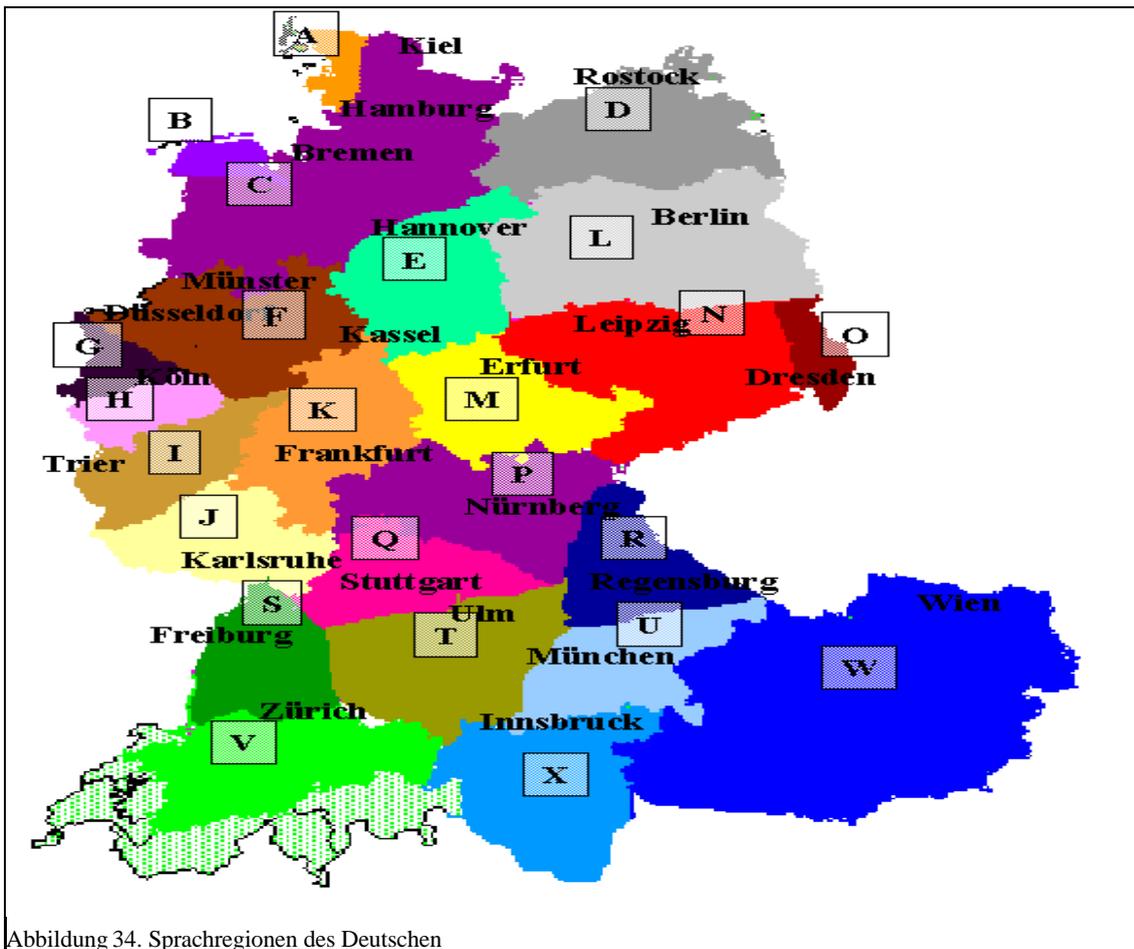


Abbildung 34. Sprachregionen des Deutschen

Kürzel	Dialekt	Anzahl der Teilnehmer
A	nordfriesisch	0
B	ostfriesisch	2
C	nordniedersächsisch	3
D	pommersch	0
E	ostfälisch	6
F	westfälisch	8
G	niederrheinisch	8
H	mittelfränkisch	2
I	moselfränkisch	1
J	pfälzerisch	3
K	hessisch	6
L	brandenburgisch	3
M	thüringisch	2
N	obersächsisch	3
O	sorbisch	0
P	ostfränkisch	2
Q	südfränkisch	8
R	nordbairisch	2
S	niederallemanisch	0
T	schwäbisch	0
U	mittelbairisch	10
V	schweizerisch	1
W	österreichisch	1
X	tirolerisch	0

Tabelle 3. Dialektale Zurordnung der Probanden des WWW-Experiments

7.3.2.2 Abgegebene Urteile

In dem WWW-Experiment erhielten von den insgesamt 1332 abgegebenen Beurteilungen der 74 Probanden insgesamt 11 Dialoge die in dem Experiment mögliche Beurteilung „keine Bewertung“ und 10 Dialoge wurden überhaupt nicht beurteilt. Die Verteilung der verbleibenden Beurteilungen „gut“, „einigermaßen“ und „schlecht“ der 18 Dialoge ist in Abbildung 35 dargestellt, wobei auf der x-Achse mit den Kürzeln „g (=gut)“ und „s (=schlecht)“ die erwarteten Beurteilungen und auf der y-Achse die tatsächlich abgegebenen Beurteilungen angezeigt sind. Wie bereits erwähnt, basiert die erwartete schlechte Bewertung auf dem Fall, dass ein Dialog zu Kontrollzwecken künstlich zusammengestellt wurde, der B-Satz in dem Produktionsexperiment ursprünglich also eine Replik auf einen anderen A-Satz war.

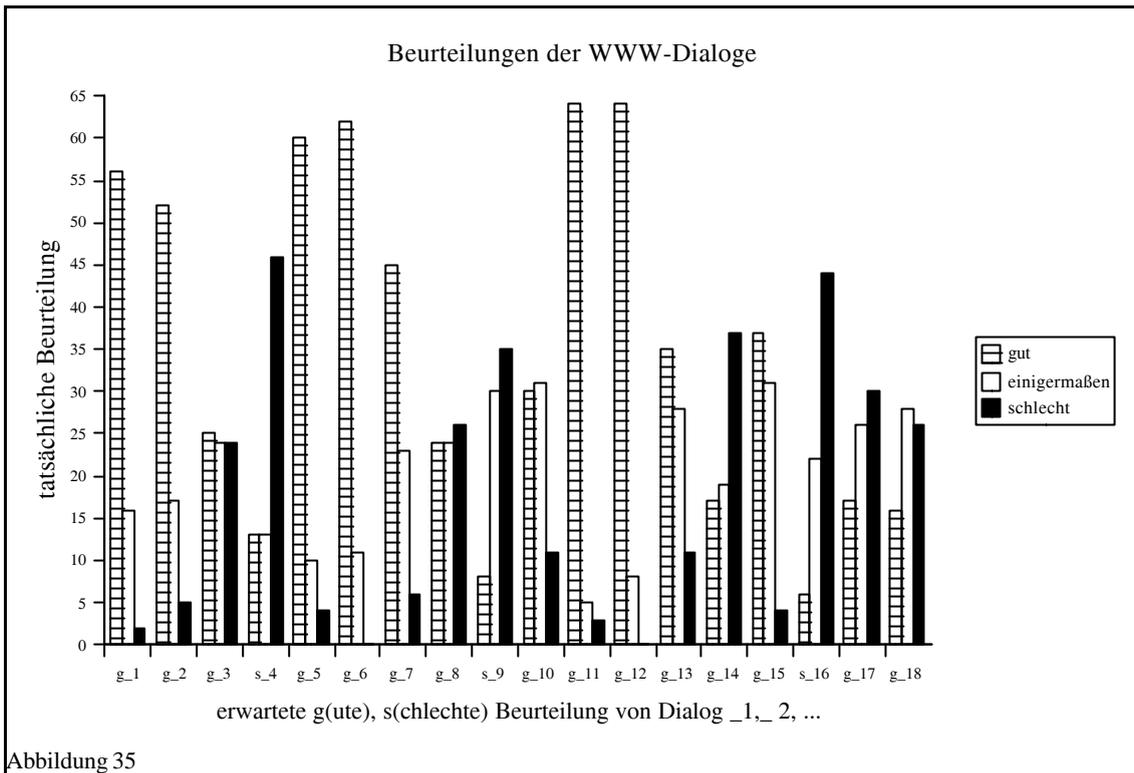


Abbildung 35

Insgesamt wurden von den Kontrolldialogen 12% gut, 30% einigermaßen und 58% schlecht bewertet. Die anderen passenden Dialoge wurden insgesamt zu 55% gut, zu 28% einigermaßen und zu 17% schlecht bewertet. Tabelle 4 gibt die Ergebnisse der passenden Dialoge des Perceptionsexperiments für die einzelnen Sprecher prozentual wieder. In Abhängigkeit von einer positiven oder negativen Interpretation des Urteils „einigermaßen“ ergibt sich dabei eine unterschiedliche Bewertung des Perceptionsexperiments; vgl. dazu die folgende Diskussion.

Beurteilung	Sprecher 1	Sprecher 2	Sprecher 3	Sprecher 4
gut	46%	61%	50%	62%
einigermaßen	39%	27%	25%	22%
schlecht	15%	11%	25%	16%
<i>gut & einigermaßen</i>	85%	88%	75%	84%
<i>einigermaßen & schlecht</i>	54%	38%	50%	38%

Tabelle 4. Prozentuale Anteile der Sprecherurteile

7.3.3 Diskussion des Perzeptionsexperiments

Batliner (1994) kommt in einer Untersuchung von Fokusstrukturen zu dem Ergebnis, dass eine Fokusmarkierung nicht unbedingt prosodisch erfolgt, sondern auch aus dem sprachlichen oder situativen Kontext abgeleitet werden kann. Der Einsatz der Kontrolldialoge in dem Perzeptionsexperiment diente gerade der Überprüfung, ob eine derartige Ableitung der entsprechenden kontextuellen Kontraste von den Versuchspersonen vorgenommen wurde.

Der Anteil von 58% Beurteilungen „schlecht“ schließt diese Interpretation zunächst eher schwach aus. Größere Sicherheit ergibt sich, wenn die Beurteilung „einigermaßen“ als negatives und nicht als positives Urteil gewertet wird, so dass die Kontrolldialoge dann insgesamt zu 88% negativ beurteilt wurden. Bei einem derartigen Prozentsatz ist nicht von einer prosodischen Markierung der Kontraste und deren Dekodierung von den Versuchspersonen auszugehen.

Die Einteilung der Bewertung „einigermaßen“ in die Kategorie der negativen Beurteilungen bedingt, dass die anderen, nicht der Kontrolle der Versuchspersonen dienenden Dialoge weitestgehend schlecht beurteilt wurden; vgl. Tabelle 4, S. 78. Aus diesem Ergebnis folgt, dass bestehende komplexe Kontraste im Diskurs, wie sie die vorliegende Arbeit untersucht, in der Regel nicht mit intonatorischen Mitteln ausgedrückt werden. Ein Ergebnis, das bereits bei der Diskussion des Produktionsexperiments ermittelt wurde (vgl. Kap. 7.2.3, S. 72).

Ein anderes Bild ergibt sich dagegen, wenn die Beurteilung „einigermaßen“ positiv gewertet wird, so dass zusammen mit der Beurteilung „gut“ vorwiegend positiv beurteilte B-Sätze vorliegen. Die konstruierten Kontraste der Dialoge wären damit in den B-Sätzen prosodisch markiert. In diesem Fall könnten die in Kapitel 7.2.2, S. 57ff, ermittelten schwach ausgeprägten akustischen Unterschiede der akzentuierten und der neutralen Varianten auch auf die vorgenommene, nunmehr ungeeignete Unterscheidung nach Satztypen zurückgeführt werden. Die multivariaten Analysen in Kapitel 9, S. 95ff, gehen dieser Frage weiter nach, in dem in gesonderten Diskriminanzanalysen zum einen nach Satztypen und zum anderen nach Sprechern unterschieden wird.

7.3.4 INTSINT-Synthese

Die Äußerungen des Produktionsexperiments wurden nach dem INTSINT-Verfahren modelliert (vgl. Kap. 6.5, S. 47f). Mit den dabei ermittelten F₀-Zielpunkten bei den geäußerten B-Sätzen, den Korrelaten der INTSINT-Annotation, wurde mit Praat eine PSOLA-Synthese durchgeführt⁵³. Die synthetischen Stimuli wurden anschließend, wie in den vorherigen Kapiteln beschrieben, als B-Sätze in dem WWW-Experiment verwendet. Die Ergebnisse dieses Experiments werden im folgenden kurz vorgestellt.

An dem WWW-Experiment mit den synthetischen B-Sätzen nahmen 48 Probanden teil, deren regionale Verteilung in Tabelle 5 aufgeführt ist. Von den 48 Probanden beurteilten 16 die synthetisierten Varianten des ersten Sprechers, und 32 Probanden beurteilten die entsprechenden Varianten des vierten Sprechers. Insgesamt wurden somit 864 Beurteilungen getroffen, von denen 20 nicht bewertet wurden oder bei denen „keine Bewertung“ angeklickt wurde. Die ver-

⁵³) Selbstverständlich hätten auch mit den Parametern der anderen phonetischen Intonationsmodelle Syntheseexperimente durchgeführt werden können. Aus Aufwandsgründen war dies jedoch nicht möglich.

bleibenden 844 Beurteilungen beinhalten dabei 134 Beurteilungen von Kontrolldialogen, von denen 11% als gut, 32% als einigermaßen und 55% schlecht beurteilt wurden. Von den 710 Beurteilungen der passenden Dialoge wurden 49% als gut, 29% als einigermaßen und 29% als schlecht beurteilt. Die Interpretation der abgegebenen Beurteilungen hängt, wie oben bereits diskutiert, von der Einteilung der Kategorie „einigermaßen“ ab, so dass an dieser Stelle ebenfalls auf die Ergebnisse der multivariaten Analysen in Kapitel 9, S. 95ff, verwiesen werden soll.

Kürzel	Dialekt	Anzahl der Teilnehmer
A	nordfriesisch	1
B	ostfriesisch	0
C	nordniedersächsisch	3
D	pommersch	0
E	ostfälisch	6
F	westfälisch	8
G	niederrheinisch	1
H	mittelfränkisch	1
I	moselfränkisch	1
J	pfälzerisch	1
K	hessisch	1
L	brandenburgisch	2
N	thüringisch	1
M	obersächsisch	2
O	sorbisch	2
P	ostfränkisch	3
Q	südfränkisch	6
R	nordbairisch	2
S	niederallemanisch	0
T	schwäbisch	0
U	mittelbairisch	2
V	schweizerisch	1
W	österreichisch	4
X	tirolerisch	0

Tabelle 5. Dialektale Zuordnung der Teilnehmer des WWW-Experiments mit synthetischen Stimuli

8. MODELLIERUNG DER SATZAKZENTREALISIERUNGEN.

Die F0-Verläufe der experimentell gewonnenen Äußerungen des Korpus MONA werden im folgenden nach den in Kapitel 6 bereits besprochenen phonetischen Intonationsmodellen modelliert. Die Modellierung beschränkt sich auf die Wortakzentsilben der beiden Akzentwörter in dem in Kapitel 7.2.2, S. 57f, erläuterten Sinne. Andere als die mit diesen Wortakzentsilben korrespondierende F0-Abschnitte werden nicht im Rahmen der einzelnen Intonationsmodelle analysiert. Die Modelle bieten bei den Akzentsilben durchgehend entsprechende Parameter an, so dass ein - statistischer - Vergleich der Modelle an dieser Stelle optimal möglich ist. Mit dieser Restriktion werden Fragen nach möglichen nicht lokalen Akzentmanifestationen a priori ausgeklammert.

8.1 KIM-Modellierung

8.1.1 Vorgehen

Der KIM-Modellierung der F0-Verläufe der experimentellen Äußerungen liegen im wesentlichen die in Peters (1999) beschriebenen Kriterien zugrunde. Peters (1999) bietet eine aktuelle Untersuchung, die KIM auf spontansprachliches Material anwendet. Die Analyse der KIM-Parameter erfolgt sowohl auditiv als auch nach dem – globalen - visuellen Eindruck des F0-Verlaufs einer Äußerung. Bei der Modellierung stehen im vorliegenden Zusammenhang die Satzakzente im Mittelpunkt des Interesses, so dass nicht alle von KIM zur Verfügung gestellten Analysemöglichkeiten verwendet werden. Von besonderem Interesse sind demnach die Gipfel- und Talkonturen, die nach dem KIM-Modell akustische Korrelate von Satzakzenten bilden. Bei der Modellierung der F0-Verläufe wurden die in Tabelle 6 beschriebenen Merkmale erfasst, wobei die Aussagen über die akustischen Eigenschaften der F0-Verläufe nicht genau quantifiziert sind.

Label	Beschreibung
früher Gipfel	wenn Iktussilbe wortinitial, dann F0(max) am Anfang der Silbe und insgesamt Fallen von F0 durch die Silbe (bzw. deren Kern)
	wenn Iktussilbe nicht wortinitial, dann Betrachtung des Fensters von Iktussilbe und deren vorhergehende Silbe; F0(max) in der vorhergehenden Silbe und Fallen von F0 durch die Iktussilbe (bzw. deren Kern) [ein Fenster von zwei Silben ist für das untersuchte Material der vorliegenden Arbeit ausreichend, d.h. die Iktussilbe steht maximal an zweiter Stelle]
mittlerer Gipfel	F0(max) der Iktussilbe in der Mitte des Silbenkerns
später Gipfel	wenn Iktussilbe wortfinal, dann F0(max) am Ende der Iktussilbe und Ansteigen von F0 durch die Iktussilbe
	wenn Iktussilbe nicht wortfinal, dann Betrachtung des Fensters von Iktussilbe und deren nachfolgende Silbe; F0(max) in der nachfolgenden Silbe, wobei F0 durch die Iktussilbe steigt
frühes Tal	F0(min) kurz vor der Iktussilbe, so dass der Anstieg durch die Silbe (und ggf. weiter) verläuft; vor F0(min) muss also ein Abfall stattfinden [in frühes Tal kann somit nicht wortinitial auftreten]
spätes Tal	bei einem späten Tal liegt F0(min) im Silbenkern [dieser Fall ist mitunter schwer von einem späten Gipfel zu unterscheiden]
'upstep'	bei Gipfeln, wenn der 2. Gipfel gleich hoch oder höher als der 1. Gipfel ('downstep' per default)
	bei Tälern, wenn F0(min) des zweiten Tals nicht tiefer als F0(min) des ersten Tals
ebene Konturen	ebene Konturen bei satzakzentuierten Silben, deren F0-Verlauf sich nicht von dem Niveau der umgebenden Silben abhebt [dieser Fall trat in dem untersuchten Korpus der vorliegenden Arbeit nicht auf]

Tabelle 6. Kriterien der KIM-Modellierung

8.1.2 Ergebnis

Die 288 akzentuierten Äußerungen des Korpus wurden an den Positionen der realisierten kontrastiven Satzakkente mit KIM-Labels annotiert. Abbildung 36 zeigt eine Äußerung des vierten Sprechers mit einem späten und einem frühen Gipfel.

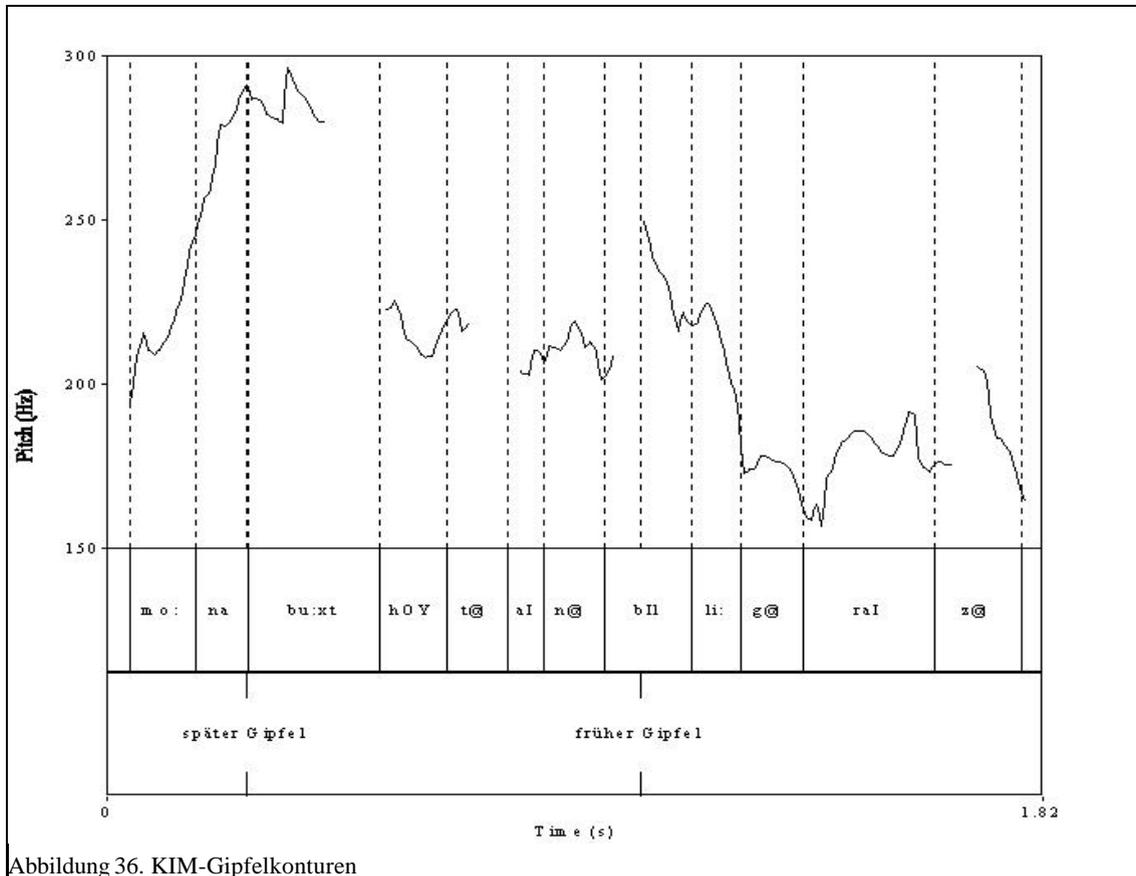


Abbildung 36. KIM-Gipfelkonturen

Tabelle 7 listet die Ergebnisse der Annotation auf (Anzahl der Label). Die Ergebnisse zeigen, dass Gipfelkonturen am häufigsten realisiert wurden. Der späte Gipfel ist bei dem ersten Akzent und der frühe Gipfel bei dem zweiten Akzent am häufigsten vertreten. Die anderen Gipfelkonturen kommen in etwa in der gleichen Häufigkeit vor. Das frühe Tal ist mit Abstand am seltensten realisiert worden.

1. Akzentposition		2. Akzentposition	
früher Gipfel	42	früher Gipfel	152
mittlerer Gipfel	38	mittlerer Gipfel	51
später Gipfel	149	später Gipfel	61
frühes Tal	20	frühes Tal	4
spätes Tal	39	spätes Tal	20

Tabelle 7. KIM-Label

Der Vergleich mit den neutralen Äußerungen erfolgt bei der KIM-Modellierung, indem bei einem Gipfel in der akzentuierten Äußerung das F0-Maximum der Silbe der korrespondierenden Akzentposition der neutralen Äußerungen gemessen wird⁵⁴. Analog wird bei einem Tal das F0-Minimum ermittelt. Die neutralen Äußerungen weisen dabei an den Akzentpositionen natürlich keine kontrastiven Satzakkzentrealisierungen auf. Aus der Abhängigkeit der Bestimmung des F0-Wertes der neutralen Variante von der entsprechenden KIM-Klassifikation der akzentuierten Variante ergeben sich verbundene Stichproben, die im folgenden zu der Anwendung von entsprechenden statistischen Tests für derartige Stichproben führen. Diese Tests werden auch bei den anderen in der vorliegenden Arbeit untersuchten Intonationsmodellen angewendet, obwohl diese spezielle Abhängigkeit bei den anderen Modellen nicht besteht. Die Anwendung der Tests für verbundene Stichproben ist allerdings auch in diesen Fällen sinnvoll, weil jeweils ein bestimmter Satz von einem Sprecher innerhalb einer experimentellen Gesamtumgebung nur eben auf zwei unterschiedliche – intonatorische – Arten gesprochen wurde.

Es werden im folgenden ein t-Test und ein Wilcoxon-Test für verbundene Stichproben durchgeführt (vgl. Clauß/Finze/Partsch (1995)). Der parametrische t-Test setzt genau genommen Normalverteilung voraus, die mit dem im folgenden durchgeführten Kolmogorov-Smirnov-Anpassungstest nur in einem Fall nicht abgelehnt werden konnte. Der t-Test wurde dennoch verwendet, weil er zum einen in erster Linie gegenüber Extremwerten empfindlich ist, die in den untersuchten Daten jedoch nicht auftreten. Zum anderen wurde er aber insbesondere deswegen durchgeführt, weil er der bekannte, übliche Test für Mittelwertvergleiche ist und zudem asymptotisch auch für andere Verteilungen gilt. Wie sich im folgenden zeigt, weichen die Ergebnisse des t-Tests kaum von den Ergebnissen des nicht-parametrischen Wilcoxon-Tests ab, der keine Normalverteilung voraussetzt (und unempfindlich gegenüber Extremwerten ist).

Tabelle 8 gibt die statistische Auswertung der KIM-Modellierung wieder. Die Tests zeigen, dass die neutralen und akzentuierten Varianten sowohl bei den mittleren Zeitwerten als auch bei den mittleren F0-Werten hochsignifikant voneinander abweichen. Die Zeitwerte sind auf die entsprechende Silbe relativiert, so dass ein Zeitwert von 0 den Anfang und ein Zeitwert von 1 das Ende einer Silbe kennzeichnet. Die Mittelwerte der F0-Werte basieren auf den originalen Hertz-Werten.

Variable	μ , n	μ , a	e (n-a)	p (t-Test)	p (Wilcoxon – Test)
t	0.609	0.462	0.23	0.000	0.000
f0	177.203	197.916	46.734	0.000	0.000

Tabelle 8. n(eutral), a(kzentuiert), μ : Mittelwert, e: Standardabweichung der Differenzen n-a, p: Signifikanz

⁵⁴) s.a. Kap. 7.2.2, S. 57f.

8.2 IPO-Modellierung

8.2.1 Vorgehen

Die F0-Verläufe der realisierten Satzakkente wurden zunächst nach dem Verfahren von d'Allessandro/Mertens in tonale Segmente eingeteilt, vgl. Kap. 6.2.2, S. 40f. Anschließend erfolgte die Abbildung der berechneten tonalen Segmente auf die Standardkonturen des IPO-Modells, vgl. Tabelle 1, S. 38. Die Einteilung in tonale Segmente ist mit der Anwendung der in Kapitel 6.2.2, S. 40f, genannten Gewichtungsfunktion verbunden, deren genaue Umsetzung von d'Allessandro/ Mertens nicht angegeben ist. Aus diesem Grund wurde in der vorliegenden Arbeit die in Anhang B, S. 108, angegebene Funktion „wtam-fkt“ mit Matlab (vgl. Anhang C, S. 110) implementiert, die die Gewichtungsfunktion im Hinblick auf die IPO-Modellierung umsetzt.

Die Anwendung von „wtam-fkt“ ist mit einer Glättung (und auch mit einer leichten Verzögerung) des F0-Verlaufs verbunden. Die Funktion gewichtet dabei das letzte Drittel des F0-Verlaufs einer Silbe stärker als den davor liegenden F0-Verlauf. Wenn die Silbe ausreichend lang ist, dann arbeitet „wtam-fkt“ in dem von d'Allessandro/Mertens angesetzten Zeitfenster von 140 ms, dessen Größe sich nach der durchschnittlichen Dauer französischer Silben von etwa 200 ms richtet⁵⁵. Dieser Wert entspricht in etwa auch der durchschnittlichen Silbendauer der Äußerungen des in dieser Arbeit verwendeten Korpus (vgl. Tabelle 9).

Akzentposition	Sprecher 1	Sprecher 2	Sprecher 3	Sprecher 4
I	248.65 ms	222.92 ms	233.24 ms	232.53 ms
II	265.00 ms	223.88 ms	233.01 ms	237.78 ms

Tabelle 9. Durchschnittliche Silbenlängen

Die mit dem Verfahren von d'Allessandro/Mertens ermittelten tonalen Segmente werden mit einem in der vorliegenden Arbeit eigens entwickelten Algorithmus auf die Standardkonturen des IPO-Modells abgebildet. Die Abbildung wird damit automatisiert und objektiviert. Die Konturen sind in Tabelle 1, S. 38, dargestellt. Wie bereits erwähnt, handelt es sich bei den tonalen Segmenten um ebene, fallende und steigende lineare F0-Verläufe, so dass also auch ebene Verläufe auf die nicht-ebenen Standardkonturen abgebildet werden können. Ebene Verläufe sind allerdings nicht Bestandteil des IPO-Modells, so dass es vor dem Hintergrund der geplanten Untersuchung des IPO-Modells nicht sinnvoll ist, die ebenen Verläufe etwa durch Absenkung der Grenzwerte für die Wahrnehmung von Tonhöhenbewegungen (vgl. Kap. 6.2.2, S. 40f) zu eliminieren.

Der Algorithmus zur Abbildung der tonalen Segmente auf die Standardkonturen ermittelt sukzessive die Zuordnung des Segments zu zwei unterschiedlichen Standardkonturen. Insgesamt wird ein Segment mit allen Konturen aus Tabelle 1, S. 38, verglichen. Die Ermittlung der Zuordnung eines Segments ist mit einer Bewertung verbunden, zu welchen der beiden gerade betrachteten Konturen das Segment besser passt. Im nächsten Schritt wird dann die besser passende Kontur mit einer weiteren Kontur im Hinblick auf das betrachtete Segment verglichen usw.

⁵⁵) Mertens/d'Allessandro (1995) arbeiten alternativ auch mit einem Zeitfenster von 50 ms.

Die Bewertung erfolgt nach den drei die Standardkonturen definierenden Kriterien der Position bzgl. Vokalanfang bzw. Ende des stimmhaften Teil, der Dauer der Kontur und der Umfang in Halbtönen. Eine Bewertung wird durch Kennzahlen ausgedrückt, so dass beispielsweise eine Segmentdauer von 130 ms im Hinblick auf eine Konturdauer A von 120 ms und auf eine Konturdauer B von 180 ms wie in (a) und (b) angegeben bewertet wird; vgl. Abbildung 37.

$$(a) \text{ Bew}_{\text{Kontur A}}(\text{Dauer}) = 1 - \text{abs}(130-120) / (\text{abs}(130-120) + \text{abs}(130-180)) = 0.83$$

$$(b) \text{ Bew}_{\text{Kontur B}}(\text{Dauer}) = 1 - \text{abs}(130-180) / (\text{abs}(130-120) + \text{abs}(130-180)) = 0.17$$

Die Bewertung drückt die 'relative Nähe' zu den beiden Vergleichswerten aus, d.h. die Dauer von 130 ms ist in einer 'besseren Nähe' von 83% bei der Dauer von 120 ms (Kontur A) und in einer 'schlechteren Nähe' von 17% bei der Dauer von 180 ms (Kontur B). Analog werden Umfang und Segmentpositionierung durch Kennzahlen bewertet, so dass die Summation der Kennzahlen für Position, Dauer und Umfang einer Standardkontur die Zuordnung eines tonalen Segments zu einer Kontur ermöglicht.

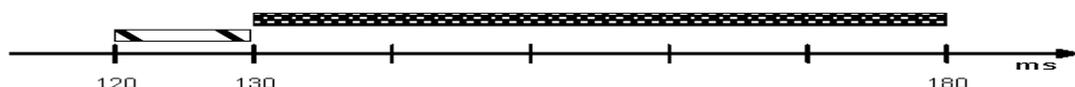


Abbildung 37. Absoluter Abstand einer Konturdauer von 130 ms zu Kontur A (120 ms) bzw. zu Kontur B (180 ms)

8.2.2 Ergebnis

Der Algorithmus der automatischen Abbildung ermittelte an den beiden Akzentpositionen in dem Korpus nur 6 der 12 möglichen Standardkonturen:

Kontur drei. Beginn 30 ms vor Vokalanfang (VA), steigend, 60 ms Dauer, 2.5 Halbtöne (HT)

Kontur vier. Variabler Beginn⁵⁶, steigend, 60 ms Dauer, 7.5 HT

Kontur fünf. Variabler Beginn (vgl. Fußnote 56), fallend, 180 ms Dauer, 7.5 HT

Kontur acht. Beginn 150 ms nach VA, 180 ms Dauer, 7.5 HT

Kontur zehn. 300 ms vor Ende des stimmhaften Bereichs (EST), steigend, 300 ms Dauer, 12.5 HT

Kontur elf (=10a). 120 ms vor EST, steigend, 120 ms Dauer, 5 HT

Von den ermittelten Konturen treten allerdings nach dem IPO-Modell die Konturen 3, 4, 5 und 8 bei Akzenten auf, so dass das Ergebnis auf den ersten Blick positiv zu bewerten ist. In Abbildung 38 sind die Häufigkeiten der ermittelten Standardkonturen unterteilt nach akzentuierten und neutralen Varianten für beide Akzentpositionen zusammengefasst wiedergegeben. Kontur 4, 5 und 8 sind am häufigsten vertreten, wobei Kontur 8 nur an zweiter Position auftritt. Die beiden Konturen sind relativ gleichmäßig über die beiden Varianten verteilt.

Insgesamt ist bis auf Kontur 3 kein deutlicher Unterschied zwischen den akzentuierten und den neutralen Varianten festzustellen. Daraus läßt sich schließen, dass die Standardkonturen

⁵⁶) Es wurde der Wert des Beginns des tonalen Segments übergeben.

nicht distinktiv verwendet werden. An dieser Stelle sei erneut darauf hingewiesen, dass das IPO-Modell lediglich die perzeptiv relevanten F0-Bewegungen experimentell untersucht. Die epiphänomenologische Zuordnung zu funktionalen Kategorien basiert auf Annahmen über die sprachliche Form der untersuchten Äußerungen.

Kontur	acht	drei	elf	fünf	vier	zehn
Typ	fallend	steigend	steigend	fallend	steigend	steigend
akzentuierte Variante	185	9	5	89	280	8
neutrale Variante	177	25	5	83	282	4

Abbildung 38. Anzahl der Standardkonturen an erster und zweiter Akzentposition

Abbildung 39 enthält die Mittelwerte der Summanden der von dem Abbildungsalgorithmus berechneten Bewertungskennzahlen für den relativen Beginn, die Dauer und den Umfang eines Segments im Hinblick auf die 'relative Nähe' zu einer Standardkontur. Der Maximalwert eines Summanden beträgt 6. Dabei ist eine Bewertung von beispielsweise $Bew_B(\text{eginn}) = 3$, $Bew_U(\text{mfang}) = 3$ und $Bew_D(\text{auer}) = 3$ so zu interpretieren, dass alle drei Kriterien gleich stark in die Bestimmung der 'relativen Nähe' eingegangen sind. Ein entsprechendes Wertetripel von etwa (1,3,3) besagt dagegen, dass der Beginn im Vergleich zu dem Umfang und der Dauer eine untergeordnete Rolle bei der Bewertung der Relation 'tonales Segment – Standardkontur' spielt.

Kontur		acht	drei	elf	fünf	vier	zehn
Variante	Daten	fallend	steigend	steigend	fallend	steigend	steigend
akzentuiert	Bew_Beginn	2,73	2,87	3,43	4	3,74	3,12
	Bew_Umfang	2,05	3,92	3,09	2,11	3,05	3,18
	Bew_Dauer	1,99	4,06	3,14	2	3,03	3,13
neutral	Bew_Beginn	2,71	2,87	3,48	4	3,63	3,25
	Bew_Umfang	2,03	4,1	2,93	2,1	3,05	3,18
	Bew_Dauer	1,99	3,98	2,86	1,97	3,04	3,17

Abbildung 39. Mittelwerte der Summanden der Bewertungskennzahlen

Wie sich in Abbildung 39 zeigt, ist bei den Standardkonturen 3, 5 und 8 im Gegensatz zu den anderen drei Konturen eine eher heterogene Verteilung der drei Bewertungszahlen zu beobachten. Aus einer derartigen heterogenen Verteilung lässt sich auf eine heterogene Verteilung der entsprechenden Merkmale der zugrunde liegenden tonalen Segmente in dem Sinne schließen, dass sie sich nicht eindeutig auf eine Standardkontur abbilden lassen. Bei homogenen Bewertungszahlen weisen die entsprechenden Merkmale dagegen eine größere Homogenität auf, so dass eine eindeutigere Abbildung auf eine Standardkontur möglich ist. Im Vergleich der akzentuierten mit den neutralen Varianten ist wiederum eine hochsignifikant unterschiedliche Ausprägung bei den Mittelwerten festzustellen, vgl. Tabelle 10.

Variable	μ , n	μ , a	e (n-a)	p (t-Test)	p (Wilcoxon – Test)
Dauer (ms)	162.621	183.645	53.85	0.000	0.000
Halbtonumfang	2.825	3.311	3.110	0.000	0.001

Tabelle 10. n(eutral), a(kzentuiert), μ : Mittelwert, e: Standardabweichung der Differenzen n-a, p: Signifikanz

8.3 Fujisaki-Modellierung

8.3.1 Vorgehen

Die Fujisaki-Modellierung nach dem in Kapitel 6.3.1 beschriebenen Modell erfolgte manuell mit einem Software-Tool zur interaktiven Berechnung und Manipulation der Phrasen- und Akzentparameter des Modells (vgl. Anhang C, S. 110, 'Displaying Tool'⁵⁷); der Screenshot des Tools in Abbildung 40 illustriert das Vorgehen, bei der stets eine Äußerung modelliert wurde.

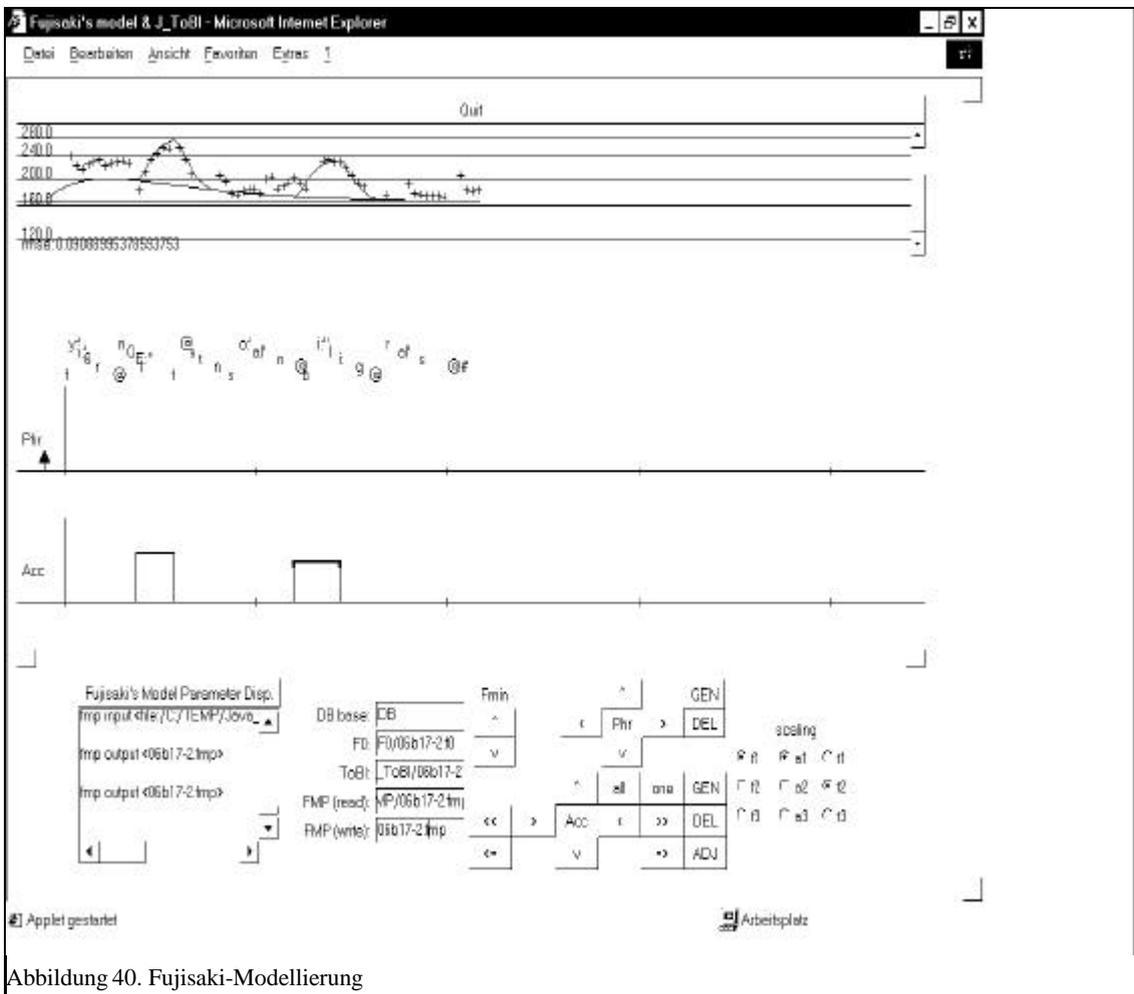


Abbildung 40. Fujisaki-Modellierung

⁵⁷) Der Java-Quellcode des Tools ist im Internet erhältlich und muss an die aktuelle Systemumgebung angepasst werden, um mit eigenen Ein- und Ausgabedateien arbeiten zu können.

Für die Modellierung der Fujisaki-Parameter wurden Dateien mit Default-Parametern für jede Äußerung erzeugt, die u.a. einen konstanten Wert $w = \ln(f_0(\min))$ für jeden Sprecher und die Zeiten der Akzentkommandos für jede Akzentsilbe enthielten. Der Wert w wurde Mixdorff (1998) folgend aus den Mimina von allen Äußerungen eines Sprechers berechnet. Er betrug für den ersten Sprecher 4.32, für den zweiten Sprecher 5.02, für den dritten Sprecher 4.26 und für den letzten Sprecher 5.18 Halbtöne. Während der Modellierung waren die SAMPA-Zeichen der Lautsegmente zu erkennen, die eine Kennzeichnung der beiden Akzentsilbenkerne beinhalteten. Die SAMPA-Zeichen der Lautsegmente waren mit dem originalen bzw. modellierten F0-Verlauf synchronisiert.

Das Tool zeigte bei jeder Veränderung der Parameter einen aktuell berechneten (modellierten) F0-Verlauf und den RMSE zu dem Originalverlauf an. Als Richtwert der Modellierung wurde dabei ein Wert von Möhler (1998: 99) von $rmse = 18.4$ Hz zugrundegelegt. Möhler hat diesen Wert bei der Evaluierung der in Kapitel 4.3 beschriebenen Abbildungsregeln des Tonsequenzmodells verwendet. Bei der Modellierung der Phrasen wurde nach Möglichkeit versucht, mit einem Phrasenkommando auszukommen. Die Akzente wurden modelliert, indem primär die Akzentamplitude und erst sekundär die Parameter der Zeitachse eines Akzentkommandos verändert wurden. Die Akzentkommandos waren mindestens 200 ms lang. Per default waren die Anfangs- und Endzeiten eines Akzentkommandos mit dem Anfang und dem Ende einer Akzentsilbe synchronisiert. Eine Justierung der Anfangs- und Endzeiten wurde vorgenommen, wenn ein Akzentkommando nicht mit den lokalen F0-Veränderungen einer Akzentmanifestation übereinstimmte. Wie in dieser Arbeit üblich, wurden die F0-Verläufe der entsprechenden Silben der akzentuierten Varianten und der korrespondierenden Silben der neutralen Varianten mit Akzentkommandos modelliert.

8.3.2 Ergebnis

Tabelle 11 gibt die Ergebnisse des Vergleichs der Mittelwerte der Akzentdauern und – amplituden der neutralen und der akzentuierten Varianten wieder, wobei der Wilcoxon-Test in diesem Fall bei den Akzentamplituden kein signifikantes Ergebnis liefert.

Variable	μ , n	μ , a	e (n-a)	p (t-Test)	p (Wilcoxon – Test)
Dauer (sec)	0.289	0.313	0.131	0.000	0.000
Amplitude	0.389	0.475	0.194	0.009	0.415

Tabelle 11. n(neutral), a(akzentuiert), μ : Mittelwert, e: Standardabweichung der Differenzen n-a, p: Signifikanz

8.4 Tilt-Modellierung

8.4.1 Vorgehen

Das Tilt-Modell ist als Teil der 'Edinburgh Speech Tools' (vgl. Anhang C, S. 110) implementiert und benötigt nur die entsprechend aufbereiteten Eingabedateien. Neben den Dateien mit den symbolischen Informationen über die Zeitgrenzen der Satzakzente gehören dazu die Dateien mit den F0-Verläufen. Die F0-Verläufe werden im dem automatisierten Tilt-Modell vollständig interpoliert. Eine derartige Interpolation erscheint zumindest bei der Beschränkung auf Akzentsilben, wie sie die vorliegende Arbeit vornimmt, gerechtfertigt.

Dem implementierten Tilt-Modell können u.a. zwei Parameter für den Suchbereich des 'pattern matching' – Algorithmus in der F0-Umgebung der zu analysierenden intonatorischen Ereignisse übergeben werden. Der Parameter 'limit' bestimmt den zeitlichen Bereich vor und nach einem Ereignis, in welchem der Algorithmus die möglichen 'rises' und 'falls' in dem F0-Abschnitt sucht⁵⁸. Der Parameter 'range' teilt den Suchbereich innerhalb eines Ereignisses auf, wobei er von einer durchschnittlichen Dauer der eingegebenen Ereignisse ausgeht. Wie sich herausstellte, konnten bei dem vorliegenden Korpus mit einer Belegung der Variablen von limit gleich 0.1 und von range gleich 0.5 überzeugende Ergebnisse erzielt werden. Der erste Wert besagt, dass 0.1 ms vor und 0.1 ms nach einem Ereignis gesucht wird. Nach dem zweiten prozentualen Wert wird die Suche genau auf die erste und die zweite Hälfte eines Ereignisses aufgeteilt. Wie bereits in Kapitel 6.7, S. 50f, Punkt (d) erwähnt, können die kontinuierlichen Tilt-Parameter nach Taylor (2000: 15) in die in Tabelle 12 dargestellten Klassen eingeteilt werden.

Tilt-Parameter	Beschreibung des F0-Verlaufs
+1.0	Anstieg
+0.5	Anstieg plus partieller Fall
±0.0	Anstieg plus vollständiger Fall
-0.5	partieller Anstieg plus Fall
-1.0	Fall

Tabelle 12. Tilt-Parameter

⁵⁸) Die berechneten Werte für eine F0-Bewegung sind im einzelnen einsehbar. Nach Auskunft von Paul Taylor ist es an der Universität Edinburgh Praxis, unplausible Werte auf Null zu setzen. Somit wird z.B. ein Fall mit einem positiven Wert < 0.5 auf Null gesetzt; bei einem Wert ≥ 0.5 'mutiert' der Fall zu einem Anstieg. Die vorliegende Arbeit schließt sich dieser Praxis an.

8.4.2 Ergebnis

Tabelle 13 enthält die statistischen Werte der Tilt-Modellierung, wobei der Tilt-Parameter und der berechnete F0-Wert berücksichtigt wurden. Der F0-Wert ist das F0-Maximum am Anfang eines Falls bzw. das Maximum am Ende eines Anstiegs. In diesem Fall liefern der t-Test und der Wilcoxon-Test nur bei den F0-Werten ein positives Ergebnis.

Variable	μ , n	μ , a	e (n-a)	p (t-Test)	p (Wilcoxon – Test)
f0 (max)	166.620	182.427	48.729	0.000	0.000
tilt-Parameter	-0.233	-0.227	0.613	0.972	0.975

Tabelle 13. n(neutral), a(kzentuiert), μ : Mittelwert, e: Standardabweichung der Differenzen n-a, p: Signifikanz

8.5 INTSINT-Modellierung

8.5.1 Vorgehen

Die INTSINT-Modellierung besteht aus zwei Arbeitsschritten, aus der Bestimmung der Zielpunkte des Modells in einem F0-Verlauf und aus der Zuweisung der INTSINT-Label zu den F0-Zielpunkten. Der erste Schritt erfolgt manuell, indem zunächst automatisch berechnete Zielpunkte korrigiert werden. Die Korrektur basiert zum einen auf dem visuellen und zum anderen auf dem auditiven Eindruck von einer synthetischen Äußerung, deren F0-Verlauf dem modellierten F0-Verlauf entspricht.

Bei den Akzentsilben wurden nur Zielpunkte verzeichnet, wenn auch welche ermittelt werden konnten. Es gibt somit Akzentsilben, die kein INTSINT-Label, also keinen Ton im Sinne des INTSINT-Modells aufweisen. Gleichmaßen können bei einer Akzentsilbe aber auch mehrere verschiedene Label auftreten. Um die Zielpunkte möglichst objektiv zu bestimmen, wurden die synthetischen Äußerungen in einem Perzeptionsexperiment verwendet, welches die perzeptive und funktionale Relevanz der Zielpunkte bestätigte (vgl. Kap. 7.3.4, S. 79f). Abbildung 41 illustriert die Bestimmung der F0-Zielpunkte (Kreise), die manuell gelöscht, hinzugefügt oder verschoben werden können.

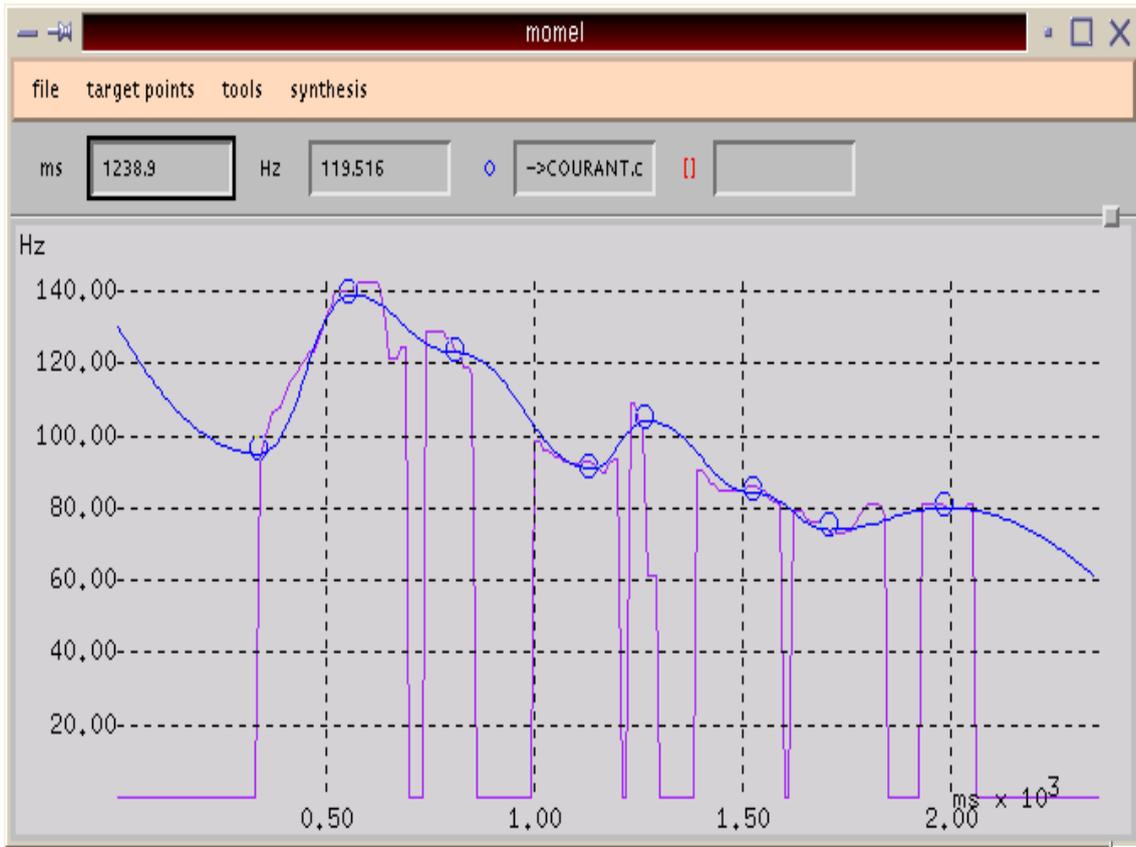


Abbildung 41. Bestimmung von F0-Zielpunkten

8.5.2 Ergebnis

Eine Auswertung der INTSINT – Modellierung ist nur bei Akzentpositionen sinnvoll, denen auch ein Ton zugewiesen wurde. Insgesamt erhielten 354 von insgesamt möglichen 1152 Akzentpositionen keinen Ton (Label „x“). Die anderen Akzentpositionen weisen mindestens einen und maximal zwei Töne auf (Label 1, 2, s.u.). Abbildung 42 gibt die 'tonlosen' Akzentpositionen unterteilt nach akzentuierter und neutraler Variante sowie nach Akzentposition wieder. Die neutralen Varianten weisen insgesamt häufiger keinen Ton (keinen F0-Zielpunkt) auf als die akzentuierten Varianten.

V a r i a n t e	P o s i t i o n	D a t e n	
a k z e n t u i e r t	A 1	A n z a h l - L a b e l x	6 7
	A 2	A n z a h l - L a b e l x	7 0
n e u t r a l	A 1	A n z a h l - L a b e l x	1 0 8
	A 2	A n z a h l - L a b e l x	1 0 9

Abbildung 42. A1: erste, A2: zweite Akzentposition

In Abbildung 43 sind die ermittelten absoluten Töne Top, Mid und Bottom sowie die relativen Töne Higher, Same, Lower, Upstepped und Downstepped an den beiden Akzentpositionen unterschieden nach akzentuierter und neutraler Variante sowie für Label 1 und 2 (s.o.) dargestellt. Von den absoluten Tönen sind an der ersten Akzentposition Top und Mid am häufigsten vertreten. Bei den relativen Tönen kommt Lower am häufigsten vor. Im Vergleich mit den relativen Tönen sind beide Gruppen von Tönen in etwa gleich stark vertreten.

Anzahl	Label	Ton	A1, Label 1	A1, Label 2	A2, Label 2	A2, Label 2
akzentuiert	B	Bottom	1	2	9	8
	D	Downstepped	18	1	39	11
	H	Higher	15	6	69	4
	L	Lower	54	11	11	1
	M	Mid	50	0	0	0
	S	Same	14	2	11	7
	T	Top	56	10	74	2
	U	Upstepped	13	6	5	0
neutral	B	Bottom	2	0	16	8
	D	Downstepped	19	1	50	3
	H	Higher	12	4	34	2
	L	Lower	38	3	33	3
	M	Mid	38	0	0	0
	S	Same	12	2	23	2
	T	Top	46	8	10	0
	U	Upstepped	13	0	13	1

Abbildung 43. Anzahl der ermittelten INTSINT-Töne

In Tabelle 14 sind die Ergebnisse der statistischen Analyse der INTSINT-Modellierung angegeben. Dabei wurden über die Zeitpunkte und über die F0-Werte der INTSINT-Label einer Akzentposition aus Gründen des Vergleichs von akzentuierter und neutraler Variante der Mittelwert gebildet, weil innerhalb einer Akzentposition zum Teil mehrere Label auftreten (siehe Abbildung 43). Die Tabelle zeigt, dass bei den Zeitwerten im Gegensatz zu den Frequenzwerten kein signifikanter Unterschied festzustellen ist.

Variable	μ , n	μ , a	e (n-a)	p (t-Test)	p (Wilcoxon – Test)
t	0.438	0.412	0.215	0.996	0.896
f0	175.037	186.970	57.547	0.016	0.013

Tabelle 14. n(neutral), a(kzentuiert), μ : Mittelwert, e: Standardabweichung der Differenzen n-a, p: Signifikanz

8.6 Diskussion der Modellierungen

Die in den Kapiteln 8.1 bis 8.5 vorgestellten Verfahren zur Modellierung der F0-Verläufe von kontrastiven Satzakkenten im Deutschen unterscheiden sich zunächst in dem Grad der Automatisierung. Die KIM- und die Fujisaki-Modellierung sind manuelle Verfahren, während die in der vorliegenden Arbeit vorgenommene Umsetzung des IPO-Modells sowie das Tilt-Modell vollständig automatisiert und damit reproduzierbar sind. Die beiden automatisierten Verfahren setzen dabei symbolische Information in Form von Silbengrenzen bzw. Akzentpositionen voraus.

Tabelle 15 fasst die Ergebnisse der statistischen Tests zusammen. KIM und das IPO-Modell kommen bei der Trennung der entsprechenden Mittelwerte der akzentuierten und der neutralen Varianten zu den besten Ergebnissen. Das Fujisaki-Modell erzielt hingegen nur bei der Zeitinformation ein signifikantes Ergebnis. Bei der Frequenzinformation liefert ausschließlich der Wilcoxon-Test ein nicht signifikantes Ergebnis, wobei allerdings auch der t-Test im Vergleich mit KIM und mit dem IPO-Modell schlechter ausfällt. Das Tilt- und das INTSINT-Modell sind komplementär im Hinblick auf die Signifikanzen der Zeit- und der Frequenzinformation. Das Tilt-Modell kommt bei der Frequenzinformation zu schlechten Ergebnissen, während bei der Zeitinformation das INTSINT-Modell negativ ausfällt, d.h. die statistischen Tests konnten in diesen Fällen keine Unterschiede bei den Mittelwerten im Hinblick auf die Unterscheidung 'akzentuiert – neutral' feststellen.

p	KIM		IPO		Fujisaki		Tilt		INTSINT	
	Z	F	Z	F	Z	F	Z	F	Z	F
t-Test	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.972	0.996	0.016
Wilcoxon	0.000	0.000	0.000	0.001	0.000	0.415	0.000	0.975	0.896	0.013

Tabelle 15. p: Signifikanz, Z: p der Zeitinformation, F: p der Frequenzinformation

Die Analyse von Mittelwerten liefert kein differenziertes Bild der untersuchten phonetischen Intonationsmodelle, so dass in dem folgenden Kapitel multivariate Analysen, d.h. Diskriminanzanalysen für die einzelnen Modelle durchgeführt. Anhand der Ergebnisse dieser Analysen lassen sich die Leistungen der einzelnen Modelle besser beurteilen.

9. MULTIVARIATE ANALYSE DER SATZAKZENTMODELLIERUNGEN

Die in Kapitel 8 gewonnenen intonatorischen Modellparameter der realisierten kontrastiven Satzakkente des MONA-Korpus werden im folgenden verschiedenen linearen Diskriminanzanalysen mit dem Ziel unterzogen, die Modellparameter mit der quantitativ deutlichsten Trennung der - modellierten - Satzakkente von den korrespondierenden Abschnitten der neutralen Äußerungsvarianten⁵⁹ zu ermitteln. Von diesen Parametern ist anzunehmen, dass sie die Grundfrequenzverläufe von kontrastiven Satzakkenten im Standarddeutschen im Vergleich zu den Parametern der anderen in dieser Arbeit untersuchten phonetischen Intonationsmodelle empirisch am besten beschreiben.

Backhaus et al. (2000: 146ff) folgend, bildet die Diskriminanzanalyse eine Diskriminanzfunktion für die einzelnen Modellparameter X_i eines phonetischen Intonationsmodells der Form:

$$Y = b_0 + b_1 X_1 + \dots + b_N X_N$$

Die Diskriminanzfunktion ermöglicht die Zuordnung der Parameterwerte der Modelle zu einer der beiden Gruppen, der entsprechenden Abschnitte der akzentuierten und der neutralen Äußerungsvarianten. Die Funktion ist eine Optimierungsfunktion, die bei ihrer Berechnung dem folgenden Kriterium genügen muss:

$$y = \max \left(\frac{(\text{Streuung zwischen den Gruppen})}{(\text{Streuung innerhalb der Gruppen})} \right)$$

Der Wert y wird auch als Eigenwert bezeichnet; aus dem Eigenwert wird die normierte kanonische Korrelation c berechnet, die die Güte der entsprechenden Gruppentrennung eines phonetischen Intonationsmodells ausdrückt:

$$c = \sqrt{\left(\frac{(\text{erklärte Streuung})}{\text{Gesamtstreuung}} \right)} = \sqrt{\left(\frac{y}{(1+y)} \right)}$$

Die kanonische Korrelation liegt zwischen 0 und 1, wobei ein möglichst hoher Wert c die optimale Trennung der beiden Gruppen - akzentuiert vs. neutral - kennzeichnet. Die Signifikanz p kennzeichnet schließlich die Irrtumswahrscheinlichkeit, mit der man sagen kann, dass sich die beiden Gruppen unterscheiden. Ihre Berechnung basiert auf einem χ^2 - Test, dessen Durchführung wiederum die Berechnung von Wilks' Lambda λ voraussetzt; ein möglichst kleiner λ - Wert drückt dabei eine gute Trennung der Gruppen aus:

$$\lambda = \frac{1}{(1+y)} = 1 - c^2$$

⁵⁹) Vgl. Kap. 7.2.2, S. 57ff.

Die Diskriminanzanalysen wurden mit SPSS durchgeführt (vgl. Kap. 10, S. 110). Als Eingabedaten die entsprechenden Modellparameter der 288 akzentuierten und der 288 neutralen Äußerungen des MONA - Korpus. Die prozentuale Angabe der kreuzvalidierten Fälle besagt, wie viele Fälle der Eingabe bei der Diskriminanzanalyse korrekt einer Gruppe zugeordnet werden konnten. Kreuzvalidierung ist ein Verfahren bei der Diskriminanzanalyse, bei dem die Gruppenzugehörigkeit eines Falles auf der Basis einer Diskriminanzfunktion untersucht wird, dessen Berechnung diesen Fall gerade nicht berücksichtigt.

Die Anzahl der gültigen Fälle N ist die Menge der Datensätze, die eine Diskriminanzanalyse in dem vorliegenden Zusammenhang verarbeitet hat. Im Idealfall handelt es sich um 576 Datensätze, also 288 von den akzentuierten und 288 von den neutralen Äußerungen. Eine Unterscheidung nach Satztypen (vgl. Kap. 9.1, S. 97f) teilt die 576 Datensätze in 192 Äußerungen mit normaler Wortstellung (Satztyp 1), in 192 Äußerungen mit dem syntaktischen Konstruktionstyp der Ellipse (Satztyp 2) und in 192 Äußerungen mit topikalisierte Wortstellung ein (Satztyp 3). Von den 192 Äußerungen eines Satztyps sind 96 akzentuierte und 96 neutrale Äußerungen. Eine Unterscheidung nach den vier Sprechern des MONA - Korpus liefert entsprechend eine Aufteilung des Korpus in Blöcke mit je 144 Äußerungen (vgl. Kap. 9.2, S. 99f).

Die Anzahl der gültigen Fälle beträgt nicht bei jedem phonetischen Intonationsmodell $N = 576$. Bei dem Modell Fujisaki und Tilt ist dies auf fehlerhafte Berechnungen der Software zurückzuführen (vgl. Kap. 8.3, S. 88ff, und Kap. 8.4, S. 90ff). Der relativ geringe Wert von N resultiert bei dem Modell INTSINT (vgl. Kap. 8.5, S. 91ff) hingegen aus berechneten Wendepunkten, die außerhalb der Silben der kontrastiven Satzakkente der akzentuierten Äußerungen und deren Entsprechungen bei den neutralen Äußerungen liegen (vgl. Kap. 7.2.2, S. 57ff.). Aus Gründen der optimalen Vergleichbarkeit gingen diese Wendepunkte nicht in die Diskriminanzanalyse ein. Damit wurden nur Wendepunkte berücksichtigt, die auch Korrelate von kontrastiven Satzakkenten bildeten.

Diese Einschränkung gilt nicht für die F0- Minima und -Maxima der Gipfel und Täler des Modells KIM (vgl. 8.1, S. 81ff). Im Gegensatz zu INTSINT, dessen Wendepunkte des modellierten Grundfrequenzverlaufs losgelöst von sprachlichen Einheiten wie etwa Silben oder Silbenkernen analysiert werden⁶⁰, baut KIM bei der Zuweisung von Gipfel- und Talkonturen auf derartigen sprachlichen Strukturen auf. Demnach ist etwa ein F0 - Maximum einer Gipfelkontur sicher eine akustische Manifestation (im Sinne des KIM - Modells) eines kontrastiven Satzakkents. Die Diskriminanzanalysen des vorliegenden Kapitels wurden darüber hinaus auch auf die akustischen Parameter der univariaten Analyse des MONA - Korpus angewendet (vgl. Kap. 7.2.2, S. 57ff.); diese Parameter bilden im folgenden das phonetische Intonationsmodell MONA.

⁶⁰) In diesem Sinne ist INTSINT gewissermaßen ein Verfahren der - reinen - Grundfrequenztranskription.

9.1 Unterscheidung nach Satztypen

Die linearen Diskriminanzanalysen (LDA) wurden zunächst für die einzelnen Satztypen des MONA - Korpus durchgeführt (s.a. Kap. 7.2.2, S. 57ff). Die Ergebnisse sind im einzelnen in Tabelle 16 wiedergegeben.

	LDA	KIM	IPO	TILT	Fujisaki	INTSINT	MONA
Satztyp 1	gültige Fälle N	192	192	184	168	80	192
	Korrelation c	0,13	0,51	0,39	0,29	0,3	0,5
	Wilks' Lambda	0,98	0,739	0,834	0,917	0,911	0,751
	Signifikanz	0,56	0,000	0,001	0,028	0,068	0,000
	Kreuzval. Fälle	53,1%	71,9%	54,8%	55,2%	47,9%	63,5%
Satztyp 2	gültige Fälle N	192	192	188	166	124	192
	Korrelation c	0,211	0,585	0,610	0,359	0,322	0,599
	Wilks' Lambda	0,956	0,658	0,628	0,871	0,896	0,642
	Signifikanz	0,073	0,000	0,000	0,001	0,004	0,000
	Kreuzval. Fälle	54,7%	76,0%	75,3%	57,3%	53,1%	68,8%
Satztyp 3	gültige Fälle N	192	192	189	173	86	192
	Korrelation c	0,188	0,449	0,491	0,367	0,190	0,526
	Wilks' Lambda	0,965	0,798	0,759	0,865	0,964	0,724
	Signifikanz	0,147	0,000	0,000	0,000	0,387	0,000
	Kreuzval. Fälle	56,8%	67,7%	67,4%	61,5%	49,5%	66,7%
Mittelwert	Kreuzval. Fälle	54,866	71,866	65,833	58,000	58,266	66,330

Tabelle 16

Tabelle 16 ist zusammenfassend zu entnehmen (vgl. Mittelwert der kreuzvalidierten Fälle), dass die korrekte Gruppenzuordnung - die kreuzvalidierten Fälle - insgesamt relativ gering ist; selbst die maximal richtige Unterscheidung der 'akzentuierten' und der 'neutralen' Fälle von 76% bei Satztyp 2 des IPO - Modell ist nicht als besonders gutes Ergebnis zu werten. KIM fällt in diesem Zusammenhang bei allen drei Satztypen besonders schlecht aus, dessen Wahrscheinlichkeiten von 53,1% bis 56,8% nur knapp über dem Raten liegen. Eine ähnliche Bewertung ist bei INTSINT vorzunehmen. Dabei läßt sich die in diesem Zusammenhang bestehende Vergleichbarkeit von KIM und INTSINT schon aus den ähnlichen Signifikanzwerten der beiden Modelle ablesen. Abbildung 44 stellt die Ergebnisse im Überblick dar, wobei neben der Unterscheidung nach Satztypen auch der Prozentsatz der kreuzvalidierten Fälle über alle 576 Äußerungen angegeben ist.

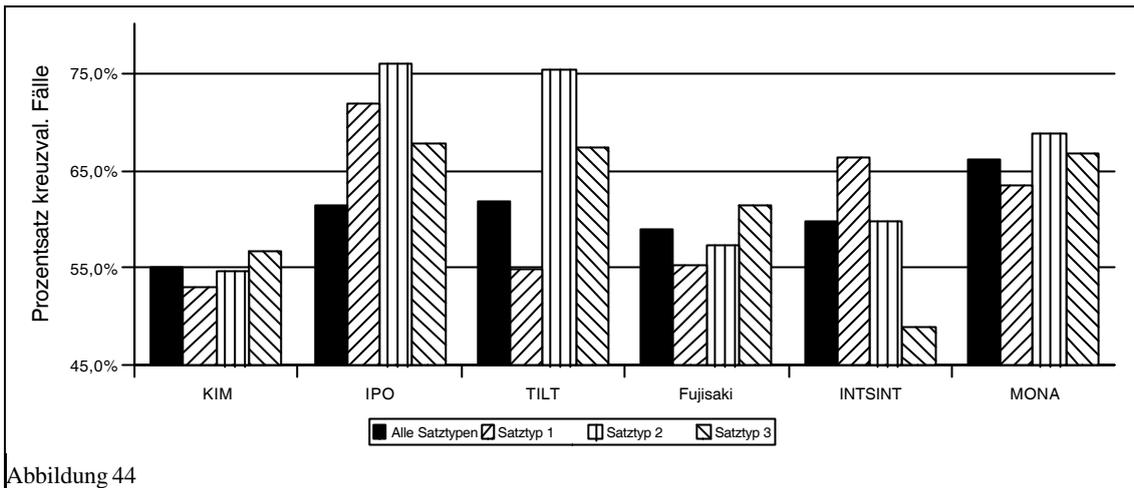


Abbildung 44

Wie Abbildung 44 zeigt, führt eine Unterscheidung nach Satztypen im Gegensatz zu der Zusammenfassung von allen 576 Äußerungen insgesamt eher zu besseren Ergebnissen. Zudem sind die Ergebnisse der einzelnen Intonationsmodelle für die drei Satztypen in etwa jeweils gleich, d.h. es gibt keine eindeutigen satztypspezifischen Leistungen der Intonationsmodelle. Abbildung 45 fasst die Ergebnisse für den kanonischen Korrelationskoeffizienten der einzelnen Diskriminanzanalysen der verschiedenen phonetischen Intonationsmodelle vergleichbar zusammen.

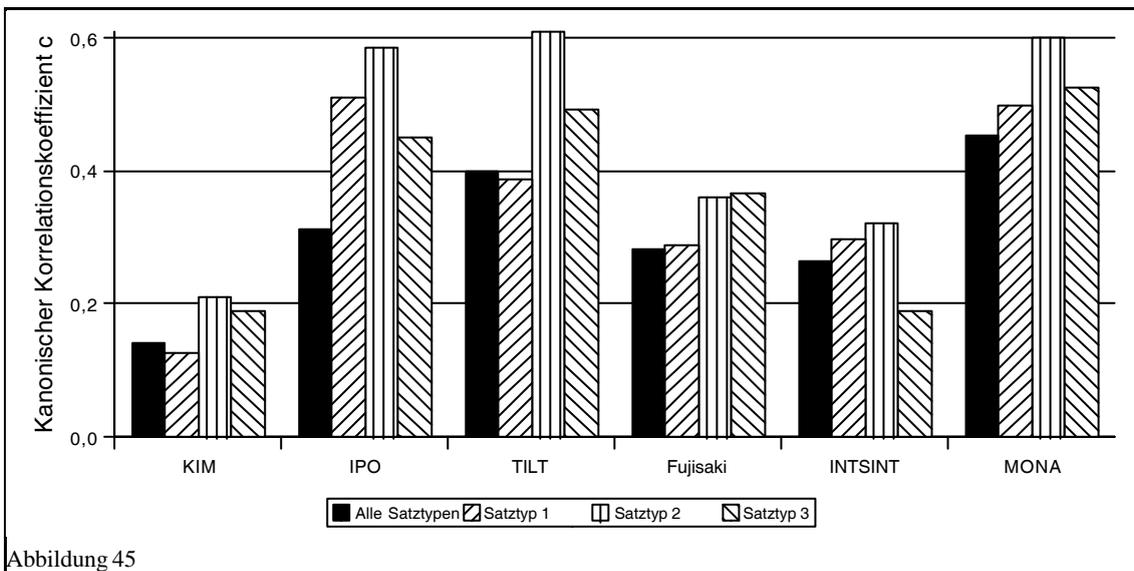


Abbildung 45

Die insgesamt eher schlechten Leistungen der untersuchten Intonationsmodelle sollen im folgenden verbessert werden, indem Batliner (1991) folgend eine Unterscheidung nach Sprechern vorgenommen wird.

9.2 Unterscheidung nach Sprechern

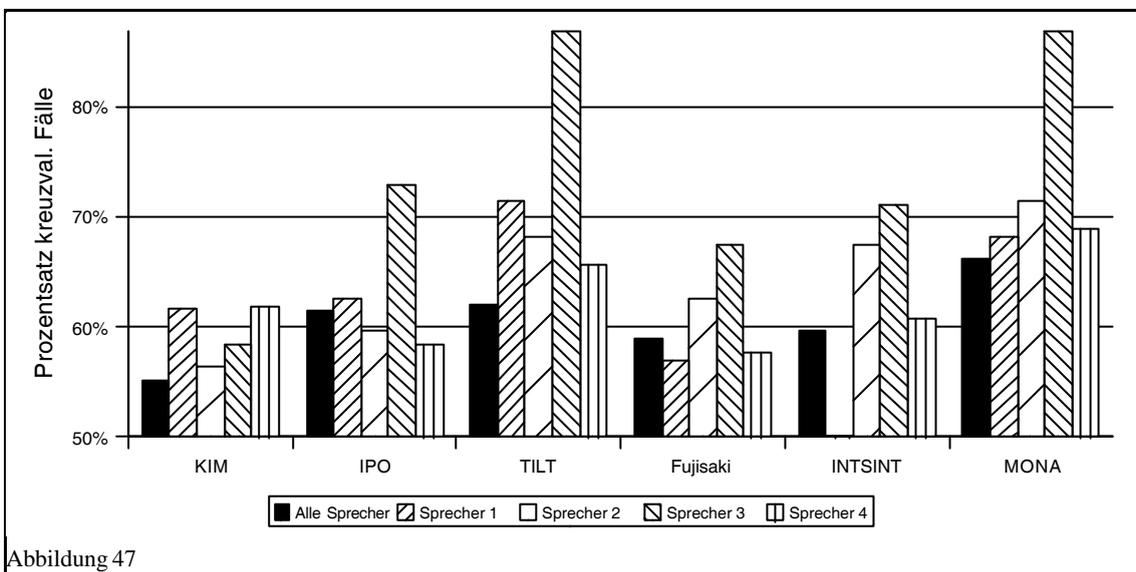
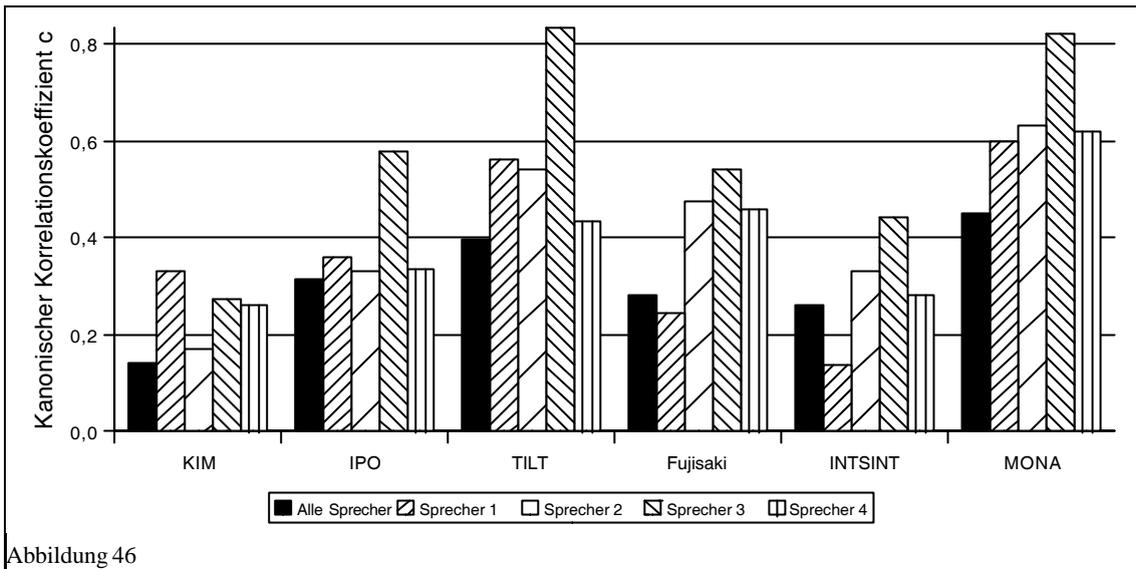
Tabelle 17 gibt die Ergebnisse der linearen Diskriminanzanalysen der Parameter der einzelnen Intonationsmodelle unterschieden nach Sprechern wieder; die vier Sprecher des MONA - Korpus sind in Tabelle 2, S. 57, kurz charakterisiert.

	LDA	KIM	IPO	TILT	Fujisaki	INTSINT	MONA
Sprecher 1	gültige Fälle N	144	144	143	121	63	144
	Korrelation c	0,33	0,36	0,56	0,24	0,14	0,6
	Wilks' Lambda	0,89	0,870	0,685	0,941	0,981	0,641
	Signifikanz	0	0,013	0,000	0,313	0,767	0,000
	Kreuzval. Fälle	61,6%	62,5%	71,5%	56,9%	42,9%	68,1%
Sprecher 2	gültige Fälle N	144	144	140	131	83	144
	Korrelation c	0,171	0,332	0,543	0,477	0,333	0,633
	Wilks' Lambda	0,971	0,890	0,705	0,773	0,889	0,599
	Signifikanz	0,389	0,040	0,000	0,000	0,025	0,000
	Kreuzval. Fälle	56,3%	59,7%	68,1%	62,5%	67,5%	71,5%
Sprecher 3	gültige Fälle N	144	144	135	131	83	144
	Korrelation c	0,272	0,578	0,835	0,540	0,441	0,824
	Wilks' Lambda	0,926	0,666	0,302	0,709	0,805	0,321
	Signifikanz	0,029	0,000	0,000	0,000	0,001	0,000
	Kreuzval. Fälle	58,3%	72,9%	86,8%	67,4%	71,1%	86,8%
Sprecher 4	gültige Fälle N	144	144	143	124	61	144
	Korrelation c	0,261	0,336	0,434	0,461	0,282	0,619
	Wilks' Lambda	0,932	0,887	0,811	0,787	0,920	0,616
	Signifikanz	0,043	0,035	0,002	0,000	0,189	0,000
	Kreuzval. Fälle	61,8%	58,3%	65,7%	57,6%	60,7%	68,8%
Mittelwert	Kreuzval. Fälle	58,500	63,350	73,025	61,125	60,375	73,800

Tabelle 17

Die Unterscheidung nach Sprechern liefert im Vergleich zu der Unterscheidung nach Satztypen generell etwas bessere Werte bei den einzelnen Intonationsmodellen, wobei die Unterschiede zwischen den Modellen aber grundsätzlich erhalten bleiben. Auch werden nur selten gut zu beurteilende Ergebnisse erzielt, also Ergebnisse, die deutlich über 80% liegen. Bei INTSINT ist erneut auf die Signifikanzwerte hinzuweisen, insbesondere bei dem ersten und dem vierten Sprecher. Bei den Parametern dieser beiden Sprecher ist nicht von einer Unterscheidung der Gruppen 'akzentuiert' und 'neutral' auszugehen.

Die Ergebnisse aus Tabelle 17 sind wiederum für die kreuzvalidierten Fälle und den kanonischen Korrelationskoeffizienten in Abbildung 46 und 47 zusammengefasst. Für die Sprecher 1, 3 und 4 liegt der kanonische Korrelationskoeffizient der sprecherbezogenen Diskriminanzfunktionen über den entsprechenden sprecherunabhängigen Werten. Lediglich für Sprecher 1 ergeben sich bei dem Fujisaki- und dem INTSINT-Modell kleinere Korrelationskoeffizienten. Auffällig ist zudem, dass die Werte des dritten Sprechers fast immer maximal innerhalb eines Modells sind. Die Unterscheidung der Äußerungen des MONA - Korpus nach Sprechern liefert demnach relevantere Ergebnisse als die Unterscheidung nach Satztypen, die in Kapitel 9.1, S. 97f, vorgenommen wurde.



9.3 Diskussion der multivariaten Analysen

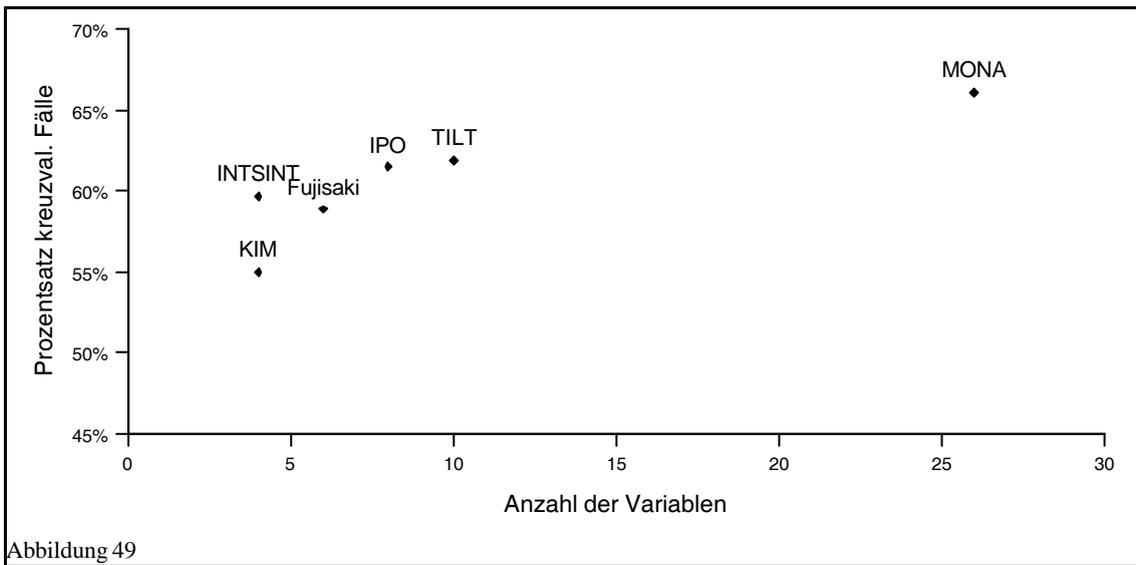
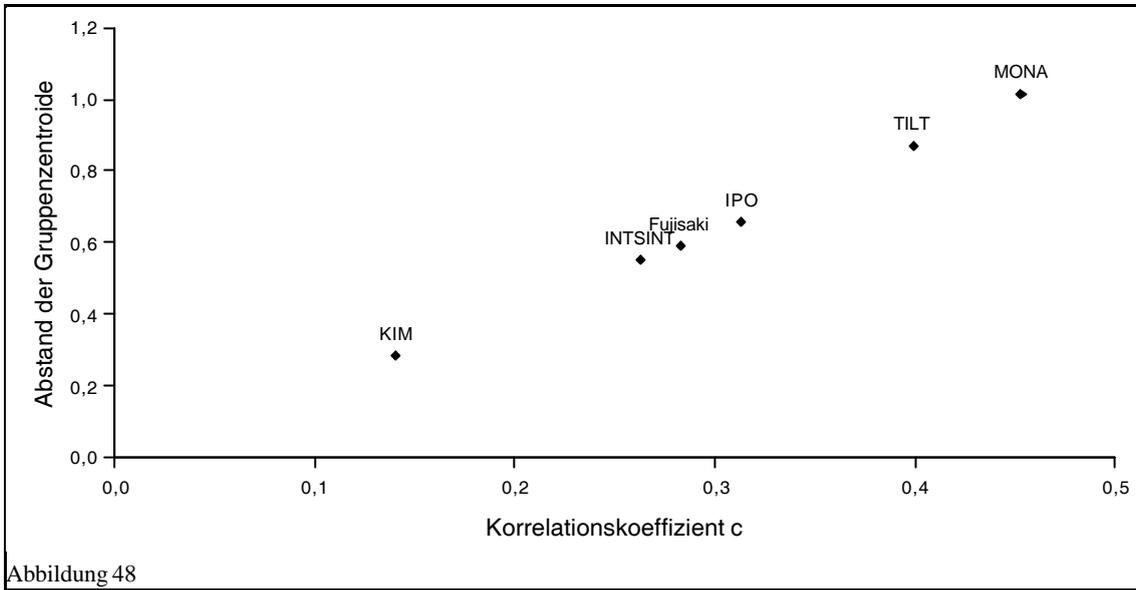
Die besseren Ergebnisse der linearen Diskriminanzanalysen mit der Unterscheidung nach Sprechern gegenüber der Unterscheidung nach Satztypen sprechen zunächst gegen die Annahme eines sprachsystematischen Zusammenhangs zwischen der intonatorischen und der syntaktischen Form eines Satzes (im Sinne einer potentiellen Äußerung). Intonatorische Formen haben demnach vor allem eine relevante sprecherspezifische Ausprägung. Wie in den Kapiteln 9.1 und 9.2 gezeigt, sind zumindest bei den einzelnen Sprechern deutliche Verbesserungen der Ergebnisse der Diskriminanzanalysen zu erkennen. Bei der Unterscheidung nach Satztypen konnten dagegen vergleichbare Differenzen nicht ermittelt werden. Eine weitere Unterscheidung bei den Diskriminanzanalysen etwa nach Sprechern und Satztypen wurde wegen der dabei bestehenden geringen Anzahl an Fällen nicht vorgenommen; auch unterblieb an dieser Stelle eine im Hinblick auf mögliche Deklinationseffekte gleichermaßen denkbare Unterscheidung nach erstem und zweiten Akzent.

Die eher schlechten Ergebnisse der Diskriminanzanalysen insgesamt weisen hingegen auf die Inadäquatheit der untersuchten phonetischen Intonationsmodelle hin. Eine genauere Betrachtung der Ergebnisse der verschiedenen Diskriminanzanalysen ergibt an dieser Stelle, dass ein linearer Zusammenhang zwischen der Anzahl der Parameter und des Prozentsatzes der korrekten (kreuzvalidierten) Fälle besteht; diesbezüglich fasst Tabelle 18 die Ergebnisse von Diskriminanzanalysen über alle Äußerungen ohne Unterscheidung nach Satztypen oder Sprechern entsprechend zusammen.

LDA	KIM	IPO	TILT	Fujisaki	INTSINT	MONA
Anzahl Variable	4	8	10	6	4	26
gültige Fälle N	576	576	561	507	290	576
Korrelation c	0,140	0,313	0,399	0,283	0,263	0,453
Wilkes Lambda	0,980	0,902	0,841	0,920	0,931	0,795
Signifikanz	0,023	0,000	0,000	0,000	0,000	0,000
Kreuzval. Fälle	55,0%	61,5%	61,9%	58,9%	59,7%	66,1%

Tabelle 18

Tabelle 18 zeigt, je mehr Variablen ein Modell besitzt, desto besser ist die Diskriminanzfunktion und dessen Klassifizierungsleistung. In Abbildung 48 sind zwei Gütewerte der berechneten Diskriminanzfunktion in Abhängigkeit voneinander eingetragen, der Abstand der Gruppenzentroide und der Korrelationskoeffizient. Gruppenzentroide sind die Mittelwerte der Diskriminanzwerte Y der einzelnen Gruppenelemente (vgl. die Diskriminanzfunktion auf S. 95). Ein möglichst großer Abstand der beiden Gruppenzentroide gilt dabei als Index für die Güte der Trennung der entsprechenden Modellparameter. Es zeigt sich, dass Modell KIM die schlechteste und das Modell MONA die beste Trennung bei der Diskriminanzanalyse aufweist. Abbildung 49 schließlich veranschaulicht den Zusammenhang zwischen der Anzahl an Variablen und dem Prozentsatz der korrekt klassifizierten Fälle.



10. ABSCHLIEBENDE DISKUSSION

Die Intonation (des Deutschen) ist einer allgemeinen Charakterisierung folgend in dem Bereich zwischen abstrakten, historisch gewachsenen Konventionen einer Sprachgemeinschaft und konkreten, physikalischen Manifestationen der gesprochenen Sprache der Gemeinschaft angesiedelt. Die Analyse der systematischen Beziehungen zwischen den konventionell bestimmten sprachlichen Bedeutungen der Intonation und deren akustischen Korrelaten ist dabei ein komplexes und - experimentell - bekanntermaßen schwer zu fassendes Phänomen. Die vorliegende Arbeit stellt den akustisch-phonetischen Aspekt in den Mittelpunkt und geht von einer semantisch konstanten Funktion aus. Damit besteht die Möglichkeit, verschiedene phonetische Intonationsmodelle bei der Beschreibung der korrespondierenden lautsprachlichen Formen der semantischen Funktion direkt miteinander zu vergleichen.

Die in der vorliegenden Arbeit untersuchten phonetischen Intonationsmodelle quantifizieren die akustische Manifestation von (bestimmten) semantischen Funktionen der Intonation. Eine derartige Quantifizierung ist in dem anwendungsbezogenen Kontext der Sprachtechnologie immanent erforderlich⁶¹. In der automatischen Sprachverarbeitung werden sprachakustischen Formen und syntaktisch-semantisch analysierbare symbolische Repräsentationen algorithmisch aufeinander abgebildet; dies setzt eine quantitativ explizite und regelhafte Formulierung der Zusammenhänge zwischen der sprachakustischen und der symbolischen Ebene voraus. Die Intonation ist in diesem Zusammenhang eines der notorisch großen Probleme, dessen Lösung mit der aktuell starken Entwicklung von automatischen Dialogsystemen zunehmend an Bedeutung gewinnt.

In der vorliegenden Arbeit wurde eine zweifellos wichtige dialogische Funktion der Intonation genauer untersucht, der intonatorische Ausdruck von bestimmten Kontrasten im Diskurs. Ein besonderes Anliegen war dabei die explizite Berücksichtigung von verschiedenen syntaktischen Bedingungen, deren Einfluss auf die Intonation in der (theoretisch orientierten) Linguistik stark diskutiert werden. Die Berücksichtigung trägt der Forderung in dem vorliegenden Zusammenhang der Sprachtechnologie nach der Nutzbarkeit von linguistischem Wissen Rechnung. Wie die experimentellen Ergebnisse aus Kapitel 7 und 9 zeigen, ist die sprecherspezifische Ausprägung der Intonation ein relativ deutliches Merkmal bei der Kennzeichnung von intonatorischen Formen. Die entsprechende Berücksichtigung von syntaktischen Kriterien ermittelte dagegen keine spezifischen intonatorischen Form. Ein weiteres, in diesem Zusammenhang interessantes Ergebnis der vorliegenden Arbeit ist, dass zumindest in den relativ kurzen Äußerungen der konstruierten Dialoge (vgl. Anhang A) der intonatorische Ausdruck von zwei Kontrasten eher ungewöhnlich ist. Damit ist auch im alltäglichen Sprachgebrauch davon auszugehen, dass derart komplexe Botschaften im Diskurs mit anderen als intonatorischen Mitteln ausgedrückt werden.

Die in Kapitel 9 berechneten, insgesamt eher schlechten Klassifikationsergebnisse der verschiedenen Diskriminanzanalysen lassen das Problem einer quantitativen Ausformulierung der Intonation (des Deutschen) weiterhin bestehen. Die in der vorliegenden Arbeit untersuchten phonetischen Intonationsmodelle konnten die Formen von kontrastiven Satzakzenten bei den - zumindest experimentell intendierten - komplexen Intonationsmustern der vorliegenden Arbeit nicht überzeugend analysieren. Aus den in diesem Zusammenhang zu erwähnenden entsprechend besseren Ergebnissen, von denen etwa Altmann/Batliner/Oppenrieder (1989) bei

⁶¹) Carstensen et al. (2001) geben einen aktuellen Überblick über die verschiedenen Aspekte der Sprachtechnologie.

Untersuchung von Äußerungen mit genau einem Satzakzent (Fokus) berichten, ist allerdings zu schließen, dass noch eine der vorliegenden Arbeit vergleichbare Untersuchung von Äußerungen mit einem kontrastiven Satzakzent durchzuführen wäre. Dieser methodische Abschluss bei der Untersuchung der phonetischen Intonationsmodelle müsste selbstverständlich auch andere intonatorische Funktionen als des kontrastiven Satzakzents beinhalten. Darüber hinaus ist natürlich auch anderes Sprachmaterial als im Labor elizitierte Sprache zu untersuchen. Aus Gründen eines angemessenen experimentellen Aufwandes konnten diese Analysen in der vorliegenden Arbeit nicht geleistet werden.

Bei der quantitativen Ausformulierung der Intonation bietet natürlich auch der in Kapitel 9.3 formulierte positive lineare Zusammenhang zwischen der Anzahl der berücksichtigten Parameter und der Klassifikationsleistung keine Lösung im Sinne einer Algorithmisierung der oben genannten Abbildung zwischen der sprachakustischen und der symbolischen Ebene. Eine qualitative Bewertung der Leistung der einzelnen Parameter bleibt somit grundsätzlich Bestandteil einer sprachakustisch quantitativen Bestimmung der Intonation (des Deutschen). Ein zentraler Nachteil der phonetischen Intonationsmodelle ist in diesem Zusammenhang zweifellos auch die Beschränkung auf den Grundfrequenzverlauf. Andere akustisch-phonetische Eigenschaften wie Dauer-, Intensitäts- oder Formanteigenschaften berücksichtigen die untersuchten Intonationsmodelle nicht. Bei der Modellierung von Dauereigenschaften der Grundfrequenzverläufe dagegen werden diese gegenüber den Frequenzeigenschaften der Verläufe von den einzelnen Modellen zumeist nur implizit behandelt.

Zukünftige Untersuchungen der lautsprachlichen Realisierung der Intonation sollten somit weitere Parameter wie etwa die oben genannten behandeln. Besonders interessant ist dabei auch die genauere Untersuchung von bestimmten Interaktionen wie beispielsweise die Interaktion der lokalen und globalen Komponenten eines Grundfrequenzverlaufs. Von generellem Interesse ist in diesem Zusammenhang zudem die - weitere - Automatisierung von Modellierungen, so dass zur Verbesserung der empirischen Basis auf umfangreichere Sprachkorpora zugegriffen werden kann. Die in der vorliegenden Arbeit entwickelte Abbildung von tonalen Segmenten auf IPO-Standardkonturen ist eine Komponente einer derartigen Automatisierung (vgl. Kap. 8.2.1).

ANHANG A. DIALOGSTIMULI

Akzentpositionen	Dialoge
Subjekt (=S), temporales adverbiales Adjektiv	A01. Klaus bucht in den nächsten beiden Tagen eine Reise, und Mona auch. Ich glaube, Klaus bucht morgen eine Reise. B01-1. Mona bucht heute eine Reise. B01-2. Mona heute. B01-3. Heute bucht Mona eine Reise.
S, attributives Adjektiv der adverbialen PGr (=PGr)	A02. Klaus bucht heute für einen seiner beiden Söhne eine Reise, und Mona auch. Ich glaube, Klaus bucht für seinen jüngsten Sohn eine Reise. B02-1. Mona bucht heute für den ältesten Sohn eine Reise. B02-2. Mona für den ältesten Sohn. B02-3. Für den ältesten Sohn bucht Mona heute eine Reise.
S, Nominal der adverbialen PGr	A03. Klaus bucht heute für eines seiner beiden Kinder eine Reise, und Mona auch. Ich glaube, Klaus bucht für seine Tochter eine Reise. B03-1. Mona bucht heute für den Sohn eine Reise. B03-2. Mona für den Sohn. B03-3. Für den Sohn bucht Mona heute eine Reise.
S, attributives Adjektiv des Akkusativobjekts (= AO)	A04. Klaus bucht heute eine Reise, und Mona auch. Ich glaube, Klaus bucht eine teure Reise. B04-1. Mona bucht heute eine billige Reise. B04-2. Mona eine billige Reise. B04-3. Eine billige Reise bucht Mona heute.
S, erster Teil des Kompositums des Nomens des AO	A05. Klaus bucht heute eine Reise, und Mona auch. Ich glaube, Klaus bucht eine Flugreise. B05-1. Mona bucht heute eine Schiffsreise. B05-2. Mona eine Schiffsreise. B05-3. Eine Schiffsreise bucht Mona heute.
S, Nomen der lokalen adverbialen PGr	A06. Klaus bucht heute eine Reise, und Mona auch. Ich glaube, Klaus bucht eine Reise nach Schottland. B06-1. Mona bucht heute eine Reise nach England. B06-2. Mona nach England. B06-3. Nach England bucht Mona heute eine Reise.
Prädikat, attributives Adjektiv der adverbialen PGr	A07. Mona geht heute für einen ihrer beiden Söhne in ein Reisebüro. Ich glaube, sie storniert für ihren jüngsten Sohn eine Reise. B07-1. Mona bucht heute für ihren ältesten Sohn eine Reise. B07-2. Bucht für ihren ältesten Sohn. B07-3. Für ihren ältesten Sohn bucht Mona heute eine Reise.
Prädikat, Nominal der adverbialen PGr	A08. Mona geht heute für eines ihrer beiden Kinder in ein Reisebüro. Ich glaube, sie storniert für ihre Tochter eine Reise. B08-1. Mona bucht heute für ihren Sohn eine Reise. B08-2. Bucht für ihren Sohn. B08-3. Für ihren Sohn bucht Mona heute eine Reise.
Prädikat, attributives Adjektiv des AO	A09. Mona geht heute in ein Reisebüro, um sich um eine Reise von Klaus zu kümmern. Ich glaube, sie storniert eine teure Reise. B09-1. Mona bucht heute eine billige Reise. B09-2. Bucht eine billige Reise. B09-3. Eine billige Reise bucht Mona heute.

Akzentpositionen	Dialoge
Prädikat, erster Teil des Kompositums des Nomens des AO	A10. Mona kümmert sich heute um eine Flug- oder Schiffsreise von Klaus. Ich glaube, sie storniert eine Flugreise. B10-1. Mona bucht heute eine Schiffsreise. B10-2. Bucht eine Schiffsreise. B10-3. Eine Schiffsreise bucht Mona heute.
Prädikat, Nomen der lokalen adverbialen PGr	A11. Mona kümmert sich heute um eine Reise von Klaus. Ich glaube, sie storniert eine Reise nach Schottland. B11-1. Mona bucht heute eine Reise nach England. B11-2. Bucht nach England. B11-3. Nach England bucht Mona heute eine Reise.
temporales adverbiales Adjektiv, attributives Adjektiv der adverbialen PGr	A12. Mona bucht in den nächsten zwei Tagen für einen ihrer beiden Söhne eine Reise. Ich glaube, sie bucht morgen für ihren jüngsten Sohn eine Reise. B12-1. Mona bucht heute für ihren ältesten Sohn eine Reise. B12-2. Heute für ihren ältesten Sohn. B12-3. Für ihren ältesten Sohn bucht Mona heute eine Reise.
temporales adverbiales Adjektiv, Nominal der adverbialen PGr	A13. Mona bucht heute oder morgen für eines ihrer beiden Kinder eine Reise. Ich glaube, sie bucht morgen für ihre Tochter eine Reise. B13-1. Mona bucht heute für ihren Sohn eine Reise. B13-2. Heute für ihren Sohn. B13-3. Für ihren Sohn bucht Mona heute eine Reise.
temporales adverbiales Adjektiv, attributives Adjektiv des AO	A14. Mona bucht heute oder morgen eine Reise, deren Kosten sie noch nicht kennt. Ich glaube, sie bucht morgen eine teure Reise. B14-1. Mona bucht heute eine billige Reise. B14-2. Heute eine billige Reise. B14-3. Eine billige Reise bucht Mona heute.
temporales adverbiales Adjektiv, erster Teil des Kompositums des Nomens des AO	A15. Mona bucht heute oder morgen noch eine Reise. Ich glaube, sie bucht morgen eine Flugreise. B15-1. Mona bucht heute eine Schiffsreise. B15-2. Heute eine Schiffsreise. B15-3. Eine Schiffsreise bucht Mona heute.
temporales adverbiales Adjektiv, Nomen der lokalen adverbialen PGr	A16. Mona bucht heute oder morgen noch eine Reise. Ich glaube, sie bucht morgen eine Reise nach Schottland. B16-1. Mona bucht heute eine Reise nach England. B16-2. Heute nach England. B16-3. Nach England bucht Mona heute eine Reise.
attributives Adjektiv der adverbialen PGr, attributives Adjektiv des AO	A17. Mona bucht heute für einen ihrer beiden Söhne eine Reise. Ich glaube, sie bucht heute für ihren jüngsten Sohn eine teure Reise. B17-1. Mona bucht heute für ihren ältesten Sohn eine billige Reise. B17-2. Für ihren ältesten Sohn eine billige Reise. B17-3. Eine billige Reise bucht Mona heute für ihren ältesten Sohn.
attributives Adjektiv der adverbialen PGr, erster Teil des Kompositums des Nomens des AO	A18. Mona bucht heute für einen ihrer beiden Söhne eine Reise. Ich glaube, sie bucht für ihren jüngsten Sohn eine Flugreise. B18-1. Mona bucht heute für ihren ältesten Sohn eine Schiffsreise. B18-2. Für ihren ältesten Sohn eine Schiffsreise. B18-3. Eine Schiffsreise bucht Mona heute für ihren ältesten Sohn.

Akzentpositionen	Dialoge
attributives Adjektiv der adverbialen PGr,	A19. Mona bucht heute für einen ihrer beiden Söhne eine Reise. Ich glaube, sie bucht für ihren jüngsten Sohn eine Reise nach Schottland.
Nomen der lokalen adverbialen PGr	B19-1. Mona bucht heute für ihren ältesten Sohn eine Reise nach England. B19-2. Für ihren ältesten Sohn nach England.
Nominal der adverbialen PGr, attributives Adjektiv des AO	B19-3. Eine Reise nach England bucht Mona heute für ihren ältesten Sohn. A20. Mona bucht heute für eines ihrer beiden Kinder eine Reise. Ich glaube, sie bucht für ihre Tochter eine teure Reise.
Nominal der adverbialen PGr, erster Teil des Kompositums des Nomens des AO	B20-1. Mona bucht heute für ihren Sohn eine billige Reise. B20-2. Für ihren Sohn eine billige Reise. B20-3. Eine billige Reise bucht Mona heute für ihren Sohn.
Nominal der adverbialen PGr, erster Teil des Kompositums des Nomens des AO	A21. Mona bucht heute für eines ihrer beiden Kinder eine Reise. Ich glaube, sie bucht für ihre Tochter eine Flugreise. B21-1. Mona bucht heute für ihren Sohn eine Schiffsreise. B21-2. Für ihren Sohn eine Schiffsreise. B21-3. Eine Schiffsreise bucht Mona heute für ihren Sohn.
Nominal der adverbialen PGr, Nomen der lokalen adverbialen PGr	A22. Mona bucht heute für eines ihrer beiden Kinder eine Reise. Ich glaube, sie bucht für ihre Tochter eine Reise nach Schottland.
attributives Adjektiv des AO, Nomen der lokalen adverbialen PGr	B22-1. Mona bucht heute für ihren Sohn eine Reise nach England. B22-2. Für ihren Sohn nach England. B22-3. Nach England bucht Mona heute für ihren Sohn eine Reise.
erster Teil des Kompositums des Nomens des AO, Nomen der lokalen adverbialen PGr	A23. Mona bucht heute eine Reise, die entweder nach England oder nach Schottland gehen soll. Ich glaube, sie bucht eine teure Reise nach Schottland. B23-1. Mona bucht heute eine billige Reise nach England. B23-2. Eine billige Reise nach England. B23-3. Nach England bucht Mona heute eine billige Reise.
erster Teil des Kompositums des Nomens des AO, Nomen der lokalen adverbialen PGr	A24. Mona bucht heute entweder eine Flug- oder eine Schiffsreise. Ich glaube, sie bucht eine Flugreise nach Schottland.
AO, Nomen der lokalen adverbialen PGr	B24-1. Mona bucht heute eine Schiffsreise nach England. B24-2. Eine Schiffsreise nach England. B24-3. Nach England bucht Mona heute eine Schiffsreise.

ANHANG B. GEWICHTUNGSFUNKTION

```
function w = wtam_fkt(ti, f0, f0zeitstart, f0zeitende, alpha, ipol, wtam_fenster)
% Gewichtung des F0-Verlaufs einer Silbe
if length(ti) < 2
    w = f0; return
end

% Splinefunktion anlegen; Annahme: max. alle 10ms ein Messwert, neue Zeitpunkte jede ms
t_spline = f0zeitstart:ipol:f0zeitende;
if t_spline(length(t_spline)) ~= f0zeitende
    t_spline = [t_spline f0zeitende]
end
f0_spline = csapi(ti, f0, t_spline);

% Berechnung eines gewichteten f0-Wertes für jeden Messwert-Zeitpunkt ti
w = [f0(1)];
for j = 2:length(ti)
    % bestimme Zeitreihe für Interpolation auf Basis der interpolierten und tatsächlich gemessenen Werte:
    f0_j = []; ti_j = [];

    % suche den Startindex des für die Berechnung des Integrals, d.h. den ersten Messwert f0, der im Fenster
    % enthalten ist; bei wtam_fenster = 0 werden immer alle vorherigen Messwerte verwendet
    j1 = 1;
    if wtam_fenster > 0
        while ti(j1) < ti(j) - wtam_fenster
            % suche ersten Messwert, der im Fenster liegt
            j1 = j1 + 1;
        end
    end

    % suche Startindex der Spline-Reihe:
    k1 = 1;
    if wtam_fenster > 0
        while t_spline(k1) < ti(j) - wtam_fenster
            % suche ersten Messwert, der im Fenster liegt
            k1 = k1 + 1;
        end
    end

    j_i = j1; k_i = k1;

    while t_spline(k_i) < ti(j)
        if t_spline(k_i) > ti(j_i)
            f0_j = [f0_j f0(k_i)]; % Messwert in F0-Reihe einfügen
            ti_j = [ti_j ti(j_i)]; % Messwert in Zeitreihe einfügen
            j_i = j_i + 1;
        end
        if t_spline(k_i) ~= ti(j_i)
```

```

    f0_j = [f0_j f0_spline(k_i)]; %Splinewert an F0-Reihe anhängen
    ti_j = [ti_j t_spline(k_i)]; %Splinewert an Zeitreihe anhängen
end
k_i = k_i + 1;
if k_i > length(t_spline)
    break;
end
end
end
f0_j = [f0_j f0(j)]; ti_j = [ti_j ti(j)];

%Bestimmung der zu integrierenden Reihe
g = exp( alpha .* (ti(j)- ti_j)); %Gewichtungsfunktion bezogen auf Messwert ti(j)
f0_g = g .* f0_j; %gewichteten Messwerte

%Integral
integral_oben = trapz(ti_j, f0_g); integral_unten = trapz(ti_j, g);
w = [w integral_oben/integral_unten];

end
return;

```

ANHANG C. VERWENDETE SOFTWARE

In dem experimentellen Teil dieser Arbeit wurde die folgende Software verwendet, die zum Teil an die Systemumgebung angepasst werden musste

Displaying Tool for Fujisaki's model & J_ToBI für die Berechnung der Fujisaki-Parameter, s.u. http://www.itl.atr.co.jp/department/dept2/fj_ToBI/

Edinburgh Speech Tools für die Tilt-Analyse, s.u.
http://www.cstr.ed.ac.uk/projects/speech_tools/

Matlab, Version 5.3, für die Berechnung der Gewichtungsfunktion und der tonalen Segmentierung bei der Umsetzung des IPO-Modells

Mes Signaux package, Speech Signal processing tools(v 1.0) für die MOMEL/INTSINT – Analyse und Synthese, s.u. http://www.lpl.univ-aix.fr/ext/projects/mes_signaux.htm/

Perl für die WWW-Programmierung und für Routinen der Dateikonvertierung

Praat für die Sprachsignalanalyse und -synthese, s.u. <http://www.fon.hum.uva.nl/praat/>

SPSS for Windows, Version 10.0.5, für statistische Analysen

Visual Basic, Professional Version 6.0, für das Programm des Produktionsexperiments

ANHANG D. LITERATUR

- Adrians, L.M.H. (1991), Ein Modell deutscher Intonation. Dissertation, Technische Universität Eindhoven.
- Altmann, H. /Batliner, A. /Oppenrieder, W. (Hg.) (1989): Zur Intonation von Modus und Fokus im Deutschen. Linguistische Arbeiten 234. Tübingen: Niemeyer.
- Auer, P. (1991), Zur More in der Phonologie. Zeitschrift für Sprachwissenschaft 10, Heft 1, 3-36.
- Auer, P./Gilles, P./Peters, J./Selting, M. (2000): Intonation regionaler Varietäten des Deutschen. Vorstellung eines Forschungsprojekts. In: Stellmacher, D. (Hg), Dialektologie zwischen Tradition und Neuansätzen. Beiträge der Internationalen Dialektologentagung, Göttingen, 19.-21.Okt. 1998, Stuttgart: Steiner, 222-239.
- Backhaus, K./Erichson, B./Wulff, P./ Weiber, R. (2000), Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Neunte, überarbeitete und erweiterte Auflage. Berlin: Springer Verlag.
- Bagshaw, P. C. (1994), Automatic prosodic analysis for computer aided pronunciation teaching. Dissertation, Universität Edinburgh.
- Bannert, R./Schmidt, W./Smude, P. (1994), Intonationserkennung. In: Hoepelman, J./Machate, J. (Hg), Modellbildung für die Auswertung der Fokusintonation im gesprochenen Dialog (MAFID). Beiträge zur Dialogforschung, Bd. 7. Tübingen: Niemeyer, 1-52.
- Batliner, A. (1991), Deciding upon the Relevancy of Intonational Features for the Marking of Focus: a Statistical Approach. Journal of Semantics 8, 171-189.
- Batliner, A. (1994), Prosody, Focus, and Focal Structure: Some Remarks on Methodology. Verbmobil Report 58.
- Batliner, A./Buckow, J./Niemann, H./Nöth, E./Warnke, V. (2000a), The Prosody Module. In: Wahlster, W. (ed.), Verbmobil: Foundations of Speech-to-Speech Translation. Berlin: Springer Verlag, 106-121.
- Batliner, A./Huber, R./Niemann, H./Nöth, E./Spilker, J./Fischer, K. (2000b), The Recognition of Emotion. In: Wahlster, W. (ed.), Verbmobil: Foundations of Speech-to-Speech Translation. Berlin: Springer Verlag, 122-130.
- Batliner, A./Nöth, E./Möbius, B./Möhler, G. (2000c), Prosodic models and speech recognition: towards the common ground. Proceedings of Prosody 2000, Krakau.
- Beckman, M. E. (1997), A Typology of Spontaneous Speech. In: Sagisaka, Y./Campbell, N./Higuchi, N. (eds.), Computing Prosody. Computational Models for Processing Spontaneous Speech. Berlin: Springer, 7-26.
- Beckman, M./Pierrehumbert, J. (1986), Intonational Structure in Japanese and English. Phonology Yearbook 3, 255-309.
- Blaauw, E. (1995), On the Perceptual Classification of Spontaneous and Read Speech. Dissertation, Universität Utrecht.
- Boersma, P. (1993): Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 17: 97-110.
- Bolinger, D. (1972). Accent is predictable (if you're a mindreader). Language 48, 633-644.
- Bosch, P./van der Sandt, R. (1999) (eds.), Focus. Linguistic, Cognitive, and Computational Perspectives. Cambridge: Cambridge University Press.
- Brindöpke, Ch./Schaffranietz, B. (1999), Ein Transkriptionssystem für die Sprachmelodie des Deutschen. Linguistische Berichte 179, 286-306.
- Burger, S. (1997), Transliteration spontansprachlicher Daten - Lexikon der Transliterationskonventionen - VERBMOBIL II. Verbmobil Technisches Dokument 56.
- Burger, S./Schiel, F. (1995), RVG 1 - A Database for Regional Variants of Contemporary German. Proceedings of LREC, Granada 1995.

- Burger, S./Weilhammer, K./Schiel, F./Tillmann, H.G. (2000), *Verbmobil Data Collection and Annotation*. In: Wahlster, W. (ed.), *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin: Springer, 537-549.
- Carstensen, K.U./Ebert, Ch./Endriss, C./Jekat, S./Klabunde, R./Langer, H. (2001), *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag.
- Chun, D.M. (1998), *Signal Analysis Software for Teaching Discourse Intonation*. *Language Learning & Technology*, Vol. 2, No. 1, 61-77, [on-line available] <http://llt.msu.edu>.
- Clauß, G./Finze, F.-R./Partsch, L. (1995), *Statistik für Soziologen, Pädagogen, Psychologen und Mediziner*. Band 1, 2., überarbeitete und erweiterte Auflage. Thun/Frankfurt am Main: Verlag Harri Deutsch.
- d'Allessandro, Ch./Mertens, P. (1995), *Automatic pitch contour stylization using a model of tonal perception*. *Computer Speech and Language* 9, 257-288.
- Delin, J./Zacharski, R. (1997), *Pragmatic Determinants of Intonation Contours for Dialogue Systems*. In: *International Journal of Speech Technology* 1, 109-120.
- Duden Aussprachewörterbuch (1990), (= Duden, Band 6), 3., Auflage. Mannheim: Dudenverlag.
- Ehlich, K./Rehbein, J. (1979): *Erweiterte Halbinterpretative Arbeitstranskriptionen (HIAT 2): Intonation*. *Linguistische Berichte* 59, 51-75.
- Féry, C. (1993), *German Intonational Patterns*. *Linguistische Arbeiten* 285. Tübingen: Niemeyer.
- Fox, A. (2000), *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Oxford: Oxford University Press.
- Fuchs, A. (1984), *Deaccenting and default accent*. In: Gibbon, D./ Richter, H. (eds.), *Intonation, Accent and Rhythm. Studies in Discourse Phonology*. Berlin: de Gruyter, 134-164.
- Fujisaki, H. (1988), *A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour*. In: Fujimura, O. (Hg.): *Vocal physiology: voice production, mechanisms and functions*, 347-355. New York: Raven
- Geumann, A./Oppermann, D./Schaeffler, F. (1997), *The conventions for phonetic transcription and segmentation of German used for the Munich Verbmobil Corpus*. *Verbmobil Memo* 127.
- Gibbon, D. (1998), *Intonation in German*. In: Hirst, D./Di Cristo, A. (eds.), *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Grabe, E./Nolan, F./Farrar, K.J. (1998), *IViE - A Comparative Transcription system for Intonational Variation in English*. *Proceedings ICSLP Sydney 1998*, Vol. 4, 1259-1262.
- Gussenhoven, C. (1984), *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
- Gussenhoven, C./Rietveld, A.C.M. (1988), *Fundamental frequency declination in Dutch: testing three hypotheses*. *Journal of Phonetics* 16, 355-369.
- Halliday, M.A.K. (1967), *Intonation and Grammar in British English*. The Hague: Mouton.
- Helfrich, H. (1985), *Satzmelodie und Sprachwahrnehmung. Psycholinguistische Untersuchungen zur Grundfrequenz*. Berlin: de Gruyter.
- Hess, W. (1996), *Neuere Entwicklungen in der Sprachsynthese*. *Verbmobil Report* 174.
- Heuft, B. (1999), *Eine prominenzbasierte Methode zur Prosodieanalyse und -synthese*. Frankfurt am Main: Peter Lang.
- Hirst, D. (1999), *The Symbolic Coding of Segmental Duration and Tonal Alignment: An Extension to the INTSINT System*. *Proceedings Eurospeech 1999*.
- Hirst, D./Di Cristo, A. (1998), *A survey of intonation systems*. In: Hirst, D./Di Cristo (eds.), *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1-44.
- Hirst, D./Di Cristo, A./Espesser, R. (2001), *Levels of representation and levels of analysis for the description of intonation systems*. In: Horne, M. (ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Press. [hier Manuskript von Daniel Hirst verwendet].
- Höhle, T. N. (1982), *Explication für 'normale Betonung' und 'normale Wortstellung'*. In: Abraham, W. (Hg.): *Satzglieder im Deutschen. (Studien zur deutschen Grammatik 15)*, 75-153.
- House, D. (1990), *Tonal Perception of Speech*. Lund: Lund University Press.

- Huber, D. (1988), Aspects of the communicative function of voice in text intonation. Dissertation, Universität Göteborg.
- Isaënenko, A.V./Schädlich, H.J. (1966), Untersuchungen über die deutsche Satzintonation. *studia grammatica* VII.
- Jacobs, J. (1983), Fokus und Skalen. Zur Syntax und Semantik der Gradpartikel im Deutschen. *Linguistische Arbeiten* 138. Tübingen: Niemeyer.
- Jacobs, J. (1984), Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte* 91, 25-58.
- Jacobs, J. (1988), Fokus-Hintergrund-Gliederung und Grammatik. In: Altmann, H. (Hg.), *Intonationsforschungen*. *Linguistische Arbeiten* 200. Tübingen: Niemeyer, 89-134.
- Jacobs, J. (1997), I-Topikalisierung. *Linguistische Berichte* 168, 91-133.
- Klinghardt, H. (1927), *Übungen im deutschen Tonfall*. Für Lehrer und Studierende, auch für Ausländer. Leipzig: Quelle und Meyer.
- Kohler, K. J. (1983), F0 in Speech Timing. In: Kohler, K. J./Hoequist, Ch. E. (eds.), *Studies in Speech Timing*. *Arbeitsberichte des Instituts für Phonetik, Universität Kiel*, Nr. 20, 55-97.
- Kohler, K.J. (1994), *Lexica of the Kiel PHONDAT Corpus*. Read Speech. *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung der Universität Kiel* 27.
- Kohler, K.J. (1991), A Model of German Intonation. In: Kohler, K.J. (ed.), *Studies in German Intonation*. *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung der Universität Kiel* 25, 295-360.
- Kohler, K. (1997), Modelling Prosody in Spontaneous Speech. In: Sagisaka, Y./Campbell, N./Higuchi, N. (eds.), *Computing Prosody. Computational Models for Processing Spontaneous Speech*. New York: Springer, 188-210.
- Kohler, K. J./van Dommelen, W. A. (1986), Prosodic Effects on Lenis/Fortis Perception: Preplosive Fo and LPC Synthesis. *Phonetica* 43, 70-75.
- Kohonen, T. (1995), *Self-Organizing Maps*. Berlin: Springer.
- Kröger, B. (1996), Artikulatorische Sprachsynthese. In: *Fortschritte der Akustik, DAGA 96*, 96-99.
- Ladd, D. R. (1996), *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lambrecht, K. (1994), Information structure and sentence form: topic, focus, and the mental representations of discourse referents. Cambridge: Cambridge University Press.
- Lehiste, I./Peterson, G. E. (1961), Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America* 33, 419-425 (hier Nachdruck aus Fry, D.B. (ed.) (1976), *Acoustic Phonetics. A course in basic readings*. Cambridge: Cambridge University Press, 378-393).
- Lieb, H.-H. (1983), Akzent und Negation im Deutschen - Umriss einer einheitlichen Konzeption. Teil A. *Linguistische Berichte* 84, 1-32.
- Lemmety, L. (1999), *Review of Speech Synthesis Technology*. Dissertation, Technische Universität Helsinki.
- Levinson, S.C. (2000), *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge, Mass.: MIT Press.
- Maas, U. (1999), *Phonologie. Einführung in die funktionale Phonetik des Deutschen*. Opladen/ Wiesbaden: Westdeutscher Verlag.
- Mayer, J. (1997), *Intonation und Bedeutung. Aspekte der Prosodie-Semantik-Schnittstelle im Deutschen*. Dissertation, Universität Stuttgart.
- Mertens, P./d'Allesandro, Ch. (1995), Pitch contour stylization using a tonal perception model. *Proceedings of ICPhS 95, Stockholm*, Vol. 4, 228-223.
- Mixdorff, H. (1998), *Intonation Patterns of German - Model-based Quantitative Analysis and Synthesis of F0 contours*. Dissertation, TU Dresden.
- Mixdorff, H. (2000), A novel approach to the fully automatic extraction of Fujisaki model parameters. *Proceedings of the International Conference on Spoken Language Processing*. Beijing, China.
- Möbius, B. (1993), Ein quantitatives Modell der deutschen Intonation. Analyse und Synthese von Grundfrequenzverläufen. *Linguistische Arbeiten* 305. Tübingen: Niemeyer.

- Möbius, B./Zimmermann, A. /Hess, W. (1987), Untersuchungen zu mikroprosodischen Grundfrequenzvariationen im Deutschen. In: Tillmann, H.-G./Willée, G. (Hg.), *Analyse und Synthese gesprochener Sprache*. Hildesheim: Olms, 102-110.
- Möhler, G. (1998), *Theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese*. Dissertation, Universität Stuttgart.
- Monaghan, A.I.C. (1993), What determines accentuation? A reply to Cruttenden and Faber. *Journal of Pragmatics* 19, 559-584.
- Peters, B. (1999), *Prototypische Intonationsmuster in deutscher Lese- und Spontansprache*. Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung der Universität Kiel 34, 1-175.
- Pierrehumbert, J. (1980), *The Phonology and Phonetics of English Intonation*. Dissertation, MIT.
- Pierrehumbert, J. (1990), Phonological and phonetic representation. *Journal of Phonetics* 18, 375-394.
- Rapp, S. (1998), *Automatisierte Erstellung von Korpora für die Prosodieforschung*. Dissertation, Universität Stuttgart.
- Reetz, H. (1996), *Pitch Perception in Speech: A Time Domain Approach*. Dordrecht: Foris Publications.
- Reinicke, J. (1995), *Konzept einer Arbeitsstation zur Segmentierung und Etikettierung prosodischer Einheiten*. *Verbmobil Report* 73.
- Reinicke, J. (1996), Resynthese als Hilfsmittel bei der prosodischen Etikettierung. *Proc. ITG-Fachtagung Sprachkommunikation, Frankfurt a.M., September 1996 (= ITG-Fachbericht 139)*, 43-46.
- Reyelt, M./Batliner, A. (1994), *Ein Inventar prosodischer Etiketten für VERBMOBIL*. *Verbmobil Memo* 33.
- Richter, H. (1966a), Zur Kategorialität segmenteller Abhörtexte. In: Richter, H./Zwerner, E. (Hg.), *Gesprochene Sprache. Probleme ihrer strukturalistischen Beschreibung*. Wiesbaden (= DFG-Forschungsbericht 7), 26-48.
- Richter, H. (1966b), Anleitung zur auditiv-phänomenalen Beurteilung der suprasegmentellen Eigenschaften sprachlicher Äußerungen. In: Richter, H. /Zwerner, E. (Hg.), *Gesprochene Sprache. Probleme ihrer strukturalistischen Beschreibung*. Wiesbaden (= DFG-Forschungsbericht 7), 11-21.
- Richter, H. (1967), Zur Intonation der Bejahung und Verneinung im Hochdeutschen. In: *Satz und Wort im heutigen Deutsch (= Sprache der Gegenwart, Bd.1)*, 329-362.
- Rose, P. (1987), Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication* 6, 343-351.
- Rost, J. (1988), *Quantitative und qualitative probabilistische Testtheorie*. Bern, Stuttgart, Toronto: Verlag Hans Huber.
- Sappock, Ch. (1989), Dialog und Intonation. In: Weigand, E./Hundsnurscher, F. (Hg.), *Dialoganalyse II, Bd.2. Linguistische Arbeiten* 230, 423-434
- Schubiger, M. (1965), English Intonation and German Modal Particles: A Comparative Study. *Phonetica* 12, 65-84.
- Searle, J. (1969), *Speech acts*. Cambridge: Cambridge University Press.
- Selkirk, E.O. (1984), *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: M.I.T. Press.
- Selting, M. (1992), Phonologie der Intonation. Probleme bisheriger Modelle und Konsequenzen einer neuen interpretativ-phonologischen Analyse. *Zeitschrift für Sprachwissenschaft* 11/1, 99-138.
- Selting, M./Auer, P./Barden, B./Bergmann, J./Couper-Kuhlen, E./Günthner, S./Meier, Ch./ Quasthoff, U./Schlobinski, P./Uhmann, S. (1998), *Gesprächsanalytisches Transkriptionssystem (GAT)*. *Linguistische Berichte* 173, 91-122.
- Silverman, K. (1986), F0 segmental cues depend on intonation: The case of rise after voiced stops. *Phonetica* 43, 76-91.
- Silverman, K./Beckman, M./Pitrelli, J./Ostendorf, M./Wightman, C./Price, P./Pierrehumbert, J./Hirschberg, J. (1992), ToBI: A standard system for labeling prosody. *Proceedings ICSLP 92, Banff*, 867-870.
- Spaai, G.W.G./Hermes, D.J. (1993), A visual display for the teaching of intonation. *CALICO (Computer Assisted Language Learning and Instruction Consortium) Journal* 10, 19-30.

- Stibbard, R. (1996), Teaching English Intonation with Visual Display of Fundamental Frequency. The Internet TESL Journal, Vol. II, No. 8, [on-line available]
<http://www.aitech.ac.jp/~iteslj/Articles/Stibbard-Intonation>
- Strik, W.A.J. (1994), Physiological control and behaviour of the voice source in the production of prosody. Dissertation, Universität Nijmegen.
- Taylor, P. (2000), Analysis and Synthesis of Intonation using the Tilt Model. Journal of the Acoustical Society of America 107 (3), 1697-1714. [Online-Version verwendet]
<http://www.cstr.ed.ac.uk/~pault/papers.html>
- 't Hart, J. (1974), Discriminability of the size of pitch movements in speech. Institute for Perception Research, Annual Progress Report 9, 56-63.
- 't Hart, J./Collier, R./Cohen, A. (1990), A perceptual study of intonation. An experimental-phonetic approach to speech melody. Cambridge: Cambridge University Press.
- Thorsen, N. (1979), Interpreting Fundamental-Frequency Tracings of Danish. *Phonetica* 38, 57-78.
- Thorsen, N. (1988), Standard Danish Intonation. Annual Report of the Institute of Phonetics, Nr. 22, Universität Kopenhagen, 1-23.
- Tillmann, H.-G./Mansell, P. (1980), *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. Stuttgart: Ernst Klett Verlag.
- Uhmann, S. (1991), *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Linguistische Arbeiten 252. Tübingen: Niemeyer.
- Umeda, N. 1982, Fundamental frequency declination is situation-dependent. *Journal of Phonetics* 10, 279-290.
- van Donzel, M. (1994), How to specify focus without using acoustic features. Institute of Phonetic Sciences, University of Amsterdam, Proceedings 18, 1-17.
- Vary, P./Heute, U./Hess, W. (1998), *Digitale Sprachsignalverarbeitung*. Stuttgart: B.G. Teubner.
- Verhoeven (1994), The discrimination of pitch movement alignment in Dutch. *Journal of Phonetics* 22, 65-85.
- von Heusinger, K. (1999), *Intonation and Information Structure*. Habilitationsschrift, Universität Konstanz.
- Wahlster, W. (2000), Mobile Speech-to-Speech Translation of Spontaneous Dialogs: An Overview of the Final VerbMobil System. In: Wahlster, W. (ed.), *VerbMobil: Foundations of Speech-to-Speech Translation*. Berlin: Springer, 3-21.
- Werner, S. (2000), *Modelle deutscher Intonation: zu Vergleichbarkeit und empirischer Relevanz von Intonationsbeschreibungen*. Dissertation, Universität Joensuu, Finnland.
- Werner, S./Keller, E. (1994). Prosodic aspects of speech. In: E. Keller, E. (ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges*. Chichester: John Wiley, 23-40.
- Wunderlich, D. (1991), Intonation and Contrast. *Journal of Semantics* 8, 239-251.