

Autoren:

Niebuhr, Oliver. Seminar für Allgemeine und Vergleichende Sprachwissenschaft, Christian-Albrechts-Universität zu Kiel (Hauptautor).

Bergherr, Julia. Institut für Phonetik und Digitale Sprachverarbeitung, Christian-Albrechts-Universität zu Kiel.

Huth, Susanne. Institut für Phonetik und Digitale Sprachverarbeitung, Christian-Albrechts-Universität zu Kiel.

Lill, Cassandra. Institut für Sozialwissenschaften, Christian-Albrechts-Universität zu Kiel.

Neuschulz, Jessica. Seminar für Allgemeine und Vergleichende Sprachwissenschaft, Christian-Albrechts-Universität zu Kiel.

Titel:

Intonationsfragen hinterfragt – Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Fragesätzen mit deklarativer Syntax.

Schlüsselwörter:

Intonationsfrage, Deklarativfrage, F0, Intonation, Prosodie, Satzmodus, Phonetik

Kontaktdaten des Hauptautors:

Prof. Dr. Oliver Niebuhr
Seminar für Allgemeine und Vergleichende Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
Leibnizstraße 10
D-24098 Kiel

Tel: +49 431 880 3314

Fax: +49 431 880 7405

Email: niebuhr@linguistik.uni-kiel.de

1. Einleitung

1.1 Zur Form und Funktion der Nachfrage

In der Literatur findet sich ein breites Spektrum an Definitionen für die kommunikative Handlung der Frage. Ein verbindendes funktionales Element zwischen den meisten Definitionen ist, dass eine Frage eine Aktion des Gesprächspartners evoziert, typischerweise in Form einer lautsprachlichen Übermittlung von Informationen. Hierbei kann es entweder um neue Informationen oder lediglich um Bestätigungen von zuvor (ggf. im gleichen Diskurs) seitens des Gesprächspartners bereits thematisierter Informationen gehen (vgl. Brown et al. 1980; Gussenhoven 1984; Grice et al. 1995; Haan 2002). Das Spektrum an übermittelten Informationen wiederum reicht vom einfachen Ausdruck der Zustimmung oder Ablehnung (oder der Unwissenheit) bis hin zur Schilderung komplexer Sachverhalte.

Parallel zu der grundlegenden funktionalen Unterscheidung zwischen der Übermittlung neuer und der Bestätigung bereits thematisierter Informationen durch den Gesprächspartner lassen sich die im Deutschen auftretenden Fragen entlang einer Reihe formaler Merkmale klassifizieren. Zu den formalen Fragetypen zählen Verberststellungsfragen, Fragewortfragen, elliptische, d.h. syntaktisch unvollständige Fragen sowie Deklarativfragen. Der Begriff 'deklarativ' bedeutet hier, dass der Fragesatz morphosyntaktisch einem Aussagesatz gleicht. Weder das syntaktische Mittel der Verberststellung noch das lexikalische Mittel der Fragewörter wird zur Indikation der Frage herangezogen. Stattdessen wird postuliert, dass die Interrogativität in Deklarativfragen intonatorisch markiert wird, weswegen dieser Fragetyp aus formaler Perspektive auch als Intonationsfrage bezeichnet wird (vgl. Hentschel und Weydt 2003; Haan 2002; Rost-Roth 2003, 2006; Näheres unter 1.2). Nachfolgend wird 'deklarativ' ebenfalls in einem morphosyntaktischen Sinn mit den Eigenschaften eines selbstständigen deutschen Aussagesatzes als Referenz gebraucht. Intonation ist im engeren Sinne der wahrgenommene Tonhöhenverlauf einer Äußerung. In der akustischen Manifestation des Sprachschalls wird vor allem durch die Grundfrequenz (F0) bestimmt (vgl. House 1990; Niebuhr 2007a). Im weiteren Sinne kann sich Intonation sowohl auf die erlebte Tonhöhe als auch auf das akustische F0 beziehen. Der vorliegende Beitrag verwendet Intonation aus drei Gründen in diesem weiteren Sinne. Erstens lassen sich die Literatur und die eigene Forschung so kohärenter und anschaulicher darstellen, zweitens wird Intonation hier nicht rein phänomenologisch, sondern mit funktionalem Bezug gebraucht, was die Integration aller intonatorischen Manifestationsebenen sinnvoll macht, und drittens ist sowohl bei den eigenen als auch bei den referierten Befunden früherer Studien davon auszugehen, dass die akustischen F0-Variationen so groß sind, dass sie sich auf jeden Fall in wahrgenommenen Tonhöhenunterschieden wiederfinden.

Im Zentrum der vorliegenden Arbeit steht die *Nachfrage*. Bezogen auf die eingangs gemachte, funktionale Unterscheidung, steht bei Nachfragen die Bestätigung von zuvor bereits thematisierten Informationen im Vordergrund. Die Ausführungen von Rost-Roth (2003, 2006) aufgreifend, können dieser Basisfunktion der Nachfrage vier unterschiedliche Sprecherintentionen zugrunde liegen: der Ausdruck von Erwartungsproblemen, die Signalisierung von Verständigungsproblemen, zum Beispiel durch Bedeutungs-, Referenz- oder auditive Probleme, der Wunsch nach Vergewisserung sowie eine Gruppe nicht primär informationaler Intentionen, zu denen Zeitgewinn oder der Ausdruck von Ironie zählen.

Auch die Überbrückung von Gesprächspausen kann dieser Gruppe von Intentionen zugerechnet werden. In jedem Fall handelt es sich gemäß Haan (2002) bei Nachfragen um einen sekundären Fragetyp. Das heißt, sie können nicht lösgelöst von ihrem jeweiligen Gesprächs- oder Situationskontext auftreten.

Werden von einer funktionalen Perspektive ausgehend die formalen Eigenschaften von Nachfragen analysiert, so zeigt sich, dass dieser Fragetyp durch einen hohen Anteil an syntaktisch elliptischen Konstruktionen charakterisiert ist oder in Form von Deklarativ- bzw. Intonationsfragen realisiert wird (vgl. Rost-Roth 2003, 2006). Letzteres ist insbesondere dann der Fall, wenn Sprecher die vom Gesprächspartner unmittelbar zuvor gemachten Aussagen zwecks einer erneuten Bestätigung wiederholen. Hierin liegt auch der Grund für die häufige Bezeichnung der Nachfrage als Echofrage. Elliptische Konstruktionen beschränken sich oft auf die zu bestätigende syntaktische Phrase. Der folgende Beispieldialog illustriert die elliptische und die deklarative Nachfrage als alternative Reaktionen des Sprechers B auf die Aussage von Sprecher A.

- A: „*Er kommt aus Kanada.*“ (Aussage mit deklarativer Syntax)
- B₁: „*Er kommt aus Kanada?*“ (Nachfrage mit deklarativer Syntax)
- B₂: „*Aus Kanada?*“ (Nachfrage mit elliptischer Syntax, die nur aus der zu bestätigenden Präpositionalphrase besteht).

1.2 Forschungshintergrund: Intonation in Fragen und Aussagen

Aussagen fallen, und Fragen steigen. Diese einfache Form-Funktions-Beziehung, die sich auf das Ende eines (vollständigen oder elliptischen) Satzes bezieht, hat die Anfänge der linguistischen und phonetischen Forschung zur Rolle der Intonation in der lautsprachlichen Kommunikation geprägt (vgl. Bolinger 1978), und sie ist außerhalb der Forschung, zum Beispiel im Fremdsprachenunterricht, auch heute noch weit verbreitet. So wird etwa in Griesbach (2000:231) konstatiert: „*Fragen erkennt man bei mündlichen Äußerungen an der Frageintonation, im schriftlichen Ausdruck am Fragezeichen am Ende des Satzes*“.

Detailliertere Studien über den Zusammenhang von Intonation und Satzmodus haben jedoch im Laufe der Zeit zu einem immer differenzierteren Bild geführt. In dem durch von Essen (1964) vorgeschlagenen System standarddeutscher Satzintonationen wird beispielsweise die Auffassung vertreten, dass die Verlaufsrichtung der finalen Intonationsbewegung in Abhängigkeit von den weiteren formalen Merkmalen des Satzes beschrieben werden kann. So zeichnen sich lexikalisch markierte Fragen durch eine fallende und Verberststellungsfragen sowie syntaktische Ellipsen- und Deklarativkonstruktionen durch eine steigende Intonation aus (vgl. auch Halliday 1967; Isačenko und Schädlich 1970; Féry 1993). Dieses Konzept ist jedoch deutlich zu eng gefasst. Zwar sagen auch Stock und Zacharias (1973:138), dass final steigende Intonationen „*tatsächlich, grammatisch gesehen, häufiger mit Fragesätzen als mit anderen Satzarten*“ auftreten und dass zudem Verberststellungsfragen häufiger durch finale Anstiege gekennzeichnet sind als lexikalisch markierte Fragen. Allerdings betonen sie ebenfalls, dass es „*keine obligatorische Verbindung*“ von steigender Intonation und Verberststellungsfrage gibt (Stock und Zacharias 1973:139). Ähnlich äußert sich Siebs (1969). Diese vorrangig auf muttersprachlichen Introspektionen basierenden Feststellungen wurden von Selting (1995), Kügler (2004) und Kohler (2004) anhand empirisch-quantitativer Analysen spontansprachlicher Dialogkorpora zu verschiedenen regionalen Varietäten des Deutschen bestätigt. Selting, Kügler und Kohler finden übereinstimmend, dass etwa 20-30% der Verberststellungsfragen mit einer final fallenden Intonation realisiert

worden sind. Kohler (2004) zeigt zudem, dass ein ähnlicher Prozentsatz (ca. 25%) der lexikalischen markierten Fragen entgegen den Behauptungen von Essens (1964) final ansteigt. Für andere Sprachen wie das Holländische und das Amerikanische Englisch finden sich ähnliche prozentuale Zusammenhänge zwischen der Richtung der finalen Intonationsbewegung auf der einen und syntaktischen bzw. lexikalischen Fragemarkierungen auf der anderen Seite (vgl. Haan 2002; Ladd 1996).

Diese empirischen Befunde haben dazu geführt, dass die finale Intonationsbewegung aus dem Zusammenspiel formaler Fragemerkmale weitgehend herausgelöst wurde. Kohler (2004) zum Beispiel schlägt aufgrund systematischer funktionaler Untersuchungen mit resynthetisierten Stimuli aus dem zuvor analysierten Datenkorpus (dem 'Kiel Corpus of Spontaneous Speech', vgl. Kohler et al. 1997) vor, dass final steigende und fallende Intonationen – sowohl im Rahmen lexikalisch als auch syntaktisch markierter Fragen – einen eigenständigen attitudinalen Bedeutungsunterschied ausdrücken. Durch Fragen mit final steigender Intonation signalisiert der Sprecher sein Interesse bzw. seine Unterordnung gegenüber dem Gesprächspartner und gibt diesem damit gleichzeitig die Freiheit, Umfang und Inhalt des Redebeitrags, einschließlich der Antwort auf die Frage, den eigenen Wünschen entsprechend zu gestalten. Gegenteiliges gilt für eine Frage mit final fallender Intonation. Ähnliche Bedeutungskonzepte für den finalen Intonationskontrast werden auch von Kügler (2004) oder Peters (2005) aufgestellt. Sie finden sich in Andeutungen sogar bereits bei Siebs (1969) sowie bei Stock und Zacharias (1973). Beispielsweise signalisiert die Frage „*Würde Ihnen das passen?*“ mit einer final fallenden Intonation, dass eine positive Antwort erwartet wird, während bei einer final steigenden Intonation alle Antwortalternativen für den Gesprächspartner offen bleiben (vgl. Kohler 2004). Wenn des Weiteren eine Frage wie „*Sind Sie Heidelbergerin?*“ mit final steigender Intonation kombiniert wird, drückt dies Interesse daran aus, auch über die spezifische Frage hinaus mehr über den Gesprächspartner zu erfahren. Insofern ist dessen Redebeitrag nicht nur hinsichtlich des Inhalts, sondern ebenfalls hinsichtlich des Umfangs keinen Beschränkungen seitens des Sprechers unterworfen (vgl. Peters 2005). Im Falle einer final fallenden Intonation wird hingegen eine kurze (positive) Antwort erwartet.

Trotz dieses neuen attitudinalen Bedeutungskonzeptes für die finale Intonationsbewegung im Rahmen lexikalisch oder syntaktisch markierter Fragen herrscht weiterhin die Meinung vor, dass für Sätze mit elliptischer und deklarativer Syntax die Intonation zum Ende hin ansteigen *muss*, damit die formale Markierung als Frage sichergestellt ist und der Redebeitrag vom Gesprächspartner entsprechend verstanden werden kann. Isačenko und Schädlich (1970:32) sagen etwa „*if an utterance contains no other syntactic or lexical cue to identify it as a question [...] then only the last [...] rising tone-switch provides the necessary information to allow the hearer to identify it as a question.*“ Ähnlich äußern sich auch Stock und Zacharias (1973:137): „*Ein Satz, der strukturell als Nichtfrage erscheint, muß mit Steigkadenz gesprochen werden, wenn er eine Fragefunktion übernehmen soll*“ (vgl. auch Lieberman 1967 und Ladd 1996 zu anderen Sprachen als dem Deutschen).

Aufbauend auf der Arbeit von Petrone (2008) zum neapolitanischen Italienisch zeigen Petrone und Niebuhr (2009) jedoch, dass auch syntaktisch deklarative Sätze im Standarddeutschen am Ende mit einer fallenden Intonation auftreten und dennoch als Fragen – genauer als Nachfragen – interpretiert werden können. Bereits Brinckmann und Benz Müller (1999) haben in ihrem Dialogkorpus Fragen nachweisen können, die trotz deklarativer Syntax mit einem Intonationsabstieg am Satzende produziert wurden. Dies

sind klare Hinweise darauf, dass der attitudinale Bedeutungsunterschied, der für final fallende und steigende Intonationen in lexikalisch und syntaktisch markierten Fragen aufgestellt wurde, auch bei Deklarativfragen nicht zugunsten der Differenzierung des Satzmodus aufgegeben wird. Doch wie werden Deklarativfragen dann gegenüber vergleichbaren Aussagen gekennzeichnet, wenn nicht über die Richtung der satzfinalen Intonation? Auf Basis ihrer Produktionsdaten identifizieren Petrone und Niebuhr (2009) in einem ersten Wahrnehmungsexperiment die *Verlaufsform* der steigend-fallenden Intonationsgipfel über den akzentuierten Silben (auch 'Tonhöhenakzente' oder 'pitch accents', vgl. Ladd 1996; Peters 2004; Niebuhr 2007a,b) als einen Indikator für den Aussage-Frage-Kontrast in Sätzen mit deklarativer Syntax. Im Falle einer Frage ist der Abstieg vom Intonationsgipfel deutlich konvex. Der Anstieg kann zudem einen konkaven Verlauf aufweisen. Das genaue Gegenteil trifft auf die Verlaufsformen der steigend-fallenden Intonationsgipfel in Aussagen zu.

Darüber hinaus finden Petrone und Niebuhr (2009), dass die Differenzierung von Frage und Aussage anhand der Verlaufsform des Intonationsgipfels nicht bzw. nicht nur im Rahmen der letzten Akzentsilbe im Satz, dem sogenannten Nukleus, stattfindet, sondern sich – wie auch im neapolitanischen Italienisch – vor allem an den davor liegenden, prä nuklearen Akzenten manifestiert (vgl. Cruttenden 1990 für die Dichotomie 'nuklear' vs. 'prä nuklear'). Abbildung 1 exemplifiziert den Befund von Petrone und Niebuhr (2009) anhand der final fallenden F0-Verläufe je einer Aussage und Nachfrage, die mit dem gleichen, syntaktisch deklarativen Wortlaut „Katharina sucht 'ne Wohnung“ (Akzentsilben unterstrichen) produziert wurden. Im direkten Vergleich der beiden F0-Verläufe ist klar zu erkennen, dass insbesondere der Abstieg vom Intonationsgipfel des ersten, prä nuklearen Tonhöhenakzentes bei der Frage klar konvex und bei der Aussage klar konkav verläuft. Darüber hinaus ist der Anstieg zum Gipfel im Falle der Aussage eher konvex.

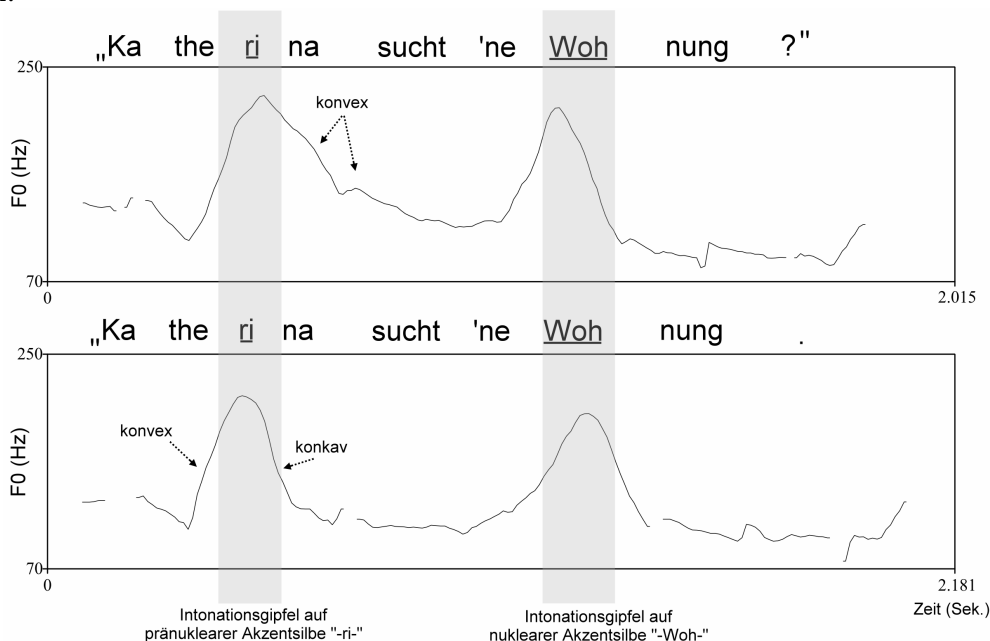


Abbildung 1: F0-Verläufe des syntaktisch deklarativen Satzes „Katharina sucht 'ne Wohnung“, produziert von einem männlichen Sprecher mit final fallender Intonation als Nachfrage (oben) und Aussage (unten). Beide Produktionen enthalten Intonationsgipfel über den grau unterlegten Akzentsilben „-ri-“ (prä nuklear) und „Woh-“ (nuklear).

Mit ihren Befunden haben Petrone und Niebuhr (2009) nicht nur die beschriebene, sukzessive Aushöhlung der simplen Gleichsetzung von Aussagen und Fragen mit satzfinal ab- und ansteigenden Intonationen zu Ende geführt. Sie haben darüber hinaus eine neue Perspektive für die Einbindung der Sprechmelodie in die charakteristischen Eigenschaften von Fragen bzw. Fragetypen eröffnet. Bislang wurde von einer zeitlichen Asymmetrie der Fragemerkmale ausgegangen. Während die lexikalische und syntaktische Markierung der Frage am *Satzanfang* stattfindet, war der Blick bezüglich der intonatorischen Markierung der Frage, sofern sie postuliert wurde, auf das *Satzende* gerichtet. Petrone und Niebuhr zeigen jedoch auf Basis der syntaktisch deklarativen Nachfrage, dass deren intonatorische Charakteristika sich nicht lokal auf das Satzende beschränken, sondern sich globaler über den Satz verteilen. An diesem Punkt setzt die vorliegende Studie die Forschung von Petrone und Niebuhr fort. Worin genau das Forschungsziel besteht, wird im nachfolgenden Abschnitt weiter ausgeführt.

1.3 Forschungsziel

Die vorliegende Studie zielt darauf ab, in akustischen und auditiven Analysen von eigens aufgenommenem, interaktionalem Sprachmaterial über die Verlaufsqualitäten der steigend-fallenden Intonationsgipfel hinaus weitere, nicht-satzfinale Eigenschaften zu bestimmen, die Aussagen und Nachfragen mit deklarativer Syntax und zum Teil gleichem Wortlaut im Standarddeutschen differenzieren. Dabei wird das Augenmerk neben der Intonation am Satzanfang auch auf globale prosodische Eigenschaften wie Rhythmus (d.h. Muster aus akzentuierten und unakzentuierten Silben), Tempo und Stimmklang gerichtet. In die Auswahl der Eigenschaften sind zwei Überlegungen eingeflossen.

Erstens sollte die Auswahl bei solchen Sprachen ansetzen, die auf nicht-satzfinale sprechmelodische Eigenschaften angewiesen sind. Entsprechend sollen die Befunde von Brinkmann und Benz Müller (1999) zum Standarddeutschen zunächst unberücksichtigt bleiben. Es gibt eine Reihe von Sprachen, in denen sowohl Fragen als auch Aussagen mit gleichgerichteten satzfinalen Intonationen verbunden sind und die zum Teil auch über keine anderen, lexikalischen oder syntaktischen Mittel zur Fragemarkierung verfügen. Wie werden in solchen Sprachen Fragen von Aussagen unterschieden? Untersuchungen zu Tonsprachen wie dem Hausa oder dem Mandarin haben gezeigt, dass die Intonation in Fragen auf einem höheren Niveau liegt, das zudem weniger schnell über den Satz hinweg absinkt als in Aussagen (vgl. Lindau 1986; Shen 1990; Zeng et al. 2004). Thorsen (1980) beschreibt für das Dänische einen ähnlichen Zusammenhang zwischen dem Satzmodus und dem Gefälle des Intonationsniveaus. Des Weiteren zeigen akustische Analysen unter anderem zum Kantonesischen, neapolitanischen Italienisch und zum Malaysischen, dass Fragen mit einer schnelleren Sprechgeschwindigkeit einhergehen als Aussagen (vgl. D'Imperio 2000; Petrone 2008; van Heuven und van Zanten 2005).

Petrone und Niebuhr (2009) haben am Beispiel der Verlaufsform von Intonationsgipfeln bereits gezeigt, dass sich intonatorische Eigenschaften von Fragen aus anderen Sprachen in Nachfragen des Standarddeutschen wiederfinden, und es ist durchaus denkbar, dass solche sprachübergreifenden Parallelen auch in puncto Intonationsniveau, -gefälle und Sprechgeschwindigkeit bestehen. In der Tat ist es eine weit verbreitete Ansicht, dass sich die sprechmelodischen Unterschiede zwischen kommunikativen Handlungen wie Frage und Aussage aus grundlegenden, biologisch-motivierten ethologischen Mustern heraus entwickelt haben, die, sofern sie nicht durch Sprachwandelprozesse verändert wurden, bis heute als sprachübergreifende Gemeinsamkeiten bestehen (vgl. Bolinger 1978; Ohala

1983; Gussenhoven 2002) und eine rudimentäre sprachübergreifende Kommunikation ermöglichen.

Die zweite Überlegung, die der Auswahl der analysierten Eigenschaften zugrundegelegen hat, war funktionaler Natur. Es geht bei der Differenzierung des Satzmodus in deklarativen Sätzen nicht um Aussage vs. Frage an sich, sondern um Aussage vs. *Nachfrage*. Wie in 1.1 umrissen, kommt aus funktionaler Sicht im Rahmen einer Nachfrage in besonders deutlicher Weise eine Verfehlung oder Unsicherheit bzw. Unterordnung des Sprechers zum Tragen. Es ist davon auszugehen, dass sich diese Konnotation in prosodischen Eigenschaften niederschlägt, durch die sich die Nachfrage von einer Aussage selbst bei vergleichbarer Morphosyntax und satzfinal fallender Intonation klar abhebt. Grice und Savino (1997) berichten etwa, dass Fragen und insbesondere Nachfragen in der Barivarietät des Italienischen stark behaucht produziert werden, was darauf zurückgeführt wurde, „*that [...] the degree of speaker confidence [...] is low.*“ (Grice und Savino 1997:32). Vor diesem Hintergrund ist es gut möglich, dass der Grad der Behauchung auch im Deutschen Aussagen von deklarativen Nachfragen unterscheidet, wobei letztere behauchter sein sollten als erstere (für 'Behauchung' vgl. 2.4.3). Des Weiteren wurde unter 1.1 erwähnt, dass syntaktischen Analysen zufolge Nachfragen nicht nur in Form von Deklarativsätzen, sondern auch in Form von syntaktischen Ellipsen vorkommen. In Einklang mit der Funktion der Nachfrage liegt der Grund für die Ellipsen in einer inhärenten Fokussierung der Nachfrage auf die zu bestätigende Information. Es ist plausibel, dass diese inhärente Fokussierung ebenfalls in solchen Nachfragen gegeben ist, die als vollständige Sätze mit deklarativer Syntax realisiert werden. Dabei könnte die inhärente Fokussierung auf die zu bestätigende Information zu einer Deakzentuierung und melodischen Einebnung der übrigen Satzbestandteile führen. Hiervon sollten vor allem pränukeare Akzente betroffen sein. Zum Beispiel ist es vorstellbar, dass der Satz „*Er kommt aus Kanada*“ (vgl. 1.1) nur als Aussage einen pränukearen Tonhöhenakzent auf „*kommt*“ aufweist, nicht jedoch als Nachfrage; „*Kanada*“, die neue bzw. zu bestätigende Kerninformation, sollte in jedem Fall einen nuklearen Tonhöhenakzent tragen. Folglich liegen in den Parametern der pränukearen Akzentdichte bzw. der in diesem Bereich gegebenen, allgemeinen melodischen Variabilität weitere potentielle prosodische Unterschiede zwischen Aussagen und Nachfragen mit deklarativer Syntax.

1.4 Hypothesen und Aufbau

Vor dem in 1.3 skizzierten, sprachübergreifenden Hintergrund und den darüber hinausgehenden funktionalen Überlegungen lassen sich die folgenden 6 Hypothesen formulieren, die in insgesamt drei sprechmelodische Phänomengruppen eingeteilt werden können: (I) Intonation und Akzent, (II) Sprechgeschwindigkeit, und (III) Stimmqualität.

Für die erste Phänomengruppe werden im Rahmen der vorliegenden Studie die folgenden Hypothesen getestet:

- (1a): Verglichen mit Aussagesätzen sind deklarative Nachfragen durch eine geringere Akzentdichte im Sinne einer kleineren Zahl pränukearer Akzente gekennzeichnet.
- (1b): Das Frequenzniveau der Intonation ist in deklarativen Nachfragen höher als in Aussagesätzen.
- (1c): Das Absinken des Intonationsniveaus verläuft in deklarativen Nachfragen langsamer als in Aussagesätzen.

- (1d): In das Intonationsniveau sind in deklarativen Nachfragen weniger starke Schwankungen eingebettet, d.h. der melodische Verlauf ist stärker eingebettet als in Aussagesätzen.

Die Hypothesen (1b)-(1d) sollen anhand des Intonationsverlaufs vom Satzanfang bis zum ersten pränukearen Akzent getestet werden. Dies geschieht aus drei Gründen: Erstens sind Messungen zum Intonationsniveau (vgl. auch 'Deklinaton', Cohen und 't Hart 1967) ohne Interferenz durch akzentbasierte Intonationsgipfel reliabler. Zweitens steht die vorgenommene Beschränkung des Messbereichs im Einklang mit der Orientierung der Studie weg von Intonationseigenschaften am Satzende und hin zu Eigenschaften des Satzanfangs. Drittens ist die Hypothese (1d) aus der inhaltlich verwandten Dreiergruppe (1b)-(1d) durch ihre gemeinsame funktionsbasierte Herleitung mit (1a, vgl. 1.3) ohnehin auf den pränukearen Bereich ausgerichtet. Durch die Beschränkungen der Messungen zu den Hypothesen (1b) und (1c) auf den gleichen Bereich können insgesamt kohärentere Aussagen zur Intonation getroffen werden.

Für die beiden Phänomengruppen (II) und (III) wurden die folgenden beiden Hypothesen überprüft:

- (2) Deklarative Nachfragen werden mit einer schnelleren Sprechgeschwindigkeit realisiert als Aussagesätze.
- (3) Deklarative Nachfragen werden mit einer behauchteren Sprechstimme realisiert als Aussagesätze.

In der Überprüfung der aufgestellten Hypothesen mittels akustischer Messungen sind Wechselwirkungen zwischen den einzelnen Parametern möglich. Beispielsweise ist es denkbar, dass Variationen in der pränukearen Akzentdichte die Sprechgeschwindigkeit beeinflussen oder umgekehrt. Die Stimmqualität könnte mit dem Frequenzniveau der Intonation korrelieren, und auch innerhalb der Phänomengruppe (I) könnten Parametervariationen miteinander interagieren. Doch auch wenn solche Wechselwirkungen möglich sind, sind sie prinzipiell weder zwingend noch in ihren Konsequenzen für die Hypothesen einheitlich und vorhersehbar. Beispielsweise bezieht sich die Messung der Sprechgeschwindigkeit auf den gesamten Satz, während die Akzentdichte nur die pränukearen Akzente abdeckt, was Wechselwirkungen bereits a priori schwächen sollte. Zudem ist, verglichen mit nuklearen Akzenten, die sprachgeschwindigkeitsrelevante Silbendauer gar nicht oder nur gering in die Bildung pränukearer Akzente involviert, sodass etwaige akzentbedingte Variationen in intra- und interindividuellen Variationen untergehen oder leicht von tatsächlichen Variationen in der Sprechgeschwindigkeit überdeckt werden können. Einflüsse des Frequenzniveaus der Intonation auf die Stimmqualität wurden zwar sprachübergreifend gefunden, fielen jedoch interindividuell so heterogen aus, dass davon auszugehen ist, dass sie sich über eine größere Personenzahl hinweg nur als erhöhtes statistisches Rauschen manifestieren (vgl. Swerts und Veldhuis 2001). Über die Hypothesen hinausgehende sprachproduktorisches Kopplungen zwischen der Stimmqualität auf der einen und der Akzentuierung sowie der Vokalqualität auf der anderen Seite wurden im Experiment kontrolliert (vgl. 2.4.3). Im Falle der Phänomengruppe (I) kann eine geringere pränukeare Akzentdichte damit einhergehen, dass ein längeres Intonationsstück vermessen wird, wodurch lautlich bedingte Grundschwankungen in der Intonation weniger stark in der Messung zum Tragen kommen und das gemessene Schwankungsniveau so niedriger ausfällt. Andererseits könnte das längere Intonationsstück über die lautlichen Grundschwankungen hinaus auch größere, intendierte Intonationsbewegungen aufweisen, die das Schwankungsniveau insgesamt

höher ausfallen lassen als bei einem kürzeren Stück. Bezüglich der Länge Intonationsstücks ist bei beiden Möglichkeiten ferner zu berücksichtigen, dass die Antwort auf die Frage, *wo* im Satz der erste prä nukleare Tonhöhenakzent auftritt, selbst bei sinkender Akzentdichte variabel ist. Letztlich ist es nicht zwangsläufig der Fall, dass ein geringeres Absinken des Intonationsniveaus zu einem insgesamt höheren Frequenzniveau der Intonation führt. Das Frequenzniveau könnte ebenso gut bei langsamerem Abfall insgesamt tiefer liegen. Zusammengefasst ist festzustellen, dass trotz möglicher, aber nicht zwingender Wechselwirkungen zwischen den Parametern der einzelnen Hypothesen letztere unabhängig voneinander bestätigt oder verworfen werden können. Dennoch sollte das Gesamtbild der Ergebnisse in die Interpretation der Hypothesen einfließen.

Für den Aufbau der Studie wurde des Weiteren der folgende Aspekt berücksichtigt. Die drei Phänomenbereiche Intonation / Akzent, Sprechgeschwindigkeit und Stimmqualität fallen im Deutschen nicht nur für Aussagen und Nachfragen mit deklarativer Syntax potentiell unterschiedlich aus. Sie variieren auch nachweislich in Abhängigkeit von Emphase. Beispielsweise zeigen Baumann et al. (2006, 2007) sowie Kügler (2008), dass Emphase im Sinne von engem bzw. kontrastivem Fokus im Standarddeutschen mit Längung – mindestens des betreffenden Wortes – sowie mit größerer melodischer Variabilität im Bereich der emphatisch akzentuierten Silbe und melodischer Einebnung außerhalb (insb. nach) der emphatisch akzentuierten Silbe verbunden ist. Niebuhr (2010) weist zudem darauf hin, dass auch Stimmqualitätseigenschaften an der Signalisierung von Emphase sowie der Differenzierung von Emphasetypen im Deutschen beteiligt sind.

Vor diesem Hintergrund wird Emphase neben dem Satzmodus als zusätzliche Variable in die Studie einbezogen. Dies betrifft sowohl die Erhebung des Sprachmaterials als auch dessen akustische und auditive Analysen. Folglich ist die Studie, wie in Tabelle 1 dargestellt, auf vier Satzbedingungen aufgebaut, die sich aus der Kreuzung der beiden jeweils zweistufigen Variablen *Satzmodus* und *Emphase* ergeben. Durch diesen Aufbau kann die Interaktion der analysierten Parameter bezüglich der Variablen *Satzmodus* und *Emphase* untersucht werden. Neben den dabei gewonnenen Einblicken in diesen bislang wenig erforschten Problembereich selbst, ist ein entscheidender Vorteil dieses Ansatzes, dass für die zu testenden Hypothesen differenziertere Aussagen getroffen werden können. So ist es beispielsweise denkbar, dass sachliche Nachfragen zwar gemäß den Hypothesen (2) und (3) tatsächlich schneller (vgl. die Zeitachsen in Abb.1) und behauchter realisiert werden als sachliche Aussagen, dass im Vergleich zu emphatischen Aussagen diese Unterschiede jedoch verschwinden oder sogar ins Gegenteil verkehrt werden. Abschließend ist darauf hinzuweisen, dass als Untersuchungsgegenstand nur Aussagen mit final fallender und Nachfragen mit final steigender Intonation herangezogen werden. Es ist nicht vorherzusehen, welche Einflüsse eine Variation der finalen Intonationsbewegung innerhalb der Satzbedingungen auf die zu analysierenden Variablen hat. Daher sollen in dieser ersten Studie nur Fälle mit den für Aussage- und deklarative Nachfragesätze prototypisch erscheinenden finalen Intonationsbewegungen untersucht werden, die zudem leichter zu elizitieren sind. Weitere Einzelheiten der Methode werden im folgenden Kapitel erläutert.

		Satzmodus	
Emphase		Sachliche Aussage	Sachliche Frage
		Emphatische Aussage	Emphatische Frage

Tabelle 1: Der viergliedrige Aufbau der Studie, resultierend aus den beiden gekreuzten, zweistufigen Variablen Satzmodus und Emphase.

2. Methode

2.1 Konstruktion der Monologe und Dialoge

In der Dialektologieforschung werden die Methoden der Datenerhebung und der Wert der mit ihnen gewonnenen Einblicke schon seit langem kontrovers diskutiert (vgl. Wagener 1986). Dies ist vor allem dem Umstand geschuldet, dass sich das lautsprachliche Ausdrucksverhalten (u.a. hinsichtlich nicht-standardsprachlicher Formen) in informellen Alltagsgesprächen von dem in formellen und gezwungenen Laborsituationen unterscheidet. Man sollte das (lautsprachliche) Verhalten eines Individuums in einem sozial und situativ „alltäglichen“ Kontext beobachten; aber ein solcher Kontext ist mit den Mitteln und Zielen der Beobachtung weitgehend unvereinbar. Wagener (1986) schlägt vor, dieses auch aus anderen empirischen Disziplinen bekannte „Beobachter-Paradoxon“ in der Weise aufzulösen, dass es für sprachliches Ausdrucksverhalten nicht eine einzige, optimale „Natürlichkeit“ gibt. Folglich erheben auch die in der Dialektologie häufig durchgeführten 'Tonbandinterviews' valide Daten, die allerdings nur einen Ausschnitt des natürlichen Sprachverhaltens des jeweiligen Individuums wiedergeben. Was genau die einzelnen Ausschnitte vom Gesamtverhalten der Individuen abdecken, muss für jeden Einzelfall (vergleichend) diskutiert werden.

Den Ansichten von Wagener (1986) ist grundsätzlich zuzustimmen. Allerdings ist der Ausgangspunkt von Datenerhebungen in der phonetischen Forschung nicht ganz derselbe wie in der Dialektologie. Im Vordergrund stehen nicht allein die Probleme der Gewinnung und Repräsentativität von Sprachdaten. Vielmehr werden, wie in der vorliegenden Studie, Sprachdaten eigens zur Überprüfung spezifischer Hypothesen erhoben. Dabei liegt das Forschungsinteresse zudem typischerweise in der Standardvarietät einer Sprache bzw. im Vergleich zwischen Standardvarietäten. Aus beiden Gründen basieren phonetische Studien in der Regel auf der Elizitation vorab verschrifteter und dann in Form kurzer Monologe gelesener Einzelsätze. Sie bieten einen frei wählbaren (insbesondere hinreichend großen) Stichprobenumfang und überdies die größtmögliche Kontrolle bekannter Einflussfaktoren auf die zu untersuchenden Phänomene. Intensive Diskussionen über die externe Validität und damit über die Generalisierbarkeit der mit gelesenen Einzelsätzen gewonnenen Erkenntnisse sind in der phonetischen Forschung noch vergleichsweise jung und rar (z.B. Anderson et al. 1991; Brink et al. 1998; Iriando et al. 2007), ebenso wie neue methodische Ansätze und genauere Einblicke in die phonetischen Unterschiede zwischen Lese- und Spontansprache. Den entscheidenden Anstoß zu diesen Entwicklungen gab der Forschungsfortschritt und die damit einhergehende Ausweitung und gleichzeitige Verfeinerung der Forschungsziele. Das Ausmaß und die Vielfalt der Muster in der Reduktion von Wortaussprachen etwa sind über gelesene Einzelsätze kaum untersuchbar. Gleiches gilt für die formale und funktionale Vielfalt intonatorischer und emphatischer Kategorien; und auch im Falle etwaiger Unterschiede zwischen den prosodischen Details von Aussagen und deklarativen Nachfragen ist die Erhebung gelesener Einzelsätze ein

zweifelhafter Ansatz. Warum beispielsweise sollte ein Sprecher zur Disambiguierung beider Bedingungen einen artikulatorischen und/oder phonatorischen Mehraufwand betreiben, wenn weder ein kommunikativer Rahmen noch ein Kommunikationspartner vorhanden sind?

Vor diesem Hintergrund bedient sich die vorliegende Studie eines zweiteiligen Kompromisses zwischen gelesenen Einzelsätzen und Spontansprache, der zwar anders als dialektologische Interviews weiterhin von kontrollierbaren Textgrundlagen ausgeht, diese jedoch in unterschiedlichem Ausmaß von formeller und ungerichteter Kommunikation in einer einschüchternden Labor- bzw. Aufnahmesituation abhebt. Obwohl durch den Elizitationsansatz von Kohler und Niebuhr (2007) inspiriert, geht die Methode dieser Studie darüber hinaus. Auf diese Weise können die Erfahrungen aus dieser Studie für methodische Fortschritte in der phonetischen und auch in der dialektologischen Forschung von Interesse sein.

Der erste Teil des methodischen Kompromisses bestand in der Aufnahme von sowohl Monolog- als auch Dialogmaterial. Im Rahmen der Monologe wurde der konservative Weg gelesener Einzelsätze beschritten. Das Monologmaterial bestand aus vier Zielsätzen, die in ähnlicher Weise auch in die Dialoge eingebettet vorkamen und für die im Rahmen einer sachlichen Aussage jeweils zwei Akzentuierungen innerhalb der insgesamt 5 bzw. 7 Silben erwartet werden konnten. Die vier Zielsätze sind nachfolgend aufgeführt (erwartete Akzentsilben unterstrichen).

- „Du wohnst in Mailand“
- „Du wohnst in Nepal“
- „Er arbeitet als Sänger“
- „Er arbeitet als Schneider“

Alle vier Sätze verfügten über eine deklarative Syntax mit jeweils drei Konstituenten; einem nominalen Subjekt (Personalpronomen), einem Vollverb (im Präsens, Indikativ) und einem präpositionalem Objekt. Letzteres endete auf ein Substantiv, in dem die initiale Silbe als akzentuiert produziert erwartet werden konnte und das eine weitere unakzentuierte Silbe bis zum Satzende bereitstellte. Die erwarteten akzentuierten Silben selbst besaßen einen vollständig stimmhaften Reim, in dessen Nukleus etwa gleich häufig offene und nicht-offene Vokalqualitäten vorkamen. Die Anzahl offener und nicht-offener Vokale wurde auch über die Zielsätze insgesamt vergleichbar gehalten.

Jeder der vier obigen Sätze war auf konkrete Instruktionen hin nacheinander in den zwei kreuzkombinierten *Satzmodus*- und *Emphase*-Bedingungen zu produzieren, d.h. als sachliche Aussage (AS), als sachliche Frage (FS) sowie als emphatische Aussage (AE) und emphatische Frage (FE). Es wurden also von jedem Probanden insgesamt 4x4=16 Einzelsätze produziert. In diesem Zusammenhang stützen sich die Akzentuierungserwartungen für die satzfinalen Substantive „Mailand“, „Nepal“, „Sänger“ und „Schneider“ auch darauf, dass es sich hierbei um die in den Aussagen festzustellende (Rhema) bzw. die in den Fragen zu klärende Kerninformation handelt. Emphase wurde für die Probanden im Falle der Fragen als „*ungläubiges Erstaunen*“ und im Falle der Aussagen als „*vehementes Zurückweisen*“ definiert. Insofern handelt es sich um eine Art der Emphase, die sich an das in der Fachliteratur gebräuchliche Konzept des kontrastiven bzw. engen Fokus anlehnt (vgl. Baumann et al. 2006, 2007; Kügler 2008).

Im Dialogmaterial wurden die Zielsätze in einem Gesprächskontext eingebettet produziert, was per se bereits eine Bedingung darstellt, die näher an spontaner Alltagssprache ist. Um überdies zu vermeiden, dass die Zielsätze innerhalb des Dialogmaterials als solche erkannt und von den Probanden mit besonderer Aufmerksamkeit gesprochen werden konnten, wurden die Dialogzielsätze anders als die Monologsätze über die kreuzkombinierten *Satzmodus*- und *Emphase*-Bedingungen hinweg im Wortlaut variiert. Lediglich die grundlegenden Eigenschaften der Monologsätze wurden beibehalten. Hierzu gehören kurze Zielsätze aus 6-10 Silben, die eine deklarative Syntax mit den Elementen Subjekt, Verb und Objekt sowie ein satzfinale, zweisilbiges Substantiv mit erwarteter Akzentuierung auf der initialen Silbe aufwies. In den Nuklei dieser initialen Silben kamen, wie auch im Rest Zielsatzsilben, innerhalb jeder der vier *Satzmodus*- und *Emphase*-Bedingungen etwa gleich häufig offene und nicht-offene Vokalqualitäten vor. Auch die unterschiedlichen Zielsatzlängen, in Form der Anzahl der Silben, wurden über die vier *Satzmodus*- und *Emphase*-Bedingungen hinweg annähernd ausbalanciert.

Um den variableren Produktionsbedingungen und der Ausbalancierung von Zielsatzeigenschaften im Dialogmaterial Rechnung zu tragen, wurden insgesamt 8 Zielsätze für jede der vier *Satzmodus*- und *Emphase*-Bedingungen entworfen. Somit umfasste das Dialogmaterial $4 \times 8 = 32$ Zielsätze, doppelt so viele wie im Falle der Monologe. Die 32 Zielsätze wurden über zwei Dialogtexte verteilt.

Der zweite Teil des Kompromisses bei der Gewinnung des Sprachmaterials zu dieser Studie bestand darin, die Produktion der Monologe und Dialoge so weit wie möglich in einem zwanglosen und spontansprachlichen Alltagsrahmen stattfinden zu lassen, wie er auch außerhalb des Labors gegeben ist. Neben der Ergänzung der Monologe durch Dialoge, die sich Reisen, Beziehungen, Prüfungen und anderen Alltagserlebnissen widmen, war eine zentrale Maßnahme hierbei die Arbeit mit Probandenpaaren. Auf diese Weise wurden die Redebeiträge in den Dialogen sowie die Einzelsätze der Monologe immer direkt an einen Gesprächs- bzw. Ansprechpartner gerichtet. Zusätzlich waren die eingeladenen Probandenpaare durch eine langjährige Freundschaft eng verbunden. Dieses gewohnte Element schaffte eine informelle Atmosphäre und gemeinsame Kommunikationsgrundlage in der Aufnahmesituation. Ferner stammten alle Probanden auch aus dem Freundeskreis der Autoren und waren daher ebenfalls mit den Versuchsleitern vertraut.

Des Weiteren erhielten die Probanden, anders als im Rahmen der Monologe, für die Produktion der Dialoge keine expliziten metasprachlichen Instruktionen bezüglich der Existenz bestimmter Zielsätze oder der Art und Weise ihrer Realisierung. Die Produktion der Zielsätze gemäß den vier *Satzmodus*- und *Emphase*-Bedingungen wurde stattdessen über die Schaffung entsprechender semantisch-pragmatischer Kontexte gesteuert. Diese Kontexte wurden durch die Zeichensetzung in den Texten unterstützt. So waren sachliche Fragen und Aussagen am Satzende durch $\langle ? \rangle$ bzw. $\langle . \rangle$ markiert. Emphatische Fragen und Aussagen endeten in doppelten Frage- bzw. Ausrufezeichen ($\langle ?? \rangle$ und $\langle !! \rangle$). Die Kontextualisierung war ferner darauf ausgerichtet, die Fragen mit satzfinal steigender Intonation zu elizitieren. Die kreierte Kontexte für die Elizitation von *Emphase* zielten auf „*ungläubiges Erstaunen*“ bzw. „*vehementes Zurückweisen*“ ab.

Der für diese Studie gewählte Kompromiss aus den mit expliziten Instruktionen gelesenen, an einen Gesprächspartner gerichteten Einzelsätzen auf der einen Seite und den allein kontextgesteuert variierten, in Alltagsdialoge eingebetteten Zielsätzen auf der anderen Seite wurde bereits in früheren Studien von Kohler und Niebuhr (2007) sowie Niebuhr (2009, 2010) zur Erhebung von Sprachdaten für akustische Analysen angewendet und konnte überzeugend spontansprachlich klingendes Material von den befreundeten Gesprächspartnern elizitieren. Allerdings waren das Monolog- und Dialogmaterial in den Studien von Kohler und Niebuhr (2007) sowie Niebuhr (2010) Bestandteil der gleichen Stichprobe. Im Falle der vorliegenden Studie hingegen repräsentieren die Monolog- und Dialogmaterialien zwei separate, aber parallele Stichproben. Dies hat zweierlei Vorteile. Einerseits kann über den Vergleich mit den gewonnenen Monologdaten, bei denen eine explizite Instruktion hinsichtlich der experimentellen Bedingungen stattgefunden hat, die Wirksamkeit der Kontextsteuerung in den Dialogdaten evaluiert werden. Andererseits kann der Vergleich der Monolog- und Dialogdaten Hinweise darauf ergeben, inwieweit die Dialogdaten aufgrund der Kontextualisierung und Interaktion näher an realer Spontansprache sind als die Monologdaten. Sollten Anzeichen dafür gefunden werden, dass sich die gegenüber den Monologen deutlich (zeit-)aufwendigere Erhebung von Dialogdaten bezüglich spontansprachlicher Eigenschaften lohnt, wäre dies für zukünftige akustisch-phonetische Studien von großer Bedeutung.

2.2 Durchführung der Aufnahme

2.2.1 Versuchspersonen

Die 8 Teilnehmer, die für die Aufnahmen ausgewählt wurden, waren durchgehend weiblichen Geschlechts und zwischen 23 und 26 Jahre alt. Die Beschränkung auf Teilnehmerinnen geschah vor dem Hintergrund, geschlechtsspezifische Stimmqualitäten, Intonationsniveaus und Sprechgeschwindigkeiten für die akustische Analyse zu kontrollieren (Frauen haben typischerweise eine höhere und behauchtere Stimme sowie eine langsamere Sprechgeschwindigkeit als Männer, vgl. Klatt und Klatt 1990; van Borsel und de Maesschalk 2008; Simpson 1998, 2009). Alle 8 Teilnehmerinnen sind in Schleswig-Holstein geboren worden und haben bis zum Zeitpunkt der Aufnahme nicht in anderen Bundesländern gelebt. Sie wohnten zum Zeitpunkt der Aufnahme alle in Kiel, haben aber auch zuvor ausschließlich in größeren Städten mit weit über 10.000 Einwohnern in Schleswig-Holstein gewohnt (d.h. 3x Kiel, 2x Neumünster, 1x Plön, 1x Rendsburg, 1x Bad Segeberg). Die Teilnehmerinnen sind in Familien mit beiden Elternteilen und mindestens einem Bruder bzw. einer Schwester aufgewachsen, haben die Schule mit dem Abitur abgeschlossen und übten zum Aufnahmezeitpunkt entweder bereits einen Beruf aus oder befanden sich in einem höheren (ökonomischen, juristischen oder pädagogischen) Studienfachsemester. Alle Teilnehmerinnen gaben an, Plattdeutsch (in Anbetracht ihrer Herkunft z.B. Holsteinisch) zwar in engen Grenzen zu verstehen, damit jedoch nur sehr selten in Kontakt gekommen zu sein. Folglich gebrauchen sie es im Alltag weder aktiv, noch können sie es – von einigen Phraseologismen abgesehen – sprechen. Keine der Teilnehmerinnen hatte Erfahrung mit Sprachaufnahmen in einem Tonstudio.

Der weitgehend einheitliche regionale, soziale, kulturelle und linguistische Hintergrund der 8 Teilnehmerinnen sollte dazu beitragen, interindividuelle Einflussfaktoren in die akustischen Analysen zu minimieren. Bezüglich des regionalen und linguistischen Hintergrunds können alle Teilnehmerinnen zudem als norddeutsche Standardsprecherinnen klassifiziert werden.

2.2.2 Aufnahme der Monologe und Dialoge

Die Aufnahmen fanden in einer schallbehandelten Kabine des Instituts für Phonetik und Digitale Sprachverarbeitung (IPDS) der Universität Kiel statt. Vor jeder der beiden Sprecherinnen eines Paares, die mit direktem Sichtkontakt einander gegenüber saßen, wurde ein eigenes Mikrofon platziert. Die Monologe und Dialoge wurden mit einer 96kHz Abtastrate sowie einer 48bit Amplitudenauflösung digital im 'wav'-Format aufgenommen.

Im Vorfeld wurden den Teilnehmerinnen die Monologe und Dialoge, die auf separate DIN A4 Zettel gedruckt wurden, von den Versuchsleitern (den Autoren) ausgehändigt. Mit der Zielsetzung, dass die Monologe und Dialoge „natürlich und spontan gesprochen“ klingen sollten, wurden den Teilnehmerinnen wie in den Vorbildstudien von Kohler und Niebuhr (2007) sowie Niebuhr (2009, 2010) anschließend zirka 30 Minuten Zeit zum laut Lesen und Trainieren der Texte eingeräumt. In diesem Zusammenhang wurde ihnen ebenfalls mitgeteilt, dass sie den Wortlaut der Dialogtexte durch das Hinzufügen, Weglassen oder Austauschen einzelner Konstituenten (sowie durch das Integrieren von Ausrufen und Ähnlichem) ihrer persönlichen Ausdrucksweise anpassen konnten.

Die Aufnahme selbst bestand aus zwei Teilen. Zunächst wurden die Dialoge insgesamt zweimal produziert, wobei die Teilnehmerinnen am Ende des ersten Durchgangs die Sprecherrollen tauschten. Dann hat jede der beiden Teilnehmerinnen im Anschluss an die Nennung und Erläuterung der vier *Satzmodus*- und *Emphase*-Bedingungen das Monologmaterial gesprochen. Um Imitationen zu vermeiden, wurden die erläuterten Bedingungen von den Versuchsleitern nicht mittels Beispielen illustriert. Wie in 2.1 beschrieben, war die jeweils passive Teilnehmerin bei der Produktion der Monologe ebenfalls anwesend, und die Monologe wurden in Richtung auf den Gesprächspartner produziert, um die kommunikative Situation weiterhin aufrecht zu erhalten. Die Monologe wurden bewusst den Dialogproduktionen nachgestellt, mit dem Ziel, dass die Teilnehmerinnen die Sprechweise der Dialoge, die einen natürlicheren und spontansprachlicheren Rahmen boten, in den Monologen übernehmen. Darüber hinaus wurde mit der Dialog-Monolog-Reihenfolge ausgeschlossen, dass die konkreten funktionalen Instruktionen, die den Einzelsatzproduktionen im Rahmen der Monologe zugrundelagen, die kontextgesteuerten Produktionen der Zielsätze in den Dialogen beeinflussen konnten. Einschließlich der Instruktions- und Trainingsphasen dauerte eine komplette Aufnahmesitzung etwa 70-90 Minuten.

2.3 Annotation der Monologe und Dialoge

Die aufgenommenen Monologe und Dialoge bildeten einen Sprachdatenkorpus von annähernd einer Stunde. Innerhalb dieses Datenkorpus wurde von den Autoren dieses Beitrags, unter der Leitung des ausgebildeten Phonetikers Oliver Niebuhr, eine teilweise segmentelle und prosodische Annotation vorgenommen. Die Annotation konzentrierte sich auf die Zielsätze und wurde auf der Grundlage der *TextGrid*-Funktion in *Praat* durchgeführt (vgl. Boersma 2001). Abbildung 2 veranschaulicht die Annotation anhand einer Bildschirmdarstellung des Zielsatzes „*Er arbeitet als Schneider*“, der von Sprecherin IHE als emphatische Aussage im Rahmen der Monologe produziert wurde.

Für die segmentelle Annotation wurde das SAMPA Zeicheninventar herangezogen (vgl. Wells 1997). Segmentgrenzen wurden, unterstützt durch den Gehörseindruck, auf der Grundlage von Oszillogramm und FFT-Breitbandspektrogramm bestimmt. Die SAMPA-

Zeichen wurden in einer phonetischen (d.h., nicht phonematischen) Interpretation verwendet. Konsonanten und Vokale wurden zusätzlich durch <C_> und <V_> eingeleitet; <aV_> weist auf akzentuierte Vokale hin.

Die prosodische Annotation, die in einem separaten *TextGrid* stattfand, umfasste die Markierung von akzentuierten Silben und den mit ihnen assoziierten, phonologischen Intonationskonturen (Tonhöhenakzente). Die Annotation erfolgte auf Basis von PROLAB (vgl. Kohler 1997), dem prosodischen Symbolinventar des Kieler Intonationsmodells KIM (vgl. Kohler 2006). Der Komplexität und bekannten Subjektivität der Aufgabe Rechnung tragend (vgl. Syrdal und McGory 2000) wurde sowohl die Entscheidung, welche Silben akzentuiert waren, als auch die Entscheidung, welche Intonationskonturen mit diesen Akzentuierungen einhergingen, von mindestens zwei Versuchsleitern nach dem Verfahren des *interobserver agreement* durchgeführt. Für die Bestimmung der Akzentsilben wurde ausschließlich der Gehörseindruck herangezogen, während in die Bestimmung der Intonationskontur auch die F0-Analyse in *Praat* einfluss. Der in Abbildung 2 dargestellte Zielsatz enthält zum Beispiel zwei Akzentsilben, „ar-“ und „Schnei-“, die jeweils mit mittleren Intonationsgipfeln und den Akzentstärken <2> (standard) bzw. <3> (emphatisch) realisiert wurden (vgl. Kohler 2006).

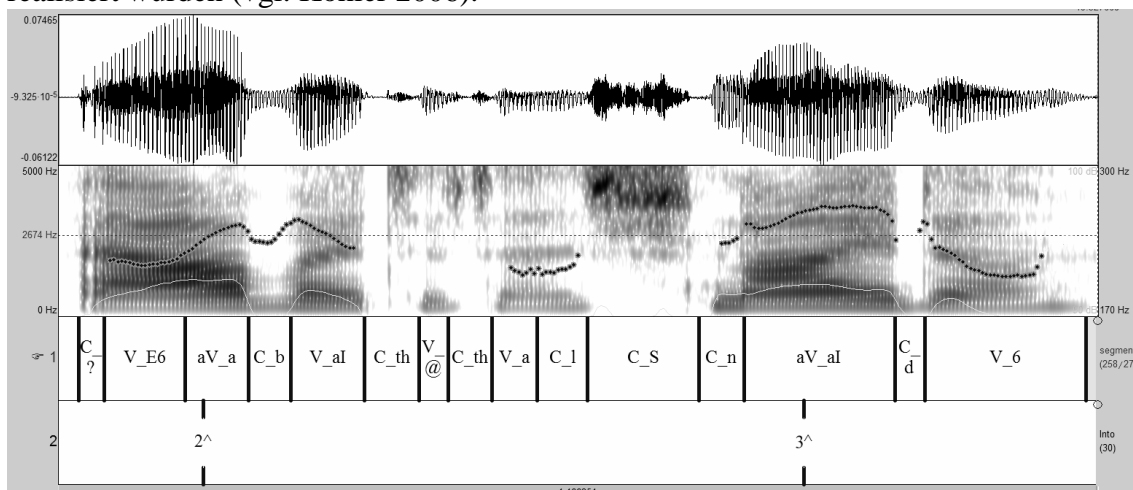


Abbildung 2: Bildschirmansicht (in Graustufen umgewandelt) des annotierten Zielsatzes „Er arbeitet als Schneider“ (Akzentsilben unterstrichen). Von oben nach unten sind folgende Fenster dargestellt: Oszillogramm, FFT-Breitbandspektrogramm (0-5000Hz) mit darübergelegter F0-Analyse (dunkelgraue Linie, 170-350Hz), segmentell-phonetische Annotation auf SAMPA-Basis, prosodische Annotation anhand von PROLAB. Der Zielsatz wurde als emphatische Aussage im Monologmaterial von IHE produziert.

2.4 Messungen

2.4.1 Intonation

Für die Dialoge wurden anhand der prosodischen Annotation zunächst pro Zielsatz zwei silbenbasierte und daher ganzzahlige Distanzwerte ermittelt: Die Silbenzahl zwischen dem Satzanfang und der ersten akzentuierten Silbe sowie die Silbenzahl zwischen der letzten prä nuklearen Akzentsilbe und der nuklearen Akzentsilbe. Beide Distanzmaße zielen von verschiedenen Ansatzpunkten her kommend darauf ab, die Anzahl der prä nuklearen Akzente und damit letztlich Unterschiede in der Akzentdichte zwischen den Aussage- und Fragesätzen aufzudecken. Da alle Zielsätze dieser Untersuchung eine neue Information

entweder direkt vermitteln oder kontrastiv kommentieren, kann davon ausgegangen werden, dass es in jedem Fall mindestens einen nuklearen Satzakzent geben wird, der sich auf dem satzfinalen Substantiv, das diese neue/kontrastierte Information betrifft, manifestiert. Da im Falle der Monologe die Zielsätze über alle experimentellen Bedingungen hinweg konstant waren, wurde die Akzentdichte einfach über die Gesamtzahl der Akzentsilben pro Zielsatz bestimmt. Dies geschah ebenfalls anhand der prosodischen Annotation.

Was die ergänzende akustische F0-Analyse anlangt, so wurden in dem unakzentuierten Abschnitt zwischen dem Satzanfang und der ersten Akzentsilbe des Satzes drei Verlaufseigenschaften bestimmt: das durchschnittliche F0-Niveau (in Hz) sowie mit Blick auf die Intonationsvariabilität die F0-Standardabweichung der Messwerte (in Hz) und die durchschnittliche Steilheit des F0-Gefälles. Letztgenannter Wert wurde zwecks der Vergleichbarkeit mit Angaben früherer Studien in Halbtonschritten pro Sekunde (HTS/Sek.) berechnet. Für die ersten beiden akustischen Parameter erlaubte die Beschränkung auf weibliche Probanden eine Messung in Hertz. Alle Messungen wurden für die zuvor entsprechend isolierten Zielsatzabschnitte anhand von in *Praat* implementierten Analysealgorithmen durchgeführt. Die F0-Analyse selbst geschah über ein Autokorrelationsverfahren mit den Standardeinstellungen in *Praat*. Die Ausgaben der Analyse wurden manuell bezüglich etwaiger Analysefehler (stimmhaft/stimmlos-Klassifikationen, Oktavfehler, etc.) durchgesehen und korrigiert.

2.4.2 Sprechgeschwindigkeit

Zur Schätzung der Sprechgeschwindigkeit wurde für jeden Zielsatz die Silbenrate, d.h. die Anzahl der Silben pro Sekunde, berechnet. Das Maß hat sich in zahlreichen Studien als reliabel und perzeptionsnah erwiesen (Bezooijen 1984; Roach 1998). Dabei bildete die Dauer des Zielsatzes, die sich durch die segmentelle Annotation ergab, den zeitlichen Bezugsrahmen. Die Anzahl der Silben wurde dann aufgrund der vollen (d.h. kanonischen) Formen der im jeweiligen Zielsatz enthaltenen Wörter bestimmt. Insofern flossen Elisionen und andere Reduktionen, die bei der tatsächlichen Realisierung der Wortsequenzen in den Zielsätzen zu einer Verringerung der Silbenzahl führen können, als Erhöhung der Silbenrate in die Messung ein. Dies lässt sich am Beispiel des Satzes „*Hast Du einen Moment Zeit*“ aus Kohler (1998) verdeutlichen. Der Satz besteht aus 7 Silben, die sich über einen Zeitraum von etwa 2,5 Sekunden erstrecken. Bei einer alltagssprachlichen Realisierung kann dieser Satz zu der viersilbigen Form [has mom'en ts'art^h] mit einer Gesamtdauer von etwas über einer Sekunde reduziert werden. In der vorliegenden Studie wäre für die Silbenrate dieses Satzes die verkürzte Gesamtdauer zu allen 7 Silben in Bezug gesetzt worden. Eventuelle Pausen in den Zielsätzen wurden vor der Berechnung der Silbenzahl pro Sekunde von der Satzdauer subtrahiert und blieben folglich bei der Messung unberücksichtigt.

Als Ergänzung zur Anzahl der Silben pro Sekunde wurde für jeden Zielsatz ein ΔC -Wert berechnet. Er repräsentiert die Standardabweichung der Konsonantendauern innerhalb des Zielsatzes und ließ sich ebenfalls aus der segmentellen Annotation ableiten. Das Maß ΔC wurde von Ramus et al. (1999) eigentlich zur Vermessung und Klassifikation der Sprechrhythmen verschiedener Sprachen vorgeschlagen. Das Maß hat sich jedoch in der empirischen Anwendung als sehr anfällig für Variationen in der Sprechgeschwindigkeit herausgestellt. Entsprechend konstatieren Ramus et al. (1999:117) “[t]he usefulness of

variables such as [...] ΔC may well be limited to corpora where speech rate is strictly controlled” (vgl. auch Barry et al. 2003; Dellwo und Wagner 2003; Dellwo et al. 2006). Die sprachübergreifend in Rhythmusstudien beobachtete sensible Reaktion von ΔC auf Variationen in der Sprechgeschwindigkeit wird in der gegenwärtigen Studie zum Anlass genommen, diese relativ einfach zu berechnende Größe gezielt zur Messung der Sprechgeschwindigkeit einzusetzen und die Befunde mit denen der Silbenrate zu vergleichen. In diesem Zusammenhang ist darauf hinzuweisen, dass sowohl die Silbenzahl pro Sekunde als auch ΔC auf der Ebene des akustischen Signals angesiedelt sind und nur in gewissem Umfang mit der wahrgenommenen Sprechgeschwindigkeit korrelieren (vgl. Pfitzinger 1999; Dellwo et al. 2006; Koreman 2006). Selbst innerhalb der akustischen Messungen können die Ergebnisse je Bezugsgrößen sehr unterschiedlich ausfallen (vgl. Roach 1998).

2.4.3 Stimmqualität

Die Stimmqualität wurde über die Amplitudendifferenz in dB zwischen den ersten beiden Harmonischen (H1-H2) des Sprachsignals bestimmt. Diese H1-H2-Differenz hat sich sprachübergreifend als reliables Maß insbesondere für den Grad der Behauchung der Sprechstimme herausgestellt (vgl. Klatt und Klatt 1990). Behauchung bedeutet, dass an den Stimmlippen im Kehlkopf zwei Signale erzeugt werden. Während die Stimmlippen nur lose aneinander geführt werden, strömt eine hohe Luftmenge durch die gebildete glottale Enge. Dadurch schwingen die Stimmlippen (v.a. im vorderen Bereich am Schildknorpel), d.h. es entsteht Stimmhaftigkeit, ein quasi-periodisches Signal. Gleichzeitig bleiben die Stimmlippen während eines Schwingungsvorgangs ganz oder teilweise (v.a. im hinteren Bereich an den Stellknorpeln) geöffnet, sodass die mit hoher Geschwindigkeit an ihnen entlang strömende Luft turbulent wird und gleichzeitig ein Rauschsignal generiert. Das Rauschen in Verbindung mit den wenig dynamischen und zeitlich annähernd symmetrischen Öffnungs- und Schließbewegungen der Stimmlippen hebt die erste Harmonische (H1) an und schwächt alle höheren Harmonischen des quasi-periodischen Signals ab.

Die H1-H2-Messungen wurden auf der Basis von FFT-Schmalbandspektrogrammen (Analysefenster mit 50Hz Breite) durchgeführt, die über einen Zeitraum von 20ms gemittelt wurden. Im Gegensatz zu den übrigen Messgrößen, für die jeweils ein Wert pro Zielsatz ermittelt wurde, bezogen sich die H1-H2-Messungen auf alle Vokale innerhalb der Zielsätze. Entsprechend hing die Anzahl der Messwerte pro Zielsatz von der Anzahl der Vokale im Zielsatz ab.

Formanten manifestieren sich in der Anhebung der Amplituden der Harmonischen, die im Bereich der Resonanzfrequenzen des Ansatzrohres liegen. Folglich wird die H1-H2-Differenz nicht nur durch den Grad der Behauchung, sondern auch durch die Lage des ersten Formanten und damit durch die Vokalqualität beeinflusst (vgl. Klatt und Klatt 1990). Die Lage des ersten Formanten korreliert wiederum mit dem Öffnungsgrad des Vokals. Besonders stark können sich die H1-H2-Differenzen zwischen offenen und nicht-offenen Vokalen unterscheiden. Daher wurde dieser Einflussfaktor in den Zielsätzen ausbalanciert (vgl. 2.1) und sollte somit lediglich zur Streuung der Messwerte, nicht jedoch zu systematischen Unterschieden beitragen. Letztere sind damit als Veränderungen des Behauchungsgrads der Sprechstimme interpretierbar.

Über jeden Vokal hinweg wurden an drei Stellen H1-H2-Differenzen in den gemittelten FFT-Schmalbandspektrogrammen bestimmt: 10ms nach dem Vokalbeginn, in der Vokalmittle und 10ms vor dem Vokalende. Die Vokalgrenzen wurden durch die zuvor angefertigte segmentelle Annotation vorgegeben.

2.5 Statistische Auswertung

Für jede Messgröße der Untersuchung wurde, separat für die Monologe und Dialoge, eine univariate Varianzanalyse (ANOVA) durchgeführt. Die zwei experimentellen Variablen *Satzmodus* und *Emphase* gingen als feste Zwischensubjektfaktoren mit jeweils zwei Faktorstufen (Aussage vs. Frage und sachlich vs. emphatisch) in die Analysen ein. Im Falle der Stimmqualität kam ein dritter Zwischensubjektfaktor, *Akzent*, hinzu, der ebenfalls durch zwei Faktorstufen (akzentuierter vs. unakzentuierter Vokal) konstituiert wurde. Diese Maßnahme trägt dem Umstand Rechnung, dass Akzentsilben mit anderen Bewegungsabläufen des Sprechapparates verbunden sind als nicht-akzentuierte Silben (vgl. Hermes et al. 2008). Es kann daher nicht ausgeschlossen werden, dass sich auch in puncto Stimmqualität akzentuierungsbedingte Unterschiede ergeben. Erste dahingehende Hinweise wurden außerhalb des Deutschen bereits gefunden (vgl. Ní Chasaide und Gobl 2004). Insgesamt wurden 16 einzelne ANOVAs berechnet. Das Signifikanzniveau wurde angesichts der moderaten Stichprobengrößen auf $p \leq 0,05$ festgesetzt. Die Resultate der statistischen Auswertung sind im nachfolgenden Kapitel zusammengefasst.

3. Ergebnisse

Zwei der insgesamt 8 aufgenommenen Sprecherinnen mussten von der weiteren Datenanalyse ausgeschlossen werden. Eine der ausgeschlossenen Sprecherinnen (aus Kiel) hat sich innerhalb der Zielsätze oft versprochen oder gezögert, ohne neu anzusetzen. Des Weiteren hat sie die Zielsätze zum Teil in unvorhergesehener Weise (z.B. auf Funktionswörtern) akzentuiert oder die Fragesätze mit terminal fallender Intonation realisiert. Während diese Verhaltensweisen auf eine Überforderung mit der Aufnahmesituation hindeuten, war für die zweite Sprecherin (aus Plön) das Gegenteil der Grund für den Ausschluss. Die Sprechweise der zweiten Sprecherin war sehr expressiv und emotional überzogen und – insbesondere im Bereich der Zielsätze – von häufigem Lachen überlagert oder unterbrochen, wodurch die Aufnahmen (zumindest für die Zwecke der vorliegenden Untersuchung) ungeeignet waren.

Die übrigen 6 Sprecherinnen kamen hingegen mit der Aufnahmesituation und der Produktionsaufgabe gut zurecht und haben Monologe und Dialoge produziert, die nach einvernehmlicher Meinung der Autoren als alltagssprachlich klingend charakterisiert werden können. Dabei wurde von der Möglichkeit der lexikalischen Abänderung der Dialogtexte selten und vor allem an deren Rändern Gebrauch gemacht. In den Zielsätzen selbst fanden keine lexikalischen Abänderungen statt. Audiobeispiele können unter http://www.linguistik.uni-kiel.de/Niebuhr_index.html heruntergeladen werden. Die durch das satzfinale Substantiv repräsentierte, zu erfragende bzw. die neu zu vermittelnde Information erhielt innerhalb der Frage- und Aussagesätze erwartungsgemäß jeweils eine nukleare Akzentuierung. Ferner wurden, unterstützt durch die entworfenen semantisch-pragmatischen Kontextrahmen, alle Fragesätze mit terminal steigender und alle Aussagesätze mit terminal fallender Intonation realisiert. Daher brauchte aus diesem – wie auch aus anderen Gründen – kein Datensatz der 6 Sprecherinnen von der weiteren Analyse

ausgeschlossen zu werden. Folglich basierten die für die einzelnen Messgrößen berechneten ANOVAs (vgl. 2.5) für die Monologe auf einer Stichprobengröße von jeweils 48 Aussage- und Fragesätzen bzw. 48 sachlichen und emphatischen Sätzen. Für die Dialoge lagen pro Satzbedingung 96 Fälle vor. Da im Rahmen der Stimmqualität Mehrfachmessungen innerhalb jedes Zielsatzes stattfanden, erhöhten sich die Stichprobengrößen auf 196 (148+48) Messwerte für die Monologe und 264 (192+72) Messwerte für die Dialoge.

Die nachfolgenden Abschnitte präsentieren die Ergebnisse der akustischen Analysen getrennt nach Phänomenbereichen und deren Messgrößen. Einleitend hierzu lässt sich sagen, dass der *Satzmodus* einen signifikanten Einfluss auf alle Messgrößen ausgeübt und zudem signifikant mit der *Emphase* der Sätze interagiert hat, für die sich nicht durchgängig ein eigenständiger Effekt zeigte. Generell waren die parametrischen Unterschiede zwischen den experimentellen Variablen in den Dialogen ausgeprägter als in den Monologen.

3.1 Anzahl und Verteilung akzentuierter Silben

Beide experimentellen Variablen, *Satzmodus* und *Emphase*, hatten in den berechneten ANOVAs einen signifikanten Einfluss auf die Anzahl und Verteilung der akzentuierten Silben innerhalb der Zielsätze. Was die Anzahl der akzentuierten Silben anlangt, so finden sich in den 48 Sätzen der Dialogbedingung im Durchschnitt zwar insgesamt mehr akzentuierte Silben als in den 48 Sätzen der Monologbedingung (2,26 vs. 1,69 Akzentsilben), was auch darauf zurückzuführen ist, dass die Zielsätze in den Dialogen länger waren und insb. mehr Inhaltswörter (Verben und Substantive) aufwiesen. Die Einflüsse der experimentellen Variablen waren jedoch in den Monolog- und Dialogbedingungen vergleichbar. So wiesen die Fragesätze signifikant weniger akzentuierte Silben auf als die Aussagen. In sachlichen Sätzen etwa fanden sich für die Fragen durchschnittlich nur etwa halb so viele Akzentsilben wie für die Aussagen (Monologe: 1,22 vs. 2,17 Akzentsilben, $F_{(1,92)}=20,899$; $p<0,001$; Dialoge: 1,58 vs. 2,96 Akzentsilben, $F_{(1,92)}=26,744$; $p<0,001$).

In den emphatischen Sätzen fällt der Unterschied zwischen den Akzentzahlen der Fragen und Aussagen geringer aus, was darauf zurückgeht, dass in den emphatischen Aussagesätzen deutlich weniger Akzentsilben vorkamen als in den sachlichen Aussagesätzen (Monologe: 1,71 vs. 2,17 Akzentsilben; Dialoge: 2,50 vs. 2,96 Akzentsilben), während die Akzentzahlen der emphatischen und sachlichen Fragen ähnlich groß ausfielen (Monologe: 1,40 vs. 1,22 Akzentsilben; Dialoge: 1,84 vs. 1,58 Akzentsilben). Entsprechend ergaben die ANOVAs zu den Akzentzahlen der Monolog- und Dialogsätze keine eigenständigen Effekte der Variable *Emphase* (Monologe: $F_{(1,92)}=3,275$; $p=0,074$; Dialoge: $F_{(1,92)}=0,058$; $p=0,996$), es fanden sich jedoch signifikante Interaktionen der Variablen *Satzmodus* und *Emphase* (Monologe: $F_{(1,92)}=6,754$; $p=0,011$; Dialoge: $F_{(1,92)}=6,673$; $p=0,015$).

Wie in 2.4.1 beschrieben, wurden für die Dialoge aufgrund der komplexeren und heterogeneren Strukturen der hierin enthaltenen Zielsätze zwei weitere akzentbezogene Distanzmaße ermittelt: Die Silbenzahl zwischen dem Satzanfang und der ersten akzentuierten Silbe sowie die Silbenzahl zwischen der letzten prä nuklearen Akzentsilbe und der nuklearen Akzentsilbe. Gemäß den hierzu berechneten ANOVAs wurden diese

beiden Distanzmaße ebenfalls durch den *Satzmodus* beeinflusst. Im Einklang mit der in Fragesätzen insgesamt geringen Anzahl akzentuierter Silben fiel die Distanz zwischen dem Satzanfang und der ersten Akzentsilbe in Fragen deutlich größer aus als in Aussagen. Während im Falle einer Frage durchschnittlich mehr als zwei unakzentuierte Silben (2,27) auf den Satzanfang folgten, stand bei den Aussagen nur etwa eine Silbe (1,07) zwischen dem Satzanfang und der ersten Akzentsilbe ($F_{(1,92)}=16,197$; $p<0,001$). Dieser Unterschied fand sich in vergleichbarer Größe zwischen den beiden Ausprägungen der Variable *Emphase*, die entsprechend keinen eigenen signifikanten Effekt hatte ($F_{(1,92)}=0,026$; $p=0,873$). Auch für die Distanz zwischen der letzten prä nuklearen und der nuklearen Akzentsilbe ergab sich lediglich ein Effekt des *Satzmodus*. Wie auch im anderen Distanzmaß wiesen Fragen eine um etwa eine unakzentuierte Silbe größere Distanz auf als Aussagen. Während bei Fragen im Mittel mehr als drei unakzentuierte Silben (3,29) von der letzten prä nuklearen bis zur nuklearen Akzentsilbe vergingen, waren es war im Falle der Aussagen nur knapp zwei (1,98) unakzentuierte Silben ($F_{(1,92)}=27,959$; $p<0,001$).

3.2 F0-Verlauf am Satzanfang

Um einen Eindruck von den Kontureigenschaften des Intonationsverlaufs zwischen dem Satzanfang und der ersten Akzentsilbe zu gewinnen, wurden drei Maße herangezogen: das Durchschnitts-F0 als Repräsentant des globalen Intonationsniveaus sowie die Standardabweichung des F0 und die Steilheit des F0-Gefälles als Repräsentanten der Variabilität der Intonation. Die Ergebnisse der F0-Messungen sind – von den tatsächlichen Kontureigenschaften der einzelnen Sätze abstrahiert – in Abbildung 3 am Beispiel der Dialogdaten zusammengefasst schematisch dargestellt.

3.2.1 Durchschnitts-F0

Wie auch in Abbildung 3 zu sehen ist, sank das durchschnittliche Intonationsniveau von den 48 Aussage- zu den 48 Fragesätzen hin um gute 30Hz (bzw. 3 HTS) ab. Dieses Absinken in Abhängigkeit vom *Satzmodus* hat sich sowohl in den Mono- als auch in den Dialogen als signifikanter Effekt herausgestellt (Monologe: $F_{(1,92)}=259,820$; $p<0,001$; Dialoge: $F_{(1,92)}=551,952$; $p<0,001$). Darüber hinaus wiesen auch die emphatischen Sätze gegenüber den sachlichen ein tieferes Durchschnitts-F0 auf dem unakzentuierten Konturstück zwischen dem Satzanfang und der ersten Akzentsilbe auf. Allerdings wurde dieser Effekt der *Emphase*, der mit etwa 10Hz (bzw. 0,8 HTS) nur knapp halb so groß ausfiel wie der des Satzmodus, nur im Rahmen der Dialoge signifikant (Monologe: $F_{(1,92)}=1,756$; $p=0,188$; Dialoge: $F_{(1,92)}=27,879$; $p<0,001$). Ferner wurden in den Monologen und Dialogen signifikante Interaktionen zwischen den Variablen *Satzmodus* und *Emphase* gefunden, die darauf zurückgehen, dass das Durchschnitts-F0 in Verbindung mit sachlichen Fragen und Aussagen stärker absinkt als mit emphatischen (Monologe: $F_{(1,92)}=52,113$; $p<0,001$; Dialoge: $F_{(1,92)}=7,098$; $p=0,009$, vgl. Abb. 3). Letztlich ist für den Parameter des Durchschnitts-F0 anzumerken, dass dessen Werte im Bereich der Dialoge zirka 10-15Hz (bzw. 0,8-1,5 HTS) über denen der Monologe lagen. Wie ein separater t-Test für unabhängige Stichproben mit den 96 Werten beider Sprechbedingungen gezeigt hat, handelt es sich auch hierbei um einen klar signifikanten Effekt ($t=-8,852$, $df=95$, $p<0,001$).

3.2.2 Standardabweichung des F0

Das erste Maß zur Abschätzung der Intonationsvariabilität unterschied sich signifikant in Abhängigkeit vom *Satzmodus*. Die Richtung dieses Effekts wurde allerdings durch die Variable *Emphase* bestimmt. Diese folglich ebenfalls signifikante Interaktion beider

Variablen fand sich für die Mono- und für die Dialoge (Monologe: $F_{(1,92)}=33,455$; $p<0,001$; Dialoge: $F_{(1,92)}=11,292$; $p=0,001$). Unter der sachlichen Bedingung reduzierte sich die Standardabweichung des F0 von der Aussage zur Frage in den Monologen um knappe 15% (31,583 vs. 27,125 Hz, $F_{(1,92)}=12,964$; $p=0,001$). In den Dialogen betrug die Reduktion sogar über 20% (41,791 vs. 32,666 Hz, $F_{(1,92)}=34,796$; $p<0,001$; vgl. Abb. 3). Die emphatische Bedingung hingegen führte zu einem leichten Anstieg der Standardabweichung des F0 von der Aussage zur Frage, insbesondere im Falle der Monologe (29,625 vs. 35,543 Hz). Die Variable Emphase selbst hat zu keinen eigenständigen signifikanten Effekten geführt. Allerdings fand sich für die Dialoge eine gegenüber den Monologen um knapp 20% erhöhte F0-Standardabweichung, die sich in einem separaten t-Test für unabhängige Stichproben als signifikant herausgestellt hat (durchschnittliche Standardabweichungen Mono- vs. Dialoge: 30,068 vs. 36,677Hz; $t=-7,753$, $df=95$, $p<0,001$).

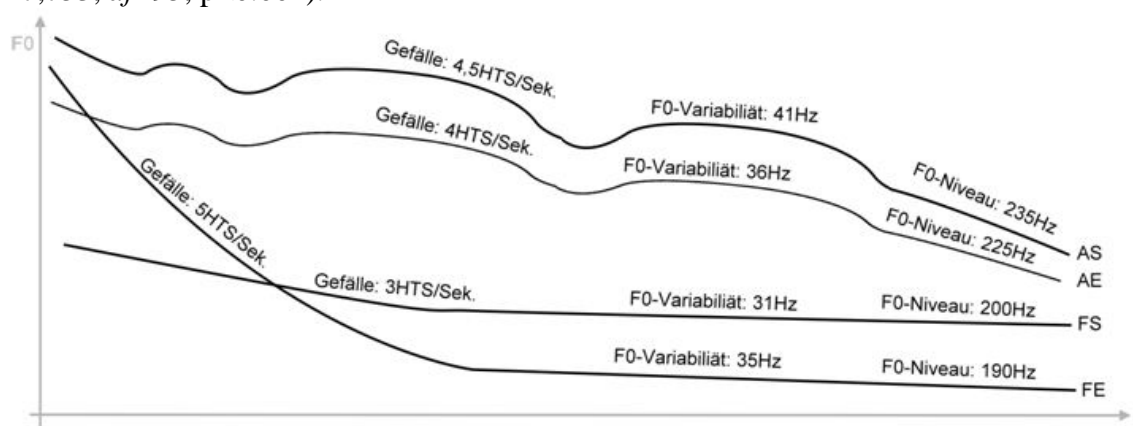


Abbildung 3: Vereinfachte schematische Darstellung der Eigenschaften der F0-Veränderungen (y-Achse) über die Zeit (x-Achse) zwischen dem Satzanfang und der ersten Akzentsilbe anhand des Durchschnitts-F0 (in Hz), der F0-Standardabweichung (in Hz) und des F0-Gefälles (in Halbtonschritten pro Sekunde, HTS/Sek.). Die Darstellung zeigt für die Dialogdaten die Mittelwerte der drei Messgrößen separat für die Variablen Satzmodus (Aussagen vs. Fragen, A vs. F) und Emphase (sachlich vs. emphatisch, S vs. E). Eine höhere Intonationsvariabilität wird durch eine Welligkeit der F0-Kontur angezeigt, die, ebenso wie der zeitliche Verlauf des F0-Gefälles, nicht den tatsächlich produzierten Kontureigenschaften aller Sätze entsprechen muss.

3.2.3 Steilheit des F0-Gefälles

Das F0-Gefälle zwischen dem Satzanfang und der ersten Akzentsilbe wurde signifikant durch die Variable *Satzmodus* beeinflusst. Ähnlich wie zuvor bei der Standardabweichung war jedoch die Richtung dieser Beeinflussung von der Variable *Emphase* abhängig. Von den sachlichen Aussagen zu den sachlichen Fragen wurde das F0-Gefälle um gut 1,0-1,5 HTS pro Sekunde flacher (Monologe: 4,333 vs. 3,412 HTS/Sek., $F_{(1,92)}=7,072$; $p=0,009$; Dialoge: 4,412 vs. 2,958 HTS/Sek., $F_{(1,92)}=3,779$; $p=0,046$; vgl. auch Abb. 3). Von den emphatischen Aussagen zu den emphatischen Fragen wurde es um etwa einen Halbtonschritt pro Sekunde steiler (Monologe: 4,525 vs. 5,354 HTS/Sek.; Dialoge: 4,041 vs. 4,912 HTS/Sek.). Die emphatische Frage hat mit gut 5 HTS pro Sekunde das steilste, die sachliche Frage mit nur etwa 3 HTS pro Sekunde das flachste F0-Gefälle. In der Folge ist auch die Interaktion zwischen den Variablen *Satzmodus* und *Emphase* signifikant (Monologe: $F_{(1,92)}=46,994$; $p<0,001$; Dialoge: $F_{(1,92)}=70,973$; $p<0,001$). Die Variable

Emphase selbst wurde hingegen nur im Rahmen der Dialogsätze signifikant ($F_{(1,92)}=30,309$; $p<0,001$). Abgesehen von den überkreuzten Einflüssen des *Satzmodus* und der *Emphase* auf das F0-Gefälle konnte auch festgestellt werden, dass sich die Aussage- und Fragesätze in den Dialogen durch ein generell um etwa einen Halbtonschritt pro Sekunde niedrigeres F0-Gefälle auszeichneten als in den Monologen. Dieser Effekt hat sich in einem gesondert durchgeführten t-Test für unabhängige Stichproben als signifikant erwiesen (durchschnittliche F0-Gefälle Mono- vs. Dialoge: 4,752 vs. 4,071 HTS/Sek.; $t=5,798$, $df=95$, $p<0,001$).

3.3 Sprechgeschwindigkeit

3.3.1 Silbenzahl pro Sekunde

Das für die Sprechgeschwindigkeit etablierte Maß der Silbenzahl pro Sekunde (Silb/Sek.) zeigte ebenfalls signifikante Einflüsse der experimentellen Variablen. Im Falle der Monologe war jedoch lediglich ein Effekt des *Satzmodus* festzustellen. Dieser war dergestalt, dass die 48 Fragesätze im Mittel mit einer etwa 10% höheren Anzahl von Silben pro Sekunde produziert wurden als die 48 Aussagesätze (5,239 vs. 4,675 Silb/Sek.; $F_{(1,92)}=4,223$; $p=0,043$). In der Dialogbedingung fiel dieser Unterschied zwischen den jeweils 48 Frage- und Aussagesätzen noch größer aus. Gegenüber den Aussagen war die Silbenzahl pro Sekunde in den Fragen um etwa 20% höher (4,912 vs. 4,005 Silb/Sek.; $F_{(1,92)}=197,443$; $p<0,001$). Die genannten Mittelwertsunterschiede werden aus Abbildung 4 ersichtlich. Die Unterschiede mögen klein anmuten, liegen aber deutlich oberhalb der empirisch geschätzten Wahrnehmungsschwelle (vgl. Klatt und Cooper 1975).

Zudem ergaben sich für die Dialogbedingung auch ein signifikanter Effekt der Variablen *Emphase* auf die Silbenzahl pro Sekunde sowie eine signifikante Interaktion zwischen den Variablen *Satzmodus* und *Emphase*. Wie ebenfalls in Abbildung 4 zu erkennen ist, war die Silbenzahl pro Sekunde im Durchschnitt für emphatische Sätze niedriger als für sachliche (4,178 vs. 4,600 Silb/Sek.; $F_{(1,92)}=78,705$; $p<0,001$). In der signifikanten Interaktion zwischen *Satzmodus* und *Emphase* in der Dialogbedingung kommt zum Ausdruck, dass der Anstieg der Silbenzahl pro Sekunde von Aussagen zu Fragen bei sachlichen Sätzen größer ausfällt als bei emphatischen (Sachlich: 4,208 vs. 4,966; Emphatisch: 3,957 vs. 4,249; $F_{(1,92)}=4,557$; $p=0,035$).

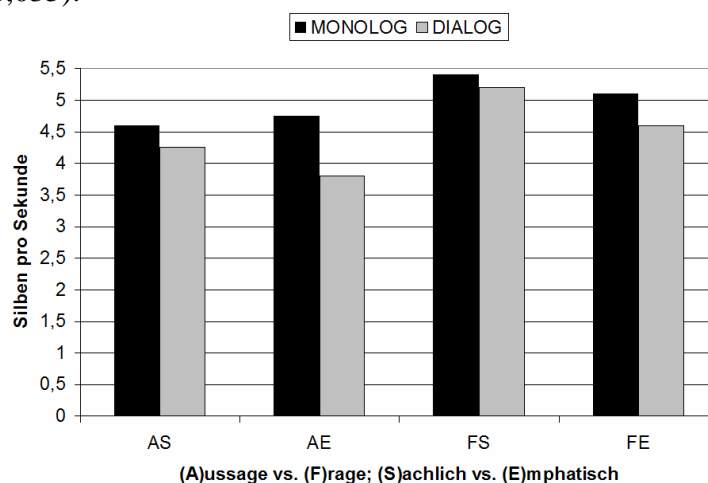


Abbildung 4: Mittelwerte der Silben pro Sekunde für die jeweils 48 Sätze zu den beiden Variablen Satzmodus (Aussage vs. Frage, A vs. F) und Emphase (sachlich vs. emphatisch, S vs. E). Die grauen und schwarzen Balken stellen die Dialog- bzw. Monologsätze dar.

Mit Blick auf Abbildung 4 kann ferner festgehalten werden, dass die Anzahl der Silben pro Sekunde für die Dialoge konsistent etwa eine halbe Silbe pro Sekunde unter der der Monologe lag. Gemäß einem separaten t-Tests für unabhängige Stichproben, der auf der Grundlage der jeweils 96 Sätze in beiden Produktionsbedingungen durchgeführt wurde, handelt es sich hierbei um einen statistisch signifikanten Befund (durchschnittliche Silbenzahl pro Sekunde Mono- vs. Dialoge: 4,927 vs. 4,389; $t=6,305$, $df=95$, $p<0,001$).

3.3.2 ΔC

Auch für die Standardabweichungen der Konsonantendauern (in ms) innerhalb der 48 Frage- und Aussagesätze (ΔC) in den Monolog- und Dialogbedingungen finden sich geringe, aber klar signifikante Effekte. In den Monologen nimmt ΔC von den Aussagen zu den Fragen signifikant ab (32,17 vs. 30,02 ms; $F_{(1,92)}=7,113$; $p=0,009$). Diese mittlere Abnahme konzentriert sich jedoch auf die sachlichen Sätze (Sachlich: 33,38 vs. 29,08 ms; Emphatisch: 30,96 vs. 31,35 ms), was sich in einer signifikanten Interaktion zwischen *Satzmodus* und *Emphase* widerspiegelt ($F_{(1,92)}=7,351$; $p=0,005$) und darüber hinaus dazu führt, dass die Variable *Emphase* per se nicht signifikant wird ($F_{(1,92)}=0,114$; $p=0,737$).

Was die Dialogsätze anlangt, so fällt, wie im Falle der Silbenzahl pro Sekunde zuvor, der Unterschied zwischen Fragen und Aussagen deutlicher aus als in der Monologbedingung. Bezüglich ΔC bedeutet dies, dass der Wert von der Aussage zur Frage hin mehr als doppelt so stark absinkt (31,63 vs. 27,25 ms; $F_{(1,92)}=37,762$; $p<0,001$). Eine weitere Parallele zur Silbenzahl pro Sekunde besteht darin, dass sich, anders als in der Monologbedingung, in der Dialogbedingung auch für ΔC ein separater signifikanter Effekt der Variablen *Emphase* gezeigt hat. Demnach waren die sachlichen Sätze durch einen im Mittel kleineren ΔC -Wert gekennzeichnet als die emphatischen (28,46 vs. 30,42 ms; $F_{(1,92)}=7,566$; $p=0,007$). Da zudem die ΔC -Abnahme von Aussagen zu Fragen in den sachlichen Sätzen umfangreicher ist als in den emphatischen Sätzen, kommt es zu einer signifikanten Interaktion zwischen den Variablen *Satzmodus* und *Emphase* (Sachlich: 35,71 vs. 25,13 ms; Emphatisch: 29,38 vs. 27,54 ms; $F_{(1,92)}=76,042$; $p<0,001$).

3.4 Stimmqualität

Die Stimmqualität, im Speziellen der Grad der Behauchung der Stimme, wurde über die Differenz der ersten beiden Harmonischen des Sprachsignals (H1-H2, in dB) geschätzt. Für jeden Vokal wurden in jedem Zielsatz drei Differenzwerte bestimmt, jeweils einer 10ms nach dem Vokalbeginn, in der Vokalmitte und 10ms vor dem Vokalende. Folglich variierte die Anzahl der Messwerte, die einen Zielsatz repräsentierten, mit der Anzahl der darin enthaltenen Vokale. Abweichend von den anderen gemessenen Parametern wurden die Zielsätze dann jedoch nicht nur entlang der beiden experimentellen Variablen *Satzmodus* und *Emphase* gruppiert und statistisch anhand von separaten ANOVAs für die Mono- und Dialoge (hinsichtlich H1-H2-Mittelwertunterschieden) ausgewertet. Vielmehr wurde den ANOVAs eine weitere zielsatzinterne Variable, *Akzent*, hinzugefügt, mittels derer die H1-H2-Werte der akzentuierten und nicht akzentuierten Silben getrennt behandelt wurden. Die Ergebnisse der H1-H2-Messungen sind in Form der Mittelwerte für die drei zweistufigen Variablen in den Abbildungen 5(a)-(b) graphisch dargestellt.

Wie in den Abbildungen 5(a)-(b) zu erkennen ist, haben sich für die Fragen im Mittel etwa 3-4dB höhere H1-H2-Werte ergeben als für die Aussagen. Auch die emphatischen Sätze wiesen 2-3dB höhere H1-H2-Mittelwerte auf als die sachlichen. Beide dB-Unterschiede

sind deutlich signifikant, sowohl für die Monologe (*Satzmodus*: $F_{(1,616)}=93,163$; $p<0,001$; *Emphase*: $F_{(1,616)}=65,046$; $p<0,001$) als auch für die Dialoge (*Satzmodus*: $F_{(1,1056)}=193,820$; $p<0,001$; *Emphase*: $F_{(1,1056)}=145,070$; $p<0,001$). Entsprechend diesen Effekten fanden sich für die emphatischen Fragen über die Vokale hinweg insgesamt die höchsten H1-H2-Werte (Monologe: 10,333 dB, Abb. 5a; Dialoge: 12,075 dB, Abb. 5b); für die sachlichen Aussagen waren die H1-H2-Werte im Durchschnitt am niedrigsten (Monologe: 2,878 dB, Abb.5a; Dialoge; 1,435 dB, Abb.5b).

Für die Zusatzvariable *Akzent* ist festzustellen, dass sich die akzentuierten Silben insgesamt durch höhere H1-H2-Mittelwerte auszeichneten als die nicht akzentuierten. Mit etwa 3dB fällt dieser Effekt ähnlich groß aus wie der zwischen sachlichen und emphatischen Sätzen. Der signifikante Effekt der Variablen *Akzent* prägte sich in den Dialogsätzen noch deutlicher und konsistenter aus ($F_{(1,1056)}=146,146$; $p<0,001$) als in den Monologsätzen ($F_{(1,616)}=25,999$; $p<0,001$). Zudem fällt der Anstieg der H1-H2-Mittelwerte von den unakzentuierten zu den akzentuierten Silben im Rahmen der emphatischen Frage- und Aussagesätze zirka 2-3dB größer aus als im Rahmen der sachlichen. Dies gilt für die Monologe und insbesondere auch für die Dialoge und resultierte in signifikanten Interaktionen der Variablen *Emphase* und *Akzent* (Monologe: $F_{(1,616)}=10,825$; $p<0,001$; Dialoge: $F_{(1,1056)}=12,334$; $p<0,001$).

Letztlich ist darauf hinzuweisen, dass in den Dialogen ein global etwa 5,5dB niedrigeres H1-H2-Niveau gemessen wurde als in den Monologen. Ein separater t-Test für unabhängige Stichproben auf Basis aller H1-H2-Werte der Mono- und Dialoge zeigte, dass es sich hierbei um einen klar signifikanten Unterschied handelt (6,203 vs. 11,758 dB; $t=8,950$; $df=458$; $p<0,001$).

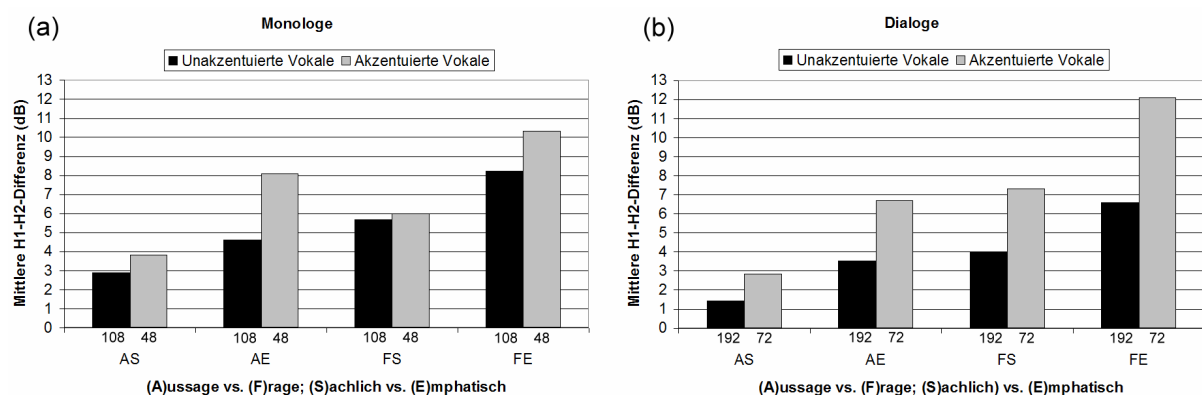


Abbildung 5: Mittelwerte der H1-H2-Differenzen (in dB) für die drei Variablen Satzmodus (Aussage vs. Frage, A vs. F), Emphase (sachlich vs. emphatisch, S vs. E) und Akzent (akzentuierte vs. nicht akzentuierte Silben, graue vs. schwarze Balken). Die linke Abbildung (a) zeigt die Monologe, die rechte Abbildung (b) die Dialoge. Die numerischen Angaben unter den Balken verweisen auf die Stichprobengrößen.

4. Diskussion

4.1 Hypothesen und Schlussfolgerungen

Die der phonetischen Untersuchung zugrunde gelegten Hypothesen (2)-(3) haben sich bestätigt. Deklarative Nachfragen wurden gegenüber vergleichbar lautenden Aussagen mit höherer Sprechgeschwindigkeit und größerer Behauchung gesprochen. Darüber hinaus

wiesen die deklarativen Nachfragen im Einklang mit Hypothese (1a) weniger pränukeare Akzente auf als die vergleichbaren Aussagen. Die Hypothese (1c) zum Absinken des Intonationsniveaus ging mit den Ergebnissen nur teilweise konform. Ein langsames Absinken des Intonationsniveaus bei Fragen konnte satzinitial nur im Bereich der sachlichen Äußerungen gefunden werden. Emphatische Fragen hingegen waren durch ein schnelleres Absinken des satzinitialen Intonationsniveaus gekennzeichnet als emphatische Aussagen. Der Einfluss der *sachlich-emphatisch*-Dimension auf die Satzmodusunterschiede wird in Abschnitt 4.4 weiter diskutiert. Hypothese (1b) letztlich konnte durch die Daten nicht bestätigt werden. Zwar wurde ein systematischer Unterschied zwischen den deklarativen Nachfragen und den dazugehörigen Aussagen im Intonationsniveau gefunden. Die Nachfragen lagen jedoch nicht auf einem höheren, sondern auf einem tieferen Niveau als die Aussagen. Mögliche Gründe für den unerwarteten Befund werden in 4.4 eruiert.

Unabhängig von den weiterführenden Diskussionen legen die Ergebnisse den Schluss nahe, dass das formbezogene Konzept der 'Intonationsfrage' inadäquat ist. Zwar war Intonation an der phonetischen Unterscheidung zwischen Aussagen und Deklarativfragen in mehrfacher Weise beteiligt, die Kennzeichnung des Satzmodus basierte jedoch auf einem vielschichtigeren Einsatz prosodischer Mittel, zu dem ebenso Sprechgeschwindigkeit, Stimmqualität und in Form der Akzentuierungsmuster auch Sprechrhythmus zählten. Darüber hinaus geht das Formkonzept der 'Intonationsfrage' von einer inadäquaten zeitlichen Domäne aus. Während die lexikalische und syntaktische Markierung von Fragen satzinitial stattfindet, sollen deklarative Nachfragen nur anhand der phonetischen Markierung am Satzende als Frage erkennbar sein. Unsere Untersuchung korrigiert diese zeitliche Asymmetrie zwischen den verschiedenen formalen Fragetypen. Das gefundene Spektrum prosodischer Unterschiede betrifft den Satz als Ganzes, und es ist anzunehmen, dass deklarative Nachfragen dadurch ebenfalls bereits am Satzanfang von Aussagen hinreichend klar differenziert sind. Dass prinzipiell eine frühzeitige, satzinitiale Markierung von Fragen für eine erfolgreiche Kommunikation von Nutzen ist, ist unbestreitbar.

Einhergehend mit der empirischen Widerlegung des Formkonzeptes der 'Intonationsfrage' lässt die breit gefächerte prosodische Unterscheidung zwischen deklarativen Aussagen und Fragen ausreichend Spielraum für die Annahme, dass die Richtung der satzfinalen Intonationsbewegung auch in deklarativen Fragen nicht mit der Signalisierung des Satzmodus funktional belastet ist, sondern ebenso wie in lexikalisch und syntaktisch markierten Fragen zum Ausdruck eines attitudinalen Bedeutungsunterschieds verwendet werden kann, wie er unter 1.2 umrissen wurde (vgl. Kohler 2004; Peters 2005). Hierzu passt, dass Petrone und Niebuhr (2009) bereits deklarative Fragen mit satzfinal fallender Intonation in Sprachkorpora des Deutschen finden konnten (vgl. Abb.1). Insgesamt ist einschränkend darauf hinzuweisen, dass die getroffenen Schlussfolgerungen einer Unterstützung durch Wahrnehmungsexperimente bedürfen, die die einzelnen prosodischen Unterschiede zwischen Aussagen und Deklarativfragen auf ihre Nutzung durch den Hörer hin testen und das Zusammenwirken dieser Unterschiede mit der Form und Funktion der satzfinalen Intonationsbewegung eingehender untersuchen.

4.2 Kritische Anmerkungen

Im Einklang mit den Ausführungen in 1.4 unterstützen die Ergebnisse die Annahme, dass die einzelnen akustischen Messgrößen eigenständige Beiträge zum Gesamtbild der

Ergebnisse geleistet haben. Signifikante Befunde sind keine Epiphänomene anderer signifikanter Befunde. Wären zum Beispiel die Unterschiede in der Sprechgeschwindigkeit ausschließlich eine Folge signifikant unterschiedlicher pränuklearer Akzentdichten, dann hätten sich für die emphatischen und sachlichen Fragen bei vergleichbarer pränuklearer Akzentdichte keine signifikanten Sprechgeschwindigkeitsunterschiede ergeben. Zudem zeichneten sich emphatische und sachliche Fragen trotz ihrer vergleichbaren pränuklearen Akzentdichte durch deutlich unterschiedliche Intonationsvariabilitäten und –gefälle im pränuklearen Bereich bis zur ersten Akzentsilbe aus. Intonationsniveau und –gefälle haben untereinander ebenfalls zu unabhängigen Befunden geführt. Es gibt keinen Anlass dazu, eine Wechselwirkung im Sinne einer positiven Korrelation anzunehmen. So unterschieden sich beispielsweise sachliche Aussagen von emphatischen Aussagen auf der einen und emphatischen Fragen auf der anderen Seite nicht im Intonationsgefälle, aber erheblich und signifikant im Intonationsniveau. Darüber hinaus lag das Intonationsniveau sachlicher Fragen deutlich unter den Niveaus der sachlichen und emphatischen Aussagen, obwohl das Intonationsgefälle nicht steiler, sondern flacher ausfiel. Hinsichtlich der sprachproduktorisches Kopplung von Intonationsniveau und Stimmqualität zeigen die Ergebnisse, dass im Falle der emphatischen Aussagen und sachlichen Fragen der größte Unterschied im Intonationsniveau mit dem kleinsten (und nicht signifikanten) Unterschied in der Behauchung zusammenfällt. Auch hier sind die Befunde der einen Messgröße also nicht aus den Befunden der anderen Messgröße vorhersagbar. Wie in 4.4 skizziert werden wird, könnte aber dennoch eine gewisse Beziehung zwischen beiden Messgrößen bestehen, die jedoch nicht als Automatismus im Sinne einer sprachproduktorisches Kopplung zu verstehen ist, sondern als funktional oder sprechökonomisch intendierte Handlung.

Des Weiteren muss darauf hingewiesen werden, dass die Ergebnisse dieser Studie, den ihr zugrundeliegenden Sprecherinnen folgend, erst einmal nur für die im zentralen Schleswig-Holstein gesprochene Varietät des Standarddeutschen gelten. Trotz noch vieler empirischer Lücken ist auf dem gegenwärtigen Stand der Forschung bereits klar, dass verschiedene regionale Varietäten des Deutschen Aussagen (vgl. Auer et al. 2000; Gilles 2005; Selting 2005) sowie Fragewort- und Verberststellungsfragen (Schädlich und Eras 1970; Peters 2004, 2006) in zum Teil sehr unterschiedlicher Weise realisieren. Ähnlich wie im Englischen (vgl. Grabe 2004) können – bei ansonsten konstanten semantisch-pragmatischen Bedeutungen – in einigen Varietäten Aussagen am Satzende intonatorisch ansteigen und Fragen am Satzende abfallen. Auch in ihren Intonationsvariabilitäten und den Dynamiken von An- und Abstiegen weichen regionale Varietäten zum Teil erheblich voneinander ab (vgl. Gilles 2005; Kleber und Rathcke 2008). Darüber zeigen die Untersuchungen von Leemann und Siebenhaar (2006), dass regionale Varietäten im Deutschen grundsätzlich auch durch spezifische Sprechgeschwindigkeiten gekennzeichnet sein können. Zusammengefasst ist es von Vornherein offensichtlich, dass eine Ausweitung der vorliegenden Studie auf Standard- und Dialektsprecherinnen und –sprecher aus anderen regionalen Regionen nicht zu quantitativ vergleichbaren Ergebnissen führen muss. Überdies ist es höchst unwahrscheinlich, dass selbst die qualitativen Unterschiede alle varietätenunabhängig erhalten bleiben, da unterschiedliche Varietäten unterschiedlichen Spielraum für satzmodusbezogene Variationen der untersuchten akustischen Parameter lassen. Auch im Italienischen, Dänischen und den Tonsprachen, die in die Hypothesen eingeflossen sind, wurden nicht alle hier gefundenen qualitativen Unterschiede beobachtet. Einige qualitative Unterschiede mit struktureller Natur oder ansonsten geringer funktional Last könnten sich jedoch im deutschsprachigen Raum als überregional robust erweisen.

Dies betrifft zum Beispiel die pränuclare Akzentdichte, das Intonationsniveau und die Behauchung.

Ferner kann nicht einmal mit Sicherheit gesagt werden, dass die Fragestellung dieser Studie für andere Varietäten des Deutschen als die untersuchte überhaupt gleichermaßen naheliegend wäre. Dass Deklarativfragen nämlich ebenso wie Aussagen satzfinal eine terminal fallende Intonation zeigen können und folglich auf anderem Wege voneinander differenziert werden müssen, wurde bislang nur für norddeutsche Standardsprecherinnen und -sprecher dokumentiert. Insgesamt sollten die Befunde der vorliegenden Studie daher zum Anlass genommen werden, (a) die prosodische Kodierung von Deklarativfragen vor dem Hintergrund der in 1.2 beschriebenen kommunikativen Funktionen überregional eingehender zu erforschen und dabei (b) über die reine Intonation und Unterschiede am Satzende hinauszugehen.

4.3 Weitergehende Erkenntnisse

4.3.1 Die dominierende Rolle der Intonation

Wenn es um die suprasegmentelle Signalisierung kommunikativer Bedeutungen geht (attitudinale Bedeutungen, Fokuskategorien, Emphase, Dialogsteuerung, Satzmodus etc.), ist das phonetische Augenmerk beinahe automatisch auf die Intonation gerichtet. Einzelne hohe und tiefe Tonhöhenakzente oder Grenztöne werden auf einer phonologischen Grundlage mit bestimmten Bedeutungen verknüpft (vgl. Grice und Baumann 2002; Féry 1993; Pierrehumbert und Hirschberg 1990). Die vorliegende Studie unterstreicht erneut, dass diese phonetisch-phonologische Perspektive zu eng ist. Mit suprasegmentellen Mitteln ausgedrückte Bedeutungen basieren ebenso wie die Signalisierung der Segmente auf multifaktoriellen Parametervariationen und dem Prinzip der redundanten Kodierung. Selbst die Trennung zwischen der segmentellen und suprasegmentellen Ebene lässt sich nicht aufrecht erhalten. So zeigt Niebuhr (2007a, 2009, 2010), dass vermeintlich rein intonatorisch ausgedrückte, attitudinale und emphatische Bedeutungen auch über Variationen in der Qualität, Dauer und Energie von Frikativen und Vokalen signalisiert werden. Nachfolgende Studien müssen dieses Geflecht phonetischer Parameter und ihr Zusammenspiel grundsätzlich stärker berücksichtigen und über <H>s und <L>s hinausgehen.

4.3.2 Die Methode der Datenerhebung

Zusätzlich zu den hypothesenbezogenen Schlussfolgerungen hat diese Studie weitere Erkenntnisse erbracht, die für die zukünftige phonetische Forschung von Interesse sind. Dies betrifft vor allem die Methode der Sprachdatenerhebung, bei der zweierlei Wege beschritten wurden. Zum einen wurden die Zielsätze von den Versuchspersonen – auf für die phonetische Forschung traditionelle Weise – als Abfolgen von Einzelsatzmonologen und mit der expliziten Instruktion produziert, sachliche Aussagen und Fragen bzw. Aussagen mit „*vehementem Zurückweisen*“ und Fragen mit „*ungläubigem Erstaunen*“ auszudrücken. Zum anderen wurden Dialogtexte entworfen, in denen diese vier Satzmodus- und Emphasebedingungen allein über den erzeugten semantisch-pragmatischen Kontext elizitiert werden sollten, sodass die Versuchspersonenpaare lediglich instruiert wurden, die Dialoge so natürlich und spontan wie möglich klingen zu lassen. Diese Instruktion wurde durch verschiedene Maßnahmen unterstützt. Die Versuchspersonenpaare bekamen die Dialoge eine halbe Stunde vor Beginn der Aufnahme zum Üben und mit dem Hinweis ausgehändigt, dass sie Wörter und Formulierungen

auslassen oder austauschen durften, um die Texte ihrer persönlichen Ausdrucksweise anzupassen. Die Dialoge waren zudem inhaltlich auf Alltagsaspekte ausgerichtet und in einem informellen Sprechstil verschriftet. Der informelle Charakter der Aufnahme wurde des Weiteren dadurch gefördert, dass die Versuchspersonen eines Dialogs nicht nur untereinander eng befreundet, sondern auch mit den Aufnahmeleitern bekannt waren.

In den Arbeiten von Kohler und Niebuhr (2007) und Niebuhr (2009, 2010) wurde der Erfolg der kontextgesteuerten Elizitation von Sprachdaten über Dialoge, die von einer Textgrundlage ausgehend informell produziert wurden, bereits betont. Diese Einschätzung kann durch die parallel erhobenen Monolog- und Dialogstichproben in der vorliegenden Studie erstmalig empirisch untermauert werden. So haben die explizit instruierten Monologe und die kontextgesteuert elizitierten Dialoge zu qualitativ vergleichbaren phonetischen Unterschieden zwischen den Zielsätzen der Satzmodus- und Emphasebedingungen geführt. Das heißt, die Schaffung entsprechender semantisch-pragmatischer Kontexte kann tatsächlich explizite funktionale Instruktionen ersetzen. Darüber hinaus haben die Dialoge insbesondere für die emphasebezogenen Unterschiede in den Bereichen der satzinitialen F0-Muster, der Sprechgeschwindigkeit und der Stimmqualität noch deutlichere phonetische Unterschiede generiert als die Monologe. In letzteren wurde der Faktor *Emphase* nicht durchgängig signifikant. Dies unterstützt die Forderung von Kohler und Niebuhr (2007), dass insbesondere phonetische Studien zu expressiven und attitudinalen Bedeutungen – sofern sie einen lesesprachlichen Ansatz verfolgen – besser von Dialogen als von Monologen ausgehen sollten.

Frühere Studien zu phonetischen Vergleichen zwischen Lese- und Spontansprache (z.B. Terminabsprachen und Nacherzählungen im Dialog) haben ergeben, dass Spontansprache sich durch ein höheres und langsamer absinkendes Intonationsniveau, eine erhöhte Intonationsvariabilität sowie durch eine geringere Sprechgeschwindigkeit und Behauchung auszeichnet als die Lesesprache (z.B. Mixdorff und Pfitzinger 2005; Hirschberg 2000; Swerts et al. 1996). In eben diesen Parametern wiesen separaten t-Tests zufolge auch die Monologe und Dialoge der vorliegenden Studie generelle Unterschiede auf, die parallel zu denen der Lese- und Spontansprache verliefen. Das heißt, die Dialoge waren insgesamt echter Spontansprache auf globaler phonetischer Ebene ähnlicher als die Monologe. Damit existiert erstmalig ein empirisches Argument für den bereits von Kohler und Niebuhr geschilderten Gehörseindruck, den auch die Autoren der vorliegenden Studie teilen: Die erhobenen Dialoge klingen in weiten Teilen wie Spontansprache. Versuchspersonen sind in der Lage, die einzige Instruktion „*klinge so natürlich und spontan wie möglich*“ intuitiv in angemessener Weise umzusetzen. Natürlich wird mit dieser Schlussfolgerung nicht behauptet, dass diese Umsetzung immer und in allen Aspekten Spontansprache imitiert. Die Dialoge erfüllen nicht die Kriterien natürlicher Gespräche (vgl. Zwirner 1964) und würden von Prins und Bastiaanse (2004) als Semispontansprache klassifiziert werden. Somit wird Spontansprache als eigener Untersuchungsgegenstand nicht obsolet. Bei vergleichbaren Möglichkeiten, phonetische Variablen zu kontrollieren, sind Dialoge Monologen in puncto Validität und Generalisierbarkeit der Befunde aber offensichtlich klar überlegen. Folglich zahlt sich der im Rahmen einer Dialogerhebung beträchtliche, zeitliche und organisatorische Mehraufwand aus.

Einschränkend ist das Problem unterschiedlicher Sprecherkompetenzen zu berücksichtigen. Obwohl es sich um ein grundsätzliches Problem bei der Erhebung und

Auswertung von Sprachdaten handelt, kommt es in der beschriebenen Dialogmethode besonders zum Tragen, da die Versuchspersonen anders als in Monologen und echter Spontansprache nicht einfach explizite Instruktionen umsetzen bzw. schlichtweg frei sprechen müssen. Die beschriebene Dialogmethode verlangt, dass sich die Versuchspersonen in eine Alltagssituation hineinversetzen, von der Textgrundlage abstrahierend einen spontansprachlichen Sprechstil annehmen und dabei gleichzeitig die semantisch-pragmatischen Zusammenhänge im Text richtig erkennen und umsetzen sowie an einen Dialogpartner gerichtet kommunizieren. Ähnlich hohe Ansprüche gelten auch für Synchronsprecher. Dementsprechend ist, wie bei Kohler und Niebuhr (2007) geschehen, eine Vorauswahl der Sprecher erforderlich. Obwohl für die vorliegende Studie ebenfalls eine Vorauswahl getroffen wurde, mussten 2 Sprecherinnen von der phonetischen Analyse ausgeschlossen werden.

Eine weitere Einschränkung besteht darin, dass sich die hier angewandte Methode der Elizitations nicht für alle Forschungsrichtungen gleichermaßen gut eignet. Im Rahmen der dialektologischen Forschung ist es vermutlich nicht günstig, von einer Textgrundlage auszugehen, da anzunehmen ist, dass diese das Auftreten nicht-standardsprachlicher Formen eher unterdrückt. Aufgrund anderer Forschungsziele besteht in der Dialektologie wohl auch kein durchgängiger Bedarf an Textgrundlagen. Einzelne Elemente der Methode, wie etwa das Hinzuziehen eines befreundeten Dialogpartners in die Aufnahmesituation und das Lösen einer sprachlich gesteuerten Aufgabe im Gespräch (vgl. auch die Map-Task-Methode, Arnorld et al. 1991) scheinen aber durchaus übertragbar zu sein und könnten zukünftige „Tonbandinterviews“ verbessern bzw. sie phonetisch und allgemein sprechstilistisch bereichern. Die dialektologische Forschung hat schon vor geraumer Zeit die Notwendigkeit gesehen, über „*geschickte Erhebungstechnik*“ (Wagener 1986:323) informelle, ungezwungene Passagen im Rahmen von Gesprächen zu elizitieren und dabei erkannt, dass dafür die Aufnahmeleiter im Rahmen ihrer Aufnahmen in den Hintergrund treten und die Probanden im Gespräch durch eine Aufgabe von der Aufnahmesituation abgelenkt werden sollten (vgl. auch Menge 1982). Wie bereits zu Beginn von 2.1 erwähnt, ist der Argumentation Wagener (1986) zuzustimmen, dass jede Aufnahmemethode auf ihre Weise eine Facette des natürlichen sprachlichen Ausdrucksverhaltens von Sprecherinnen und Sprechern erfasst. Das bedeutet jedoch nicht, dass methodischer Fortschritt in Richtung auf spontansprachliches Verhalten überflüssig ist.

4.3.3 Das Maß ΔC

Das Maß ΔC wurde ursprünglich von Ramus et al. (1999) als eine Komponente in der Vermessung und Klassifikation der Sprechrhythmen verschiedener Sprachen eingeführt. Es hat sich jedoch in der Anwendung auf konkrete Daten als besonders anfällig für Variationen in der Sprechgeschwindigkeit herausgestellt. Aufgrund dessen wurde ΔC in der vorliegenden Studie gezielt für die Messung von Sprechgeschwindigkeitsunterschieden herangezogen. Wie die Auswertung der Messungen zeigte, hat ΔC in der Tat zu Ergebnissen geführt, die mit denen der etablierten und als besonders valide und robust geltenden Silbenrate konform gehen. Demzufolge ist ΔC eine zur Silbenrate geeignete Alternative für die Bestimmung der Sprechgeschwindigkeit. Zwar setzen beide Maße eine vorherige Segmentierung des Sprachsignals voraus. Im Gegensatz zur Abgrenzung von Silben, bei der phonologische und/oder perzeptorische Faktoren eine gewichtige Rolle spielen, ist eine Segmentierung auf Einzellautebasis, wie ΔC sie erfordert, aber sehr viel einfacher und über Sprachen und Bearbeiter hinweg auch konsistenter durchführbar.

4.4 Der Faktor Emphase

Der Wechsel von der sachlichen zur emphatischen Produktion der Zielsätze hatte einen klaren Effekt auf alle analysierten Parameter. So wiesen emphatische Aussagen verglichen mit sachlichen weniger pränukeare Akzente bei gleichzeitig tieferer und ebenerer satzinitialer Intonationskontur auf. Die Sprechgeschwindigkeit der emphatischen Aussagen war gegenüber der der sachlichen Aussagen herabgesetzt, der Grad der Behauchung fiel hingegen höher aus. Die gefundenen Charakteristika passen in das phonetische Profil des engen bzw. kontrastiven Fokus im Deutschen (vgl. 1.4), auf das die emphatischen Aussagen in den Monologen und Dialogen abzielten.

Analog zu den Aussagen wurden die emphatischen Fragen im Vergleich zu ihren sachlichen Pendanten ebenfalls stärker behaucht und mit geringerer Sprechgeschwindigkeit sowie tieferem satzinitialen Intonationsniveau gesprochen. Das Intonationsniveau fiel aber anders als im Falle der Aussagen bei den emphatischen Fragen nicht flacher, sondern steiler ab als bei den sachlichen Fragen. Die Intonationsvariabilität der emphatischen war gegenüber der der sachlichen Fragen nicht herabgesetzt, sondern leicht erhöht. Zudem wiesen emphatische Fragen nicht weniger pränukeare Akzente auf als sachliche, sondern etwa gleich viele. Die stärkere Behauchung und die durch segmentelle Dehnungen reduzierte Sprechgeschwindigkeit stimmen mit früheren Befunden zum Ausdruck von Überraschung im Deutschen überein (vgl. Gartenberg und Panzlaff-Reuter 1991; Niebuhr 2007a,b). Das abgesenkte Intonationsniveau passt jedoch nicht in das empirische Profil der Überraschung (vgl. Kehrein 2002).

Eine Erklärung hierfür kann in der starken Behauchung liegen. Eine tiefe Intonation erleichtert den Ausdruck von Behauchung. So ist aus Tonsprachen zum Beispiel bekannt, dass Tieftönigkeit zusammen mit Behauchung auftritt (Jones 1973; Michaud und Mazaudon 2006; Keating und Eposito 2007), die dann sogar den eigentlichen Ton ersetzen kann. Bei der Überlegung, dass die Probanden dem Ausdruck von starker Behauchung den Vorzug vor einer hohen Intonation gegeben haben könnten, muss ebenfalls berücksichtigt werden, dass es sich bei den betreffenden Sätzen um Fragen handelt. Die Beziehung zwischen Überraschung und erhöhtem Intonationsniveau wurde jedoch primär an Aussagesätzen festgemacht. Es ist gut möglich, dass in Fragen eine Beziehung zwischen Intonation und Überraschung gar nicht besteht; oder aber sie besteht in anderer Weise. So endete zum Beispiel der satzfinale Intonationsanstieg in den emphatischen Fragen in Halbtonschritten ausgedrückt mindestens 25% höher als in den sachlichen.

Der Ausdruck von Behauchung kann auch der Grund dafür sein, dass die Fragen entgegen der Hypothese (1b) nicht mit einer höheren, sondern mit einer tieferen satzinitialen Intonation realisiert wurden. Ähnlich wie bei der Emphase zuvor, muss bei dieser Möglichkeit jedoch ebenfalls einschränkend bedacht werden, dass die Hypothese (1b) nicht aus Nachfragen, sondern aus Fragen anderer Funktion und überdies aus anderen Sprachen abgeleitet wurde. Die in der Hypothese zum Ausdruck gebrachte sprach- und funktionsübergreifende Parallele im Intonationsniveau existiert vielleicht nicht. Diese Einschränkung betont zugleich, dass, obwohl Fragen sich durch die gemeinsame ethologische Basis der Fragehandlung sprachübergreifend phonetisch ähneln (vgl. 1.3), sprachspezifische phonetische Merkmale nicht vernachlässigt werden dürfen.

Der zuvor angedeuteten, limitierten Generalisierbarkeit der Befunde ist ein wichtiger Aspekt hinzuzufügen. Die Einbeziehung der *sachlich-emphatisch*-Dimension in die

vorliegende Untersuchung hat entscheidend zur Relativierung der prosodischen Muster beigetragen. Brinckmann und Benz Müller (1999) verbinden in ihrer Untersuchung zur Intonation verschiedener formaler Fragetypen im Deutschen etwa mit der deklarativen Frage bestimmte Eigenschaften wie eine tief einsetzende Intonation, die nach dem Satzanfang nur langsam absinkt. Unsere Analyse stützt dieses Ergebnis, allerdings nur für die sachlichen und nicht für die emphatischen Fragen. Letzteres zeigt das Problem von Pauschalisierungen, die bestimmten kommunikativen Funktionen bestimmte phonetische Merkmale zuweisen. Phonetische Merkmale sind kontextsensitiv und zugleich Ausdrucksmittel verschiedener Funktionen. So hat unsere Analyse beispielsweise auch ergeben, dass sachliche Aussagen durchaus an die Sprechgeschwindigkeit emphatischer Fragen heranreichen können, und emphatische Aussagen überschneiden sich in ihrer Behauchung deutlich mit sachlichen Fragen. Inwieweit Hörer in der Lage sind, diese phonetische Interaktion von Satzmodus und Emphase aufzulösen, muss in Wahrnehmungsexperimenten getestet werden. Weitere Experimente werden im nachfolgenden Ausblick skizziert.

5. Ausblick

Durch die Untersuchung der sprechmelodischen Unterschiede von deklarativen Nachfragen und Aussagen haben sich weiterführende Fragestellungen herauskristallisiert. Die zentralen Fragen betreffen die Relevanz der satzinitialen und globalen prosodischen Merkmale für die Aussage- und Frageidentifikation durch den Hörer und die kommunikative Bedeutung von satzfinal fallenden und steigenden Intonationsbewegungen in deklarativen Fragen.

Überdies stellt sich die Frage der Generalisierbarkeit der für die deklarativen Fragen gefundenen multiparametrischen prosodischen Eigenschaften auf andere Fragetypen. Sind lexikalisch und/oder syntaktisch markierte Fragen zum Beispiel auch durch eine schnellere Sprechgeschwindigkeit, andere Rhythmik, größere Behauchung und tiefer verlaufende, ebenere Intonation am Satzanfang gekennzeichnet als vergleichbare Aussagen? Oder verschwinden diese systematischen prosodischen Unterschiede zwischen den Satzmodi, da bereits eine hinreichende lexikalische oder syntaktische Markierung gegeben ist und die phonetischen Mittel für andere kommunikative Aufgaben frei werden? Treten andere prosodische Unterschiede an deren Stelle, wie die Daten von Brinckmann und Benz Müller (1999) vermuten lassen? In 1.3 und 4.4 wurde angedeutet, dass ein Teil der gefundenen Unterschiede zwischen deklarativen Fragen und Aussagen der besonderen Funktion der Nachfrage und damit verbundenen phonetisch manifesten ethologischen Mustern geschuldet sein könnte (vgl. Gussenhoven 2002). Dieser Umstand spricht gegen eine Generalisierbarkeit der Ergebnisse über alle Fragetypen im Deutschen.

Vor diesem Hintergrund stellt sich auch die Frage nach situativen, soziolinguistischen und regionalen Variationen in den Ausprägungen der gefundenen prosodischen Unterschiede. Mit der Emphase wurde die Relevanz eines situativen Faktors bereits belegt. Weiter ist zu bedenken, dass die vorliegende Studie aus Gründen messphonetischer Konsistenz ausschließlich auf Frauen basierte. Unter der Berücksichtigung, dass Männer aufgrund der physiologischen Eigenschaften ihres Sprechapparates unter anderem eine weniger behauchte Stimme aufweisen, sind geschlechtsspezifische Eigenarten bei der phonetischen Frage-Aussage-Differenzierung denkbar. In jedem Fall gilt es, eine Pauschalisierung

konkreter phonetischer Merkmale zu vermeiden. Dabei müssen in eine differenziertere Analyse der phonetischen Merkmale auch spontansprachliche Daten mit einbezogen werden. Eine Ausweitung der Forschungsfrage auf andere Dialekte des Deutschen muss damit einhergehen zu überprüfen, inwieweit hierin deklarative Fragen mit satzfinal fallender Intonation bzw. der satzfinalen Intonation einer ansonsten funktional vergleichbaren Aussage überhaupt vorkommen können.

6. Danksagungen

Die Autoren dieses Beitrags möchten Alexander Werth, Ulrike Mosel, Michael Elementaler und ihren drei anonymen Gutachtern für ihre hilfreichen Kommentare im Rahmen der Untersuchung und des Verfassens dieses Beitrags herzlich danken.

7. Literatur

- ANDERSON, A.; BADER, M.; BARD, E.; BOYLE, E.; DOHERTY, G. M.; GARROD, S.; ISARD, S.; KOWTKO, J.; MCALLISTER, J.; MILLER, J.; SOTILLO, C.; THOMPSON, H. S.; WEINERT, R. (1991): The HCRC Map Task Corpus. *Language and Speech* 34, 351-366.
- AUER, P.; GILLES, P.; PETERS, J.; SELTING, M. (2000): Intonation regionaler Varietäten des Deutschen. Vorstellung eines Forschungsprojekts. In: D. Stellmacher (Hrsg.): *Dialektologie zwischen Tradition und Neuansätzen. Beiträge der Internationalen Dialektologentagung*. Stuttgart: Steiner, 222-239.
- BARRY, W. J.; ANDREEVA, B.; RUSSO, M.; DIMITROVA, S.; KOSTADINOVA, T. (2003): Do Rhythm Measures Tell us Anything about Language Type? *Proc. 15th ICPhS, Barcelona, Spain*, 2693–2696.
- BAUMANN, S.; GRICE, M.; STEINDAMM, S. (2006): Prosodic marking of focus domains – categorical or gradient? *Proc. 3rd international conference of speech prosody, Dresden, Germany*, 301-304.
- BAUMANN, S.; BECKER, J.; GRICE, M.; MÜCKE, D. (2007): Tonal and articulatory marking of focus in German. *Proc. 16th ICPhS, Saarbrücken, Germany*, 1029-1032.
- BOERSMA, P. (2001): Praat, a system for doing phonetics by computer. *Glott International* 5, 341-345.
- BEZOOIJEN, R. A. M. G. VAN (1984): *Characteristics and recognizability of vocal expressions of emotions*. Dordrecht: Foris.
- BOLINGER, D. (1978): Intonation across languages. In J. H. Greenberg (Hrsg.): *Universals of human language, Phonology V. 2*. Stanford: Stanford University Press, 471-523.
- BRINK, J.; WRIGHT, R.; PISONI, D. B. (1998): Eliciting speech reduction in the laboratory: Assessment of a new experimental method. *Research on Spoken Language Processing* 22, 395-420.
- BRINCKMANN, CAREN; BENZMÜLLER, RALF (1999): The Relationship between Utterance Type and F0 Contour in German. *Proceedings 6th European Conference on Speech Communication and Technology (EUROSPEECH)*, Budapest, Hungary, 21-24.
- BROWN, G.; CURRIE, K.; KENWORTHY, J. (1980): *Questions of Intonation*. London: Croom Helm.
- CHASAIDE, AILBHE NÍ; GOBL, CHRISTER (2004): Voice Quality and *f*₀ in Prosody: Towards a Holistic Account. *Talk. Proc. 2nd International Conference on Speech Prosody, Nara, Japan*, 189-196.
- CHEN, AOJU (2005): Universal and language-specific perception of paralinguistic

intonational meaning. Doktorarbeit, Radboud University Nijmegen. Utrecht: LOT.

COHEN, A.; 'T HART, J. (1967): On the anatomy of intonation. *Lingua* 19, 177–192.

CRUTTENDEN, ALAN (1990): The origin of nucleus. *Journal of International Phonetic Association* 20, 1-8.

DELLWO, VOLKER; FERRAGNE, EMMANUEL; PELLEGRINO, FRANÇOIS (2006): The perception of intended speech rate in English, French, and German by French speakers. *Proc. 3rd International Conference of Speech Prosody, Dresden, Germany*, 101-104.

DELLWO, VOLKER; WAGNER, PETRA (2003): Relations between language rhythm and speech rate. *Proc. 15th ICPhS, Barcelona, Spain*, 471-474.

D'IMPERIO, M. (2000): The Role of Perception in Defining Tonal Targets and their Alignment. Doktorarbeit, The Ohio State University, OH, USA.

FÉRY, CAROLINE (1993): German Intonational patterns. Tübingen: Niemeyer.

GARTENBERG, ROBERT; PANZLAFF-REUTER, C. (1991): Production and perception of f0 peak patterns in German. *Kiel: IPDS, AIPUK* 25, 29-115.

GILLES, PETER (2005): Regionale Prosodie im Deutschen. Variabilität in der Intonation von Abschluss und Weiterweisung. de Gruyter: Berlin/New York.

GRABE, ESTHER (2004): Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles, J. Peters, J. (Hrsg.): *Regional Variation in Intonation. Linguistische Arbeiten*. Tuebingen, Niemeyer, 9-31.

GRICE, MARTINE; R. BENZMÜLLER; M. SAVINO; B. ANDREEVA (1995): The intonation of queries and checks across languages: data from map task dialogues. *Proc. 13th ICPhS, Stockholm, Sweden*, 648-651.

GRICE, MARTINE; SAVINO, MICHELINA (1997): Can pitch accent type convey information status in yes-no questions? *Proc. ACL97 Workshop on Concept-to-Speech Generation Systems, Madrid, Spain*, 29-38.

GRICE, MARTINE; BAUMANN, STEFAN (2002): Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267-298.

GRIESBACH, HEINZ (2002): *Bauplan Deutsch. Übungsgrammatik mit Satzbauhelfer*. München: Klett.

GUSSENHOVEN, CARLOS (1984): On the grammar and semantics of sentence accents. Dordrecht: Foris.

GUSSENHOVEN, CARLOS (2002): Intonation and interpretation: Phonetics and Phonology. *Proc. 1st International Conference of Speech Prosody, Aix-en-Provence, France*, 47-57.

HAAN, J. (2002): *Speaking of Questions: An Exploration of Dutch Question Intonation*. Doktorarbeit, Graduate School of Linguistics, Utrecht: LOT.

HALLIDAY, M.A.K. (1967): *Intonation and Grammar in British English*. The Hague: Mouton.

HENTSCHEL, ELKE; WEYDT, HARALD (2003): *Handbuch der deutschen Grammatik*. Berlin/New York: deGruyter.

HERMES, ANNE; BECKER, J.; MÜCKE, D.; BAUMANN, S.; GRICE, M. (2008): Articulatory gestures and focus marking in German. *Proc. 4th International Conference of Speech Prosody, Campinas, Brazil*: 457-460.

HIRSCHBERG, JULIA (2000): A corpus-based approach to the study of speaking style. In: M. Horne (Hrsg.): *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*. Dordrecht: Kluwer Academic Publishers, 335–350.

HOUSE, DAVID (1990): Tonal perception in speech. *Travaux de l'institute de linguistique de Lund* 24, 7-163.

ISACENKO, A.V.; SCHÄDLICH, H.-J. (1970): *Untersuchungen über die deutsche*

- Satzintonation. Berlin: Akademie-Verlag.
- OHALA, JOHN J. (1973): The physiology of tone. *Southern California Occasional Papers in Linguistics*. 1, 1-14.
- IRIONDO, I.; PLANET, S.; CLAUDI SOCORÓ, J.; ALÍAS, F.; MONZO, C.; & MARTÍNEZ, E. (2007): Expressive speech corpus validation by mapping subjective perception to automatic classification based on prosody and voice quality. *Proc. 16th ICPhS, Saarbrücken, Germany*, 2125-2128.
- KEATING, P.A.; EPOSITO, CH. (2007): Linguistic Voice Quality. *UCLA Working Papers in Phonetics* 105, 85-91.
- KEHREIN, ROLAND (2002): The prosody of authentic emotions. *Proc. 1st International Conference of Speech Prosody, Aix-en-Provence, France*, 423-426.
- KLATT, DENNIS. H.; COOPER, W.E. (1975): Perception of segment duration in sentence contexts. In A. Cohen, S.G. Nooteboom (Hrsg.): *Structure and Process in Speech Perception*. Springer: New York, 69-89.
- KLATT, DENNIS H.; KLATT, LAURA C. (1990): Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87, 820-857.
- KLEBER, FELICITAS; RATHCKE, TAMARA (2008): More on the “segmental anchoring” of prenuclear rises: Evidence from East Middle German. *Proc. Speech Prosody 2008, Campinas, Brazil*, 123–126.
- KOHLER, KLAUS J. (1995): *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.
- KOHLER, KLAUS J.; PÄTZOLD, MATTHIAS; SIMPSON, ADRIAN P. (1997): From the acoustic data collection to a labelled speech data bank of spoken Standard German. Kiel: IPDS, AIPUK 32.
- KOHLER, KLAUS J. (1998). The disappearance of words in connected speech. *ZAS Working Papers in Linguistics* 11, 21-34.
- KOHLER, KLAUS J. (2004): Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions. In: G. Fant / H. Fujisaki / J. Cao / Y. Xu (Hrsg.): *From traditional phonology to modern speech processing – Festschrift for Professor Wu Zongji’s 95th Birthday*. Beijing: Foreign Language Teaching and Research Press, 205-215.
- KOHLER, KLAUS J. (2006): Paradigms of experimental prosodic analysis: from measurement to function. In S. Sudhoff, Lenertová, Meyer, Pappert, Augurzky, Mleinek, Richter (Hrsg.): *Methods in Empirical Prosody Research*. Berlin/New York: De Gruyter, 132-152.
- KOHLER, KLAUS J.; NIEBUHR, OLIVER (2007): The phonetics of emphasis. *Proc. 16th ICPhS, Saarbrücken, Germany*, 2145-2148.
- KOREMAN, JACQUE (2006): Perceived speech rate: The effects of articulation and speaking style in spontaneous speech. *Journal of the Acoustical Society of America* 119, 582-596.
- KÜGLER, FRANK. (2004): Do we know the answer? Variation in yes-no-question intonation. In: S. Fischer / R. Van de Vijver / R. Vogel (Hrsg.): *Experimental studies in linguistics 1*. Potsdam: Potsdam University Press, 9-29.
- KÜGLER, FRANK. (2008). The role of duration as a phonetic correlate of focus. *Proc. 4th International Conference on Speech Prosody, Campinas, Brazil*, 591-594.
- LADD, D. ROBERT (1996): *Intonational Phonology*. Cambridge: CUP.
- LEEMANN, ADRIAN; SIEBENHAAR, BEAT (2007): Intonational and temporal features of Swiss German. *Proc. 16th ICPhS. Saarbrücken, Germany*, 957-960.
- LIEBERMAN, P. (1967): *Intonation, perception, and language*. Cambridge: MIT Press.

- LINDAU, M. (1986): Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America* 80, 757-764.
- MENGE, HEINZ H. (1982): Was ist Umgangssprache? Vorschläge zur Behandlung einer lästigen Frage. *Zeitschrift für Dialektologie und Linguistik* 40, 52-63.
- MICHAUD, ALEXIS; MAZAUDON, MARTINE (2006): Pitch and voice quality characteristics of the lexical word-tones of Tamang, as compared with level tones (Naxi data) and pitch-plus-voice-quality tones. *Proc. 3rd Speech Prosody 2006, Dresden Germany*, 823-826.
- MIXDORFF, HANSJÖRG; PFITZINGER, HARTMUT R. (2005): Analysing fundamental frequency contours and local speech rate in map task dialogs. *Speech Communication* 46, 310-325.
- NIEBUHR, O. (2007a): *Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und Empirische Untersuchungen*. Berlin/New York: De Gruyter.
- NIEBUHR, O. (2007b): The signalling of German rising-falling intonation categories - The interplay of synchronization, shape, and height. *Phonetica* 64, 174-193.
- NIEBUHR, OLIVER. (2009): Intonation segments and segmental intonations. *Proc. Interspeech 2009, Brighton, UK*, 2435-2438.
- NIEBUHR, OLIVER. (2010): On the phonetics of intensifying emphasis in German. *Phonetica* 67, 170-198.
- OHALA, JOHN. J. (1983): Cross-language use of pitch: An ethological view. *Phonetica* 40, 1-18.
- PETERS, JÖRG (2004): *Intonatorische Variation im Deutschen. Studien zu ausgewählten Regionalsprachen*. Habilitationsschrift Universität Potsdam.
- PETERS, JÖRG (2005): Intonation. In: *Duden. Die Grammatik*. Mannheim: Dudenverlag, 95-128.
- PETERS, JÖRG (2006): Dialektintonation. *Osnabrücker Beiträge zur Sprachtheorie* 71, 177-201.
- PETRONE, CATERINA. (2008) : *Le rôle de la variabilité phonétique dans la représentation des contours intonatifs et de leur sense*. Doktorarbeit, Université de Provence, France.
- PETRONE, CATERINA.; NIEBUHR, OLIVER. (2009): The role of the prenuclear F0 region in the perception of German questions and statements. *Proc. Conference on Phonetics and Phonology in Iberia (PaPI), Las Palmas, Spain*. www.ipds.uni-kiel.de/on/downloads/Petrone_Niebuhr.ppt.
- PFITZINGER, HARTMUT R. (1999): Local Speech Rate Perception In German Speech. *Proc. 14th ICPhS, San Francisco, USA*, 893-896.
- PIERREHUMBERT, JANET; HIRSCHBERG, JULIA (1990): The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, M. Pollack (Hrsg.): *Intentions in Communication*. Cambridge: MIT Press, 271-311.
- PRINS, R.; BASTIAANSE, R. (2004): Analysing the spontaneous speech of aphasic speakers. *Aphasiology* 18, 1075-1091.
- RAMUS, FRANCK; NESPOR, MARINA; MEHLER, JACQUES (1999): Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.
- ROACH, PETER (1998): Some languages are spoken more quickly than others. In: L. Bauer. / P. Trudgill (Hrsg.): *Language Myth*. London: Penguin, 150-158.
- ROST-ROTH, MARTINA (2003): Fragen-Nachfragen-Echofragen. *Formen und Funktionen von Interrogationen im gesprochenen Deutsch*. *Linguistik Online* 13, 1-43.
- ROST-ROTH, MARTINA (2006): *Nachfragen. Formen und Funktionen äußerungsbezogener Interrogationen*. Berlin/New York: DeGruyter.
- SCHÄDLICH, HANS-JOACHIM; ERAS, HEINRICH (1970): *Vergleichende Untersuchungen über*

die Satzintonation in deutschen Dialekten. Proc. 6th ICPHS, Prague, Czech Republic, 793-796.

SELTING, MARGRET (1995): Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation. Tübingen: Niemeyer.

SELTING, MARGRET (2005): Variation der Intonation: Unterschiede zwischen Standard und Stadtsprache am Beispiel des Berlinerischen. In: L.M. Eichinger / W. Kallmeyer (Hrsg.): Standardvariation. Wie viel Variation verträgt die deutsche Sprache? Berlin, New York: DeGruyter, 247–277.

SHEN, XAIO-NAN SUSAN (1990): The prosody of Mandarin Chinese. Berkeley: UCLA Press

SIEBS, THEODOR (1969): Deutsche Aussprache. Reine und gemäßigte Hochlautung mit Aussprachewörterbuch. Berlin/New York: de Gruyter.

SIMPSON, ADRIAN P. (1998): Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. Kiel: IPDS, AIPUK 33.

SIMPSON, ADRIAN P. (2009): Breathiness differences in male and female speech. Is H1-H2 an appropriate measure? In: P. Branderud / H. Traunmüller (Hrsg.): Proceedings of Fonetik, Stockholm, Sweden, 172-175.

STOCK, EBERHARD., ZACHARIAS, C. (1973): Deutsche Satzintonation. Leipzig: VEB Verlag Enzyklopädie.

SWERTS, MARC; STRUNGERT, EVA; HELDNER, MATTIAS (1996): F0 Declination in read-aloud and spontaneous Speech. Proc. International Conference of Spoken Language Processing, Rhodes, Greece, 1501-1504.

SWERTS, M.; VELDHUIS, R. (2001): The effect of speech melody on voice quality. Speech Communication 33, 297-303.

SYRDAL, ANN K.; MCGORY, JULIA (2000): Inter-Transcriber reliability of ToBI prosodic labeling. Proc. International Conference of Spoken Language Processing, 235-238.

THORSEN, NINA (1980): A study of the perception of sentence intonation: Evidence from Dutch. Journal of the Acoustical Society of America 67, 1014-1030.

VAN BORSEL, JOHN; DE MAESSCHALC, DOROTHY. (2008): Speech rate in males, females, and male-to-female transsexuals. Clinical Linguistics & Phonetics 22, 679-685.

VON ESSEN, OTTO (1964): Grundzüge der hochdeutschen Satzintonation. Ratingen: Hehn.

VAN HEUVEN, VINCENT; VAN ZANTEN, E. (2005): Speech rate as a secondary prosodic characteristic of polarity questions in three languages. Speech Communication, 47, 87-99.

WAGENER, PETER (1986): Sind Spracherhebungen paradox? Über die Möglichkeit, natürliches Sprachverhalten wissenschaftlich zu erfassen. In: A. Schöne (Hrsg.): Akten des VII. IVG-Kongresses. Göttingen 1985, Band 4. Tübingen: Niemeyer, 319-327.

WELLS, JOHN C. (1997): SAMPA computer readable phonetic alphabet. In: D. Gibbon / R. Moore / R. Winski (Hrsg.): Handbook of Standards and Resources for Spoken Language Systems. Berlin/New York: deGruyter.

ZENG, X.; MARTIN, P.; BOULAKIA, G. (2004): Tones and Intonation in Declarative and Interrogative Sentences in Mandarin. Proc. International Conference on Tonal Aspects of Languages, Beijing, China, 235-238.

ZWIRNER, E. (1964): Anleitung zu sprachwissenschaftlichen Tonbandaufnahmen.

Lautbibliothek der Deutschen Mundarten, Band 31. Göttingen: Vandenhoeck & Ruprecht.

Kurzzusammenfassung

Intonationsfragen sind definiert als Fragen, die nicht syntaktisch oder lexikalisch, sondern phonetisch markiert sind. Die Markierung ist in Form eines Intonationsanstiegs am Satzende konzeptualisiert. Beobachtungen in spontansprachlichen Dialogen, in denen auch deklarative Äußerungen mit satzfinal fallender Intonation als Fragen verwendet und erkannt wurden, widersprechen dem Formkonzept der Intonationsfrage. Angesichts dessen wurde dieser deklarative Fragetyp, dessen Kernfunktion die Nachfrage ist, in der vorliegenden Studie einer detaillierten phonetischen Analyse unterzogen, die sowohl sachlich als auch emphatisch geäußerte Sätze berücksichtigte und auf gezielt erhobenen lesesprachlichen Mono- und Dialogen basierte. Die Ergebnisse zeigen, dass sich die elizitierten Nachfragen von vergleichbaren deklarativen Aussagesätzen – abgesehen von der Richtung der satzfinalen Intonationsbewegung – dadurch unterscheiden, dass die Nachfragen global schneller und behauchter, mit weniger prä nuklearen Akzenten und am Satzanfang mit einer tieferen, eingeebneten Intonation realisiert wurden. Diese multifaktoriellen phonetischen Unterschiede galten jedoch nicht absolut, sondern zwischen gleichermaßen sachlichen oder emphatischen Aussage- und Fragesätzen, was den zentralen Stellenwert einer kontextbezogenen Interpretation phonetischer Muster unterstreicht. Trotz ausstehender Wahrnehmungsexperimente muss das Formkonzept der Intonationsfrage als inadäquat bezeichnet werden, sowohl hinsichtlich seiner Beschränkung auf Intonation als auch hinsichtlich seiner Beschränkung auf das Satzende.

Kontaktinformationen des Hauptautors:

Prof. Dr. Oliver Niebuhr
Seminar für Allgemeine und Vergleichende Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
Leibnizstraße 10
D-24098 Kiel

Tel: +49 431 880 3314

Fax: +49 431 880 7405

Email: niebuhr@linguistik.uni-kiel.de