

**Weiterführende  
Untersuchungen  
zu prosodischen Grenzen  
in deutscher Spontansprache**

Benno Peters



## 1 Einleitung

In der sprachlichen Kommunikation wird die syntaktische, semantische und pragmatische Gliederung des Gesprochenen durch eine Hierarchie syntagmatischer prosodischer Gruppierungen (prosodische Phrasierung) vermittelt, indem Sprecher an bestimmten Positionen ihrer Äußerung phonetisch gesteuerte Einschnitte (prosodische Grenzen) erzeugen. Prosodische Grenzen werden durch unterschiedliche Bündelungen der phonetischen Faktoren Satzakkentuierung, Melodiemuster, Phonationstyp und Zeitstrukturierung getragen. Hierbei gibt es eine Beziehung zwischen der Stärke semantisch-pragmatischer Einschnitte und der phonetischen Manifestation prosodischer Grenzen. Die Ausprägung prosodischer Grenzen steht auch in enger Verbindung zur interaktiven Dialogstruktur. So können Sprecher durch phonetische Mittel ihre Bereitschaft anzeigen, ihren Redebeitrag (*Turn*) zu beenden und dem Hörer das Wort zu übergeben. Ein dritter, funktional abzugrenzender Bereich, in dem prosodische Grenzen vorkommen, ist der Bereich der unflüssigen Sprechweise. Hier spiegeln die prosodischen Grenzen den Planungsprozess auf der Sprecherseite wider. In allen Fällen muss der Hörer die vom Sprecher produzierten prosodischen Grenzen anhand ihrer phonetischen Struktur interpretieren und funktional zuordnen. Prosodische Grenzen sind somit entscheidend für den Erfolg sprachlicher Interaktion zwischen Sprecher und Hörer.

Die funktionale Zuordnung anhand phonetischer Kriterien wird einerseits über Produktionsuntersuchungen (sprecherbezogen) und andererseits über Perzeptionsuntersuchungen (hörerbezogen) analysiert. Zentraler Untersuchungsgegenstand des vorliegenden Projekts ist die phonetische Extensionalisierung und Skalierung der prosodischen Kohäsion bzw. Separierung aufeinander folgender sprachlicher Phrasierungseinheiten, einschließlich Redebeiträgen bei Sprecherwechsel und prosodischen Gruppierungen in unflüssiger Sprache. Ziele sind die Modellierung und die automatische subklassifizierende Annotation der prosodischen Phrasierung. Ausgangspunkt dieser Untersuchung ist die in Peters et al. 2005a erarbeitete Korpusanalyse zu Vorkommen und Bündelung folgender phonetischer Merkmale an prosodischen Grenzen:

- Pausen und Atmungsphasen,
- melodische Verläufe,

- F0-Reset,
- segmentelle Längung.

Auf dieser Basis ist für prosodische Grenzen in flüssiger Sprache folgende Einteilung vorgenommen worden:

**Phrasengrenze 1 (PG1):** Die phonetische Separierung an der prosodischen Grenze wird durch temporale Merkmale (segmentelle Längung) und/oder tonale Merkmale hervorgerufen (Reset und/oder starke melodische Bewegung am Phrasenende). Es tritt keine Pause auf.

**Phrasengrenze 2 (PG2):** An der prosodischen Grenze tritt eine kurze oder mittlere Pause oder eine Atmungsphase (unter 700 ms) auf.

**Phrasengrenze 3 (PG3):** Die prosodische Grenze weist eine lange Pause oder Atmungsphase (über 700 ms) auf. Tonale Merkmale (Reset und/oder starke melodische Bewegung am Phrasenende) sind nicht vorhanden.

**Phrasengrenze 4 (PG4):** Die prosodische Grenze ist mit einer langen Pause oder Atmungsphase (über 700 ms) und zusätzlich mit mindestens einem tonalen Merkmal realisiert.

Diese vier Grenztypen zeigen einen von PG1 zu PG4 aufsteigenden Grad phonetischer Separierung, kodiert durch Merkmalsbündel, die in Anzahl und Stärke der phonetischen Merkmale variieren. Peters et al. 2005a belegen anhand der inhaltlichen Interpretation einzelner Korpusbeispiele, dass der zunehmende Grad der phonetischen Separierung der Klassen PG1-PG4 an prosodischen Grenzen in vielen Fällen mit einem zunehmenden Grad inhaltlicher Separierung zwischen benachbarten Phrasen einhergeht.

Ziel der vorliegenden Arbeit ist die empirische Validierung und eventuelle Modifikation dieses Klassifikationssystems. Das erfordert zum einen korpusbasierte experimentelle Untersuchungen des Zusammenhangs zwischen den Klassen PG1-PG4 und einer aufsteigenden perceptiven Trennung. Zum anderen wird evaluiert, ob die Klassen PG1 bis PG4 in systematischer Verbindung zu einer zunehmenden inhaltlichen Separierung in spontaner Sprache stehen. Diese Evaluation dient dazu, Divergenzen und Konvergenzen zwischen inhaltlicher und phonetischer Kohä-

sion bzw. Separierung zu belegen und das Klassifikationssystem entsprechend zu optimieren.

Diese Forschungsziele schließen folgende Arbeitsschritte ein:

- Ermittlung der Häufigkeitsverteilung über die Klassen PG1-PG4 im untersuchten Korpus
- Perzeptorische Skalierung phonetisch signalisierter Kohäsion/Separierung in Wahrnehmungsexperimenten. Es werden in einem Resyntheseverfahren die phonetischen Merkmalsbündel an prosodischen Grenzen systematisch in dem Maße variiert, in dem sich phonetische Variation in den Korpusdaten zeigt. Versuchspersonen beurteilen den Grad der perzeptorischen Separierung, den phonetisch unterschiedliche Merkmalsbündel hervorrufen.
- Interpretative Überprüfung des Zusammenhangs zwischen phonetischer und inhaltlicher Kohäsion bzw. Separierung.

Diese Arbeitsschritte beziehen sich auf den Phänomenbereich der flüssigen Phrasengrenzen im Inneren von Redebeiträgen. Aber auch das Ende jedes Redebeitrags in einem Gespräch ist eine prosodische Grenze, die durch phonetische Merkmalsbündel gekennzeichnet ist. Weitere Untersuchungen umfassen deshalb die Abgrenzung zwischen turninternen und turnfinalen prosodischen Grenzen anhand phonetischer und funktionaler Kriterien. Prosodische Grenzen an Turnübergängen werden phonetisch analysiert und hinsichtlich ihrer interaktionalen Bedeutung differenziert. Drei Typen interaktional unterschiedlicher Sprecherwechsel werden herausgearbeitet: 'gezielte Sprecherwechsel', 'akzeptierte Sprecherwechsel' und 'erzwungene Sprecherwechsel'.

Die Analyse und Klassifikation der turnfinalen prosodischen Grenzen umfasst folgende Arbeitsschritte:

- Labelbasierte Zusammenstellung der phonetischen Merkmale von Phrasengrenzen bei Sprecherwechseln unterschiedlichen Typs.
- Perzeptionsexperimente zur Bestimmung der interaktionalen Bedeutung der Parameter 'melodischer Verlauf' und 'Längung' in turnfinaler Position.

- Funktionale Interpretation von Korpusbeispielen und Herstellung des Bezugs zwischen den gefundenen syntaktischen und inhaltlichen Strukturen und der phonetischen Realisierung turnfinaler prosodischer Grenzen.

Der letzte Untersuchungsbereich, der im Rahmen der Analyse prosodischer Grenzen in spontaner Sprache untersucht wird, sind prosodische Grenzen, die im Zusammenhang mit Störungen der Sprechflüssigkeit stehen. Grenzen dieser Art sind in Spontansprache häufig und spiegeln mehr den kognitiven Planungsprozess seitens des Sprechers wider als die inhaltliche und syntaktische Struktur von Äußerungen. Ziel ist eine funktionale Abgrenzung flüssiger und unflüssiger prosodischer Grenzen, wobei wiederum über Korpusanalysen funktionale Strukturen mit phonetischen Parametern verbunden werden. Eine zentrale Rolle in der phonetischen Signalisierung prosodischer Grenzen in unflüssiger Sprache spielt die segmentelle Dauerstruktur. Daraus ergeben sich folgende Untersuchungen:

- Labelbasierte Zusammenstellung der phonetischen Merkmale von prosodischen Grenzen in unflüssiger Sprache.
- Perzeptorische Überprüfung der Wahrnehmung und der funktionalen Einordnung von segmenteller Längung bei der Signalisierung von Produktionsstörungen.

Die Integration der funktional verschiedenen Bereiche, in denen prosodische Grenzen vorkommen, und die Kombination aus sprecher- und hörerbezogenen Untersuchungsmethoden führen letztlich zu einem komplexen Überblick über den Phänomenbereich der prosodischen Grenzen.

## **2 Methoden**

Dieser Abschnitt führt kurz in die grundlegenden methodischen Anforderungen der Arbeit und die gewählten methodischen Verfahren ein, die zur Anwendung kommen. Detaillierte Beschreibungen finden sich in den Methodenteilen der nachfolgenden Kapitel.

Die Zielsetzung der Arbeit führt zu drei Arbeitsbereichen, für die methodische Verfahren entwickelt werden:

- (1) Phonetische Korrelate prosodischer Grenzen werden über Häufigkeitsverteilungen beschrieben. Einige der untersuchten Merkmale können direkt aus der prosodischen Etikettierung übernommen werden, andere werden durch Messverfahren aus der segmentellen Etikettierung abgeleitet. Die Häufigkeitsverteilungen der Merkmale können herangezogen werden, um drei Grenztypen auf phonetischer Basis voneinander zu differenzieren: turninterne flüssige Grenzen, turninterne unflüssige Grenzen und turnfinale Grenzen. Weiterhin werden innerhalb dieser Grenztypen Subklassen gebildet. Die Analysen basieren auf Produktionsdaten aus spontansprachlichen Gesprächen und beziehen verschiedene phonetisch annotierte Korpora ein (Peters 2005a).
- (2) Beziehungen zwischen der phonetischen Realisierung prosodischer Grenzen und sprachlichen Funktionen werden anhand einer interpretativen Korpusanalyse untersucht.
- (3) Perzeptorische Aspekte phonetischer Separierung und deren semantisch-pragmatische und interaktionale Funktion werden experimentell untersucht.

Diese Ziele sind sehr unterschiedlich und nur über die Kombination verschiedener Methoden erreichbar. Für den ersten Bereich wird das unter Peters 2005a beschriebene Korpus verschiedener spontansprachlicher Dialoge herangezogen. Die prosodische Etikettierung der untersuchten Dialoge umfasst etwa 9000 Kennzeichnungen prosodischer Grenzen. Für die statistische Beschreibung der phonetischen Merkmale von Phrasengrenzen wird anhand der Etikettierung an jeder etikettierten Grenze durch automatische Datenbankabfragen ein Bündel der vorhandenen phonetischen Phrasierungsmerkmale abgeleitet. Die phonetischen Parameter prosodischer Grenzen, die aus der Etikettierung gewonnen werden können, haben unterschiedlichen Status: einige Merkmale (z.B. melodische Konturen oder F0-Reset) stammen aus der prosodischen Etikettierungsebene und gehen auf perzeptiv orientierte Beurteilungen durch trainierte Etikettierer zurück. Auf andere Parameter (z.B. Pausen und Atmungsphasen) kann direkt über die segmentelle Etikettierung zugegriffen werden. Die Analyse der finalen Längung vor prosodischen Grenzen hingegen beruht auf einem automatisierten mathematischen Verfahren, das anhand der segmentellen Etikettierung für jede prosodische Grenze den Grad der finalen Längung ableitet.

Die Korpusuntersuchung liefert zunächst einen Überblick über die Art und Häufigkeit einzelner phonetischer Merkmale sowie deren Bündelungen an Phrasengrenzen. Es können aber anhand der automatisch abfragbaren Information auch schon strukturelle und funktionale Subklassifizierungen vorgenommen werden, da die Etikettierung Informationen über die Position einer Phrasengrenze im Dialog (turnintern vs. turnfinal) sowie über das Vorkommen von Produktionsstörungen (syntaktische Abbrüche, Zögerungslängung und Häsitationspartikeln sind gekennzeichnet) enthält.

Die Zuordnung bestimmter Bündelungen zur pragmatischen, semantischen und interaktionalen Struktur der Dialoge erfordert ein Hinausgehen über automatische Korpusabfragen, da auf diesem Weg nur wenig Information über die Äußerungs- und Dialogstruktur gewonnen werden kann. Es ist vielmehr notwendig, eine große Zahl von Beispielen aus dem Korpus herauszugreifen und in Bezug auf sprachliche Funktionen im Dialogkontext zu interpretieren. So können verschiedene Typen inhaltlicher und struktureller Grenzen in der Interpretation erfasst und phonetischen Ausprägungen prosodischer Grenzen zugeordnet werden.

Die Kombination von Korpusabfrage und Interpretation liefert noch keine Erkenntnisse über die Einordnung verschiedener Merkmalskombinationen auf perzeptorischen Skalen zwischen Kohäsion und Separierung. Deshalb werden für bestimmte Merkmalskombinationen, die sich im Korpus zeigen, Perzeptionsexperimente zur Wahrnehmung und Funktion dieser phonetischen Muster durchgeführt. Diese Experimente liefern Erkenntnisse über die Perzeption von Pausen verschiedener Länge und ihre perzeptorische Interaktion mit anderen Merkmalen oder beziehen sich auf den perzeptorischen Zusammenhang zwischen melodischen Mustern und segmenteller Längung an prosodischen Grenzen und bei Sprecherwechsel im Dialog.

Die Entwicklung der Perzeptionsexperimente steht in enger Verbindung mit den Erkenntnissen aus der Korpusanalyse (automatische Abfragen und Interpretation), so dass sich die Vorgehensweisen und die Ergebnisse der drei methodischen Stränge gegenseitig ergänzen. Die Forschungsergebnisse der methodischen Stränge sollen am Ende ein Gesamtbild des Phänomenbereichs ergeben.

Da die Experimente auf die Aufdeckung von Form-Funktionszusammenhängen ausgerichtet sind, ist es nicht sinnvoll, die Ergebnisse lediglich über die gefundenen Signifikanzen (z.B. auf dem 5%-Niveau) zu beschreiben, denn das gewählte Messverfahren ist sehr sensitiv, und schon bei kleinen Differenzen zwischen Mittelwerten ergeben sich signifikante

Unterschiede. Es ist jedoch zu beachten, dass ein kleiner Unterschied in der Beurteilung phonetisch unterschiedlicher Stimuli nicht unbedingt mit funktional unterschiedlichen Klassen einhergeht. Häufig sind über die Stimulusserien kontinuierlich steigende oder fallende Urteile zu beobachten, die sich von Stimulus zu Stimulus signifikant unterscheiden. Dennoch ist bei kleinen messbaren Unterschieden nicht automatisch eine funktionale Relevanz gegeben. Hieraus ergibt sich folgendes Vorgehen.

In der Auswertung der intervallskalierten Daten aus den Perzeptionsexperimenten kommt eine Kombination deskriptiver und prüfstatistischer Verfahren zum Einsatz, die eine übersichtliche Präsentation der experimentellen Ergebnisse ermöglicht und Interpretierbarkeit gewährleistet. In deskriptiven Statistiken (meist Darstellungen von Mittelwerten) können klar und übersichtlich die Auswirkungen der Parametermanipulationen auf die Urteile der Versuchspersonen herausgearbeitet werden. Diese Art der Darstellung ist besonders geeignet, verschiedene Effektstärken aufzuzeigen und starke Effekte von schwachen Effekten zu trennen. Ergänzend wird für jedes Experiment eine mehrfaktorielle Varianzanalyse (ANOVA) mit Messwiederholung berechnet, um zu einer prüfstatistischen Absicherung der Beobachtungen zu gelangen. In das prüfstatistische Verfahren gehen aufgrund der Struktur der Experimente lediglich Messwiederholungsfaktoren ein. Notwendige Freiheitsgradkorrekturen werden nach Greenhouse-Geisser (1959) vorgenommen. Bei einem Niveau unter 5% Fehlerwahrscheinlichkeit wird der Terminus 'signifikant' verwendet, bei einem Niveau unter 1% der Terminus 'hochsignifikant'.

### 3 Datenbasis

Folgende Daten gehen in die Korpusanalyse ein:

- das spontansprachliche Korpus der CD-ROM Vol. 2 und Vol. 3 (IPDS 1995, 1996) aus dem Verbmobil-Szenario der Terminabsprache: 77 Dialoge aus 11 Dialogsitzungen mit 22 Sprechern (13 männlich, 9 weiblich), 2,5 Stunden, 25.000 fortlaufende Wörter, 5100 etikettierte prosodische Grenzen.
- 6 überlappende Dialoge aus dem Video Task-Szenario (LINDENSTRASSE-Korpus), (Peters 2001, 2005a): 12 Sprecher, 4 männlich, 8 weiblich, 1,5 Stunden, 13.500 fortlaufende Wörter, 3900 etikettierte prosodische Grenzen. Diese Daten werden insbesondere für die Differenzierung zwischen flüssiger turninterner Phrasierung und Phrasierung an

Äußerungsgrenzen eingesetzt. Für die Daten des Terminabsprachekorpus wurde ein Aufnahmeverfahren benutzt, bei dem die Sprecher wechselseitig nur dann aufgenommen wurden, solange sie einen Knopf gedrückt hielten. Diese Daten sind deshalb bezüglich der Turnübergaben unnatürlich und werden in diesem Analysebereich nur unter Vorbehalt in die Analysen einbezogen.

Die gesamte Datenbasis liegt orthographienah verschriftet, manuell segmentiert und prosodisch annotiert vor. Für die prosodische Etikettierung wurde das Transkriptionssystem PROLAB (Kohler 1995, Peters und Kohler 2004) verwendet, das auf dem Kieler Intonationsmodell KIM (Kohler 1991a; Peters 1999) basiert. Weiterhin gibt es in den Daten aus dem Video Task-Szenario (LINDENSTRASSE-Korpus) zusätzliche Etikettierungen, die das Ende von Redebeiträgen und die Anfangs- und Endpositionen von überlagerndem Sprechen markieren.

## **4 Turninterne flüssige prosodische Grenzen**

### **4.1 Labelbasierte akustisch-phonetische Analyse der Produktionsdaten**

Dieser Abschnitt befasst sich mit der Verteilung der phonetischen Phrasierungssignale im Korpus und gibt anhand von Häufigkeitsverteilungen einen Überblick über Vorkommen und Ausprägung der Merkmale Pause/Atmen, melodischer Verlauf, segmentelle Längung, F<sub>0</sub>-Reset und Phonation. Zuerst wird ein Überblick über Häufigkeit und Ausprägung der einzelnen Merkmale gegeben, später über ihre Bündelungen. Es werden sowohl die Einzelmerkmale als auch deren Bündelungen dargestellt, da in der Literatur häufig Bezug auf einzelne phonetische Parameter genommen und weniger auf Merkmalskombinationen eingegangen wird und diese Arbeit auch einen Bezug zu diesen Publikationen herstellen soll. Weiterhin ist im ersten Schritt eine separate Darstellung der Einzelmerkmale und bestimmter funktionaler Aspekte isolierter phonetischer Strukturen übersichtlicher und führt z.T. zu klareren Ergebnissen als die Analyse der Bündelungen.

Im Rahmen der labelbasierten Abfrage der phonetischen Merkmale an prosodischen Grenzen werden sämtliche Ergebnisse lediglich durch deskriptive Statistiken dargestellt. Auf eine prüfstatistische Analyse wird verzichtet, da es hier nicht um den Nachweis geringfügiger Unterschiede

zwischen verschiedenen Verteilungen geht, sondern zunächst um eine globale Beschreibung der phonetischen Eigenschaften prosodischer Grenzen. Des Weiteren sollen die Verteilungen mit funktionalen Kategorien assoziiert werden, und dies erfordert so deutliche Unterschiede in den Verteilungen, dass das Belegen leichter, aber trotzdem signifikanter Unterschiede hier keinen Sinn macht und starke Unterschiede auch ohne Prüfstatistik deutlich sichtbar werden.

Die Analyse der Einzelmerkmale sowie ihrer Kombinationen ist die Basis für eine phonetisch motivierte Klassenbildung einerseits und andererseits der Ausgangspunkt für die Überführung der im Korpus beobachteten phonetischen Variation in Stimulusmaterial für Perzeptionsexperimente.

- Nicht für jede Phrasengrenze kann automatisch ein Merkmalsbündel abgeleitet werden, in dem Information über den melodischen Verlauf, die finale Längung, F0-Reset und Pause/Atmen vorhanden ist. Dies hat verschiedene Gründe.
  - (1) Melodische Verläufe am Phrasenende werden nur dann in der prosodischen Etikettierung erfasst, wenn die Phrase mindestens einen Satzakzent enthält. Im Korpus gibt es aber häufig abgebrochene Phrasen, die keinen Satzakzent enthalten. In diesen Fällen kann automatisch keine Information über den melodischen Verlauf an der prosodischen Grenze erfasst werden.
  - (2) Die Berechnung der finalen Längung wird über einen Vergleich von Segmentdauern im Phraseninneren und an Phrasengrenzen durchgeführt. Um einen aussagekräftigen Messwert abzuleiten, wird die Berechnung für jeden Segmenttyp und für jeden Sprecher getrennt durchgeführt und später durch ein Normalisierungsverfahren in einen segment- und sprecherunabhängigen Wert überführt. Wenn für einen Sprecher bestimmte Segmente in einer der beiden Positionen zu selten vorkommen, kann kein Wert für diese Segmente errechnet und somit keine Angabe zur Längung im finalen Reim gemacht werden.
  - (3) Bei bestimmten melodischen Konstellationen am Phrasenende werden im Rahmen der prosodischen Etikettierung keine Beurteilung eines nachfolgenden F0-Resets vorgenommen. Für diese prosodischen Grenzen fehlt folglich die Angabe über das Vorhan-

densein eines F0-Resets. Für die Gesamtdarstellung der Merkmalsbündel wird nur Bezug auf die prosodischen Grenzen genommen, für die alle Merkmale automatisch abgeleitet werden können.

Durch diese Einschränkung kommt es zu einer Anzahl von insgesamt 3399 flüssigen turninternen Phrasengrenzen aus beiden Korpora, für die automatisch ein komplettes Merkmalsbündel abgeleitet werden kann.

- Für einige Fragestellungen müssen zusätzliche Etikettierungen oder interpretative Beurteilungen vorgenommen werden. Wegen des großen damit verbundenen Arbeitsaufwands können nur Datenausschnitte diesem Verfahren unterzogen werden.

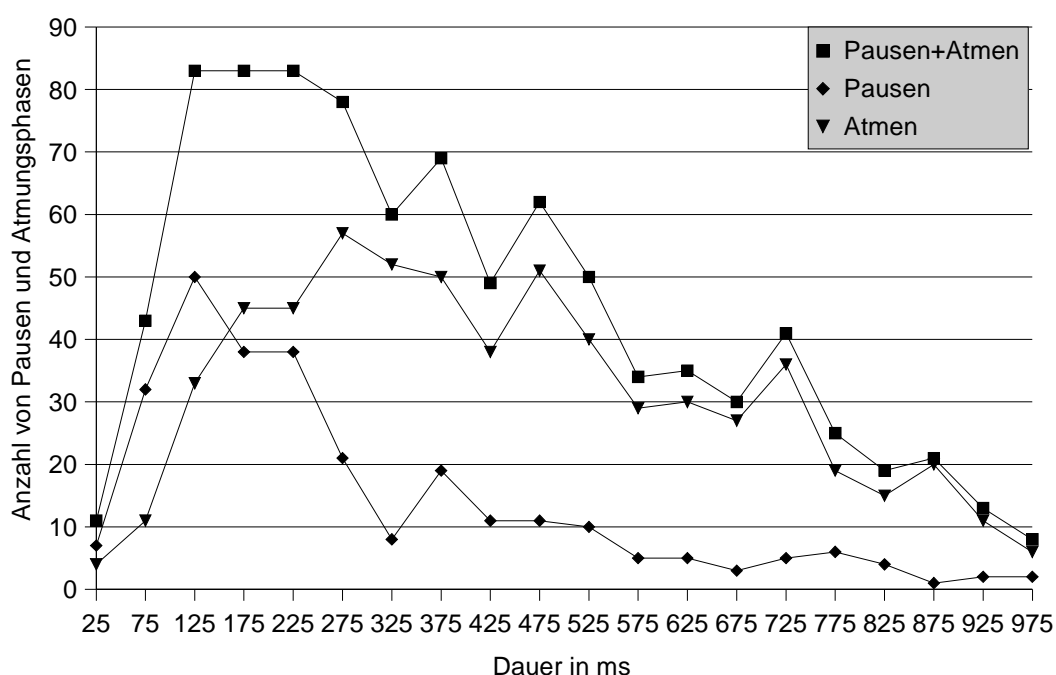
Entsprechend wird für jeden Bereich der Korpusanalyse angegeben, welcher Teil der Datengrundlage den Ausgangspunkt der jeweiligen Teiluntersuchung bildet.

#### **4.1.1 Pausen und Atmungsphasen an turninternen prosodischen Grenzen**

Pausen und Atmungsphasen werden anhand der Terminabsprachedaten untersucht. Es stellt sich die Frage, ob anhand der Analyse von Pausen- und Atmungsdauern eine Trennung in mehrere Klassen (z.B. lange vs. kurze Pausen) erreicht werden kann. Ein Hinweis auf eine solche Klassenbildung wäre eine bimodale Verteilungen von Pausen- und Atmungsdauern. Butcher (1981) weist eine solche Bimodalität in lesesprachlichen Daten nach.

Über die automatische Analyse der segmentellen Etikettierung ist keine eindeutige Trennung zwischen Pausen und Atmungsphasen zu erreichen. Dies hat zwei Gründe: Erstens kann ein als Atmen etikettierter Signalabschnitt neben einem hörbaren Atmungsgeräusch auch noch eine Pause enthalten und zweitens kann nicht-hörbares Atmen nicht anhand des akustischen Signals detektiert werden. So können als Pausen etikettierte Abschnitte auch mit nicht hörbarer Atmungsaktivität verbunden sein. Trotz dieser Einschränkungen wird im Rahmen der Untersuchungen jeder als Atmen etikettierte Abschnitt als 'Atmungsphase' angesehen und dessen Dauer bestimmt. Genauso werden etikettierte Pausen rein anhand der Etikettierung gemessen.

Abbildung 1 zeigt die Dauern der Pausen und Atmungsphasen im Terminabsprachekorpus. Es gehen 283 Pausen und 662 Atmungsphasen in die Darstellung ein. Die Pausen- und Atmungsdauern sind für die Abbildung in Klassen mit einer Breite von 50 ms eingeteilt. Die Zahlen auf der x-Achse zeigen den Mittelwert der jeweiligen Klasse. Für die Abbildung sind die oberen 5% der Werte ausgespart, da in diesem Bereich eine breite Streuung der wenigen Fälle bis zu Werten von 2000 ms vorliegt und diese die grafische Darstellung sehr unübersichtlich machen würden.



**Abbildung 1:** Verteilung der Pausen- und Atmungsdauern im Terminabsprachekorpus.

Folgende Kennwerte der Verteilungen der Pausen- und Atmungsdauern ergeben sich (hier inklusive der oberen 5% der Werte, Angaben in ms):

**Tabelle 1:** Kennwerte der Verteilung von Pausen- und Atmungsdauern im Terminabsprachekorpus (Angaben in ms).

	Mittelwert	Standardabweichung	Median	Modalwert (Klasse von)
<b>Pausen</b>	298	236	220	100-150
<b>Atmungsphasen</b>	509	334	268	250-300

Die Verteilungen von Pausen und Atmen zeigen weder einzeln noch zusammen eine deutliche Bimodalität, sondern zusammen eher ein breites Maximum zwischen 100 und 300 ms. Pausen sind durchschnittlich kürzer als Atmungsphasen. Diese Verteilungseigenschaften bieten keinen Ansatzpunkt für eine produktionsdatengestützte Einteilung in mehrere Klassen unterschiedlich langer Pausen- und Atmungsdauern.

#### **4.1.2 Pausen und Atmungsphasen im Inneren prosodischer Phrasen**

Wenn Aussagen über die separierende Wirkung phonetischer Parameter gemacht werden sollen, stellt sich die Frage, ob und wie häufig phonetische Merkmale, die sich an etikettierten prosodischen Grenzen finden, auch im Inneren prosodischer Phrasen vorkommen. Je häufiger dies für ein Merkmal der Fall ist, desto geringer ist dessen perzeptorisch separierende Wirkung einzuschätzen. In der Etikettierung des Korpus wurden prosodische Grenzen dort gesetzt, wo ein deutlich hörbarer phonetischer Einschnitt wahrgenommen wurde.

Im gesamten Korpus (Terminabsprachen und LINDENSTRASSE) finden sich 1957 etikettierte Pausen und 3821 Atmungsphasen. Davon gibt es nur 31 Pausen (1,6% aller Pausen) und 13 Atmungsphasen (0,3% aller Atmungsphasen), an denen keine Phrasengrenze etikettiert ist. Die auditive Prüfung dieser Fälle zeigt, dass es sich in den meisten Fällen um Etikettierungsfehler handelt, da Pause/Atmen auch in den nicht mit einer Phrasengrenze etikettierten Fällen einen deutlichen perzeptorischen Einschnitt bewirkt. Pausen und Atmen sind also perzeptiv stark hervortretende akustische Korrelate prosodischer Grenzen.

#### **4.1.3 Melodische Muster an turninternen prosodischen Grenzen**

In der PROLAB-Etikettierung werden 12 phrasenfinale melodische Muster unterschieden. Diese Konturen verlaufen von der letzten satzakzentuierten Silbe bis zum Phrasenende. In der Literatur werden einige dieser Konturen als besonders prominent hinsichtlich ihrer separierenden perzeptorischen Wirkung beschrieben. Hierbei handelt es sich in allen Fällen um Konturen mit einer starken melodischen Bewegung. Die Ursache für die separierende Wirkung dieser Konturen wird in zweierlei Eigenschaften gesehen. Zum einen kann eine tonal sehr umfangreiche melodische Bewegung selbst als separierendes Element angesehen werden (Peters et al. 2005a), zum anderen enden die als separierend eingestuften Konturen

im oberen oder im unteren Bereich der Sprechstimme. Der auf solche tonalen Bewegungen folgende Einsatz von F0 liegt aber bei den tief fallenden Konturen meist höher als der vorher erreichte Endpunkt der melodischen Bewegung und bei den hoch steigenden Konturen niedriger. Dadurch entstehen tonale Brüche, die zu einer perceptiven Separierung führen (Lehiste 1975, Couper-Kuhlen 1983, Batliner et al. 1995).

Insgesamt handelt es sich um 7 der 12 Intonationsverläufe, die in der weiteren Untersuchung als separierende tonale Merkmale eingestuft werden. Die übrigen 5 sind ebene oder leicht steigende oder leicht fallende Muster, die ohne das Hinzukommen weiterer Merkmale nicht zu einem perceptorischen Einschnitt führen. Aufgrund des seltenen Vorkommens sind die vier fallend-steigenden Muster, die in der Etikettierung differenziert werden, in der Untersuchung zusammengefasst.

Tabelle 2 zeigt die Verteilung der melodischen Muster, getrennt in separierende und nicht-separierende Konturen. Da sich die Distributionen in den beiden untersuchten Korpora stark ähneln, werden die Ergebnisse in einer gemeinsamen Tabelle zusammengefasst.

**Tabelle 2:** Melodische Muster am Phrasenende (Turninterne prosodische Grenzen, Terminabsprachen und LINDENSTRASSE, n=3399).

<b>Starke melodische Bewegung</b>	<b>Anzahl</b>	<b>%</b>
Stark fallend	915	26,9
Stark fallend + sehr leicht steigend	202	5,9
Fallend-steigend	121	3,6
Stark Steigend	65	1,9
Gesamt starke melodische Bewegung	1303	38,3
<b>Schwache melodische Bewegung</b>		
Eben	562	16,5
Leicht fallend	962	28,3
Leicht bis mittelhoch steigend	438	12,9
Eben + sehr leicht steigend	23	0,7
Leicht fallend + sehr leicht steigend	111	3,3
Gesamt schwache melodische Bewegung	2096	61,7
<b>Gesamt</b>	<b>3399</b>	<b>100</b>

#### 4.1.4 Melodische Muster im Inneren prosodischer Phrasen

Bei der Anwendung des Etikettierungssystems PROLAB auf Lese- und Spontansprache zeigte sich, dass ein Teil der beobachtbaren Konturen gleichermaßen im Inneren von Phrasen (zwischen zwei Satzakkzenten) und auch am Ende von Phrasen vorkommen. Dies betrifft z.B. die leicht fallende Kontur, die sehr häufig zwischen zwei Gipfelkonturen auftritt (Peters et al. 2005b). Andere melodische Muster, z.B. fallend-steigende Konturen wurden hingegen ausschließlich am Ende prosodischer Phrasen beobachtet. Aus diesem Grund ist schon in den Richtlinien für die prosodische Etikettierung (Peters und Kohler 2004) festgelegt, dass die Etikettierung dieser Muster nur in phrasenfinaler Position vorgesehen ist. Ähnliches gilt auch für die stark steigenden und die stark fallenden + sehr leicht steigenden Konturen. In neueren, bislang informellen Experimenten mit synthetischem Sprachmaterial konnte allerdings gezeigt werden, dass kombinierte Konturen auch im Phraseninneren vorkommen können. Da aufgrund der Etikettierungspraxis von den separierenden Konturen also nur die stark fallende Kontur im Phraseninneren vorkommen kann, ist die Analyse des phraseninternen Vorkommens separierender Konturen auf diesen Konturtyp beschränkt. Dies ist auch unter dem Gesichtspunkt sinnvoll, dass die stark fallenden Konturen über zwei Drittel aller separierenden Konturen an flüssigen turninternen Phrasengrenzen ausmachen.

Die Analyse der melodischen Muster zwischen den Satzakkzenten innerhalb prosodischer Phrasen in den Terminabsprachedaten ergibt, dass von den 5103 Phrasen des Terminabsprachekorpus 459 Phrasen (9%) im Phraseninneren eine melodische Bewegung bis an die untere Grenze der Sprechstimme zeigen. 155 der 459 Fälle bilden hierbei allerdings den tiefsten Punkt einer steigenden Kontur (Talkontur). Dieser Anteil ist sehr hoch, wenn in Betracht gezogen wird, dass nur 839 der untersuchten 5103 Phrasen überhaupt eine Talkontur haben. Das stark fallende melodische Muster kann in diesem Zusammenhang als Erreichen eines initialen tiefen Zielpunktes von steigenden Konturen angesehen werden, der entscheidend für die auditive Prominenz der Talkonturen ist. Der Rest der stark fallenden Verläufe liegt zwischen zwei Gipfelkonturen im Phraseninneren (Peters et al. 2005b).

Diese Daten zeigen, dass das Konturmerkmal in seiner separierenden Wirkung hinter dem Merkmal Pause zurücksteht, da stark fallende Konturen im Gegensatz zu Pausen auch im Phraseninneren vorkommen können, ohne eine deutliche perzeptorische Separierung zu verursachen.

### 4.1.5 F0-Reset nach turninternen prosodischen Grenzen

Tabelle 3 zeigt die Verteilung von Phrasengrenzen mit und ohne Reset in den beiden Korpora. Da auch hier sehr ähnliche Zahlen für die Daten der beiden Korpora vorliegen, werden wiederum die Ergebnisse zusammengefasst.

**Tabelle 3:** F0-Reset in der nachfolgenden Phrase (Turninterne prosodische Grenzen, Terminabsprachen und LINDENSTRASSE, n=3399).

Reset	Kein Reset	Keine Entscheidung
2165 (63,7%)	610 (18%)	624 (18,4%)

Etwa zwei Drittel aller analysierten prosodischen Grenzen werden von einem Reset gefolgt, und nur bei 18% gibt es einen fortlaufenden Downstep über die Phrasengrenze hinweg. Die dritte Spalte fasst die Fälle zusammen, in denen im Rahmen der prosodischen Etikettierung keine Entscheidung über Reset an Phrasengrenzen gefällt wurde. Dies betrifft vor allem prosodische Phrasen, die steigende melodische Muster (Talkonturen oder kombinierte Konturen) enthalten. Der Hintergrund dieser fehlenden Resetentscheidung ist, dass als Bezugspunkt für die Entscheidung, ob ein Reset zwischen zwei Phrasen vorkommt, das F0-Maximum der letzten Gipfelkontur der Phrase festgesetzt ist. Dieser Bezugspunkt wird dann mit der Höhe des ersten F0-Gipfels der Folgephrase verglichen. Wenn der erste Gipfel der Folgephrase tiefer liegt, wird diese als 'kein Reset' etikettiert. Die Erfahrungen im Rahmen der prosodischen Etikettierung haben gezeigt, dass steigende Konturen den kontinuierlichen Downstep zwischen den F0-Maxima innerhalb von Phrasen oft unterbrechen und eine Art phraseninternen Reset verursachen. Deshalb ist am Ende dieser Phrasen auf eine Resetentscheidung verzichtet worden.

### 4.1.6 F0-Reset im Inneren prosodischer Phrasen

Tabelle 4 zeigt die Auftretenshäufigkeiten von prosodischen Phrasen mit durchgehendem Downstep gegenüber Phrasen, in denen der Downstep durch einen oder mehrere Upsteps unterbrochen wird. Diese phraseninternen Upsteps können als Neueinsatz des Downsteps angesehen werden und folglich als phraseninterner Reset.

Für die Analyse wird folgende Stichprobe gewählt: Alle prosodischen Phrasen der Terminabsprachedialoge, die ausschließlich Gipfel der Akzentstufe 2 ('default-Akzentstärke') enthalten (n=1019). Eine solche

Auswahl wird getroffen, um Interferenzen mit den Akzenten der Stufe 3 ('emphatische Akzentuierung'), die in der Regel ein sehr hohes F0-Maximum zeigen, auszuschließen. Auch die schwachen Akzente der Stufe 1 ('partielle Deakzentuierung') sollen ausgeschlossen werden, da diese meist nur leichte melodische Bewegungen implizieren. Die gewählte Stichprobe gewährleistet also die Vergleichbarkeit zwischen den untersuchten Gipfelkonturen.

**Tabelle 4:** Upstep bzw. Downstep in Phrasen mit nur 2er-Akzenten (Turninterne prosodische Grenzen, Terminabsprachen, n=1019).

	<b>Downstep (n=790)</b>	<b>Upstep (n=229)</b>
<b>2 Akzente (n1=743)</b>	622 (83,7%)	121 (16,3%)
<b>3 Akzente (n2=214)</b>	138 (64,5%)	76 (35,5%)
<b>4 Akzente (n3=46)</b>	22 (47,8%)	24 (52,2%)
<b>5 Akzente (n4=16)</b>	8 (50,0%)	8 (50,0%)
<b>Gesamt (n=1019)</b>	790 (77,5%)	229 (22,5%)

Die Zahlen in Tabelle 4 zeigen deutlich, dass ein phraseninterner Upstep besonders in Phrasen mit mehr als zwei Gipfelkonturen oft vorkommt. Dieser Upstep ist im Prinzip dasselbe, was an Phrasengrenzen mit 'Reset' bezeichnet wird: ein Abweichen von der Absenkung der Gipfelpositionen von Gipfel zu Gipfel. Reset kann also in spontaner Sprache seine primäre Funktion nicht in der Abgrenzung prosodischer Phrasen haben. Aus den Erfahrungen in der prosodischen Etikettierung von Spontansprache ist bekannt, dass ein Upstep innerhalb einer prosodischen Phrase meistens mit einer verstärkten Prominenz des erhöhten F0-Gipfels einhergeht. Insofern scheint das phonetische Phänomen Upstep vor allem im Bereich der Fokussierung lexikalischer Einheiten funktionalisiert zu sein und als Reset im Bereich des Neueinsatzes prosodischer Phrasen eine perzeptorisch weniger relevante Begleiterscheinung zu sein.

Trotzdem soll hier festgehalten werden, dass durchgehender Downstep zwischen Gipfelkonturen in den meisten untersuchten prosodischen Phrasen auftritt (77,5%) und damit auch in spontaner Sprache ein wichtiges phonetisches Phänomen im Rahmen der kohäsiven Gestaltung prosodischer Phrasen ist.

#### 4.1.7 Längung an turninternen prosodischen Grenzen

Für die Erfassung der segmentellen Dauerstruktur in phrasenfinalen Reimen wird ein Vergleich der Segmentdauern in phraseninternen Reimen mit den Dauern in phrasenfinalen Reimen durchgeführt. Am Ende wird ein Durchschnittswert für jeden phrasenfinalen Reim errechnet, der die Stärke der Längung relativ zu Segmenten in wortfinalen phraseninternen Reimen beschreibt. Die Grundlage für den Vergleich bilden ca. 50.000 Segmente aus phraseninternen Reimen. Ähnliche Verfahren kommen in den Untersuchungen von Ostendorf et al. (1990), Price et al. (1991) und Wightman et al. (1992) zum Einsatz. Eine genaue Beschreibung ist unter (Peters et al. 2005a) gegeben.

Die temporale Struktur der phrasenfinalen Reime wird in drei Klassen eingeteilt: Kürzung, Längung sowie weder Kürzung noch Längung. Als gekürzt werden die Reime angesehen, deren durchschnittliche Segmentdauern kleiner sind als der phraseninterne Mittelwert minus eine Standardabweichung. Als gelängt werden Reime klassifiziert, deren Segmente durchschnittlich länger sind als der phraseninterne Mittelwert plus eine Standardabweichung. Weder gelängt noch gekürzt sind alle Reime dazwischen (siehe Tabelle 5).

Es ergibt sich ein globaler Mittelwert von 64,6 ms für alle phraseninternen Segmente mit einer Standardabweichung von 20 ms. Die Klasse "Längung" (Dauern oberhalb Mittelwert plus eine Standardabweichung) würde dementsprechend bei einem Faktor von 1,31 beginnen. D.h., dass Reime in die Klasse 'Längung' fallen wenn ihre Dauer sich gegenüber der phraseninternen Dauer um mehr als das 1,3-fache vergrößert.

**Tabelle 5:** Längung am Phrasenende (Terminabsprachen und LINDENSTRASSE, n=3399).

Kürzung	Weder Längung noch Kürzung	Längung
75 (2,2%)	1087 (32%)	2237 (65,8%)

Die folgenden Tabellen zeigen die Häufigkeiten des gemeinsamen Auftretens von separierenden Konturen und den verschiedenen Längungsklassen (Tabelle 6) bzw. von Pausen und den Längungsklassen (Tabelle 7). Ziel der Analyse ist es festzustellen, ob eine Kopplung von segmenteller Längung an einen der beiden Parameter vorliegt.

**Tabelle 6:** Verteilung der Längungsklassen bei separierenden und nicht-separierenden Konturen (Terminabsprachen, n=2470).

	<b>Kürzung</b>	<b>Weder Längung noch Kürzung</b>	<b>Längung</b>
<b>Starke tonale Bewegung (n=1008 (40,8%))</b>	18 (1,8%)	317 (31,5%)	673 (66,7%)
<b>Keine starke tonale Bewegung (n=1462 (59,2%))</b>	41 (2,8%)	460 (31,5%)	961 (65,7%)

**Tabelle 7:** Verteilung der Längungsklassen an Phrasengrenzen mit und ohne Pausen (Terminabsprachen, n=2470).

	<b>Kürzung</b>	<b>Weder Längung noch Kürzung</b>	<b>Längung</b>
<b>Mit Pause/Atmungsphase (n=945 (38,3%))</b>	8 (0,8%)	143 (15,1%)	794 (84%)
<b>Ohne Pause/Atmungsphase (n=1525 (61,7%))</b>	51 (3,3%)	634 (41,6%)	840 (55,1%)

Die Zahlen in Tabelle 6 zeigen keine Verbindung zwischen der Dauer des phrasenfinalen Reims und der Art der melodischen Kontur. Für Phrasengrenzen mit Pausen hingegen gibt es ein deutlich erhöhtes Vorkommen von finaler Längung (84%), gegenüber dem durchschnittlichen Wert von 65,8% (siehe Tabelle 5) und besonders gegenüber den 55,1% gelängerter Reime an den Phrasengrenzen ohne Pause.

Pausen kommen also kaum ohne vorangehende Längung vor. Nur 15,9% aller Phrasengrenzen mit Pause haben keine vorangehende Längung. Wenn also ein Sprecher innerhalb seiner Äußerung auf eine Pause "zusteuert", ist dies mit einer Verlangsamung der artikulatorischen Abläufe verbunden. Oller (1973) und Klatt (1975) beschreiben dieses Phänomen für das Englische als 'prepausal lengthening'. Sie beobachten, dass die Silbe vor einer Pause um 60 bis 200 ms länger wird. Diese Längung betrifft vor allem den Vokal und die postvokalischen Konsonanten.

Allerdings kann der Umkehrschluss, dass finale Längung zumeist an Pausen gebunden ist, nicht gezogen werden, da an den Phrasengrenzen ohne Pause eine Längung des finalen Reims mit 55,1% ebenfalls recht häufig ist.

#### **4.1.8 Längung im Inneren prosodischer Phrasen**

Eine entsprechende Analyse des Faktors Längung für die phraseninterne Position ist schwierig, da es im Phraseninneren zu einer starken Fluktuation der Dauerverhältnisse in Abhängigkeit von den Akzentuierungsverhältnissen kommt. Häufig sind außerdem Dauerveränderungen, die in Verbindung mit Störungen der Sprechflüssigkeit stehen. Somit sind segmentelle Längungserscheinungen im Inneren von prosodischen Phrasen üblich und stehen in dieser Position nicht in Verbindung mit prosodischer Phrasierung. Wenn Längung an Phrasengrenzen in flüssiger Sprache vorkommt, dann fast immer in Kombination mit anderen phonetischen Merkmalen. Die automatische Korpusabfrage findet nur 256 prosodische Grenzen, an denen die Längung des finalen Reims das einzige extrahierbare Merkmal ist (7,5% der 3399 untersuchten Grenzen mit kompletten Merkmalsbündeln in beiden Korpora).

#### **4.1.9 Phonation an internen prosodischen Grenzen**

Neben den oben untersuchten Merkmalen finden sich in der Literatur Hinweise auf phonatorische Veränderungen am Ende prosodischer Phrasen. So beschreiben Local et al. (1985) das Auftreten von Verhauchung und Knarrstimme am Ende von Redebeiträgen. Nakatani und Hirschberg (1994) beobachten Knarrstimme in unflüssiger Sprache am Ende syntaktischer Fragmente. Sie führen weiterhin eine terminologische Trennung zwischen 'laryngealization' (an flüssigen prosodischen Grenzen) und 'glottalization' (an Abbrüchen) ein, um diese beiden Typen von Knarrstimme zu differenzieren.

Um die Häufigkeit von Knarrstimme an prosodischen Grenzen zu ermitteln, wurde jeder Fall von stark fallender Intonation im LINDENSTRASSE- Korpus hinsichtlich des Vorkommens unregelmäßiger Schwingungszyklen und Glottalverschlüssen einer auditiv-visuellen Analyse unterzogen. Im Rahmen dieser Analyse wurde deutlich, dass zwischen zwei Arten von 'Knarrstimme' zu unterscheiden ist, die sich auditiv relativ ähnlich sind, in der visuellen Darstellung in Oszillogramm und Sonagramm aber unterschieden werden können. Eine Art von Knarrstimme (in Tabelle 8 als 'Knarrstimme (unregelmäßig)' bezeichnet) zeigt Unregelmäßigkeiten in der Amplitude und der Frequenz der Schwingungszyklen. Die andere Art von Knarrstimme ('Knarrstimme (regelmäßig)') zeigt einen in Frequenz und Amplitude recht regelmäßigen Verlauf. Allerdings sinkt die Grundfrequenz hierbei in

einen sehr tiefen Bereich ab, so dass ein knarrender Gehörseindruck entsteht. Dieses Phänomen ist vor allem bei tiefen Männerstimmen zu beobachten.

Die Restkategorie enthält Fälle, in denen plosivbezogene Knarrstimme oder harte Vokaleinsätze interferieren oder Unsicherheit in der Beurteilung des Phonationstyps vorliegt. Im Deutschen (wie auch im Englischen) können orale stimmlose Plosive unter bestimmten morphologischen und phonotaktischen Rahmenbedingungen durch Glottalverschlüsse oder durch Glottalisierungen ersetzt werden (Kohler 2001). Interferenz von Glottalisierung im Zusammenhang mit harten Vokaleinsätzen tritt auf, wenn morpheminitiale Vokale mit Glottalverschluss oder Glottalisierungen produziert werden. Diese Fälle machen eine eindeutige Zuordnung schwierig und werden gekennzeichnet, damit sie später aus der Analyse ausgeschlossen werden können.

Um die zeitliche Extension der nicht-modalen Phonation zu bestimmen, wurden Zeitmarken eingefügt, die den Bereich der Knarrstimme delimitieren. Auch Anfang und Ende vorkommender Glottalverschlüsse wurden mit Zeitmarken abgegrenzt, so dass die Dauer von Glottalverschlüssen und geknarrten Signalabschnitten in deren Umgebung separat voneinander messbar sind.

**Tabelle 8:** Phonationstyp bei terminaler Intonation (LINDENSTRASSE, n=238).

<b>Modalstimme</b>	159 (66,8%)
<b>Knarrstimme (unregelmäßig)</b>	26 (10,9%)
<b>Knarrstimme (regelmäßig)</b>	20 (8,4%)
<b>Glottalverschluss</b>	-
<b>Rest (Interferenz und unsichere Beurteilung)</b>	33 (13,8%)

Die Zahlen in Tabelle 8 zeigen, dass beide Arten von Knarrstimme etwa gleich häufig vorkommen und zusammen etwa 20% der stark fallenden Intonationsmuster begleiten. Die durchschnittliche Dauer von Knarrstimme (unregelmäßig) beträgt 121 ms. In den Daten ist kein Fall zu beobachten, in dem die Knarrstimme in einem Glottalverschuss endet, wie dies an syntaktischen Abbrüchen vorkommen kann.

Es ist eine klare Sprecherabhängigkeit im Vorkommen der nicht-modalen Phonation beobachtbar. Dies betrifft sowohl das Auftreten von Knarrstimme generell - nur manche Sprecher zeigen dieses phonetische Phänomen - aber auch die Art der Knarrstimme. So häufen sich vor allem die Fälle von regelmäßiger Knarrstimme bei einigen Sprechern im Korpus. Dieser Phonationstyp ist also als sprechertypisch einzustufen.

#### **4.1.10 Diskussion zu den Einzelmerkmalen an turninternen prosodischen Grenzen**

Pausen und Atmen sind sehr eng mit prosodischen Grenzen verbunden. Sie sind somit als stark separierende phonetische Merkmale einzuordnen, die fast immer zur Wahrnehmung einer prosodischen Grenze führen. Dies ist bei den tonalen Merkmalen anders. Melodische Bewegungen, die den unteren Bereich der Sprechstimme erreichen sind zwar selten (9% aller prosodischen Phrasen haben eine solche Kontur), aber keine wirkliche Ausnahmeerscheinung. Reset bzw. das Abweichen vom Downstep ist in spontaner Sprache offensichtlich auch in phraseninterner Position häufig; fast ein Viertel der hierauf untersuchten Phrasen zeigen Abweichungen vom Downstep.

Zum Merkmal Längung kann auch ohne eine umfassende Analyse der phraseninternen Dauerstruktur gesagt werden, dass segmentelle Längung keineswegs primär mit dem Ende prosodischer Phrasen assoziiert ist, sondern auch in Verbindung mit prominenten Silben und Sprechflusstörungen vorkommt. Möglicherweise liegt die Funktion des Dauerparameters hauptsächlich in der Erzeugung auditiver Prominenz auf satzakzentuierten Silben und in der phonetischen Signalisierung von Planungsproblemen.

#### **4.1.11 Merkmalsbündel an turninternen Grenzen**

Nach der separaten Beschreibung einzelner phonetischer Merkmale an turninternen Phrasengrenzen soll nun deren Bündelung analysiert werden. Dies schließt folgende Arbeitsschritte ein:

- Analyse der Merkmalsbündel im LINDENSTRASSE-Korpus, analog zu den Terminabsprachedaten (Peters et al. 2005a). Dieser Arbeitsschritt soll zeigen, welche Unterschiede und Gemeinsamkeiten die phonetischen Merkmalskombinationen an prosodischen Grenzen in Daten aus einem anderen Korpus zeigen.
- Automatische subklassifizierende Zuordnung der flüssigen turninternen Phrasengrenzen zu den Klassen PG1 bis PG4 (vorerst nur für die Terminabsprachedaten).

Der Vergleich zwischen den Merkmalsbündeln im Terminabsprachekorpus und in den LINDENSTRASSE-Daten zeigt sehr ähnliche Verteilungen.

Deshalb sind die Ergebnisse für beide Korpora in Tabelle 9 zusammengefasst. Für die Darstellung sind die Fälle ohne Resetentscheidung der Kategorie '-R' zugeordnet.

**Tabelle 9:** Merkmalsbündel an turninternen flüssigen prosodischen Grenzen ohne nachfolgende Hässitation und ohne Überlagerungen (Terminabsprachen und LINDENSTRASSE, n=3399).

	-	<i>L</i>	<i>P</i>	<i>K</i>	<i>LP</i>	<i>PK</i>	<i>LK</i>	<i>LPK</i>	<i>Ges.</i>
<b>+R</b>	339	418	112	237	377	90	249	343	2165
<b>%</b>	10	12,3	3,3	7	11,1	2,6	7,3	10,1	63,7
<b>-R</b>	227	301	54	81	268	22	153	128	1234
<b>%</b>	6,7	8,9	1,6	2,4	7,9	0,7	4,5	3,8	36,5

Auf der Basis der Verteilung in den Terminabsprachedaten in Kombination mit der Interpretation von Korpusbeispielen nehmen Peters et al. 2005a die Einteilung in die vier Kategorien PG1-PG4 vor (siehe 1).

Es ergibt sich weiterhin eine Klasse von Phrasengrenzen (n=171), für die bei der automatischen Merkmalsextraktion keine phonetischen Korrelate festgestellt werden können. Im Rahmen einer Untersuchung dieser Fälle zeigt sich, dass diese Phrasengrenzen häufig in Verbindung mit Register- oder Sprechgeschwindigkeitswechsel stehen und/oder mit Talkonturen enden, die nicht bis in den oberen Bereich der Sprechstimme ansteigen. Da diese Fälle immerhin 6,9% aller turninternen flüssigen Phrasengrenzen ausmachen, werden sie bei der Validierung des von Peters et al. (2005a) erarbeiteten Klassifikationssystems mit einbezogen und in die sich ergebende Neustrukturierung des Systems integriert. In den nachfolgenden Texten und Tabellen werden diese Fälle einer Klasse zugeordnet, die PG0 genannt wird.

Unter Hinzuziehung der in 4.1.1 gemessenen Atmungs- und Pausendauern im Terminabsprachekorpus kann die Matrix der Merkmalsbündel um den jeweiligen Atmungs- oder Pausenwert ergänzt werden. Dies ermöglicht die automatische Zuordnung jeder flüssigen turninternen Phrasengrenze zu einer der Kategorien PG1-PG4 (bzw. der Restkategorie PG0). Folgende Zuordnung ergibt sich:

**Tabelle 10:** Verteilung über die fünf PG-Kategorien PG0-PG4 (Terminabsprachen, n=2470).

<b>PG0</b>	171 (6,9%)
<b>PG1</b>	1354 (54,8%)
<b>PG2</b>	770 (31,2%)
<b>PG3</b>	106 (4,3%)
<b>PG4</b>	69 (2,8%)

Da die Analyse der gemessenen Verteilung der Merkmalsbündel über die Klassen keine Möglichkeit einer funktionalen Einordnung gibt, wird der Form-Funktions-Beziehung in den Abschnitten 4.2 und 4.4 mit Hilfe von Perzeptionsexperimenten und Textinterpretationen nachgegangen.

## 4.2 Perzeptionsexperimente zu turninternen prosodischen Grenzen

### 4.2.1 Wahrnehmung phonetischer Kohäsion an internen prosodischen Grenzen (Experiment 1a und 1b)

Die untersuchten Produktionsdaten zeigen ein hohes Maß an Variabilität in der phonetischen Signalisierung von Phrasengrenzen. Sowohl die Zusammensetzung der Merkmalsbündel als auch die Ausprägung der Einzelmerkmale variieren stark. Es ist eines der Ziele dieser Arbeit, die unterschiedlich ausgeprägten flüssigen Phrasengrenzen auf einer perzeptorischen Skala zwischen Kohäsion und Separierung aneinandergrenzender Phrasen anzuordnen. Deshalb werden zwei Experimente mit dem Ziel durchgeführt, eine systematische Beziehung zwischen verschiedenen Zusammensetzungen von Merkmalsbündeln und dem Grad der perzeptorischen Kohäsion/Separierung herzustellen. Hierzu werden experimentell die perzeptorischen Effekte von Pausen, finaler Längung und verschiedenen melodischen Mustern hinsichtlich ihrer separierenden Wirkung innerhalb unterschiedlicher Merkmalskombinationen untersucht und der Bezug zu den Klassen PG0-PG4 hergestellt.

Die Ergebnisse aus den Perzeptionsexperimenten werden dann in die funktionale Interpretation von Korpusbeispielen einbezogen. So kann über die Kombination von Produktions- und Perzeptionsdaten ein komplexes Bild erarbeitet werden, das sowohl die Hörer- als auch die Sprecherseite berücksichtigt.

Da Pilotexperimente gezeigt haben, dass F0-Reset keinen Einfluss auf die wahrgenommene Kohäsion bzw. Separierung zwischen prosodischen Phrasen hat, wird dieser Parameter nicht in die systematische Variation des Stimulusmaterials einbezogen. Diese Entscheidung wird auch durch die Korpusanalyse gestützt, die zeigt dass Reset zwischen benachbarten Gipfelkonturen auch im Inneren prosodischer Phrasen häufig vorkommt und so nicht als typisches Merkmal prosodischer Phrasierung angesehen werden kann (siehe 4.1.5 und 4.1.6).

Das Stimulusmaterial der Perzeptionsexperimente geht von einer natürlich produzierten Äußerung aus, die ohne eine Phrasengrenze produziert wurde. Die Äußerung lautet: *Ich möchte eine kleine Portion mit viel Salat*. In den Experimenten werden in systematischer Variation an der Grenze zwischen *Portion* und *mit* drei Parameter verändert:

- Vorhandensein und Dauer von Pausen zwischen *Portion* und *mit*,
- die Längung des Reims von *Portion*,
- melodische Muster auf dem Wort *Portion*.

So entstehen Stimuli mit unterschiedlichen phonetischen Merkmalsbündeln im Übergang zwischen *Portion* und *mit*. Diese Merkmalsbündel repräsentieren typische, in den Produktionsdaten beobachtete phonetische Strukturen.

Im Hörtest sollen Probanden zu jedem Stimulus auf einer siebenstufigen bipolaren Skala entscheiden, wie stark getrennt ihnen die zwei Äußerungsteile (*Ich möchte eine kleine Portion* und *mit viel Salat*) erscheinen. Die Skala rangiert hierbei von "nicht getrennt" bis "extrem getrennt".

Das erste Teilexperiment soll vor allem die Rolle von Pausen in der wahrgenommenen Separierung zwischen prosodischen Phrasen beleuchten und Erkenntnisse über die Interaktion von Pausen mit anderen Merkmalen der Phrasierung bringen. Das zweite Teilexperiment konzentriert sich mehr auf die Wahrnehmung von finaler Längung und melodischen Mustern.

### 4.2.1.1 Hypothesen

Folgende Hypothesen werden überprüft:

#### **Hypothese 1**

Jeder der drei variierten Parameter beeinflusst die Stärke der wahrgenommenen Trennung zwischen den Äußerungsteilen.

#### **Hypothese 2a**

Pausen sind am stärksten für den Eindruck von Trennung verantwortlich. Die verschiedenen Pausendauern führen zu unterschiedlich wahrgenommenen Trennungsgraden auf der Skala.

#### **Hypothese 2b**

Durch die Urteile können Pausenklassen herausgearbeitet werden, in denen trotz unterschiedlicher Dauer ein vergleichbarer Grad an perzeptorischer Separierung erreicht wird.

#### **Hypothese 3**

Starke tonale Bewegungen, vor allem stark fallende Melodiemuster an der Phrasengrenze haben einen separierenden Effekt. Der stark fallende melodische Verlauf wird deshalb zu größerer perzeptorischer Separierung führen als der ebene und der leicht fallende Verlauf.

In einem Pilotexperiment zu dieser Studie zeigte sich die Tendenz, dass Kombinationen von finaler Längung und starken melodischen Bewegungen in Stimuli ohne Pausen oder mit kurzen Pausen zu einer Zunahme der wahrgenommenen Separierung führen. Bei Stimuli mit langen Pausen hingegen sind die Urteile der Versuchspersonen praktisch unabhängig von Längung oder starken melodischen Bewegungen. Hieraus ergibt sich Hypothese 4.

#### **Hypothese 4**

Pausen ab einer bestimmten Dauer maskieren die anderen phonetischen Parameter in ihrer trennenden Funktion. Wenn sich dieser Befund absichern lässt, bietet sich die Möglichkeit eine Kategoriengrenze zwischen kurzer und langer Pause zu etablieren, nämlich an dem Punkt, wo die Pausen beginnen, sich maskierend auf andere phonetische Merkmale auszuwirken.

#### 4.2.1.2 Stimulusmaterial

Der Testsatz *Ich möchte eine kleine Portion mit viel Salat.* wird vom Autor gesprochen und in einer schallgedämpften Aufnahmekabine mit 44,1 kHz digital aufgezeichnet. Ausgehend von diesem Satz werden phonetische Manipulationen der Parameter 'finale Längung', 'phrasenfinaler Melodieverlauf' und 'Pausendauer' durchgeführt. Der Satz zeigt in der Ausgangsäußerung keine interne prosodische Grenze. Er hat Satzakkente auf den Wörtern *Portion* und *Salat* und ist mit einem mittleren Sprechtempo gesprochen. Die Änderungen der phonetischen Parameter betreffen im Fall der Längung und des melodischen Verlaufs den Reim von *Portion*. Die unterschiedlich langen Pausen werden an der Wortgrenze zwischen *Portion* und *mit* hereingeschnitten. Durch diese Manipulationen entstehen an dieser Stelle verschieden ausgeprägte prosodische Einschnitte, die die Äußerung in zwei Teile untergliedern.

##### 4.2.1.2.1 Längung des finalen Reims

Die Dauer des finalen Reims von *Portion* wird in drei Stufen vergrößert. Ausgehend von einer kurzen Reimdauer der Originaläußerung, die nicht zur Wahrnehmung einer prosodischen Grenze führt, wird der Reim der Äußerung mit dem PSOLA-Algorithmus (Moulines und Carpentier 1990) in PRAAT (Boersma und Weenink) sukzessive vergrößert. Drei Längungsfaktoren werden angewendet: 1,2 und 1,4 und 1,6. Da aus Untersuchungen zur finalen Längung bekannt ist, dass die Domäne der phrasenfinalen Längung bei flüssiger Phrasierung in der Regel der phrasenfinale Reim ist (Klatt 1975, Whightmann et al. 1992) wird die Daueränderung auf die Segmente des Reims *Portion* angewendet. Durch dieses Verfahren ergeben sich folgende Dauern im Reim von *Portion*:

- Ausgangsstimulus ohne wahrnehmbare Längung (Faktor 1), Reimdauer: 230 ms,
- Erste Längungsstufe (Faktor 1,2), Reimdauer: 275 ms,
- Zweite Längungsstufe (Faktor 1,4), Reimdauer: 325,
- Dritte Längungsstufe (Faktor 1,6), Reimdauer: 370 ms.

Keiner der so manipulierten Stimuli mit den Längungsfaktoren 1 bzw. 1,2 und 1,4 vermittelt den Eindruck einer zögerlichen oder unflüssigen Sprechweise, wie dies bei einer noch stärkeren Dehnung des finalen Reims auftreten kann. Der Stimulus mit der stärksten finalen Längung (Faktor 1,6) kann hingegen als leicht zögerlich interpretiert werden.

#### 4.2.1.2.2 Verlauf des phrasenfinalen melodischen Musters

Die Ausgangsäußerung hat zwei deutliche Satzakkente, die durch zwei F<sub>0</sub>-Gipfel mit einer leichten Einbuchtung dazwischen realisiert sind. Ein Gipfelpunkt liegt bei 130 Hz, in der Mitte des Silbennukleus von *Portion*. Der zweite, breitere Gipfel beginnt auf *viel* und erstreckt sich bis in den Silbennukleus der satzakzentuierten Silbe von *Salat* bei etwa 120 Hz. Damit liegt ein Downstep von etwa 8% zwischen den beiden Gipfeln vor. Danach beginnt ein starkes Absinken der Grundfrequenz bis zum Äußerungsende.

Abbildung 2 zeigt die drei F<sub>0</sub>-Muster, die über die PSOLA-Komponente von PRAAT generiert wurden.

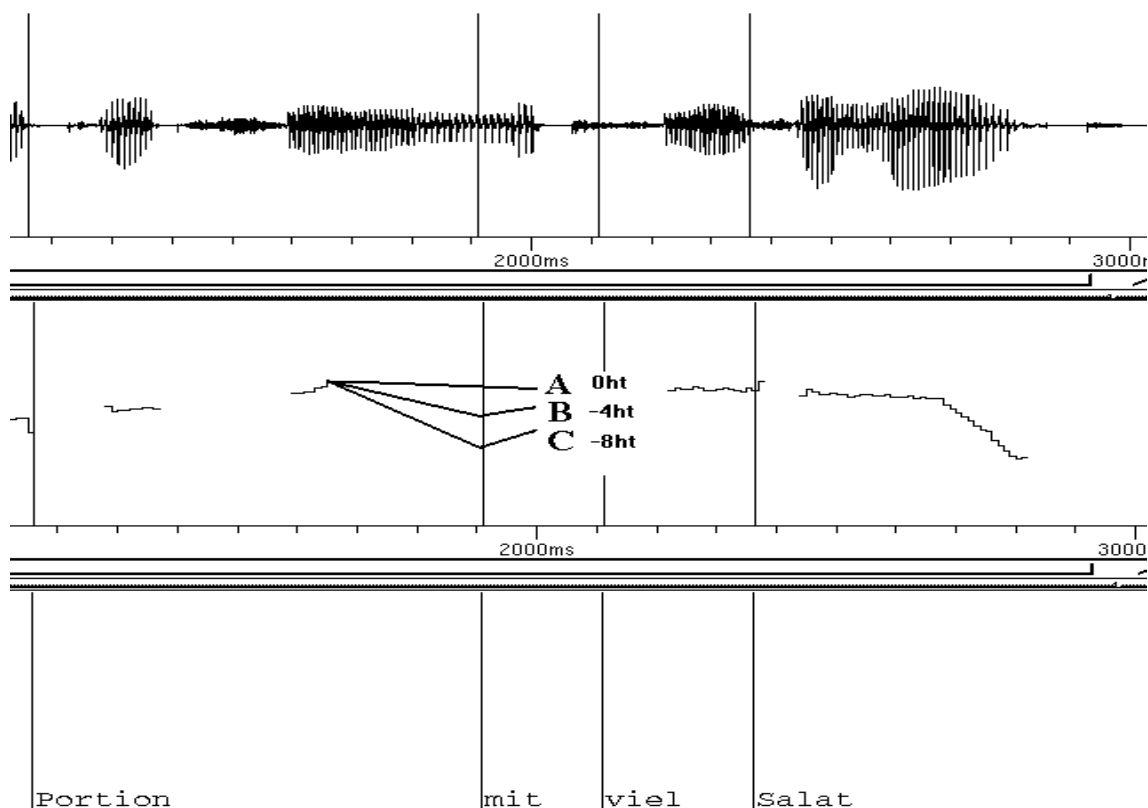


Abbildung 2: Melodische Muster im Stimulusmaterial.

Die drei resynthetisierten melodischen Muster repräsentieren drei Kategorien phrasenfinaler Intonationsmuster, die in der prosodischen Etikettierung mit KIM bzw. PROLAB differenziert werden.

Der Verlauf A ist eine direkte Verbindung zwischen dem Gipfel auf *Portion* und dem Gipfel auf *Salat*. Diese Art der Gipfelkonkatenation wird als Hutmuster bezeichnet ('t Hart et al. 1990; Peters et al. 2005b).

Der Verlauf B weist ein Absinken von F0 nach dem Gipfel auf *Portion* auf, das seinen Endpunkt am Ende des alveolaren Nasals bei 105 Hz erreicht. Das entspricht einer tonalen Differenz von vier Halbtönen zum vorangehenden Gipfelpunkt. Zum Beginn des nachfolgenden bilabialen Nasals von *mit* beginnt ein linearer Wiederanstieg, der sich bis zum Vokal von *viel* fortsetzt und bei 120 Hz endet.

Für Verlauf C wird der Endpunkt des Absinkens nach dem Gipfel auf *Portion*, ausgehend vom Endpunkt der Kontur B, nochmals um vier Halbtöne herabgesetzt. So entsteht ein Endpunkt der Bewegung bei 82 Hz. Der Wiederanstieg verläuft wiederum mit dem Vokal in *viel* als Zielpunkt, nur entsprechend steiler.

In den untersuchten Produktionsdaten finden sich diese drei Muster mit großer Häufigkeit an flüssigen turninternen Grenzen. Die ebene Kontur tritt mit einer Häufigkeit von 16,5% auf, die leicht fallende mit 28,3% und die stark fallende mit 26,9% (siehe Tabelle 2). Damit haben diese drei Muster die höchsten Vorkommenshäufigkeiten vor allen anderen melodischen Verläufen und werden deshalb im Stimulusmaterial abgebildet.

#### 4.2.1.2.3 Pausen

Um das Kontinuum von einer nicht hörbaren Pause zu einer sehr langen Pause perzeptorisch zu untersuchen, werden Pausen verschiedener Dauer in den Ausgangsstimulus hereingeschnitten. Es werden sieben Pausendauern gewählt. Ausgehend von der kleinsten Pause mit 50 ms wird eine Multiplikation mit dem Faktor 1,6 vorgenommen. Folgende Pausendauern ergeben sich (in ms): 50, 80, 130, 210, 330, 560, 890.

Diese Spannbreite zwischen der kürzesten und der längsten Pause (50 ms - 890 ms) entspricht in etwa der im Korpus beobachteten Variation von Pausen- und Atmungsdauern. Pausen im Inneren von Redebeiträgen oberhalb einer Dauer von 900 ms kommen im Korpus kaum vor; nur etwa 0,5% der Pausen- und Atmungsdauern im Terminabsprachekorpus liegen in diesem Bereich.

Das Hereinschneiden von Pausen in Äußerungen, in denen ursprünglich keine Pause vorhanden ist, kann sehr leicht zum Eindruck eines unnatürlichen, ruckartigen Einschnitts im Signal führen, da Pausen in natürlicher Sprache in der Regel von einer Abnahme der Intensität und von vorangehender segmenteller Längung begleitet sind. Weiterhin kommt es auch über Wortgrenzen hinweg zu Koartikulation zwischen benachbarten Lauten. Wenn diese koartikulierten Laute künstlich voneinander getrennt werden, entsteht ein unnatürlicher Bruch an dieser Stelle. Wegen der genannten Gründe verursacht andererseits aber auch das Herausschneiden von ursprünglich vorhandenen Pausen oft den Eindruck von Unnatürlichkeit. Bei Pilotversuchen zu diesem Experiment wurde speziell darauf geachtet, dass nicht der Eindruck eines technischen Artefakts entsteht. Hierbei zeigte sich, dass im Übergang zwischen zwei Wörtern manche Lautkombinationen recht gut auseinandergeschnitten und Pausen eingefügt werden können, ohne dass ein zu starker, technisch bedingter Einschnitt entsteht. So kann ohne weiteres eine Pause in den Übergang zwischen dem wortfinalen Nasal in *Portion* und dem wortinitialen Nasal von *mit* hereingeschnitten werden. Der Eindruck von Unnatürlichkeit ist hierbei nur sehr gering. Bei sehr genauem Abhören der Schnittstelle kann lediglich im finalen alveolaren Nasal von *Portion* eine leichte Bewegung in den bilabialen Verschluss festgestellt werden. Andere Lautkombinationen erwiesen sich als wesentlich schlechter trennbar.

Die Parametermanipulationen in den Bereichen Längung des finalen Reims (4 Längungsstufen), phrasenfinale Melodiemuster (3 Konturverläufe) und Pausendauer (keine Pause und sieben Pausendauern) und die systematische Kombination aller Merkmale ergibt 96 Stimuli mit verschiedenen Merkmalskombinationen.

#### 4.2.1.3 Durchführung von Experiment 1a

Aus den erzeugten Stimuli wird eine Audiodatei erzeugt, die im Hörexperiment präsentiert wird. Die Testserie umfasst drei Wiederholungen jedes Stimulus. Die Abfolge aller Stimuli ist randomisiert. Nach je zehn Stimuli ist eine längere Pause und ein Piepton. Zwischen zwei Stimuli ist innerhalb der Blöcke eine Pause von vier Sekunden, in der die Probanden das jeweilige Urteil abgeben können. Die gesamte Stimulusserie ist 40 Minuten lang.

Vor Beginn des Experiments wird den Probanden eine Aufnahme mit Instruktionen für den Versuch präsentiert. Dieser Präsentationsmodus hat

zwei Gründe: Erstens ist es wichtig, dass die Instruktionen in den verschiedenen Datenerhebungssitzungen identisch sind, um auszuschließen, dass Unterschiede in der Instruierung die Ergebnisse zwischen verschiedenen Sitzungen beeinflussen. Zweitens sind in die Instruktionen verschiedene Audiobeispiele eingebunden, die bei einer mündlichen Präsentation nicht exakt reproduziert werden könnten. Die Instruktionen umfassen eine kurze Einführung in das Konzept der Äußerungsgliederung durch prosodische Grenzen und einige Audiobeispiele, die unterschiedlich ausgeprägte Phrasengrenzen veranschaulichen. Weiterhin wird die konkrete Fragestellung und der Ablauf des Experiments erklärt (siehe Anhang).

Nach dem Abspielen der Instruktionen werden zehn Stimuli präsentiert, die das Spektrum der gesamten phonetischen Variation repräsentieren. Die Versuchspersonen geben für jeden Stimulus ein Urteil auf einer Skala ab. Danach gibt es eine Unterbrechung mit der Möglichkeit, Fragen zum Ablauf zu stellen. In keiner der Datenerhebungssitzungen gab es grundlegende Verständnisfragen.

Die Versuchspersonen beantworten im Experiment für jeden Stimulus folgende Frage *Wie stark sind der erste und der zweite Teil der Äußerung voneinander getrennt?* Für die Befragung wird eine 7-stufige bipolare Antwortskala gewählt, die folgendermaßen aufgebaut ist:

<b>nicht getrennt</b>	<b>mittel</b>					<b>extrem getrennt</b>
<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>

Skalen dieser Art werden *Equal appearing interval scales (EAI)* genannt und gehen auf Messungen in der experimentellen Psychologie zurück (Thurston und Chave 1929, Likert 1932). Sie sind geeignet, um die Urteile von Versuchspersonen als skalierte Messwerte aufzunehmen und werden in neuerer Zeit auch zunehmend häufiger in phonetischen Experimenten, vor allem im Bereich der Prosodieforschung, eingesetzt (de Pijper und Sanderman 1994, Grabe et al. 1997, Caspers 2000, Rietveld et al. 2002, Chen 2003).

An dem Experiment nahmen insgesamt 14 Versuchspersonen in zwei getrennten Datenerhebungssitzungen teil. Wie in auch in allen weiteren Versuchen sind die Versuchspersonen zum größten Teil Studierende der Universität Kiel. Alle Versuchspersonen sind deutsche Muttersprachler.

#### 4.2.1.4 Ergebnisse aus Experiment 1a

In diesem Abschnitt werden die oben formulierten Hypothesen anhand der Ergebnisse des Perzeptionstests überprüft. Die Varianzanalyse liefert folgende Ergebnisse.

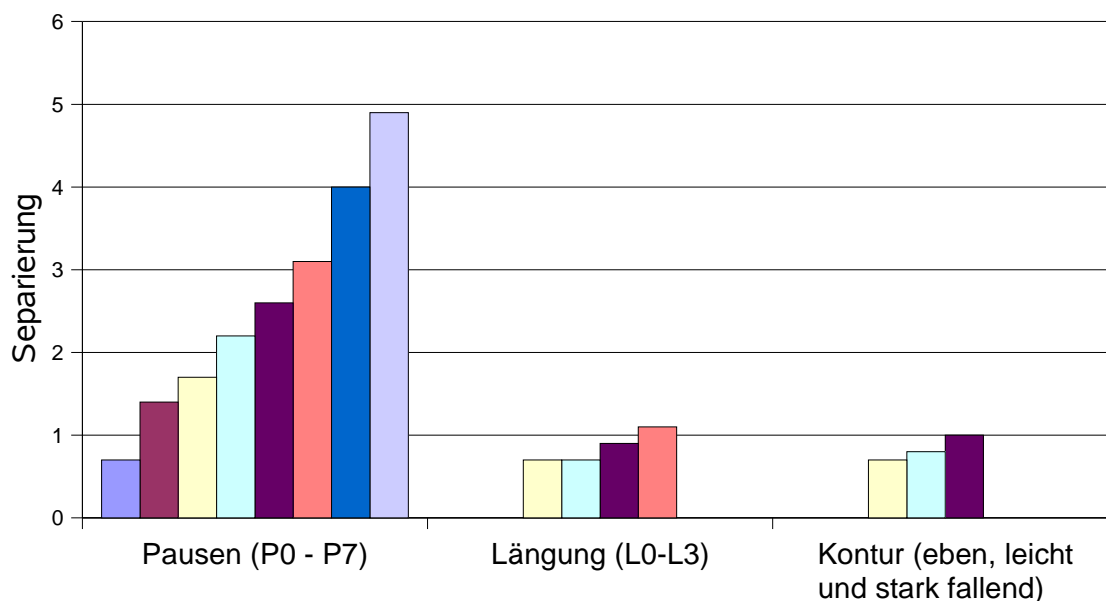
**Tabelle 11:** Ergebnisse der Varianzanalyse für Experiment 1a.

<i>Faktor / Interaktion</i>	<i>df</i>	<i>F</i>	<i>p</i>
Längung des finalen Reims	2,13	8,57	0,001
Finale Kontur	1,29	1,91	0,185
Pause	2,11	250,86	0,000
Längung * Kontur	4,16	1,84	0,132
Längung * Pause	6,97	1,26	0,279
Kontur * Pause	6,06	2,3	0,042
Längung*Kontur*Pause	8,53	2,5	0,014

#### **Hypothese 1:**

Jeder der drei variierten Parameter beeinflusst die Stärke der wahrgenommenen Trennung zwischen den Äußerungsteilen.

Abbildung 3 zeigt die Auswirkungen der manipulierten Parameter Pause, Längung und Konturverlauf auf die wahrgenommene Separierung zwischen den Äußerungsteilen. Um eine anschauliche Darstellung zu erhalten, die den Effekt jedes einzelnen Parameters, getrennt von der Variation der anderen Parameter herausarbeitet, wird für alle drei Parameter von dem Stimulus mit den geringsten Anzeichen phonetischer Separierung ausgegangen. Dies ist der Stimulus ohne Pause, ohne finale Längung und mit einem ebenen F<sub>0</sub>-Verlauf. Für die Darstellung der Pausen werden die Parameter Längung und Kontur konstant gehalten, d.h. in der Pausenserie verändert sich nur die Dauer der Pausen. Ebenso wird für Längung und Kontur vorgegangen. Um die perzeptorische Wirkung der vier Längungsstufen zu zeigen, sind hier die Urteile der Stimuli ohne Pause und mit ebenem Konturverlauf abgebildet. Die Variable 'Kontur' wird im konstanten Kontext ohne Pause und ohne Längung dargestellt. Diese Darstellungsweise ist auch unter Einbeziehung der prüfstatistischen Ergebnisse gerechtfertigt, die zeigen dass die Interaktion zwischen den getesteten Parametern nur sehr gering ist.



**Abbildung 3:** Wahrgenommene Separierung für die Pausenstufen, die Längungsstufen und die Konturen. In der Abbildung entspricht auf der y-Achse ein Wert von 0 einer maximalen Kohäsion zwischen den Äußerungsteile, ein Wert von 6 einer maximalen Trennung.

Abbildung 3 zeigt deutlich, dass jede Pausenstufe sukzessive zu einer Zunahme der wahrgenommenen Separierung führt. Die Effekte der Längung des finalen Reims und des phrasenfinalen Konturverlaufs hingegen sind sehr schwach ausgeprägt. Die prüfstatische Bearbeitung zeigt, dass Pausen hochsignifikant die Urteile beeinflussen ( $F=250,86$ ,  $p=0,000$ ; siehe Tabelle 11). Die Berechnung von Einzelkontrasten zeigt, dass der Unterschied von jeder Stufe zur nächsten hochsignifikant ist. Längung hat ebenfalls einen hochsignifikanten Einfluss ( $F=8,57$ ,  $p=0,001$ ; siehe Tabelle 11), der Einfluss der Kontur hingegen ist nicht signifikant ( $F=1,91$ ,  $p=0,185$ ; siehe Tabelle 11). Hierbei zeigt die deskriptive Darstellung den großen Unterschied zwischen den Urteilen jeder Pausenstufe und die minimalen Unterschiede der Serien, in denen Längung und Konturmerkmal verändert wurden. Hypothese 1 kann also nur zum Teil bestätigt werden.

### Hypothese 2a:

Pausen sind am stärksten für den Eindruck von Trennung verantwortlich. Die verschiedenen Pausendauern führen zu unterschiedlich wahrgenommenen Trennungsgraden auf der Skala.

Hypothese 2a wird durch die Ergebnisse bestätigt.

### **Hypothese 2b:**

Durch die Urteile können Pausenklassen herausarbeitet werden, in denen trotz unterschiedlicher Dauer ein vergleichbarer Grad an perzeptorischer Separierung erreicht wird.

Die Hypothese 2b kann nicht bestätigt werden. Es kommt von jeder Pausenstufe zur nächsten zu einem hochsignifikanten Anstieg der wahrgenommenen Separierung. Allerdings zeigt sich, dass die Effektstärke variiert: Der Unterschied zwischen Stimuli ohne Pause (P0) zur ersten Pausendauer (P1) und die Differenzen von P5 nach P6 und P6 nach P7 sind größer als die Unterschiede im mittleren Bereich.

### **Hypothese 3:**

Starke tonale Bewegungen, vor allem stark fallende Melodiemuster an der Phrasengrenze haben einen separierenden Effekt. Der stark fallende melodische Verlauf wird deshalb zu größerer perzeptorischer Separierung führen als der ebene und der leicht fallende Verlauf.

Diese Hypothese kann durch die statistische Analyse der Ergebnisse nicht belegt werden, obwohl sich in den Daten ein Trend in Richtung stärkerer Separierung des stark fallenden Musters abzeichnet.

### **Hypothese 4:**

Pausen ab einer bestimmten Dauer maskieren die anderen phonetischen Parameter in ihrer trennenden Funktion. Wenn sich dieser Befund absichern lässt, bietet sich die Möglichkeit eine Kategoriengrenze zwischen kurzer und langer Pause zu etablieren, nämlich an dem Punkt, wo die Pausen beginnen, sich maskierend auf andere phonetische Merkmale auszuwirken.

Da der separierende Effekt der Konturverläufe und der finalen Längung selbst ohne das Vorhandensein von Pausen sehr schwach ausfällt, kann dieser Hypothese anhand der Ergebnisse nicht weiter nachgegangen werden. Auch die Varianzanalyse zeigt keine systematischen Interaktionen zwischen den Parametern.

#### **4.2.1.5 Diskussion zu Experiment 1a**

Die Ergebnisse des Versuchs sind in zweierlei Hinsicht überraschend. Zum einen wegen des geringen Einflusses von Längung und Konturver-

lauf auf die Urteile der Versuchspersonen und zum anderen wegen der fehlenden Interaktion zwischen den getesteten Parametern.

Die Ursache für dieses Ergebnis liegt möglicherweise im Aufbau des Experiments 1a, denn nur 12 der 96 verwendeten Stimuli haben keine Pause. Aus diesem Grund konzentrieren sich die Versuchspersonen möglicherweise vor allem auf das Vorkommen von Pausen und deren unterschiedliche Dauern und geben nur dort, wo Pausen zwischen den Äußerungsteilen vorkommen, ein Urteil für deutliche Separierung ab. Ein solches Verhalten kann die Effekte von Konturverlauf und Längung sehr gering ausfallen lassen. Um diese Annahme zu überprüfen, wird in Experiment 1b eine Auswahl aus den 96 Stimuli des ersten Experiments getroffen, die ein Gleichgewicht zwischen Stimuli mit und Stimuli ohne Pause herstellt (siehe 4.2.1.6).

Weiterhin zeigen die Ergebnisse, dass die Versuchspersonen bereits die erste Pausenstufe (50 ms) klar von den Stimuli ohne Pause differenzieren. Diese Ergebnisse stehen im Widerspruch zu verschiedenen Experimenten in denen untere Schwellenwerte für die Wahrnehmung von Sprechpausen im Bereich zwischen 80 und 250 ms angesetzt werden (Butcher 1981, Goldmann Eisler 1968, Grosjean und Deschamps 1975). Die Ursache für diesen Befund liegt m.E. vor allem in der Art der Befragung in den verschiedenen Experimenten. In den Experimenten, die einen Schwellenwert belegen, werden die Versuchspersonen meist direkt nach dem Vorhandensein einer Pause befragt. Da die Versuchspersonen aber meist weder eine klare Vorstellung von Signalpausen noch ein ausgebildetes analytisches Hörvermögen haben, führt diese Art der Befragung zu zwei Artefakten: Erstens werden z.T. Urteile für das Vorhandensein von Pausen abgegeben, obwohl gar keine Pause vorhanden ist, und zweitens werden Urteile gegen das Vorhandensein von Pausen abgegeben, obwohl Pausen vorhanden sind.

Die Nicht-Übereinstimmung zwischen Signalpausen und Pausenwahrnehmung ist auch in neueren Experimenten häufig zu beobachten (z.B. Megyesi und Gustafson-Capkova 2002). Bereits 1937 schreiben Zwirner und Zwirner (S. 114):

*"Andererseits gibt es [...] Fälle, in denen die Wahrnehmung von Sprechpausen (vielleicht auch ihre Intention) einerseits, ihre Registrierung und Messung andererseits nicht übereinstimmen, in denen einwandfrei messbare Pausen nicht wahrgenommen werden (vielleicht auch nicht intendiert) werden und andere Fälle, in denen man derartige Pausen wahrzunehmen glaubt (vielleicht aufgrund syntaktischer oder sprachmelodi-*

*scher Faktoren), ohne daß sich solche Pausen messend feststellen lassen."*

Die Ursache für dieses Phänomen ist, dass die Kategorie 'Pause' bei naiven Hörern nicht mit dem phonetischen Konzept der Signalpause übereinstimmt. In der signalphonetischen Betrachtungsweise des Phonetikers ist eine Pause ein Zeitintervall, in dem die Amplitude des Sprachsignals gegen Null geht (abgesehen von Hintergrundgeräuschen). Signalpausen, die durch Plosivverschlüsse entstehen, sind bei dieser Definition ausgenommen. Die Beurteilung eines Signalabschnitts als Pause durch naive Hörer hingegen ist nicht allein durch tatsächliche Stille im Signal bedingt. Das Verhalten von Versuchspersonen in verschiedenen Perzeptionsexperimenten lässt vielmehr darauf schließen, dass naive Hörer verschiedene Phänomene vermengen, wenn sie über das Auftreten von 'Pausen' im Stimulusmaterial befragt werden. So können auch syntaktische und suprasegmentelle Merkmale des Gesprochenen zur Wahrnehmung von 'Pausen' führen. Versuchspersonen scheinen also eher über Zäsuren unterschiedlichen phonetischen oder linguistischen Ursprungs als über tatsächliche Signalpausen zu urteilen, wenn sie direkt nach Pausen befragt werden. Auf der anderen Seite werden kurze Pausen u.U. nicht als Pausen beurteilt, wenn sie nicht von anderen Merkmalen wie Längung begleitet werden. Da aber im hier durchgeführten Experiment nicht direkt nach Pausen gefragt wird, sondern nach der Stärke der Zäsur, zeigt sich, dass auch schon kurze Pausen Zäsuren verursachen, die durchaus wahrgenommen werden.

#### **4.2.1.6 Stimulusmaterial und Durchführung von Experiment 1b**

Während Experiment 1a vor allem auf die Wahrnehmung von Pausen ausgerichtet ist, steht die perzeptorische Wirkung der melodischen Muster und der phrasenfinalen Längung im Zentrum von Experiment 1b. Wie in Experiment 1a gezeigt wurde, steht die separierende Wirkung tonaler Brüche und phrasenfinaler Längung deutlich hinter dem Grad der Separierung zurück, die selbst kurze Pausen verursachen. Die Ursache hierfür liegt aber, wie bereits beschrieben, möglicherweise im Übergewicht der Stimuli mit Pausen im ersten Experiment. Experiment 1b arbeitet deshalb mit einer Auswahl von Stimuli aus Experiment 1a und stellt ein Gleichgewicht zwischen Stimuli mit und ohne Pausen her.

Folgende Stimuli aus Experiment 1a werden ausgewählt und in derselben Weise wie in Experiment 1a präsentiert und auf der 7-stufigen Skala beurteilt:

- 3 Konturtypen: eben, leicht fallend, stark fallend (wie in Experiment 1a),
- 3 Längungsfaktoren: keine Längung des finalen Reims (Längungsfaktor 1), Längungsfaktor 1,2 und Längungsfaktor 1,4. Der Längungsfaktor 1,6 entfällt, da so stark gelängte Reime möglicherweise als zögerliches Sprechen wahrgenommen werden und sich so die Bereiche der flüssigen und unflüssigen Phrasierung in diesem Experiment mischen würden.

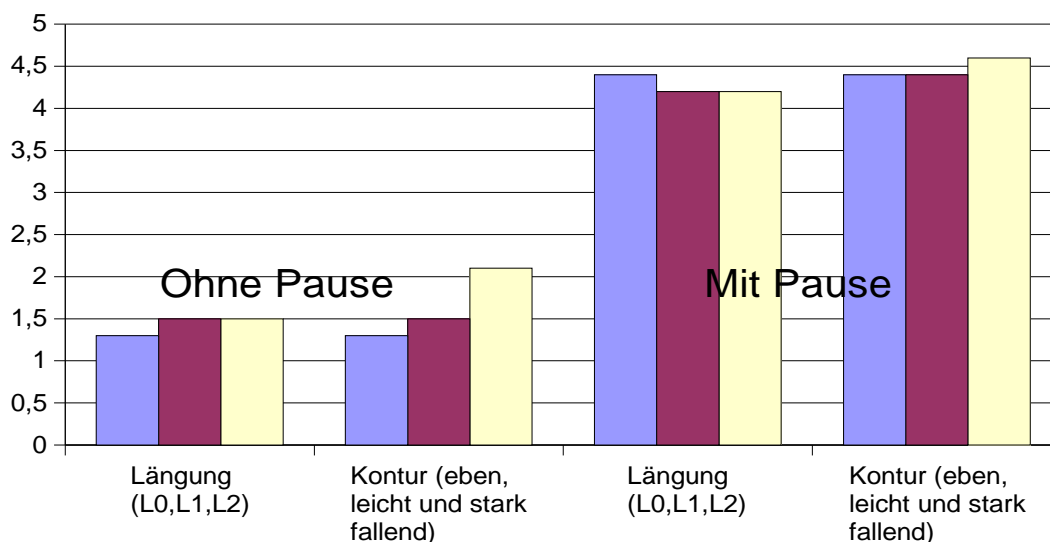
Die resultierenden 9 Stimuli (drei melodische Muster und drei Längungsstufen) werden entweder ohne Pause oder mit einer Pause von 210 ms präsentiert. Diese Pausendauer entspricht etwa der mittleren Pausendauer in den untersuchten Korpusdaten (siehe 4.1.1, Tabelle 1). So ergibt sich eine Anzahl von 18 verschiedenen Stimuli. Die Testserie umfasst zehn Wiederholungen jedes Stimulus. Der sonstige Ablauf ist parallel zu Experiment 1a. Die Instruktionen für das Experiment sind identisch mit den Instruktionen für Experiment 1a und befinden sich im Anhang. An dem Experiment nahmen 24 Versuchspersonen in drei getrennten Datenerhebungssitzungen teil. Die Hypothesen entsprechen mit Ausnahme der Hypothesen 2b und 4, die nicht noch einmal getestet werden, denen aus Experiment 1a.

#### 4.2.1.7 Ergebnisse aus Experiment 1b

**Tabelle 12:** Ergebnisse der Varianzanalyse für Experiment 1b.

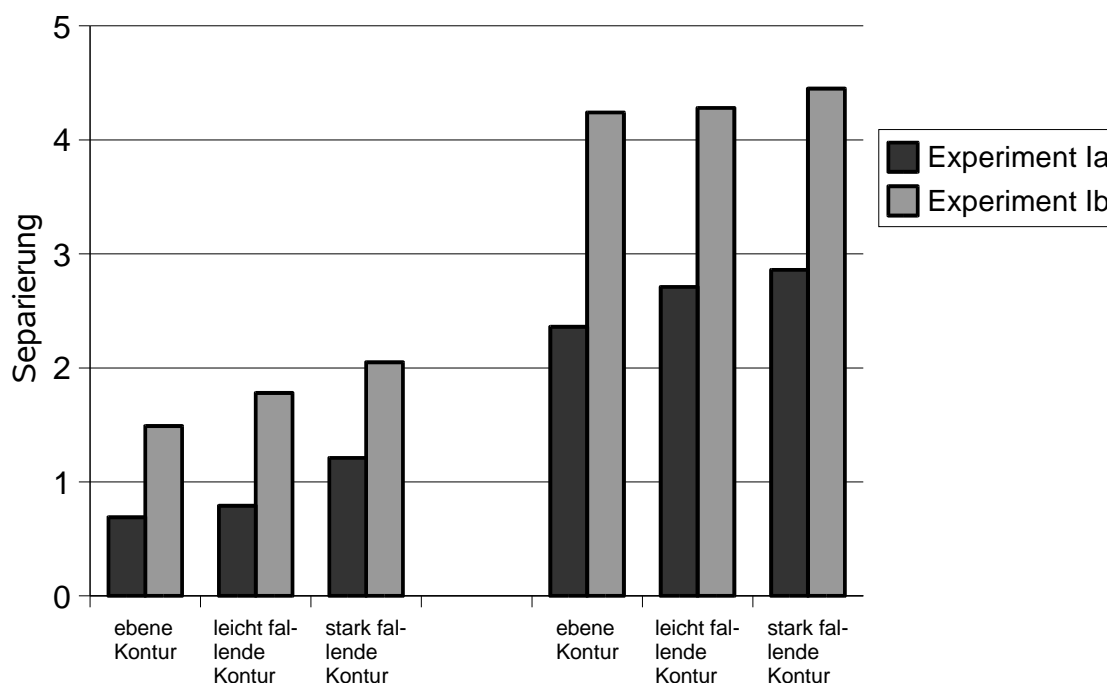
<i>Faktor / Interaktion</i>	<i>df</i>	<i>F</i>	<i>p</i>
Längung des finalen Reims	1,53	2,69	0,095
Finale Kontur	1,49	22,4	0,000
Pause	1,0	303,58	0,000
Längung * Kontur	3,13	1,9	0,134
Längung * Pause	1,89	6,8	0,003
Kontur * Pause	1,59	7,14	0,004
Längung*Kontur*Pause	2,99	2,8	0,047

Abbildung 4 zeigt die perzeptorische Wirkung der Parameter 'Längung' und 'Kontur'; links in Stimuli ohne und rechts in Stimuli mit Pause. Um den Effekt der Einzelparameter möglichst isoliert zeigen zu können, wird bei der Darstellung des Effekts der Längung für alle drei Stufen von einem ebenen melodischen Verlauf am Phrasenende ausgegangen. Für die Darstellung der Auswirkung der phrasenfinalen Kontur wird für alle Konturen der Kontext der Längungsstufe L0 (keine Längung) gewählt. In der Abbildung entspricht auf der y-Achse ein Wert von 0 einer maximalen Kohäsion zwischen den Äußerungsteilen, ein Wert von 6 einer maximalen Trennung.



**Abbildung 4:** Wahrgenommene Separierung für die Längungsstufen und die Konturen für die Stimuli mit und ohne Pause in Experiment 1b.

Die Varianzanalyse (Tabelle 12) zeigt, dass im Gegensatz zu Experiment 1a hier die Konturen einen hochsignifikanten Effekt haben ( $F=22,4$ ,  $p=0,000$ ), während der Effekt der Längung nicht signifikant ist ( $F=2,69$ ,  $p=0,095$ ). Außerdem ist Interaktion zwischen dem Vorhandensein von Pausen und der Wahrnehmung der Konturen signifikant: Der Unterschied zwischen den drei Konturen in Stimuli ohne Pause ist stärker als in Stimuli mit Pausen ( $F=7,14$ ,  $p=0,04$ ). Dies deutet auf den Maskierungseffekt von Pausen hin.



**Abbildung 5:** Vergleich der wahrgenommenen Separierung zwischen zwei identischen Stimulusserien in den Experimenten 1a und 1b. Alle Stimuli haben eine Längung des finalen Reims mit Faktor 1,2. Links sind die Ergebnisse für die Serie ohne Pause dargestellt und rechts die Ergebnisse für die Serie mit einer Pause von 215ms.

Im Vergleich zwischen den Bewertungen der identischen Stimuli in beiden Experimenten zeigt sich ein deutlicher Einfluss des unterschiedlich zusammengesetzten Stimulusmaterials. Die Urteile für alle Stimuli fallen in Experiment 1a deutlich niedriger aus. D.h., die Stimuli ohne Pause verschieben sich in Experiment 1b auf der Skala um etwa einen Skalenpunkt in Richtung der extremen Separierung, die Stimuli mit Pause um etwa zwei Skalenpunkte. Diese Befunde zeigen die Abhängigkeit der Hörerurteile von der Gesamtzusammensetzung des Stimulusmaterials (Abbildung 5). Da es in Experiment 1a sehr viele Stimuli gibt, die z.T. sehr lange Pausen zwischen den Äußerungsteilen enthalten, orientieren sich die Versuchspersonen an diesen extremen Trennungsgraden und setzen die kohäsiveren Realisierungen zu diesen stark separierten Äußerungen in Beziehung. So wandern die Urteile in Experiment 1b alle in Richtung stärkerer Separierung, da extrem separierte Stimuli in diesem Experiment fehlen. Die erhaltenen Urteile zu Kohäsion und Separierung sind also als relative Werte zu betrachten, die im Kontext der Experimente eine Vergleichsmöglichkeit zwischen der

Wahrnehmung verschiedener Merkmalsbündel bieten, nicht aber als absolute Messwerte auf natürliches Material übertragen werden können.

#### 4.2.1.8 Diskussion zu Experiment 1a und 1b

In allen Konturtripeln ohne Pausen in Experiment 1b gibt es einen hochsignifikanten Effekt des Konturverlaufs. Der Grad der wahrgenommenen Trennung steigt von der ebenen Kontur zur leicht fallenden Kontur und weiter zur stark fallenden Kontur. Dieser Effekt wird deutlich schwächer bei den Stimuli mit Pausen. Längung hat fast keinen systematischen Einfluss auf die wahrgenommene Separierung. Bei den Stimuli ohne Pausen lässt sich von Längungen mit Faktor 1 zu Längungen mit Faktor 1,4 bei den ebenen und leicht fallenden Konturen eine leichte Zunahme der Separierung feststellen. Bei den Stimuli mit Pausen ist keine Beziehung zwischen Längungsstufen und wahrgenommener Separierung zu erkennen. Es zeigt sich also ein leichter Maskierungseffekt der Pausen. Insgesamt sind jedoch die Interaktionen zwischen den getesteten Parametern weit geringer als erwartet.

Die Annahme, dass die Zusammensetzung des Stimulusmaterials in Experiment 1b im Vergleich zu Experiment 1a die separierende Wirkung unterschiedlicher Konturen deutlicher herausarbeiten kann, bestätigt sich, da der Effekt der Kontur in Experiment 1b hochsignifikant ist ( $F=22,4$ ,  $p=0,000$ ). Es muss allerdings beachtet werden, dass der Einfluss des melodischen Verlaufs auf die Urteile im Vergleich zu dem der Pausen nur gering ist (siehe Abbildung 4) und in den Experimenten 1a und 1c nicht signifikant wird. Trotzdem lässt sich Hypothese 3 durch das Experiment 1b prüfstatisch absichern.

In Experiment 1b ist der Effekt der Längung nicht signifikant ( $F=2,69$ ,  $p=0,095$ ). Dieses Ergebnis legt nahe, die Längung des finalen Reims aus dem Klassifikationssystem für flüssige prosodische Grenzen herauszulassen. Dies gilt allerdings nur dann, wenn die Längung einen Faktor von etwa 1,6 nicht überschreitet, da stärkere Längungen des finalen Reims von Hörern als Hinweis auf eine unflüssige Sprechweise interpretiert werden. Das führt im Rahmen der Klassifikation unterschiedlicher prosodischer Grenzen zu einer anderen funktionalen Einordnung (vgl. 6.2, Experimente 3a und 3b).

## 4.2.2 Einfluss syntaktischer Grenzen auf die wahrgenommene Kohäsion (Experiment 1c)

Experiment 1c erweitert die Fragestellung nach dem Zusammenhang zwischen der phonetischen Struktur prosodischer Grenzen und der wahrgenommenen Kohäsion zwischen Äußerungsteilen auf den Einfluss syntaktischer Strukturen. D.h. es wird untersucht, ob prosodische Grenzen an syntaktischen Grenzen anders beurteilt werden als prosodische Grenzen innerhalb syntaktischer Einheiten. Diese Fragestellung geht zurück auf Experimente von Boomer und Dittmann (1962) und Boomer (1965), in denen unterschiedliche Wahrnehmungsschwellen für Pausen in Abhängigkeit von ihrer Position in der syntaktischen Struktur festgestellt wurden. Pausen innerhalb syntaktischer Konstituenten wurden von den Versuchspersonen schon bei geringeren Pausendauern erkannt als Pausen am Ende syntaktischer Einheiten. Hieraus ergibt sich die Frage, ob nicht nur Pausen, sondern auch F0-Bewegungen und unterschiedliche segmentelle Dauerstrukturen bzw. Bündelungen dieser Merkmale in Abhängigkeit von ihrer Position in der syntaktischen Struktur unterschiedlich wahrgenommen werden.

Die zweite Erweiterung, die Experiment 1c vornimmt, ist die Untersuchung der wahrgenommenen Kohäsion in Abhängigkeit von Zögerungslängung am Phrasenende. Es werden in Experiment 1c Stimuli einbezogen, die eine so starke Längung des finalen Reims aufweisen, dass der Eindruck einer unflüssigen Sprechweise entsteht. Wie in Experiment 3a und 3b nachgewiesen werden kann, führt zunehmende Längung des phrasenfinalen Reims sukzessive zur Wahrnehmung einer unflüssigen Sprechweise (siehe 6.2). Besonders die Längung des finalen Reims mit Faktor 2,2 führt zu einer Beurteilung als stark unflüssig. Aus diesem Grund werden in das Stimulusmaterial auch Stimuli aufgenommen, deren finaler Reim mit den Faktoren 1,8 und 2,2 gelängt ist. Hierdurch soll die Wahrnehmung phonetischer Kohäsion bei Störungen im Sprechfluss untersucht werden. Eine ausführlichere Diskussion dieser Fragestellung findet sich in Abschnitt 6.2.

### 4.2.2.1 Hypothesen

#### Hypothese 1

Der Grad der wahrgenommenen Separierung ist bei gleicher Merkmalsbündelung um so höher, je niedriger die syntaktische Grenze ist.

## Hypothese 2

Leichte bis mittlere segmentelle Längung hat keinen deutlichen Einfluss auf die wahrgenommene Separierung. Starke segmentelle Längung führt zu deutlicher wahrgenommener Separierung.

### 4.2.2.2 Stimulusmaterial

Experiment 1c ist den Experimenten 1a und 1b ähnlich. Es wird wiederum die phonetische Struktur am Ende einer Teiläußerung systematisch variiert, indem die Parameter 'Längung', 'melodischer Verlauf' und 'Pause' verändert werden. Durch diese phonetische Variation der Merkmalsbündel am Ende des ersten Äußerungsteils entstehen 24 phonetische Varianten, die jeweils mit drei verschiedenen Äußerungsfortsetzungen gepaart werden. Je nach Art der Fortsetzung ist der syntaktische Einschnitt nach der ersten Teiläußerung unterschiedlich. Der erste Äußerungsteil ist für alle Paarungen *Ich hätte gern' 200 Gramm*. Die erste Fortsetzung ist *Marzipankugeln*, die zweite Fortsetzung *mit viel Nüssen* und die dritte Fortsetzung *Mehr ist schlecht für meine Diät*. Durch Paarung 1 (*Ich hätte gern' 200 Gramm - Marzipankugeln*) entsteht nach *Gramm* eine niedrige syntaktische Grenze innerhalb der Substantivgruppe *zweihundert Gramm Marzipankugeln*, zwischen der partitiven Apposition *zweihundert Gramm* und dem Bezugswort *Marzipankugeln*. Bei Paarung 2 (*Ich hätte gern' 200 Gramm - mit viel Nüssen*) entsteht eine höhere syntaktische Grenze zwischen *Gramm* und der Präpositionalphrase in attributiver Funktion *mit viel Nüssen*. Die dritte Paarung (*Ich hätte gern' 200 Gramm. - Mehr ist schlecht für meine Diät.*) führt nach *Gramm* zu einer noch höheren Grenze zwischen zwei Hauptsätzen.

Die Versuchspersonen beurteilen im Experiment für jede Paarung den Grad der wahrgenommenen Separierung zwischen den Äußerungsteilen. Dieser Aufbau bietet die Möglichkeit zu prüfen, wie die Wahrnehmung identischer phonetischer Varianten durch den syntaktischen Kontext beeinflusst wird.

Das Stimulusmaterial wird vom Autor gesprochen und in einer schalldämpften Aufnahmekabine mit 44,1 kHz digital aufgezeichnet. Jede der Äußerungsfortsetzungen wird zusammen mit dem Äußerungsbeginn, ohne eine intervenierende prosodische Grenze zwischen den Äußerungsteilen, gesprochen. Folgende Äußerungen werden aufgenommen:

- *Ich hätte gern' 200 Gramm Marzipankugeln.*
- *Ich hätte gern' 200 Gramm mit viel Nüssen.*
- *Ich hätte gern' 200 Gramm. Mehr ist schlecht für meine Diät.*

Bei der Weiterverarbeitung wird zuerst aus einer der Äußerungen die Sequenz *Ich hätte gern' 200 Gramm* herausgeschnitten und hinsichtlich der finalen Merkmalsbündel systematisch variiert. Dann wird aus den drei Äußerungen jeweils die zweite Teiläußerung isoliert und mit den Varianten des ersten Äußerungsteils gepaart.

Folgende Kriterien gehen in die Auswahl des verwendeten Sprachmaterials ein:

- Durch das Zusammenfügen der Teiläußerungen sollen nach der ersten Teiläußerung unterschiedlich hohe syntaktische Grenzen entstehen.
- Die Schnittstelle zwischen dem Ende des ersten Äußerungsteils und dem Beginn des zweiten Äußerungsteils soll von der segmentellen Struktur her geeignet sein, die Teiläußerungen zusammenzuschneiden, ohne dass ein unnatürlicher Bruch an dieser Stelle entsteht, und sie soll ebenfalls geeignet sein, eine intervenierende Pause hineinzuschneiden, ohne dass dies unnatürlich klingt. Hierfür ist der Übergang zwischen den bilabialen Nasalen am Ende der ersten Teiläußerung und am Beginn aller drei Folgeäußerungen besonders gut geeignet, da aufgrund der Identität der finalen und initialen Laute keine deutlichen artikulatorischen Veränderungen auftreten, die beim Schneiden an dieser Stelle unnatürlich klingen würden.
- Die Äußerungen sollen alle in einem vergleichbaren situativen Kontext geäußert werden können. Das macht es den Versuchspersonen vermutlich leichter, sich in die Aufgabenstellung hineinzusetzen. In diesem Fall könnte der situative Kontext ein Verkaufsgespräch an einem Stand mit Süßigkeiten sein.

#### **4.2.2.2.1 Längung des finalen Reims**

Die Dauer des finalen Reims von *Gramm* wird in vier Stufen vergrößert. Ausgehend von der Reimdauer der Originaläußerung ohne prosodische

Grenze nach *Gramm* werden in PRAAT die Längungsfaktoren 1,4 und 1,8 und 2,2 angewendet. Es ergeben sich folgende Reimdauern:

- Keine Längung (Faktor 1, Originaläußerung), Reimdauer: 170 ms,
- Erste Längungsstufe (Faktor 1,4), Reimdauer: 240 ms,
- Zweite Längungsstufe (Faktor 1,8), Reimdauer: 305 ms,
- Dritte Längungsstufe (Faktor 2,2), Reimdauer: 375 ms.

Die erste Reimdauer (Faktor 1) kann als ungelängte Variante des finalen Reims angesehen werden, da in der Ausgangsäußerung, aus der sie entstammt, nach *Gramm* keine prosodische Grenze, sondern eine direkte Äußerungsfortsetzung folgt. Auch der auditive Eindruck deutet nicht auf finale Längung hin. Die zweite Reimdauer (Längungsstufe 1,4) entspricht der mittleren phrasenfinalen Längung, die in den Produktionsdaten an den etikettierten prosodischen Grenzen gemessen wurde. Die zwei höheren Längungsstufen (1,8 und 2,2) liegen deutlich oberhalb des Mittelwerts. Stimuli mit Längungen oberhalb des Faktors 1,4 verursachen zunehmend den Eindruck einer unflüssigen Sprechweise (siehe 6.2).

#### 4.2.2.2 Verlauf des phrasenfinalen melodischen Musters

Die Ausgangsäußerung hat einen Satzakkzent auf *Zweihundert*, realisiert mit einem F0-Gipfel, der von 115 Hz auf 130 Hz ansteigt. Ausgehend von diesem Gipfel werden drei linear verlaufende melodische Verläufe erzeugt, die sich bis an das Ende von *Gramm* erstrecken. Die drei melodischen Muster sind mit denen der Experimente 1a und 1b vergleichbar. Es wird zum einen ein ebenes Muster generiert, das den Gipfel auf *Zweihundert* durch eine F0-Kontur ohne Einbuchtung mit dem Gipfel der zweiten Teiläußerung verbindet, der in allen drei Varianten bei 120 Hz liegt. Die Satzakkzentpositionen in den zweiten Teiläußerungen sind folgende: *Marzipankugeln, mit viel Nüssen, mehr ist schlecht für meine Diät*. Die zweite resynthetisierte Variante ist ein leicht fallendes Muster, das vom Gipfel auf *Zweihundert* bis zum Ende von *Gramm* eine Differenz von 4 Halbtönen aufweist und bei 103 Hz endet. Die dritte Variante fällt über denselben Zeitraum um 8 Halbtöne und endet bei 82 Hz. Für die zweite und dritte Variante gilt, dass nach *Gramm*, also ab Beginn der zweiten Teiläußerungen, ein Aufstieg zur

konstanten Gipfelposition bei 120 folgt. Die zweite Teiläußerung endet in allen drei Varianten mit stark fallender Intonation (vgl. Abbildung 2).

#### **4.2.2.2.3 Pausen**

Durch die Kombinationen der Teiläußerungen und die Variationen der Parameter 'Längung' und 'Kontur' am Ende der ersten Teiläußerung entstehen 36 unterschiedliche Stimuli (4 Längungsstufen x 3 Konturen x 3 Äußerungsfortsetzungen). In jeden dieser Stimuli wird zwischen die erste und zweite Teiläußerung eine Pause von 210 ms geschritten. Diese Pausendauer liegt im Bereich der mittleren Pausendauer, die in den untersuchten Produktionsdaten gemessen wurde. Die Stimuli mit Pause werden separat abgespeichert. So entsteht neben den 36 Stimuli ohne Pause eine ansonsten identische Serie von 36 Stimuli mit interner Pause.

Der Schnittpunkt für das Einfügen der Pause wird in allen Fällen so gesetzt, dass der initiale Nasal der zweiten Teiläußerung vier Perioden lang ist. Da die Schnittstelle sich innerhalb eines langen bilabialen Nasals befindet, kann hier eine Pause hineingeschnitten werden, ohne dass koartikulatorische Abläufe durch die Pause unterbrochen werden und so ein unnatürlicher Gehörseindruck entsteht. Dies wird zusätzlich durch die geringe Intensität des Nasals begünstigt.

#### **4.2.2.3 Durchführung von Experiment 1c**

Aus den 72 Stimuli wird eine randomisierte Abfolge generiert und wie in den Experimenten 1a und 1b präsentiert. Im Experiment 1c wird auf eine Stimuluswiederholung verzichtet, da die Experimente 1a und 1b bereits gezeigt haben, dass die Versuchspersonen bei Stimuluswiederholungen relativ konsistente Urteile abgeben. Die gesamte Serie ist 9 Minuten lang.

Vor der Präsentation und Beurteilung der Stimuli werden den Versuchspersonen die Instruktionen für das Experiment vorgespielt. Im Prinzip handelt es sich um die selben Anweisungen und Erläuterungen wie in den Experimenten 1a und 1b. Allerdings ist eine Passage hinzugefügt worden, in der auf die drei unterschiedlichen Sätze eingegangen wird, die durch das Zusammenfügen der ersten Teiläußerung mit den drei nachfolgenden Äußerungsfortsetzungen entstehen. Die Instruktionen befinden sich im Anhang.

Die Versuchspersonen beurteilen den Grad der Trennung zwischen den angrenzenden Äußerungsteilen wieder auf der 7-stufigen Skala. An dem Experiment nehmen 17 Versuchspersonen teil.

#### 4.2.2.4 Ergebnisse aus Experiment 1c

**Tabelle 13:** Ergebnisse der Varianzanalyse für Experiment 1c.

<i>Faktor / Interaktion</i>	<i>df</i>	<i>F</i>	<i>p</i>
Syntax	1,91	0,03	0,976
Längung des finalen Reims	1,53	8,47	0,003
Finale Kontur	1,42	1,7	0,208
Pause	1	162,3	0,000
Syntax * Längung	4,69	2,71	0,029
Syntax * Kontur	3,26	1,49	0,227
Syntax * Pause	1,91	3,58	0,042
Längung * Kontur	3,95	0,82	0,519
Längung * Pause	1,9	29,16	0,000
Kontur * Pause	1,92	0,53	0,588
Syntax * Längung * Pause	3,8	0,72	0,577
Syntax * Längung * Kontur	5,79	2,6	0,024
Syntax * Pause * Kontur	3,26	0,88	0,464
Längung * Pause * Kontur	4,24	0,47	0,771
Syntax * Pause * Kontur * Längung	5,77	0,87	0,517

Die folgenden Abbildungen zeigen die Effekte der variierten phonetischen und syntaktischen Parameter auf die Urteile der Versuchspersonen. Der Einfluss der syntaktischen Struktur auf die wahrgenommene Separierung ist für keinen der getesteten Parameter signifikant ( $F=0,03$ ,  $p=0,976$ ; siehe Tabelle 13). Hypothese 1 wird deshalb verworfen.

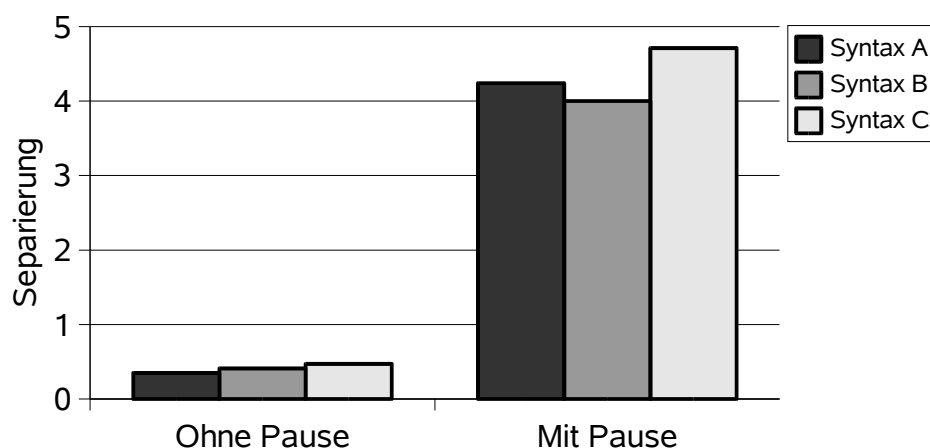
Um den perzeptorischen Effekt der Einzelparameter möglichst isoliert zeigen zu können, ist jeder phonetische Parameter in einer separaten Abbildung dargestellt. Für die Darstellungen gilt, dass hier die Mittelwerte von Stimuli einbezogen werden, in denen die jeweils nicht relevanten Parameter die am wenigsten trennende Ausprägung haben. Das bedeutet, dass bei der Darstellung des Effekts der Pausen von Stimuli ohne finale Längung und mit einer ebenen Kontur am Phrasenende ausgegangen wird. Die abgebildeten Mittelwerte für die vier Längungsstufen kommen aus Stimuli ohne Pause und mit einem ebenen melodischen Verlauf am Phrasenende. Für die Darstellung der phrasenfinalen Konturen wird für alle melodischen Muster der phonetische Kontext ohne Pause und einer

Längung mit Faktor 1,0 (keine Längung) gewählt.

In den den folgenden drei Abbildungen entspricht auf der y-Achse ein Wert von 0 einer maximalen Kohäsion zwischen den Äußerungsteile, ein Wert von 6 einer maximalen Trennung.

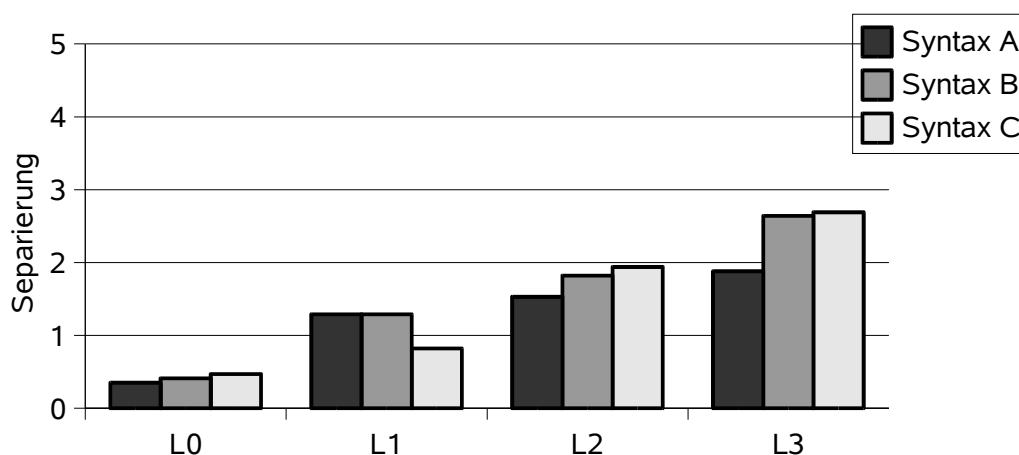
Die syntaktischen Kontexte sind folgende:

- Syntax A: *Ich hätte gern' 200 Gramm Marzipankugeln.*
- Syntax B: *Ich hätte gern' 200 Gramm mit viel Nüssen.*
- Syntax C: *Ich hätte gern' 200 Gramm. Mehr ist schlecht für meine Diät.*



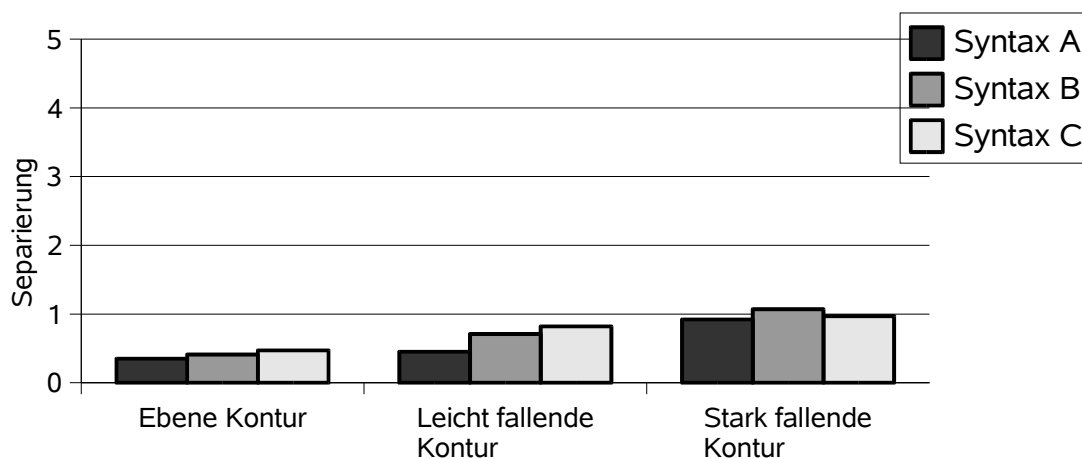
**Abbildung 6:** Wahrgenommene Separierung für Stimuli mit und ohne Pausen.

Abbildung 6 zeigt den wiederum hochsignifikanten Effekt der Pausen (siehe Tabelle 13).



**Abbildung 7:** Wahrgenommene Separierung für die vier Längungsstufen.

Abbildung 7 zeigt die signifikante Auswirkung der vier Längungsstufen ( $F=8,47$ ,  $p=0,003$ , siehe Tabelle 13): Von Faktor 1 bis Faktor 2,2 ist ein kontinuierliches Ansteigen der wahrgenommenen Separierung zu beobachten. Die Unterschiede zwischen den Beurteilungen der vier Längungsstufen sind im Vergleich zu den Experimenten 1a und 1b deutlich. Der Grund hierfür ist, dass im Experiment 1c eine stärkere Variation der Längung des finalen Reims einbezogen wird und die einzelnen Längungsfaktoren hierbei weiter auseinander liegen als in 1a und 1b. Hypothese 2 wird verworfen; denn bei der gewählten Zusammensetzung des Stimulusmaterials zeigen alle Längungsstufen im Vergleich zur nächsthöheren Stufe eine deutliche Zunahmen der wahrgenommenen Separierung.



**Abbildung 8:** Wahrgenommene Separierung für die drei melodischen Muster.

Die in Abbildung 8 gezeigte Auswirkung des Konturverlaufs ist gering und statistisch nicht signifikant (siehe Tabelle 13).

#### 4.2.2.5 Diskussion zu Experiment 1c

Die Ergebnisse zeigen, dass die Relation der prosodischen zur syntaktischen Grenze keinen bedeutenden Einfluss auf die Urteile der Versuchspersonen hat. Die Versuchspersonen trennen also anscheinend zwischen der phonetischen Kohäsion und der syntaktischen Kohäsion und konzentrieren sich in ihrer Beurteilung auf die phonetischen Veränderungen in den Stimuli. Somit wird auch Hypothese 1 verworfen. Es zeigt sich vielmehr ein Trend in den Urteilen, der der Hypothese 1 gegenläufig ist: Die Urteile fallen an niedrigen syntaktischen Grenzen etwas niedriger aus als an hohen syntaktischen Grenzen. Möglicherweise wird unabhängig von der phonetischen Struktur also auch die syntaktische Separierung beurteilt, die unter der Bedingung Syntax A geringer ist.

### 4.3 Interpretative Analysen und Neuordnung des Kategoriensystems

Weder die labelbasierte Korpusanalyse noch die Perzeptionsexperimente geben Auskunft über die Beziehung zwischen der semantisch-pragmatischen Struktur komplexer Redebeiträge und deren phonetischer Manifestation durch unterschiedliche prosodische Grenzen. Um Erkenntnisse in diesem Bereich zu gewinnen, wird eine große Anzahl von Redebeiträgen einem interpretativen Verfahren unterzogen, in dem unter anderem beurteilt wird, ob sich an einer vorgefundenen prosodischen Grenze ein starker inhaltlicher Einschnitt, ein schwacher Einschnitt oder gar kein inhaltlicher Einschnitt befindet. Diese grobe Klassifikation dient bei der Validierung des Klassifikationssystems prosodischer Grenzen dazu, Divergenzen und Konvergenzen zwischen inhaltlicher und phonetischer Separierung zu belegen und das Klassifikationssystem entsprechend zu optimieren.

Folgende Kriterien gelten bei der interpretativen Beurteilung der inhaltlichen Struktur. Auf der niedrigsten Stufe inhaltlicher Separierung (IB-, *keine inhaltliche Blockgrenze*) kommt es am Phrasenende nicht zum Abschluss einer inhaltlichen Einheit. Die Phrase ist also kein mögliches geplantes Äußerungsende, sondern ist auf inhaltliche Komplettierung ausgerichtet. Die größte vergebene Grenze (IB+) wird als Abschluss

eines inhaltlichen Blocks bezeichnet. Hier ist der Abschluss einer inhaltlichen Einheit zu beobachten. Eine inhaltliche Einheit kann als abgeschlossen gelten, wenn sich die Funktion eines Phrasenblocks im Dialog (z.B. Aussage vs. Frage) ändert oder dem Gesprächsbeitrag eine inhaltliche Wendung durch neue Information gegeben wird. Auch der Abschluss einer pragmatischen Einheit, etwa einer Begrüßung, wird als hohe Grenze innerhalb von Äußerungen angesehen. IB+ repräsentiert also einen hohen Grad inhaltlicher Separierung. Die mittlere der drei Kategorien (IB0) ist eine negativ definierte Kategorie, die Fälle beinhaltet, welche weder IB- noch IB+ zugeordnet werden können. Dies betrifft eine in spontaner Sprache recht große Anzahl inhaltlicher Einschnitte.

Hinter dieser Einteilung steht das Problem, dass die inhaltliche Struktur spontaner Sprache oft nur sehr schwer erfassbar und klassifizierbar ist. Die vorgenommene Klassifikation hat den Vorteil, dass die Kategorien IB- und IB+ klar voneinander abgegrenzt sind und in den Untersuchungen zum Klassifikationssystem prosodischer Grenzen zumindest über diese beiden extremen Ausprägungen klare Aussagen getroffen werden können. Die Fälle in der mittleren Kategorie IB0 bilden ein Kontinuum zwischen den Extremen der fehlenden inhaltlichen Separierung (IB-) und dem Ende von inhaltlichen Blöcken (IB+).

Bei der Zuordnung der prosodischen Grenzen wird sowohl das akustische Signal als auch die orthographische Verschriftung des jeweiligen Gesprächsabschnitts herangezogen. In die interpretative Analyse gehen 60 längere Gesprächsbeiträge von verschiedenen Sprechern aus dem Terminabsprachekorpus ein. Bei der Auswahl der Gesprächsbeiträge wird darauf geachtet, dass die Beiträge mindestens eine hohe inhaltliche Grenze enthalten und mit einer weitgehend flüssigen Sprechweise produziert sind.

So entsteht für jede etikettierte Phrasengrenze in den interpretierten Gesprächsabschnitten eine Klassifikation hinsichtlich der Stärke der inhaltlichen Grenze. Diese Klassen inhaltlicher Kohäsion bzw. Separierung werden in der Auswertung der Daten auf die jeweiligen Merkmalskombinationen an den prosodischen Grenzen bezogen.

Die Neuordnung des Kategoriensystems kann sich nach Abschluss aller Forschungsarbeiten auf drei methodisch verschiedene Untersuchungsbereiche stützen:

- labelbasierte akustische Analysen der Merkmalsbündel in den Produktionsdaten,

- Perzeptionsexperimente,
- inhaltliche Interpretation von Äußerungen.

Die Ergebnisse der drei Bereiche werden aufeinander bezogen, wobei die Funktion einzelner Merkmale und ihre Rolle in Merkmalskomplexen beleuchtet wird. Im folgenden Absatz werden kurz die perzeptorischen Eigenschaften der phonetischen Merkmale sowie ihre inhaltliche Funktion diskutiert. Diese Darstellung mündet in einen Vorschlag für ein neues Kategoriensystem.

### 4.3.1 Pausen

Die bisherigen Untersuchungen zeigen einen sehr deutlichen Zusammenhang zwischen dem Auftreten von Pausen, der wahrgenommenen Separierung zwischen benachbarten prosodischen Phrasen und der inhaltlichen Struktur von Äußerungen.

Die Perzeptionsexperimente 1a, 1b und 1c belegen, dass Pausen zu einer stärkeren perzeptorischen Separierung führen als alle anderen untersuchten Merkmale. Eine Erhöhung der Pausendauern in 7 Schritten von 50 auf 890 ms führte im Perzeptionsexperiment in jedem Fall zu einer hochsignifikanten Zunahme der wahrgenommenen Separierung. Aus diesen Ergebnissen lässt sich keine Kategoriengrenze zwischen einer kurzen und einer langen Pause ableiten. Auch die Untersuchung der Pausendauern im gesamten Korpus deutet nicht auf separate Kategorien hin (z.B. durch eine Bimodalität der gemessenen Pausendauern).

Die Auswertung der interpretativen Daten belegt einen Zusammenhang zwischen dem Auftreten von Pausen und der inhaltlichen Struktur von Äußerungen. Folgende Zahlen zeigen, dass mit zunehmender Stärke inhaltlicher Einschnitte das Vorkommen von Pausen stark ansteigt.

- IB-: davon mit Pause/Atmen: 8 (26,6%),
- IB0: davon mit Pause/Atmen: 34 (40,0%),
- IB+: davon mit Pause/Atmen: 44 (68,8%).

Pausen verhalten sich somit entsprechend der Hypothese, dass wahrgenommene phonetische Separierung mit inhaltlicher Separierung einhergeht.

Weitere Untersuchungsschritte ziehen auch die Dauer der Pausen an verschieden hohen inhaltlichen Einschnitten in Betracht und belegen eine Tendenz zu größeren Pausendauern an starken inhaltlichen Einschnitten. Diese Tendenz ist allerdings so schwach ausgeprägt, dass sie keine Bedeutung für die Neukonzeption eines funktionsorientierten Klassifikationssystems hat. Ebenso spricht dieser Befund (wie auch die Perzeptionsexperimente 1a, 1b und 1c) gegen einen festen Pausengrenzwert und die darauf basierenden Einteilung in lange vs. kurze Pausen.

Die Untersuchung der Merkmalsbündel zeigt weiterhin, dass das Auftreten von Pausen fast immer an andere Merkmale gebunden ist. Nur bei 3,8% aller Pausen in beiden Korpora können keine weiteren Merkmale automatisch extrahiert werden.

### **4.3.2 Starke tonale Bewegungen**

Wie die Perzeptionstests zeigen, wirken starke tonale Bewegungen in deutlich geringerem Maße perzeptorisch separierend als Pausen. Die Interpretation der Korpusdaten belegt aber einen klaren Zusammenhang zwischen dem Auftreten starker tonaler Bewegungen am Phrasenende und hoher inhaltlicher Grenze und somit die Signalwirkung der Intonationsmuster bei der phonetischen Markierung von inhaltlichem Abschluss vs. inhaltlicher Kontinuität, wie folgende Verteilung zeigt.

- IB-: davon mit starker tonaler Bewegung: 6 (20%),
- IB0: davon mit starker tonaler Bewegung: 39 (45,9%),
- IB+: davon mit starker tonaler Bewegung: 48 (75%).

Unter den Konturen mit starker tonaler Bewegung finden sich vor allem melodische Muster, die bis an die untere Grenzen der Sprechstimme fallen (terminale Intonationsmuster). Besonders diese Kontur signalisiert meist das Ende einer inhaltlichen Einheit. Hinsichtlich starker tonaler Bewegungen gilt also auch der Zusammenhang zwischen phonetischer und inhaltlicher Separierung.

### 4.3.3 F0-Reset

Generell muss bei der Modifikation des Kategoriensystems die Funktion des Resets neu bewertet werden. Hierzu gibt es folgende Überlegungen: Reset ist eine Eigenschaft der nachfolgenden Phrase, während alle übrigen Merkmale in den Merkmalsbündeln Eigenschaften entweder der vorangehenden Phrase (Längung, Kontur) oder der Fuge (Pause/Atmen) zwischen den Phrasen beschreiben.

Die Verteilung von Phrasengrenzen mit und ohne Reset bezogen auf die inhaltlichen Kategorien zeigt folgendes Bild:

- IB-: n=30 davon mit Reset : 20 (66,7%),
- IB0: n=85 davon mit Reset : 51 (60,0%),
- IB+: n=64 davon mit Reset : 48 (75,0%).

Reset ist demzufolge in allen Kategorien ähnlich stark vertreten und ist somit kein inhaltlich funktionalisiertes Mittel in der syntagmatischen Strukturierung von Äußerungen in prosodische Phrasen. Die Ergebnisse von Pilotexperimenten zu den Experimenten 1a und 1b haben ausserdem gezeigt, dass die separierende Wirkung von F0-Reset, wenn überhaupt vorhanden, dann sehr gering ist. Diese Überlegungen führen dazu, den Reset aus dem Klassifikationssystem auszuschließen.

### 4.3.4 Längung des finalen Reims

In den Perzeptionsexperimenten 1a, 1b und 1c konnte gezeigt werden, dass Versuchspersonen bei zunehmender Längung des phrasenfinalen Reims nur in sehr geringem Maße eine zunehmende Separierung zwischen benachbarten Phrasen wahrnehmen. Weiterhin ist Längung in allen IB-Klassen ähnlich stark vertreten und somit keiner inhaltlichen Funktion im Rahmen der flüssigen turninternen Phrasierung zuzuordnen.

- IB-: n=30 davon mit Längung: 24 (80%),
- IB0: n=85 davon mit Längung: 59 (69,4%),
- IB+: n=64 davon mit Längung: 47 (73,4%).

Auf der Basis dieser Daten muss die Hypothese verworfen werden, dass eine Längung des phrasenfinalen Reims eine perzeptorische Separierung bewirkt, die wiederum eine inhaltliche Separierung signalisiert. Diese Überlegungen führen dazu, auch das Merkmal 'Längung' aus dem Klassifikationssystem auszuschließen.

#### 4.3.5 Zusammenfassung

Insgesamt sind phrasenfinale Konturverläufe und Pausen sprachlich klar funktional angebunden. Vor allem die stark fallenden Konturen stehen im Zusammenhang mit inhaltlichen Einschnitten in der Äußerung. Ebenso treten Pausen vor allem an hohen inhaltlichen Grenzen auf. Längung und F0-Reset zeigen hingegen im Bereich flüssiger prosodischer Grenzen keine erkennbare inhaltliche Funktion. Trotz der deutlichen funktionalen Bindung vor allem der stark fallenden Konturen steht aber die perzeptorische Separierung, die diese melodischen Verläufe verursacht, hinter der der Pausen zurück. Hier wird das Prinzip der Konvergenz semantisch-pragmatischer und phonetischer Kohäsion bzw. Separierung durchbrochen.

Ein funktional orientiertes Klassifikationssystem für flüssige prosodische Grenzen sollte also auf den Merkmalen F0-Kontur und Pause beruhen. So können für die extremen Kategorien der inhaltlichen Grenzen (IB- und IB+) jeweils prototypische Merkmalsbündel angesetzt werden:

- Für Phrasengrenzen ohne inhaltlichen Einschnitt (IB-): keine Pause, keine starke tonale Bewegung. Die perzeptorische Separierung ist durch phrasenfinale Längung und/oder Reset bedingt.
- Für Phrasengrenzen an hohen inhaltlichen Grenzen (IB+): separierende Kontur und Pause, wobei Pausen zu einem stärkeren perzeptorischen Einschnitt führen als separierende Konturen. Üblicherweise werden Phrasengrenzen mit Pausen und/oder separierenden Konturen durch F0-Reset und/oder segmentelle Längung verstärkt. Dieses Ergebnis deckt sich weitgehend mit Shriberg et al. (2000), die beobachten, dass sich im Englischen an prosodischen Grenzen mit nachfolgendem Themenwechsel meist eine Pause ('silent pause'), ein tiefer Grenzton und ein F0-Reset findet.
- Der Kategorie der mittelstarken inhaltlichen Einschnitte (IB0) ist weniger eindeutig ein bestimmtes Merkmalsbündel zuzuordnen.

#### 4.4 Klassifikationssystem I

##### Flüssige turninterne prosodische Grenzen

Aus den Resultaten der Untersuchungen in Kapitel 4 wird das in Tabelle 14 zusammengefasste Klassifikationssystem für flüssige turninterne prosodische Grenzen abgeleitet. Die PG-Klassen sind mit PGi1-PGi4 bezeichnet, wobei i für 'intern' steht, da mit diesem System die **turninterne**n Grenzen klassifiziert werden. Für die turnfinalen und die unflüssigen prosodischen Grenzen werden zwei weitere Klassifikationssysteme entwickelt und als System II und System III bezeichnet.

**Tabelle 14:** Klassifikationssystem I.

<i>Nomenklatur</i>	<i>Phonetische Definition</i>	<i>Funktion</i>
PGi1	<i>Kontur:</i> - <i>Pause:</i> -	Phrasierung innerhalb inhaltlicher Blöcke (IB-)
PGi2	<i>Kontur:</i> - <i>Pause:</i> +	Phrasierung innerhalb inhaltlicher Blöcke (IB-) und an schwachen inhaltlichen Grenzen (IB0)
PGi3	<i>Kontur:</i> + <i>Pause:</i> -	Phrasierung an schwachen oder starken inhaltlichen Grenzen (IB0 oder IB+)
PGi4	<i>Kontur:</i> + <i>Pause:</i> +	Phrasierung an starken inhaltlichen Grenzen (IB+)

Alle Merkmalsbündel des Systems I können von Längung des finalen Reims und/oder F0-Reset begleitet sein.

##### 4.4.1 Verteilungen der Phrasengrenzen über die inhaltlichen Klassen für das System I im Vergleich zum alten System

###### Verteilungen für das System I:

Die folgenden Tabellen vergleichen die Verteilungen der prosodischen Grenzen nach dem alten und dem neuen Klassifikationssystem. Um eine bessere Vergleichbarkeit zwischen den inhaltlichen Kategorien zu schaffen, werden die Daten auf eine Klassengröße von n=100 normiert.

**Tabelle 15:** Zuordnung der PG-Typen aus System I zu den inhaltlichen Grenzen (auf n=100 normiert).

	<i>IB-</i> (n=100)	<i>IB0</i> (n=100)	<i>IB+</i> (n=100)
<b>PGi1 (n=106,6)</b>	60 (56,3%)	34,1 (32%)	12,5 (11,7%)
<b>PGi2 (n=52,5)</b>	20 (38,1%)	20 (38,1%)	12,5 (23,8%)
<b>PGi3 (n=51,9)</b>	10 (19,3%)	24,7 (47,6%)	17,2 (33,1%)
<b>PGi4 (n=89)</b>	10 (11,2%)	21,2 (23,8%)	57,8 (64,9%)

**Tabelle 16:** Zuordnung der PG-Typen des alten Systems zu den inhaltlichen Grenzen (auf 100 normiert).

	<i>IB-</i> (n=100)	<i>IB0</i> (n=100)	<i>IB+</i> (n=100)
<b>PG0 (n=13,8)</b>	6,7 (48,6%)	7,1 (51,5%)	0 (0%)
<b>PG1 (n=151)</b>	66,7 (44,2%)	53,0 (35,1%)	31,3 (20,7%)
<b>PG2 (n=113,2)</b>	26,7 (23,6%)	36,5 (32,2%)	50,0 (44,2%)
<b>PG3 (n=8,7)</b>	0 (0%)	2,4 (27,6%)	6,3 (72,4%)
<b>PG4 (n=13,7)</b>	0 (0%)	1,2 (8,8%)	12,5 (91,2%)

Das System I kann die Kategorien PGi1 und PGi4 mit einer Wahrscheinlichkeit von 56,3% bzw. 64,9% den Kategorien IB- bzw. IB+ zuordnen. Oder: Mit einer Wahrscheinlichkeit von 60% ist die Kategorie IB- durch das Merkmalsbündel PG1 repräsentiert, mit einer Wahrscheinlichkeit von 57,8% die Kategorie IB+ durch das Merkmalsbündel PG4. Die Ergebnisse sind also in beiden Richtungen relativ gut. Das alte System führt zu starken Abweichungen: so sind z.B. 91,2% der PG4 in der Kategorie IB+, aber nur 12,5% der Kategorie IB+ durch PG4 repräsentiert.

#### 4.4.2 Vergleich von erwarteten und vorgefundenen Wahrscheinlichkeiten für das System I

Tabelle 17 zeigt einen Vergleich zwischen einer zufälligen Verteilung der PG-Klassen über die inhaltlichen Klassen und der gefundenen Verteilung. Es wird deutlich, dass vor allem die Verbindungen von PGi1 mit niedrigen inhaltlichen Grenzen und von PGi4 mit hohen inhaltlichen Grenzen stark von der zufälligen Verteilung abweicht. Für die Klassen PGi2 und PGi3 zeigen sich hingegen keine klaren Verbindungen, außer dass die Klasse PGi3 deutlich seltener an niedrigen inhaltlichen Grenzen vorkommt als bei zufälliger Verteilung.

**Tabelle 17:** Wahrscheinlichkeitstabelle für das System I. Jede Zelle enthält die folgenden vier Werte: (1) Zellenwahrscheinlichkeiten bei zufälliger Verteilung vs. gemessene Zellenwahrscheinlichkeiten und (2) erwartete Häufigkeiten bei zufälliger Verteilung (normiert) vs. gemessene Häufigkeiten (normiert).

	<i>IB- (n=100)</i>	<i>IB0 (n=100)</i>	<i>IB+ (n=100)</i>
<b>PGi1 (n=106,6)</b>	0.118 vs. 0,2 35,4 vs. 60	0.118 vs. 0,113 35,4 vs. 34,1	0.118 vs. 0,042 35,4 vs. 12,5
<b>PGi2 (n=52,5)</b>	0.058 vs. 0.067 17,5 vs. 20	0.058 vs. 0,67 17,5 vs. 20	0.058 vs. 0.042 17,5 vs. 12,5
<b>PGi3 (n=51,9)</b>	0.058 vs. 0,033 17,5 vs. 10	0.058 vs 0,082 17,5 vs. 24,7	0.058 vs 0,057 17,5 vs. 17,2
<b>PGi4 (n=89)</b>	0.1 vs. 0,033 30 vs. 10	0.1 vs. 0.071 30 vs. 21,2	0.1 vs. 0,193 30 vs. 57,8

#### 4.4.3 PG-Klassen des Systems I und die nicht-spezifizierten Merkmale Längung und Reset

Die folgenden Auswertungen der Merkmalsbündel zeigen die Verteilungen der Merkmale Längung und Reset über die Klassen Pgi1-PGi4 des Systems I.

**Tabelle 18:** Verteilung von Reset und Längung über die PG-Klassen des Systems I. Stichprobe: Terminabsprachen, Phrasengrenzen mit Resetentscheidung, n=2008. In jeder Zelle sind drei Zahlen angegeben: (1) absolute Häufigkeit, (2) relative Häufigkeit in Bezug auf n der PG-Klasse (Prozentangaben der Zeile addieren sich auf 100%), (3) relative Häufigkeit in Bezug auf die Gesamtstichprobe (n=2008, alle Zellen zusammen addieren sich auf 100%).

	<i>-L -R</i>	<i>-L +R</i>	<i>+L -R</i>	<i>+L +R</i>
<b>PGi1 (n= 761 (37,9%)) (-K -P)</b>	96 (12,6%) 4,8%	255 (33,5%) 12,7%	124 (16,3%) 6,2%	286 (37,6%) 14,2%
<b>PGi2 (n= 357 (17,8%)) (-K +P)</b>	13 (3,6%) 0,7%	44 (12,3%) 2,2%	65 (18,2%) 3,2%	235 (65,8%) 11,7%
<b>PGi3 (n= 524 (26,1%)) (+K -P)</b>	41 (7,8%) 2,0%	202 (38,6%) 10,1%	71 (13,6%) 3,5%	210 (40,1%) 10,5%
<b>PGi4 (n= 366 (18,2%)) (+K +P)</b>	8 (2,2%) 0,4%	60 (16,4%) 3,0%	33 (9%) 1,6%	265 (72,4%) 13,2%

Die Zahlen in Tabelle 18 zeigen für das Auftreten von Längung und/oder Reset keine eindeutige Bindung an die Merkmale Pause und/oder separierende Kontur. Ein Zusammenhang der Form "Wenn eine Pause auftritt, liegt auch Reset vor, und wenn keine Pause auftritt, liegt auch kein

Reset vor" wäre ein gutes (weiteres) Argument für das Weglassen von Reset aus dem Klassifikationssystem. Die einzigen deutlichen Zusammenhänge, die die Tabelle zeigt, sind, dass das Vorkommen einer Pause in über 80% aller Fälle mit Längung des finalen Reims einhergeht und dass ein kombiniertes Auftreten von separierender Kontur und Pause in 72,4% mit Längung und Reset zusammengeht. Allerdings gibt es auch viele Fälle, bei denen Reset und/oder Längung ohne Pause und/oder separierende Kontur vorkommen (vgl. hierzu auch Tabellen 6 und 7). Die hohe Auftretenshäufigkeit von Längung und Reset in der Klasse PGi4 deutet darauf hin, dass ein wirkliches "zum Ende kommen", das mit der Kategorie PGi4 verbunden ist, üblicherweise neben einer Pause und einer separierenden Kontur auch finale Längung und Reset impliziert.

Weiterhin ist auffällig, dass Phrasengrenzen ohne Reset und ohne Längung sehr selten sind (7,9% = 158 von 2008 Fällen). Von diesen 158 Fällen gibt es nur 21 Fälle, die auch eine Pause aufweisen. Wie bereits gezeigt werden konnte, stehen vor allem die Phrasengrenzen mit Pause und/oder separierender Kontur in Verbindung mit hohen inhaltlichen Einschnitten (IB+). In diesen Fällen scheint eine Ergänzung des Merkmalsbündels durch Längung und/oder Reset üblich zu sein.

Hiermit sind Pausen und melodische Merkmale am Phrasenende relativ unabhängige, funktional gebundene Gliederungsmerkmale, während Reset und Längung nur eine schwache funktionale Bindung im Bereich der prosodischen Phrasierung zeigen.

## 5 Turnfinale prosodische Grenzen

Wie im vorangegangenen Kapitel gezeigt wurde, sind die Positionierung und die phonetische Ausprägung prosodischer Grenzen eng verknüpft mit der Informationsstruktur von Äußerungen. Durch diese Verbindung zwischen inhaltlicher und phonetischer Struktur entsteht auch eine zentrale Bedeutung prosodischer Grenzen für die Organisation des Sprecherwechsels im Gespräch, denn Gesprächspartner orientieren sich bei Turnwechseln an der inhaltlichen Struktur des Gesprächs. In diesem Kapitel werden prosodische Muster am Ende von Turns analysiert und hinsichtlich interaktionaler Funktionen in Gesprächen eingeordnet.

Gespräche werden während des Sprechens geplant und strukturiert, und sowohl der Inhalt als auch die phonetische Struktur ergeben sich zu einem großen Teil spontan aus der Interaktion der Gesprächsteilnehmer. Die Fragestellungen, Hypothesen und Experimente dieser Arbeit basieren u.a. auf der Annahme, dass Hörer im Gespräch fortlaufend Erwartungen über den weiteren Äußerungsverlauf bilden, also einschätzen, ob der derzeitige Sprecher weitersprechen wird oder ob eine Übernahme des Rederechts sinnvoll ist. Hierbei orientieren sich die Hörer unter anderem an den phonetischen Eigenschaften der gehörten Äußerung (speziell im Bereich der Phrasengrenzen). Die Einschätzungen geschehen unbewusst und folgen dem Fluss des Gehörten in seiner phonetischen, morpho-syntaktischen und semantisch-pragmatischen Ausprägung. Die von Hörern gebildeten Erwartungen über den weiteren Äußerungs- bzw. Gesprächsverlauf werden während des Zuhörens fortlaufend modifiziert und evtl. verworfen. Diese Erwartungsbildung ist für den Aufbau von Dialogen entscheidend; denn sie ermöglicht schnelle Reaktionen seitens des Hörers auf das Gehörte und ist so eine Voraussetzung für die Strukturierung von Gesprächen in Turns der einzelnen Sprecher. Vor allem, wenn an einer prosodischen Grenze eine Pause auftritt, muss der Hörer richtig einschätzen, ob der Sprecher in seiner Rede fortfahren oder das Rederecht abgeben möchte, ob er eine inhaltliche Ergänzung des bereits Gesagten geben möchte oder ob er zum Abschluss einer thematischen Einheit gekommen ist und evtl. eine Reaktion des Gesprächspartners erwartet (z.B. durch ein Rezeptionssignal). Wenn die Einschätzung des weiteren Gesprächsverlaufs durch den Hörer nicht mit der Erwartung und der Planung des Sprechers übereinstimmt, steigt die Gefahr von Missverständnissen und unerwünschten Unterbrechungen durch den Dialogpartner.

Als empirische Belege für diesen Standpunkt lassen sich verschiedene Beobachtungen aus den untersuchten Produktionsdaten anführen. Die drei im Folgenden angeführten Phänomene zeigen die schnelle und zeitlich exakte Koordination zwischen den Redebeiträgen der Dialogpartner, die m.E. nur durch die vorausschauende Erwartungsbildung als grundlegenden Mechanismus in der Gesprächsorganisation erklärt werden kann.

- Unter den 377 untersuchten Sprecherwechseln im LINDENSTRASSE-Korpus gibt es 66 Turnübergänge (17,5%), an denen es weder zu einer Überlagerung von mehr als 100 ms, noch zu einer Pause von mehr als 100 ms zwischen den Turns kommt; vielmehr findet der vormalige Hörer äußerst exakt den Anschluss an den vorangehenden Turn. Diese präzise Koordination vieler Turnübergänge zeigt die große Sensibilität der Hörer für den Aufbau der Rede des Partners.
- Auch die zeitliche Synchronisierung von Rezeptionssignalen mit der syntagmatischen Struktur der fortlaufenden Äußerung zeigt, dass die Hörer Gesprächsverläufe genau mitverfolgen und sehr präzise, vor allem am Ende prosodischer Phrasen, durch kurze Äußerungen ihre Aufmerksamkeit bekunden. Der größte Teil der vorkommenden Rezeptionssignale befindet sich in kurzen Redepausen des derzeitigen Turninhabers.
- In den Dialogen kommt es zu insgesamt 20 sog. 'collaborative completions'. Das sind vom Hörer gegebene inhaltliche Ergänzungen dort, wo seitens des Sprechers Wortfindungsprobleme auftreten und der Hörer bereits weiß, was der Sprecher sagen will (Szczepek 2000). 'Collaborative completions' zeigen ebenfalls, dass Hörer sowohl inhaltlich als auch phonetisch den Ablauf der Äußerung mitverfolgen, hierbei Hypothesen über den weiteren Ablauf bilden und unter Umständen, wiederum mit sehr exaktem Timing, inhaltliche Ergänzungen liefern. Das Phänomen der 'collaborative completions' deutet vor allem darauf hin, dass Hörer Planungsprobleme seitens des Sprechers schnell erkennen und sich hierbei vermutlich auch auf typische phonetische Exponenten unflüssiger Sprechweise stützen.

Natürlich gibt es das "perfekte Gespräch" nicht. Ein Gespräch beinhaltet fast immer auch Missverständnisse, Verständigungsschwierigkeiten und konkurrierendes Verhalten um das Rederecht. Jeder kennt auch Situationen und Gespräche, in denen Hörer und Sprecher nur wenig aufmerksam

aufeinander reagieren. Wenn ein Hörer z.B. desinteressiert, müde oder geistig abwesend ist, kann es vorkommen, dass eine Äußerung einfach an ihm "vorbeirauscht". Trotzdem zeichnen sich Gespräche oft durch ein hohes Maß an Kooperationsbereitschaft und "interaktiver Feinfühligkeit" aus.

So ergeben sich folgende zentrale Fragen:

- Welche interaktionalen Funktionen können angesetzt werden, um die Koordination von Sprecher und Hörer zu erklären?
- Welche phonetischen Parameter prosodischer Grenzen stehen in Verbindung mit den angesetzten interaktionalen Funktionen?

Hinsichtlich der interaktionalen Funktionen wird hier mit den Konzepten 'Abschluss' und 'Weiterweisung' gearbeitet, die auch Gilles (2005:21) verwendet:

*"Bei der Gliederung von Gesprächsbeiträgen ('Turns') können zwei antagonistische Funktionskomplexe unterschieden werden: Einerseits werden spezifische linguistische Merkmale eingesetzt, um das Ende eines Gesprächsbeitrags zu signalisieren. Andererseits können spezifische Merkmale darauf hindeuten, dass der Sprecher seinen Beitrag noch nicht beendet hat. Diese abstrakten Funktionskomplexe werden im Folgenden mit den Begriffen 'Abschluss' und 'Weiterweisung' bezeichnet."*

Prototypische Abschlüsse sind am Ende von Gesprächsbeiträgen zu beobachten und zeigen sowohl syntaktische als auch semantisch-pragmatische Abgeschlossenheit. Im Bereich der Sprechmelodie herrscht weitgehende Übereinstimmung, dass es eine sprachübergreifende Tendenz zu stark fallenden melodischen Mustern an konversationellen Abschlüssen gibt (Delattre et al. 1965; Lieberman 1967; Cruttenden 1981; Bolinger 1978; Wichmann 2000). 'Weiterweisung' als Antagonist in der Funktionsopposition von 'Abschluss' und 'Weiterweisung' bedeutet, dass eine Fortsetzung des Turns über die erreichte prosodische Grenze hinaus angestrebt wird. Der derzeitige Sprecher ist also nicht bereit, das Rederecht abzugeben, obwohl ein gewisser Einschnitt im Turn vorliegt, und signalisiert dies dem Gesprächspartner, unter anderem durch prosodische Mittel. Das Konzept der Weiterweisung findet sich in vergleichbarer Form auch bei v. Essen (1964) als 'Progredienz', bei Delattre et al. (1965) und

Pierrehumbert und Hirschberg (1990) als 'continuation' und bei Selting (1995) als 'Turnhalten'.

Der Aufbau dieses Kapitels ist parallel zum Aufbau des Kapitels über turninterne prosodischen Grenzen. Als erstes wird anhand von Häufigkeitsverteilungen in den untersuchten Korpora die phonetische Struktur prosodischer Grenzen in turnfinaler Position beschrieben. Dem folgen perzeptive Untersuchungen zur Form-Funktions-Beziehung prosodischer Grenzen im Gespräch und anschließend Ergebnisse aus einem korpusbasierten interpretativen Verfahren. Die Ergebnisse aus den verschiedenen Untersuchungen werden zuletzt in ein phonetisch orientiertes Klassifikationssystem für turnfinale prosodische Grenzen umgesetzt.

## **5.1 Labelbasierte akustisch-phonetische Analyse der Produktionsdaten**

Die meisten Analysen beziehen sich auf die Daten aus dem LINDENSTRASSE-Korpus, da hier im Gegensatz zum Terminabsprachekorpus aufgrund des technischen Aufbaus die Aufnahme von überlagerndem Sprechen möglich war und dadurch hinsichtlich der Turnübergänge realitätsnahe Daten erhoben werden konnten.

Die automatisierten Korpusabfragen prosodischer Merkmale an turnfinalen prosodischen Grenzen umfassen die Parameter 'melodischer Verlauf', 'Längung des phrasenfinalen Reims', 'Pause' bzw. 'Überlagerung zwischen aufeinanderfolgenden Turns' und 'phonatorische Veränderungen'. Das Merkmal 'Reset' spielt in den Analysen keine Rolle, da nach einer turnfinalen prosodischen Grenze der aktuelle Sprecher aufhört zu sprechen und somit kein Bezugspunkt für einen nachfolgenden Reset gegeben ist. Folgende Hypothesen werden überprüft:

### **Hypothese 1**

Turnfinale Merkmalskomplexe an Äußerungsgrenzen unterscheiden sich von den Merkmalen turninterner prosodischer Grenzen. Hierbei gibt es allerdings starke Überschneidungen zwischen den Merkmalsbündeln, so dass kein Merkmalskomplex ausschließlich an einer der beiden Positionen vorkommt.

### **Hypothese 2**

Die phonetischen Merkmale turnfinaler Phrasengrenzen unterscheiden sich, je nachdem, ob es sich um Turnübergänge mit oder ohne überlagerndes Sprechen handelt.

### 5.1.1 Pausen und Atmungsphasen

Pausen (und Atmungsphasen) zwischen den Redebeiträgen verschiedener Dialogpartner sind prinzipiell etwas anderes als Pausen an prosodischen Grenzen im Inneren von Redebeiträgen. Ob eine Pause auftritt und wie lang diese Pause ist, wird im Inneren von Turns vom derzeitigen Sprecher bestimmt und steht in Zusammenhang mit der inhaltlichen Gliederung des Gesprochenen oder mit Störungen im Sprechfluss. Pausen zwischen den Redebeiträgen verschiedener Sprecher im Dialog sind eine Folge der Interaktion zwischen den Gesprächspartnern. Sie können nur entstehen, wenn Sprecher A aufhört zu sprechen und Sprecher B nach einer Pause das Rederecht übernimmt. Unter diesem Gesichtspunkt können drei Arten von Turnübergängen unterschieden werden. Erstens diejenigen mit einer Pause zwischen den Redebeiträgen der Sprecher. Zweitens Übergänge zwischen Redebeiträgen, in denen Sprecher B beginnt zu sprechen, bevor Sprecher A seine Äußerung beendet hat und es zu gleichzeitigem (überlagerndem) Sprechen kommt. Der dritte zu unterscheidende Fall liegt in der Mitte zwischen den Turnübergängen mit Pause und den Turnübergängen mit Überlagerung. Hier kommt es zu einem unmittelbaren Anschluss der Rede von Sprecher B an den Redebeitrag von Sprecher A, so dass weder eine Pause noch eine Überlagerung entsteht. Da sich bei einer präzisen akustischen Analyse der Übergänge zwischen den Sprachsignalen der Sprecher jedoch praktisch immer entweder eine Pause oder eine Überlagerung zeigt, wird für die dritte Kategorie ein oberer Grenzwert von 100 ms Pausen- bzw. Überlagerungsdauer gewählt. Dieser Wert liegt im Bereich der Dauer von einzelnen Segmenten und eine informelle auditive Analyse zeigt, dass Pausen und Überlagerungen unter 100 ms als unmittelbarer Anschluss wahrgenommen werden. Mit einem Grenzwert von 100 ms (allerdings für Pausen innerhalb von Redebeiträgen) arbeiten auch Koiso et al. (1998) und Wichmann und Caspers (2001).

Um die unterschiedlichen Synchronisationen der Redebeiträge im Korpus messen zu können, werden in einer manuellen Etikettierung sämtliche Überlagerungen und Pausen an Turnübergängen durch Zeitmarken abgegrenzt. In der späteren Verarbeitung der Etikettierdateien kann aus diesen Etiketten die Dauer von Pausen bzw. Überlagerungen berechnet und in die automatisch generierten Merkmalsbündel integriert werden.

In den Daten des LINDENSTRASSE-Korpus können für insgesamt 377 turnfinale prosodische Grenzen Merkmalsbündel abgeleitet werden, die

Informationen über die Längung des finalen Reims, das melodische Muster am Phrasenende und die Dauer von Überlagerungen oder Pausen zwischen den Turns enthalten. Die Verteilung der Sprecherwechsel über die drei oben beschriebenen Kategorien von Turnübergängen im LINDENSTRASSE-Korpus sehen folgendermaßen aus:

- Sprecherwechsel mit Pause oder Atmen zwischen den Redebeiträgen: 183 Fälle (48,5%). Die durchschnittliche Pausendauer beträgt 704 ms mit einer Standardabweichung von 702 ms.
- Sprecherwechsel mit unmittelbarem Anschluss (+-100 ms): 66 Fälle (17,5%).
- Sprecherwechsel mit Überlagerung: 128 Fälle (34%). Die durchschnittliche Dauer der Überlagerungen liegt bei 480 ms mit einer Standardabweichung von 375 ms.

Diese drei akustisch definierten Klassen, die sich in der temporalen Koordination der Turnübergänge unterscheiden, geben einen ersten Ansatz für die korpusbasierte Differenzierung funktional verschiedener Arten von Sprecherwechseln. Es wird angenommen, dass sich in der Klasse der überlagernden Übergänge vor allem Turnwechsel finden, die durch den Hörer eingeleitet werden und unter Umständen den Sprecher unterbrechen, ohne dass dies von ihm intendiert ist. Bei den Übergängen mit Pausen und mit unmittelbarem Anschluss hingegen kann eher davon ausgegangen werden, dass der Sprecher zu einem Abschluss gekommen ist, dies phonetisch signalisiert und dem Hörer das Wort überlässt. Unter diesem Gesichtspunkt wird in den folgenden Darstellungen der Merkmale und Merkmalsbündel an turnfinalen prosodischen Grenzen zwischen diesen drei Arten von Sprecherwechseln unterschieden, um zu untersuchen, ob sich andere phonetische Merkmale in Abhängigkeit von der Art der Synchronisierung der aufeinander folgenden Turns unterschiedlich ausprägen. Ist dies der Fall, können Hinweise auf interaktional unterschiedliche Funktionen der drei Arten von Turnübergängen abgeleitet werden.

### 5.1.2 Melodische Muster

Tabelle 19 zeigt die phrasenfinalen melodischen Muster turnfinaler prosodischer Grenzen im LINDENSTRASSE-Korpus. Alle hier untersuchten

Phrasengrenzen sind syntaktisch abgeschlossen und zeigen keine Zögerungslängung am Ende der Phrase.

**Tabelle 19:** Melodische Verläufe am Turnende (LINDENSTRASSE).

Melodischer Verlauf	Pause (n=183)	Unmittelbarer Anschluss (n=66)	Überlagerung (n=128)	Gesamt (n=377)
<b>Starke melodische Bewegung</b>				
Stark fallend	53 (29,0%)	23 (34,9%)	35 (27,3%)	111 (29,4%)
Stark fallend + sehr leicht steigend	12 (6,6%)	4 (6,1%)	5 (3,9%)	21 (5,6%)
Fallend-steigend	12 (6,6%)	6 (9,1%)	5 (3,9%)	23 (6,1%)
Stark Steigend	13 (7,1%)	4 (6,1%)	3 (2,3%)	20 (5,3%)
Gesamt starke melodi- sche Bewegung	90 (49,2%)	37 (56,1%)	48 (37,5%)	175 (46,4%)
<b>Schwache melodische Bewegung</b>				
Eben	25 (13,7%)	7 (10,6%)	19 (14,8%)	51 (13,5%)
Leicht fallend	20 (10,9%)	9 (13,6%)	34 (26,6%)	63 (16,7%)
Leicht bis mittelhoch steigend	37 (20,2%)	9 (13,6%)	11 (8,6%)	57 (15,1%)
Eben + sehr leicht stei- gend	3 (1,6%)	2 (3,0%)	5 (3,9%)	10 (2,7%)
Leicht fallend + sehr leicht steigend	8 (4,4%)	2 (3,0%)	11 (8,6%)	21 (5,6%)
Gesamt schwache mel. Bewegung	93 (50,8%)	29 (43,9%)	80 (62,5%)	202 (53,6%)
<b>Gesamt</b>	<b>128 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>	<b>377 (100%)</b>

In allen drei Klassen ist die stark fallende Kontur am häufigsten vertreten. Die Rangliste der häufigsten Konturen in den drei Klassen zeigt aber einen deutlichen Unterschied der Verteilungen: Während in den Klassen mit Pause und unmittelbarem Anschluss neben den stark fallenden Konturen meistens leicht bis mittelhoch steigende und ebene Konturen vorkommen, treten bei überlagernden Übergängen neben den stark fallenden Konturen vor allem leicht fallende Konturen auf. Alle starken melodischen Bewegungen zusammengenommen zeigen eine Progression von Turnübergängen mit unmittelbarem Anschluss (56,1% starke melodische Bewegungen), über Übergänge mit Pause (49,2%), hin zu den überlagernden Übergängen mit nur 37,5% starken melodischen Bewegungen.

Tabelle 20 zeigt zum Vergleich die turnfinalen melodischen Muster im Terminabsprachekorpus. Dieser Vergleich ist insofern interessant, als jeder Turnübergang in diesen Daten als gezielter Turnabschluss durch den aktuellen Sprecher eingeleitet wird – also als intendiert angesehen werden kann. Dies ist eine Folge des technischen Aufbaus der Aufnahmen: Die Sprecher konnten hierbei nur wechselseitig sprechen, solange sie einen Knopf gedrückt hielten. Das Halten des Knopfs blockierte den Kanal des Partners. Wenn ein Sprecher also seinen Turn für beendet hielt, wurde der Knopf losgelassen und der Partner konnte seinen Knopf drücken und antworten. Dieses technische Szenario ist zwar unnatürlich, bietet aber durch die Aufnahmesituation ein externes Kriterium, das es zulässt, jedes Turnende als gezielten Abschluss zu klassifizieren. In den Daten mit überlagerndem Sprechen ist dies anders, denn der Hörer kann hier jederzeit das Wort übernehmen (z.B. in einer kurzen Sprechpause des Sprechers). Die Entscheidung, ob eine Übernahme vom Sprecher erwünscht war, ist damit immer eine Frage der Interpretation.

**Tabelle 20:** Melodische Verläufe am Turnende (Terminabsprachen).

Melodischer Verlauf	Absolute und relative Häufigkeiten
<b>Starke melodische Bewegung</b>	
Stark fallend	536 (49,1%)
Stark fallend + sehr leicht steigend	161 (14,7%)
Fallend-steigend	46 (4,2%)
Stark Steigend	112 (10,3%)
Gesamt starke melodische Bewegung	855 (78,3%)
<b>Schwache melodische Bewegung</b>	
Eben	48 (4,4%)
Leicht fallend	45 (4,1%)
Leicht bis mittelhoch steigend	117 (10,7%)
Eben + sehr leicht steigend	6 (0,5%)
Leicht fallend + sehr leicht steigend	21 (1,9%)
Gesamt schwache melodische Bewegung	237 (21,6%)
<b>Gesamt</b>	<b>1092 (100%)</b>

Der Vergleich von Tabelle 19 und Tabelle 20 zeigt, dass die Verteilungen der melodischen Muster in den beiden Korpora starke Unterschiede aufweisen. Etwa die Hälfte aller turnfinalen Muster in den Terminabsprachedaten sind stark fallend und fast 80% haben starke melodische Bewegungen. In den LINDENSTRASSE-Daten gibt es über alle Kategorien knapp

30% fallende Muster und insgesamt 46,4% starke melodische Bewegungen. Die schwachen melodischen Bewegungen spielen bis auf die leicht bis mittelhoch steigenden Konturen in den Terminabsprachedaten praktisch keine Rolle, während sie in den LINDENSTRASSE-Daten zusammen fast 30% aller gefundenen Verläufe ausmachen.

Die Zahlen zeigen somit deutlich, dass bei Turnwechseln, die vom Sprecher intendiert sind, in den meisten Fällen starke melodische Bewegungen am Turnende eingesetzt werden. Da in den Daten aus dem Videotaskkorpus in allen drei akustisch definierten Klassen von Turnübergängen wesentlich weniger starke melodische Bewegungen als in den Terminabsprachedaten vorkommen, ist anzunehmen, dass jede der drei Klassen sowohl intendierte als auch nicht intendierte Sprecherwechsel enthält. Das bedeutet, dass das Kriterium von Pause vs. Überlagerung nicht hinreichend ist, eine funktionale Differenzierung intendierter vs. nicht intendierter Turnwechsel zu tragen, obwohl sich eine Tendenz in diese Richtung zeigt.

### 5.1.3 Segmentelle Längung

Das unter 4.1.7 beschriebene Messverfahren erlaubt Aussagen über das Vorhandensein von segmenteller Längung an prosodischen Grenzen. Tabelle 21 zeigt wie häufig turnfinale Reime in den drei Klassen gelängt sind.

**Tabelle 21:** Segmentelle Längung des turnfinalen Reims (LINDENSTRASSE).

	<b>Pause (n=183)</b>	<b>Unmittelbarer Anschluss (n=66)</b>	<b>Überlagerung (n=128)</b>
Kürzung	1 (0,5%)	-	2 (1,5%)
Weder Längung noch Kürzung	47 (25,7%)	18 (27,3%)	40 (31,3%)
Längung	135 (73,8%)	48 (72,7%)	86 (67,2%)
<b>Gesamt</b>	<b>183 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>

Im Gegensatz zu den melodischen Konturen gibt es keine klare Abhängigkeit zwischen den drei differenzierten Klassen von Turnübergängen und dem Auftreten von finaler Längung. In allen drei Klassen liegt der Anteil von gelängten Reimen bei rund 70%. Die Daten zeigen lediglich einen leichten Trend zu weniger Längung im Falle der Übergänge mit Überlagerung. Die gemessenen Werte sind den Ergebnissen der turnfinalen prosodischen Grenzen im Terminabsprachekorpus sehr ähnlich. Hier wurde finale Längung in 74,3% der Fälle gemessen (Peters et al. 2005a).

#### 5.1.4 Phonation

Neben den oben untersuchten Merkmalen finden sich in der Literatur Hinweise auf phonatorische Veränderungen am Ende prosodischer Phrasen. Die vorliegende Untersuchung soll zeigen, welche Rolle Knarrstimme an turnfinalen prosodischen Grenzen im LINDENSTRASSE-Korpus spielt.

Aus Untersuchungen zum Englischen (Local und Kelly 1986; Nakatani und Hirschberg 1994) ist bekannt, dass Knarrstimme an prosodischen Grenzen auftreten kann. Local und Kelly beschreiben zwei funktional unterschiedliche Bereiche, in denen dieser Phonationstyp eingesetzt wird. Zum einen enden unvollständige, vom Sprecher abgebrochene Äußerungen z.T. mit Knarrstimme. In solchen Fällen folgt häufig ein Glottalverschluss, nach dessen Ende der Sprecher seine Äußerung fortsetzt. In diesem Kontext wird die Glottalisierung als phonetisches Signal eingesetzt, um anzuzeigen, dass der Sprecher mit seinem Redebeitrag noch nicht zum Ende gekommen ist und beabsichtigt, die abgebrochene Äußerung fortzusetzen. Die Funktion der nicht-modalen Phonation ist hier also, den Anspruch auf das weitere Rederecht anzuzeigen. Local und Kelly bezeichnen diese Art der Glottalisierung deshalb als 'turn-holding-device', Nakatani und Hirschberg als 'interruption glottalization', die als 'edit-signal' das Ende eines Reparandums markiert.

Der zweite Bereich, in dem oft Knarrstimme vorkommt, ist am Ende kompletter prosodischer Phrasen, die mit terminaler Intonation (Absinken der Grundfrequenz in den untersten Bereich der Sprechstimme) produziert werden. Hier ist die Funktion gerade umgekehrt. Der Sprecher zeigt durch dieses sprachliche Verhalten an, dass er zum Ende der Äußerung gekommen ist und der Gesprächspartner das Rederecht bekommt. Folglich wird diese Verwendung der Knarrstimme von Local und Kelly als 'turn-yielding-device' bezeichnet.

Weder Local und Kelly noch Nakatani und Hirschberg treten allerdings einen Nachweis über die perzeptorische Relevanz der phonatorischen Veränderungen an. Sie leiten die postulierten Funktionen lediglich daraus ab, dass die unterschiedlichen Typen von Knarrstimme in den untersuchten Produktionsdaten komplementär verteilt auftreten. Aus unterschiedlichen Verteilungseigenschaften allein kann m.E. aber noch keine perzeptorische und funktionale Relevanz abgeleitet werden. Aus diesem Grund wird in den Perzeptionsexperimenten 2 und 3a gezielt die funktionale Zuordnung phonatorischer Veränderungen geprüft (siehe 5.2 und 6.2).

Kohler et al. (2005) untersuchen 22 Sprecher und Sprecherinnen des *Kiel Corpus of Spontaneous Speech* Vol. I und II auf vergleichbare phonetische Strukturen. Generell können die von Local und Kelly bzw. Nakatani und Hirschberg beschriebenen Verwendungen von Knarrstimme auch in den deutschen Daten des Terminabspracheszenario nachgewiesen werden, indem zwei Datensätze miteinander verglichen werden. Der erste Datensatz enthält etwa 300 prosodische Phrasen, die mit einem syntaktischen Abbruch enden, der nicht durch eine Unterbrechung durch den Dialogpartner hervorgerufen wurde. Der zweite Datensatz beinhaltet über tausend prosodische Phrasen ohne syntaktische Abbrüche, die mit terminaler Intonation enden. Die Sprachsignale beider Datensätze werden anhand der segmentellen und prosodischen Etikettierung aus dem Gesamtkorpus extrahiert. Für diese zwei Datensätze wird das Ende jeder extrahierten prosodischen Phrase hinsichtlich des Vorkommens von Knarrstimme und/oder Glottalverschlüssen manuell etikettiert.

Die Auswertung der Etikettierung zeigt, dass am Ende prosodischer Phrasen mit syntaktischen Abbrüchen in 27% der Fälle Knarrstimme vorkommt, wobei diese in etwa einem Drittel der Fälle in Verbindung mit einem Glottalverschluss steht. Die durchschnittliche Dauer des glottalisierten Signalabschnitts liegt bei 76 ms. 46% der syntaktischen Abbrüche sind von Pausen oder Atmen gefolgt. Die phonetische Realisierung der abgebrochenen Phrasen unterliegt starken individuellen Schwankungen. Der Datensatz mit den terminalen Intonationsmustern zeigt eine Kopplung dieses melodischen Musters mit Knarrstimme in insgesamt 32% der Fälle, hier allerdings in weniger als einem Prozent der Fälle in Verbindung mit einem Glottalverschluss. Eine weitere Differenzierung der terminalen Intonationsmuster bezüglich ihrer Position in der Äußerung liefert ein sehr deutliches Ergebnis: In turnfinaler Position findet sich geknarrte Stimme in 57% der Fälle, in turninterner Position hingegen nur in 27%. Die durchschnittliche Dauer der nicht-modalen Phonation beträgt 164 ms. Auch bei der Auftretenshäufigkeit von Laryngalisierung gibt es eine deutliche Sprecherabhängigkeit.

Zusammenfassend lassen sich also zwei Unterschiede zwischen Knarrstimme an Abbruchstellen und an prosodischen Grenzen mit terminaler Intonation feststellen. An Abbruchstellen alterniert die Knarrstimme mit Glottalverschlüssen, was bei terminaler Intonation kaum vorkommt. Außerdem ist die Dauer der Knarrstimme an Abbrüchen deutlich kürzer. Kohler et al. (2005) beschreiben weiterhin, dass die auditive Analyse vieler Fälle den Eindruck vermittelt, dass die Knarrstimme an Abbruchstellen eher fest und angespannt klingt, während sie am Ende termi-

naler Intonationsmuster weniger fest und entspannt klingt. Aus dieser Reihe von Beobachtungen ergibt sich eine Trennung in zwei Phänomenbereiche, erstens die Abbruchglottalisierung ('interruption glottalisation') und zweitens die phrasenfinale bzw. turnfinale Laryngalisierung ('final laryngealisation').

Da Glottalisierung und Laryngalisierung vermutlich relevante phonetische Signale für die Dialogsteuerung sind, soll nun auch das LINDENSTRASSE-Korpus auf das Vorkommen dieser phonatorischen Signale untersucht werden. Diesmal allerdings weniger in Bezug zu syntaktischen Abbrüchen, als vielmehr hinsichtlich der Funktion von Knarrstimme als phonetische Markierung von Turnabschlüssen und somit als Korrelat der Abschlussfunktion.

Da über die akustischen Daten nicht zuverlässig das Vorkommen von Knarrstimme bestimmt werden kann, ist für diesen Teil der Korpusanalyse eine weitere manuelle Etikettierung notwendig, die nach folgende Kriterien vorgeht: Für den Phänomenbereich der Knarrstimme in Verbindung mit terminalen Intonationskonturen wird im ersten Arbeitsschritt jedes Vorkommen des stark fallenden Musters anhand der prosodischen Etikettierung aus dem Korpus automatisch herausgesucht. Die Verbindung zwischen dem Auftreten von Laryngalisierung und stark fallenden melodischen Verläufen wird im Vorfeld durch das gezielte Abhören der Dialoge bestätigt. Hierbei zeigt sich, dass Laryngalisierungen fast nie in Verbindung mit anderen melodischen Konturen stehen.

Jedes Vorkommen stark fallender Intonation wird dann nach den unter 4.1.9 eingeführten Kriterien beurteilt.

**Tabelle 22:** Phonationstyp bei stark fallendem melodischen Verlauf in turnfinaler Position (LINDENSTRASSE).

	<b>Pause (n=53)</b>	<b>Unmittelbarer Anschluss (n=23)</b>	<b>Überlagerung (n=35)</b>
Modalstimme	30 (56,6%)	14 (60,9%)	27 (77,1%)
Knarrstimme (unregelmäßig)	15 (28,3%)	7 (30,4%)	4 (11,4%)
Knarrstimme (regelmäßig)	2 (3,8%)	0 (0%)	2 (5,7%)
Interferenz oder Unsicherheit	6 (11,3%)	2 (8,7%)	2 (5,7%)
Glottalverschluss	0 (0%)	0 (0%)	0 (0%)
<b>Gesamt</b>	<b>53 (100%)</b>	<b>23 (100%)</b>	<b>35 (100%)</b>

Tabelle 22 zeigt trotz der relativ geringen Anzahl von Fällen eine Tendenz zum verstärkten Vorkommen von Knarrstimme an Turnübergängen mit Pausen und mit unmittelbarem Anschluss, gegenüber Übergängen

mit Überlagerungen. Möglicherweise ist dieses gehäufte Vorkommen darauf zurückzuführen, dass in den Klassen ohne Überlagerung mehr intendierte Turnübergaben vorkommen, wobei die Laryngalisierung eines der phonetischen Merkmale der Abschlussfunktion ist. Auch der Vergleich mit Laryngalisierungen in turninterner Position belegt diese Deutung, denn hier tritt unregelmäßige Knarrstimme nur in knapp 11% der Fälle auf (siehe 4.1.9, Tabelle 8).

Die durchschnittliche Dauer von Knarrstimme beträgt über die drei Klassen von Synchronisierungen gemittelt 152 ms und ist damit etwas länger als mit 121 ms in turninterner Position. Aus dieser durchschnittlichen Dauer kann abgeleitet werden, dass die phonatorische Veränderung meist auf den finalen Reim beschränkt ist. In keinem Fall wurde ein Glottalverschluss in turnfinaler Position festgestellt.

In den Daten aus dem Terminabspracheszenario wurden allerdings an 57% aller Turnabschlüsse Laryngalisierungen gefunden. Dies legt nahe (wie schon die unterschiedlichen Verteilungen der melodischen Muster), dass viele der Turnübergänge in den LINDENSTRASSE-Daten nicht explizit durch den Sprecher eingeleitet werden.

### 5.1.5 Merkmalsbündel an turnfinalen prosodischen Grenzen

Tabelle 23 fasst die Bündelungen der Merkmale 'starke tonale Bewegung' und 'Längung' für die drei Klassen von Übergängen zusammen.

**Tabelle 23:** Merkmalsbündel an turnfinalen prosodischen Grenzen.

	<b>Mit Pause (n=183)</b>	<b>Ohne Pause (n=66)</b>	<b>Mit Überlagerung (n=128)</b>
Kein extrahierbares Merkmal	26 (14,2%)	7 (10,6%)	28 (21,9%)
Nur Längung	67 (36,6%)	22 (33,3%)	52 (40,6%)
Nur starke tonale Bewegung	22 (12,0%)	11 (16,7%)	14 (10,9%)
Längung und starke tonale Bewegung	68 (37,2%)	26 (39,4%)	34 (26,6%)
<b>Gesamt</b>	<b>183 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>

Die gefundenen Merkmalsbündel zeigen keine weiteren klaren Unterschiede zwischen den drei Klassen von Turnübergängen. Wie auch in den Tabellen 19 und 20, ist der deutlichste Unterschied zwischen den Klassen die geringere Häufigkeit starker melodischer Bewegungen in der Klasse von Turnübergängen mit Überlagerungen. Neben diesem Ergebnis fällt vor allem auf, dass die Anzahl von Fällen, in denen die automati-

sche Auswertung der Etikettierung kein Merkmal extrahieren kann, für die Klasse der überlagernden Übergänge mit 21,9% mit Abstand am größten ist.

### 5.1.6 Diskussion

Ob Hypothese 1 (Turnfinale Merkmalskomplexe an Äußerungsgrenzen unterscheiden sich von den Merkmalen turninterner prosodischer Grenzen. Hierbei gibt es allerdings starke Überschneidungen zwischen den Merkmalsbündeln, so dass kein Merkmalskomplex ausschließlich an einer der beiden Positionen vorkommt.) als bestätigt angesehen werden kann, ist maßgeblich davon abhängig, welche Vergleiche vorgenommen werden. Die Verteilung der melodischen Muster an turninternen prosodischen Grenzen (Tabelle 2) unterscheidet sich stark von der Verteilung der melodischen Muster an turnfinalen Grenzen im Terminabsprachekorpus (Tabelle 20). Dies würde für die Annahme der Hypothese sprechen. Vergleicht man allerdings die Verteilung in Tabelle 2 mit den Häufigkeiten melodischer Muster am Turnende in den LINDENSTRASSE-Daten (Tabelle 19), zeigen sich keine klaren Unterschiede. Auch die weiteren Merkmale (segmentelle Längung und Phonationstyp) lassen keine eindeutige Bestätigung von Hypothese 1 zu.

Die Auswertung der phonetischen Merkmale in den Labeldateien zeigt lediglich Tendenzen, die auf funktionale Klassen hinweisen. Die drei Klassen von Turnübergängen (mit und ohne Überlagerung bzw. mit unmittelbarem Anschluss) weisen so starke Überschneidungen der Merkmalsbündel an den Turnenden auf, dass die durchgeführte Analyse allein keine Zuordnung bestimmter phonetischer Merkmalskomplexe zur Abschluss- bzw. Weiterweisungsfunktion erlaubt. Hypothese 2 (Die phonetischen Merkmale turnfinaler Phrasengrenzen unterscheiden sich, je nachdem, ob es sich um Turnübergänge mit oder ohne überlagerndes Sprechen handelt.) kann also nur im Ansatz bestätigt werden. Folglich kann durch die Ergebnisse auch keine Trennung zwischen gezielt vom Sprecher eingeleiteten Turnübergängen und vom Hörer herbeigeführten Turnübernahmen abgeleitet werden. Die Daten deuten eher darauf hin, dass in allen drei Klassen von Übergängen sowohl Turnübergänge enthalten sind, die vom Sprecher eingeleitet sind als auch solche, die der Hörer forciert, indem er "unaufgefordert" zu sprechen beginnt (z.B. in einer kurzen Redepause).

Eine Erweiterung der Korpusabfragen könnte die Ergebnisse entscheidend verändern. Die bisherige labelbasierte Analyse der finalen melodischen Muster bezieht sich nur auf den melodischen Verlauf im Anschluss an die letzte satzakzentuierte Silbe vor der jeweiligen prosodischen Grenze. Es ist aber bekannt, dass die Synchronisierung des F0-Maximums nuklearer Gipfelkonturen bzw. des F0-Minimums nuklearer Talkonturen unabhängig von der nachfolgenden tonalen Bewegung zu semantischen Differenzierungen führt (Kohler 1991b). Besonders deutlich wird dies bei einer Positionierung des Gipfels vor der satzakzentuierten Silbe ('früher Gipfel'), im Gegensatz zu einer Positionierung nach der satzakzentuierten Silbe ('später Gipfel'). Der frühe Gipfel bringt eine retrospektive, zusammenfassende und abschließende semantisch-pragmatische Bedeutungskomponente in die Äußerung, während der späte Gipfel eröffnend wirkt, die gegebene Information als "neue Information" charakterisiert und den Eindruck von Überraschtheit hereinbringen kann. Die hierzu durchgeführten Experimente beziehen sich allerdings nicht auf dialogbezogene interaktionale Funktionen. Es ist aber anzunehmen, dass verschieden Gipfeltypen auch in der Signalisierung von interaktionalen Abschlüssen bzw. von Weiterweisung funktional relevant sind. In weiteren Arbeiten sollte also der nukleare Akzent und der nachfolgende melodische Verlauf als globales Muster betrachtet werden, dessen Funktion sich aus verschiedenen melodischen und temporalen Parametern der Gesamtkontur ableitet.

Weiterhin fehlt in der phonetischen Analyse das Merkmal 'Intensitätsverlauf', das evtl. weitere wichtige Information für die Steuerung von Turnübergaben trägt (French und Local 1983; Local und Kelly 1986). Auch eine durchgehende Analyse sämtlicher phonatorischer Merkmale wie Verhauchung und Registerwechsel könnte zu klareren Ergebnissen der Produktionsdatenanalyse führen.

## **5.2 Perzeptionsexperiment zu turnfinalen prosodischen Grenzen (Experiment 2)**

### **5.2.1 Einleitung**

Im Bereich der Dialogsteuerung wurden bisher nicht viele phonetisch ausgerichtete Perzeptionsexperimente durchgeführt. Dies liegt wahrscheinlich daran, dass neben phonetischen Faktoren auch der linguistische und situative Kontext einen entscheidenden Einfluss auf den Dialogverlauf hat und eine systematische Erfassung dieser Variablen in Experimenten kaum zu leisten ist. Trotz dieser generellen Problematik hat Caspers (2000) ein Experiment zum Turntaking im Holländischen konzipiert und durchgeführt, das eine Kombination aus der Verwendung natürlichsprachlicher Stimuli und funktionaler Beurteilungen im Experiment ermöglicht. Dieses Experiment wird von Wichmann und Caspers (2001) für das Englische reproduziert. Da in den Experimenten zu den beiden Sprachen sowohl das generelle Vorgehen als auch die Ergebnisse übereinstimmen, wird hier nur auf das Experiment zum Englischen im Detail eingegangen.

Der Ausgangspunkt für die Experimente ist, dass nach Ford und Thompson (1996) ein Sprecherwechsel im Dialog dann eintreten kann, wenn melodische, syntaktische und pragmatische Komplettierung zusammenfallen. Wichmann und Caspers (2001) überprüfen diese Hypothese, indem sie Dialogfragmente unterschiedlicher phonetischer und syntaktischer Ausprägung in zwei Experimenten hinsichtlich ihrer Funktion für die Dialogsteuerung beurteilen lassen.

Für den ersten Teil des Experiments isolieren die Autoren eine Reihe von 8-13 Sekunden langen Fragmenten aus südeinglischen Maptask-Dialogen (Anderson et al. 1991). Alle Fragmente haben die Eigenschaft, dass vor deren Beginn und nach deren Ende im Turn eine Pause auftrat. Die Fragmente sind also sog. 'interpausal units' (Koiso et al. 1998). Weiterhin werden die Stimuli so ausgewählt, dass folgende Parameter in systematischer Variation im Stimulusmaterial vertreten sind:

- syntaktische Abgeschlossenheit vs. syntaktische Unabgeschlossenheit an der Grenze des Fragments,
- nachfolgender Turnwechsel vs. Weitersprechen des aktuellen Sprechers im Ursprungsdialog.

Aus der Kombination dieser Kriterien ergeben sich vier Bedingungen hinsichtlich der syntaktischen Abgeschlossenheit und des weiteren Dialogverlaufs. Für jede der vier Bedingungen werden (soweit vorhanden) aus dem Korpus jeweils zwei Ausschnitte herausgesucht, die mit einem der folgenden fünf Intonationsmuster enden:

- 'high rise' (H\* H%)
- 'high level' (H\* %)
- 'fall-rise' (H\*L H%)
- 'fall' (H\*L L%)
- 'truncated fall' (H\*L %)

Bei der Auswahl der Stimuli wird darauf geachtet, dass die Fragmente im Ursprungsdialog keine Frage an den Gesprächspartner darstellen. Im Fall der hoch steigenden Kontur gibt es im Korpus nicht genügend Fälle, die den genannten Kriterien entsprechen. Hier wird deshalb nur mit sechs Ausschnitten gearbeitet (zwei in deklarativer, zwei in interrogativer Funktion und zwei syntaktische Fragmente). Insgesamt kommen so 38 Stimuli zusammen (statt 40 Stimuli bei kompletter Symmetrie).

Die Hörer hören jedes Fragment zweimal und beurteilen dann, wie der Dialog, aus dem das Fragment stammt, nach dem vorgespielten Ausschnitt vermutlich weitergehen wird. Es gibt drei Antwortkategorien:

- Der Sprecher spricht weiter.
- Der Sprecher spricht nach einer kurzen Bestätigung durch den Hörer weiter ('Rezeptionssignal').
- Der Turn wird vom Dialogpartner übernommen.

Die Ergebnisse des Hörtests zeigen, dass die Beurteilung der Funktion einer prosodischen Grenze stark von der syntaktischen Abgeschlossenheit an der Stelle des prosodischen Einschnitts abhängig ist. Wenn die präsentierte Äußerung syntaktisch nicht vollständig ist, vermuten die Versuchspersonen in insgesamt 98% der Urteile, dass der Sprecher weiterspre-

chen wird. In 24% der Stimuli wird hierbei angenommen, dass dem Weitersprechen ein Rezeptionssignal vorangehen könnte. Eine systematische Auswirkung der unterschiedlichen Intonationskonturen auf die Urteile lässt sich nicht erkennen.

Bei syntaktischer Abgeschlossenheit wird mit einer relativen Häufigkeit von 62% geurteilt, dass derselbe Sprecher weitersprechen wird (hiervon 30% mit vorangehendem Rezeptionssignal). Die Urteile zeigen hierbei eine klare Präferenz des Urteils 'Sprecher spricht weiter' wenn das Fragment mit H\* % endet (89%). Weiterhin ist interessant, dass bei syntaktischer Vollständigkeit für die Muster H\* H%, H\*L L% und H\*L H\* jeweils etwa die Hälfte der Urteile für einen unmittelbaren Sprecherwechsel abgegeben werden und die andere Hälfte sich auf die anderen beiden Kategorien verteilt. Das Experiment zeigt also erstens die primäre Wichtigkeit der syntaktischen Struktur bei der Beurteilung des Dialogfortgangs und zweitens die fehlende Eindeutigkeit bei der Form-Funktions-Zuordnung melodischer Muster durch Hörer. Die einzige Ausnahme bildet der Verlauf H\* %, der als klares Signal des Turnhaltens aufgefasst wird. Vor allem ist auffällig, dass auch das stark fallende Muster H\*L L% nur in etwa der Hälfte der Urteile als Turnabschluss beurteilt wird.

Im zweiten Teil des Experiments werden alle Stimuli aus dem ersten Experiment einbezogen, die einen nachfolgenden Sprecherwechsel haben. Die Stimuli werden den Versuchspersonen zusammen mit dem Anfang der Folgeäußerung zusammen präsentiert, und es ist Aufgabe der Probanden zu beurteilen, ob der gehörte Sprecherwechsel vom ersten Sprecher erwartet war. Drei Antwortkategorien stehen zur Auswahl:

- Sprecher erwartet den Wechsel,
- Sprecher erwartet weiterzusprechen,
- Unklarheit über die Erwartung des Sprechers.

Auch bei dieser Frage orientieren sich die Versuchspersonen vor allem an der syntaktischen Struktur. Bei nicht abgeschlossener Syntax werden nur 15% der Turnübergänge als erwartet eingestuft, während es bei den syntaktisch kompletten Teiläußerungen 68% sind. Unter den Turnübergängen mit syntaktischem Abschluss sticht wieder der H\* % Verlauf heraus, der nur mit 4% als Ende eines erwarteten Übergangs und mit 74% als Ende eines nicht erwarteten Übergangs beurteilt wird. H\* % zeigt

also auch hier eine relativ klare Funktion als Turnhaltesignal. Ein überraschendes Ergebnis ist, dass bei den Übergängen mit syntaktischem Abschluss bei allen anderen Mustern als H\* % mindestens 70% der Urteile für den erwarteten Übergang abgegeben werden. Das bedeutet, dass jede dieser Konturen als adäquate turnfinale Kontur für erwartete Turnübergänge angesehen wird. Diese Ergebnisse zeigen deutlich die Abhängigkeit der Urteile von der syntaktischen Struktur am Äußerungsende und die weitgehende Uneindeutigkeit der funktionalen Einordnung der Intonationsmuster.

Die Experimente haben allerdings einige Nachteile. Zum einen kann aufgrund der Stimulusgenerierung der semantisch-pragmatische Kontext der getesteten Äußerungen nicht konstant gehalten werden, obwohl er möglicherweise in der Beurteilung durch die Versuchspersonen eine Rolle spielt. Zweitens fließen in das Urteilsverhalten der Versuchspersonen unter Umständen auch noch andere phonetische Parameter als die Intonationskontur ein, ohne dass diese Parameter kontrolliert werden können (z.B. Phonationstyp, segmentelle Dauerstruktur und Intensitätsverlauf).

### **5.2.2 Erwartung von Turnübergaben an prosodischen Grenzen (Experiment 2)**

Das hier durchgeführte Perzeptionsexperiment ist eine Modifikation des experimentellen Vorgehens von Wichmann und Caspers (2001). Verschieden ausgeprägte prosodische Grenzen werden in Bezug auf ihre interaktive Bedeutung im Dialog experimentell überprüft. Die Versuchspersonen sollen für jeden Stimulus ein Urteil darüber abgeben, ob der gehörte Satz in einer Gesprächssituation eher das Ende eines Redebeitrags darstellt oder ob der Sprecher nach dem Satz noch weitersprechen wird. Es geht also darum, die Bereitschaft des Sprechers hinsichtlich eines Sprecherwechsels einzuschätzen. Das Experiment soll prüfen, inwieweit die phonetische Struktur von Hörern einbezogen wird, wenn sie Erwartungen darüber bilden, ob ein Sprecherwechsel intendiert ist. Besonders der finale melodische Verlauf scheint eine wichtige Rolle bei der Signalisierung von Abschluss vs. Weiterweisung zu spielen, denn wie bereits in der Korpusanalyse gezeigt wurde, unterscheiden sich turninterne und turnfinale Phrasengrenzen in der Verteilung der melodischen Muster deutlich (siehe 4.1.3 und 5.1.2). Unterschiede zeigen sich auch im verstärkten Auftreten von Laryngalisierung am Ende von Äußerungen. Diese Parameter und finale Länge werden experimentell hinsichtlich ihrer interaktionalen Funktion überprüft.

Um zu systematischen Untersuchungsergebnissen zu kommen, wird nicht wie bei Wichmann und Caspers (2001) mit Fragmenten aus aufgenommenen Dialogen gearbeitet, sondern ein konstanter Trägersatz systematisch und hypothesengesteuert in seinen phonetischen Eigenschaften variiert. Hierbei werden verschiedene phonetische Kontinua in mehrere Stufen unterteilt und über Resynthesen in die Stimuli eingearbeitet. So entsteht eine Vielzahl von Stimuli, die sowohl in der Kombination der Merkmale als auch in der Stärke der Einzelmerkmale variieren und so das Spektrum der im Korpus beobachteten phonetischen Variation repräsentieren. Die beiden wichtigsten Kontinua in diesem Experiment sind segmentelle Längung und der melodische Verlauf nach der letzten Akzentposition bis zum Phrasenende. Weiterhin werden phonatorische Eigenschaften am Ende der Stimuli gezielt verändert. Stimuli ohne syntaktischen Abschluss werden nicht in das Experiment einbezogen, da in dieser Hinsicht die Ergebnisse von Wichmann und Caspers eindeutig ausfallen und auch für das Deutsche keine Abweichung hiervon zu erwarten sind.

Die Versuchspersonen bekommen die phonetischen Varianten vorgespielt und bewerten auf 7-stufigen bipolaren Skalen die Wahrscheinlichkeit eines nachfolgenden Sprecherwechsels. Die Skalen sind Messskalen, die den Zusammenhang zwischen phonetischen Merkmalsbündelungen und der dadurch ausgelösten Erwartung von Hörern hinsichtlich der Äußerungsfortsetzung experimentell messbar machen. Die Skala ist folgendermaßen aufgebaut:

**Wird der Sprecher nach dem Satz noch weitersprechen?**

Ja							Nein
<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	

Durch diese Vorgehensweise werden im Vergleich zu den Experimenten von Wichmann und Caspers (2001) drei Verbesserungen erreicht:

- Die lexikalische Struktur der verwendeten Stimuli bleibt konstant. So kann ein Einfluss dieser Informationsebene auf die Urteile der Versuchspersonen ausgeschlossen werden.
- Die phonetische Struktur der Stimuli ist vollständig kontrollierbar und weist systematische Variationen verschiedener Parameter an der finalen prosodischen Grenze auf.

- Graduelle Unterschiede in der funktionalen Einordnung der Stimuli werden durch die Verwendung von Messskalen erfassbar.

Das letztendliche Ziel des Experiments ist die funktionale Zuordnung phonetisch unterschiedlicher prosodischer Grenzen, die in der Kombination mit der Interpretation von Korpusbeispielen folgende (oder ähnliche) Kategorien herausarbeiten soll:

- gezielte Turnübergabe,
- akzeptierte Turnübergabe,
- erzwungene Turnübergabe.

Natürlich bestimmt in einem Gespräch, neben der phonetischen Gestaltung des Gesprochenen, auch dessen syntaktische und lexikalische Struktur sowie der situative Kontext die Interpretation durch den Hörer. Es ist davon auszugehen, dass in einem natürlichen Gespräch alle verfügbaren kommunikativen Ressourcen, also der gesamte situative und formale Kontext, vom Hörer in die Einschätzung des weiteren Äußerungsverlauf einbezogen werden. Diese Variablen in ein Experiment einzubeziehen ist aber kaum zu leisten und zum derzeitigen Forschungsstand nicht sinnvoll. In den Experimenten soll deshalb speziell der Beitrag der phonetischen Ausprägung auf die Interpretation durch den Hörer untersucht werden.

### 5.2.2.1 Hypothesen

Folgende Hypothesen werden überprüft:

#### **Hypothese 1**

Melodische Muster an prosodischen Grenzen werden als Indikatoren für den vom Sprecher geplanten weiteren Verlauf des Dialogs interpretiert.

#### **Hypothese 2**

Stark fallende tonale Bewegungen am Phrasenende werden als Indikatoren eines nachfolgenden Sprecherwechsels interpretiert.

### **Hypothese 3**

Eben verlaufende und leicht bis mittelhoch steigende Konturen werden als Indikatoren einer geplanten Äußerungsfortsetzung interpretiert.

### **Hypothese 4**

Stark steigende Muster signalisieren eine Frage an den Gesprächspartner und werden somit als Hinweis auf einen nachfolgenden Sprecherwechsel interpretiert.

### **Hypothese 5**

Starke finale Längungen, die auf Planungsschwierigkeiten bzw. eine unflüssige Sprechweise hindeuten, senken die Erwartung eines Sprecherwechsels im Vergleich zu schwachen Längungen.

### **Hypothese 6**

Endet der Stimulus mit einer stark fallenden melodischen Bewegung und Laryngalisierung, verstärkt sich die Erwartung eines nachfolgenden Sprecherwechsels gegenüber dem Stimulus mit stark fallender melodischer Bewegung ohne Knarrstimme.

### **Hypothese 7**

Endet die Variante mit einem Glottalverschluss, führt dies eher zur Erwartung einer Äußerungsfortsetzung als bei vergleichbaren Varianten ohne finalen Glottalverschluss.

## **5.2.2.2 Stimulusmaterial**

### **5.2.2.2.1 Generierung des Ausgangsstimulus**

Als Trägersatz für die Resynthesen wird folgender Satz vom Autor gesprochen, unter Studiobedingungen aufgenommen und mit 44,1 kHz digitalisiert: *Das ist ja nun nicht so angenehm.* Dieser Trägersatz würde in einem realen Dialog eine Vielzahl von Fortsetzungsmöglichkeiten bieten. Er ist syntaktisch komplett und durch die semantisch-pragmatische Struktur weder auf eine Turnfortsetzung noch auf einen Turnabschluss festgelegt. Ein anderer Satz könnte die Urteile stark in die eine oder andere Richtung beeinflussen, wenn er z.B. auf der inhaltlichen oder syntaktischen Ebene bestimmte Fortführungen suggeriert. In so einem Fall würde die Interpretation wesentlich stärker von nicht-phonetischen Faktoren bestimmt werden. Trotzdem kann davon ausgegangen

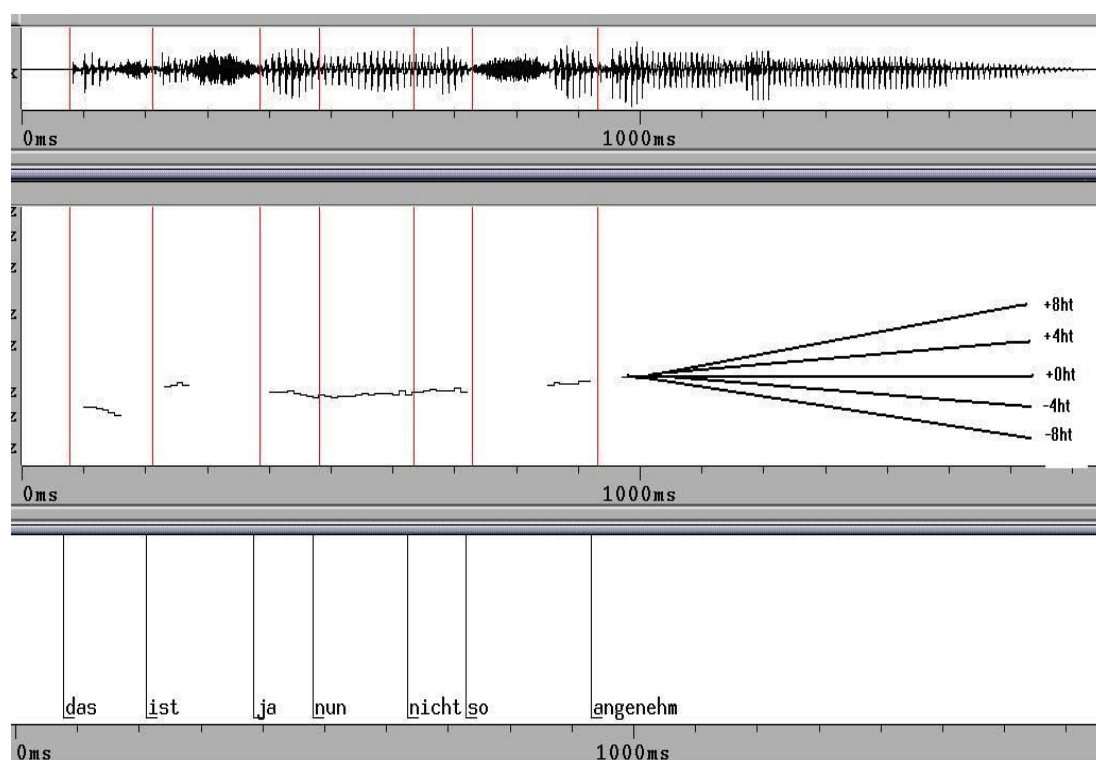
werden, dass das phonetische Signalisierungssystem bei einem stärker funktional gebundenen Trägersatz in ähnlicher Weise funktioniert wie bei dem ausgewählten, relativ neutralen Stimulus, nur dass Hörer sich bei Vorhandensein eindeutiger syntaktischer oder inhaltlicher Hinweise auf den Äußerungsfortgang viel stärker auf die Informationen stützen würden, die auf der syntaktischen und inhaltlichen Ebene zugänglich sind. D.h., die funktionale Last bei der Interpretation der interaktionalen Intention des Sprechers würde sich in so einem Fall eher auf syntaktische und/oder wortsemantische Parameter als auf die phonetische Struktur konzentrieren. Der Trägersatz zeigt in der Ausgangsform für die Resynthesen folgende phonetische Charakteristika:

- Tonhöhengipfel auf dem satzakzentuierten *angenehm* bei 111 Hz,
- Dauer des finalen Reims (*angenehm*): 250 ms (wurde mit Faktor 0,7 in seiner Dauer verkürzt, um die in der natürlichen Produktion enthaltene finale Längung zu eliminieren),
- Endpunkt der finalen Kontur bei 70 Hz,
- Durchgehend modale Stimmqualität.

#### 5.2.2.2.2 Konturverlauf am Phrasenende

Fünf finale Konturverläufe werden in äquidistanten Schritten von vier Halbtönen resynthetisiert. Der Bezugspunkt für die variierte tonale Differenz ist bezogen auf den tonalen Unterschied zwischen dem 111 Hz Gipfel auf *angenehm* und den F0-Endpunkten am Phrasenende. Es ergeben sich folgende Verläufe (Abbildung 9):

- stark fallend                    -8 Ht (von 111 auf 70 Hz fallend),
- leicht fallend                    -4 Ht (von 111 auf 88 Hz fallend ),
- eben                                -0 Ht (gleichbleibend bei 111 Hz),
- leicht steigend                    +4 Ht (von 111 auf 140 Hz steigend),
- stark steigend                    +8 Ht (von 111 auf 176 Hz steigend).



**Abbildung 9:** Resynthetisierte melodische Muster.

Die gewählten Konturverläufe orientieren sich an den Kategorien des Kieler Intonationsmodells. Dort werden insgesamt 12 phrasenfinale Melodiemuster differenziert. Eine experimentelle Überprüfung aller 12 melodischen Muster kann hier jedoch nicht geleistet werden. Deshalb wird nur eine Auswahl der in KIM postulierten Muster in die Experimente einbezogen. Die Auswahl beschränkt sich auf 5 Konturen, die einem linearen Verlauf folgen (ausgehend von der letzten satzakzentuierten Silbe der Phrase). Zwei der Konturen sind fallend, zwei steigend und eine verläuft eben auf dem Niveau der vorangehenden satzakzentuierten Silbe. Diese fünf gewählten Muster sind in den etikettierten Produktionsdaten sehr zahlreich vertreten und machen zusammen etwa 90% aller melodischen Muster am Phrasenende aus. Die restlichen Konturen sind vor allem kombinierte Konturen (fallend-steigend) (Kohler 1991a, Peters 1999) und pseudoternale Konturen (Peters 2000).

### 5.2.2.2.3 Längung des finalen Reims

Ausgehend von einem Stimulus ohne finale Längung werden folgende Längungsfaktoren auf den finalen Reim von *angenehm* angewendet.

- Faktor 1,0 (Reimdauer: 250 ms),
- Faktor 1,4 (Reimdauer: 350 ms),
- Faktor 1,8 (Reimdauer: 450 ms),
- Faktor 2,2 (Reimdauer: 550 ms).

Wie im Perzeptionsexperiment 3a und 3b (siehe 6.2) gezeigt wird, werden Reime mit einer Längung ab Faktor 1,8 als klare Indikatoren einer unflüssigen Sprechweise wahrgenommen. Die Längungsstufen 1,8 und 2,2 werden in das Experiment aufgenommen, um eventuelle Wechselwirkungen zwischen der Wahrnehmung einer unflüssigen Sprechweise und der interaktionalen Funktion festzustellen. Ebenfalls kann unter 6.2 belegt werden, dass sich in natürlicher Sprache Längungserscheinungen auch über einen größeren Bereich als den finalen Reim ausdehnen können. Dies steht meist im Zusammenhang mit Planungsschwierigkeiten. Um zu prüfen, ob solche ausgedehnten segmentellen Längungen einen Einfluss auf die Vorhersage von Sprecherwechseln haben, wird eine Serie von Stimuli generiert, in der sich die Längung über das gesamte letzte Wort der Äußerung (*angenehm*) ausdehnt.

Folgende Längungsfaktoren werden auf das letzte Wort der Äußerung angewendet:

- Faktor 1,0 (Wortdauer: 515 ms),
- Faktor 1,4 (Wortdauer: 720 ms),
- Faktor 1,8 (Wortdauer: 930 ms),
- Faktor 2,2 (Wortdauer: 1130 ms).

Auch diese verschieden gelängten Varianten werden mit allen fünf Melodiemustern gepaart (siehe 5.2.2.2.2).

#### 5.2.2.2.4 Phonatorische Veränderungen

Wie die Korpusuntersuchungen zeigen, spielen auch phonatorische Veränderungen an prosodischen Grenzen eine Rolle. Sowohl an syntaktischen Abbrüchen als auch an flüssigen turninternen und turnfinalen Grenzen tritt Knarrstimme auf. Allerdings ist die durchschnittliche Dauer des geknarrten Signalabschnitts an syntaktischen Abbrüchen kürzer als bei flüssigen Grenzen (83 ms vs. 121 ms). Auch endet die Knarrstimme an syntaktischen Abbrüchen in einigen Fällen mit einem Glottalverschluss, was an flüssigen Grenzen nicht zu beobachten ist. Weiterhin zeigen die Analysen, dass die Knarrstimme an syntaktischen Abbrüchen fast immer aus einem ebenen oder leicht fallenden F<sub>0</sub>-Verlauf heraus einsetzt, während phrasenfinale Knarrstimme in flüssiger Sprache immer mit stark fallender Intonation gekoppelt ist. Um die funktionale Zuordnung unterschiedlicher Arten von Knarrstimme zu testen, werden in das Stimulusmaterial auch Varianten mit und ohne finalen Glottalverschluss aufgenommen.

Das technische Vorgehen hierbei ist für beide Arten der Knarrstimme gleich. Bei der Aufnahme des Rohmaterials für die Stimulusgenerierung werden vom Sprecher des Testsatzes zusätzlich mehrere Versionen des Satzes gesprochen, die am Ende der Äußerung entweder eine abrupt einsetzende Glottalisierung mit anschließendem Glottalverschluss haben oder einen stark fallenden Intonationsverlauf, der mit Laryngalisierung endet. Die Stimuli entstehen, indem das Ende der Äußerungen mit Knarrstimme an den Ausgangsstimulus angehängt wird. Die Dauerverhältnisse werden anschließend exakt an die übrigen Stimuli angepasst.

Zur Generierung der Stimuli mit Glottalisierung wird an die Stimuli der Längungsstufe 1,4 für alle Konturen 100 ms Glottalisierung aus einer natürlichen Produktion an den Basisstimulus angehängt, nachdem zuerst 100 ms Modalstimme weggeschnitten wurden. Das entspricht dem im Korpus gemessenen Mittelwert von 95 ms Knarrstimme bei Glottalisierung an syntaktischen Abbrüchen.

Laryngalisierung wird in das Material aufgenommen, indem vom Stimulus mit Längungsstufe 1,4 und stark fallender Kontur die Sequenz *angenehm* abgeschnitten und dann ein anderes *angenehm* angehängt wird, das eine stark fallende Kontur mit finaler Laryngalisierung enthält. Die Dauerverhältnisse werden exakt an die Stimuli der Längungsstufe 1,4 angepasst. Die Laryngalisierung ist 230 ms lang. Das ist länger als die durchschnittliche gemessene Dauer im Korpus (152 ms). Die Laryngalisierung wird nur an die stark fallende Kontur angehängt, da auch im

Korpus eine solche phonatorische Veränderung nur in diesem Kontext auftritt und sie mit allen anderen Konturen unnatürlich klingt. Der laryngalisierte Abschnitt ist von leichter Verhauchung begleitet.

Der Grund, weshalb die phonatorischen Veränderungen nicht systematisch mit allen Längungsstufen gepaart werden, ist, dass dadurch die Anzahl der Stimuli zu groß werden würde und dies die Versuchspersonen überlasten könnte. Der Längungsfaktor 1,4 entspricht einer unauffälligen aber klar vorhandenen Längung des phrasenfinalen Reims. Die so erzeugten Varianten können dann in den Ergebnissen auch nur innerhalb der Serie mit entsprechender Längungsstufe verglichen werden.

#### **5.2.2.2.5 Überblick über das gesamte Stimulusmaterial**

Durch die Manipulationen ergeben sich folgende Stimuli:

- Basisstimuli: vier Längungsfaktoren x fünf Konturen = zwanzig Stimuli,
- Ausgedehnte Längung (auf das ganze letzte Wort): drei Längungsfaktoren x fünf Konturen = fünfzehn Stimuli,
- Glottalisierung: ein Längungsfaktor x fünf Konturen = fünf Stimuli,
- Laryngalisierung: ein Längungsfaktor x eine Kontur = ein Stimulus,
- Gesamt:  $20 + 15 + 5 + 1 = 41$  Stimuli.

#### **5.2.2.3 Durchführung von Experiment 2**

Die Stimuli werden mit dreifacher Wiederholung randomisiert und zu einer Audiodatei zusammengeschnitten. Zwischen den Stimuli ist immer eine Pause von vier Sekunden. Nach jeweils 10 Stimuli ist eine längere Pause und ein Piepton. In den Pausen zwischen den Stimuli geben die Versuchspersonen ihre Urteile ab. Zu Beginn der Sitzung werden vorher aufgenommene Instruktionen abgespielt, die den Ablauf und die Aufgabenstellung erläutern (siehe Anhang). An dem Experiment nahmen 14 Personen teil.

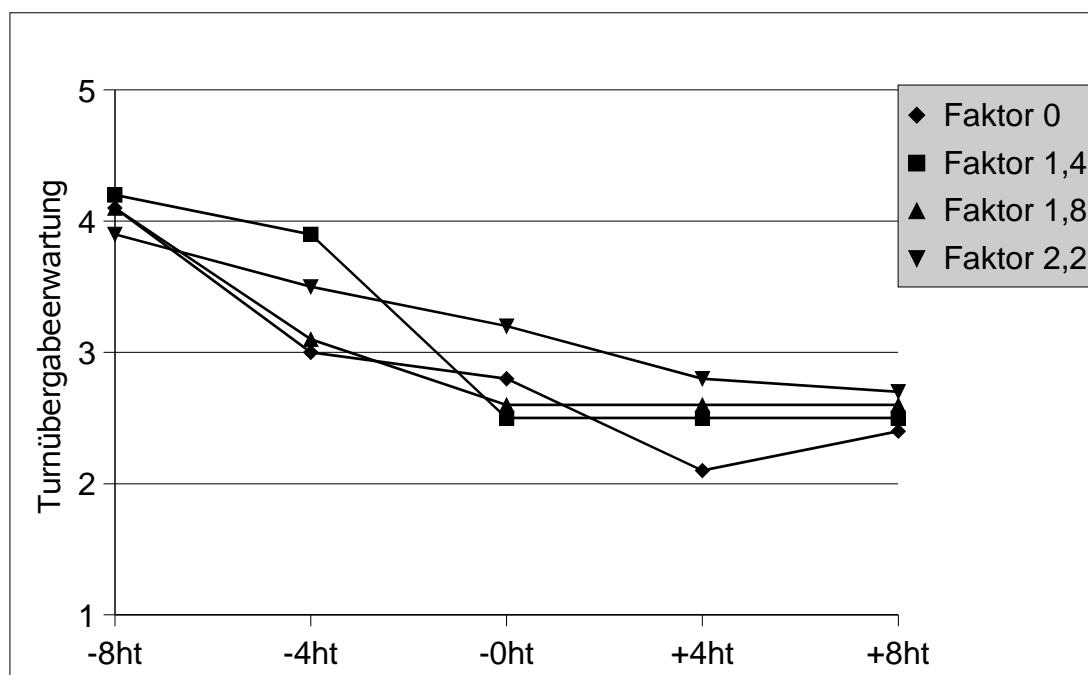
### 5.2.2.4 Ergebnisse aus Experiment 2

Tabelle 24 fasst die Ergebnisse der Varianzanalyse zusammen. In einigen Fällen sind in Abhängigkeit von der jeweiligen Fragestellung weitere Einzelkontraste zwischen bestimmten Stimuli berechnet worden. Diese Ergebnisse sind in den nachfolgenden Text integriert.

**Tabelle 24:** Ergebnisse der Varianzanalyse von Experiment 2.

<i>Faktor / Interaktion</i>	<i>df</i>	<i>F</i>	<i>p</i>
Längung des finalen Reims	1,17	0,34	0,604
Längung des finalen Wortes	1,32	0,6	0,495
Finale Kontur (über gelängtem Reim)	1,42	4,27	0,041
Finale Kontur (über gelängtem Wort)	1,66	4,81	0,024
Phonation	1	0,021	0,887
Längung (Reim) * Kontur (Reim)	6,18	1,65	0,141
Längung (Wort) * Kontur (Wort)	4,29	1,93	0,114
Phonation * Kontur (Reim)	2,95	1,4	0,258

Im Folgenden werden die oben formulierten Hypothesen anhand der experimentellen Ergebnisse abgearbeitet. Die y-Achse der folgenden Abbildungen ist entsprechend der von 0 bis 6 skalierten Urteile aufgebaut. In den Abbildungen werden die Werte 0 und 6 weggelassen, da die dargestellten Mittelwerte sich nie in diesen Bereichen bewegen. Ein niedriger Wert auf der y-Achse bedeutet, dass die Versuchspersonen eher für eine Fortsetzung des Turns gestimmt haben; ein hoher Wert bedeutet, dass der Eindruck eines abgeschlossenen Turns entstanden ist. Anhand der Abbildung 10 können die ersten fünf Hypothesen überprüft werden:



**Abbildung 10:** Turnübergabeerwartung für die Stimulusserie mit unterschiedlich stark gelangten finalen Reimen und fünf finalen melodischen Mustern.

### Hypothese 1

Melodische Muster an prosodischen Grenzen werden als Indikatoren für den vom Sprecher geplanten weiteren Verlauf des Dialogs interpretiert.

Die Abbildung zeigt einen deutlichen Einfluss des melodischen Verlaufs auf die Urteile der Versuchspersonen, der prüfstatistisch signifikant ist (siehe Tabelle 24). Der finale Konturverlauf ist also relevant für die Erwartungen über den weiteren Dialogverlauf. Die Varianzanalyse zeigt im Rahmen der Berechnung von Einzelkontrasten, dass hierbei die Unterschiede zwischen der stark fallenden Kontur (-8ht) und allen anderen Konturen signifikant werden (schon der Unterscheid zwischen der stark fallenden und der leicht fallenden Kontur führt zu einem F-Wert von 5,4 und einem P-Wert von 0,036 alle weiteren Vergleiche haben folglich höhere F-Werte). Ebenfalls signifikant wird der Unterschied zwischen der leicht fallenden Kontur und allen weiteren Mustern (leicht fallend vs. eben:  $F=8,9$ ,  $p=0,011$ ). Die ebenen und steigenden Muster werden untereinander nicht signifikant unterschieden.

### Hypothese 2

Stark fallende tonale Bewegungen (-8ht) am Phrasenende werden als Indikatoren eines nachfolgenden Sprecherwechsels interpretiert.

Auch diese Hypothese wird durch die Ergebnisse belegt. Es zeigt sich, dass die stark fallende Kontur vor allen anderen melodischen Mustern den höchsten Wert erreicht. Allerdings ist zu beachten, dass kein Mittelwert deutlich über den Wert 4 hinausgeht. D.h., dass die Urteile keine eindeutige Einschätzung, sondern nur eine Tendenz zeigen.

### **Hypothese 3**

Eben verlaufende (-0ht) und leicht bis mittelhoch steigende Konturen (+4ht) werden als Indikatoren einer geplanten Äußerungsfortsetzung interpretiert.

Die Urteile zeigen, dass die ebenen und die leicht bis mittelhoch steigenden Muster eher als Signal für eine geplante Äußerungsfortsetzung angesehen werden als das stark fallende Muster. Allerdings bewegen sich die Mittelwerte der Urteile für die ebene und die leicht bis mittelhoch steigende Kontur um den Wert 3, der genau auf der Mitte der Skala liegt. Nur wenige Mittelwerte liegen deutlich darunter. Dieses Ergebnis drückt also eher eine Unentschiedenheit der Versuchspersonen hinsichtlich der funktionalen Zuordnung aus als ein klares Votum für eine erwartete Turnfortsetzung. Auch im Vergleich mit den Ergebnissen von Wichmann und Caspers (2001), die für den hoch-eben endenden Verlauf im Englischen eine Wahrscheinlichkeit von 89% für eine unmittelbare Turnfortsetzung erhalten haben, fallen die Ergebnisse uneindeutig aus. Allerdings ist zu beachten, dass Wichmann und Caspers mit einer 'forced choice'-Befragung gearbeitet haben.

### **Hypothese 4**

Stark steigende Muster (+8ht) signalisieren eine Frage an den Gesprächspartner und werden somit als Hinweis auf einen nachfolgenden Sprecherwechsel interpretiert.

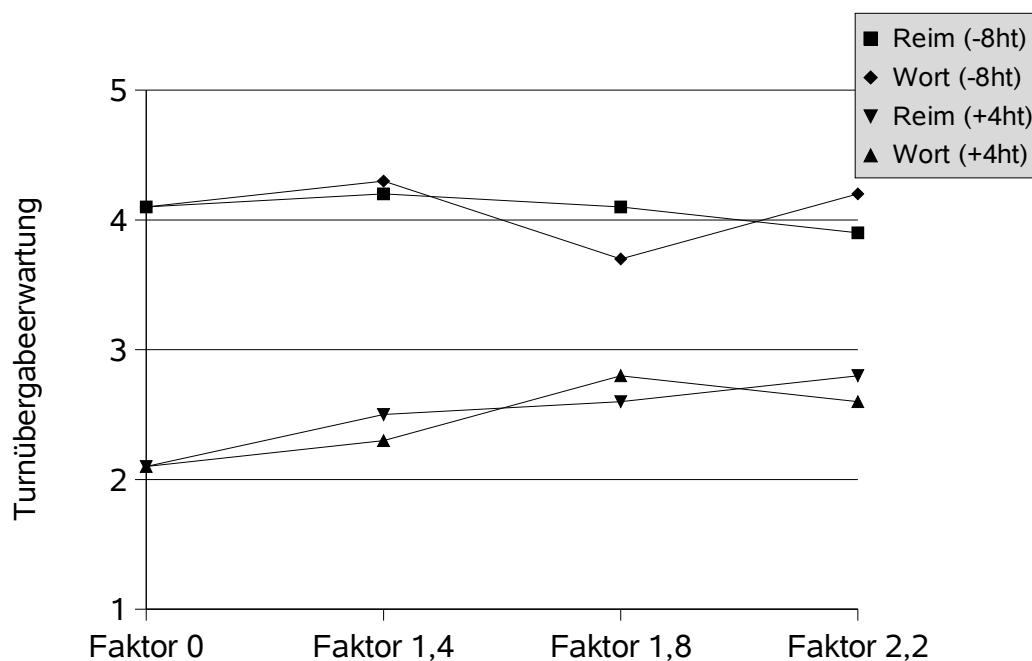
Das stark steigende Muster erreicht ein ähnliches Antwortprofil wie das leicht bis mittelhoch steigende Muster und wird somit nicht als Indikator für eine Frage an den Gesprächspartner gedeutet, da dies mit der Erwartung eines nachfolgenden Sprecherwechsels einhergehen müsste. Hierfür gibt es zwei mögliche Erklärungen: Der Trägersatz (*Das ist ja nun nicht so angenehm.*) hat eine deklarative Syntax. Deklarative Sätze können zwar durch steigende Intonationsmuster als konversationelle Fragen realisiert werden, allerdings enthält der Satz die Partikeln *ja* und *nun* in direkter Folge. Diese Konstruktion in Verbindung mit einer Frage an den

Gesprächspartner wäre sehr ungewöhnlich und hat wahrscheinlich die Wahrnehmung als Frage verhindert. Eine zweite Erklärung, die auch ergänzend zur ersten Erklärung gelten kann, ist, dass für die intonatorische Markierung eines deklarativen Satzes nicht nur der finale melodische Verlauf eine Rolle spielt, obwohl dies das auffälligste Merkmal ist, sondern auch andere Parameter (z.B. eine global angehobene Tonhöhe).

### **Hypothese 5**

Starke finale Längungen, die auf Planungsschwierigkeiten bzw. eine unflüssige Sprechweise hindeuten, senken die Erwartung eines Sprecherwechsels im Vergleich zu schwachen Längungen.

Insgesamt zeigen die Ergebnisse in Abbildung 10 keine klaren Auswirkung der verschiedenen Längungsstufen des finalen Reims auf die Einschätzungen über den Sprecherwechsel. Auch die Prüfstatistik ergibt keinen signifikanten Effekt der Längung. Im Perzeptionsexperiment 3a (siehe 6.2.1) kann hingegen eindeutig gezeigt werden, dass die Äußerungen mit final stark gelängten Reimen und Wörtern (Faktor 1,8 und 2,2) als sehr unflüssig wahrgenommen werden. Abbildung 11 zeigt ergänzend einen Vergleich zwischen dem stark fallenden und dem leicht bis mittelhoch steigenden Muster unter den verschiedenen Längungsbedingungen. Die beiden melodischen Muster wurden gewählt, weil sie hinsichtlich der interaktionalen Bedeutung die deutlichsten funktionalen Tendenzen zeigen. Die Abbildung illustriert, dass sich für die leicht bis mittelhoch steigende Kontur bei zunehmender Längung die Urteile zur Mitte der Skala verschieben, also zunehmende Unsicherheit über den weiteren Gesprächsverlauf ausgedrückt wird. Hierbei spielt es keine Rolle, ob nur der finale Reim oder das gesamte letzte Wort gelängt ist. Für die stark fallende Kontur gibt es keinen systematischen Einfluss der Dauerrelationen. Hypothese 5 wird folglich verworfen.



**Abbildung 11:** Turnübergabeerwartung für die stark fallende Kontur (-8ht) und die leicht steigende Kontur (+4ht) unter den vier Längungsbedingungen. Getrennt für gelängte finale Reime und gelängte finale Wörter (drei Silben).

### Hypothese 6

Endet der Stimulus mit einer stark fallenden melodischen Bewegung und Laryngalisierung, verstärkt sich die Erwartung eines nachfolgenden Sprecherwechsels gegenüber dem Stimulus mit stark fallender melodischer Bewegung ohne Knarrstimme.

Die Beurteilung für den laryngalisierten Stimulus liegt im Mittel bei 4,2. Der in der Dauerstruktur gleiche Stimulus mit stark fallender Intonation (-8ht) ohne Laryngalisierung wird hinsichtlich der Turnübergabeerwartung mit einem Wert von 4,1 praktisch gleich eingeschätzt. Hypothese 6 wird verworfen.

### Hypothese 7

Endet die Variante mit einem Glottalverschluss, führt dies eher zur Erwartung einer Äußerungsfortsetzung als bei vergleichbaren Varianten ohne finalen Glottalverschluss.

Die Urteile zeigen bei Stimuli mit sonst gleichen Merkmalsbündeln keine signifikanten Unterschiede zwischen den Stimuli mit und ohne finalen Glottalverschluss. Hypothese 7 wird somit verworfen.

### 5.2.2.5 Zusammenfassung und Diskussion zu Experiment 2

Die Versuchspersonen bilden systematische Urteile und können offensichtlich mit der Aufgabenstellung umgehen. Allerdings zeigt die deutlich höhere Standardabweichung als in den Experimenten 1a, 1b und 1c, dass die verschiedenen Hörer die Stimuli weniger einheitlich beurteilen.

Es kann gezeigt werden, dass der melodische Verlauf am Äußerungsende einen Einfluss auf die Einschätzung des weiteren (hypothetischen) Dialogverlaufs hat, während die segmentellen Dauerrelationen hier fast keine Rolle spielen. Vor allem die stark fallende Kontur wird mit Turnabschlüssen assoziiert und die steigenden Konturen mit Turnfortsetzungen. Allerdings bewegen sich die Mittelwerte der Urteile kaum aus dem Bereich zwischen 2 und 4 heraus. Das bedeutet, dass im Mittel keine eindeutigen funktionalen Zuordnungen getroffen wurden, sondern sich lediglich Tendenzen zeigen. Dies hat mehrere Gründe:

- Das individuelle Urteilsverhalten der Versuchspersonen variiert stark. Während die durchschnittliche Standardabweichung in den Experimenten zur Kohäsion (Experimente 1a, 1b, 1c) bei etwa 1 liegt und für alle Stimuli etwa gleich ist, liegt sie im Experiment zur Turnübergabewahrscheinlichkeit durchschnittlich deutlich höher und variiert für die einzelnen Stimuli zwischen 0,8 und 2,3. Die Urteile für die einzelnen Wiederholungen der Stimuli innerhalb einzelner Versuchspersonen hingegen sind relativ konsistent.
- Die wiederholte und kontextfreie Präsentation der Stimuli ist eine unnatürliche Situation und das Urteilsverhalten kann nicht als direktes Abbild der in einer Dialogsituation ablaufenden Hypothesenbildung angesehen werden. Für die Versuchspersonen ist es möglicherweise schwierig, die Einschätzungen über einen Dialogverlauf, die im normalen Gespräch unbewusst ablaufen, bewusst und ohne kontextuelle Hinweise zu treffen.
- Dem phonetischen Signalisierungssystem liegt kein striktes 1:1-Verhältnis zwischen Form und Funktion zugrunde. Die Ergebnisse deuten eher auf eine unterstützende Funktion der phonetischen Komponente hin, die erst in Kombination mit syntaktischen und semantischen Strukturen zu eindeutigen interaktionalen Funktionalisierungen führt.

- Die phonetische Struktur von Äußerungen bietet noch eine Reihe von Variationsmöglichkeiten, die für das Dialogsteuerungssystem eine Rolle spielen können, aber nicht in das Stimulusmaterial integriert wurden. In natürlichen Dialogen ist z.B. in turnfinaler Position oft eine deutliche Intensitätsabnahme zu beobachten. Auch unterschiedliche Synchronisierungen von nuklearen Gipfel- und Talpositionen könnten in weitere Experimente einbezogen werden, um zu deutlicheren Urteilen zu gelangen.

Die Urteile zeigen, dass das phonetische Signalisierungssystem zwar funktionale Relevanz besitzt, jedoch kaum eindeutige Zuordnungen hinsichtlich des Gesprächsfortgangs zulässt. Ein klares Ergebnis ist, dass die Veränderungen in der Phonation keine Auswirkungen auf die Urteile haben. Dies ist überraschend, denn die beiden Typen von Knarrstimme zeigen in den Produktionsdaten Verteilungen, die auf eine funktionale Bindung hinweisen. Knarrstimme mit Glottalverschluss ist nur an syntaktischen Abbruchstellen zu beobachten, während weich auslaufende Laryngalisierung gehäuft an prosodischen Grenzen vor Sprecherwechseln vorkommt. Hier scheinen typische Merkmale der Produktion in der Perzeption nicht mit funktionalen Kategorien assoziiert zu werden. Unter diesem Gesichtspunkt muss die von Local und Kelly (1986) postulierte funktionale Differenzierung zwischen 'turn-holding-device' (Glottalverschluss) und 'turn-yielding-device' (Laryngalisierung) noch einmal kritisch geprüft werden. Aufgrund der experimentellen Ergebnisse liegt nahe, dass die perzeptive Relevanz dieser Parameter von den Autoren überschätzt wird. So ist auch deren generelle Ansicht, dass jedes phonetische Detail für den Kommunikationsprozess von Bedeutung ist, kritisch zu überdenken. Natürlich beziehen sich die Aussagen von Local und Kelly auf das Englische, da aber die von ihnen beschriebenen Phänomene völlig vergleichbar auch in den untersuchten deutschen Daten vorliegen, ist anzunehmen, dass auch die perzeptiven Ergebnisse auf das Englische übertragbar sind.

Ein Einwand, der seitens der Conversation Analysis (CA) gegen diese Kritik erhoben werden könnte, ist, dass es lediglich im durchgeführten Perzeptionsexperiment nicht möglich ist, die perzeptorische Relevanz phonatorischer Veränderungen zu belegen, obwohl diese in natürlichen Gesprächen durchaus vorhanden ist und nur durch die Methoden der sequentiellen Analyse belegt werden kann. Eine solche Argumentation würde allerdings die experimentelle Phonetik insgesamt in Frage stellen und m.E. eine einseitige und dem Forschungsgegenstand nicht angemessene

sene Fixierung auf interpretative Verfahren demonstrieren. Allerdings ist sicher richtig, dass ein Perzeptionsexperiment immer eine Abstraktion einer tatsächlichen Gesprächssituation ist und Versuchspersonen in diesem situativen Kontext u.U. anders auf sprachliches Material reagieren als in einem natürlichen Gespräch. Besonders da in dem Experiment ein bewusstes Urteil über Erwartungen gefordert wurde, die sonst unbewusst gebildet werden, ist nicht sicher, dass das Experiment die kommunikativen Funktionen phonatorischer Veränderungen adäquat erfassen kann. Dass für die melodischen Muster die erwarteten funktionalen Zuordnungen von den Probanden vorgenommen wurden, ist allerdings ein Indiz dafür, dass das Experiment geeignet ist, funktionale Zusammenhänge herzustellen.

### 5.3 Interpretation von Korpusbeispielen

Wie bereits bei den turninternen prosodischen Grenzen soll die Subklassifikation turnfinaler prosodischer Grenzen durch die Interpretation von Korpusbeispielen gestützt werden. Hierfür werden jeweils 40 Beispiele der drei unterschiedlich synchronisierten Turnübergänge aus dem LINDENSTRASSE-Korpus herausgegriffen (Übergänge mit Pause, Übergänge mit unmittelbarem Anschluss und Übergänge mit Überlagerung). An jeder der 120 finalen prosodischen Grenzen wird entschieden, ob am Turnübergang das Ende eines möglichen Satzes erreicht wird und ob ein möglicher inhaltlicher Abschluss vorliegt. Weiterhin wird entschieden, ob das Turnende mit einer Frage an den Gesprächspartner einhergeht.

Durch dieses Verfahren soll eine funktionale Subklassifikation in verschiedene Typen von Turnübergängen gestützt werden, denn es hat sich bereits durch die automatische Merkmalsextraktion gezeigt, dass eine Subklassifikation auf rein akustisch/phonetischer Basis wegen der starken Überschneidung der Merkmalsbündel in den drei unterschiedlich synchronisierten Turnübergängen nicht sinnvoll ist. Auch das durchgeführte Perzeptionsexperiment kann keine eindeutigen funktionalen Bindungen herausarbeiten. Das gewählte interpretative Verfahren bietet die Möglichkeit festzustellen, ob die drei Arten der Übergänge in ihrer semantisch-pragmatischen Struktur unterschiedlich sind. Diese Herangehensweise unterscheidet sich durch den statistischen Ansatz von den Einzelfallbeschreibungen der Conversation Analysis (dt.: 'Gesprächsanalyse', z.B. Selting 1995). Folgende Hypothesen werden geprüft:

**Hypothese 1**

Turnübergänge mit Überlagerung zeigen häufiger inhaltliche und/oder syntaktische Unabgeschlossenheit als unmittelbare Übergänge oder Übergänge mit Pausen.

**Hypothese 2**

Fragen an den Gesprächspartner sind bei überlagernden Turnübergängen seltener zu erwarten als in den anderen beiden Klassen.

Wenn die Interpretation diese Erwartungen bestätigt, kann das Ergebnis herangezogen werden, um Turnübergänge ohne Überlagerung als gezielt vom Sprecher eingeleitet zu klassifizieren und Übergänge mit Überlagerung als vom Hörer eingeleitete Turnwechsel.

**5.3.1 Ergebnisse des interpretativen Teils**

Die Interpretation der 120 Turnwechsel liefert folgende Ergebnisse:

**Tabelle 25a:** Inhaltliche Abgeschlossenheit.

	<i>Pause (n=40)</i>	<i>Unmittelbarer An- schluss (n=40)</i>	<i>Überlagerung (n=40)</i>
Inhaltlich abgeschlossen	36 (90%)	40 (100%)	39 (97,5%)

**Tabelle 25b:** Syntaktische Abgeschlossenheit.

	<i>Pause (n=40)</i>	<i>Unmittelbarer Anschluss (n=40)</i>	<i>Überlagerung (n=40)</i>
Syntaktisch abgeschlossen	39 (97,5%)	38 (95%)	37 (92,5%)

**Tabelle 25c:** Fragefunktion.

	<i>Pause (n=40)</i>	<i>Unmittelbarer Anschluss (n=40)</i>	<i>Überlagerung (n=40)</i>
Frage an den Gesprächspartner	11 (27,5%)	8 (20%)	9 (22,5%)

Die Zahlen der Tabellen 25a, b und c zeigen keine verwertbaren Unterschiede zwischen den drei Typen von Turnübergängen. An fast allen Turnwechseln wird vom übergebenden Sprecher sowohl die syntaktische als auch die inhaltliche Struktur komplettiert. Das durchgeführte Interpretationsverfahren liefert also keine Hinweise auf funktionale Unterschiede zwischen den drei unterschiedlich synchronisierten Übergängen. Beide gestellten Hypothesen werden verworfen.

Lediglich die finalen melodischen Muster zwischen den Übergängen mit und ohne Frage an den Partner zeigen deutliche Unterschiede in der Verteilung, wie Tabelle 26 zeigt. Unter den Turnwechseln mit Frage gibt es wesentlich mehr steigende Muster, und es ist auffällig, dass die melodische Kontur, die am häufigsten in Verbindung mit Fragen steht, die leicht bis mittelhoch steigende Kontur ist. Schwerpunkte bei den Wechseln ohne Frage liegen bei den stark fallenden und den leicht fallenden Verläufen. Es ist aber wichtig zu bemerken, dass auch an Turnabschlüssen, die mit einer Frage an den Dialogpartner verbunden sind, fast alle melodischen Muster vorkommen, obwohl nur 28 Fragen in die Untersuchung eingehen. Auch die Fragefunktion scheint also nicht an bestimmte melodische Muster gebunden zu sein.

**Tabelle 26:** Melodische Verläufe am Turnende (LINDENSTRASSE).

Melodischer Verlauf	Keine Frage (n=92)	Frage (n=28)
Stark fallend	35 (38,0%)	2 (7,1%)
Stark fallend + sehr leicht steigend	6 (6,5%)	0 (0%)
Fallend-steigend	5 (5,4%)	4 (14,3%)
Stark Steigend	0 (0%)	6 (21,4%)
Eben	13 (14,1%)	1 (3,6%)
Leicht fallend	23 (25,0%)	2 (7,1%)
Leicht bis mittelhoch steigend	7 (7,6%)	11 (39,3%)
Eben + sehr leicht steigend	2 (2,2%)	1 (3,6%)
Leicht fallend + sehr leicht steigend	1 (1,1%)	1 (3,6%)

### 5.3.2 Diskussion der interpretativen Ergebnisse

Es gibt mehrere Erklärungen, warum die Verteilungen in den drei getrennt analysierten Klassen so ähnlich sind und es kaum zu abgebrochenen inhaltlichen und syntaktischen Strukturen kommt. Als erstes ist die generell sehr einvernehmliche und kooperative Einstellung der Dialogpartner zu nennen. Die Gesprächspartner kennen sich gut und achten darauf, dem Gegenüber nicht zu forsch ins Wort zu fallen. Beide sind an der Lösung einer gemeinsamen Aufgabe interessiert und die Gespräche geben keinen Anlass zu Uneinigkeiten, die ein forsches Unterbrechen des Partners hervorrufen könnten. Möglicherweise wird das kooperative Verhalten der Versuchspersonen zusätzlich durch die Aufnahmesituation verstärkt, da häufiges Unterbrechen des Partners als unhöflich angesehen werden kann und dies in dieser besonderen Gesprächssituation speziell vermieden wird.

Beim Abhören der Dialoge entsteht dennoch an vielen Turnübergängen der Eindruck, dass die Übergabe vom derzeitigen Sprecher nicht geplant war und der Hörer trotzdem an einer geeigneten Stelle zu reden beginnt. Die so unterbrochenen Sprecher machen aber kaum Anstrengungen den Turn zu halten, z.B. indem sie trotz der Unterbrechung weitersprechen. Sie bringen bei überlagernden Einsätzen des Hörers meist nur noch die begonnene syntaktische Konstruktion zu Ende und begeben sich anschließend in die Hörerrolle. Dieses Zuendebringen ist bei Unterbrechungen möglicherweise die übliche Verhaltensweise, im Gegensatz zu einem sofortigem Abbruch innerhalb der begonnenen syntaktischen Struktur, zumal die unterbrechenden Hörer bei überlagerndem Redebeginn meist einen Punkt in der Äußerung wählen, an dem die inhaltliche und syntaktische Struktur kurz vor einem möglichen Abschluss steht. Auch die gemessene durchschnittliche Dauer der Überlagerungen liegt mit 480 ms niedrig und stützt die obige Argumentation.

Auf der anderen Seite müssen Turnwechsel mit Pausen nicht unbedingt gezielt eingeleitete Übergaben sein. Die Tabelle 25a zeigt, dass es hier in vier Fällen zu einem Sprecherwechsel kommt, ohne dass ein inhaltlicher Abschluss erreicht wird. Die Erklärung hierfür ist wahrscheinlich, dass die Sprecher in diesen Fällen eine Pause machen, um den weiteren Äußerungsverlauf zu planen, und dass der Hörer in dieser Pause das Wort ergreift und den Turn übernimmt, obwohl der Sprecher dies nicht geplant hat.

#### **5.4 Vergleich zwischen turnfinalen und turninternen prosodischen Grenzen**

Die folgenden Tabellen zeigen die Verteilungen melodischer Muster, finaler Längung und phonatorischer Veränderungen im Vergleich zwischen turnfinalen und turninternen prosodischen Grenzen. Diese drei Parameter sind durchgehend vergleichbar (im Gegensatz zu F0-Reset) und der Vergleich soll zeigen, ob sich die funktionalen Unterschiede in prosodischen Strukturen widerspiegeln.

**Tabelle 27:** Vergleich melodischer Verläufe an turnfinalen und turninternen prosodischen Grenzen.

Melodischer Verlauf	Turnfinal mit Pause (n=183)	Turnfinal mit unmittelbarem Anschluss (n=66)	Turnfinal mit Überlagerung (n=128)	Turnintern (n=3399)
<b>Starke melodische Bewegung</b>				
Stark fallend	53 (29,0%)	23 (34,9%)	35 (27,3%)	915 (26,9%)
Stark fallend + sehr leicht steigend	12 (6,6%)	4 (6,1%)	5 (3,9%)	202 (5,9%)
Fallend-steigend	12 (6,6%)	6 (9,1%)	5 (3,9%)	121 (3,6%)
Stark Steigend	13 (7,1%)	4 (6,1%)	3 (2,3%)	65 (1,9%)
Gesamt starke melodische Bewegung	90 (49,2%)	37 (56,1%)	48 (37,5%)	1303 (38,3%)
<b>Schwache melodische Bewegung</b>				
Eben	25 (13,7%)	7 (10,6%)	19 (14,8%)	562 (16,5%)
Leicht fallend	20 (10,9%)	9 (13,6%)	34 (26,6%)	962 (28,3%)
Leicht bis mittelhoch steigend	37 (20,2%)	9 (13,6%)	11 (8,6%)	438 (12,9%)
Eben + sehr leicht steigend	3 (1,6%)	2 (3,0%)	5 (3,9%)	23 (0,7%)
Leicht fallend + sehr leicht steigend	8 (4,4%)	2 (3,0%)	11 (8,6%)	111 (3,3%)
Gesamt schwache mel. Bewegung	93 (50,8%)	29 (43,9%)	80 (62,5%)	2096 (61,7%)
<b>Gesamt</b>	<b>128 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>	<b>3399 (100%)</b>

**Tabelle 28:** Vergleich phonatorischer Veränderungen an turnfinalen und turninternen prosodischen Grenzen mit stark fallender Intonation.

	Turnfinal mit Pause (n=53)	Turnfinal mit unmittelbarem Anschluss(n=23)	Turnfinal mit Überlagerung (n=35)	Turnintern (n=238)
Modalstimme	30 (56,6%)	14 (60,9%)	27 (77,1%)	159 (66,8%)
Knarrstimme (unregelmäßig)	15 (28,3%)	7 (30,4%)	4 (11,4%)	26 (10,9%)
Knarrstimme (regelmäßig)	2 (3,8%)	0 (0%)	2 (5,7%)	20 (8,4%)
Interferenz oder Unsicherheit	6 (11,3%)	2 (8,7%)	2 (5,7%)	33 (13,8%)
Glottalverschluss	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<b>Gesamt</b>	<b>53 (100%)</b>	<b>23 (100%)</b>	<b>35 (100%)</b>	<b>238 (100%)</b>

**Tabelle 29:** Vergleich segmenteller Längung des turnfinalen Reims an turnfinalen und turninternen prosodischen Grenzen.

	<b>Turnfinal mit Pause (n=183)</b>	<b>Turnfinal mit unmittelbarem Anschluss (n=66)</b>	<b>Turnfinal mit Überlagerung (n=128)</b>	<b>Turnintern (n=3399)</b>
Kürzung	1 (0,5%)	0 (0%)	2 (1,5%)	75 (2,2%)
Weder Längung noch Kürzung	47 (25,7%)	18 (27,3%)	40 (31,3%)	1087 (32,0%)
Längung	135 (73,8%)	48 (72,7%)	86 (67,2%)	2237 (65,8%)
<b>Gesamt</b>	<b>183 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>	<b>3399 (100%)</b>

Alle dargestellten Verteilungen sind sich generell sehr ähnlich. Unterschiede zeigen sich lediglich in einer erhöhten Häufigkeit starker melodischer Bewegungen zwischen turnfinalen Grenzen ohne Überlagerung und turninternen Grenzen sowie in einer erhöhten Häufigkeit von Knarrstimme in der turnfinalen Position. Dies bedeutet, dass über die phonetische Struktur prosodischer Grenzen keine klare Zuordnung zu Turnabschlüssen vs. Turnfortsetzungen erreicht werden kann.

## 5.5 Zusammenfassung und abschließende Diskussion zu den turnfinalen prosodischen Grenzen

Die Ergebnisse aus den drei Analysebereichen (automatische Merkmalsextraktion, Perzeptionsexperimente und interpretative Analyse) ergeben ein komplexes Bild und tragen einiges zur Klärung der untersuchten Fragestellungen bei. Sie werfen aber auch neue Fragen und Probleme auf. Es sei vorangestellt, dass die Ergebnisse eine Aufstellung phonetisch definierter Kategorien turnfinaler prosodischer Grenzen mit eindeutiger funktionaler Zuordnung nicht zulassen, da die Variationsbreite prosodischer Strukturen im Korpus zu groß ist und auch in den Experimenten die Urteile der Versuchspersonen keine klaren Kategorien zeigen.

Als erster Diskussionspunkt stellt sich die Frage, ob es ein striktes, phonetisch kodiertes System gibt, das zwischen turninternen und turnfinalen Grenzen unterscheidet und so die Funktionen 'Weiterweisung' vs. 'Abschluss' trägt. Die Untersuchungen zeigen, dass die segmentelle Dauerstruktur zu dieser Funktionsopposition kaum beiträgt, sondern vor allem starke melodische Bewegungen Turnabschlüsse markieren. Hierbei macht die stark fallende Kontur den größten Teil der starken melodischen Bewegungen aus. In den Terminabsprachedaten gibt es am Ende von etwa 80% der Turns starke melodische Konturen, während es an

prosodischen Grenzen im Turninneren nur knapp 40% sind. Auch das Perzeptionsexperiment zeigt eine Bindung der stark fallenden Kontur mit dem Abschluss von Redebeiträgen. Die Untersuchungen belegen aber auf der anderen Seite auch, dass es im Inneren von Turns zu einer Vielzahl von prosodischen Grenzen kommt, die phonetische Merkmale tragen, die typisch für die turnfinale Position sind. Besonders häufig sind diese Strukturen zu finden, wenn im Inneren von Redebeiträgen semantisch-pragmatische Blöcke abgeschlossen werden. Auch Gilles (2005:21) beobachtet in seinen Daten solche Abschlüsse ohne nachfolgenden Sprecherwechsel:

*"Turn-interne Abschlüsse markieren z.B. das Ende von Erzählungen oder Sachverhaltsdarstellungen oder kleinere erzähltechnische, thematische oder argumentative Einschnitte (Ende eines Erzählschritts, eines Themas, eines Subthemas oder eines Einschubs)."*

Um also besser mit den Konzepten 'Abschluss' vs. 'Weiterweisung' arbeiten zu können, ist m.E. eine Untergliederung der beiden Funktionen in 'inhaltsbezogen' vs. 'dialogbezogen' sinnvoll. So können Abschlüsse im Inneren von Turns auf einer inhaltlichen Ebene als abschließend klassifiziert werden, ohne dass sie auch zu einem interaktionalen Abschluss mit Sprecherwechsel führen. Die Unterteilung in 'inhaltsbezogene' und 'dialogbezogene' Abschlüsse und die damit verbundene Möglichkeit inhaltlicher Abschlüsse ohne dialogbezogene Abschlussfunktion ist zudem sinnvoll, wenn man von Gesprächssituationen ausgeht, in denen Sprecherwechsel nicht üblich sind. Wenn z.B. ein Vortrag gehalten wird, ist innerhalb der fortlaufenden Rede häufig mit inhaltlichen Abschlüssen zu rechnen, die lediglich die thematische Gliederung (z.B. in Absätze) unterstützen, aber dialogbezogen keineswegs dazu anregen sollen, dass ein Zuhörer das Wort übernimmt.

Auch für das Konzept der Weiterweisung ist zu erwarten, dass die dialogbezogene und die inhaltsbezogene Ebene nicht immer parallel laufen; denn eine inhaltliche Weiterweisung kann zwar mit dem Wunsch nach einer Turnfortsetzung einhergehen, aber auch hier ist ein Auseinandergehen der beiden funktionalen Aspekte vorstellbar, z.B. wenn ein Sprecher ein Thema eröffnet und dem Hörer dann den Turn übergibt, damit dieser Ergänzungen vornehmen kann. Eventuell ist es sinnvoll, dieses Konzept von inhaltlicher Weiterweisung und interaktionalem Abschluss auch auf Fragen zu beziehen, denn eine Frage wirft oft ein Thema auf, ist

also weiterweisend, aber ist in interaktionaler Hinsicht abschließend, da der Turn mit der Frage abgeschlossen wird.

Eine rein phonetisch ausgerichtete Differenzierung zwischen inhaltsbezogenen und dialogbezogenen Abschlüssen und Weiterweisungen ist m.E. allerdings nicht möglich; denn die phonetischen Merkmale der funktionalen Klassen zeigen starke Überschneidungen, und für Hörer ist vermutlich vor allem aus dem inhaltlichen und situativen Gesamtkontext ableitbar, zu welchem Zeitpunkt z.B. ein dialogbezogener Abschluss erreicht wird und eine Übernahme des Rederechts möglich ist oder erwartet wird. Hier liegt auch eines der großen Probleme für experimentelle Untersuchungen der Dialogsteuerung. Es kann nie ganz klar bestimmt werden, inwieweit die Versuchspersonen Urteile über die inhaltliche oder die dialogbezogene Ebene abgeben. Ebenso ist der Einfluss des lexikalischen und situativen Kontextes auf die Urteile schwer abschätzbar.

Die eben diskutierten Funktionen sind Kategorien der Sprecherseite. D.h., es wird angenommen, dass der Sprecher einen (groben) Plan für die Struktur seines Turns hat, zu einem bestimmten Zeitpunkt einen dialogbezogenen Abschluss erreicht, dies inhaltlich und phonetisch zum Ausdruck bringt und der Hörer ohne Verzögerung in die Sprecherrolle wechselt. Diese Sichtweise ist allerdings idealisiert. In gleichberechtigten Alltagsdialogen sind oft Sprecherwechsel an Positionen zu beobachten, die keinerlei Merkmale eines Abschlusses erkennen lassen. Hier entsteht meistens der Eindruck, dass der Hörer eine unterbrechende Turnübernahme durchführt. Aus diesem Grund muss in die Klassifikation von Sprecherwechseln auch die Aktivität des Hörers und somit eine Kombination der Perspektiven von Sprecher und Hörer einfließen. Um dieses Ziel zu erreichen, sollte eine Unterteilung in 'gezielte Turnübergaben', 'erlaubte Turnübernahmen' und 'umstrittene Turnübernahmen' auf empirischer Basis belegt werden. Als Ausgangspunkt hierfür stand die Annahme, dass Turnübergänge mit überlagerndem Sprechen eher in die Kategorie der 'erlaubten' oder der 'umstrittenen' Übergänge fallen, während Übergänge mit unmittelbarem Anschluss oder einer Pause zwischen den Redebeiträgen durch den Sprecher gezielt eingeleitet werden. Diese Annahme müsste sich dadurch bestätigen lassen, dass an den überlagernden Übergängen zumeist Merkmale der interaktionalen Weiterweisung zu finden sind (z.B. inhaltliche und syntaktische Unabgeschlossenheit, schwache melodische Bewegungen), während die nicht-überlagernden Übergänge Merkmale von Abschlüssen tragen (z.B. inhaltliche und syntaktische Abgeschlossenheit, starke melodische Bewegungen).

Die Untersuchungsergebnisse bestätigen eine strikte Trennung dieser Kategorien jedoch nicht. An nahezu allen untersuchten Turnübergängen finden sich sowohl (mögliche) syntaktische Abschlüsse als auch Endpunkte von inhaltlich als komplett interpretierbaren Einheiten. Auch die melodischen Muster, die auf der phonetischen Ebene entscheidend für Abschluss vs. Weiterweisung sind, stützen die angestrebte Einteilung nur zum Teil. Zwar liegt der Anteil der starken melodischen Bewegungen an den überlagernden Übergängen mit 37,5% am niedrigsten, jedoch an den Übergängen mit unmittelbarem Anschluss auch nur bei 56,1% und an denen mit Pause bei 49,2%. Hieraus lässt sich ableiten, dass unter den überlagernden Übergängen weniger gezielt durch den Sprecher eingeleitete Übergaben sind. Auf der anderen Seite gibt es aber auch mehr als ein Drittel überlagernder Übergänge mit starken melodischen Mustern. Ebenso muss bei den nicht-überlagernden Übergängen der hohe Anteil von finalen prosodischen Grenzen ohne starke melodische Bewegung berücksichtigt werden, der eine funktionale Zuordnung als 'gezielte Turnübergaben' nicht stützt, da hier der melodische Abschluss fehlt.

Es ist also davon auszugehen, dass sich in jeder Klasse der verschiedenen synchronisierten Übergänge sowohl gezielte als auch erlaubte Übergänge befinden. Das systematische Abhören der Dialoge lässt weiterhin die Aussage zu, dass wirklich umstrittene Übergänge in den Daten kaum vorkommen, da die Dialoge sehr kooperativ verlaufen und die unterbrochenen Sprecher sich fast immer ohne den Versuch, das Rederecht zu halten, in die Hörerrolle begeben. Die unterbrechenden Hörer ihrerseits wählen die Position ihrer Unterbrechung so, dass der derzeitige Sprecher dort einen (möglichen) inhaltlichen und syntaktischen Abschluss erreicht hat. Dies scheint sogar bei den überlagernden Übernahmen der Fall zu sein. Die Übernahmeposition ist hier in den meisten Fällen so gewählt, dass der unterbrochene Sprecher die begonnene inhaltliche und syntaktische Struktur während der Überlagerung noch komplettieren kann. Dies gilt aber keineswegs für jede Art von Gespräch, sondern ist als spezifische Eigenschaft der LINDENSTRASSE-Dialoge anzusehen. Hier herrscht anscheinend Einvernehmen darüber, dass Unterbrechungen ein normaler Bestandteil der Dialogstrukturierung sind und der Unterbrechende üblicherweise sofort das Rederecht bekommt.

Aus diesen Beobachtungen und Interpretationen der Dialogdaten und aus den experimentellen Ergebnissen ergeben sich einige neue Aspekte hinsichtlich interaktionaler Muster im Dialog und der Rolle der Phonetik in der Dialogsteuerung. Es zeigt sich (vor allem durch das Perceptionsexperiment), dass die funktionale Bindung intonatorischer Konturen und

begleitender phonatorischer Veränderungen hinsichtlich der Turnübergangsbekundung schwächer ausgeprägt ist als erwartet. Die Ergebnisse deuten eher auf eine unterstützende Funktion der phonetischen Komponente hin, die erst in Kombination mit syntaktischen und semantischen Strukturen zu interaktionalen Funktionalisierungen führt. Dies bedeutet, dass für eine aussagekräftige Klassifikation von einem erweiterten Merkmalsbündel ausgegangen werden muss, das auch die semantische und syntaktische Struktur einbezieht. Hierdurch entsteht ein wesentlich weiterer Rahmen, in dem die Gesprächspartner ihr interaktionales Verhalten steuern und Einschätzungen über adäquate Reaktionen im Gespräch vornehmen.

Aber auch dieses erweiterte Bündel führt m.E. nicht zu einem starren System, das Sprecherwechsel bei bestimmten Bündelungen "vorschreibt" bzw. an anderer Stelle "verbietet". Sprecher und Hörer zeigen im Dialog oft ein großes Maß an Flexibilität und gegenseitiger Anpassungsfähigkeit. So kann das grundlegende Prinzip des Turntakings als 'erlaubend' und nicht als 'vorschreibend' charakterisiert werden. Dieser erlaubende Charakter lässt sich durch die folgenden zwei Situationen illustrieren, die intuitiv gut nachvollziehbar sind.

- (1) Ein Sprecher schließt eine Äußerung durch melodische, syntaktische und semantische Vervollständigung ab und macht eine kurze Pause. Der Hörer kann an dieser Stelle nun das Wort ergreifen. Oft wird er es aber nicht tun, sondern z.B. durch ein Rezeptionssignal andeuten, dass er weiter in der Hörerrolle bleiben möchte. Üblicherweise wird der Sprecher dann seine Rede fortsetzen (es sei denn er hat eine direkte Frage an den Partner gestellt).
- (2) Ein Sprecher lässt in seiner Rede kein komplexes Merkmalsbündel entstehen, das einen interaktionalen Abschluss erkennen lässt. Dennoch hat der Hörer das Bedürfnis "einzuhaken", weil er z.B. etwas Wichtiges ergänzen möchte. Der Hörer wird die Ergänzung in vielen Fällen als unterbrechende Turnübernahme beginnen, und der vormalige Sprecher wird dies üblicherweise akzeptieren.

Anders ist die Situation, wenn eine direkte Frage an den Gesprächspartner gestellt wird. In diesem Fall liegt ein eindeutiger gezielter Turnübergang vor. Der Komplex der Frage wird aber durch die vorgenommenen Untersuchungen nicht explizit analysiert. Hier sind weiterführende Untersuchungen notwendig.

Die obigen Erklärungsversuche entfernen sich von der rein phonetischen Fragestellung und machen deutlich, dass der Versuch einer wirklichen Erklärung interaktionaler Strukturen die Grenzen der experimentellen Phonetik zeigt. In diesem Zusammenhang soll noch einmal auf das Forschungsparadigma der Gesprächsanalyse hingewiesen werden, das über kontextuelle Interpretationen versucht, Zugang zu funktionalen Kategorien im Gespräch zu finden, aber kaum zu statistischen Validierungen ihrer Ergebnisse gelangt. M.E. ist eine Kombination experimenteller und interpretativer Verfahren ein geeigneter Weg, sich dem Phänomen "Gespräch" weiter zu nähern, wobei das in diesem Abschnitt gewählte Interpretationsverfahren offensichtlich noch der Verfeinerung bedarf.

## **5.6 Klassifikationssystem II - Turnfinale prosodische Grenzen**

Das folgende Klassifikationssystem (Tabelle 30) setzt die Ergebnisse der Produktionsdatenanalyse und der Perzeptionsexperimente in Form-Funktionsbeziehungen unterschiedlicher turnfinaler prosodischer Grenzen um. Aus der Datenlage ergeben sich sechs zu differenzierende Klassen turnfinaler prosodischer Grenzen. Da die phonetischen Strukturen an prosodischen Grenzen stark variieren, kann keine eindeutige Form-Funktions-Beziehung angesetzt werden. Die aufgestellten Kategorien sind deshalb als prototypische Kernbereiche zu verstehen, von denen es bedingt durch individuelle und situative Variation zu deutlichen Abweichungen kommen kann.

Da die Ergebnisse verschiedener Untersuchungen zeigen, dass die syntaktische Struktur ein wichtiger Informationsträger für die interaktionale Strukturierung von Gesprächen ist, wird sie mit in die Definition verschiedener prosodischer Grenzen einbezogen. Aus der Literatur und aus informellen Korpusuntersuchungen ist zudem bekannt, dass auch Intensitätsverläufe Träger interaktional relevanter Information sein können. Aus diesem Grund wird für einige Kategorien auch die Intensitätsstruktur in das Merkmalsbündel einbezogen.

Das t in den Klassen PGt1-PGt6 zeigt an, dass es sich um turnfinale prosodische Grenzen handelt.

**Tabelle 30:** Klassifikationssystem II.

<b>Nomenklatur</b>	<b>Merkmalsbündel</b>	<b>Funktionen</b>
PGt1	<i>Stark fallende melodische Bewegung, evtl. Intensitätsabnahme und Veränderung der Stimmqualität, Keine Überlagerung, Syntaktischer Abschluss</i>	Gezielte Turnübergabe: inhaltlicher Abschluss, interaktionaler Abschluss
PGt2	<i>Steigende melodische Bewegung, Keine Überlagerung, Syntaktischer Abschluss</i>	Gezielte Turnübergabe mit Frage an den Gesprächspartner: inhaltliche Weiterweisung, interaktionaler Abschluss
PGt3	<i>Stark fallende melodische Bewegung, Pause oder direkter Anschluss Syntaktischer Abschluss wird während der Überlagerung erreicht</i>	Erlaubte Turnübernahme: inhaltlicher Abschluss, interaktionale Weiterweisung
PGt4	<i>Keine starke melodische Bewegung Überlagerung, Syntaktischer Abschluss / Kein syntaktischer Abschluss</i>	Erlaubte Turnübernahme: inhaltliche Weiterweisung, interaktionale Weiterweisung
PGt5	<i>Keine starke melodische Bewegung Überlagerung, Kein syntaktischer Abschluss, Glottalverschluss</i>	Erlaubte temporäre Turnübernahme: inhaltliche Weiterweisung, interaktionale Weiterweisung
PGt6	<i>Keine starke melodische Bewegung, Überlagerung, Hohe Intensität während der Überlagerung, Syntaktischer Abschluss / Kein syntaktischer Abschluss</i>	Erzwungene Turnübernahme: inhaltliche Weiterweisung, interaktionale Weiterweisung

## 6 Prosodische Grenzen in unflüssiger Sprache

In spontaner Sprache kommt es oft zu Planungsproblemen und damit zu Äußerungen, die Fehler enthalten. Diese Fehler (z.B. inhaltlich falsche Aussagen, syntaktische Fehlkonstruktionen oder artikulatorische Fehler) werden häufig von den Sprechern selbst unmittelbar bemerkt. Es ist aus den untersuchten Produktionsdaten bekannt, dass solche Fehler in der Planung und/oder der Ausführung eines Redebeitrags oft prosodische Grenzen verursachen, die dadurch entstehen, dass der Sprecher zögert oder "sich selbst unterbricht". Neben eindeutigen Fehlern kommt es durch Probleme bei der simultan zum Sprechen ablaufenden Äußerungsplanung auch immer wieder zu Verzögerungen wie Pausen und segmentellen Längungen, die nicht wie in flüssiger Sprache durch die inhaltliche Gliederungsfunktion erklärt werden können, aber z.T. deutliche perzeptive Einschnitte verursachen. Daraus erwächst die Frage, wie sich prosodische Grenzen in flüssiger Sprache von prosodischen Grenzen in unflüssiger Sprache unterscheiden und welche phonetischen Merkmale von Hörern als Hinweise auf Planungsprobleme bzw. eine unflüssige Sprechweise wahrgenommen werden. Denn für Hörer ist es wichtig, im Gespräch zwischen Einschnitten, die inhaltliche und syntaktische Strukturen widerspiegeln, und solchen, die aufgrund von Produktionsproblemen auftreten, zu unterscheiden. Geschieht dies nicht, kann das Verstehen eines Redebeitrags und eine angemessene Hörerreaktion im Rahmen eines flüssigen Gesprächsablaufs erheblich erschwert werden.

Der Entstehung von Sprechflussstörungen liegen unterschiedliche Ursachen zugrunde. Folgende Probleme können zu einer unflüssigen Sprechweise führen:

- **Phonetische Versprecher:** Während des Sprechens kann es zu Lautvertauschungen, Fehlartikulationen und anderen artikulatorischen Problemen kommen.
- **Syntaktische Versprecher:** Ein bereits zum Teil ausgesprochener Satz enthält agrammatische Strukturen oder lässt sich nicht sinnvoll ohne eine Verletzung grammatischer Regeln fortsetzen.
- **Inhaltliche Versprecher:** Ein Wort, das nicht der Intention des Sprechers entspricht, wird versehentlich ganz oder zum Teil ausgesprochen und eine Korrektur wird notwendig. In diese Kategorie fallen auch die sog. Freudschen Versprecher (Freud z.B. 1901/1971).

- Wortfindungsprobleme: Ein Wort, das eigentlich bekannt ist, kann nicht rechtzeitig aus dem Gedächtnis abgerufen werden.
- Syntaktische Planungsprobleme: Eine korrekte syntaktische Fortsetzung kann nicht so schnell geplant werden, dass die derzeitige Äußerung flüssig fortgesetzt werden kann.

Die beschriebenen möglichen Ursachen unflüssigen Sprechens haben eines gemeinsam: Der Sprecher weiß im Grunde, was er sagen will, hat aber Probleme bei der Gestaltung der intendierten Äußerung (artikulatorisch, syntaktisch, lexikalisch oder inhaltlich). Hierdurch kann es zu einer unflüssigen Sprechweise kommen, die sich u.U. durch starke prosodische Einschnitte, also prosodische Grenzen, manifestiert.

Unabhängig davon, welche Ursache unflüssiges Sprechen hat, kann m.E. davon ausgegangen werden, dass Hörer Unflüssigkeiten wahrnehmen und dadurch Rückschlüsse auf den mentalen Sprachplanungsprozess des Sprechers möglich werden: Der Planungsprozess wird z.T. durch die phonetische Struktur der Äußerung offenbart, da Planungsprobleme durch bestimmte Exponenten im Signal repräsentiert werden. Das Nachvollziehen des Planungsprozesses hilft dem Hörer bei der inhaltlichen und syntaktischen Interpretation des bereits Gehörten und bei der Einschätzung des weiteren Äußerungs- bzw. Gesprächsverlaufs.

Das grundlegende Problem in der Differenzierung flüssiger und unflüssiger prosodischer Grenzen ist, dass in natürlicher Sprache für beide Typen der phonetischen Separierung ähnliche phonetische Parameter zum Einsatz kommen. Vor allem Pausen und Atmungsphasen sowie segmentelle Längungen sind sowohl an flüssigen als auch an unflüssigen Grenzen zu finden. Andere Phänomene sind allerdings typisch für prosodische Grenzen, die durch Planungsprobleme entstehen (z.B. Häsitationspartikeln). Auch auf der lexikalischen und syntaktischen Ebene finden sich Korrelate unflüssiger Sprechweise: Es kann hier zu Wortwiederholungen und syntaktischen Abbrüchen kommen.

Das folgende Kapitel umfasst eine Analyse der im Korpus gefundenen Merkmale unflüssiger prosodischer Grenzen und stellt über Perzeptionsexperimente Beziehungen zwischen akustisch/phonetischen und perzeptiven Größen bei der Wahrnehmung von Sprechflussstörungen her. Unflüssige prosodische Grenzen werden hierbei vor allem in turninterner Position analysiert, denn es soll primär die phonetische Struktur von unflüssigen Grenzen untersucht werden, die Sprecher ohne direktes Zutun des Gesprächspartners produzieren und weniger Unflüs-

sigkeiten, die z.B. durch Unterbrechungen durch den Partner entstehen können.

### 6.1 Labelbasierte akustisch-phonetische Analyse der Produktionsdaten

Für die Analyse der phonetischen Merkmale unflüssiger prosodischer Grenzen stellt sich zuerst die Frage, wie im Rahmen einer labelbasierten Korpusabfrage der Zugriff auf unflüssige Grenzen vorgenommen werden kann, denn in der Etikettierung ist keine explizite Trennung zwischen flüssigen und unflüssigen Grenzen enthalten. Da es für die Analyse der phonetischen Merkmale aber notwendig ist, diese Trennung zu erreichen, müssen Kriterien gefunden werden, um gezielt auf prosodische Grenzen zuzugreifen, die im Zusammenhang mit Sprechflussstörungen stehen. Hierfür werden drei Kriterien gewählt:

- In der Etikettierung sind syntaktische Abbrüche gekennzeichnet. Wenn an der Position eines syntaktischen Abbruchs eine Phrasengrenze etikettiert ist, kann davon ausgegangen werden, dass es sich nicht um eine gezielte, inhaltlich gliedernde Grenze handelt, sondern dass es an dieser Position der Äußerung syntaktische Planungsprobleme gab, die zu dem prosodischen Einschnitt geführt haben. Die Klasse der prosodischen Grenzen an syntaktischen Abbrüchen bildet also die erste Gruppe untersuchter prosodischer Grenzen.
- Häsitationspartikeln wie *äh* oder *ähm* sind in den Etikettierdateien markiert. Diese Partikeln werden normalerweise dann von Sprechern produziert, wenn der weitere Äußerungsverlauf geplant wird und dies nicht schnell und flüssig während des fortlaufenden flüssigen Sprechens bewältigt werden kann. Prosodische Grenzen, an denen solche Partikeln vorkommen, stehen daher ebenfalls in Verbindung mit Planungsproblemen und bilden eine weitere Gruppe der Analyse.
- Der dritte Hinweis auf Unflüssigkeit an prosodischen Grenzen liegt in der Etikettierung der segmentellen Dauerstruktur. Die Etikettierer haben im Rahmen der segmentellen Etikettierung an den Positionen, wo sie eine deutliche Zögerungslängung wahrgenommen haben, das Etikett *z:* eingefügt. Diese Kennzeichnung ist häufig im Korpus zu finden und zwar dort, wo einzelne oder mehrere Segmente ungewöhnlich lange Dauern aufweisen. Die Klasse von prosodischen Grenzen,

an denen eine solche Markierung gesetzt wurde, bildet den dritten Untersuchungsbereich bei der automatischen Korpusabfrage.

In den folgenden Abschnitten werden die extrahierbaren phonetischen Merkmale dieser drei Klassen von unflüssigen prosodischen Grenzen getrennt dargestellt (LINDENSTRASSE-Korpus). Um zu einer klaren Trennung der drei automatisch abfragbaren Analysebereiche zu kommen, wird für jede Klasse das Auftreten der anderen beiden Merkmale ausgeschlossen, wenn also z.B. ein syntaktischer Abbruch zusammen mit einer nachfolgenden Häsitationspartikel vorkommt, wird diese prosodische Grenze in keine der drei Klassen einbezogen. Auf diese Weise können die drei Analysebereiche klar voneinander getrennt werden. In die Auswertung werden nur prosodische Grenzen einbezogen, an denen im Rahmen der Etikettierung eine phrasenfinale Kontur zugeordnet wurde. In vielen Fällen gilt dies nicht, da bei Phrasenfragmenten, die durch eine prosodische Grenze abgebrochen werden, bevor eine satzakzentuierte Silbe in der Phrase auftritt, in der Etikettierung auf eine Kennzeichnung des finalen Melodiemusters verzichtet wird. Folgende Klassen ergeben sich:

- 204 Fälle von prosodischen Grenzen an syntaktischen Abbrüchen (ohne Zögerungslängung und ohne nachfolgende Häsitationspartikel),
- 204 Fälle von prosodischen Grenzen mit nachfolgender Häsitationspartikel (ohne Zögerungslängung und ohne syntaktischen Abbruch),
- 129 Fälle von prosodischen Grenzen mit Zögerungslängung (ohne nachfolgende Häsitationspartikel und ohne syntaktischen Abbruch).

Im LINDENSTRASSE-Korpus kommen an insgesamt 288 prosodischen Grenzen Kombinationen der drei Merkmale vor, die nicht in die Analyse eingehen. Folgende Kombinationen treten auf:

- Syntaktischer Abbruch und Häsitationspartikel: 126 Fälle,
- Syntaktischer Abbruch und Zögerungslängung: 74 Fälle,
- Zögerungslängung und Häsitationspartikel: 55 Fälle,
- Syntaktischer Abbruch, Zögerungslängung und Häsitationspartikel: 33 Fälle.

Sicher gibt es im Korpus auch eine Vielzahl prosodischer Grenzen, die durch Verzögerungen im simultanen Planungsprozess entstehen, ohne eines der oben genannten Merkmale zu zeigen. Auf solche Grenzen könnte lediglich durch eine umfangreiche Interpretation prosodischer Grenzen zugegriffen werden. Deshalb beziehen sich alle nachfolgenden Analysen in diesem Abschnitt auf die oben eingeführten Klassen.

### 6.1.1 Melodische Muster

Tabelle 31 zeigt die etikettierten melodischen Muster an den drei Typen unflüssiger prosodischer Grenzen.

**Tabelle 31:** Melodische Verläufe an turninternen unflüssigen Grenzen (LINDENSTRASSE).

Melodischer Verlauf	Syntaktischer Abbruch (n=204)	Häsiationspartikel (n=204)	Zögerungslängung (n=129)	Gesamt (n=537)
<b>Starke melodische Bewegung</b>				
Stark fallend	19 (9,3%)	26 (12,8%)	13 (10,1%)	58 (10,8%)
Stark fallend + sehr leicht steigend	2 (1,0%)	2 (1,0%)	1 (0,8%)	5 (0,9%)
Fallend-steigend	1 (0,5%)	8 (3,9%)	1 (0,8%)	10 (1,9%)
Stark Steigend	0 (0%)	3 (1,5%)	0 (0%)	3 (0,6%)
Gesamt starke melodische Bewegung	22 (10,8%)	39 (19,1%)	15 (11,6%)	76 (14,2%)
<b>Schwache melodische Bewegung</b>				
Eben	47 (23%)	61 (29,9%)	22 (17,1%)	130 (24,2%)
Leicht fallend	125 (61,3%)	85 (41,7%)	80 (62,0%)	290 (54,0%)
Leicht bis mittelhoch steigend	3 (1,5%)	12 (5,9%)	3 (2,3%)	18 (3,4%)
Eben + sehr leicht steigend	0 (0%)	0 (0%)	3 (2,3%)	3 (0,6%)
Leicht fallend + sehr leicht steigend	7 (3,4%)	7 (3,4%)	6 (4,6%)	20 (3,7%)
Gesamt schwache mel. Bewegung	182 (89,2%)	165 (80,9%)	114 (88,4%)	461 (85,9%)
<b>Gesamt</b>	<b>204 (100%)</b>	<b>204 (100%)</b>	<b>129 (100%)</b>	<b>537 (100%)</b>

Die Verteilungen in den drei Klassen sind ähnlich. In allen Klassen ist der Anteil der starken melodischen Bewegungen an der prosodischen Grenze relativ klein, wobei fast alle Fälle starker melodischer Bewegun-

gen stark fallende Muster sind. Der auffälligste Unterschied ist, dass die leicht fallenden Konturen an syntaktischen Abbrüchen und an Grenzen mit Zögerungslängung über 60% der Fälle ausmachen, während es für die Klasse mit nachfolgender Häsitationspartikel nur knapp 42% sind, sich hier aber ein erhöhter Anteil ebener Konturen findet. Dieser erhöhte Anteil resultiert evtl. daraus, dass Äußerungsteile vor Häsitationspartikeln häufiger flüssig und inhaltlich komplett abgeschlossen werden und die Planungsschwierigkeiten einen weiteren nachfolgenden inhaltlichen Block betreffen. Bei syntaktischen Abbrüchen hingegen ist das Planungsproblem jedoch eindeutig im aktuellen inhaltlichen Block lokalisiert, da ein inhaltlicher Abschluss ohne syntaktischen Abschluss nicht zu erwarten ist.

### **6.1.2 Pausen**

An den drei Typen von unflüssigen prosodischen Grenzen finden sich Pausen und Atmungsphasen mit folgenden Anteilen: Grenzen mit syntaktischen Abbrüchen: 36,8% (75 Fälle), Grenzen mit nachfolgenden Häsitationspartikeln: 21,1% (43 Fälle), Grenzen mit Zögerungslängung: 28,7% (37 Fälle). Dieses Ergebnis zeigt, dass Sprechpausen in vielen Fällen als Korrelate unflüssiger Sprechweise vorkommen. Möglicherweise ist die etwas geringere Häufigkeit von Pausen an prosodischen Grenzen mit Häsitationspartikeln und Zögerungslängung damit zu erklären, dass diese phonetischen Phänomene eine Art Ersatz für Pausen darstellen und dadurch deutlich die momentanen Planungsprobleme signalisieren und das Rederecht sichern.

### **6.1.3 Längung**

Zunächst war auch eine automatische labelbasierte Quantifizierung der segmentellen Längung an unflüssigen prosodischen Grenzen geplant. Nach eingehenden Voruntersuchungen unflüssiger Grenzen wurde diese Vorhaben jedoch eingestellt. Hierfür gibt es folgende Gründe:

- Die segmentelle Dauerstruktur an unflüssigen prosodischen Grenzen unterliegt einer starken Variation in der Position der gelängten Abschnitte. Während sich die segmentelle Längung an flüssigen prosodischen Grenzen relativ zuverlässig auf den finalen Reim beschränkt, dehnt sich Zögerungslängung teilweise über ganze Silbenketten aus oder betrifft einzelne Segmente mitten in Wörtern. Die segmentelle

Etikettierung erlaubt keine genaue Positionsbestimmung der etikettierten Zögerungslängungen. Wenn z.B. einem Wort das Label *z:* folgt, das wahrgenommene Zögerungslängung markiert, ist nicht ableitbar, ob die Längung das ganze Wort oder nur das letzte Segment betrifft. Messungen, die auf dieser Basis durchgeführt werden könnten, wären im Nachhinein kaum interpretierbar und geben wenig Aufschluss über tatsächliche Dauerstrukturen. Ein Beispiel für diese Problematik sind die Dauerstrukturen von drei unakzentuierten Realisierungen des Wortes *weil*, die die Sprecherin APE im Dialog 103a innerhalb eines Ausschnitts von fünf Sekunden unflüssiger Sprechweise mit einer starken Variation in der segmentellen Dauerstruktur produziert. Tabelle 32 zeigt die Segmentdauern der drei Realisierungen. Die Realisierung in Zeile 1, im Gegensatz zu den beiden anderen, wirkt nicht gezögert. Die Zahlen zeigen, dass in Realisierung 2, im Vergleich zu Realisierung 1, vor allem der Silbenonset und der Silbennukleus gelängt sind, in Realisierung 3 hingegen der Onset und der Offset der Silbe. Eine Mittelwertbildung über die gesamte Silbe würde also keinen geeigneten Messwert zur Beschreibung der Dauerstruktur ergeben, da die Unterschiede in der Position der gelängten Segmente hierdurch nicht erfasst werden könnten.

**Tabelle 32:** Segmentdauern von drei Realisierungen des Wortes *weil* (in ms).

	[v]	[ai]	[l]
Realisierung 1	30	140	70
Realisierung 2	180	415	110
Realisierung 3	100	130	230

- Weiterhin ist in den Daten eine große individuelle Fluktuation in Art und Häufigkeit segmenteller Längungen bei Produktionsstörungen beobachtbar. Einzelne Sprecher scheinen deutlich unterschiedliche Strategien zu verfolgen, wenn sie Probleme mit der Äußerungsplanung haben. Einige Sprecher verwenden viele Häsitationspartikeln, andere statt dessen ungefüllte Pausen, dritte wiederum starke segmentelle Dehnungen. Eine wirklich aussagekräftige Untersuchung müsste folglich zuerst das Verhalten unterschiedlicher Sprecher untersuchen, bevor verallgemeinernde Messungen über Korpora mit vielen Sprechern vorgenommen werden.

- Der dritte Punkt, der eine quantitative Untersuchung segmenteller Längung bei Sprechflussstörungen fragwürdig macht, hängt mit der generellen Konzeption der vorliegenden Arbeit zusammen. Ziel der Arbeit ist die Analyse phonetischer Merkmalsbündel an prosodischen Grenzen. Im Korpus tritt aber etikettierte Zögerungslängung auch an vielen Positionen auf, an denen keine prosodische Grenze etikettiert wurde. Insgesamt sind im LINDENSTRASSE-Korpus 403 Fälle von Zögerungslängung etikettiert. Davon stehen 282 Fälle in Verbindung mit einer unmittelbar folgenden prosodischen Grenze. Dass in vielen Fällen keine prosodische Grenze etikettiert wurde, liegt daran, dass die vorhandene Längung nicht zu einem deutlichen perzeptorischen Einschnitt geführt hat. Auch die Perzeptionsexperimente 1a, 1b und 1c (siehe 4.2) belegen, dass segmentelle Längung allein nicht zu einer deutlichen Separierung führt. Eine inhaltlich sinnvolle Untersuchung von Zögerungslängung müsste also auch die Längungen mit einbeziehen, die ohne prosodische Grenzen vorkommen und würde so zu einer erheblichen Erweiterung der ursprünglichen Fragestellung führen.

#### 6.1.4 Phonation

Wie unter 5.1.4 beschrieben, ist an syntaktischen Abbruchstellen z.T. Knarrstimme zu beobachten. Tabelle 33 zeigt die Häufigkeiten verschiedener phonatorischer Ausprägungen an prosodischen Grenzen mit syntaktischen Abbrüchen. Für die Analyse werden sämtliche syntaktische Abbruchstellen, die mit einer prosodischen Grenze zusammenfallen einer manuellen akustisch-phonetischen Analyse unterzogen (siehe 4.1.9). Diese Analyse bezieht sowohl turninterne als auch turnfinale syntaktische Abbrüche ein. Die untersuchten Fälle sind nach ihrer Position im Turn getrennt dargestellt. Die Anzahl der syntaktischen Abbrüche liegt hier höher als in den vorigen Abschnitten, da hier auch die prosodischen Grenzen einbezogen werden, für die kein komplettes Merkmalsbündel abgeleitet werden kann.

**Tabelle 33:** Phonationstyp an syntaktischen Abbrüchen (LINDENSTRASSE).

	<i>Turninterne syntaktische Abbrüche (n=544)</i>	<i>Turnfinale syntaktische Abbrüche (n=85)</i>	<i>Gesamt (n=629)</i>
Modalstimme	69,1%	71,8%	69,5%
Knarrstimme (unregelmäßig)	6,3%	4,7%	6,0%
Knarrstimme (regelmäßig)	3,1%	5,9%	3,5%
Glottalverschluss	5,5%	12,9%	6,5%
Rest (Interferenz und unsichere Beurteilung)	16,0%	4,7%	14,5%
<b>Gesamt</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

Die durchschnittliche Dauer von Knarrstimme an Abbruchstellen beträgt 96 ms. Die durchschnittliche Dauer von Glottalverschlüssen an Abbruchstellen beträgt 393 ms mit einer hohen Standardabweichung von 743 ms. Der melodische Verlauf an syntaktischen Abbruchstellen ist in der Regel leicht fallend (62,4%) oder eben (23,6%). Stark fallende Konturen treten nur in 9,5% aller Fälle auf. Glottalisierungen an den Abbruchstellen setzen zum größten Teil aus leicht fallenden oder ebenen Konturen heraus ein (52,2% bzw. 34,8%).

Die Verteilungen zeigen, dass das Vorkommen von Knarrstimme an syntaktischen Abbrüchen zwar nicht selten ist, dass Knarrstimme (mit oder ohne Glottalverschluss) aber auch nicht der Regelfall ist. Nimmt man alle drei Kategorien zusammen, die mit phonatorischen Veränderungen in Verbindung stehen (Knarrstimme regelmäßig/unregelmäßig und Glottalverschluss) ergibt sich für die turninterne Position eine relative Häufigkeit von 14,9% und für die turnfinale ein Anteil von 23,5%. Der erhöhte Anteil in turnfinaler Position geht vor allem auf den größeren Anteil turnfinaler Glottalverschlüsse zurück. Im Gegensatz zu Knarrstimme an flüssigen prosodischen Grenzen kann Knarrstimme an Abbruchstellen mit Glottalverschlüssen alternieren. Insoweit decken sich die Ergebnisse mit den Befunden von Kohler et al. 2005. Allerdings ist in den Daten aus dem LINDENSTRASSE-Korpus ein hoher Anteil von Glottalverschlüssen in turnfinaler Position festzustellen. Dies ist kaum in Einklang zu bringen mit der Interpretation von Local und Kelly (1986), die gehaltene Glottalverschlüsse als 'turn-holdig-device' funktional einordnen. Hieraus ergibt sich, dass Glottalverschlüsse vor allem in turninterner Position zu finden sein müssten. Die Daten zeigen aber eine gegenläufige Tendenz.

Es scheint vielmehr so zu sein, dass die hier beobachteten Glottalverschlüsse in einem anderen Zusammenhang vorkommen. Neun der zwölf beobachteten Glottalverschlüsse in turnfinaler Position treten während einer Überlagerung durch den Gesprächspartner auf. Der ursprüngliche Sprecher beendet den Turn also mit einem Glottalverschluss ohne die begonnene syntaktische Struktur abzuschließen. Dies lässt die Interpretation zu, dass Sprecher Glottalverschlüsse z.T. einsetzen, um sich selbst zu unterbrechen, wenn der Partner überlagernd zu sprechen beginnt. Die auditive Überprüfung dieser Fälle zeigt, dass in einigen Fällen der Glottalverschluss so lange gehalten wird, bis der unterbrochene Sprecher das Wort wieder übernehmen kann. Aus diesem Verhalten erklärt sich auch die große Standardabweichung der Verschlussdauern. In einem gewissen Sinn lässt sich dieses Verhalten dann auch wieder mit der Funktion des Turnhaltens in Verbindung bringen, denn die Selbstunterbrechung mit Glottalverschluss ist hier eine Art vorläufiger Turnübergabe und der Sprecher signalisiert durch die Art des Abschlusses, dass er nicht fertig gesprochen hat und bei nächster Gelegenheit wieder den Turn übernehmen wird. Der Sprecher schneidet sich gewissermaßen selbst die Luft ab und begibt sich mit durch den Verschluss aufgestauter Luft in "Wartstellung". Gelingt es dem Sprecher nicht, wieder an das Rederecht zu kommen, findet sich in einigen Fällen ein hörbares Lösungsgeräusch des Glottalverschlusses, das von einem deutlichen Ausatmen begleitet wird und anzeigt, dass der Sprecher den Plan der unmittelbaren Wiederübernahme des Rederechts aufgegeben hat.

Da das Vorkommen von Glottalisierung und Glottalverschluss an Abbruchstellen keineswegs die Regel ist, muss in die Interpretation und funktionale Einordnung von Glottalverschlüssen an syntaktischen Abbruchstellen ein weiterer Aspekt einbezogen werden, der sich aus der morphophonologischen Struktur des Deutschen ergibt: Wenn im Deutschen morpheminitial ein Vokal steht, erfordert dies in der phonetischen Realisierung (zumindest in satzakzentuierter Position) den so genannten 'harten Vokaleinsatz'. Dieser harte Einsatz kann entweder durch einen dem Vokal voran gestellten Glottalverschluss oder durch Glottalisierung des Vokalonsets realisiert werden. Es ist also möglich, dass die Knarrstimme am Ende syntaktischer Abbrüche bereits einen nachfolgenden harten Vokaleinsatz ankündigt, die eigentliche Realisierung des geplanten Wortes aber ausbleibt. Diesem Problem wird in der Analyse dadurch begegnet, dass alle syntaktischen Abbrüche, denen ein Morphem mit Vokal im Onset folgt, der Restkategorie zugeordnet wird und nicht mit in die eigentliche Auswertung eingeht. Für die verbleibenden Fälle kann

aber nicht ausgeschlossen werden, dass ein Sprecher zum Zeitpunkt des Abbruchs bereits eine Äußerungsfortführung geplant hatte, die mit einem Vokal beginnt und die sich bereits am Ende des syntaktischen Fragments durch Glottalisierung ankündigt, der Sprecher dann aber diese Fortführung verwirft und die Äußerung mit einem Wort ohne Vokal im Onset fortsetzt. Fälle dieser Art auszuschließen, ist grundsätzlich nicht möglich, da Vorhersagen über eine geplante und wieder verworfene Struktur nicht aus den Daten ableitbar sind.

### **6.1.5 Zusammenfassung**

Die phonetischen Repräsentationen unflüssiger Sprechweise sind sehr variabel. Zwei typische Ausprägungen können anhand der Daten belegt werden, zum einen prosodische Grenzen mit vorangehender Zögerungslängung und zum anderen prosodische Grenzen ohne Zögerungslängung aber mit nachfolgender Häsitationspartikel. Für beide Klassen gilt, dass der melodische Verlauf vor der prosodischen Grenze meist leicht fallend oder eben verläuft. Möglicherweise sind die beiden Ausprägungen sprechertypisch. Allerdings sind auch Kombinationen von Zögerungslängung und Häsitationspartikeln zu beobachten. Beide Strategien führen dazu, dass über die gedehnten Segmente Zeit für den Planungsvorgang gewonnen wird und dem Hörer zugleich signalisiert wird, dass der Sprecher beabsichtigt weiterzusprechen. Auch Knarrstimme (z.T. mit Glottalverschluss) tritt in Verbindung mit syntaktischen Abbrüchen auf. Ein durchgehender systematischer Einsatz dieses phonetischen Parameters zeigt sich jedoch nicht.

## **6.2 Perzeptionsexperimente zur Sprechflüssigkeit (Experiment 3a und 3b)**

Die folgenden Experimente zeigen den Zusammenhang zwischen der phonetischen Struktur von Äußerungen und der Wahrnehmung von unflüssiger Sprechweise. Das experimentelle Vorgehen ist vergleichbar mit den Experimenten zu den flüssigen turninternen prosodischen Grenzen und den turnfinalen Grenzen. Die im Korpus beobachtete phonetische Variation wird systematisch in Stimulusserien überführt, und Versuchspersonen beurteilen die Stimuli auf Antwortskalen. In den Experimenten zur unflüssigen Sprechweise wird eine Skala zwischen den Polen 'zögerlich' und 'nicht zögerlich' verwendet. Das erste Experiment überprüft die perzeptorischen Einflüsse der segmentellen Dauerstruktur und des melo-

dischen Verlaufs auf die Wahrnehmung von unflüssiger Sprechweise. Im zweiten Experiment wird zusätzlich die Wechselwirkung zwischen syntaktischer und prosodischer Struktur bei der Wahrnehmung von Sprechflussstörungen untersucht.

## **6.2.1 Experiment zur Wahrnehmung von unflüssiger Sprechweise (Experiment 3a)**

### **6.2.1.1 Hypothesen**

Folgende Hypothesen werden geprüft:

#### **Hypothese 1**

Längung des phrasenfinalen Reims führt zur Wahrnehmung von unflüssiger Sprechweise, wenn sie einen Schwellenwert überschreitet. Wird die Längung oberhalb des Schwellenwertes noch vergrößert, verstärkt sich die wahrgenommene Unflüssigkeit weiter.

#### **Hypothese 2**

Dehnt sich die phrasenfinale Längung über einen größeren Bereich als den phrasenfinalen Reim aus, führt dies bei einem vergleichbaren Längungsfaktor stärker zur Wahrnehmung von unflüssiger Sprechweise.

#### **Hypothese 3**

Werden in einer flüssig gesprochenen prosodischen Phrase einzelne Segmente im Phraseninneren gelängt, führt dies zur Wahrnehmung von unflüssiger Sprechweise.

#### **Hypothese 4**

Starke melodische Bewegungen dämpfen den Eindruck von Unflüssigkeit ab. D.h., dass bei ansonsten gleicher Merkmalskombination die ebenen, leicht fallenden und leicht steigenden Muster eher mit Unflüssigkeit assoziiert werden als die stark fallenden oder stark steigenden.

#### **Hypothese 5**

Glottalverschlüsse am Ende der Äußerung verstärken die Wahrnehmung unflüssiger Sprechweise.

### 6.2.1.2 Stimulusmaterial

Die verwendeten Stimuli sind identisch mit den Stimuli aus dem Experiment zur Turnübergabeerwartung (siehe 5.2.2.2). Lediglich eine zusätzliche Serie wird in den Versuch aufgenommen, in der einzelne Segmente im Inneren der Äußerung gelängt werden. Es folgt eine kurze Zusammenfassung.

Der Trägersatz ist: *Das ist ja nun nicht so angenehm.*

Folgende melodische Verläufe werden äußerungsfinal resynthetisiert:

- stark fallend            -8 Ht (von 111 auf 70 Hz fallend)
- leicht fallend         -4 Ht (von 111 auf 88 Hz fallend )
- eben                     -0 Ht (gleichbleibend bei 111 Hz)
- leicht steigend        +4 Ht (von 111 auf 140 Hz steigend)
- stark steigend         +8 Ht (von 111 auf 176 Hz steigend)

Ausgehend von einem Stimulus ohne finale Längung werden folgende Längungsfaktoren auf den finalen Reim von *angenehm* angewendet.

- Faktor 1,0 (Reimdauer: 250 ms)
- Faktor 1,4 (Reimdauer: 350 ms)
- Faktor 1,8 (Reimdauer: 450 ms)
- Faktor 2,2 (Reimdauer: 550 ms)

Folgende Längungsfaktoren werden auf das letzte Wort der Äußerung (*angenehm*) angewendet:

- Faktor 1,0 (Wortdauer: 515 ms)
- Faktor 1,4 (Wortdauer: 720 ms)
- Faktor 1,8 (Wortdauer: 930 ms)
- Faktor 2,2 (Wortdauer: 1130 ms)

Weiterhin wird eine Serie mit finaler Glottalisierung, die in einem Glottalverschluss endet einbezogen (mit Längungsfaktor 1,4 und allen melodischen Mustern) und ein Stimulus mit stark fallender Intonation und Längungsfaktor 1,4, der in Laryngalisierung ausläuft.

Eine weitere Serie kommt hinzu, indem der Nasal in *nicht* mit Faktor 2 bzw. Faktor 3 gelängt wird. Diese Längung wird auf die Stimuli mit

leicht gelängten finalen Reimen (Faktor 1,4) angewendet, wobei alle 5 finalen Konturen einbezogen werden. So entstehen weitere 10 Stimuli. Diese Serie wird in das Experiment aufgenommen, da vergleichbare Fälle im Korpus oft beobachtbar sind. In den meisten dieser Fälle ist keine prosodische Grenze etikettiert und auch in den resynthetisierten Stimuli entsteht nicht der Eindruck eines deutlichen prosodischen Einschnitts.

Das Stimulusmaterial im Überblick:

- Basisstimuli: vier Längungsfaktoren x fünf Konturen = zwanzig Stimuli
- Ausgedehnte Längung (über das ganze letzte Wort): drei Längungsfaktoren x fünf Konturen = fünfzehn Stimuli.
- Glottalisierung: ein Längungsfaktor x fünf Konturen = fünf Stimuli
- Laryngalisierung: ein Längungsfaktor x eine Kontur = ein Stimulus
- Wortinitiale phraseninterne Längung: ein Längungsfaktor x fünf Konturen x zwei internen Längungsfaktoren = zehn Stimuli
- Gesamt:  $20 + 15 + 5 + 1 + 10 = 51$  Stimuli

### **6.2.1.3 Durchführung von Experiment 3a**

Die Stimuli werden mit dreifacher Wiederholung randomisiert und zu einer Audiodatei zusammengeschnitten. Zwischen den Stimuli ist immer eine Pause von vier Sekunden. Nach jeweils 10 Stimuli ist eine längere Pause und ein Piepton. In den Pausen zwischen den Stimuli geben die Versuchspersonen ihre Urteile ab.

An dem Experiment nahmen 14 Personen teil. Die Durchführung wurde mit der Durchführung von Experiment 2 kombiniert. Zu Beginn der Sitzung werden vorher aufgenommene Instruktionen abgespielt, die den Ablauf und die Aufgabenstellung erläutern (siehe Anhang).

Hinsichtlich der Sprechflüssigkeit und ihrer Beurteilung ist m.E. von einem Kontinuum von flüssiger zu unflüssiger Sprache auszugehen. Diese Sichtweise beruht auf der Analyse der bereits untersuchten Produktionsdaten, die zeigt, dass prosodische Grenzen in unflüssiger Sprache

zum einen nur sehr geringe phonetische Anzeichen von Unflüssigkeit zeigen, zum anderen aber auch Einschnitte vorkommen, die extreme Anzeichen von Unflüssigkeit aufweisen. Auch aus der Sicht des Sprechers können Produktionsprobleme als graduell abgestuft betrachtet werden. Die Suche nach einem Wort oder nach einer syntaktischen Konstruktion kann mehr oder weniger lange dauern und dadurch mehr oder weniger starke Unflüssigkeit in der Rede bedingen. Auch gibt es Sprecher, die generell den Eindruck einer stark unflüssigen Sprechweise hinterlassen, Sprecher, die durchgehend sehr flüssig reden, und Sprecher, deren Sprechweise zwischen diesen Polen liegt. So ist anzunehmen, dass diese Unterschiede von Hörern rezipiert werden und die Interpretation hinsichtlich der Sprechflüssigkeit sich daran orientiert, welche und wieviele Merkmale unflüssiger Sprechweise eine Äußerung enthält und wie stark die einzelnen Merkmale sich ausprägen. Aus diesem Grund wird auch bei der experimentellen Untersuchung wieder mit einer Skala zwischen flüssiger und unflüssiger Sprechweise gearbeitet, die diese graduellen Abstufungen erfassen kann und die folgendermaßen aufgebaut ist:

**Hört sich die Äußerung zögerlich an?**

Ja						Nein
<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>

### 6.2.1.4 Ergebnisse aus Experiment 3a

Tabelle 34 fasst die Ergebnisse der Varianzanalyse zusammen. In einigen Fällen sind in Abhängigkeit von der jeweiligen Fragestellung weitere Einzelkontraste zwischen bestimmten Stimuli berechnet worden. Diese Ergebnisse sind in den nachfolgenden Text integriert.

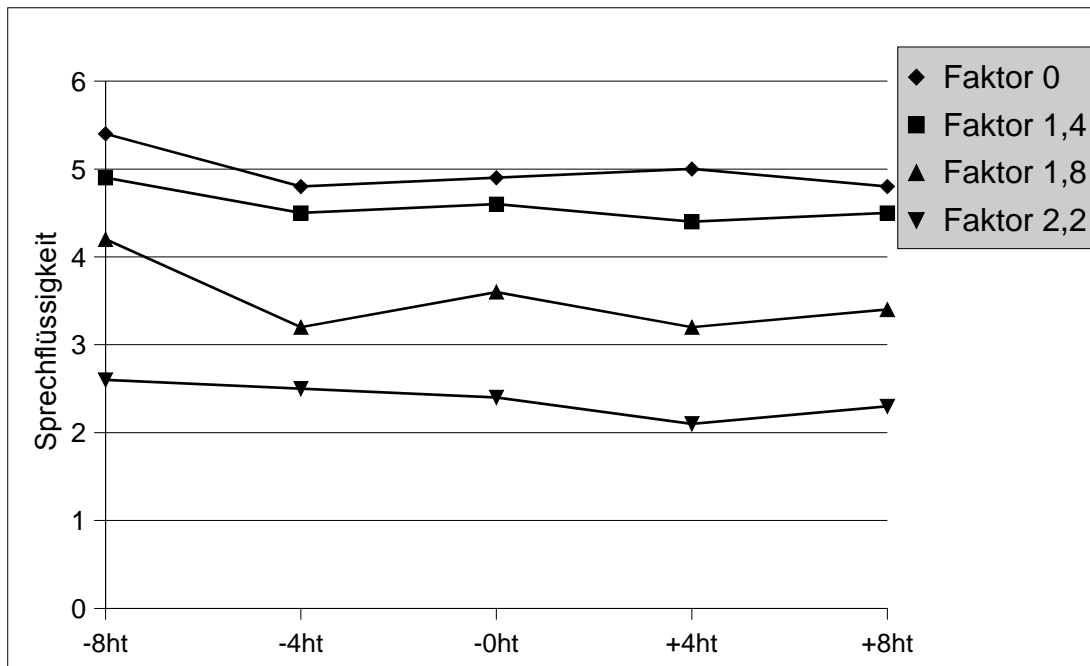
**Tabelle 34:** Ergebnisse der Varianzanalyse von Experiment 3a.

<i>Faktor / Interaktion</i>	<i>df</i>	<i>F</i>	<i>p</i>
Längung des finalen Reims	1,39	100,41	0,000
Längung des finalen Wortes	1,12	112,13	0,000
Phraseninterne Längung	1,67	86,5	0,000
Finale Kontur (über gelängtem Reim)	1,7	4,9	0,022
Finale Kontur (über gelängtem Wort)	2,61	0,655	0,56
Phonation	1	1,5	0,242
Längung (Reim) * Kontur (Reim)	5,89	0,8	0,571
Längung (Wort) * Kontur (Wort)	3,38	1,83	0,149
Phonation * Kontur (Reim)	2,7	1,35	0,276

Die in 6.2.1.1 formulierten fünf Hypothesen werden anhand deskriptiver und prüfstatischer Auswertungen der Hörerurteile im Folgenden überprüft. Wie bereits in den anderen Experimenten dieser Untersuchung stützt sich die Prüfung der Hypothesen vor allem auf direkte Vergleiche der Mittelwerte über die 14 Versuchspersonen für die unterschiedlichen Stimuli und Stimulusserien. Die Werte auf den y-Achsen der Grafiken sind folgendermaßen zu interpretieren: Ein Wert von 6 bedeutet den geringsten Grad von Zögerlichkeit, ein Wert von 0 maximale Zögerlichkeit.

#### **Hypothese 1:**

Längung des phrasenfinalen Reims führt zur Wahrnehmung von unflüssiger Sprechweise, wenn sie einen Schwellenwert überschreitet. Wird die Längung oberhalb des Schwellenwertes noch vergrößert, verstärkt sich die wahrgenommene Unflüssigkeit weiter.



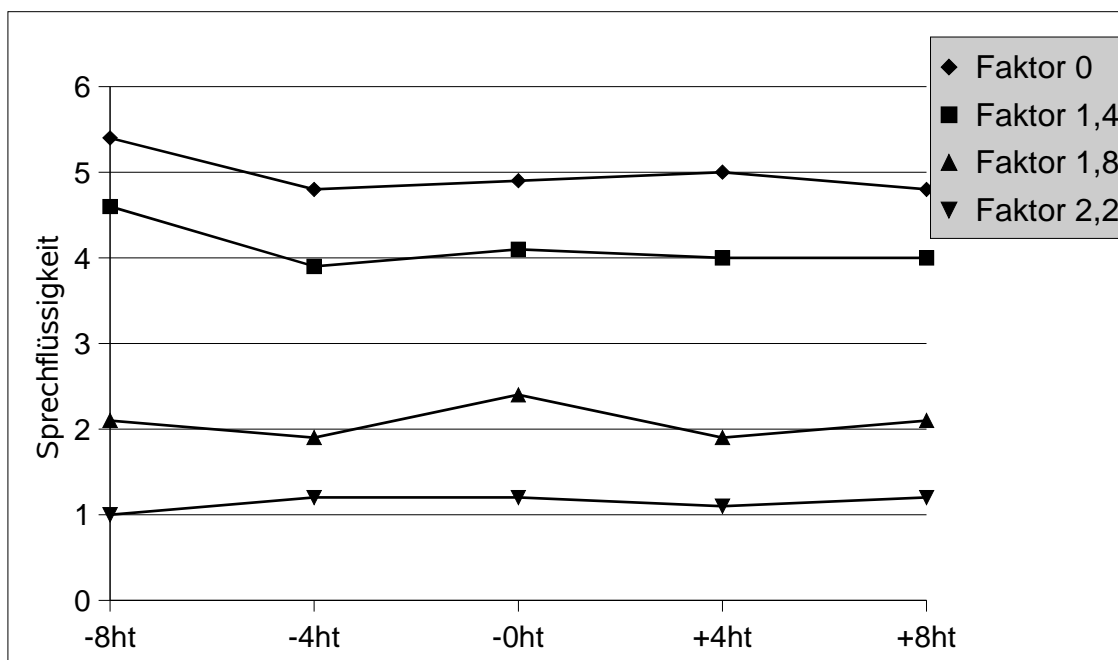
**Abbildung 12:** Wahrgenommene Sprechflüssigkeit für die Stimulusserie mit vier unterschiedlich stark gelängten Reimen und fünf melodischen Mustern am Phrasenende.

Die Dauer des finalen Reims ist eindeutig an der Wahrnehmung unflüssiger prosodischer Grenzen beteiligt (hochsignifikant,  $F=100,41$ ,  $p=0,000$ ; siehe Tabelle 34). Abbildung 12 zeigt, dass die Stimuli mit der Längungsstufe 1 durchgehend als flüssig beurteilt werden. Ein Erreichen des Maximalwerts von 6 ist in Experimenten mit Antwortskalen nicht zu erwarten, da einige Versuchspersonen sich anscheinend scheuen die äußersten Ränder der Skala in ihre Antworten einzubeziehen. So bleiben die Mittelwerte solcher Experimente fast immer unter den Maximalwerten. Die Serie der Längungsstufe 1,4 liegt leicht unterhalb der ungelängten Serie, ist aber noch immer als durchgehend flüssig einzustufen. Der Längungsfaktor von 1,4 entspricht der im Korpus gemessenen durchschnittlichen Längung an turninternen Grenzen ohne syntaktische Abbrüche, Häsitationspartikeln oder Zögerungslängung. Eine weitere Vergrößerung der Reimdauer um 0,4 auf Faktor 1,8 bringt einen deutlichen Sprung der Urteile in Richtung Zögerlichkeit. Die Urteile sind im Mittel etwas oberhalb der Skalenmitte angesiedelt, bewegen sich also klar von der Wahrnehmung als 'nicht-zögerlich' weg. Die noch stärkere Dehnung des finalen Reims (Faktor 2,2) bringt erneut eine deutliche Zunahme der Urteile in Richtung der zögerlichen Sprechweise.

Die Ergebnisse sprechen nicht eindeutig für einen Schwellenwert bei der Wahrnehmung von Unflüssigkeit. Sie belegen eher eine kontinuierliche Zunahme der wahrgenommenen Unflüssigkeit mit zunehmender Reimdauer. Allerdings liegen die Urteile für die Serie der Längungsfaktoren 1 und 1,4 sehr dicht zusammen, während die weiteren hierzu linearen Vergrößerungen der Reimdauer auf die Faktoren 1,8 und 2,2 deutlich stärkere Änderungen der Urteile verursachen. Dieses Urteilsverhalten gibt einen Hinweis auf einen Schwellenwert bei der Wahrnehmung von Unflüssigkeit. Dies wird durch die Berechnung von Einzelkontrasten deutlich: Der Vergleich zwischen Stimuli mit Längungen der Faktoren 1 und 1,4 führt zu einem F-Wert von 12,6 und ist damit hochsignifikant ( $p=0,004$ ). Der Vergleich zwischen Längungsfaktor 1,4 und 1,8 ist ebenfalls hochsignifikant, aber mit einem wesentlich höheren F-Wert von 68 ( $p=0,000$ ).

### Hypothese 2:

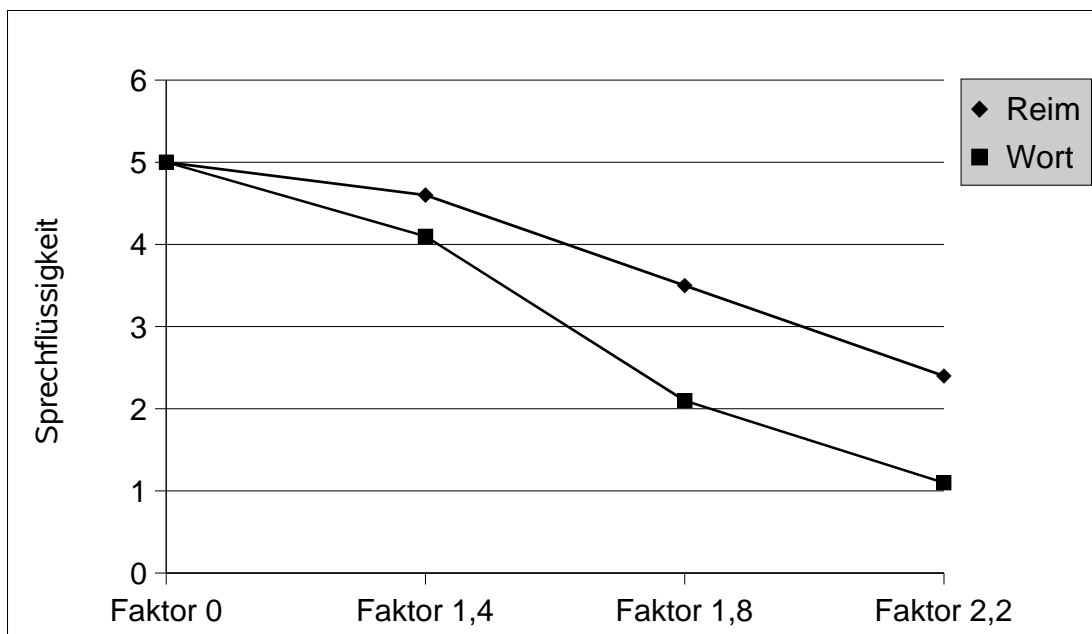
Dehnt sich die phrasenfinale Längung über einen größeren Bereich als den phrasenfinalen Reim aus, führt dies bei einem vergleichbaren Längungsfaktor stärker zur Wahrnehmung von unflüssiger Sprechweise.



**Abbildung 13:** Wahrgenommene Sprechflüssigkeit für die Stimulusserie mit vier unterschiedlich starken Längungen des letzten Wortes des Stimulus (drei Silben) und fünf melodischen Mustern am Phrasenende.

Im Vergleich zwischen Abbildung 12 und Abbildung 13 zeigt sich deutlich, dass die Wahrnehmung von Zögerlichkeit ansteigt, wenn sich die Längung am Ende der Äußerung statt über den finalen Reim über das gesamte letzte Wort ausdehnt.

Abbildung 14 stellt diesen Unterschied heraus, indem, gemittelt über alle melodischen Muster, die Serien mit Längung über dem finalen Reim den Serien mit Längung über dem gesamten finalen Wort gegenübergestellt werden. Die Unterschiede sind hochsignifikant ( $F=22,5$ ,  $p=0,000$ ).

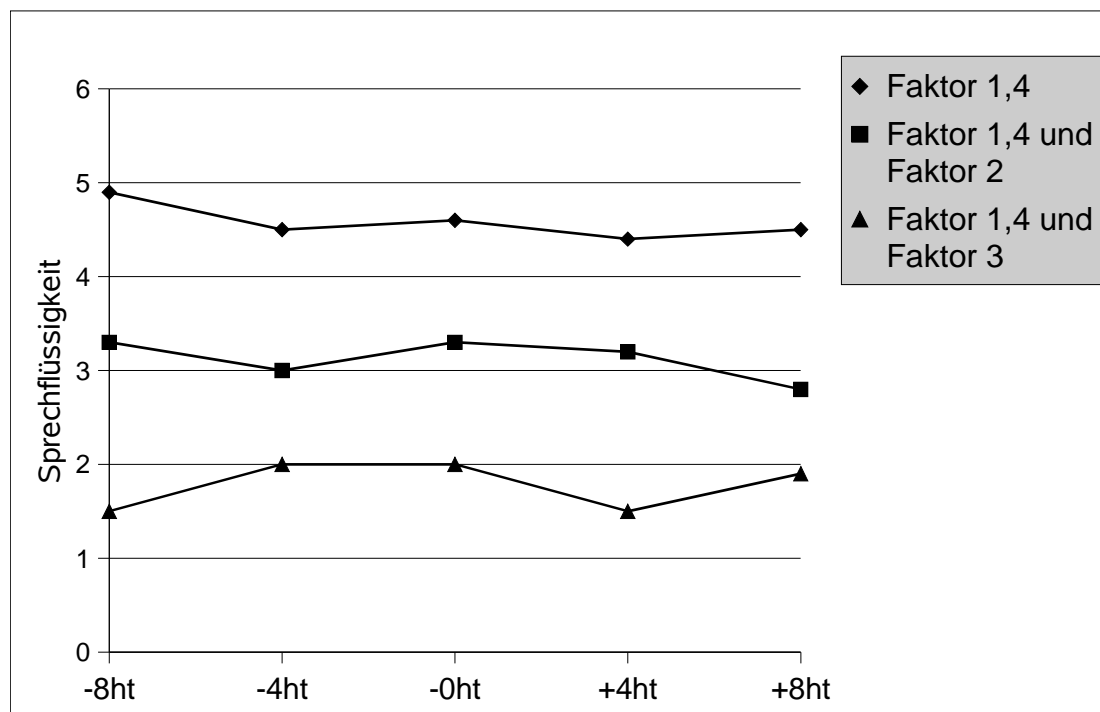


**Abbildung 14:** Wahrgenommene Sprechflüssigkeit im Vergleich zwischen der Stimulusserie mit gelängtem finalen Reim gegenüber der Serie mit Längung des letzten Wortes. Gemittelt über alle melodischen Muster.

Diese Ergebnisse lassen die Schlussfolgerung zu, dass für flüssige prosodische Grenzen die Domäne der finalen Längung auf den letzten Reim vor der Phrasengrenze beschränkt ist und sich der Grad der finalen Längung nicht über eine etwa eineinhalbfache Dauervergrößerung ausprägt. Wenn von diesen Werten deutlich abgewichen wird, kommt es zur Wahrnehmung einer zögerlichen Sprechweise.

### Hypothese 3:

Werden in einer flüssig gesprochenen prosodischen Phrase einzelne Segmente im Phraseninneren gelängt, führt dies zur Wahrnehmung von unflüssiger Sprechweise.



**Abbildung 15:** Wahrgenommene Sprechflüssigkeit im Vergleich zwischen einer Stimulusserie mit leichter Längung (Faktor 1,4) des finalen Reims und zwei Serien mit zusätzlicher Längung des Nasals in *nicht* um Faktor 2 bzw. Faktor 3.

Abbildung 15 zeigt, dass bei gleicher finaler Längung die Dauervergrößerung einzelner Segmente im Äußerungsinneren zu einer ansteigenden Wahrnehmung als unflüssig führt (hochsignifikant,  $F=86$ ,  $p=0,000$ ; siehe Tabelle 34). Die Signalisierung unflüssiger Sprechweise ist also keineswegs auf die Positionen prosodischer Grenzen beschränkt.

#### Hypothese 4:

Starke melodische Bewegungen dämpfen den Eindruck von Unflüssigkeit ab. D.h., dass bei ansonsten gleicher Merkmalskombination die ebenen, leicht fallenden und leicht steigenden Muster eher mit Unflüssigkeit assoziiert werden als die stark fallenden oder stark steigenden.

Die stark fallenden Konturen werden bei gleichen Dauerverhältnissen als flüssiger beurteilt. Für die stärkste Längungsstufe ist der Effekt nicht mehr zu beobachten. Die stark steigenden Konturen zeigen keinen Einfluss auf die Urteile. Generell zeigt sich, dass der Einfluss des Parameters 'Kontur' auf die wahrgenommene Sprechflüssigkeit gering ist (siehe Ab-

bildung 13 und 15). Der dämpfenden Einfluss der stark fallenden Kontur ist allerdings hochsignifikant stärker als der der leicht fallenden Kontur (und so auch allen anderen Konturen). Der Einzelkontrast ergibt hier Werte von  $F=21,1$  und  $p=0,000$ . Dieser dämpfende Effekt der stark fallenden Kontur ist somit dafür verantwortlich, dass der Einfluss der Variable 'Kontur' insgesamt signifikant wird ( $F=4,9$ ,  $p=0,022$ ; siehe Tabelle 34).

### **Hypothese 5:**

Glottalverschlüsse am Ende der Äußerung verstärken die Wahrnehmung unflüssiger Sprechweise.

Die Hypothese wird verworfen, da sich kein signifikanter Einfluss der Variable 'Glottalverschluss' auf die Urteile zeigt (siehe Tabelle 34). Auch dieses Ergebnis belegt, wie schon das Experiment 2 (siehe 5.2) die geringe perzeptorische Relevanz von Glottalverschlüssen an prosodischen Grenzen.

### **6.2.1.5 Zusammenfassung und Diskussion zu Experiment 3a**

Die Versuchspersonen reagieren fein auf Veränderungen in der Dauerstruktur und bringen perzeptorisch auffällige segmentelle Längung mit Störungen der Sprechflüssigkeit in Verbindung. Ab Längungen von Faktor 1,8 des finalen Reims tritt eine deutliche Beurteilung der Stimuli als unflüssig auf. Wird die Längung über einen größeren Bereich als den finalen Reim ausgedehnt (hier die letzten drei Silben) verstärkt sich die wahrgenommene Unflüssigkeit ab Längungsfaktor 1,8 gegenüber den Stimuli, in denen nur der Reim gelängt wurde. Eine leichte Längung im Bereich von Faktor 1,4 führt zu einer schwachen Anhebung der Unflüssigkeitsurteile gegenüber den ungelängten Varianten. Die Urteile belegen für die leichte finale Längung aber insgesamt die Wahrnehmung einer flüssigen Sprechweise. Die Experimente legen nahe, dass als Domäne der flüssigen finalen Längung der finale Reim vor der prosodischen Grenze anzusehen ist. Eine Verlagerung der gelängten Segmente weg vom finalen Reim deuten auf eine unflüssige Sprechweise hin. Diese Schlussfolgerung bezieht allerdings nicht die Längungserscheinungen ein, die für satzakzentuierte Silben beobachtbar sind. Hierzu müssten separate Versuche durchgeführt werden.

Die melodischen Muster am Phrasenende spielen in der Einordnung als 'zögerlich' vs. 'nicht-zögerlich' kaum eine Rolle. Lediglich das stark

fallende Muster führt in einigen Fällen zur Beurteilung als weniger zögerlich als temporal identische Stimuli mit anderen finalen Melodiemustern.

## **6.2.2 Experiment zur Wahrnehmung unflüssiger Sprechweise an unterschiedlichen syntaktischen Strukturstellen (Experiment 3b)**

In den Produktionsdaten kommen unflüssige prosodische Grenzen oft vor, ohne dass sie mit hohen syntaktischen Grenzen zusammenfallen. Die folgende experimentelle Untersuchung soll zeigen, inwieweit die Position einer prosodischen Grenze in Relation zur syntaktischen Struktur die Wahrnehmung als flüssige bzw. unflüssige Grenze beeinflusst. Hierfür werden Stimulusserien generiert, die prosodische Merkmalsbündel verschiedener Art an unterschiedlichen syntaktischen Strukturstellen enthalten. Die Versuchspersonen beurteilen dann, wie im vorangegangenen Experiment, den Grad der wahrgenommenen Unflüssigkeit (siehe 6.2.1.3).

### **6.2.2.1 Hypothesen**

Folgende Hypothese wird geprüft:

Gleiche Bündelungen phonetischer Merkmale erzeugen an niedrigen syntaktischen Grenzen stärker den Eindruck von Unflüssigkeit als an hohen syntaktischen Grenzen.

### **6.2.2.2 Stimulusmaterial**

Das verwendete Stimulusmaterial ist identisch mit dem Material aus Experiment 1c (siehe 4.2.2.2).

### **6.2.2.3 Durchführung von Experiment 3b**

Die Durchführung ist identisch mit Experiment 3a zur wahrgenommenen Zögerlichkeit (siehe 6.2.1). An dem Experiment nehmen 17 Versuchspersonen teil.

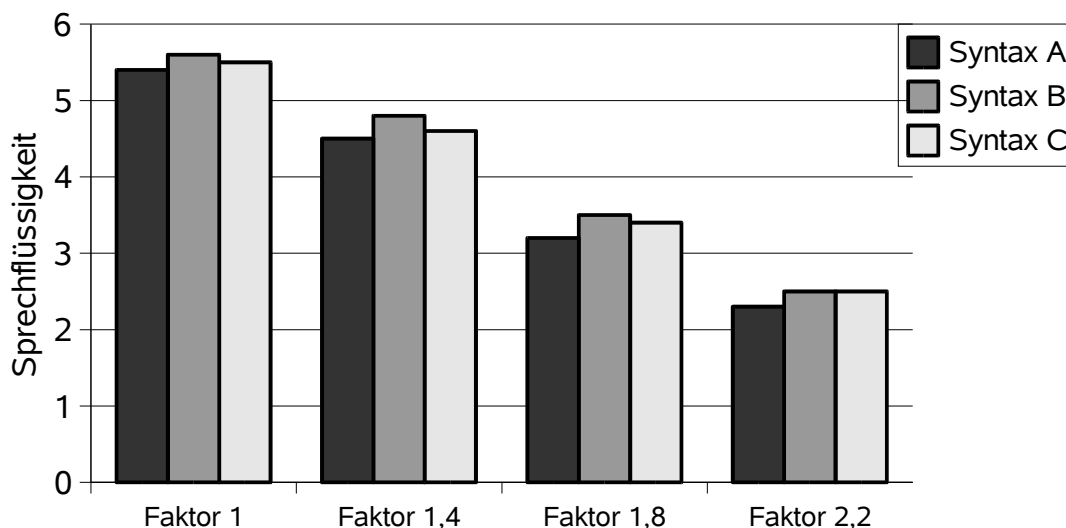
### 6.2.2.4 Ergebnisse aus Experiment 3b

Tabelle 35 fasst die Ergebnisse der Varianzanalyse zusammen. In einigen Fällen sind in Abhängigkeit von der jeweiligen Fragestellung weitere Einzelkontraste zwischen bestimmten Stimuli berechnet worden. Diese Ergebnisse sind in den nachfolgenden Text integriert.

**Tabelle 35:** Ergebnisse der Varianzanalyse von Experiment 3b.

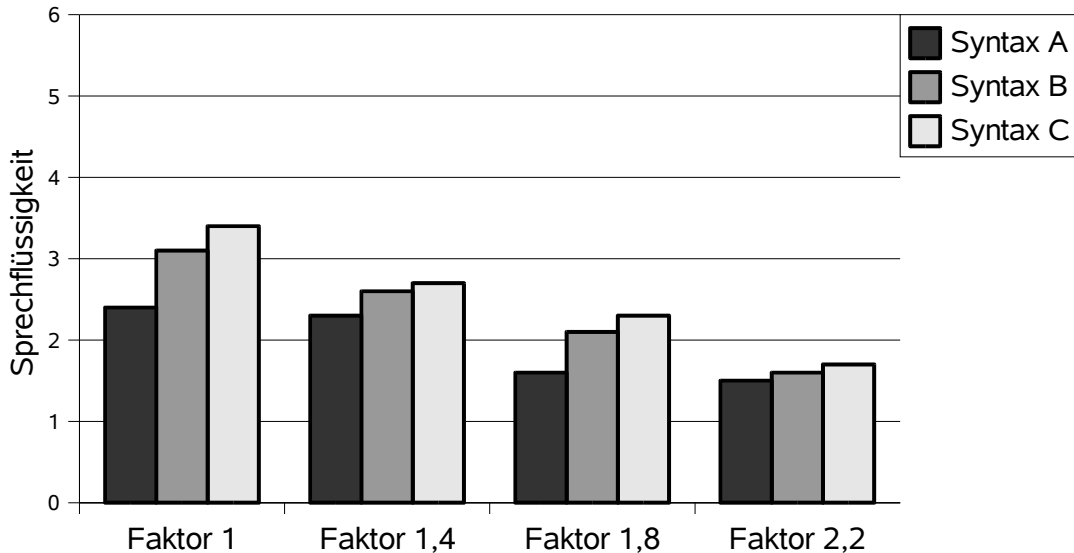
<i>Faktor / Interaktion</i>	<i>df</i>	<i>F</i>	<i>p</i>
Syntax	1,76	5,2	0,015
Längung des finalen Reims	1,59	73,62	0,000
Finale Kontur	1,88	5,02	0,015
Pause	1	26,95	0,000
Syntax * Längung	4,14	0,91	0,468
Syntax * Kontur	3,68	0,69	0,594
Syntax * Pause	1,7	2,61	0,1
Längung * Kontur	3,56	2,74	0,043
Längung * Pause	2,22	24,78	0,000
Kontur * Pause	2	0,2	0,822
Syntax * Längung * Pause	4,46	0,59	0,69
Syntax * Längung * Kontur	7,43	1	0,435
Syntax * Pause * Kontur	2,78	0,42	0,724
Längung * Pause * Kontur	4,81	1,84	0,117
Syntax * Pause * Kontur * Längung	6,25	0,93	0,481

Wie bereits im vorangegangenen Experiment, hat auch in diesem Experiment der melodische Verlauf an den prosodischen Grenzen nur schwache Auswirkungen auf die wahrgenommene Sprechflüssigkeit, obwohl er statistisch auch hier signifikant wird. Aus diesem Grund wird hier nicht weiter auf die melodischen Verläufe eingegangen. In den Abbildungen 16 und 17 sind Mittelwerte über alle melodischen Muster gebildet worden, um zu einer breiteren Datenbasis für die Aussagen über die Auswirkung von Pausen und finaler Längung auf die Urteile der Versuchspersonen zu kommen.



**Abbildung 16:** Wahrgenommene Sprechflüssigkeit für vier Längungsbedingungen unter drei syntaktischen Bedingungen. Stimuli ohne Pausen. Die melodischen Muster sind zusammengefasst.

Abbildung 16 zeigt eine deutliche Auswirkung der Längung des finalen Reims auf die wahrgenommene Sprechflüssigkeit (hochsignifikant,  $F=73,62$ ,  $p=0,000$ ; siehe Tabelle 35). Die Ergebnisse sind denen aus Experiment 3a ähnlich. Die Auswirkung der syntaktischen Strukturstelle auf die Urteile sind relativ gering, aber trotzdem signifikant ( $F=5,2$ ,  $p=0,015$ ; siehe Tabelle 35). Es zeigt sich eine Wahrnehmung stärkerer Unflüssigkeit unter der ersten syntaktischen Bedingung: Wenn also die prosodische Grenze innerhalb einer syntaktischen Einheit vorkommt steigen bei gleichen Merkmalen die Urteile für eine zögerliche Sprechweise leicht an. Die Berechnung von Einzelkontrasten zeigt hochsignifikante Unterschiede zwischen Syntax A und B ( $F=11,7$ ,  $p=0,003$ ) und Syntax A und C ( $F=9,5$ ,  $p=0,009$ ). Syntax B vs. C ergibt kein signifikantes Ergebnis ( $F=0,008$ ,  $p=0,928$ ), (siehe Tabelle 35).



**Abbildung 17:** Wahrgenommene Sprechflüssigkeit für vier Längungsbedingungen unter drei syntaktischen Bedingungen. Stimuli mit Pausen. Die melodischen Muster sind zusammengefasst.

Abbildung 17 zeigt im Vergleich zu Abbildung 16, dass Pausen den Eindruck einer unflüssigen Sprechweise deutlich verstärken. Bereits für die Längungsfaktoren 1 und 1,4 tendieren die Urteile stark in Richtung unflüssig, wenn eine Pause interveniert. Für die Serien mit Pausen lässt sich auch der vorhergesagte Einfluss der syntaktischen Struktur auf die Urteile durchgehend beobachten. Je schwächer der syntaktische Einschnitt an der prosodischen Grenze ist, desto stärker unflüssig erscheint bei gleicher Merkmalsbündelung die Sprechweise. Die geprüfte Hypothese wird folglich angenommen.

### 6.2.2.5 Zusammenfassung und Diskussion zu Experiment 3b

Es zeigt sich, dass prosodische Grenzen im Inneren syntaktischer Einheiten den Eindruck unflüssiger Sprechweise verstärken. Die Versuchspersonen erwarten also im Rahmen einer flüssigen Sprechweise möglicherweise die Konvergenz syntaktischer und prosodischer Phrasierung. Tritt diese Konvergenz nicht ein, wird die Äußerung als unflüssig eingestuft.

Die starke Auswirkung von Pausen auf die Urteile ist überraschend, vor allem, weil Pausen ein häufiges Korrelat prosodischer Grenzen auch in völlig flüssiger Sprache sind. Eine mögliche Erklärung für den starken Effekt der Pausen ist die kurze Äußerungsdauer, die das Vorkommen ei-

ner inhaltlich strukturierenden internen Pause eher unwahrscheinlich macht und die Interpretation als Häsitationspause nahe legt.

### **6.3 Zusammenfassung und Diskussion zu den unflüssigen prosodischen Grenzen**

Die Faktoren, die die Wahrnehmung von unflüssiger Sprechweise verursachen, sind vielfältig und verwoben. Die Perzeptionsuntersuchungen zeigen, dass vor allem starke segmentelle Längung den Eindruck unflüssiger Sprechweise verursacht, besonders wenn diese innerhalb syntaktischer Einheiten auftritt. Die Analyse der Produktionsdaten belegt zudem einen hohen Anteil von prosodischen Grenzen, an denen Häsitationspartikeln vorkommen. Auch hier kann davon ausgegangen werden, dass dies zur Wahrnehmung von Unflüssigkeit führt. Weiterhin sind einige Fälle im Korpus auffindbar, die Knarrstimme und Glottalverschlüsse in Verbindung mit syntaktischen Abbrüchen aufweisen. Eine perzeptorische Relevanz dieses Phänomens kann allerdings nicht nachgewiesen werden.

An den untersuchten unflüssigen Grenzen liegt der Anteil leicht fallender und ebener F<sub>0</sub>-Muster insgesamt bei etwa 80%. Starke melodische Bewegungen an unflüssigen Grenzen sind relativ selten. Sie treten in Verbindung mit finaler Zögerungslängung in etwa 10% der Fälle auf und in Verbindung mit Häsitationspartikeln in etwa 20% der Fälle. Hieraus ergibt sich die Vermutung, dass der erhöhte Anteil stark fallender Konturen vor Häsitationspartikeln damit zusammenhängt, dass vor der prosodischen Grenze mit fallender Kontur ein inhaltlicher Abschluss erreicht wurde und die Häsitationspartikel signalisiert, dass trotz dieses Abschlusses der Turn gehalten werden soll, aber noch Planungsprobleme hinsichtlich der Struktur des Folgenden bestehen. Insofern sollte eine generelle Trennung angestrebt werden, die zwischen Planungsproblemen unterscheidet, die sich auf eine neue inhaltliche Einheit beziehen, und solchen die innerhalb von inhaltlichen und/oder syntaktischen Einheiten auftreten.

Durch die gefundenen phonetischen Merkmale sind die untersuchten unflüssigen prosodischen Grenzen relativ klar von flüssigen Grenzen getrennt. Allerdings gehen in die korpusbasierte Analyse unflüssiger Grenzen von vornherein nur prosodische Grenzen ein, die entweder Zögerungslängung, Häsitationspartikeln oder syntaktische Abbrüche zeigen. Denn es gibt keine anderen Möglichkeiten, labelbasiert auf unflüssige Grenzen zuzugreifen, als über diese Parameter. Es ist aus diesem Grund wahrscheinlich, dass es viele andere prosodische Grenzen im Korpus

gibt, die durch Verzögerungen im Sprachplanungsprozess entstanden sind, die aber keine der oben genannten Merkmale tragen. Hierfür spricht auch, dass in den Terminabsprachedialogen etwa ein Drittel aller etikettierten Pausen an Positionen vorkommt, an denen kein Interpunktionszeichen vorliegt. Dies deutet darauf hin, dass viele der Pausen nicht mit hohen syntaktischen Grenzen zusammenfallen. Das letzte Perzeptionsexperiment (3b) zeigt, dass Versuchspersonen das Vorkommen prosodischer Grenzen ohne hohe syntaktische Grenzen als Anzeichen von Unflüssigkeit bewerten, also ein Zusammenfallen von prosodischer und syntaktischer Phrasierung erwarten. So liegt nahe, dass es, zumindest bei einigen Sprechern, eine Strategie gibt, Planungsschwierigkeiten nicht durch Zögerungslängung und/oder Häsitationspartikeln zu signalisieren, sondern die entstehenden prosodischen Einschnitte in ihrer phonetischen Form ähnlich den flüssigen Grenzen zu gestalten. Etikettierte Grenzen dieser Art sind aufgrund des Analyseverfahrens zusammen mit den inhaltlich motivierten, gliedernden flüssigen Grenzen in Kapitel 4 untersucht worden. Hier eine Trennung zwischen inhaltlich motivierten und durch Planungsprozesse entstandenen Grenzen herbeizuführen, kann lediglich durch inhaltliche Interpretationen und/oder syntaktische Analysen erreicht werden. Für eine funktionale Zuordnung phonetisch verschiedener Typen unflüssiger Grenzen fehlen bislang ebenfalls klare Hinweise. Auch hierfür können weitere Form-Funktions-Zusammenhänge m.E. am besten durch Interpretationen im Gesamtkontext einer Äußerung herausgearbeitet werden. Es ist also auf dem derzeitigen Wissensstand nicht möglich systematische Form-Funktions-Beziehungen unterschiedlicher unflüssiger prosodischer Grenzen aufzustellen.

## **7 Abschließende Diskussion und Ausblick**

Die Ergebnisse der verschiedenen methodischen Stränge dieser Arbeit zeigen die Komplexität des untersuchten Phänomenbereichs. Es ist gelungen, durch die Kombination von automatischer Merkmalsabfrage mit Perzeptionsexperimenten und interpretativen Analysen Bezüge zwischen phonetischen Merkmalsbündeln an prosodischen Grenzen und inhaltlichen und dialogbezogenen Funktionen herzustellen. Allerdings kann weder eine scharfe Grenze zwischen den flüssigen turninternen und den turnfinalen prosodischen Grenzen, noch zwischen den flüssigen und unflüssigen Grenze gezogen werden, noch können funktionale Klassen innerhalb dieser Bereiche eindeutig mit phonetischen Exponenten assoziiert werden. Folglich mussten viele der Ausgangshypothesen verworfen oder modifiziert werden.

Dass eine 1:1-Zuordnung zwischen kommunikativer Funktion und phonetischer Struktur nicht belegt werden kann, liegt vor allem an der hohen Variabilität in der phonetischen Ausprägung spontaner Sprache. Zum einen verfolgen verschiedene Sprecher unterschiedliche Strategien in der Realisierung prosodischer Grenzen, zum anderen kommt es auch innerhalb der Äußerungen einzelner Sprecher zu erheblicher Variation in der Realisierung prosodischer Grenzen, ohne dass diese Variation immer auch mit funktionalen Unterschieden verbunden ist. Trotzdem – oder gerade deshalb – sind die Ergebnisse der Untersuchungen aussagekräftig, denn sie erfassen aus verschiedenen Blickwinkeln die weitgehend fehlende Kategorialität innerhalb des untersuchten Phänomenbereichs, und gerade hier liegt m.E. eine wichtige Perspektive für die phonetisch orientierte Untersuchung von Gesprächen: Es ist notwendig die phonetische Variation immer im Gesamtkontext einer Gesprächssituation zu beurteilen und in Abhängigkeit vom Gesamtkontext funktionale Strukturen anzusetzen. Dieser Kontext umfasst auf der einen Seite typische linguistische Strukturen wie Morphologie, Syntax und Semantik, auf der anderen Seite aber auch situative, stilistische und emotionale Einflüsse. Dieses Prinzip der kumulativen Evidenz sprachlicher Funktionen als Ausgangspunkt für weitere Untersuchungen zu wählen, erfordert neue methodische Paradigmen und eine weitere Öffnung der Phonetik gegenüber ihren psychologischen und soziologischen Nachbardisziplinen. Nur auf diesem Weg kann ein so komplexer Vorgang wie ein Gespräch beschrieben und erklärt werden, denn alles deutet darauf hin, dass auch Sprecher und Hörer in Unterhaltungen alle verfügbaren Ressourcen nutzen, um kommunikative Funktionen zu übermitteln bzw. zu entschlüsseln. Insofern

wird auch phonetische Variation vom Hörer immer in kontextueller Abhängigkeit interpretiert. Unter diesem Gesichtspunkt trennen die in dieser Arbeit vorgenommenen Systematisierungen unterschiedlicher prosodischer Grenzen zwar verschiedene phonetische Ausprägungen und ordnen sie funktionalen Strukturen zu, die Kategorien sind aber als prototypische Kernbereiche zu verstehen von denen es oft zu Abweichungen kommt. Diese Abweichungen entstehen durch die starke phonetische Variabilität, die durch das Prinzip der kumulativen Signalisierung funktionaler Strukturen möglich wird, ohne dass hierdurch die inhaltliche und interaktionale Struktur eines Gesprächs regellos oder zufällig wird.

Die starke beobachtete phonetische Variation kann möglicherweise auch dadurch erklärt werden, dass sich in einem Gespräch mindestens drei Ebenen, die zu prosodischen Einschnitten führen, überlagern und teilweise untrennbar miteinander vermischen: die inhaltlich gesteuerte prosodische Phrasierung, die Ebene der Dialogsteuerung und die prosodischen Grenzen, die durch Planungsschwierigkeiten in die Rede einfließen. Eine Modellierung des Hörerverhaltens könnte diese drei Ebenen voneinander trennen, indem angenommen wird, dass Hörer bei ihrer Interpretation prosodischer Grenzen drei separate Beurteilungen treffen, die jeweils skaliert vorgenommen werden. Erstens eine inhaltsbezogene Einordnung zwischen Abschluss und Weiterweisung, zweites eine dialogbezogene Einordnung zwischen Abschluss und Weiterweisung und drittens eine Einschätzung hinsichtlich des Grades an Unflüssigkeit, der durch die aktuelle prosodische Grenze zum Ausdruck kommt. Aus dem Zusammenspiel dieser Beurteilungen ergibt sich der Raum für die Hypothesen über den weiteren Verlauf des Gesprächs.

## Literatur

- Anderson, A.H., M. Bader, E.G. Bard, E. Boyle, G.M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C.F. Sotillo, H.S. Thompson und R. Weinert (1991). The HCRC Map Task Corpus. *Language and Speech* 34 (4), 351–366.
- Batliner, A., R. Kompe, A. Kießling, E. Nöth, H. Niemann und U. Kilian (1995). The prosodic marking of phrase boundaries: Expectations and results. In A.J. Rubio Ayoso und J.M. López Soler, (Hrsg.) *Speech Recognition and Coding: New Advances and Trends*, NATO ASI Series F. 147, pp. 325–328. Berlin, Heidelberg, New York: Springer.
- Boersma, P. und D. Weenink. *Praat: doing phonetics by computer*. URL: [www.fon.hum.uva.nl/praat](http://www.fon.hum.uva.nl/praat) .
- Bolinger, D. (1978). Intonation across languages. In J.H. Greenberg (Hrg.) *Universals of Human Language. Vol. 2: Phonology*. pp. 471–524. Stanford, Cal.: Stanford University Press.
- Boomer, D. (1965). Hesitations and grammatical encoding. *Language and Speech* 8, 148–158.
- Boomer, D. und A. Dittmann (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech* (5), 215–220.
- Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communicative functions. *AIPUK* 15. Kiel: IPDS.
- Caspers, J. (2000). Experiments on the meaning of four types of single-accent intonation patterns in Dutch. *Language and Speech* 43(2), 127–161.
- Chen, A.J. (2003). Language Dependence in Continuation Intonation. *Proc. 15th ICPHS, Barcelona*, Volume 1, pp. 1069–1072.
- Couper-Kuhlen, E. (1983). Intonatorische Kohäsion. Eine makroprosodische Untersuchung. *Zeitschrift für Literaturwissenschaft und Linguistik* 49, 74–100.
- Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics* 17, 77–91.

- de Pijper, J.R. und A.A. Sanderman (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *JASA* 96 (4), 2037–2047.
- Delattre, P., E. Poenack und C. Olsen (1965). Some characteristics of German intonation for the expression of continuation and finality. *Phonetica* 13, 134–161.
- Essen, O. v. (1964). *Grundzüge der hochdeutschen Satzintonation*. Ratingen, Düsseldorf: Henn.
- Ford, C.E. und S.A. Thompson (1996). Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff und S. A. Thompson (Hrsg.). *Interaction and Grammar*. Studies in Interactional Sociolinguistics 13, pp. 134–184. Cambridge: Cambridge University Press.
- French, P. und J. Local (1983). Turn-competitive incomings. *Journal of Pragmatics* 7, 701–715.
- Freud, S. (1904/1971). *Zur Psychopathologie des Alltagslebens: Über Vergessen, Versprechen, Vergreifen, Aberglauben und Irrtum*. Frankfurt/M.: Fischer Bücherei.
- Gilles, P. (2005): *Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung*. Linguistik – Impulse und Tendenzen 6. Berlin, New York: de Gruyter.
- Goldman Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London, New York: Academic Press.
- Grabe, E., C. Gussenhoven, J. Haan, E. Marsi, und B. Post (1997). Preaccentual pitch and speaker attitude in Dutch. *Language and Speech* 41(1), 63–85.
- Greenhouse, S.W. und S. Geisser (1959). On methods in the analysis of profile data. *Psychometrika* 32, 95–112.
- Grosjean F. und A. Deschamps (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica* 31, 144–184.
- IPDS (1995). *The Kiel Corpus of Spontaneous Speech*, Volume 1, CD-ROM#2. Kiel: IPDS.

- IPDS (1996). *The Kiel Corpus of Spontaneous Speech*, Volume 2, CD-ROM#3. Kiel: IPDS.
- Klatt, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3, 129–140.
- Kohler, K.J. (1991a). A model of German intonation. In K. J. Kohler (Hg.), *Studies in German Intonation*, AIPUK 25, pp. 295–360, Kiel: IPDS.
- Kohler, K.J. (1991b). Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. In K. J. Kohler (Hg.), *Studies in German Intonation*, AIPUK 25, pp. 115–186.
- Kohler, K.J. (1995). PROLAB – the Kiel system of prosodic labelling. *Proc. 13<sup>th</sup> ICPPhS*, Stockholm, Volume 3, pp. 162–165.
- Kohler, K.J. (2001). Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German? In N. Grønnum und J. Rischel (Hrsg.). *To Honour Eli Fischer-Jørgensen: Festschrift on the Occasion of her 90<sup>th</sup> Birthday, February 11<sup>th</sup>, 2001*, Travaux du Cercle Linguistique de Copenhagen, 31, pp. 174–211, Copenhagen: Reizel.
- Kohler, K.J., B. Peters und T. Wesener (2005). Phonetic exponents of disfluency in German spontaneous speech. In K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, pp. 185–201, Kiel: IPDS.
- Koiso, H., Y. Horiuchi, S. Tutiya, A. Ichikawa und Y. Den (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs. *Language and Speech* 41(3-4), 295–321.
- Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen und S. Nooteboom (Hrsg.), *Structure and Process in Speech Perception*, pp. 195–206. Heidelberg: Springer.
- Liebermann, P. (1967). *Intonation, Perception, and Language*. Research Monograph 38. Cambridge/Mass.: MIT Press.
- Likert, R. (1931). A technique for the measurement of attitudes. *Archives of Psychology*. New York: McGraw-Hill.

- Local, J., W.H.G. Wells und M. Sebba (1985). Phonology for Conversation. Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9, 309–330.
- Local, J. und J. Kelly (1986). Projection and ‘silences’: Notes on phonetic and conversational structure. *Human Studies* 9, 185–204.
- Megyesi, B. und S. Gustafson-Capkova (2002). Production and perception of pauses and their linguistic context in read spontaneous speech in Swedish. *Proc. ICSLP, Denver*, pp. 2153–2156.
- Moulines, E. und F. Carpentier (1990): Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9 (5-6), 453–467.
- Nakatani, C.H. and J. Hirschberg (1994). A corpus-based study of repair cues in spontaneous speech. *JASA* 95(3), 1603–1616.
- Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. *JASA* 54(4), 1235–1247.
- Ostendorf, M., P. Price, J. Bear und C. Wightman (1990). The use of relative duration in syntactic disambiguation. In *Proc. 3<sup>rd</sup> DARPA Workshop on Speech and Natural Language*, pp. 26–31.
- Peters, B. (1999). Prototypische Intonationsmuster in deutscher Lese- und Spontansprache. In K.J. Kohler (Hg.), *Phrase-level Phonetics and Phonology of German*, AIPUK 34, pp. 1–175. Kiel: IPDS.
- Peters, B. (2000). Individuelle und geschlechtsspezifische Unterschiede in der prosodischen Gestaltung deutscher Lese- und Spontansprache. In H.D. Schlosser (Hg.) *Sprache und Kultur*, Forum Angewandte Linguistik 38, pp. 153–162, Frankfurt/M.: Lang.
- Peters, B. (2001). 'Video Task' oder 'Daily Soap Szenario' : Ein neues Verfahren zur kontrollierten Elizitation von Spontansprache. URL: [http://www.ipds.uni-kiel.de/pub\\_exx/bp2001\\_1/Linda21.html](http://www.ipds.uni-kiel.de/pub_exx/bp2001_1/Linda21.html)
- Peters, B. und K.J. Kohler (2004). *Trainingsmaterialien zur prosodischen Etikettierung mit dem Kieler Intonationsmodell KIM*. URL: [www.ipds.uni-kiel.de/kjk/pub\\_exx/bpkk2004\\_1/TrainerA4.pdf](http://www.ipds.uni-kiel.de/kjk/pub_exx/bpkk2004_1/TrainerA4.pdf)

- Peters, B. (2005a). The database – The Kiel Corpus of Spontaneous Speech. In K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, pp. 1–6, Kiel: IPDS.
- Peters, B., K.J. Kohler und T. Wesener (2005a). Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, pp. 143–184, Kiel: IPDS.
- Peters, B., K.J. Kohler, und T. Wesener (2005b). Melodische Satzakkentmuster in prosodischen Phrasen deutscher Spontansprache – Statistische Verteilung und sprachliche Funktion. In K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, pp. 7–54, Kiel: IPDS.
- Pierrehumbert, J. und J. Hirschberg (1990). The meaning of intonational contours in the interpretation of discourse. In J. Cohen, P. Morgan und M. Pollack (Hrsg.), *Intentions in Communication*, pp. 271–311. Cambridge/Mass.: MIT Press.
- Price, P.J., M. Ostendorf, S. Shattuck-Hufnagel und C. Fong (1991). The use of prosody in syntactic disambiguation. *JASA* 90(6), 2956–2970.
- Rietveld, T., J. Haan, L. Heijmans und C. Gussenhoven (2002). Explaining attitudinal ratings of Dutch rising contours: Morphological structure vs. the Frequency Code.” *Phonetica* 59(2-3), 180–194.
- Selting, M. (1995). *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer.
- Shriberg, E., A. Stolcke, D. Hakkani-Tür und G. Tür (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication* 32, 127–154.
- Szczepek, B. (2000). Formal Aspects of collaborative productions in English conversation. *InLiSt* 17.
- t'Hart, J.R., R. Collier, und A. Cohen (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press

- Thurston, L.L. und E.J. Chave (1929) *The Measurement of Attitude*. Chicago: University of Chicago Press.
- Wichmann, A. (2000). *Intonation in Text and Discourse. Beginnings, Middles and Ends*. Studies in Language and Linguistics. Harlow: Pearson Education.
- Wichmann, A. und J. Caspers (2001). Melodic cues to turn-taking in English: evidence from perception. In *Proc. 2<sup>nd</sup> SIGdial Workshop on Discourse and Dialogue, Aalborg*. pp. 180–185.
- Wightman, C.W., S. Shattuck-Hufnagel, M. Ostendorf, P.J. Price, (1992). Segmental duration in the vicinity of prosodic phrase boundaries. *JASA 91 (3)*, 1707 – 1717.
- Zwirner, E. und K. Zwirner (1937). Phonometrischer Beitrag zur Frage der Lesepausen. *Arch. Néerl. Phon. Exp. 13*, pp. 111–128.

## Anhang

Der Anhang enthält die schriftlichen Vorlagen für die Instruktionen der Perzeptionsexperimente. Die Präsentation der Instruktionen vor den Experimenten erfolgte als vorgelesene Version von CD.

### Einleitung Experiment 1a, 1b – Kohäsion an turninternen prosodischen Grenzen

In dem folgenden Hörexperiment geht es um die phonetische Gliederung von Äußerungen. Äußerungen können phonetisch als eine globale Einheit gesprochen werden, wie in dem folgenden Fall:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." ohne interne Pause, ohne finale Längung und mit ebener Kontur.*

Hier werden sie wahrscheinlich alle sagen, dass diese Äußerung phonetisch nicht in verschiedene Teile untergliedert ist. Äußerungen können aber auch so gesprochen werden, dass eine ebensolche Untergliederung phonetisch angezeigt wird, wie in dem folgenden Beispiel:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." mit kurzer interner Pause, leichter finaler Längung und stark fallender Kontur.*

Wenn sie die beiden Beispiele jetzt noch mal hintereinander hören, werden sie im zweiten gegenüber dem ersten eine deutliche Trennung wahrnehmen.

*Es folgen die beiden obigen Audiobeispiele.*

Diese Trennung kann von unterschiedlicher Stärke sein. Bei dem folgenden Beispiel urteilen sie sicher, dass es sich um einen starken Grad der Trennung handelt:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." mit langer interner Pause, starker finaler Längung und stark fallender Kontur.*

Im folgenden Beispiel aber eher um einen schwachen Grad der Trennung:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." mit ohne interne Pause, leichter finaler Längung und leicht fallender Kontur.*

In dem nun folgenden Hörexperiment möchten wir von Ihnen Auskunft darüber haben, wie sie die Stärke der phonetischen Trennung in Äußerungen wahrnehmen. Die Skala dieser Trennung rangiert zwischen dem Fehlen einer Trennung an einem Ende und einer extremen Trennung am anderen Ende. In der Mitte dieser Skala ist ein mittlerer Grad der Trennung. Die gesamte Skala wird eingeteilt in sieben Stufen von 0 (keine Trennung) bis 6 (extremer Trennungsgrad); der Wert drei entspricht einem mittleren Trennungsgrad. Sie werden im Experiment in vielfacher Wiederholung einen Satz hören, der immer die

gleichen Wörter enthält. Er variiert aber in der Aussprache und zwar im Grad der Trennung in zwei Äußerungsteile.

Zum Ablauf des Experiments: Auf dem Ihnen vorliegenden Fragebogen sehen Sie Blöcke von jeweils zehn Zeilen, die in sieben Antwortspalten unterteilt sind. Bitte geben Sie für jede der aufeinander folgenden Äußerungen in der entsprechenden Zeile ein Urteil darüber ab, wie stark Sie die Trennung innerhalb der Äußerung wahrnehmen. Nach jeder vorgespielten Äußerung ist eine Pause, in der Sie Ihr Urteil abgeben können. Bitte kreuzen Sie für jede Äußerung auf der entsprechenden Zeile den Grad der wahrgenommenen Trennung an. Nach jeweils zehn Äußerungen kommt eine längere Pause und ein Piepton. Um Ihnen den Ablauf besser zu verdeutlichen werden jetzt 10 Probebeispiele vorgespielt, die Sie bitte in Block 1 beurteilen. Danach können noch Fragen zum Ablauf gestellt werden.

Vielen Dank im Voraus für Ihre Mitarbeit.

### **Einleitung Experiment 1c – Kohäsion an turninternen prosodischen Grenzen**

Die Einleitung für Experiment 1c ist identisch mit dem obigen Text. Es ist allerdings eine kurze Passage geändert worden, die sich darauf bezieht, dass hier drei unterschiedliche Sätze beurteilt werden sollen.

### **Einleitung Experiment 2 - Turnübergabeerwartung**

Das nun folgende Experiment ist völlig parallel zu dem eben durchgeführten Experiment aufgebaut. Sie bekommen wieder den selben Satz in vielfacher Wiederholung zu hören und sollen wieder Beurteilungen zu dem Gehörten auf Skalen abgeben.

Die Fragestellung hierbei ist aber eine ganz andere als eben. Sie sollen nun eine Einschätzung darüber abgeben, wie sich wohl der Sprecher nach dem geäußerten Satz in einer Gesprächssituation verhalten würde. Ob er noch weitersprechen wird oder ob er seine Äußerung mit diesem Satz beendet und das Wort dem Gesprächspartner überlässt.

Hierfür können Sie sich vorstellen, dass der Satz in einem zwanglosen Gespräch zwischen Freunden geäußert wird. Z.B. als Kommentar zu einer Schilderung eines Missgeschicks oder als Reaktion auf die Beschreibung eines Zahnarztbesuchs. Nach diesem Kommentar - *Das ist ja nun nicht so angenehm* - gibt es für den Sprecher zwei Möglichkeiten: Weitersprechen oder nicht weitersprechen. Das Experiment hat zum Ziel herauszufinden welche phonetischen Eigenschaften von Äußerungen dazu dienen dem Hörer zu signalisieren, ob noch weitergesprochen werden soll oder ob eine Äußerung zum Ende gekommen ist.

Auch in diesem Experiment wird mit Skalen gearbeitet, damit auch Urteile gegeben werden können, die eine gewisse Unsicherheit in der Einschätzung des Gesprächsfortgangs ausdrücken. Wenn Sie also klar der Meinung sind, dass der Sprecher noch weitersprechen wird, kreuzen Sie bitte die 0 an. Wenn Sie denken, dass der Sprecher am Ende seiner Äußerung angekommen ist, kreuzen Sie die 6 an. Die 3 in der Mitte der Skala würde eine völlige Unentschiedenheit bedeuten. Bitte scheuen Sie sich

auch in diesem Experiment nicht davor die Extreme der Skala mit in ihre Beurteilungen einzubeziehen. Das Experiment dauert etwa 15 Minuten.

Nun folgt wieder ein Block mit 10 Wiederholungen, damit Sie sich an die Aufgabenstellung gewöhnen können. Danach können noch Fragen gestellt werden.

### **Einleitung Telexperiment 3a: Sprechflüssigkeit**

Im folgenden Hörexperiment geht es um den Unterschied zwischen einer zögerlichen und einer flüssigen Sprechweise. Jeder kennt es aus dem Alltag, dass man manchmal schnell und sicher die eigenen Gedanken äußern kann und zusammenhängend und flüssig spricht. Manchmal aber fallen einem die richtigen Worte nicht gleich ein und man kommt ins Stocken - so entsteht eine zögerliche Sprechweise.

In dem nun folgenden Hörexperiment geht es darum, dass sie Äußerungen hinsichtlich der Sprechweise beurteilen. Also eine Einschätzung darüber abgeben, ob es sich um eine flüssige oder eher um eine unflüssige Sprechweise handelt. Sie werden im Experiment in vielfacher Wiederholung einen Satz hören, der immer die gleichen Wörter enthält, sich aber in phonetischen Details unterscheidet.

Auf den Antwortbögen vor Ihnen sehen Sie eine Antwortskala mit sieben Stufen. Die Skala rangiert von 0 bis 6. Nehmen Sie bitte für jede Wiederholung des Satzes eine Beurteilung hinsichtlich der Flüssigkeit der Äußerung vor. Wenn Sie den Eindruck einer völlig flüssigen Sprechweise haben kreuzen Sie bitte die 6 an. Wenn der gehörte Satz einen sehr zögerlichen Eindruck macht, kreuzen Sie die 0 an. Viele der Äußerungen werden aber vermutlich zwischen diesen Extremen liegen. Machen sie Ihr Kreuz dann an einer Stelle der Skala, die für Ihr Empfinden diesem mittleren Grad der Flüssigkeit entspricht. Bitte scheuen Sie sich nicht auch die äußeren Bereiche der Skala mit auszunutzen. D.h. wenn Sie den Eindruck einer deutlich zögerlichen Sprechweise haben, kreuzen sie ruhig die 0 am linken Ende der Skala an. Wenn Ihnen die Äußerung völlig flüssig erscheint wählen Sie die 6.

Nach jeder vorgespielten Äußerung ist eine Pause, in der Sie Ihr Urteil abgeben können. Nach jeweils zehn Äußerungen kommt eine längere Pause und ein Piepton. Das ganze Experiment dauert etwa 15 Minuten. Um Ihnen den Ablauf besser zu verdeutlichen werden jetzt 10 Probebeispiele vorgespielt, die Sie bitte im ersten Antwortblock - dem Testblock - beurteilen. Danach können noch Fragen zum Ablauf gestellt werden.

Vielen Dank im Voraus für Ihre Mitarbeit.

### **Einleitung Telexperiment 3a: Sprechflüssigkeit**

Die Einleitung für Experiment 3b ist identisch mit dem obigen Text. Es ist allerdings eine kurze Passage geändert worden, die sich darauf bezieht, dass hier drei unterschiedliche Sätze beurteilt werden sollen.