Dagmar Barth-Weingarten
Beatrice Szczepek Reed (Hgg.)

Prosodie und Phonetik in der Interaktion

Prosody and phonetics in interaction

Verlag für Gesprächsforschung

Mannheim: Verlag für Gesprächsforschung 2014
http://www.verlag-gespraechsforschung.de
ISBN 978 - 3 - 936656 - 60 - 2

Alle Rechte vorbehalten.

© Verlag für Gesprächsforschung, Dr. Martin Hartung, Mannheim 2014

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

Inhalt/ Contents

Dagmar Barth-Weingarten/ Beatrice Szczepek Reed
Prosodie und Phonetik in der Interaktion – *Prosody and phonetics in interaction*: Hinführung, Termini, Methoden…… 4

Stephanie Köser
Phrasen-finale Phonationsänderungen und ihre Rolle beim *turn taking*..... 20

Uwe Küttner
Rhythmic analyses as a proof-procedure? – An initial observation on rhythmicity and projection..... 46

Leendert Plug
On (or not on) the 'upgrading–downgrading continuum': The case of 'prosodic marking' in self-repair 70

Maxi Kupetz
'Mitfühlend sprechen': Zur Rolle der Prosodie in Empathiedarstellungen 87

Jürgen Trouvain/ Khiet P. Truong
Towards unravelling prosodic characteristics of speaker-overlapping laughing in conversational speech corpora 115

Ines Bose/ Stephanie Kurtenbach
Stimmlich-artikulatorischer Ausdruck in der Kita: Erzieherinnen betrachten mit zwei- und vierjährigen Kindern ein Bilderbuch 136

Beatrice Szczepek Reed
Prosodic, lexical and sequential cues for assessments with German *süß*: Assemblages for action and public commitment 162

Beatrix Schönherr
Kooperative Gesten und Prosodie als theatrale Gestaltungsmittel 187

Elizabeth Couper-Kuhlen
Prosody as dialogic interaction 221

Anhang/ *Appendix*: Transkriptionskonventionen/ *transcription conventions*
GAT2 (deutsch) 252
GAT2 (*English*) 255

Prosodie und Phonetik in der Interaktion – *Prosody and phonetics in interaction*: Hinführung, Termini, Methoden

Dagmar Barth-Weingarten und Beatrice Szczepek Reed

1. Hinführung

Wo immer Menschen miteinander sprechen, spielen Prosodie und Phonetik eine Rolle. Wir können nichts sagen, ohne eine bestimmte Sprechmelodie, ohne Lautstärke, ein bestimmtes Sprechtempo und eine bestimmte Artikulationsweise. Selbst Flüstern weist eine, wenn auch geringe, Lautstärke (ca. 30 dB, Laver 1994: 502) auf, es wird schnell(er) oder langsam(er) und mehr oder weniger deutlich geflüstert. Kurz: Prosodisch-phonetische Phänomene sind allgegenwärtig (*co-constitutive*). Allein aus diesem Grund sollten sie Bestandteil der Analyse gesprochener Sprache sein.

Zum zweiten können sie distinktiv (*distinctive*) sein, d.h. prosodische und phonetische Merkmale können Bedeutung, aus interaktionaler Perspektive insbesondere Handlungen, voneinander unterscheiden. Selting (2010: 6) illustriert das mit dem eingängigen Beispiel

```
a. was,
b. was.
c. <<h>!WAS!?>
```

Während das Fragewort in a. mit steigender Sprechmelodie, normaler Lautstärke und Länge zumeist ein Problem im akustischen Verständnis anzeigt und dessen Reparatur initiieren kann, kontextualisiert b. mit denselben Lautstärke- und Tempomerkmalen aber fallender Sprechmelodie ein referentielles Problem. Die Variante c. hingegen – höher, lauter und eventuell auch länger – verweist auf ein Problem bezüglich der Erwartungen der SprecherIn, indem sie Affekte wie Überraschung nahelegt. Die prosodisch-phonetische Realisierung einer Äußerung kann somit grundlegend für ihre Interpretation sein. Auch dies macht ihre systematische Analyse gleichsam unabdingbar für die Untersuchung gesprochener Sprache.

Gleichzeitig ist die Beschreibung des Gebrauchs prosodisch-phonetischer Merkmale in der Interaktion ein noch junges Forschungsgebiet. Zwar wurde ihre Bedeutung in der Konversationsanalyse schon früh erkannt. So verwiesen bereits Sacks/ Schegloff/ Jefferson (1974: 722) auf die Rolle der Prosodie zur Konstitution von Turnkonstruktionseinheiten (*turn-constructional units*, TCU): "any word can be made into a 'one-word' unit-type ... via intonation". Auch würdigte eine Reihe von Arbeiten die Rolle von Prosodie und Phonetik (vgl., z.B., Jefferson 1986, 1988 zu Pausen, Lerner 1996 zur Prosodie in komplexen Äußerungen, Schegloff 1982 zu prosodischen Zögerungssignalen sowie 1997, 1998 zur Rolle von Tonhöhe bei der Organisation des Sprecherwechsels, in der Gesprächseröffnung und in der Handlungskonstitution). Eine systematischere Untersuchung von Prosodie und Phonetik im Gespräch erfolgt jedoch erst seit Mitte der 1980er Jahre. In Auseinandersetzung mit Prämissen und Befunden der klassischen Phonetik und Phonologie legte die Yorker *Phonology/ Phonetics for Conversation* Wert auf

die impressionistische und parametrische Analyse natürlicher interaktionaler Daten. Daraus ist eine Vielzahl von Arbeiten entstanden, sowohl zur Beschreibung prosodisch-phonetischer Phänomene (vgl. z.B. Local/ Kelly/ Wells 1986, Ogden 2006, Curl 2005, Plug 2011) als auch zur Theoriebildung (vgl. z.B. Local/ Walker 2004, 2005, Walker 2014). Prosodie war auch fester Bestandteil von Untersuchungen im Rahmen der Kontextualisierungstheorie (vgl. z.B. Gumperz 1982, Auer/ di Luzio 1992, Couper-Kuhlen 1993, Auer/ Couper-Kuhlen/ Müller 1999). Hier wird Prosodie als einer von einer Vielzahl von Kontextualisierungshinweisen verstanden (vgl. Couper-Kuhlen 1998: 9, Auer/ Di Luzio 1992). Eine Intonationskontur etwa hat keine Bedeutung *per se*, sondern legt vor dem Hintergrund der anderen sprachlichen Dimensionen sowie dem interaktionalen und situativen Kontext eine mögliche Interpretation nahe. Diese Auffassung unterscheidet gesprächsforscherische Ansätze z.B. vom Verständnis von Intonation als Grammatik (vgl. z.B. die *tune*-Analysen von Sweet 1906, Armstrong/ Ward 1926, Pike 1945, Pierrehumbert 1980, Halliday 1994), aber auch von Ansätzen, die die Bedeutung der Intonation für den Informationsfluss betonen (z.B. du Bois et al. 1993, Chafe 1994). Neben der Kontextualisierungstheorie hat sich die Gesprächsforschung in Deutschland mit Prosodie und Phonetik in jüngster Zeit vor allem aus interaktionslinguistischer Perspektive beschäftigt. Das Ziel der Interaktionalen Linguistik (vgl. z.B. Selting/ Couper-Kuhlen 2000), den systematischen Gebrauch sprachlicher Ressourcen zur Erreichung interaktionaler Ziele zu beschreiben, spiegelt sich in der systematischen Untersuchung prosodisch-phonetischer Ressourcen wider. Inzwischen demonstriert eine Vielzahl interaktionslinguistischer Arbeiten (vgl. z.B. die Beiträge in Couper-Kuhlen/ Selting 1996, Selting/ Couper-Kuhlen 2001, Couper-Kuhlen/ Ford 2004, Barth-Weingarten/ Reber/ Selting 2010) überzeugend den Wert und den Fortschritt prosodisch-phonetischer Untersuchungen. Im Ergebnis dessen wenden sich zunehmend auch etablierte Ansätze interaktionslinguistischen Herangehensweisen und Erkenntnissen zu (vgl. z.B. die Beiträge in Barth-Weingarten/ Dehé/ Wichmann 2009, auch Dilley/ Shattuck-Hufnagel/ Ostendorf 1996, Zellers/ Post 2012).

So haben sich Prosodie und Phonetik in der Interaktion zu einem produktiven und international anerkannten Forschungsgebiet entwickelt. Schwerpunkte lagen bisher u.a. auf der Strukturierung gesprochener Sprache, der Redebeitragsorganisation (*turn-holding/ -yielding*), der Konstitution größerer Beiträge (*multi-unit turns*), der Handlungsorganisation, der Kontextualisierung von kommunikativen Gattungen, Stilen und Affektivität, unter Beachtung ausgewählter Aspekte geographischer und soziolektaler Variation, sowie auf Aspekten des kindlichen Spracherwerbs (für einen Überblick vgl. z.B. Selting 2010). Die Ergebnisse der Forschung haben gezeigt: Die Gestaltungsmöglichkeiten prosodisch-phonetischer Ressourcen sind vielfältig, und die Systematizität ihrer Nutzung durch die TeilnehmerInnen ist faszinierend.

Dennoch ist das Potenzial der Untersuchung von Prosodie und Phonetik in der Interaktion bei weitem nicht ausgeschöpft. Dies beginnt bei der Untersuchung einzelner Parameter, wie etwa Stimmqualität (vgl. Ogden 2004), und reicht bis hin zur Bestimmung der Kontextualisierungskraft prosodisch-phonetischer Parametercluster. So ist etwa die Frage nach der Rolle von Prosodie und Phonetik bei der Konstitution von Einheiten nach wie vor aktuell (vgl. z.B. Szczepek Reed/ Raymond 2013), ebenso wie bei der sequentiellen Organisation (Couper-Kuhlen

2004, Curl 2004, Szczepek Reed 2009a, b). Hinzu kommen Fragen wie die nach dem Zusammenspiel prosodisch-phonetischer Parameter mit anderen sprachlichen Dimensionen, insbesondere visuellen Ressourcen (Li 2013, 2014a, b), nach der möglichen Hierarchisierung dieser Ressourcen, d.h. kann eine sprachliche Dimension eine andere überlagern; wenn ja, wann, wie und zu welchem Zweck (vgl. z.B. Schegloff 1996, Selting 2000, 2005). Welche Rolle spielen Prosodie und Phonetik beim Ausdruck von Affektivität (vgl. z.B. Selting 2012, Couper-Kuhlen 2012) und Epistemizität (Heritage 1998, 2012)? Welche Parameter variieren wiederum abhängig von geographischen und sozialen Variablen (vgl. z.B. Gilles 2005)?

Bei allem Forschungsbedarf muss sich die Untersuchung und Beschreibung prosodisch-phonetischer Ressourcen auch mit einer Reihe theoretischer und praktischer Problembereiche auseinandersetzen (vgl. auch Selting 2010). Die Frage nach Transkriptionsziel, Transkribierbarkeit gesprochener Sprache und Rolle des Transkripts stellt sich dabei schon ganz zu Anfang prosodisch-phonetischer Untersuchungen (vgl. Walker 2014a, b, Auer 2014). In theoretischer Hinsicht kommt etwa hinzu die ganz grundsätzliche Frage, nach dem Umgang mit Konzepten, Termini und Befunden der etablierten prosodisch-phonetischen/ phonologischen Forschung. Dies betrifft Forschungstraditionen wie die Britische Schule (z.B. Cruttenden 1997, Crystal 1969, 1975) und damit verwandte pädagogische Ansätze (z.B. Armstrong/ Ward 1926, O'Connor/ Arnold 1961, Brazil/ Coulthard/ Johns 1980) ebenso wie korpuslinguistische Untersuchungen (z.B. Knowles 1991, Wichmann 2000, Mindt 2001), die systemisch-tonetische Tradition (z.B. Halliday 1967, Tench 1996), den autosegmentalen Ansatz (z.B. Pierrehumbert 1980, Ladd 2008), akustische (z.B. 't Hart/ Collier/ Cohen 1990) und konfigurationale Ansätze (z.B. Bolinger 1986, 1989). Hier werden häufig die gleichen hörbaren Phänomene untersucht, allerdings meist mit elizitierten, experimentellen, monologischen oder sogar intuitiven Daten. Inwieweit können oder müssen darauf basierende Erkenntnisse als Basis für die gesprächsforscherische Arbeit genutzt werden? Welcher Austausch ist möglich? Was können und was müssen interaktionslinguistische Ansätze diesbezüglich bieten?

Insbesondere für NeueinsteigerInnen stellen sich auf der anderen Seite auch Fragen ganz praktischer Art. Prosodisch-phonetische Analysen beziehen eine Vielzahl von Parametern ein. Jeder Parameter an sich stellt eine Herausforderung bezüglich seiner Konzeptualisierung und Untersuchung dar. Auditive Analysen müssen (ein)geübt werden, und die Perzeption an sich unterliegt einer Vielzahl von Faktoren, selbst bei geübten AnalytikerInnen. Akustische Analysen hingegen können mit Hilfe von PRAAT (Boersma/ Weenink 2014, http://www.fon.hum.uva.nl/praat/) scheinbar einfach erstellt werden, sie stellen allerdings hohe Ansprüche an Messmethoden und die Interpretation der Ergebnisse (vgl. auch Selting 2010: 25 und z.B. Szczepek Reed 2010a).

Selbst der Versuch nur eine dieser Fragen befriedigend zu beantworten, ginge weit über den Rahmen des vorliegenden Bandes hinaus. Was er aber kann, ist Lösungsansätze aufzeigen, die die einzelnen BeiträgerInnen für sich gewählt haben. Zudem hoffen wir damit zu illustrieren, wie gewinnbringend gerade das Gebiet Prosodie und Phonetik in der Interaktion aufgrund seiner zahlreichen bisher noch wenig erforschten Aspekte für NeueinsteigerInnen sein kann. Dieser Band möchte insbesondere ihnen den Weg ebnen. Er richtet sich an GesprächsforscherInnen, die die Herausforderungen prosodisch-phonetischer Analysen annehmen möchten.

Für sie sollen zunächst grundlegende Begriffe und Konzepte einschließlich ihrer jeweiligen in der englischsprachigen Literatur gängigen Entsprechungen vorgestellt und so der Einstieg erleichtert werden. Nicht nur für sie gibt der Band dann mit seiner Vielzahl von Untersuchungsaspekten aber auch einen Einblick in die Bandbreite möglicher Themen, die eine systematische Beschäftigung mit den Phänomenen dieses Gebiets eröffnet. Damit trägt der Band auch dazu bei, aktuelle Befunde zu Prosodie und Phonetik in der Interaktion zu sammeln und den bisherigen Forschungsergebnissen zur Seite zu stellen. Weiterhin kann die Mischung aus interaktionslinguistischen, sprechwissenschaftlichen und phonetisch-phonologischen Herangehensweisen an interaktionale Daten anregen, die jeweils eigenen Methoden bezüglich ihrer Analysestärke und Erklärungskraft im Vergleich mit anderen zu reflektieren. Nicht zuletzt bietet er mit seinem gesprächsforscherisch weiten Fokus einen, wenn auch nicht umfassenden, Einblick in die aktuelle Forschung zu Prosodie und Phonetik in der Interaktion v.a. im deutschsprachigen Raum.

Im Folgenden findet sich zunächst eine Einführung in den Gegenstandsbereich sowie Termini und Konzepte der Prosodie- und Phonetikforschung. Daran schließt sich ein kurzer Überblick über die Forschungsmethoden an, in denen sich die Beiträge des vorliegenden Bandes verorten. Mit einem Überblick über die einzelnen Beiträge schließt diese Hinführung.

2. Termini und Konzepte

2.1. Gegenstandsbereich

Unter den grundlegenden Fragen, die sich bei der Untersuchung von Prosodie und Phonetik in der Interaktion stellen, ist nicht zuletzt die nach dem Gegenstandsbereich. Was genau gehört eigentlich zu 'Prosodie' und 'Phonetik'? Was davon muss in die Untersuchung einbezogen werden?

Der interaktionslinguistischen Perspektive verpflichtet verwenden wir beide Begriffe ganz bewusst im Zusammenhang, teilweise sogar mit Bindestrich ('prosodisch-phonetisch'). In der Vergangenheit wurden sie deutlich voneinander abgegrenzt: Phonetik (*phonetics*) beschäftigte sich mit phonetischen Segmenten, d.h. einzelnen Lauten (vgl. Fig. 1, phonetisch$_1$). Diese wurden bezüglich der Art und Weise ihrer Bildung beschrieben und oft genug direkt der Phonologie (*phonology*) zugeführt, d.h. abstrahiert in ein System der (möglichen und bedeutungsunterscheidenden) Laute einer Sprache. Von den Segmenten abgegrenzt wurden die suprasegmentalen Phänomene, also alles, was Lautkombinationen (Silben, Worte, Phrasen, Äußerungen) betraf. Hier fand sich auch die Prosodie (*prosody*), von Crystal (1969) als eines von drei Teilsystemen der Suprasegmentalia (*suprasegmentals*) dargestellt.

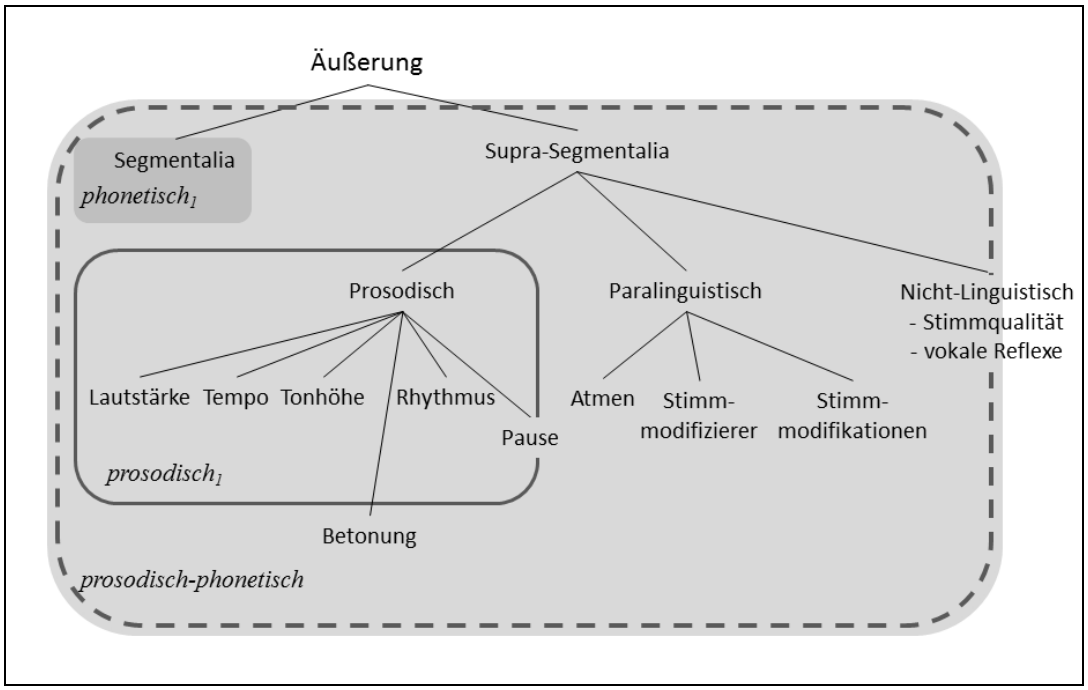


Fig. 1: Hörbare Komponenten von Äußerungen und ihre Einordnung in Forschungsgebiete (in Erweiterung von Crystal 1969: 131 und Couper-Kuhlen 1986: 4)

Als 'prosodisch' wurden dabei die Vokaleffekte der allgegenwärtigen Parameter Tonhöhe, Lautstärke und Tempo, also auch Rhythmus, bezeichnet, die in grundsätzlich variabler Beziehung mit den betroffenen Worten stehen (:5, prosodisch$_1$). Paralinguistisch (*paralinguistic*) sind die Vokaleffekte, die anders als mit den Stimmbändern produziert werden und nicht immer beobachtbar sind, also Atmen, Stimmmodifizierer (*voice qualifiers*) wie Flüstern und Glottalisierung, und Stimmmodifikationen (*voice qualifications*) wie Lachen und Weinen. Zu den nicht-linguistischen Phänomenen zählte Crystal permanente, idiosynkratische Merkmale, wie eine durchweg knarrende Stimme (*creaky voice*) und unwillkürliche Phänomene (*vocal reflexes*) wie Husten.

In der Praxis erwies sich diese Unterscheidung allerdings nicht immer als praktikabel. Insbesondere Pausen und Betonung ließen sich nicht befriedigend zuordnen oder fielen laut Crystals Definition sogar aus dem Gegenstandsbereich der Prosodie heraus. Eine Lösung fand sich mit Firths (1957) wesentlich weiterem Verständnis. Er bezeichnet mit *prosodies* alle syntagmatischen Beziehungen zwischen Silben, Wörtern und Äußerungen, die nicht von deren Struktur bestimmt und als kommunikative Signale genutzt werden. Dieses weite Verständnis (prosodisch$_2$) schließt im Ergebnis nicht nur Crystals prosodische und paralinguistische Mittel ein, sondern auch manche nicht-linguistischen, so lange sie als kommunikative Praktiken genutzt werden (vgl. z.B. Heath 2002 zu Husten als *dramatic demonstration*).

Ein solch weites Verständnis liegt auch den interaktionslinguistischen Untersuchungen von Prosodie und Phonetik in der Interaktion zugrunde. Natürlicherweise, denn ein Ansatz, der mit der Konversationsanalyse von der Prämisse "order at all points" (Heritage 1984: 241-242) ausgeht, kann kein hörbares Detail *a priori* ausschließen, eingedenk Jeffersons (2004) Warnung "one cannot know what one will find until one finds it" (:15). Die Untersuchungen sollten "all audible aspects

in and of speech that are produced by the human vocal apparatus" (Ford/ Couper-Kuhlen 2004: 3) einschließen, die potenziell interaktional signifikant sind (ibid.), also sowohl den originären Gegenstandsbereich der Phonetik (Segmentalia) als auch den der Prosodie (Suprasegmentalia) betreffen (siehe auch Szczepek Reed 2010a). In ihrer Breite nähern sich damit die interaktionslinguistischen Untersuchungen auch an den impressionistischen Yorker Ansatz (vgl. auch Local/ Walker 2005) an.

2.2. Termini und methodische Herangehensweisen

Im Folgenden werden zentrale prosodische Parameter und relevante Fachbegriffe der Prosodie- und Phonetikforschung kurz dargestellt. Detaillierte Einführungen in prosodische Grundbegriffe finden sich z.B. bei Couper-Kuhlen (1986) und Szczepek Reed (2010a), zu phonetischen Konzepten aus interaktionaler Perspektive siehe z.B. Ogden (2009).

Der in der Prosodieforschung wohl am häufigsten genannte Terminus ist der der *Intonation*. Er dient einigen Ansätzen als Oberbegriff für Frequenzbewegungen, Betonung und Rhythmus; in der interaktionslinguistischen Literatur bezieht er sich meist ausschließlich auf Veränderungen in der *Grundfrequenz* (*pitch* oder (*fundamental*) *frequency*). Hier wird oft auch von der (Intonations-) Kontur (*pitch contour*) gesprochen. Letztere benennt eine phrasenhafte, abgeschlossene Tonhöhenbewegungsgestalt. Frequenz, oder Tonhöhe, ist das hörbare Resultat der Vibration der Stimmlippen, Frequenz-/ Tonhöhenänderungen ein Ergebnis veränderter Vibrationsgeschwindigkeiten der Stimmlippen: grob gesagt, je schneller die Schwingungen, desto höher wird ein Ton vom menschlichen Ohr empfunden. Frequenz, gemessen in Hertz, bezieht sich hierbei auf die Anzahl der Stimmlippenschwingungen pro Sekunde. So entsteht zum Beispiel eine Tonhöhe von 400Hz durch 400 Schwingungen pro Sekunde. Im in der Gesprächsforschung gängigen Transkriptionssystem GAT (vgl. auch GAT 2 in Selting et al. 2009, Couper-Kuhlen/ Barth-Weingarten 2011) werden Intonationsbewegungen am Ende von Einheiten durch Satzzeichen angegeben (z.B. '?' für stark ansteigende Tonhöhe) und im Feintranskript auch für einzelne Silbenakzente markiert.

Die Intonation wird weithin nicht nur als einer von mehreren suprasegmentalen Aspekten des Sprechens betrachtet (vgl. Abschnitt 2.1), sondern spielt auch eine Rolle bei der Strukturierung gesprochener Sprache. In diesem Zusammenhang wird von Intonationsphrasen (*intonation(al) phrases*) gesprochen; weitere gebräuchliche Termini sind unter anderem *Intonationseinheit* (*intonation unit*), prosodische Einheit, *tone unit* und *tone group*. Während in nicht-interaktionalen Ansätzen sowohl die Existenz als auch die Relevanz dieser Einheit weitgehend unhinterfragt akzeptiert wird, gibt es eine Reihe von interaktionslinguistischen Arbeiten, die sich mit der Intonationsphrase als Teilnehmerkategorie auseinandersetzen (vgl. z.B. Barnwell 2013, Barth-Weingarten z.B. 2011a, 2012, 2013, Szczepek Reed 2010b, 2010c, 2012). Aufgrund dieser Forschung wird klar, dass Phrasen unterhalb der Einheit des Redebeitrags für TeilnehmerInnen deutlich relevant sind; dass es für deren Abgrenzung und interpretatives Verständnis aber differenzierterer phonetischer, prosodischer und handlungsrelevanter Analysen bedarf als bisher angenommen. Intonationseinheiten sind zudem grundlegend für

das Transkriptionssystem GAT, denn es stellt Intonationseinheiten auf jeweils getrennten Zeilen dar.

Die Tonhöhe von SprecherInnen wird weiterhin auch dann relevant, wenn es darum geht, die Tonlage einer SprecherIn, eines Gesprächs oder eines einzelnen Redebeitrags zu bestimmen. Hier wird oft von (*Sprech)Register* oder *Tonhöhenumfang* (in Abhängigkeit von der Domäne engl. *pitch register, pich span* (der Redeeinheit)*, pitch range* (der SprecherIn) gesprochen. Im GAT-System wird der Wechsel in eine vergleichsweise hohe Tonlage durch <<h> > und in eine vergleichsweise tiefe Tonlage mit <<l> > markiert. Der Vergleich mit vorhergehenden Redebeiträgen ist hierbei zentral. Ob ein Redebeitrag als hoch oder tief analysiert werden kann, hängt immer von der individuellen Sprechgewohnheit des/ der SprecherIn und dem lokalen Gesprächskontext ab.

Ein weiterer zentraler prosodischer Parameter ist das *Sprechtempo*. In phonetischen Arbeiten wird hier zwischen Artikulationstempo (*articulation rate*, Segmente pro Sekunde ohne Pausen), Sprechtempo (*speech rate*, Segmente pro Sekunde mit Pausen oder Silbenrate pro Sekunde), Silbenlänge oder Anzahl von Pausen unterschieden. Messungen spielen also eine große Rolle. Im interaktionslinguistischen Ansatz wird meist vom impressionistischen Höreindruck ausgegangen, im Transkript durch <<all> > (allegro, d.h. schnelles Sprechen) und <<len> > (lento, d.h. langsames Sprechen, bzw. <<acc> > (accelerando, d.h. schneller werdend) und <<decc> > (deccelerando, d.h. langsamer werdend) markiert. Detailliertere Untersuchungen arbeiten jedoch auch hier zunehmend mit akustischen Messungen, etwa mithilfe von PRAAT (Boersma/ Weenink 2014, http://www.fon.hum.uva.nl/praat/). Zum Tempo gehört auch die *Längung* von Silben und Segmenten (*lengthening*). Sie ist allerdings ohnehin wichtiger Bestandteil interaktionaler Analysen, da auch sie Kontextualisierungshinweise (vgl. Abschnitt 1) geben kann, etwa bezüglich der Strukturierung gesprochener Sprache oder auch im Hinblick auf Affektivität. In GAT wird Silbenlängung durch Doppelpunkt(e) innerhalb der Silbe angegeben, z.B. *kla::r*.

Ebenso wie Längung spielen auch *Pausen* im spontanen Gespräch eine Rolle, die über das Sprechtempo hinausgeht. Hier wird zum einen zwischen gefüllten und ungefüllten Pausen unterschieden, wobei beide Arten unterschiedliche Konsequenzen für die Organisation des Rederechts haben. Zum anderen muss darauf hingewiesen werden, dass Pausen, genauso wie andere prosodische Parameter auch, der subjektiven Perzeption unterworfen sind. Ob eine Pause als lang oder kurz empfunden wird, hängt von einer Anzahl Faktoren ab. Rhythmus etwa lebt von einer bestimmten zeitlichen Anordnung von Silben (vgl. hierunter). Je nach Sprache bestehen dann zum Beispiel gewisse Erwartungen hinsichtlich des genauen Zeitpunkts, zu dem Silben in Folge platziert werden müssen, um eine rhythmische Struktur zu erhalten. Es ist deshalb manchmal notwendig, die Produktion einer Silbe oder eines Wortes leicht zu verzögern, um den 'beat' zu treffen. Wenn hierzu Pausen notwendig sind, werden diese selten als solche von den Hörern wahrgenommen. Ein weiterer Faktor, der in der Analyse von Pausen relevant ist, ist der Handlungskontext. Gewisse Handlungen machen eine besonders schnelle Erwiderung erwartbar, so zum Beispiel gewisse affektgeladene Äußerungen. Wenn Rezipienten hier nicht sofort reagieren, kann dies als Pause und somit als Verzögerung gehört werden, auch wenn in anderen Kontexten eine rein zeitlich ähnlich lange Pause unbemerkt geblieben wäre.

Wie bereits angesprochen wurde, können Silben beim Sprechen rhythmisch strukturiert werden. Grob lassen sich dabei drei Arten von *Sprechrhythmus* (*rhythm*) unterscheiden: *stress-timing* (z.B. Englisch, Deutsch), *syllable-timing* (z.B. Französisch, Spanisch) und *mora-timing* (z.B. Japanisch). Ein großer Teil der interaktionslinguistischen Arbeiten zu Rhythmus analysiert englische Daten, und orientiert sich somit am stress-timing. Hier werden betonte Silben als rhythmische Schläge (*beats*) empfunden. Selten handelt es sich hierbei wirklich um Isochronie – menschliches Sprechen hält keinem Metronom stand. Trotzdem kann in flüssigen Teilen eines Redebeitrags eine gewisse Regelmäßigkeit bei der Produktion von Akzenten und Betonungen beobachtet werden. Insgesamt variieren rhythmische Strukturen aber stark von Sprache zu Sprache, von SprecherIn zu SprecherIn und von Redebeitrag zu Redebeitrag und sie hängen zudem von Handlungen ab (vgl. z.B. Couper-Kuhlen 1993) und können deshalb nur als Tendenzen betrachtet werden (z.B. Dauer 1983, Miller 1984).

Der dritte Parameter, der für Beschreibungen prosodischer Merkmale zentral ist, ist die *Lautstärke* (*volume*). Veränderungen in der Lautstärke können verschiedene Domänen betreffen: einzelne Silben, einzelne Wörter oder einen Redebeitrag. Zudem kann die Sprechgewohnheit einer SprecherIn variieren. Lautstärke bezeichnet den Höreindruck der Intensität (*intensity*). Letzterer entsteht durch die Stärke des Luftstromes beim Sprechen. GAT markiert einzelne Phrasen oder Beiträge mit <<f> > (forte) und <<p> > (piano), bzw. <<cres> > (crescendo, d.h. lauter werdend) und <<dim> > (diminuendo, leiser werdend). Größere Lautstärke wird z.B. in affektgeladenen Passagen gewählt oder zur Verteidigung des Rederechts, sie kann aber auch zur Fokussierung einzelner Silben beitragen, die dann vergleichsweise lauter als andere in ihrer Umgebung empfunden werden. Oft erfährt die betroffene Silbe zusätzlich einen Tonhöhensprung und/ oder Längung. Während es sich im ersten Fall um reine *Betonung* (*stress*) handelt, wird der zweite Fall als *Akzent* (*accent*) beschrieben. Beides wird in der GAT Notation durch Großbuchstaben markiert.

Neben prosodischen Parametern werden in der interaktionalen Linguistik, wie oben bereits erwähnt, zunehmend ursprünglich phonetische Parameter hinzugezogen. Dies betrifft oft den Bereich der Stimmqualität. Stimmqualitäten, die eine deutliche interaktionale Rolle zu spielen zu scheinen, sind *Knarrstimme* (zusammenfassend oft *creak, creaky voice*) (Ogden 2001, 2004) und *behauchte Stimme* (*breathy voice*) (Szczepek Reed 2006, 2010a). Unmarkierte/ neutrale Stimmqualität wird als *modal* (*modal voice*) bezeichnet. Bei der Knarrstimme kommt es zu starken Unregelmäßigkeiten in den Stimmlippenschwingungen, die typischerweise in eine sehr tiefe Tonhöhe resultieren. Knarrstimme kann lokal auftreten, etwa zur Strukturierung gesprochener Sprache, sie kann aber auch Merkmal einer Sprechgewohnheit, eines physiologischen Zustands oder einer Varietät (z.B. Amerikanisches Englisch) sein. Bei der behauchten Stimme schließen sich die Stimmlippen nicht ganz, und es kommt zu einem zusätzlichen Luftstrom beim Sprechen. Auch hier muss lokales Auftreten von Sprechgewohnheit unterschieden werden.

Obwohl jeder dieser Parameter für sich beschrieben werden kann, ist es im interaktionslinguistischen Ansatz unabdingbar, sie als Teil der interaktionalen Umgebung zu betrachten. Dies gilt für den Gebrauch der Parameter durch einzelne SprecherInnen. So ist besonders beim SprecherInnenwechsel eine Orientierung an

den prosodisch-phonetischen Parametern der vorherigen SprecherIn(nen) erwartbar. Zugleich, und grundsätzlich, gilt dies aber auch für die Einbettung der Parameter(änderungen) im lokalen, sequenziellen Kontext. Nur vor dem Hintergrund der Handlungen und Aktivitäten der TeilnehmerInnen können die Parameter ihre Wirkkraft entfalten und im Zusammenspiel mit anderen kommunikativen Ressourcen, etwa visuellen, bestimmte Bedeutungen kontextualisieren. Diese Perspektive auf die lokale Handlungsumgebung kennzeichnet die interaktionale Herangehensweise an Phonetik und Prosodie (vgl. etwa die Beiträge von Kupetz, Küttner und Schönherr in diesem Band). Der in gewisser Weise "radikalste" Ansatz in dieser Richtung ist der der dialogischen Prosody (*dialogic prosody*) (vgl. Couper-Kuhlen in diesem Band).

Daneben finden sich weitere Ansätze, die sich mit Phonetik und Prosodie in der Interaktion beschäftigen, zum Teil aber andere method(olog)ische Schwerpunkte setzen.

In der phonetischen/ phonologischen Tradition etwa spielt neben detaillierter, parametrischer phonetischer Beschreibung akustische Evidenz eine größere Rolle. In diese Tradition ordnen sich auch Methoden wie die Untersuchung quasi-spontansprachlicher, elizitierter Daten, Perzeptionstests sowie die Anwendung statistischer Tests auf größere Datenmengen ein (vgl. die Beiträge von Köser und Trouvain/ Truong in diesem Band).

Die Yorker *Phonology for Conversation* hingegen versucht, traditionell phonetische/ phonologische Methoden und Untersuchungsgegenstände mit der Konversationsanalyse zu verbinden, um die Relevanz der Kategorien für die TeilnehmerInnen zu prüfen (vgl. den Beitrag von Plug in diesem Band).

Die Sprechwissenschaft unterscheidet sich von den anderen Herangehensweisen in ihrer dezidiert angewandten Ausrichtung. Sie hat eine eigene Forschungstradition, was sich in der Verwendung von teilweise anderen Termini (vgl. z.B. stimmlich-artikulatorischer Ausdruck, einschließlich Stimmklang, Sprechstimmlage, Melodieumfang und -verlauf (Intervalle), Sprechspannung, Sinnkern) niederschlägt. Sie verwendet eher gestalthaft-beschreibende Konzepte (vgl. *motherese* in Bose/ Kurtenbach in diesem Band).

Allen Beiträgen gemeinsam ist aber, dass sie versuchen, Prosodie und Phonetik in der Interaktion zu beschreiben und damit offen zu legen, welche Ressourcen die TeilnehmerInnen nutzen, um sich zu verständigen und ihre kommunikativen Ziele zu verfolgen. Das Faszinierende daran ist, welche Systematizität sich hierbei finden lässt, wenn man die Herausforderungen, die der Untersuchungsgegenstand für die AnalytikerIn bereithält, annimmt.

3. Überblick über die Beiträge

Dieser Band enthält weitgehend die Beiträge der 17. Arbeitstagung zur Gesprächsforschung "Prosodie im Gespräch", die vom 20. bis 22. März 2013 am Institut für Deutsche Sprache in Mannheim stattfand (vgl. aber auch Bergmann 2008, 2012, Gerstenberg 2014, Ogden 2013). Sie wurden ergänzt um den Beitrag von Leendert Plug. Methodologisch sind die Beiträge breit gefächert. Ebenso wird das Konzept der Interaktivität, und damit verbunden die Reichweite der phonetischen und prosodischen Phänomene, von den AutorInnen unterschiedlich weit gefasst.

Stephanie Kösers Kapitel beschäftigt sich mit Knarrstimme (*creak*) und Aushauchen am Ende von Redebeiträgen im Deutschen und fragt, ob diese phonetischen Merkmale eine Rolle beim Sprecherwechsel spielen. Anhand von elizitierten Alltagsgesprächen findet Köser in mehr als zwei Dritteln der Fälle, dass das Ende von Redebeiträgen nicht modal realisiert wird und entweder von Knarrstimme, Aushauchen oder beidem begleitet wird. Kösers Studie ist eine der ersten, die die Relevanz der Stimmqualität für das Deutsche auf der Basis empirischer, interaktionaler Daten belegt.

Uwe Küttners Beitrag hat die rhythmische Analyse spontaner englischsprachiger Gesprächsdaten zum Inhalt. Durch detaillierte Analysen der rhythmischen Akzentverteilung in *demonstrative* wh-*cleft*-Konstruktionen arbeitet Küttner heraus, dass SprecherInnen die Fortführung ihres Redebeitrags durch rhythmisiertes Sprechen projizieren können. Damit liefert Küttner starke Evidenz für die rhythmische Integriertheit von prosodischen Einheiten und die Relevanz rhythmischer Strukturen als Teilnehmerkategorie.

Leendert Plug untersucht in seinem Artikel die Relevanz und Anwendbarkeit der aktuell häufig gebrauchten Begriffe *upgraded* und *downgraded* anhand niederländischer Gesprächsdaten. Plug zeigt, dass im Fall von Selbstreparaturen Tonhöhe und Intensität stark korrelieren und gemeinsam für den Höreindruck der prosodischen Markierung verantwortlich sind. Seine Analyse macht jedoch auch deutlich, dass Markierungen verschiedener prosodischer Parameter unterschiedliche interaktionale Funktionen haben. Deshalb ist die Nutzbarkeit eines Begriffes wie *upgrading* immer eine Frage des interaktionalen Kontextes.

Maxi Kupetz analysiert deutsche Alltagsgespräche im Hinblick auf die Darstellung von Empathie in der Interaktion. Sie zeigt, dass für die Gestaltung von mitfühlendem Sprechen eine Reihe prosodisch-phonetischer Merkmale typisch sind, wie zum Beispiel abnehmende Lautstärke, tiefes Tonhöhenregister, flach auslaufende Kontur, Behauchung und Silbendehnung. Kupetz' Arbeit zeigt deutlich, wie wichtig die Betrachtung aller relevanten phonetischen und prosodischen Parameter ist, da im Fall von mitfühlendem Sprechen niemals nur ein einziger Parameter eingesetzt wird.

Jürgen Trouvain und Khiet P. Truong beschäftigen sich mit sprecherüberlappendem Lachen anhand großer, transkribierter Korpora englischsprachiger Gespräche. Es wird gezeigt, dass überlappendes Lachen im Durchschnitt länger anhält und höhere Tonhöhe, größere Intensität und mehr Stimmhaftigkeit erreicht als nichtüberlappendes Lachen. Die Autoren widmen sich unter anderem den Implikationen ihrer Resultate für das 'Einladen' zum Mitlachen.

Ines Bose und Stephanie Kurtenbach behandeln Interaktionen zwischen Erzieherinnen und zwei- bis vierjährigen Kindern in einem deutschen Kindergarten. Die Autorinnen untersuchen die Gesprächskompetenz der Erzieherinnen in Bezug auf die Angemessenheit und Sensitivität ihrer Gesprächspraktiken im Umgang mit den Kindern. Die Studie zeigt, dass Erzieherinnen mit sehr expressivem, variierendem stimmlich-artikulatorischen Ausdruck weit sensibler auf das Gespräch mit Kindern eingehen und das Gespräch wesentlich erfolgreicher fokussieren als Erzieherinnen, deren Stimmausdruck sich als unflexibel und schablonenhaft darstellt.

Beatrice Szczepek Reed analysiert in ihrem Kapitel den Gebrauch von prosodisch stilisiertem *süß* als freistehendem Turn. Der Artikel zeigt, wie Gesprächs-

teilnehmerInnen lexikalische, prosodische und sequentielle Merkmale als Einheit gebrauchen, und verwendet hierfür den Terminus *assemblage*. Mit Bezug auf prosodische Stilisierung argumentiert die Autorin, dass der interaktional wichtigste Aspekt die beträchtliche Abweichung eines stilisierten Turns von der Gesprächsumgebung ist, während die präzise Form der prosodischen Abweichung eine vergleichsweise untergeordnete Rolle spielt.

Beatrix Schönherr untersucht Prosodie im Zusammenspiel mit *kooperativen Gesten* anhand zweier Theateraufführungen im Deutschen. Vier detaillierte Analysen von besonders ausdrucksvollen Szenen erlauben die Verdeutlichung von prosodisch und körperlich stark markierten Redebeiträgen und ihrer klar sequenziell geprägten Verwendung. Ein gestisches Angebot eröffnet die Möglichkeiten der Annahme und der Ablehnung, mit der gleichen Präferenz für eine Annahme wie in der verbalen Interaktion. Schönherrs Arbeit zeigt deutlich, wie unerlässlich die gemeinsame Betrachtung sprachlicher und nicht-sprachlicher Handlungen für die Gesprächsanalyse ist.

Elizabeth Couper-Kuhlen schließlich argumentiert in ihrem Kapitel für eine Betrachtung der Prosodie als sprecherübergreifend (dialogisch). Couper-Kuhlen analysiert freistehende Antwortpartikeln im Englischen wie *oh*, *alright*, *okay* auf ihre Beziehung zu vorangehenden prosodischen Mustern hin. So wird zum Beispiel argumentiert, dass *upgrading* – etwa durch eine weitere Tonhöhenspanne als der vorangegangene Redebeitrag – ikonisch mildes Interesse signalisiert, während *downgrading* durch tiefere Tonhöhe und reduzierte Lautstärke Empathie kommuniziert. Tonhöhe und Lautstärke scheinen besonders für *informing*-Sequenzen relevant zu sein, während für *requests* hauptsächlich die zeitliche Koordinierung zählt.

In ihrer Bandbreite erlauben die Beiträge einen Einblick in die derzeitige method(olog)ische und thematische Vielfalt der Forschung zu Prosodie und Phonetik in der Interaktion im deutschsprachigen Raum. Und sie machen hoffentlich Lust auf mehr...

4. Literatur

Armstrong, Lilias E./ Ward, Ida C. (1926): Handbook of English intonation. Leipzig: Teubner.

Auer, Peter (2014): There's no harm in glossing (but a need for a better understanding of the status of transcripts). In: Research on Language and Social Interaction 47 (1), 17-22.

Auer, Peter/ Di Luzio, Aldo (1992) (Hrsg.): The contextualization of language. Amsterdam: Benjamins.

Auer, Peter/ Couper-Kuhlen, Elizabeth/ Müller, Frank (1999): Language in time. The rhythm and tempo of spoken interaction. Oxford: OUP.

Barnwell, Brendan (2013): Perception of prosodic boundaries by intrained listeners. In: Szczepek Reed, Beatrice/ Raymond, Geoffrey (Hrsg.): Units of Talk – Units of Action. Amsterdam: Benjamins, 125-165.

Barth-Weingarten, Dagmar (2011a): The fuzziness of intonation units: Some theoretical considerations and a practical solution. In: InLiSt – Interaction and Linguistic Structures 51 (http://www.inlist.uni-bayreuth.de/).

Barth-Weingarten, Dagmar (2012): Cesuring in talk-in-interaction: A parametric approach to intonation units and their role in turn-taking, backchanneling, language variation and language change. Habilitationsschrift. Universität Freiburg.

Barth-Weingarten, Dagmar (2013): From "intonation units" to cesuring – an alternative approach to the prosodic-phonetic structuring of talk-in-interaction. In: Szczepek Reed, Beatrice/ Raymond, Geoffrey (Hrsg.): Units of Talk – Units of Action. Amsterdam: Benjamins, 91-124.

Barth-Weingarten, Dagmar/ Dehé, Nicole/ Wichmann, Anne (2009) (Hrsg.): Where Prosody meets pragmatics. Bingley: Emerald.

Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (Hrsg.): Prosody in interaction. Amsterdam: Benjamins.

Bergmann, Pia (2008): Regionalspezifische Intonationsverläufe im Kölnischen. Formale und funktionale Analysen steigend-fallender Konturen. Tübingen: Niemeyer.

Bergmann, Pia (2012): The prosodic design of parentheses in spontaneous speech. In: Bergmann, Pia et al. (Hrsg.): Prosody and embodiment in Interactional Grammar. (lingua et litterae series) Berlin: de Gruyter, 103-141.

Bolinger, Dwight (1986): Intonation and its parts: melody in spoken English. Stanford: Stanford University Press.

Bolinger, Dwight (1989): Intonation and its uses. Melody in grammar and discourse. Stanford: Stanford University Press.

Brazil, David/ Coulthard, Malcolm/ Johns, Catherine (1980): Discourse Intonation and Language Teaching. London: Longman.

Chafe, Wallace (1994): Discourse, consciousness and time. The flow and displacement of conscious experience in speaking and writing. Chicago: Chicago University Press.

Couper-Kuhlen, Elizabeth (1986): An Introduction to English prosody. Tübingen: Niemeyer.

Couper-Kuhlen, Elizabeth (1993): English Speech Rhythm. Form and function in everyday verbal interaction. Amsterdam: Benjamins.

Couper-Kuhlen, Elizabeth (1998): Prosody in interactional discourse. In: SKY Yearbook 1998: The yearbook of the Linguistic Association Finland, 7-40.

Couper-Kuhlen, Elizabeth (2004): Prosody and sequence organization in English conversation: The case of new beginnings. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (Hrsg.): Sound patterns in interaction . Amsterdam: Benjamins, 335-376.

Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (Hrsg.) (2004): Sound patterns in interaction. Amsterdam: Benjamins.

Couper-Kuhlen, Elizabeth/ Barth-Weingarten, Dagmar (2011): A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 12: 1-51 (www.gespraechsforschung-ozs.de).

Couper-Kuhlen, Elizabeth (2012): Exploring affiliation in the reception of conversational complaint stories. In: Peräkylä, Anssi/ Sorjonen, Marja-Leena (Hrsg.): Emotion in interaction. Oxford: OUP, 113-146.

Couper-Kuhlen, Elizabeth/ Selting, Margret (1996) (Hrsg.): Prosody in Conversation: Interactional Studies. Cambridge: CUP.
Cruttenden, Alan (1997): Intonation. Cambridge: CUP.
Crystal, David (1969): Prosodic systems and intonation in English. Cambridge: CUP.
Crystal, David (1975): The English tone of voice. Essays in intonation, prosody and paralanguage. London: Arnold.
Curl, Traci S. (2004): "Repetition" repairs: the relationship of phonetic structure and sequence organization. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (Hrsg.): Sound patterns in interaction. Amsterdam: Benjamins, 273-298.
Curl, Traci S. (2005): Practices in other-initiated repair resolution: The phonetic differentiation of 'repetitions'. In: Discourse Processes 35, 1-43.
Dauer, Rebecca M. (1983): Stress-timing and syllable-timing reanalyzed. In: Journal of Phonetics 11, 61-62.
Dilley, Laura/ Shattuck-Hufnagel, Stefanie/ Ostendorf, Mari (1996): Glottalization of word-initial vowels as a function of prosodic structure. In: Journal of Phonetics 24, 423-444.
Du Bois, John W. et al. (1993): Outline of discourse transcription. In: Edwards, Jane A./ Lampert, Martin D. (Hrsg.): Talking data: Transcription and coding in discourse research. Hillsdale, New York: Erlbaum, 45-87.
Firth, John R. (1957): Papers in Linguistics 1934-1951. Oxford: OUP.
Ford, Cecilia E./ Couper-Kuhlen, Elizabeth (2004): Conversation and phonetics: Essential connections. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (Hrsg.): Sound patterns in interaction. Cross-linguistic studies from conversation. Amsterdam: Benjamins, 3-25.
Gerstenberg, Annette (2014): The Same Old Story? Prosodic features in narratives of old age as cues for genre description. In: Cahiers de Linguistique Française 31, 87-92.
Gilles, Peter (2005): Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung. Berlin: de Gruyter.
Gumperz, John (1982): Discourse strategies. Cambridge: CUP.
Halliday, Michael A.K. (1967): Intonation and grammar in British English. The Hague: Mouton.
Halliday, Michael A.K. (1994): An Introduction to functional grammar. London: Arnold.
't Hart, Johan/ Collier, René/ Cohen, Antonie (1990): A perceptual study of intonation. An experimental-phonetic approach to speech melody. Cambridge: CUP.
Heath, Christian (2002): Demonstrative Suffering: The gestural (re)embodiment of symptoms. In: Journal of Communication 52 (3), 597-616.
Heritage, John (1984): Garfinkel and ethnomethodology. Cambridge: Polity Press.
Heritage, John (1998): Oh-prefaced responses to inquiry. In: Language in Society 27. 291-334.
Heritage, John (2012): Epistemics in action: Action formation and territories of knowledge. In: Research on Language and Social Interaction 45 (1), 1-29.
Jefferson, Gail (1986): Notes on 'latency' in overlap onset. In: Human Studies 9, 153-183.
Jefferson, Gail (1988): On the sequential organization of troubles talk in ordinary conversation. In: Social Problems 35 (4), 418-442.

Jefferson, Gail (2004): Glossary of transcript symbols with an introduction. In: Lerner, Gene (Hrsg.): Conversation Analysis. Studies from the first generation. Amsterdam: Benjamins, 13-31.

Knowles, Gerry (1991): Prosodic labelling: The problem of tone group boundaries. In: Johansson, Stig/ Stenström, Anna-Brita (Hrsg.): English computer corpora: Selected papers and research guide. Berlin, New York: Mouton de Gryuter, 149-163.

Ladd, Robert D. (2008): Intonational phonology. Cambridge: CUP.

Laver, John (1994): Principles of phonetics. Cambridge: CUP.

Lerner, Gene H. (1996): On the 'semi-permeable' character of grammatical units in conversation: conditional entry into the turn space of another speaker. In: Ochs, Elinor/ Schegloff, Emanuel A./ Thompson, Sandra A. (Hrsg.): Interaction and grammar. Cambridge: CUP, 238-276.

Li, Xiaoting (2013): Language and the body in the construction of units in Mandarin face-to-face interaction. In: Szczepek Reed, Beatrice/ Raymond, Geoffrey (Hrsg.): Units of talk – Units of action. Amsterdam: Benjamins, 343-375.

Li, Xiaoting (2014a): Multimodality, interaction and turn-taking in Mandarin conversation. Amsterdam: Benjamins.

Li, Xaoting (2014b): Learning and participant intervening questions in Mandarin conversations. In: Journal of Pragmatics 67, 34-60.

Local, John/ Walker, Gareth (2004): Abrupt-joins as a resource for the production of multi-unit, multi-action turns. In: Journal of Pragmatics 36, 1375-1403.

Local, John/ Walker, Gareth (2005): Methodological imperatives for investigating the phonetic organization and phonological structures of spontaneous speech. In: Phonetica 62, 120-130.

Local, John/ Kelly, John/ Wells, William H.G. (1986): Towards a phonology of conversation: turn-taking in Tyneside English. In: Journal of Linguistics 22, 411-437.

Miller, M. (1984): On the perception of rhythm. In: Journal of Phonetics 12, 75-83.

Mindt, Ilka (2001): Intonation im Lancaster/ IBM Spoken English Corpus. Falls und fall-rises, Sprecherwechsel, paratones, declination. Tübingen: Narr.

O'Connor, Joseph D./ Arnold, G.F. [sic] (1961/1963 (1973)): Intonation of colloquial English. Longman: London.

Ogden, Richard (2001): Turn transition, creak and glottal stop in Finnish talk-in-interaction. In: Journal of the International Phonetic Association 31, 139-152.

Ogden, Richard (2004): Non-modal voice quality and turn-taking in Finnish. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (Hrsg.): Sound patterns in interaction. Amsterdam: Benjamins, 29-62.

Ogden, Richard (2006): Phonetics and social action in agreements and disagreements. In: Journal of Pragmatics 38, 1752-1775.

Ogden, Richard (2009): An Introduction to English Phonetics. Edinburgh: Edinburgh University Press.

Ogden, Richard (2013): Clicks and percussives in English conversation. In: Journal of the International Phonetic Association 43 (3), 299-320.

Pierrehumbert, Janet B. (1980): The phonetics and phonology of English intonation. PhD, MIT.

Pike, Kenneth L. (1945): The intonation of American English. Ann Arbor, Michigan: University of Michigan Publications.

Plug, Leendert (2011): Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. In: Journal of Phonetics 39 (3), 289-297.

Sacks, Harvey/ Schegloff, Emanuel A./ Jefferson, Gail (1974): A simplest systematics for the organization of turn-taking for conversation. In: Language 50, 696-735.

Schegloff, Emanuel A. (1982): Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In: Tannen, Deborah (Hrsg.): Analyzing discourse: Text and talk. Washington: Georgetown University Press, 71-93.

Schegloff, Emanuel A. (1996): Turn-organization: one intersection of grammar and interaction. In: Ochs, Elinor/ Schegloff, Emanuel A./ Thompson, Sandra A. (Hrsg.): Interaction and grammar. Cambridge: CUP, 52-133.

Schegloff, Emanuel A. (1997): Practices and actions: Boundary cases of other-initiated repair. In: Discourse Processes 23, 499-545.

Schegloff, Emanuel A. (1998): Reflections on studying prosody in talk-in-interaction. In: Language and Speech, 41, 235-263.

Selting, Margret (2000): The construction of units in conversational talk. In: Language in Society 29, 477-517.

Selting, Margret (2005): Syntax and prosody as methods for the construction and identification of turn-constructional units in conversation. In: Hakulinen, Auli/ Selting, Margret (Hrsg.): Syntax and Lexis in conversation. Studies on the use of linguistic resources in talk-in-interaction. Amsterdam: Benjamins, 17-44.

Selting, Margret (2010): Prosody in interaction: State of the art. In: Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (Hrsg.): Prosody in interaction. Amsterdam: Benjamins, 3-40.

Selting, Marget (2012): Complaint stories and subsequent complaint stories with affect displays. In: Journal of Pragmatics 44, 387-415.

Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion 10, 353-402 (www.gespraechsforschung-ozs.de).

Selting, Margret/ Couper-Kuhlen, Elizabeth (2000): Argumente für die Entwicklung einer 'interaktionalen Linguistik'. In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion (ISSN 1617-1837) 1, 76-95 (www.gespraechsforschung-ozs.de).

Selting, Margret/ Couper-Kuhlen, Elizabeth (2001a): Forschungsprogramm 'Interaktionale Linguistik'. In: Linguistische Berichte 187, 257-287.

Sweet, Henry (1906): A primer of phonetics. Oxford: Clarendon Press.

Szczepek Reed, Beatrice (2006): Prosodic orientation in English conversation. London: Palgrave Macmillan.

Szczepek Reed, Beatrice (2009a): Prosodic orientation: A practice for sequence organization in broadcast telephone openings. In: Journal of Pragmatics 41 (6), 1223-1247.

Szczepek Reed, Beatrice (2009b): FIRST or SECOND: Establishing sequential roles through prosody. In: Barth-Weingarten, Dagmar/ Dehé, Nicole/ Wichmann, Anne (2009) (Hrsg.): Where Prosody meets pragmatics. Bingley: Emerald, 205-222.

Szczepek Reed, Beatrice (2010a): Analysing Conversation: An Introduction to Prosody. Basingstoke: Palgrave Macmillan.

Szczepek Reed, Beatrice (2010b): Intonation phrases in natural conversation: A participants' category? In: Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (Hrsg.): Prosody in interaction. Amsterdam: Benjamins, 191-212.

Szczepek Reed, Beatrice (2010c): Units of interaction: Tone units or turn constructional phrases? In: Delais-Roussarie, Elisabeth (Hrsg.): Conference proceedings: Interface discourse and prosody. Paris, 9-11 September 2009. 351-363 (http://makino.linguist.jussieu.fr/idp09/actes_fr.html, Last access: 30-09-2014).

Szczepek Reed, Beatrice (2012): Prosody, syntax and action formation: Intonation phrases as 'action components'. In: Bergmann, Pia et al. (Hrsg.): Prosodic and visual resources in interactional grammar. Berlin, New York: de Gruyter, 142-170.

Szczepek Reed, Beatrice/ Raymond, Geoffrey (Hrsg.): Units of talk – Units of action. Amsterdam: Benjamins, 91-124.

Tench, Paul (1996): The intonation systems of English. London: Cassell.

Walker, Traci (2014a): Form ≠ function: The independence of prosody and action. In: Research on Language and Social Interaction 47 (1), 1-16.

Walker, Traci (2014b): The independence of phonetic form and interactional accomplishments. In: Research on Language and Social Interaction 47 (1), 23-27.

Wichmann, Anne (2000): Intonation in text and discourse. Beginnings, middles and ends. London: Longman.

Zeller, Margaret/ Post, Brechtje (2012): Combining formal and functional approaches to topic structure. In: Language and speech 55 (1), 119-139.

Dagmar Barth-Weingarten
Englische Sprache der Gegenwart
Institut für Anglistik und Amerikanistik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
Germany
dbarth [at] uni-potsdam.de

Beatrice Szczepek Reed
Department of Education
Derwent College
University of York
YO10 5DD
York
United Kingdom
beatrice.szczepek.reed [at] york.ac.uk

Phrasen-finale Phonationsänderungen und ihre Rolle beim *turn taking*

Stephanie Köser

Abstract
Gegenstand dieser phonetisch ausgerichteten Untersuchung ist die Frage, ob *creak* oder andere Änderungen in der Phonation am Ende einer Intonationsphrase im Deutschen eine Rolle beim *turn taking* spielen.
Drei Varianten von Phonationsänderungen lassen sich unterscheiden: (1) Laryngalisierung (*creak*): Die Stimmlippen schwingen typischerweise langsam und unregelmäßig. (2) Aushauchen: In der schwächsten Form ein Auslaufen der Stimmlippenschwingungen; stärker wahrnehmbar als Behauchung oder gar Flüstern. (3) Eine Kombination aus beiden.
Vor allem Laryngalisierung mit Aushauchen überwiegt vor Redebeitragswechseln oder an übergangsrelevanten Stellen (TRPs). Außerdem kommen einige Phonationsarten bevorzugt in Bündeln mit anderen phonetischen Merkmalen vor, Laryngalisierung zum Beispiel mit tiefer Grundfrequenz und Aushauchen mit anschließendem Atmen.
Die Veränderungen können zumeist durch allgemeine Entspannung erklärt werden: Ein Nachlassen des subglottalen Drucks und der Stimmlippenspannung begünstigt sowohl Laryngalisierung als auch Behauchung/ Aushauchen.

Phonation – Stimmqualität – Laryngalisierung – Redebeitragswechsel – Redebeitragsfortführung – Aushauchen – Behauchung

This study investigates whether creak and other shifts in phonation play a role in turn taking in German. In our corpus, three different phonation types occur frequently at the end of intonation phrases: (1) Laryngealization (creak): a slow and irregular vibration of the vocal folds. (2) Breathiness: in its weak form, a gradual tapering off of vocal cord vibration, in its strong form, breathy voice or whisper. (3) A combination of both.
The third variant dominates before speaker changes and at transition-relevance places. Moreover, some of these phonatory settings often occur together with other phonetic cues, for example laryngealization with low fundamental frequency, and breathiness is followed by breathing.
Overall muscle relaxation may explain many of these changes: A reduction of subglottal pressure and vocal cord tension leads to both laryngealization and breathiness.

phonation – voice quality – laryngealization – creak – turn taking – turn continuation – breathiness

1. Einleitung

1.1. Untersuchungsgegenstand

Aussagen wie das folgende Zitat von John Laver (1980: 126), einem bekannten Phonetiker, findet man immer wieder in der Literatur:

> […] in the paralinguistic regulation of interaction (Laver 1976: 351), speakers of Received Pronunciation often use creak or creaky voice, simultaneously with a falling intonation, as a signal of completion of their turn as speaker, yielding the floor to the listener.

Laver bezieht sich damit auf das Englische. Systematische Untersuchungen zur Rolle der Laryngalisierung oder gar anderer Phonationsarten beim *turn taking* sind aber eher selten, auch für das Deutsche. Deshalb soll in diesem Beitrag auf mögliche Änderungen der Phonation am Ende von Intonationsphrasen (IPs) im Deutschen genauer eingegangen werden.

Unter Phonation versteht man die Stimmbildung: die Anregung des Sprachschalls durch das Schwingen der Stimmlippen. Je nach Form der Stimmlippenschwingungen kann man unterschiedliche Phonationsarten unterscheiden, wie zum Beispiel Laryngalisierung oder Behauchung (vgl. Laver 1994: 184ff. oder Eckert/ Laver 1994: 45ff.).

Die vorliegende Untersuchung widmet sich neben modaler Phonation vor allem zwei Phonationsarten: (1) Laryngalisierung und (2) einer Gruppe von Schwingungsmodi, die hier unter dem Begriff "Aushauchen" zusammengefasst werden. Wie bei der Beschreibung der Ergebnisse noch näher erläutert wird, können diese auch miteinander kombiniert werden.

- Modalstimme: Die Stimmlippen schwingen regelmäßig und berühren sich in ihrer gesamten Länge. Es gibt kein hörbares Reibegeräusch, das auf eine unvollständige Annäherung (Adduktion) der Stimmlippen hinwiese (siehe Laver 1980: 109ff.). In der vorliegenden Studie ist der Begriff der Modalstimme jedoch etwas weiter gefasst: Es ist die neutrale Stimme eines Sprechers gemeint; leichte Abweichungen von oben beschriebener Modalstimme, die nicht punktuell auftreten, sondern zur Stimmqualität des Sprechers gehören, werden nicht berücksichtigt.
- Laryngalisierung: Die Stimmlippen schwingen unregelmäßig, langsam und mit niedriger Amplitude. Sie berühren sich für eine längere Zeit als in der Modalstimme. Die Stimme klingt tief und knarrig. Grundsätzlich kann Laryngalisierung die Folge von sehr hoher Anspannung im Bereich der Kehlkopfmuskulatur sein, aber auch von zu niedriger Spannung. Für diese Art der Phonation sind in der Literatur auch alternative Begriffe wie *creak*, Glottalisierung, *vocal fry* und andere zu finden. Für einen Literaturüberblick über die unterschiedlichen Bedeutungsnuancen siehe zum Beispiel Köser (2005: 17ff.) oder Laver (1994: 194ff.).
- Aushauchen: In der schwächsten Ausprägung ist im akustischen Signal eine stete Abnahme der Amplitude zu erkennen; die Stimmlippen werden hier stetig schwächer angenähert. Stärkere Ausprägungen sind Behauchung, Flüster-

stimme oder gar Flüstern, bei dem die Stimmlippen sich nicht mehr berühren, wie bei stimmlosen Lauten.

Für eine genauere Beschreibung der einzelnen Phonationsarten ist Laver (1980, 1991) oder Eckert/ Laver (1994) zu empfehlen, da die vorliegende Arbeit hauptsächlich auf Lavers System der Beschreibung und Einteilung der Phonationsarten (*phonatory settings*) basiert.

Um eine Vorstellung zu geben, wie die einzelnen Stimmqualitäten klingen, im Folgenden einige Beispiele, von einer weiblichen Sprecherin produziert. Der Trägersatz entspricht dem Beispielsatz aus Eckert/ Laver (1994), die ebenfalls Tonbeispiele für unterschiedliche Stimmqualitäten anbieten.

1. Modale Phonation
2. Geringe Berührung der Stimmlippen; von behauchter Stimme bis zu Flüstern immer weniger Berührung:
   Behauchte Stimme

   Flüsterstimme

   Flüstern
3. Unterschiedliche Formen von Laryngalisierung/ *creak*:

   Creaky voice

   Creak, geringe Spannung *Creak*, hohe Anspannung

Es geht in diesem Artikel um die Phonation an ganz spezifischen Stellen im Gespräch: (a) vor einem Redebeitragswechsel (also am Ende eines Redebeitrags) und (b) vor der Fortführung des Redebeitrags durch den aktuellen Sprecher (also innerhalb des Redebeitrags).

Der Begriff "Redebeitrag" ist hier wie folgt definiert: Er umfasst alles, was ein Sprecher sagt, während er das Rederecht innehat. Kurze Hintergrundkommentare werden wie Rezipientensignale als Hörersignale aufgefasst, die keinen Redebeitragsstatus haben. Insofern kann es auch zu einer Fortführung eines Redebeitrags kommen, wenn formal kurzzeitig ein Sprecherwechsel stattgefunden hat – wenn zum Beispiel der aktuelle Sprecher eine Intonationsphrase beendet hat, aber erst nach einer Pause weiterspricht, in der der Gesprächspartner ein Hörersignal produziert. Dieses Konzept eines Redebeitrags ist mit dem von Fiehler (2005: 1232ff.) in der DUDEN Grammatik beschriebenen Gesprächsbeitrag vergleichbar, und der hier beschriebene Redebeitragswechsel entspricht dem Beitragswechsel im DUDEN.

Redebeitragswechsel können nach Pause, nahtlos anschließend oder überlappend stattfinden, Weiterführungen nach einer Pause oder nahtlos (in beiden Fällen ist jedoch gleichzeitig die Überlappung mit einem Rezipientensignal denkbar). Die folgende Graphik stellt die drei wichtigsten zeitlichen Abläufe bei einem Redebeitragswechsel dar:

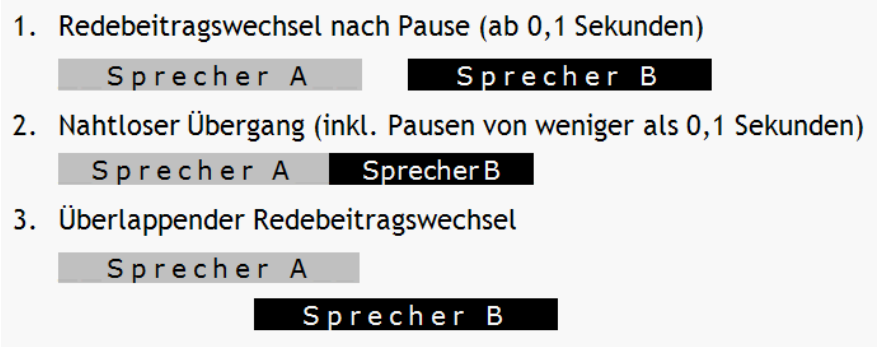


Abbildung 1: Möglichkeiten des zeitlichen Ablaufs von Redebeitragswechseln

Abbildung 2 zeigt, wie sich bei Fortführung von Redebeiträgen die nachfolgende IP zeitlich anschließen kann. Für die Weiterführung des Redebeitrags in Überlappung mit einem Hörersignal des Gesprächspartners werden exemplarisch zwei mögliche Abläufe verdeutlicht:

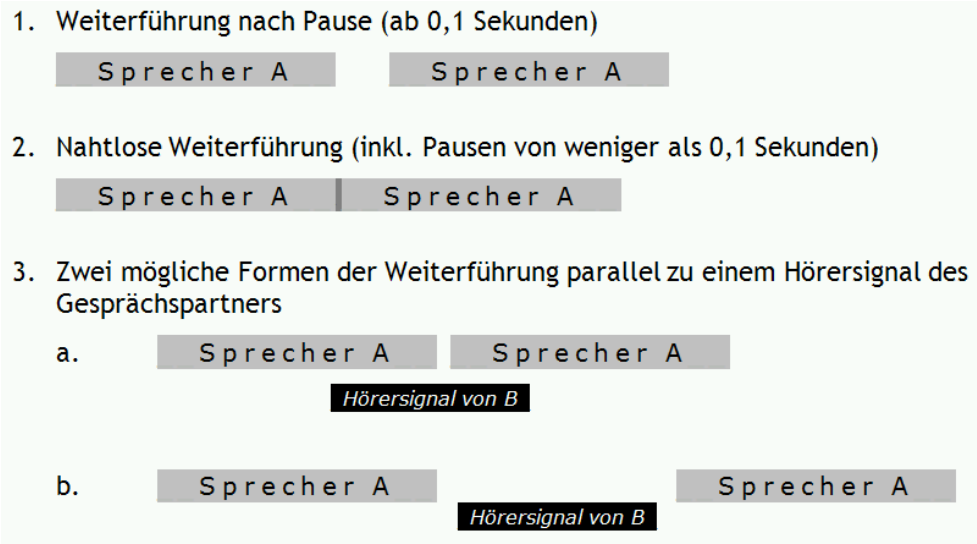


Abbildung 2: Möglichkeiten des zeitlichen Ablaufs von Redebeitragsfortführungen

1.2. Forschungsstand

Die phrasen-finale Phonation war bisher eher selten Gegenstand von Untersuchungen spontansprachlicher Daten und zum *turn taking*. An dieser Stelle soll exemplarisch auf drei Studien eingegangen werden, die zeigen, dass vor allem Laryngalisierung im Kontext von Redebeitragswechseln vorkommt.

In Ogden (2001, 2004) wird die Phonation in finnischen Gesprächen analysiert. Ogden konnte nachweisen, dass die Phonation hier beim *turn taking* eine Rolle spielt: *Creak* tritt vor allem an Stellen auf, die sich durch syntaktische, pragmatische und prosodische Abgeschlossenheit auszeichnen; aber nicht nur *creak*, sondern auch Behauchung und Flüstern oder die Abfolge von *creak* und anschließender Behauchung. Er schreibt diesen Phonationsarten deshalb die Funktion zu, den Turn abzugeben.

In einer Beschreibung des amerikanischen Englisch von Barth-Weingarten (2009) wird vor allem auf Glottalisierung[1] beim *turn taking* eingegangen. Sie ist Teil eines Clusters von sogenannten "response-allowing features". Auf diese Merkmale, zu denen neben der finalen Glottalisierung unter anderem steile Tonhöhenbewegungen zu einer deutlich hohen oder tiefen Tonhöhe und der Übergang in Ein- oder Ausatmen zählen, folgt gewöhnlich irgendeine Form von *response* von Seiten des Gesprächspartners. Andere Phonationsarten werden nicht behandelt. Hört und sieht man sich jedoch einige ihrer Beispiele an, so fällt auf, dass auch hier gelegentlich nach der Glottalisierung noch eine Art kurzes Flüstern folgt, das in Ausatmen übergeht.[2]

Für das Deutsche hat Peters (2006) einige phonetische Merkmale von Phrasengrenzen untersucht – darunter auch Laryngalisierung. Er hat Grenzen am Ende von Turns und innerhalb von Turns miteinander verglichen. In seinen Daten ist Laryngalisierung am Ende zahlreicher und auch von längerer Dauer als bei turninternen Phrasengrenzen. Zudem tritt sie primär bei tief fallender Intonationskontur auf. Dass es neben Laryngalisierung auch eine Phonationsform, die er "Verhauchung" nennt (2006: 84), am Phrasenende geben kann und dass diese mit Laryngalisierung kombiniert sein kann, wird lediglich am Rande behandelt.

Bei Peters werden nur wenige phonetische Parameter an Phrasengrenzen untersucht und Phonation zum Beispiel nur in Form von Laryngalisierung bei tief fallenden Intonationskonturen berücksichtigt. Aus diesem Grund soll die folgende Studie einen etwas tieferen Einblick in das Phonationsverhalten deutscher SprecherInnen am Phrasenende, in Abhängigkeit von der Organisation des *turn takings*, geben.

1.3. Untersuchungsfragen

Forschungen wie die eben genannten Studien und erste eigene Beobachtungen anhand spontansprachlicher Daten sind der Ausgangspunkt für die vorliegende Untersuchung, die folgende Untersuchungsfragen im Fokus hat:

1. Spielen Änderungen der Phonation gegen Ende einer Intonationsphrase eine Rolle bei der Organisation von Fortführung oder Wechsel von Redebeiträgen im Deutschen?
2. Im Speziellen: Kann anhand der vorliegenden Daten bestätigt werden, dass Laryngalisierung/ *creak* ein Signal für einen möglichen Redebeitragswechsel ist?
3. Kookkurrieren die finalen Phonationsarten mit anderen sprachlichen Merkmalen?
4. Gibt es Unterschiede zwischen den Sprechern?

[1] "Glottalisierung" wird gelegentlich als alternativer Begriff für Laryngalisierung verwendet, kann aber auch als Oberbegriff für Laryngalisierung und Glottalverschluss verstanden werden.

[2] Ich möchte mich an dieser Stelle bei Dagmar Barth-Weingarten und Richard Ogden dafür bedanken, dass sie mir ihre Tonbeispiele zur Verfügung gestellt haben.

2. Methodisches Vorgehen, Daten

Die Basis der Untersuchung bilden gemäß den Prämissen der interaktionalen Linguistik möglichst natürliche Gespräche (mit gewissen unten genannten Einschränkungen): Es handelt sich um neun deutschsprachige (elizitierte) Alltagsgespräche zwischen jeweils zwei GesprächspartnerInnen. Insgesamt wurden sechs SprecherInnen untersucht, drei Frauen und drei Männer. Die SprecherInnen sind Studierende, DoktorandInnen und ein habilitierter wissenschaftlicher Mitarbeiter im Alter zwischen 24 und 55 Jahren.

Die SprecherInnen wurden vor dem Gespräch darüber informiert, dass ihre sprachliche Interaktion phonetisch analysiert werden soll (aber nicht, dass es im Speziellen um das *turn taking* geht). Weiterhin wurden ihnen universitätsnahe Themen als Gesprächsthemen vorgeschlagen (zum Beispiel die zu diesem Zeitpunkt aktuelle Umstellung auf das neue Studiensystem oder die Einführung von Studiengebühren), woran sie sich aber nicht strikt halten mussten und auch nicht immer gehalten haben; es treten häufiger Passagen mit persönlichen Erlebnissen auf.

Die Gespräche sind so organisiert, dass sowohl Sprecher unterschiedlichen Geschlechts als auch gleichgeschlechtliche Partner miteinander sprechen. In einigen Gesprächen befinden sich die InteraktionspartnerInnen auf derselben Stufe im akademischen Lebenslauf (zum Beispiel zwei Studierende), in anderen auf unterschiedlichen Ebenen (wenn zum Beispiel ein Doktorand mit dem habilitierten Mitarbeiter spricht).

Die Sprecher saßen parallel nebeneinander, konnten sich während des Gesprächs aber nicht sehen, da sie durch eine Posterwand getrennt waren (um die Gestik und Mimik als Hinweise für das *turn taking* auszuschließen). Von jedem Sprecher wurde das akustische Signal mit Hilfe eines Headset-Mikrofons aufgenommen; darüber hinaus zeichnete ein Elektroglottograph (EGG)[3]die Stimmlippenschwingungen auf, um eine genauere Analyse der Phonation vornehmen zu können.

An den Gesprächen wurden diverse phonetische Analysen durchgeführt, unter Berücksichtigung interaktional-linguistischer Verfahren der Gesprächsanalyse und der Kontextualisierungstheorie.

Auswahl der Daten: Die im Folgenden beschriebenen Ergebnisse basieren auf der Analyse von insgesamt 273 Fällen: Es wurden 129 Intonationsphrasen (IPs) am Ende eines Redebeitrags und 144 IPs vor Weiterführung eines Redebeitrags (also beitragsinterne IPs) untersucht. Ein Beitrag wurde dann als beendet erachtet, wenn der Gesprächspartner zu sprechen beginnt und sowohl die Prosodie des neuen Redebeitrags als auch der nachfolgende Gesprächsverlauf zeigen, dass er das

[3] Mit Hilfe eines Elektroglottographen wird der relative Kontakt der Stimmlippen gemessen: Am Hals werden rechts und links vom Kehlkopf jeweils eine (oder zwei) Elektrode angelegt. Von der einen Elektrode wird ein hochfrequenter Wechselstrom mit geringer Voltzahl ausgegeben. An der anderen Seite wird gemessen, wie viel von dem gesendeten Strom ankommt. Die EGG-Signale geben so vor allem die Impedanz des Stromflusses bei den Stimmlippen wieder: Luft bietet mehr Impedanz als das Gewebe, weshalb bei hohem Kontakt der Stimmlippen geringere Impedanz vorliegt und somit mehr bei der Empfänger-Elektrode ankommt. Für weitere Einzelheiten siehe zum Beispiel Marasek (1997).

Rederecht dafür hat (oder sich im Falle von kompetitiven Überlappungen erkämpft hat).

Frage-Antwort-Sequenzen wurden nicht in die Analysen aufgenommen. Die Beschränkung lag weiterhin auf Redebeitragswechseln und -fortführungen, die unproblematisch sind. Das heißt, es sind keine Missverständnisse bezüglich des *turn takings* oder unangemessen lang erscheinende Verzögerungen erkennbar. Ebenso gibt es im nachfolgenden Gesprächsverslauf keine Hinweise darauf, dass die aktuelle SprecherIn direkt nach einem Wechsel versucht, wieder zu Wort zu kommen, was zeigen könnte, dass er oder sie noch weitersprechen wollte. Die Weiterführung gestaltet sich auch nicht schleppend, was ein Hinweis darauf wäre, dass er oder sie eigentlich einen Redebeitragswechsel einleiten wollte. Da eine unvollständige Syntax eine stark weiterweisende Kraft besitzt, war eine weitere Voraussetzung bei der Auswahl der zu analysierenden Redebeitragsfortführungen, dass die IPs syntaktisch an einem möglichen Endpunkt angekommen sind.

Wie diese Auswahlverfahren zeigen, wurde die Reaktion des Gesprächspartners grundsätzlich bei den Analysen mitberücksichtigt. Denn es wird davon ausgegangen, dass die GesprächspartnerInnen ihre Redebeiträge gemeinsam, interaktiv herstellen. Deshalb wurde in Anlehnung an interaktional-linguistische Studien jeder einzelne Fall im Detail im sequenziellen Verlauf des Gesprächs betrachtet, auch wenn später versucht wurde, über statistische Verfahren Regelmäßigkeiten im Auftreten einzelner phonetischer Merkmale zu ermitteln.

Analyseparameter: Die nach den oben genannten Kriterien ausgewählten Intonationsphrasen wurden im Hinblick auf unterschiedliche phonetische, lexikalische, pragmatische und syntaktische Parameter analysiert. Zu den phonetischen Parametern zählen unter anderem die phrasen-finale Intonationskontur, Atmen, Phonation und auf der segmentellen Ebene der Sprache die Aspiration[4] phrasenfinaler Plosive (Verschlusslaute). Sie wurden akustisch und ohrenphonetisch analysiert und ausgewählten Kategorien zugeordnet. So konnten Häufigkeiten ausgezählt (Quantifizierung) und einfache statistische Hypothesen-Überprüfungen mit Hilfe von χ^2–Tests vollzogen werden.

Die Analyse der Phonation wurde auf drei Ebenen durchgeführt: Die Stimmqualität bzw. ihre Veränderungen über die Intonationsphrase hinweg wurden auditiv beurteilt. Darüber hinaus wurden sowohl das akustische Signal als auch das EGG-Signal auf Hinweise für einen nicht-modalen Phonationsmodus geprüft. Diese "audio-visuelle" Methodik bringt einige Schwächen mit sich, wie zum Beispiel eine mangelnde Objektivierung, ermöglicht auf der anderen Seite aber auch die Erfassung minimaler Veränderungen. Ferner kann man so den unterschiedlichen "modalen" Phonationsmodi der einzelnen Sprecher gerecht werden (siehe oben). Denn die objektiv gleichen akustischen Messwerte können bei dem einen Sprecher bereits ein Hinweis auf nicht-modale Phonation sein, bei dem anderen noch innerhalb seiner neutralen Phonation liegen.

Nach einer ersten Sichtung der Phonationsweisen am Ende der IPs wurden vier Kategorien herausgearbeitet: (1) Modale/ neutrale Phonation bis zum Ende der IP, (2) Laryngalisierung (gegebenenfalls gefolgt von einer kurzen modal stimmhaften

[4] Aspiration: Nach der Lösung des für Plosive charakteristischen Verschlusses im Mundraum entweicht die hinter dem Verschluss angestaute Luft geräuschhaft ("h"-artig).

Phase oder einem stimmlosen Konsonanten), (3) Aushauchen und (4) die Kombination von Laryngalisierung und Aushauchen.

Bei der Bestimmung von Laryngalisierung ist Folgendes zu beachten: Sie tritt nicht nur äußerungsfinal auf, sondern spielt im Deutschen auch eine große Rolle als initiales Grenzsignal bei Wörtern, die mit einem Vokal beginnen (auch "harter Stimmeinsatz" genannt) – zum Beispiel in dem Verb *ist*, das man oft am Ende eines Satzes findet. Andererseits ersetzen einige Sprecher Plosive durch einen Glottalverschluss oder Laryngalisierung – wie in dem Wort "finden".[5] Konnte nicht ausgeschlossen werden, dass es sich um eine dieser Varianten handelt, wurde die Laryngalisierung nicht als phrasenfinales Ereignis behandelt.

Neben der Phonetik fanden zum Beispiel das Anhängsel (*question tag*) *ne*, typische als Endsignal verwendete Lexeme wie ein abschließendes *naja*, das Vorhandensein einer übergangsrelevanten Stelle und das Auftreten von parallelen Hörersignalen Berücksichtigung. Übergangsrelevante Stellen wurden danach bestimmt, ob prosodisch[6], syntaktisch, semantisch und inhaltlich ein möglicher Abschlusspunkt erreicht ist. Das Konzept ist mit dem "Complex Transition Relevance Place" von Ford/ Thompson (1996: 153ff.) vergleichbar.

Statistik: Den folgenden Ergebnissen liegen, soweit es die Fallzahlen zulassen, inferenzstatistische Vergleiche zugrunde: Mit Hilfe von χ^2–Tests werden die Häufigkeiten der jeweiligen Phonationsarten pro Ausprägung einer anderen Variable verglichen; zum Beispiel wird geprüft, ob die Verteilung über die vier Phonationskategorien sich signifikant zwischen den beiden Positionen im Gespräch (Ende oder Weiterführung eines Redebeitrags) unterscheidet. Die Signifikanzen werden für zweiseitige (ungerichtete) Hypothesentests angegeben.

In die Statistik gingen nicht alle 273 untersuchten Intonationsphrasen ein, da in seltenen Fällen die Phonation, die Intonation oder das Atmen nicht mit Sicherheit bestimmt werden konnten. Das ist zum Beispiel der Fall, wenn ein Sprecher am Ende seiner Äußerung lachend spricht oder direkt im Anschluss lacht. Der Vergleich von Redebeitragsende und -fortführung in Abschnitt 3.2 basiert daher auf 268 IPs, die in Abschnitt 3.4 angestellten Vergleiche auf 265 IPs.

3. Ergebnisse

3.1. Nicht-modale Phonation am Ende von Intonationsphrasen

Über Redebeitragswechsel und Redebeitragsfortführungen gemittelt enden mehr als die Hälfte aller Intonationsphrasen nicht-modal. Laryngalisierung ist mit 22% die häufigste nicht-modale Phonationsart; reines Aushauchen kommt eher selten vor.

Einige typische turn-final auftretende lexikalische Einheiten sind davon besonders betroffen: Kein Fall von abschließendem *ja* oder *naja* bleibt bis zum Ende

[5] Beide Arten der Glottalisierung im Deutschen wurden von Klaus Kohler detailliert untersucht; siehe zum Beispiel Kohler (1994). Zur Rolle des wortinitialen harten Stimmeinsatzes bei dem Anschluss neuer Turnkonstruktionseinheiten siehe Szczepek Reed (2013).

[6] Die Prosodie, die hier berücksichtigt wurde, umfasst den Tonhöhenverlauf (Intonationskontur), Lautstärke, Sprechgeschwindigkeit und Akzentuierung.

modal; besonders Laryngalisierung mit anschließendem Aushauchen, oder etwas seltener reines Aushauchen sind hier typische Phonationsarten. Ein ähnliches Bild ergibt sich für das Anhängsel (*question tag*) *ne*, das in diesem Korpus am Ende und vor der Weiterführung des Redebeitrags vergleichbar häufig vorkommt.

Vor allem eine Form der Phonationsänderung wird im Folgenden eine große Rolle spielen: die Kombination von Laryngalisierung und Aushauchen – in den meisten Fällen sequentiell in dieser Reihenfolge, gelegentlich aber auch in Form von *whispery creak* (Knarren mit gleichzeitigem Geräuschanteil aufgrund des Flüsterns). Da bisher nicht geklärt werden konnte, ob Laryngalisierung oder Aushauchen auditiv überwiegt, wird diese Kategorie separat behandelt und nicht zu Laryngalisierung hinzugezählt.

Eine besondere Variante des Aushauchens nach Laryngalisierung ist eine Art Nachwelle im Oszillogramm. Sie besteht meist nur aus drei bis acht Perioden. Die Wellenform ist fast sinusförmig und von geringer Amplitude. Im EGG-Signal gibt es entweder keinen Hinweis auf Stimmlippenschwingungen oder es ist nur ein minimaler Ansatz von Schwingung zu sehen. Von Redi/ Shattuck-Hufnagel (2001) werden vergleichbare Signale als eine Variante von Glottalisierung beschrieben und als "glottal squeak" bezeichnet. Sie sprechen von *squeak* – Quietschen – weil sie nach der Laryngalisierung einen Sprung in eine hohe Tonhöhe wahrnehmen. Misst man die Grundfrequenz (F0), ist sie in der Tat meistens höher als vor der Laryngalisierung. Die globale Intonationskontur wird jedoch nicht unbedingt als fallend-steigend oder gar steigend gehört, sondern häufig als fallend.

Die Änderung der Phonation ist in der vorliegenden Studie von eher kurzer Reichweite und häufig nicht stark ausgeprägt. Dies gilt vor allem für Aushauchen. Aber auch Laryngalisierung ist zu 61% auf das finale Lautsegment oder die finale Silbe begrenzt. Lediglich die Kombination von Laryngalisierung und Aushauchen erstreckt sich oft über mehr als eine Silbe. In den deutschen Gesprächen, die Peters (2006) analysiert hat, ist die Dauer der Laryngalisierung vergleichbar kurz; am Ende eines Turns etwas länger als innerhalb eines Turns. Im Finnischen scheint sich die Phonation bereits weiter vor dem Ende der Phrase zu verändern; denn Ogden (2001, 2004) hat nur solche Veränderungen markiert, die mindestens eine Silbe lang sind. Nicht-modale Phonation auf den letzten zwei, drei Silben ist dort keine Seltenheit. In den von Barth-Weingarten (2009) beschriebenen englischen Telefongesprächen überschreitet die Laryngalisierung zwar ebenso selten eine Silbe, aber sie ist viel stärker ausgeprägt als in den hier beschriebenen Daten.

3.2. Vergleich von Redebeitragswechsel und Redebeitragsfortführung

Vergleicht man die Intonationsphrasen am Redebeitragsende mit denjenigen vor Fortführung, lassen sich die vier Phonationsarten bezüglich der Verteilung höchst signifikant unterscheiden, χ^2 (3, N = 268)= 21,00, $p < 0{,}001$. Der Unterschied zwischen den beiden Positionen liegt vor allem in der Kategorie Laryngalisierung mit Aushauchen, χ^2 (1, N = 268) = 16,36, $p < 0{,}001$. Die Kombination dieser beiden Phonationsarten ist sehr selten, wenn der Sprecher danach weiter spricht, tritt jedoch in mehr als jeder vierten IP am Redebeitragsende auf. Laryngalisierung oder

Aushauchen alleine kommen in beiden gesprächsstrukturellen Positionen ähnlich häufig vor. Die Verteilung über die vier Phonationsarten ist im folgenden Balkendiagramm (Abbildung 3) dargestellt:

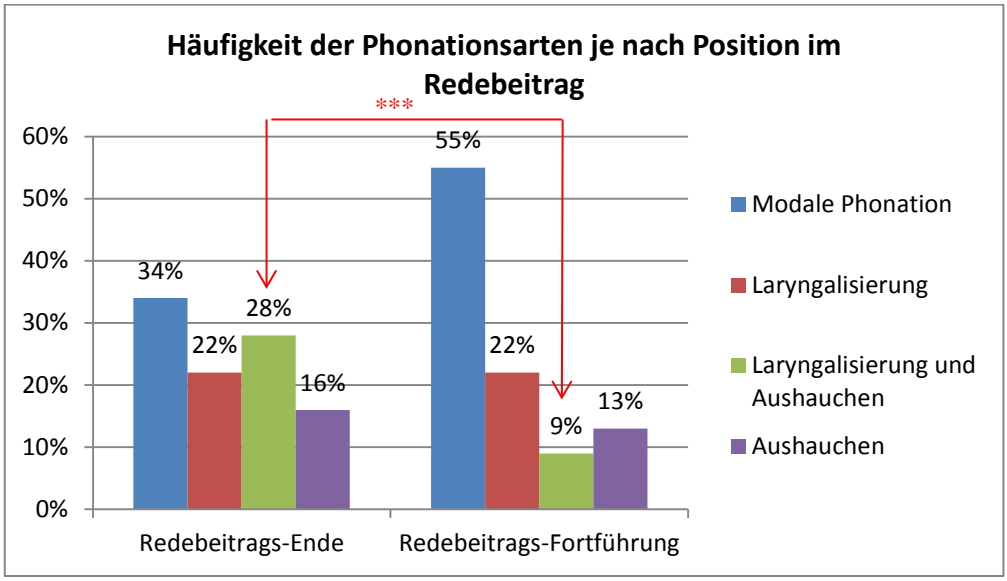


Abbildung 3: Häufigkeit der Phonationsarten je nach Position im Redebeitrag

Nun soll für jede nicht-modale Phonationskategorie vor einem Redebeitragswechsel ein Beispiel gegeben werden. Die Ausschnitte sind nach den Konventionen von GAT 2.0 (vgl. Selting et al. 2009) transkribiert; die relevante Zeile ist jeweils durch einen Pfeil davor gekennzeichnet.

(1) Gespräch zwischen MD_{md} und ND_{ws}, finale Laryngalisierung

(Sprecherin ND_{ws} glaubt nicht, dass man mit dem Bachelor-Abschluss eine Arbeit in der Wirtschaft findet. MD_{md} hingegen argumentiert, dass in der Wirtschaft ein fachliches Schmalspur-Wissen ausreichend für einen Berufseinstieg ist.)

```
    1   ND_ws:   also dEs als nen `ABschluss anzusehen,=
    2            =mit dEm du: °h (0.4)
    3            in die: ARbeitswelt; (0.1)
    4            <<creaky> Eintauchen> KANNST,
    5            °h und_äh? (0.1)
    6            des wär_ja schön_un_gUt wenn das ´GInge;=
    7            =aber ich glaube nIch dess
    8            dess des WIRklich- °h (0.1)
->  9            [äh !CHAN!cen] <<creaky>birgt>.
   10   MD_md:   [t_h°         ]
   11   ND_ws:   °h[h]
   12   MD_md:     [i]ch WEISS nich;=
   13            =also ich hab Irgendwo ma_n beRICH_jesehn-
```

In dem ersten Ausschnitt unterhält sich eine Studentin (ND_{ws}) mit einem Doktoranden (MD_{md}) über die Chancen des neu eingeführten Bachelor-Studiengangs.

Die Studentin argumentiert emphatisch gegen das neue System. Zeile 6 stellt zwar einen syntaktisch vollständigen Satz dar, die Formulierung macht jedoch ein *aber* erwartbar, so dass eine Weiterführung wahrscheinlich ist. Die Sprecherin beendet diese IP außerdem mit modaler Phonation und schließt ihren *aber*-Satz in Zeile 7 schnell an (das *aber* ist am Wortanfang glottalisiert). Nachdem sie sich so das Rederecht weiter gesichert hat, beendet sie nach kurzem Häsitieren und Einatmen ihren Redebeitrag schließlich mit einer Zusammenfassung ihrer Meinung. In Zeile 9 weisen Prosodie, Syntax und semantisch-inhaltlicher Verlauf auf eine übergangsrelevante Stelle hin. Die Sprecherin endet mit tief fallender Intonation sowie leichter Dehnung und Laryngalisierung auf der finalen Silbe. Der Plosiv am Ende des Wortes ist darüber hinaus deutlich aspiriert. In einer Varietät des Englischen, dem Tyneside English, haben Local et al. (1986) Aspiration phrasen-finaler Plosive als turn-finales Merkmal beschrieben.

Bereits simultan mit der Weiterführung von ND_{ws} atmet der Doktorand ein (Zeile 10), lässt sie dann aber ihren Redebeitrag zu Ende führen. Direkt nach ihrem anschließenden Einatmen beginnt er seinen Redebeitrag, ohne dass die Studentin einen Hinweis darauf gibt, dass sie eigentlich noch nicht am Ende angekommen ist.

Die folgende Abbildung stellt einen kleinen Ausschnitt des Sprachsignals des Wortes *birgt* (Zeile 9) dar. Dieses und die folgenden Bilder sind wie folgt aufgebaut: Ganz oben im Bild ist das Signal des Elektroglottographen (EGG) abgebildet: Die Annäherung der Stimmlippen (Adduktion) zeigt sich in der (meist steilen) Aufwärtsbewegung der Wellenform; verläuft die Welle nach unten, lösen

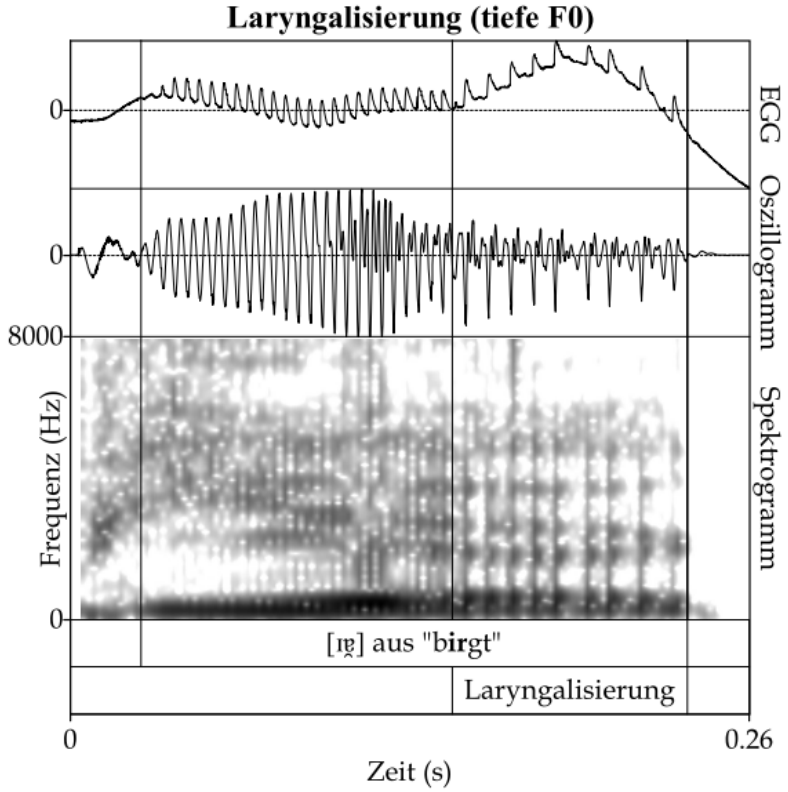


Abbildung 4: Sprachsignal mit Laryngalisierung am Ende des Wortes *birgt*

sich die Stimmlippen wieder voneinander (Abduktion).[7] Die beiden Darstellungen darunter repräsentieren das akustische Signal: In der Mitte das Oszillogramm und unten das Spektrogramm. Im Spektrogramm wird auf der x-Achse die Zeit, auf der y-Achse die Frequenz und im Schwärzungsgrad die Intensität (je dunkler, umso mehr Intensität) wiedergegeben.

Anhand des Signals in Abbildung 4 kann man erkennen, dass sich die Stimmlippen während der Laryngalisierung in dem Wort *birgt* nur langsam auseinander bewegen (die beiden letzten Konsonanten des Wortes, die nicht abgebildet sind, sind stimmlos). Mit der verzögerten Abduktion geht eine Verlangsamung der Schwingungsfrequenz und somit der Grundfrequenz einher. Beides sind typische Merkmale von Laryngalisierung. Die starke Absenkung der F0 äußert sich im Spektrogramm im stetig größer werdenden Abstand zwischen den Impulsen. Der nächste Gesprächsausschnitt ist ein Beispiel für finales Aushauchen am Ende eines Redebeitrags (in Zeile 5).

(2) Gespräch zwischen MD_{md} und MP_{mw}, finales Aushauchen

(Sprecher MP_{mw} erzählt, dass in seinem Fach bisher nur ein Bachelor-Studiengang eingerichtet wurde, der Master noch in Planung ist. Er ist der Meinung, dass die Umstellung der Studiengänge in Bachelor und Master nicht viel Neues bringt. Die Studierenden müssen auf jeden Fall die Grundlagen lernen und im Masterstudiengang ebenso wie früher im Hauptstudium eine Spezialisierung und Vertiefung vornehmen.)

```
    1   MPmw:   un_das WIRD ja auch;
    2           °hh (0.1) mit SIcherheit für den: äh:
                mAsterstudiengang nachher nich(t) `ANders sein,
    3           als es nach der ZWIschenprüfung.
    4           °h aus MEIner sicht in_der; (0.1)
->  5           °hh in_der äh_h° im magIsterstudiengang der FALL
                <<aushauchend> war>. h°
    6           (0.4)
    7   MDmd:   ja ja.
    8           (0.4)
    9           [°h                    ]
   10   MPmw:   [<<pp, aushauchend> hm>]
```

Dieser Ausschnitt stammt aus einem Gespräch zwischen einem habilitierten wissenschaftlichen Mitarbeiter (MP_{wm}) und dem Doktoranden aus Ausschnitt 1. Der habilitierte Mitarbeiter argumentiert, dass die Umstrukturierung der Studiengänge nichts an den Inhalten geändert hat, die die Studierenden lernen müssen. Er spricht stockend, mit vielen Atempausen. Zeile 2 ist ein syntaktisch vollständiger Satz, der auch inhaltlich ein Ende markieren könnte. Prosodie und Phonation weisen jedoch auf eine Weiterführung hin: Die finale Intonationskontur verläuft leicht fallend-steigend und die Phonation ist bis zum Ende modal (beziehungsweise für diesen Sprecher neutral).

[7] Beim Sprechen bewegt sich der Kehlkopf häufig leicht nach oben und unten. Diese vertikale Kehlkopfbewegung äußert sich in langsamen Verschiebungen der Welle um die Nulllinie herum und muss hier nicht beachtet werden.

In Zeile 5 führt MP_{mw} schließlich seinen Satz zu Ende. Dieses Satzende geht mit diversen phonetischen Merkmalen einher, die mögliche Endsignale darstellen: Die Intonationskontur fällt bis in die Tiefe und die Lautstärke nimmt sukzessive ab. Auf den letzten beiden Silben weist die Phonation zunehmend mehr Rauschanteile auf, bis zu vollständigem Flüstern beziehungsweise einem stimmlosen Vokal am Ende von *war*. Das Aushauchen geht direkt in hörbares Ausatmen über. Das ist für diesen Sprecher, der allgemein eine leicht raue Stimme hat, ein typischer phrasen-finaler Phonationsmodus.[8]

Prosodisch, syntaktisch und inhaltlich ist MP_{mw} in Zeile 5 an einem möglichen Abschlusspunkt angekommen. Nach einer kurzen Pause von 0,4 Sekunden reagiert MD_{md} zuerst nur mit einem Rezipientensignal (*ja ja*), beginnt dann aber kurz darauf seinen Redebeitrag (überlappend mit einem ganz leisen *hm* von MP_{wm}).

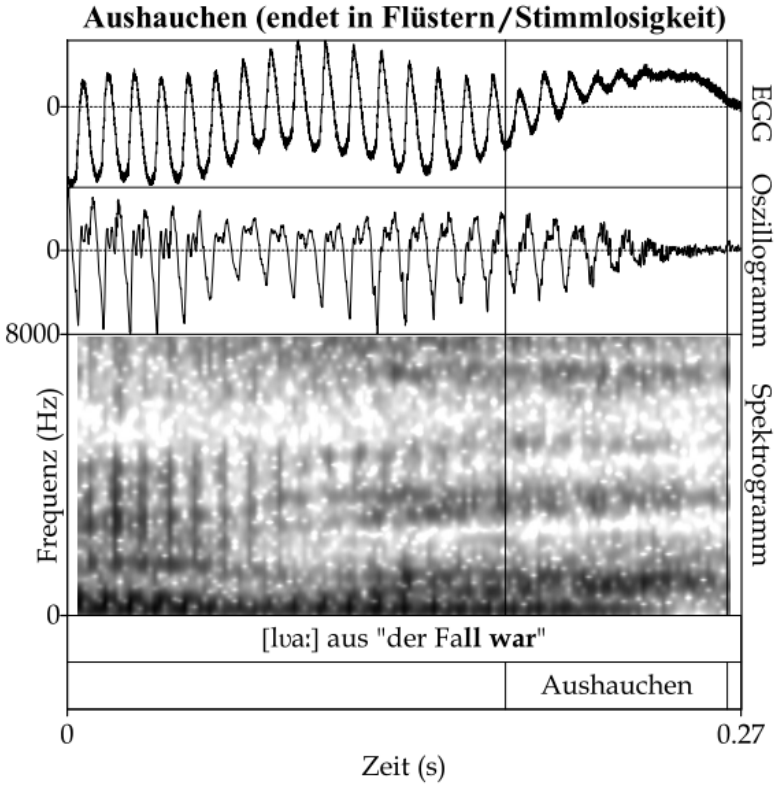


Abbildung 5: Sprachsignal mit Aushauchen am Ende des Wortes *war*

Abbildung 5 zeigt das Aushauchen von Zeile 5. Am Ende des Wortes *war* weist das EGG-Signal eine zunehmend sinusförmigere und amplitudenschwächere Wellenform auf: Die Stimmlippen werden immer schwächer adduziert und berühren sich nur noch sehr kurz, wie bei starker Behauchung, bis am Ende keine Berührung mehr zu erkennen ist.

Der folgende Gesprächsausschnitt (3) ist ein Beispiel für die Folge von Laryngalisierung und Aushauchen am Ende eines Redebeitrags:

[8] Er haucht auch auf dem Rezipientensignal "hmhm", auf das nichts weiter folgt, sehr oft in dieser Art aus (siehe zum Beispiel Transkript 4, Zeilen 2 und 5).

(3) Gespräch zwischen ND_{ws} und CK_{ms}, Laryngalisierung mit anschließendem Aushauchen

(Sprecher CK_{ms} sieht bei der Umstellung der Vorlesungszeiten und der Modularisierung der Studiengänge ein Problem: Die Umstellung wird nicht einheitlich durchgeführt und gewisse äußere Faktoren wie veränderte Busnutzung werden nicht berücksichtigt.)

```
    1   CKms:   ich mEin we_manˀ°h
    2           heut_auch tEilweise doZENten <<creaky> fragt>,
    3           °hh wIe das jetz (0.3) !IS! mitˀ (0.3) Ihrem
                jEweiligen FACH, (0.7)
    4   NDws:   °h (0.2) °h   (0.1)
    5   CKms:   !WANN! das modulariSIERT wird- (0.1)
    6   NDws:   m_hm (0,1)
    7   CKms:   ((schnalzt)) oder wie dasˀgAnze geSCHEhen soll,
    8           <<aushauchend>da:>(1.0) krieg_ma halt AU net_f:
                <<pp>wˀ>wIrklich ne sInnvolle
                ANTwort_<<aushauchend>s[o>.]
    9   NDws:   [ja:;]
   10           (0.1)
   11   CKms:   ((lacht schnaubend durch die Nase))
-> 12           [°hh (0.4) <<p, final creaky, aushauchend> naJA;>]
   13           hh°
   14   NDws:   [°hhh                                          ]
   15           (0.3)
   16   CKms:   ((schluckt))
   17   NDws:   also Ich glaub_die geFAHR is_halt dess_ma:-ˀ
```

In diesem Beispiel spricht die Studentin ND_{ws} aus Auszug 1 mit einem männlichen Studenten (CK_{ms}). Der Student, der mit dem Ablauf der Umstellung der Studiengänge noch etwas unzufrieden ist, äußert in den Zeilen 1-8 eine Kritik an den Dozenten, die auch keine Hilfe für die Studierenden sind. In Zeile 8 ist im Hinblick auf Prosodie, Syntax und semanto-pragmatische Aspekte ein möglicher Abschlusspunkt erreicht; so wird der Sprecher auf der Ebene der Prosodie zum Beispiel leiser und die Intonationskontur fällt bis in die Tiefe. Das finale *so* wird sogar leicht aushauchend realisiert. ND_{ws} reagiert jedoch nur mit einem simultanen Rezipientensignal (*ja* in Zeile 9). Das anschließende leichte Lachen von CK_{ms} kann einerseits unterstreichen, dass er nicht mehr weitersprechen möchte, kann aber auch Verlegenheit ausdrücken, da er weiß, dass die Aufnahmeleiterin ebenfalls als Dozentin in seinem Nebenfach tätig ist.

Nach einer kurzen Pause, in der seine Gesprächspartnerin keinen Hinweis gibt, dass sie selbst etwas sagen möchte, schiebt er die typische Abschluss-Formulierung *naJA* nach (Zeile 12). Das *naJA* wird mit fallender Intonation und leise realisiert; es ist am Ende ganz leicht laryngalisiert und anschließend deutlich ausgehaucht. Direkt im Anschluss atmet der Sprecher durch die Nase aus.

Simultan mit dem Nachschub atmet ND_{ws} ein und beginnt nach einer kurzen Pause von 0,3 Sekunden etwas zögerlich ihren Redebeitrag.

Abbildung 6 stellt das Signal am Ende des Wortes *naJA* dar. In der Mitte des Bildes sind kurze Irritationen im Schwingungsverhalten der Stimmlippen zu erkennen (diesmal ohne verzögerte Abduktionsphase); anschließend schwingen sie aus. Im letzten Teil weist das EGG-Signal nur noch auf minimale Schwingungen hin.

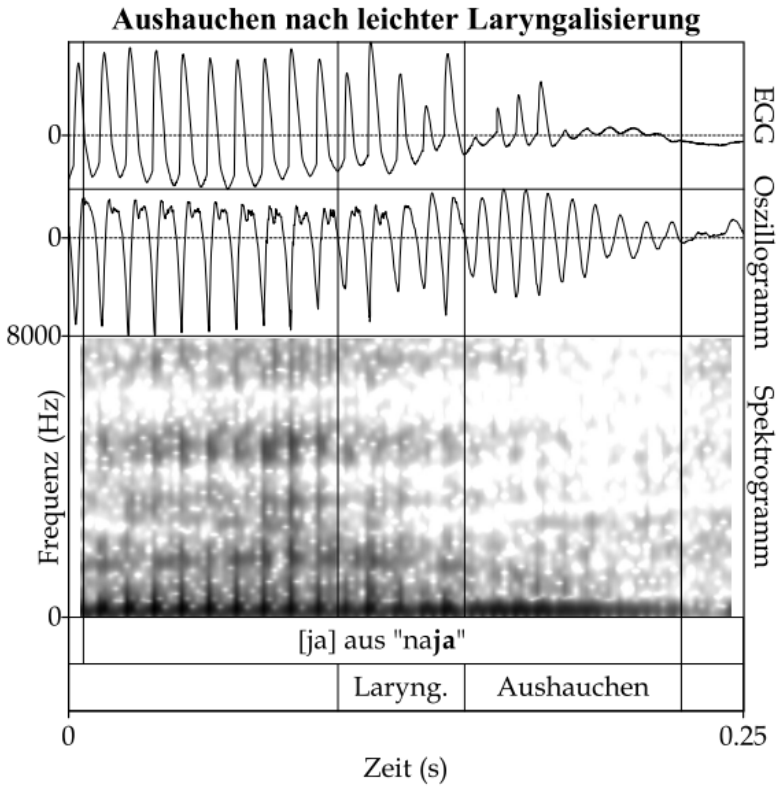


Abbildung 6: Sprachsignal mit Laryngalisierung und Aushauchen auf dem Wort *naJA*

An früherer Stelle wurde bereits erwähnt, dass es eine Sonderform von Laryngalisierung in Kombination mit Aushauchen gibt: Der aushauchende Teil zeigt sich im Oszillogramm als kurze Schallwelle von niedriger Amplitude, während im EGG-Signal nur minimale Schwingungen der Stimmlippen zu sehen sind. Diese Form tritt in den oben beschriebenen Gesprächsausschnitten nicht auf. Abbildung 7 soll verdeutlichen, wie ein Signal dieser Art aussehen kann. Es handelt sich dabei um ein abschließendes *HJA* aus dem Gespräch der Doktorandin SL_{wd} mit dem wissenschaftlichen Mitarbeiter MP_{mw}. Es bildet den Abschluss einer von mehreren Häsitationen und Nachschüben geprägten Äußerung. Transkript 4 stellt den kurzen Vorkontext dar. Im Anschluss an diese Äußerung beginnt MP_{mw} einen neuen Redebeitrag in Form einer Frage, die einen Themenwechsel einleitet. Da es lediglich um die Verdeutlichung dieser Sonderform von Laryngalisierung mit Aushauchen geht, wird an dieser Stelle auf eine ausführliche Fallbeschreibung verzichtet.

(4) Gespräch zwischen SL_{wd} und MP_{mw}, Laryngalisierung mit anschließendem Aushauchen

(Die Doktorandin SL_{wd} berichtet, wie sich mit der Umstellung in den Bachelor-Studiengang in ihrem Fachbereich der Druck für die Studierenden erhöht hat. Sie konkretisiert den Druck: Es gibt strengere Kontrollen und alle Noten gehen in den Abschluss ein.)

```
   1   SLwd:   w[eil die:                ] äh halt strEnger:ʔ wie
               gesagt kontrolLIERT werden-=
   2   MPmw:   [<<final aushauchen> m_hm>]
   3   SLwd:   =weil alle: ʔnOten auch EINgehen-
   4           und äh[m:- h°                    ]
   5   MPmw:         [<<final aushauchend> m_hm>]
-> 6   SLwd:   (0.1) ((schmatzt)) <<final creaky u. aushauchend>
               HJA.>
   7           (0.2)
   8   MPmw:   °h und_sie ham JETZ die:-
```

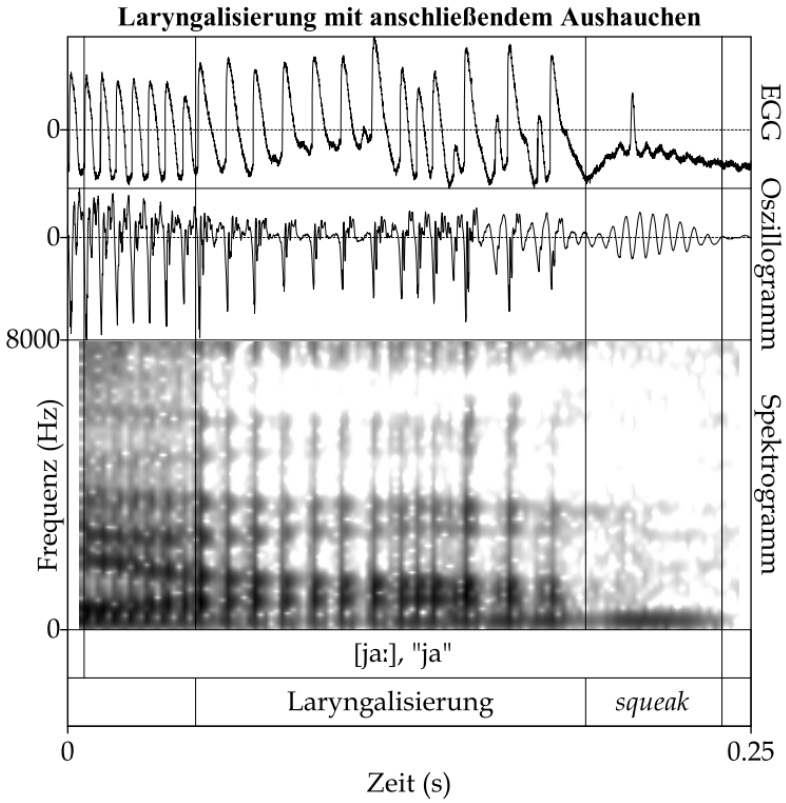


Abbildung 7: Laryngalisierung mit anschließendem Aushauchen in Form von *glottal squeak*

3.3. Analyse abweichender Fälle

Auch wenn der Unterschied im Vorkommen nicht-modaler Phonation zwischen Ende und Fortführung eines Redebeitrags signifikant ist, enden dennoch auch 45% der untersuchten IPs vor dem Weitersprechen nicht-modal. Betrachtet man diese Fälle genauer, zeigen sich einige systematische Auffälligkeiten.

Zum Beispiel bedeutet die Tatsache, dass der Gesprächspartner im Anschluss an die Äußerung keinen eigenen Beitrag begonnen hat, nicht zwangsläufig, dass der aktuelle Sprecher weitersprechen wollte. Es kann trotzdem eine übergangsrelevante Stelle (TRP) vorliegen, ein Angebot zum Redebeitragswechsel, das vom Partner nicht aufgenommen wurde.

Ein Vergleich der Phonation vor Redebeitragsweiterführung je nachdem, ob ein TRP vorhanden ist oder nicht, ergibt signifikante Unterschiede: Nicht-modale Phonation ist an übergangsrelevanten Stellen signifikant häufiger zu finden (χ^2 (3, N = 139) = 13,79, $p < 0{,}05$), vor allem Laryngalisierung mit Aushauchen. Laryngalisierung, mit oder ohne anschließendes Aushauchen, kommt an Stellen, die keine TRPs darstellen, eher selten vor.

Die Fortführung eines Redebeitrags trotz nicht-modaler Phonation könnte andererseits auch dadurch erklärt werden, dass der Gesprächspartner zwar eine Reaktion gezeigt hat, aber nur in Form eines Hörersignals. In der Tat treten Hörersignale (Rezipientensignale oder auch kurze Hintergrundkommentare) nach Laryngalisierung, vor allem, wenn ihr ein Aushauchen folgt, häufiger auf als nach modal endenden Intonationsphrasen (nach 69% der Fälle von Laryngalisierung mit anschließendem Aushauchen folgt ein Hörersignal).

Das häufige Auftreten nicht-modaler Phonation an übergangsrelevanten Stellen und das erhöhte Vorkommen von Hörersignalen nach einem Wechsel in eine andere Phonationsart legen nahe, dass der Übergang in eine nicht-modale Phonation am Ende von Phrasen nicht willkürlich ist und wahrscheinlich vom Interaktionspartner auch nicht als willkürlich empfunden wird.

Dennoch sollte erwähnt werden, dass in den vorliegenden Daten ein Drittel der Intonationsphrasen vor Weiterführung, die nicht-modal enden, durch keines der beiden Merkmale erklärt werden kann: Diese IPs befinden sich nicht an einer übergangsrelevanten Stelle und gehen ebenso nicht mit einem Hörersignal einher.

3.4. Kookkurrenzen mitanderen Faktoren

Die Phonation darf selbstverständlich nicht losgelöst von anderen sprachlichen Phänomenen betrachtet werden. Die folgenden Beispiele zeigen, dass sie eng mit anderen phonetischen Ereignissen verbunden ist.

Die wichtigste Kookkurrenz im Englischen wurde bereits im Eingangszitat von Laver (1980: 126) erwähnt: Fallende Intonationskonturen, vor allem tief fallende, gehen häufig mit Laryngalisierung (ohne oder mit anschließendem Aushauchen) einher. Zwei Drittel der Intonationsphrasen mit fallender Kontur enden nicht-modal; vor allem mit Laryngalisierung, sowohl einfacher Laryngalisierung als auch mit anschließendem Aushauchen. Auch in Peters (2006) treten die Belege für Laryngalisierung primär bei tief fallenden Melodiemustern auf.

Bei steigender Intonation hingegen überwiegt eine modale Phonation bis zum Phrasenende; gelegentlich kommt es zu Aushauchen.

Abbildung 8 zeigt die Verteilung der vier Phonationskategorien bei fallenden und steigenden Tonhöhenbewegungen. Die deutliche Präferenz für Laryngalisierung (gegebenenfalls mit nachfolgendem Aushauchen) bei fallenden und für modale Phonation bei steigenden Konturen wird hier sehr deutlich.

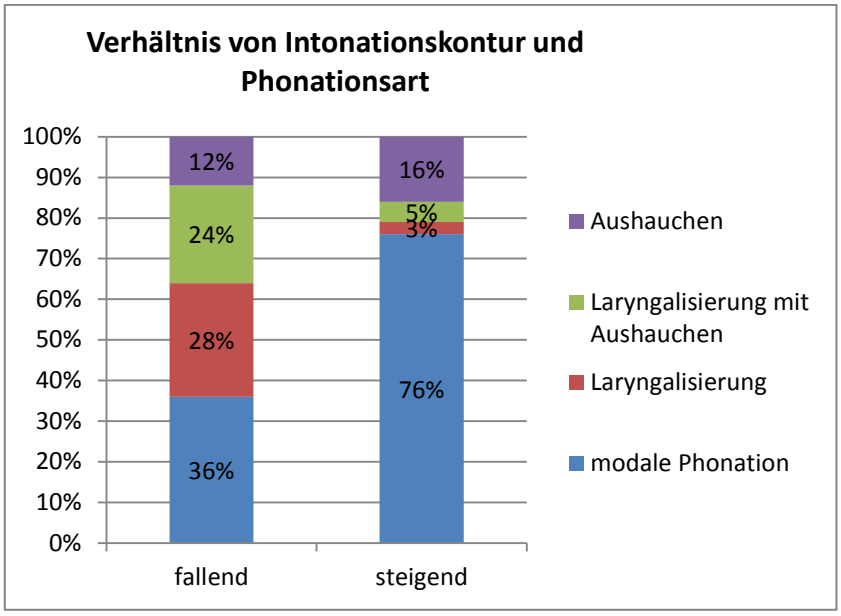


Abbildung 8: Verteilung der vier Phonationsarten bei fallenden und steigenden Intonationskonturen

Nur in seltenen Fällen tritt Laryngalisierung bei hoher F0 auf, zum Beispiel in Form eines diplophonen Signals. Bei Diplophonie schwingen zwei Teile der Stimmlippen in zwei unterschiedlichen Frequenzen. Ein Beispiel ist in Transkript 3, Zeile 2 zu hören: Sprecher CK_{ms} endet mit einer leicht steigenden Intonation und Diplophonie auf dem letzten Wort *fragt*. In Abbildung 9 ist diese Diplophonie

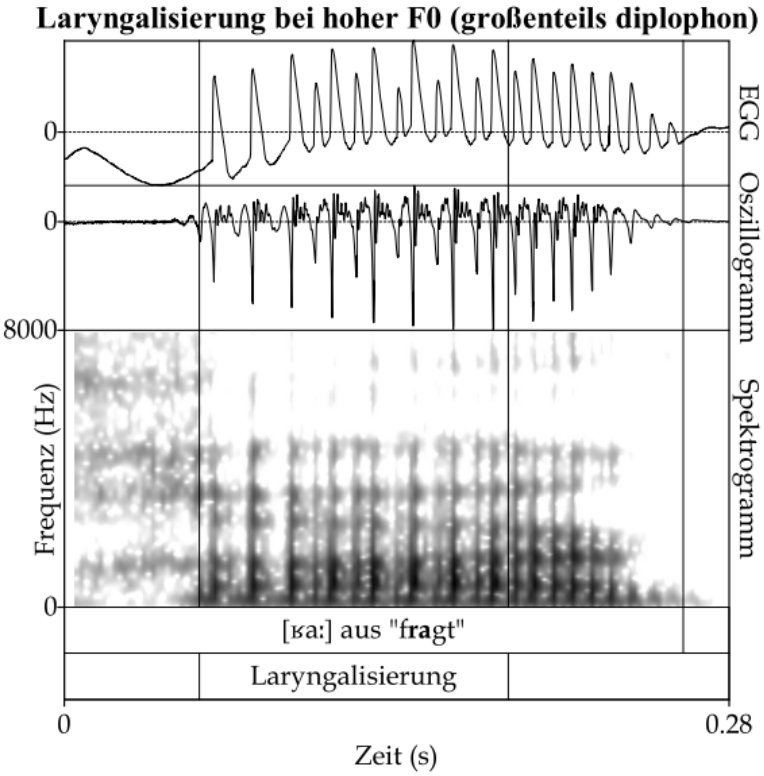


Abbildung 9: Sprachsignal mit Laryngalisierung bei hoher F0

an dem Wechsel eines kleineren und eines größeren Peaks im EGG-Signal zu erkennen.

Das Auftreten von *creak* im Zusammenhang mit steigender Intonation wird ebenso von Ogden (2004: 56) an einem Beispiel im Finnischen demonstriert und von Redi/ Shattuck-Hufnagel (2001: 426) für das gelesene Amerikanisch-Englische beschrieben.

Eine weitere Bündelung von prosodischen Ereignissen betrifft das Atmen. Vor allem Ausatmen tritt nach Aushauchen oder Laryngalisierung mit Aushauchen häufiger auf als nach den anderen beiden Phonationsarten. Nach 76% der laryngalisierten und nach 70% der modal endenden IPs ist kein Atmen im Anschluss hörbar; hingegen folgt auf 32% der rein ausgehauchten und 22% der nach Laryngalisierung ausgehauchten Phrasen ein Ausatmen; ähnlich häufig wird dort im Anschluss eingeatmet.

3.5. Unterschiede zwischen den Sprechern

Ob eine Intonationsphrase modal oder nicht-modal endet, ist außerdem sprecherabhängig. Ein Vergleich zwischen den fünf Sprechern, die an mehr als einem Gespräch beteiligt waren, bringt höchst signifikante Unterschiede für Laryngalisierung mit Aushauchen (χ^2 (4, N = 252) = 21,21, $p < 0{,}001$) und signifikante Unterschiede für Aushauchen (χ^2(4, N = 252) = 14,85, $p < 0{,}05$) zum Vorschein. So bevorzugt zum Beispiel die Studentin ND_{ws} aus Gesprächsbeispiel 1 eher Laryngalisierung mit anschließendem Aushauchen, der Student CK_{ms} aus Gesprächsausschnitt 3 Laryngalisierung (sowohl am Ende als auch vor Weiterführung des Redebeitrags!). Der wissenschaftliche Mitarbeiter MP_{mw}, von dem das Beispiel für Aushauchen entnommen ist, haucht andererseits weit mehr aus als alle anderen Sprecher; dies passt zu seiner Grundstimmqualität, die allgemein etwas rau ist.

Peters (2006) beschreibt ebenfalls interindividuelle Unterschiede in seinen Gesprächsdaten. Untersuchungen zu äußerungsfinaler Laryngalisierung bei gelesenen Daten (zum Beispiel Redi/ Shattuck-Hufnagel 2001, Klatt/ Klatt 1990 oder Slifka 2006) erwähnen ebenfalls mehr oder weniger stark ausgeprägte Unterschiede zwischen den Sprechern.

Aber die untersuchten SprecherInnen unterscheiden sich nicht nur darin, welche Phonationsart sie am Ende von Intonationsphrasen am häufigsten verwenden. Oben wurde gezeigt, dass über alle Sprecher gemittelt ein signifikanter Unterschied in der Häufigkeit von Laryngalisierung mit Aushauchen zwischen Redebeitragsende und -fortführung besteht; dieser Unterschied ist aber nicht bei jedem Sprecher gleichermaßen stark ausgeprägt. Während die eben genannte Studentin ND_{ws} am Redebeitragsende deutlich mehr Laryngalisierung mit Aushauchen produziert, ist diese Phonationskategorie bei den Sprechern MD_{md} und MP_{mw} nur geringfügig stärker am Ende von Redebeiträgen vertreten.

4. Diskussion

4.1. Physiologisches Erklärungsmodell

Ein allgemeines Nachlassen der Spannung im pulmonalen und laryngalen Bereich könnte eine Erklärung für die Phonationsänderungen bieten:

- Laryngalisierung: Wenn der subglottale Druck aufgrund sich entspannender Atemmuskeln und gleichzeitig die innere Spannung der Stimmlippen und die Spannung der Kehlkopfmuskulatur nachlässt, brauchen die Stimmlippen länger, bis sie auseinander gesprengt werden. Zudem unterstützen diese Bedingungen ein unregelmäßiges Schwingen der Stimmlippen.
- Aushauchen: Wenn die Spannung der Kehlkopfmuskeln, die die Stimmlippen adduzieren, zu schwach ist, können sie nicht mehr effizient schwingen. Berühren sich Stimmlippen nur noch teilweise, klingt die Stimme behaucht; kommt es zu keinem Kontakt mehr wie bei stimmlosen Lauten, spricht man von Flüstern. In der schwächsten Form des Aushauchens – dem Auslaufen der Schallwelle – nimmt der Kontakt der Stimmlippen und die Schwingungsamplitude stetig ab.

Ein allgemeiner Trend der abnehmenden Lautstärke und F0 zum Ende einer Äußerung passt zu der These einer nachlassenden Spannung. Parallele Ergebnisse zeigen Untersuchungen zur Gestik oder Körperhaltung beim *turn taking*. Duncan (1972) nennt zum Beispiel als Hinweise zur Abgabe des Turns die Beendigung und die Entspannung von Handgesten.

Wenn man davon ausgeht, dass die Ursache für äußerungsfinale Phonationsänderungen in einer allgemeinen Entspannung liegt, ist die Frage naheliegend, ob es sich um ein universelles Phänomen handelt. Während Phonationsmodi, die dem hier beschriebenen Aushauchen entsprechen, nicht häufig in der Literatur genannt werden, wird finale Laryngalisierung für diverse Sprachen beschrieben (wenn auch nicht zwangsläufig im Zusammenhang mit *turn taking*, sondern allgemeiner als äußerungsfinales Ereignis). Henton (1989) schreibt zum Beispiel, dass sie nicht nur bei der Analyse des Britischen Englisch äußerungsfinalen *creak* gefunden hat, sondern auch im *American English* und im Französischen. Sie kommt deshalb zu folgendem Schluss: "This phenomenon, coupled with Hyman's (1988) recent results for West African languages, may make intonational creaky voice a candidate for a phonetic universal."

Eventuelle sprachen-spezifische Unterschiede im Vorhandensein oder in der Ausprägung nicht-modaler Phonation könnten dadurch begründet werden, dass es in bestimmten Sprachen zu einer konventionalisierten Verstärkung oder Abschwächung des natürlichen Hangs zur Entspannung kommen kann. Bei den finnischen Sprechern, die Ogden (2001, 2004) untersucht hat, könnte die Abschwächung weniger ausgeprägt sein als bei den hier und in Peters (2006) untersuchten deutschen Sprechern.

Einige Funde nicht-modaler Phonation können jedoch schwer durch Entspannung erklärt werden: Aushauchen kommt häufig mit steigender Intonation vor, gelegentlich sogar Laryngalisierung. Für eine hohe Grundfrequenz wird norma-

lerweise eine erhöhte laryngale Spannung benötigt. Diese Fälle können durch andere physiologische Prozesse interpretiert werden:

- Laryngalisierung: Es gibt Probleme mit der schnellen Spannung und Dehnung der Stimmlippen, um die hohe F0 zu erreichen. Folge ist das unregelmäßige Schwingen der Stimmlippen.
- Aushauchen: Es schwingen lediglich die Randkanten der Stimmlippen, ähnlich wie bei einer Falsettstimme. Wie bei Falsett zeichnet sich das akustische Signal von Aushauchen durch wenige Obertöne aus.

4.2. Behandlung der Kategorie Laryngalisierung mit Aushauchen

Der größte Unterschied zwischen Intonationsphrasen am Redebeitragsende und solchen vor Redebeitragsweiterführung besteht in der Kategorie Laryngalisierung in Kombination mit Aushauchen; ebenso zwischen übergangsrelevanten Stellen und Stellen, an denen kein TRP vorliegt. Ist es jedoch überhaupt gerechtfertigt, sie als eigenständige Form der Phonationsänderung aufzufassen? Oder handelt es sich hierbei um eine spezifische Ausprägung von Laryngalisierung, wie Redi/ Shattuck-Hufnagel (2001) es sehen?

Die Frage muss wohl überlegt sein. Denn: Ob Laryngalisierung mit anschließendem Aushauchen zur finalen Laryngalisierung oder zum Aushauchen gezählt wird, hat Einfluss auf die Signifikanz der diversen Tests. Ein Beispiel: Der Unterschied zwischen Ende und Fortführung eines Redebeitrags ist in Bezug auf reine Laryngalisierung oder Aushauchen nach modaler Phonation nicht signifikant. Die Tests werden aber für die jeweilige Phonationsart höchst signifikant, sobald die Folge aus beiden Änderungen zu der einen oder der anderen Kategorie gezählt wird.

Vor allem zwei Aspekte sprechen für eine Einordnung zur Laryngalisierung: Zum einen ähnelt die Verteilung der Intonationskonturen, die mit dieser Kategorie einhergehen, eher der von Laryngalisierung: Einfach steigende oder gleichbleibende Konturen kommen selten vor. Je nachdem, wie analytisch man die globale Tonhöhenbewegung hört, ist es jedoch möglich, dass das Aushauchen im Anschluss an die Laryngalisierung noch einmal als leichter Tonhöhenanstieg wahrgenommen wird; die globale Kontur würde dann eventuell als fallend-steigend eingestuft werden. Zum anderen ist das Aushauchen – zumindest bewusst – oft kaum wahrnehmbar, nur mit Kopfhörern wirklich zu hören; Laryngalisierung scheint hingegen prominenter zu sein, selbst wenn nur wenige Schwingungsperioden betroffen sind.

4.3. Frage der Signalfunktion

Wie in Abschnitt 3.2 beschrieben, kommt signifikant mehr Laryngalisierung mit anschließendem Aushauchen an den Stellen vor, nach denen ein Redebeitragswechsel stattfindet. Der Effekt verstärkt sich noch, wenn man nicht Redebeitragsende und -fortführung miteinander vergleicht, sondern danach differenziert, ob eine übergangsrelevante Stelle vorliegt oder nicht. In diesen Tests zeigt auch Laryngalisierung einen signifikanten Unterschied. Daraus könnte man schließen,

dass diese beiden Phonationskategorien dafür geeignet sind, den Wunsch, das Rederecht abzugeben, zu kontextualisieren. Für diese Interpretation spricht außerdem das vermehrte Auftreten von Hörersignalen nach Laryngalisierung beziehungsweise Laryngalisierung mit Aushauchen.

Aber worauf reagiert der Gesprächspartner wirklich? Ist es die nicht-modale Phonation, die ihn dazu veranlasst, selbst etwas zu sagen? Oder bildet das Beibehalten modaler Phonation bis zum Ende den Interpretationsrahmen für das Halten des Rederechts? In Selting (1995: 195) wird so argumentiert: Der unmarkierte Fall ist der Abschluss eines Redebeitrags; um zu zeigen, dass man weitersprechen möchte, benötigt es spezifischer Signale. Im Bereich der Intonation bedeutet das zum Beispiel, dass eine fallende Kontur unmarkiert ist und eine steigende oder gleichbleibende Kontur Weiterweisung markiert. Ob es sich mit der Phonation ähnlich verhält, ist schwer nachweisbar. Ein möglicher Hinweis wären zum Beispiel Unklarheiten in der Zuschreibung des Rederechts, wenn die Phonation bis zum Ende modal bleibt; dies könnte sich in mehr und/oder längeren Pausen vor einem Redebeitragswechsel äußern. Zwar sind die Pausen nach Modalstimme in der Tat etwas länger, aber die Häufigkeit von Pausen im Vergleich zu nahtlosen oder überlappenden Anschlüssen ist in diesen Fällen nicht signifikant größer.

Ein weiteres Problem bei der Bestimmung des Signalwertes eines phonetischen Merkmals ist das häufige gemeinsame Auftreten mit anderen Parametern. Ein Beispiel: Bei fallender F0 lässt auch die Spannung der Stimmlippen nach, eine wichtige Voraussetzung für die entspannte Form von *creak*. Für eine langsamere Schwingungsfrequenz (bei tiefer F0) ist außerdem ein geringerer Atemdruck nötig, wodurch es zu Unregelmäßigkeiten in den Schwingungen kommen kann. Zeigt also wirklich die Laryngalisierung, dass an dieser Stelle ein mögliches Ende erreicht ist, oder nicht vielmehr die damit einhergehende tief fallende Intonationskontur? Oder das Bündel aus Laryngalisierung und tief fallender Intonation?

Ein anderes Beispiel: Die Phänomene, die in diesem Artikel unter den Begriff "Aushauchen" zusammengefasst werden, sind übliche laryngale Vorbereitungsgesten für anschließendes Atmen. Ohne diese kurze behauchte Phase müssten die Stimmlippen in kürzester Zeit von einem regelmäßigen Schwingungsmodus in die Atemstellung (bei der die Stimmlippen weit abduziert sind) wechseln. Reagiert der Interaktionspartner in diesem Fall auf das Aushauchen oder auf das anschließende Atmen?

Eine Möglichkeit, der Frage, ob die phrasen-finale Phonationsart als Kontextualisierungshinweis dient, ein wenig näher zu kommen, ist ein Perzeptionstest mit "naiven" Hörern. Er könnte mehr Klarheit darüber bringen, ob die phonetischen Merkmale, die als relevant erachtet werden, auch einen Signalwert für andere Hörer haben. Ein mögliches Vorgehen wäre, einzelne Äußerungen aus den analysierten Daten dem Hörer anzubieten; dieser hat die Aufgabe zu beurteilen, ob der Gesprächspartner nach dieser Äußerung einen neuen Redebeitrag beginnt oder der aktuelle Sprecher weiterspricht. Auf diese Weise haben die Testpersonen ähnliche Bedingungen wie der wirkliche Gesprächspartner in diesem Gespräch, der an dieser Stelle entscheiden musste, das Rederecht zu übernehmen oder den Sprecher weitersprechen zu lassen.

Nachteil an diesem Verfahren ist die Konfundierung der Phonation mit anderen relevanten Merkmalen. Peters (2006) hat deshalb mit Hilfe von resynthetisierten

Stimuli getestet, ob finale Laryngalisierung zusätzlich zu fallender Intonation an Phrasengrenzen den Hörern bei der Entscheidung, ob an dieser Stelle ein Sprecherwechsel möglich ist oder nicht, hilft. Die Resynthese von Sprachsignalen ermöglicht es, einzelne Parameter, wie Tonhöhe oder Stimmqualität, separat zu verändern. Peters konnte, nebenbei bemerkt, in diesen Experimenten leider keinen zusätzlichen Signalwert für Laryngalisierung nachweisen.

4.4. Beantwortung der Untersuchungsfragen

Untersuchungsfrage 1: *Spielen Änderungen der Phonation gegen Ende einer Intonationsphrase eine Rolle bei der Organisation von Fortführung oder Wechsel von Redebeiträgen?*
Vor Redebeitragswechsel ist deutlich mehr Laryngalisierung in Kombination mit nachfolgendem Aushauchen zu finden als vor der Weiterführung eines Redebeitrags. Und sollte es zu keinem Redebeitragswechsel kommen, so folgt in vielen Fällen zumindest ein Rezipientensignal oder ein kurzer Kommentar des Interaktionspartners.

Nicht-modale Phonation tritt aber auch vor Redebeitragsfortführung auf und viele Redebeitragswechsel finden nach modaler Phonation statt. Insofern scheint die Phonation ein eher schwaches Signal zu sein, das eventuell seine Stärke eher im Bündel mit der Intonation und anderen Merkmalen beweist.

Untersuchungsfrage 2: *Kann anhand der vorliegenden Daten bestätigt werden, dass Laryngalisierung ein Signal für einen möglichen Redebeitragswechsel ist?*
Laryngalisierung kann dabei helfen, Phrasengrenzen zu ermitteln (zusammen mit anderen Grenz-Markierungen); diese müssen jedoch nicht zwangsläufig mit der Grenze eines Redebeitrags zusammenfallen. Sie scheint im Deutschen kein guter Hinweis auf einen möglichen Redebeitragswechsel zu sein. Denn sie tritt nicht selten an Stellen auf, an denen keine übergangsrelevante Stelle erreicht ist und auch kein Redebeitragswechsel danach stattfindet. Des Öfteren folgt nicht einmal ein Hörersignal.

Untersuchungsfrage 3: *Kookkurrieren die finalen Phonationsarten mit anderen sprachlichen Merkmalen?*
Ja, bestimmte Phonationsarten kommen bevorzugt mit anderen prosodischen Merkmalen vor. Laryngalisierung kann zum Beispiel kaum losgelöst von der Intonationskontur betrachtet werden; sie dominiert bei tiefer F0. Ein anderes Beispiel: Endet eine IP mit Aushauchen, wird meistens danach ein- oder ausgeatmet; nach modaler Phonation oder gar Laryngalisierung ist im Vergleich dazu Atmen eher selten.

Untersuchungsfrage 4: *Gibt es Unterschiede zwischen den Sprechern?*
Ja, die Sprecher unterscheiden sich in der Tendenz zu nicht-modaler Phonation am Ende einer Phrase. Nicht jeder verwendet gleichermaßen häufig Laryngalisierung oder Aushauchen.

Aber nicht nur in der Präferenz einer Phonationsart gibt es Differenzen, sondern auch in der phonatorischen Unterscheidung zwischen Redebeitragsende und

-weiterführung. Der signifikante Unterschied in Bezug auf Laryngalisierung in Kombination mit Aushauchen zeigt sich nicht bei jedem Sprecher in gleichem Maße.

4.5. Anwendungsmöglichkeiten

Unter der Voraussetzung, man schreibt bestimmten nicht-modalen Phonationsmodi die Fähigkeit zu, ein mögliches Ende anzuzeigen – oder dem Beibehalten einer modalen Phonation die Fähigkeit, das Rederecht zu wahren – könnte man dieses Wissen zum Beispiel in der Sprachtechnologie (vor allem im Hinblick auf die Mensch-Maschine-Interaktion) gewinnbringend einsetzen: Man stelle sich ein Computersystem vor, das schneller erkennt, ob der User noch weitersprechen möchte oder auf eine Reaktion des Systems wartet. Unnötige Überlappungen, die dem Menschen unhöflich vorkommen und seine Akzeptanz des Systems verringern, könnten ebenso verhindert werden wie das Gefühl, etwas falsch gemacht zu haben, weil das System nicht in adäquater Zeit reagiert. Aber auch umgekehrt ist es für den menschlichen Benutzer manchmal schwierig abzuschätzen, wann von ihm eine Eingabe erwartet wird.

Die Interaktion zwischen Mensch und Maschine ist ein sehr spezieller Fall; die zwischenmenschliche Kommunikation ist manchmal ebenso konfliktbehaftet, was sich zum Beispiel in häufigen Unterbrechungen äußert. Ein sehr prominentes Beispiel wurde von Geoffrey Beattie und Kollegen näher untersucht (siehe zum Beispiel Beattie 1982): Margaret Thatcher wurde nachgesagt, dass sie in Interviews mehr unterbrochen wurde als ihre männlichen Kollegen. Beattie stellt die Vermutung auf, dass nicht nur die Tatsache, dass sie eine Frau ist, dazu führt, sondern auch ihre Neigung, ihre Grundfrequenz zu senken, um sich besser behaupten zu können. Diese häufig tiefe Intonation sei von den Interviewpartnern als Signal für ein mögliches Ende des Turns missverstanden worden. Tatsächlich weisen viele der untersuchten Unterbrechungen typische Endsignale wie zum Beispiel fallende Intonation am Ende des Satzes auf.

Auf mögliche Phonationsänderungen am Ende von IPs übertragen wäre folgendes Szenario denkbar: Wenn IP-finale Laryngalisierung und finales Aushauchen als mögliche Endsignale von den Rezipienten aufgefasst werden, wird ein Sprecher, der grundsätzlich am Ende einer IP nicht-modal phoniert, von diesen eventuell öfter unterbrochen. Eine Sensibilisierung dieses Sprechers für das Thema kann dann eine sinnvolle Maßnahme sein, um ihm die Kommunikation zu erleichtern.

5. Literaturverzeichnis

Barth-Weingarten, Dagmar (2009): When to say something – Some observations on prosodic-phonetic cues to the placement and types of responses in multi-unit turns. In: Barth-Weingarten, Dagmar/ Dehé, Nicole/ Wichmann, Anne (eds.): Where Prosody meets pragmatics. Bingley: Emerald, 143-181.

Beattie, Geoffrey (1982): Turn-taking and interruption in political interviews: Margaret Thatcher and Jim Callaghan compared and contrasted. In: Semiotica 39 (1/2), 93-114.

Duncan, Starkey (1972): Some signals and rules for taking speaking turns in conversations. In: Journal of Personality and Social Psychology 23 (2), 283-292.

Eckert, Hartwig/ Laver, John (1994): Menschen und ihre Stimmen. Aspekte der vokalen Kommunikation. Weinheim: Beltz.

Fiehler, Reinhard (2005): Gesprochene Sprache, In: DUDEN – Die Grammatik, 7. Aufl., Mannheim: Bibliographisches Institut Mannheim, 1175-1252.

Ford, Cecilia E./ Thompson, Sandra A. (1996): Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: Ochs, Elinor/ Schegloff, Emanuel A./ Thompson, Sandra (eds.): Interaction and Grammar. Cambridge: Cambridge University Press, 134-184.

Henton, Caroline (1989): Sociophonetic aspects of creaky voice. In: Journal of the Acoustical Society of America 86, 26.

Klatt, Dennis/ Klatt, Laura (1990): Analysis, synthesis, and perception of voice quality variations among female and male talkers. In: Journal of the Acoustical Society of America 87 (2), 820-857.

Kohler, Klaus (1994): Glottal Stops and Glottalization in German. In: Phonetica 51, 38-51.

Köser, Stephanie (2005): ['zOl?n] oder 'sollten'? Zur Glottalisierung alveolarer Plosive im Deutschen. In: PHONUS 9, 1-115. Saarbrücken: Institut für Phonetik, Universität des Saarlandes.

Laver, John (1980): The phonetic description of voice quality. Cambridge: Cambridge University Press.

Laver, John (1994): Principles of Phonetics. Cambridge: Cambridge University Press.

Local, John/ Kelly, John/ Wells, William (1986): Towards a phonology of conversation: Turn-taking in Tyneside English. In: Journal of Linguistics 22: 411-437.

Marasek, Krzysztof (1997): Electroglottographic Description of Voice Quality. Habilitationsschrift, Universität Stuttgart. Erschienen in: Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS) 3 (2).

Ogden, Richard (2001): Turn transition, creak and glottal stop in Finnish talk-in-interaction. In: Journal of the International Phonetic Association 31, 139-152.

Ogden, Richard (2004): Non-modal voice quality and turn-taking in Finnish. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (eds): Sound Patterns in Interaction. Amsterdam: Benjamins, 29-62.

Peters, Benno (2006): Form und Funktion prosodischer Grenzen im Gespräch. Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät der Christian-Albrechts-Universität zu Kiel. (http://eldiss.uni-kiel.de/macau/ receive/dissertation_diss_2078, letzter Zugriff: 28.11.2013)

Redi, Laura/ Shattuck-Hufnagel, Stefanie (2001): Variation in the realization of glottalization in normal speakers. In: Journal of Phonetics 29, 407-429.

Selting, Margret (1995): Prosodie im Gespräch. Tübingen: Niemeyer.

Selting et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 10, 353-402.

Slifka, Janet (2006): Some Physiological Correlates 2to Regular and Irregular Phonation at the End of an Utterance. In: Journal of Voice 20 (2), 171-186.

Szczepek Reed, Beatrice (2013): Glottalisation and word linking as resources for multi-unit turn construction in German talk-in-interaction: Initial observations. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 14, 8-30.

6. Anhang: Abkürzungsverzeichnis

EGG	Elektroglottograph
F0	Grundfrequenz
IP	Intonationsphrase
CK_{ms}	männlicher Student CK
MD_{md}	männlicher Doktorand MD
MP_{mw}	wissenschaftlicher Mitarbeiter MP
ND_{ws}	weibliche Studentin ND
SL_{wd}	weibliche Doktorandin SL

Stephanie Köser
Universität des Saarlandes
Fachbereich Computerlinguistik und Phonetik
Saarbrücken
Germany
skoeser [at] coli.uni-saarland.de

Rhythmic analyses as a proof-procedure? – An initial observation on rhythmicity and projection*

Uwe-A. Küttner

Abstract
This paper reports a problematic case of unequivocally evidencing participant orientation to the projective force of some turn-initial demonstrative *wh*-clefts (DCs) within the framework of Conversation Analysis (CA) and Interactional Linguistics (IL). Conducting rhythmic analyses appears helpful in this regard, in that they disclose rhythmic regularities which suggest a speaker's orientation towards a projected turn continuation. In this particular case, rhythmic analyses can therefore be shown to meaningfully complement sequential analyses and analyses of turn-design, so as to gather additional evidence for participant orientations. In conclusion, I will point to possibly more extensive relations between rhythmicity and projection and proffer a tentative outlook for the usability of rhythmic analyses as an analytic tool in CA and IL.

Interactional Linguistics – Speech Rhythm and Rhythmic Analysis – Method – Participant Orientation – Turn-Constructional Units – Projection – Demonstrative Clefts

Dieser primär interaktional-linguistisch ausgerichtete Beitrag berichtet von einem Problemfall im Nachweis von Teilnehmerorientierung an den Turn-Fortführung projizierenden Eigenschaften einiger turn-initialer "Demonstrative-Cleft-Konstruktionen" (DCs). Es wird gezeigt, dass rhythmische Analysen diesbezüglich hilfreich sein können, da sie im konkreten Fall rhythmische Regularitäten erkennen lassen, die eine sprecherseitige Orientierung an einer projizierten Turn-Fortführung nahelegen. Vorausgegangene sequenzielle Analysen und Analysen von Turn-Design werden somit bedeutsam durch rhythmische Analysen ergänzt, um zusätzliche Evidenzen für Teilnehmerorientierungen zu erfassen. Abschließend werde ich auf potenziell weitreichendere Zusammenhänge zwischen Rhythmizität und Projektion in der Interaktion hinweisen und einen provisorischen Ausblick hinsichtlich der möglichen Verwendung rhythmischer Analysen als Analysewerkzeug in der Konversationsanalyse und der Interaktionalen Linguistik geben.

Interaktionale Linguistik – Sprechrhythmus und Rhythmische Analyse – Methode – Teilnehmerorientierung – Turnkonstruktionseinheiten – Projektion – Demonstrative Clefts

---

* I am profoundly indebted to the editors of this volume Dagmar Barth-Weingarten and Beatrice Szczepek Reed for their continued support and helpful comments on this paper. Without their enormous patience, this paper would not have been possible. I also want to thank Elizabeth Couper-Kuhlen for accompanying the initial steps of this work and introducing me to the study of conversational speech rhythm. Thanks are also due to the researchers at the *Centre of Excellence in Research on Intersubjectivity in Interaction* at the University of Helsinki, where the idea for this paper was developed and first presented. All remaining errors and imperfections are, of course, my own.

1. Introduction

One of the primary methodical concerns for analysts working in the fields of Conversation Analysis (CA) and Interactional Linguistics (IL) is to evidence participant orientation to the phenomenon under investigation (cf. *inter alia*, e.g. Sidnell 2010: 23-35, Sidnell 2013: 79-82, Barth-Weingarten 2008: 86-89). Grounding one's descriptions and interpretations of a particular phenomenon in the participants' conduct serves to ensure that these descriptions and interpretations are not merely analytic artifacts – "empirically observable" they may be from an analyst's perspective – but that they are 'real' and employable resources or practices for the interactants themselves. To do this for a particular target turn, conversation analysts and interactional linguists typically rely on the participants' subsequent conduct, for it is there that *their* understandings of what the target turn was meant to achieve is most commonly displayed. This is referred to as the "next-turn-proof-procedure" (cf. Sacks/ Schegloff/ Jefferson 1974: 728f., Levinson 1983: 321, Hutchby/ Wooffitt 1998: 15).

As useful as this procedure has proven for the empirical analysis of talk-in-interaction, it is not always unproblematic. One problem may arise if the phenomenon in question is not evidently oriented-to by participants on each and every single occasion of its occurrence. An analyst may then face problems in satisfying CA's/IL's methodic *desideratum* of evidencing participant orientation to the phenomenon. The present paper reports on one instance of this problem that arose during an engagement with turn-initial demonstrative *wh*-clefts (see section 2). While some of the turn-initial demonstrative *wh*-clefts in the corpus were intuitively characterizable as projector constructions (i.e. lexico-syntactic constructions which project turn continuation beyond themselves, cf. Günthner 2011), the available cases posed difficulties in unequivocally evidencing participants' orientation to turn continuation as having been *projected*, rather than contingently produced. In attempting to come to terms with these difficulties, a consideration of the rhythmic structure of the talk turned out to be helpful. Rhythmic analyses of the demonstrative *wh*-clefts and the subsequent turn-constructional units (TCUs) yielded striking regularities in terms of their rhythmic integration. It will be proposed that these regularities can be used analytically to complement preceding sequential analyses and analyses of turn-design in order to evidence participants' orientation to locally projected next TCUs.

The paper is structured as follows. After a brief outline of the data and method used (section 2), the analytic problem will be exposed (section 3). This will be done in two steps: First, by reviewing the notion of projection in talk-in-interaction more generally (section 3.1.), and second, by highlighting the specific analytic problem as it appeared in the context of my engagement with turn-initial demonstrative *wh*-clefts (section 3.2.). Subsequently, the approach to conversational speech rhythm used to deal with this problem will be summarized (section 4). This is followed by an illustration of how the results from the rhythmic analyses can be used to analytically complement sequential analyses and analyses of turn-design in evidencing participants' orientation to locally projected next TCUs (section 5). Finally, I will summarize and discuss the results and their possible implications for future research (section 6).

2. Data and method

This paper uses methods, models and assumptions from the fields of Conversation Analysis (cf. Sidnell 2010, Sidnell/ Stivers 2013) and Interactional Linguistics (cf. Couper-Kuhlen/ Selting 2001, Barth-Weingarten 2008). Within these approaches, social interactions are viewed as achieved orderly products of the participants' conduct. Accordingly, both, CA and IL, aim to empirically describe the resources and practices participants use to mutually organize their conduct from an emic (i.e. the participants') perspective. As has been pointed out in the introduction, the main methodic consequence of this endeavor is that an analyst has to provide evidence in the form of participant orientation(s) to the phenomenon or practice under investigation, which empirically underpin(s) the analyses and interpretations proposed. While CA's focus lies mainly with discovering the principles underlying the organization of action(s) in social interactions, IL focuses on the linguistic aspects of interactional resources and practices and is guided by the assumption that linguistic structures have evolved as habitualized solutions to recurrent interactional problems.

The specific focus of this paper, which is rather interactional linguistic in nature, lies on a lexico-syntactic construction that has been called demonstrative *wh*-cleft in the literature. To my knowledge, this term has first been used by Biber et al. (1999: 691) to encompass a recurrent cleft-like structure, which was found to be particularly frequent in conversation. A basic constructional schema for demonstrative *wh*-clefts can be seen in figure 1.

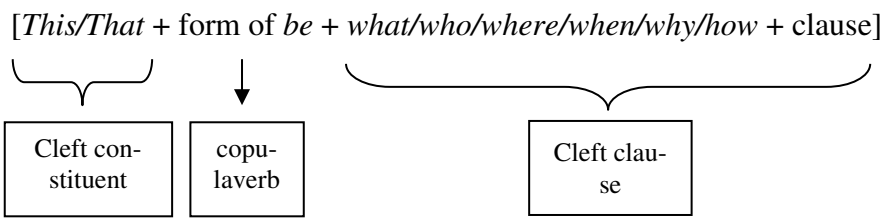


Figure 1. Basic constructional schema of demonstrative *wh*-clefts.

The data used for this study consist of roughly 14 hours of British and American English informal telephone conversations. From this corpus, all turn-initial[1] demonstrative *wh*-clefts featuring the demonstrative pronoun *that* as a cleft constituent have been collected. The data set yielded 50 instances of turn-initial demonstrative *wh*-clefts with the pronoun *that* (henceforth: DCs), a limited subset of which was potentially characterizable as projecting turn continuation.

[1] 'Turn-initial' is to be understood rather loosely in this context. The term, as it is used here, includes both, instances that constitute proper turn-beginnings (cf. Schegloff 1996) as well as cases in which the demonstrative *wh*-cleft is the first *clausal* TCU of a turn, irrespective of whether it is preceded by prosodically integrated prefaces (typically *oh* and *well*) or minimal, prosodically non-integrated token-responses to prior turns (typically *oh* and *yeah*). Space constraints preclude an account for this decision here, but can be found in Küttner (in prep.).

3. On projection and a concomitant analytic problem

3.1. Projection revisited

In its broadest sense, projection can be loosely understood as the foreshadowing of something that comes after the projecting item (cf. Streeck 1995: 87). Given that projection helps participants to anticipate the future development of mutually and ongoingly constructed courses of actions, and thereby to adjust and organize their conduct, it is a very common (almost ubiquitous) feature of talk-in-interaction. Thus, unsurprisingly perhaps, the notion of projection has figured prominently in many studies in CA and IL, and it has been shown to operate on many different levels in talk-in-interaction.

Among them are, for example, the system of preference organization, where the presence of certain elements (e.g. turn-initial prefaces like *well*) has been shown to foreshadow an upcoming dispreferred (Pomerantz 1984a) or non-straightforward response (Schegloff/ Lerner 2009), and the level of sequence organization ('action projection'), where certain utterances may be used to project a future course of action (see for instance Schegloff 2007 on *pre*-sequences and Schegloff 1980 on *pre-pre*s).

More importantly for this paper, projection is also a substantial feature on the level of turn-construction. For instance, during their emergence, TCUs project their upcoming possible completion. In other words, their upcoming possible completion is projectable for participants (cf. Sacks/ Schegloff/ Jefferson 1974: 702). This is an important feature of turn-construction for the organization of turn-taking in conversation, because, upon the possible completion of a TCU, speaker transition becomes a relevant option for participants (cf. ibid.). Generally speaking, the design of a TCU (and therefore potentially of a turn) as possibly complete involves a complex interplay of syntactic, pragmatic, and prosodic resources (cf. Ford/ Thompson 1996, Selting 1998). Accordingly, speakers can mobilize syntactic (cf. for example Auer 1996, 2002), pragmatic (cf. *inter alia* Streeck 1995, Ford/ Thompson 1996), and prosodic (cf. Auer 1996, Selting 1998, Barth-Weingarten 2009) resources to project turn continuation rather than turn completion.

Moreover, some lexico-syntactic constructions, for example basic pseudoclefts/*wh*-clefts (cf. Hopper 2008, Hopper/ Thompson 2008) and N *be (that)* constructions (cf. Günthner 2011), have been shown to specifically function as projector constructions, i.e. as projecting turn continuation with another discourse unit (at least one, but potentially more subsequent TCUs). Interestingly, participants commonly 'use' syntactic 'blank' slots to achieve the projection of turn continuation with these constructions.[2] This is different with the DCs dealt with in

[2] The projected material is regularly found in the complement slot of the 'traditional' syntactic structure (e.g. *What happened was* + $_{\text{projected}}$[TCU(s)/discourse segment], *The thing is (that)* + $_{\text{projected}}$[TCU/discourse segment]). What warrants a treatment of the projected material *as* having been projected is the fact that the TCUs following the projector construction are regularly neither syntactically, nor prosodically integrated with the projector construction (see Hopper 2008, Hopper/ Thompson 2007, Günthner 2011 for more elaborate treatments). The projection of a continuation is thereby achieved by exploiting participants' shared knowledge of a language's syntactic structure (see also Auer 2002).

this paper, which regularly constitute possibly complete TCUs of their own. Their status as possibly complete TCUs, after which speaker transition may be a relevant option (cf. Sacks/ Schegloff/ Jefferson 1974) contributes crucially to the analytic problem of unequivocally warranting a treatment of them as projector constructions. This problem – along with the problem of finding evidence for participants' orientation to the DCs *as* projector constructions – will be laid out in greater detail in the following section.

3.2. The analytic problem

As has been mentioned in section 2, a delimited subset of the DCs in the corpus was, albeit rather intuitively, characterizable as projector constructions.[3] Unlike other projector constructions however (see section 3.1.), these DCs are syntactically, and often also intonationally, possibly complete TCUs. Hence, if they really project a next unit, this may be seen to be attributable to their semantic-pragmatic properties. What is striking about most of these DCs is their local, contextual semantic-pragmatic underspecification, vagueness, or perhaps even better 'genericness'. Such 'genericness' is a prototypical design feature of projector constructions (cf. Streeck 1995: 89, Günthner 2011). Fragment (1) is a case in point.

(1) Driving test (Holt:SO88:1:5)[4]
(Susan has just reported that she failed in her driving test. Gordon had failed in his own driving test earlier.)

```
   01   Gor:   s:o you gonna reapPLY now;
   02          (0.4)
   03   Sus:   hm::;=I'll have to tAke it in BRI:GHton;=
   04          =an` gEt to know BRI:GHton;=you kn[Ow,]
   05   Gor:                                     [.tc]h YEA::H;=
-> 06          thAt's: what I:'M gonna dO:;
```

Here, Gordon inquires into Susan's plans for a future retake of her driving test in segment 01 (*s:o you gonna reapPLY now;*), which, after a brief pause, is met by Susan with a reporting of the strings attached (*I'll have to tAke it in BRI:GHton; an' gEt to know BRI:GHton;=you knOw,* segments 03-04). Following a minimal acknowledgment of Susan's response (*YEA::H;*, segment 05), Gordon moves to an announcement of his own plans for the future with a DC in segment 06 (*thAt's: what I:'M gonna dO:;*). This announcement claims that his plans are identical, or at least sufficiently comparable, with Susan's plans. Claiming this may be seen to do some empathic positioning on Gordon's side in response to Susan's reporting of

[3] The projected next discourse unit is typically circumscribed due to the fact that the initial demonstrative pronoun referentially relates the upcoming predication of the DC back to the previous turn, which imposes discourse-unit-length as well as topical constraints as to what can legitimately be positioned or introduced with and after the DC (cf. Küttner, in prep.).

[4] The transcription as well as later rhythmic notations follow the adapted GAT2 conventions for English (cf. Couper-Kuhlen/ Barth-Weingarten 2011).

a trouble, in that it invokes an identical experience (cf. Jefferson 1988, Kupetz 2014 and this volume, see also footnote 13).

In terms of its semantic-pragmatic 'genericness', Gordon's DC exhibits striking similarities with some of the basic pseudo clefts Hopper (2008, see also Hopper/ Thompson 2007) has shown to function as projector constructions: Firstly, the precise reference of the demonstrative pronoun *that* remains opaque, even after the completion of the DC. Its reference is not unambiguously retrievable from the sequential context. Secondly, the subsequent predication in the cleft clause, which features a semantically empty predicate (*dO:*), does not contribute to the specification of the reference made with the initial demonstrative pronoun either. (What is it that Gordon is going to do? Is he simply going to retake his driving test too? Is he going to retake it in Brighton? Is he going to get to know Brighton [possibly in order to retake his driving test there]?). Therefore, the DC merely implements a (perhaps designedly) unspecific announcement of Gordon planning a future action which is claimed to be congruent with Susan's plans. Following Hopper (2008), the semantic-pragmatic vagueness in Gordon's turn-design may be seen to provide leverage for characterizing this DC as projecting turn continu-ation (the two arrows in the fragment are meant to indicate the possible projection of a continuation set up therewith). As can be seen from the continuation of this sequence in fragment (1a), Gordon then indeed resolves this contextual underspecification in the subsequent TCU(s).

(1a) Driving test (Holt:SO88:1:5 ctd.)

```
   05   Gor:                                        [.tc]h YEA::H;=
   06          =thAt's: what I:'M gonna dO:;
-> 07          <<flat pitch>i'm gonna tAke it in NEWcAstle,
-> 08          After i've had a cOuple of
->             <<hyperarticulated>L:E[:ssons.]>>
   09   Sus:                      <<p>[had le?]
   10          yeah;>
```

As it turns out, Gordon's announcing claim has somewhat overstated the congruence of Susan's plans with his own plans (segments 07-08). Their plans resemble each other only on the rather 'abstract' basis of retaking a driving test at some other place.[5]

Finally, one could argue that there is a third feature which contributes to setting up the projection of a next TCU, namely the narrow focus accentuation (cf. Wells 2007: 117) on the subject of the cleft clause: *that's: what I:'M gonna dO:*. While

[5] It could be argued that the contextual knowledge the participants share plays a crucial role here. For example, it could well be that Susan knows that Gordon is going to move up to the Newcastle area, or that she knows for sure that Gordon is not going to retake his test in Brighton, and that she *recognizes* his announcing claim as *being 'misfitted'* on these grounds. However, such contextual knowledge is not retrievable from the data, and it would render other aspects of Gordon's announcement problematic (e.g. in making it uninformative and therefore possibly coming off as 'bragging' or 'self-attentive', a problematic action in the context of troubles talk). Needless to say perhaps, future plans might change, which could in principle be a basis for Gordon to produce this claim in a very 'fitted way'.

not necessarily projective in itself, such an accentuation pattern may prefigure an upcoming contrast or a shift in perspective (Barth-Weingarten 2009: 2278ff.), which may be explicated in subsequent TCUs. In this case, the shift is from *Susan*'s failed driving test and *her* plans to deal with this failure to *Gordon*'s plans.

Even if the analysis of Gordon's DC as projecting a next TCU may have been convincing so far, it should be noted that all of the foregoing are but technical considerations from an analyst's point of view. And these considerations are far from incontestable. In an alternative analysis, the DC *could*, upon its completion, arguably stand for itself and open up a transition-relevance place: a space at which speaker transition becomes a legitimate possible next action (TRP, cf. Sacks/ Schegloff/ Jefferson 1974, Ford/ Thompson 1996). Gordon's DC in segment 06 is possibly complete syntactically and (potentially) intonationally (with a final pitch movement falling-to-mid).[6] Moreover, from an on-line perspective, and taking the real time emergence of turns-at-talk into account (cf. Auer 2007, 2009), it *could* also stand for itself semantically and pragmatically. If Gordon really were to retake his driving test in Brighton, too, a turn claiming such a congruent situation with *thAt's: what I:'M gonna dO:* would be a perfectly appropriate response.

In sum then, from a linguistic point of view, Gordon's DC shows properties that would, in principle, allow for a characterization of it as projecting turn continuation (its semantic-pragmatic underspecification). Yet, from an 'on-line' perspective, this DC might as well have been possibly complete for participants. That is to say that, so far, there is no unequivocal data-*internal* evidence for the characterization of the DC as really having *foreshadowed* the subsequent TCUs *for the participants*.

To show that an item *does* project 'more-to-come' *for participants* is typically done by taking recipient behavior into account. One prototypical piece of evidence is the absence of turn-taking attempts from a recipient in the immediate aftermath of a projector construction. This would apply to fragment (1a), because Susan does not attempt to take the floor after Gordon's DC. Yet, this type of evidence could also be a coincidence of a lack of immediate uptake from a co-participant, as in fragment (2), where a contingently constructed multi-unit turn emerges as a result.

(2) Rummy cube (CFEngn6899))[7]

(Ana and her mother are talking about a rummy cube game.)

```
01   Mom:   you knOw;=↑I have One.
02          (0.7)
```

[6] Admittedly, a final pitch movement falling-to-mid is not as strong an indication of designed possible turn completion as a final pitch movement falling-to-low. Thus, the final pitch movement falling-to-mid could potentially also be analyzed as indexing an upcoming continuation. However, conversationalists recurrently use final pitch contours falling-to-mid as designed turn-completions. See for instance Gordon's declarative question *s:o you gonna reapPLY now;* in segment 01 of fragment (1), which is also done with a final pitch movement falling-to-mid and is designed to be complete. Similarly, Sarah's response in segment 03 of fragment (6) is done with a final pitch movement falling-to-mid to implement a designedly complete response. See also Szczepek Reed (2004).

[7] Call Friend data has been accessed via TalkBank (MacWhinney 2007).

```
   03   Mom:   WE have One.
   04   Ana:   ^YEA:H.
   05   Mom:   °hh but thAt was gIven to the ^THREE: of
               <<creaky>yOu:.>
   06          (2.6)
-> 07   Mom:   <<l>At some POINT.>
   08          (2.2)
   09  ⎧Mom:   ↑prO:bably One YEA:R,
   10  ⎪       (0.7)
   11  ⎪Mom:   I bought it (.) f:or hE:r-
-> 12  ⎨       (0.8)
   13  ⎪Mom:   fo? <<l, p>you knOw Eric jA:mes and JOH:N;=
   14  ⎪       =she bought it for YOU:;>=
   15  ⎩       =sOmething like THA:[T-]
   16   Ana:                       [YE]A:H,=
   17          =i THINK so.
```

Even without a sequential analysis of the actions being performed here, Mom's turn component in segment 07 (*<<l>At some POINT.>*) is hardly analyzable as having been projected by the prior turn-component in segment 05 (*°hh but thAt was gIven to the ^THREE: of <<creaky>yOu:.>*), which is designed to be complete syntactically, intonationally – with a final pitch movement falling-to-low as well as a shift in voice quality to final creaky voice – and pragmatically. To the contrary, Mom's TCU in segment 07 is an instance of a turn-increment (cf. Couper-Kuhlen/ Ono 2007) in pursuit of a response (Pomerantz 1984b). So rather than having been provided for by Mom's TCU in segment 05, the incremental TCU in segment 07 is the *result* of an absent turn-taking attempt from Ana. To put it the other way around, the absence of a turn-taking attempt, in this case, has not been *achieved* through the local projection of a subsequent TCU. Instead, it *causes* the turn-extension.

The same holds true for Mom's incrementally added, anecdotal telling (segments 09-15). While the component TCUs of this elaborating anecdote are achieved by continuative practices (e.g. final pitch movements rising to mid or staying level in segments 09 and 11; latching in segments 13-14), the entire package from segments 09-15 has not been projected by the preceding talk. In sum then, the sheer absence of recipient attempts at taking the floor does not provide satisfactory evidence for the projective force of a TCU, because it does not necessarily discriminate projected multi-unit turns from contingently constructed multi-unit turns.

Accordingly, the data in fragment (1a) remains equivocal with regard to participants' orientation to a *projection* having been done by Gordon with his DC in segment 06. On the one hand, one may view Gordon's continuation as the delivery of the projected material, taking the absence of a turn-taking attempt by Susan as evidence for its 'projectedness' (the *post-hoc* view). On the other hand, one may challenge this analysis by reference to the fact that, generally, multi-unit turns can also be contingently constructed – even without pauses between the individual units as was the case in fragment (2) – and that the absence of a turn-

taking attempt by Susan does not necessarily speak for the continuation as having been *projected* by the DC, rather than having emerged serially. That is, Gordon's continuation might be 'just' a contingently produced next TCU in a series of TCUs (the *emergent* view).

In principle, it would have been possible to solve this analytic issue by looking for comparable cases with other, more unequivocal types of recipient behavior which may evidence participants' orientation towards the projective qualities of a DC. For example,

- recipients' production of continuers, which serve as a vocalized externalization of passing up the right to talk (cf. Schegloff 1982, Jefferson 1983b, c), immediately following a DC, or alternatively
- recipients' pursuit of a continuation, where such a continuation is absent after the speaker's use of a DC, which can thereby be seen to have projected a continuation, i.e. a deviant case analysis (cf. Schegloff 1972, Sidnell 2013).

However, for the limited subset of the DCs that were potentially analyzable as having projected turn continuation, no such cases could be found in the data.

In the following, a possible analytic solution for this problem will be offered: One aspect of the design of the DCs that immediately stood out as remarkable was the recurrent rhythmicity with which the DCs were produced. Considering the rhythmic structure of the DCs and the subsequent TCUs in greater detail turned out to be helpful in addressing the issue of whether the DCs in question projected turn continuation for participants or not. In a way then, the remainder of the paper will explore the possibility of using rhythmic analyses to complement (*not* substitute!) CA/IL methods for evidencing participant orientations to locally projected TCUs.

4. On speech rhythm in conversation

Before this procedure can be illustrated, a few words concerning the approach to speech rhythm used in this paper appear due. The approach adopted here takes a perceptual *gestalt*-based perspective on speech rhythm and has mainly been advocated in Couper-Kuhlen (1993, 2009a) and Auer/ Couper-Kuhlen/ Müller 1999). This particular approach has been chosen, because it is most compatible with the basic assumptions and methods used in CA and IL, especially with regards to participants' *in situ* perceptibility of speech rhythm (see section 2 and below).[8]

English is generally said to exhibit a tendency towards a 'stress-timed' speech rhythm (see, for instance, Laver 1994: 523f). This means that stressed syllables are perceived to occur at roughly equal intervals of time. Therefore, rhythmic feet in English (a foot = a stressed syllable perceived as a rhythmic beat + all subsequent unstressed syllables up until the next stressed syllable) are perceived as having approximately the same length (cf. also Szczepek Reed 2011: 139-141).

[8] There are other useful approaches to speech rhythm, which usually make use of read aloud speech as their primary data. These approaches are commonly dealing with classifying languages into stress-timed or syllable-timed languages, respectively (see Szczepek Reed 2011: 140-146 for an introductory overview).

Interactional linguistic investigations of English speech rhythm in everyday talk have shown that conversationalists can and do manipulate the placement of prosodic prominences to form regular or irregular rhythmic *gestalts* (cf. *inter alia* Couper-Kuhlen 1993; Auer/ Couper-Kuhlen/ Müller 1999, Uhmann 1996). Within the studies conducted in this framework, speech rhythm is thus understood as "[…] a perceptual gestalt created by the quasi-regular recurrence of similar events, here accented syllables, coming at roughly equal intervals in time" (Couper-Kuhlen 2009a: 260). To establish a rhythmic *gestalt*,

> "[…] at least three prosodic prominences are necessary: the first two to establish a temporal interval […]"
>
> X----------X
>
> "[…] and the third to mark off an equivalent measure of time."
>
> X----------X----------X
>
> "A sequence of two prominences appropriately spaced may produce a gestalt-like group but is by definition not an isochronous structure."
>
> (Auer/ Couper-Kuhlen/ Müller 1999: 39;
> illustrations from Couper-Kuhlen 2009a: 260)

In keeping with these premises, this approach demands an 'auditory analysis first'-procedure, since this comes closest to how participants perceive speech rhythm *in situ*. Accordingly, the rhythmic analyses shown here are based on auditory analyses. Acoustic data (namely interval durations) have been gathered in a subsequent step and will only be provided for illustrative purposes.[9]

Using this analytic framework, the aforementioned studies contributed to a better understanding of the role of speech rhythm in everyday talk. For instance, it was found that rhythmic integration (timed responses) or non-integration (delayed responses) can function as contextualization cues within the organizations of turn-taking and preference.[10] In general, the default case (unmarked timing) is for speakers to preserve a locally established metric at turn-transitions. This can be readily seen in the rhythmic notation of a conversational fragment, taken from Couper-Kuhlen (2009a:261).

[9] With regard to acoustic data, it should be noted that Auer/ Couper-Kuhlen/ Müller (1999: 54) found no rhythmic patterns with interval durations of less than 0.23 seconds or more than 1.2 seconds. They also state that, in terms of variability in interval duration relative to the previous interval, "the border be8tween [sic!] perceptual isochrony and perceptual anisochrony must lie somewhere between 31% and 47%" (ibid.: 54).

[10] These findings are consistent with, yet at the same time exceed, the vernacular CA characterizations of preferred responses as coming forth 'immediately' and dispreferred responses as being 'delayed', because they allow for an understanding of what it means for a response to come forth 'immediately' or with "just a bit of space between the end of a prior utterance and the start of [one's] own" (Jefferson 1983a: 8). It is timed with respect to a locally established metric (if any), i.e. it preserves the said metric.

(3) Good time (quoted from Couper-Kuhlen 2009a: 261 – rhythmic notation)
(Jeff and Jill are currently in a long distance relationship.)

```
01   Jeff:   /yOu guys have been /
02           /hAving a good      /
03           /TIME, (0.1)        /
04   Jill:   /UhHUMH::,
```

Here, Jeff's B-event statement (Labov/ Fanshel 1977), functioning as a candidate upshot, in lines 01-03 (*so yOu guys have been hAving a good TIME,*) features three accented syllables on *yOu*, *hAv-* and *TIME*, which establish a local rhythmic metric. The accented syllables can therefore be said to form rhythmic beats in this rhythmic *gestalt*. Jill's confirming response in line 04 (*UhHUMH::,*) is timed with respect to this local rhythmic metric. In other words, the first accented syllable of this confirmation token continues the rhythmic pattern Jeff has established by being timed as to form the next (expectable) beat of the established rhythmic gestalt.

Breaks in rhythmic timing (i.e. marked timing) on the other hand are typically found with dispreferred responses, at sequence boundaries or in displays of surprise (cf. Auer/ Couper-Kuhlen/ Müller 1999, esp. chapters 3 & 4; Couper-Kuhlen 1993, 2009a). With these basic considerations in mind, it is now possible to illustrate how rhythmic analyses turned out to be helpful in evidencing participant orientation toward the projective qualities of the DCs in question.

5. Rhythmic analyses as a proof-procedure?

Considering the noticeable rhythmic structure of the possibly projecting DCs in greater detail with the help of such rhythmic analyses yielded striking regularities as to the rhythmic integration of the DCs with their subsequent TCUs. The conducted analyses enabled observations which suggest that a DC speaker is geared to the production of a next TCU. By putting it this way, I mean to avoid the impression of claiming that the rhythmic structure of the DCs actively contributes *in situ* to *setting up* the projection. This claim would be hard to maintain given the aforementioned findings on rhythmic timing at turn transitions. As has been pointed out above, the projective force of a DC, if any, is possibly attributable to a contextual semantic-pragmatic underspecification (or perhaps a designed misfit). Considering the rhythmic structure of the talk has been used as an analytic step in gathering additional evidence for participant orientation towards projected next TCUs here. It is *not* to be understood as a resource or practice participants use to achieve this projection. Since rhythmic integration across different speakers' talk has been found to be the default case at turn-transitions, there seems no possibility for convincingly arguing that rhythmicity at turn- or TCU-endings, i.e. at possible TRPs, can be used by participants to project more-to-come. This confinement has to be borne in mind throughout the subsequent analyses.

Let me begin to illustrate the procedure and to develop the argument with a, by now, familiar case. Fragment (1b) is a blended version of fragments (1) and (1a).

(1b) Driving test (Holt:SO88:1:5)

(Susan has just reported that she failed in her driving test. Gordon had failed in his own driving test earlier.)

```
   01   Gor:   s:o you gonna reapPLY now;
   02          (0.4)
   03   Sus:   hm::;=I'll have to tAke it in BRI:GHton;=
   04          =an` gEt to know BRI:GHton;=you kn[Ow,
   05   Gor:                                     [.tch YEA::H;=
-> 06          =thAt's: what I:'M gonna dO:;
-> 07          <<flat pitch>i'm gonna tAke it in NEWcAstle,
   08          After i've had a cOuple of
               <<hyperarticulated>L:E[:ssons.]>>
   09   Sus:                      <<p>[had le?]
   10          yeah;>
```

It may be recalled that Gordon's DC in segment 06 (*thAt's: what I:'M gonna dO:;*) was analyzable as possibly projecting the subsequent elaboration by virtue of its local, contextual 'genericness'. However, an alternative analysis – that upon its occurrence the DC could be heard as possibly complete syntactically, intonationally and pragmatically – was also considered. An analysis of the noticeable rhythmicity of the talk in segments 05-07 yields the following rhythmic structure.

(1c) Holt:SO88:1:5 rhythmic notation

		Interval duration[11] (Rhythmic tempo)	Interval variability relative to previous interval
Gor:	/ˈYEA::H /	0.48	
	/ˌthAt's: what /	0.41	-15%
	/ˈI:'M gonna /	0.37	-10%
	/ˌdO:; i'm gonna/	0.42	+14%
	/ˌtAke it in /	0.35	-17%
	/ˈNEW(ˌ)cAstle /	0.51	+46%

From the rhythmic notation, it can be seen that Gordon establishes a rhythmic *gestalt* with four beats through the regular accent placement on *YEA::H*, *thAt's*, *I:'M* and *dO:*, respectively. Given the aforementioned findings on unmarked rhythmic timing at turn transitions (see section 4), and the fact that the last element of the DC (the verb *dO:* in the cleft clause) is accented, one could expect a slight pause after the DC to enable Susan to place a possible response *on time*, if Gordon were to design this as a possible turn ending (such pauses usually remain

[11] In accordance with the approach to speech rhythm adopted here (see section 4), interval durations have been measured from the vowel onset of the current accented syllable to the vowel onset of the next accented syllable (cf. Couper-Kuhlen 1993, Auer/ Couper-Kuhlen/ Müller 1999).

untranscribed in ordinary transcripts). Instead, Gordon keeps talking through this possible transition space and places the next beat of the established rhythmic *gestalt* firmly within the next TCU, namely on its predicate (*tAke*). In conjunction with this observation, it is noteworthy that the beginning of Gordon's next TCU (*i'm gonna tAke it in NEWcAstle*) is 'omissible'. That is to say, it would have been possible for him to design these two TCUs in the following way.

(1d) invented variation on ex. Driving test

```
   05   Gor:   .tch YEA::H;=
-> 06          =thAt's: what I:'M gonna dO:;
-> 07          tAke it in NEWcAstle,
```

Ceteris paribus, this would have yielded the following rhythmic structure:

(1e) invented variation on ex. Driving test, rhythmic notation

```
     Gor:   /'YEA::H        /
            /,thAt's: what  /
            /'I:'M gonna    /
  ->        /,dO:;  (0.3)   /
            /,tAke it in    /
            /'NEW(,)cAstle     /
```

This possible realization would have freed the transition space – note the 0.3 second pause – and would have provided Susan with the opportunity to produce a receipting response *on time*. Arguably, an 0.3 second pause after a syntactically, intonationally and pragmatically possibly complete TCU could even have served as a cue for Susan that a receipting response is now due from her.

In sum,the rhythmic analysis of the DC in segment 06 suggests that (at least) Gordon appears to be oriented towards continuation here. He may even be said to block a possible *timed* incoming from Susan early on by virtue of continuing to speak and thereby tightly integrating his DC and the subsequent TCU rhythmically. Moreover, no marked continuative practices are employed in this case, like a rush-through (cf. Schegloff 1988, Walker 2010) or a final pitch movement rising to mid or staying level, for example. The rhythmic analysis of the DC could be seen to provide data-*internal* evidence for a participant orientation (namely Gordon's) towards continuation. This would then complement the preceding sequential analysis and the analysis of the DC's turn-design features (the contextual semantic-pragmatic underspecification).

A similar point can be made for fragment (4), in which Deena and Mark are talking about the personal 'moral' value of money in light of having to pay for their children's respective marriages.

(4) Assets (Holt:May88:2:4)

```
   01   Mar:   <<h>WE::LL; ((glottal hold))
```

```
   02          I don't knO:w;=
   03          i wE view> the f::::Act;=
   04          =that your KIDS are your ASsets;=rEally-=
   05          =and we'd rAther spEnd our mOney on our KIDS;=
   06          =than wAste it on ourSELves;=
   07          =or Anything ELSE,
-> 08   Dee:   THAT is exACTly what WE said;=
   09          =i sAid to SCOTT;=
   10          =as LONG as we've gOt a bit of [mOn]ey to?
   11   Mar:                                  [hh°]
   12   Dee:   you [knO]w,=as ↑LONG as wE go? e?(.)nOugh
   13   Mar:       [hh°]
        Dee:   money-=
   14          =that if we ↑WANT anythIn:g-=
   15          =a? OUR time of lI:fe,
   16          (0.4)
   17          wE can ↓BUY it.
```

In segments 01-07, Mark states his and his wife's stance towards covering the expenses for their children's marriages, with which Deena claims to fully agree with a DC in segment 08 (*THAT is exACTly what WE said;*). As it turns out, however, Deena seems to have a slightly different position on the matter, though (or at least wishes to modify it):[12] Whereas Mark claims to be inclined to assign priority to spending their money on their children's benefit(s) rather than on themselves, it is still 'us first, and then them' for Deena and her husband (see segments 10, 12, 14-17).

Now the question is how Deena's DC in segment 08 can be seen to project the elaboration of her slightly different stance or addendum to Mark's stance. Pragmatically, as well as syntactically and intonationally, her DC could very well stand for itself as a congruent agreement. However, in terms of its turn design, it is formulated as an agreement based on a settled position through the use of a narrow focus accentuation on the subject of the cleft clause (*WE*), indexing a change in perspective, and the past tense marking on the predicate of the cleft clause (*said*). These features, while claiming a congruent position on the issue, mark this opinion as having been independently arrived at prior to the here and now – a claim to epistemic rights, which may remain to be explicated (cf. Heritage/ Raymond 2005).

It may even be possible that Deena uses the DC here, precisely because she has a slightly different position on the matter, the introduction and elaboration of which requires some 'interactional space'. Otherwise she could arguably have produced a simple "*Oh absolutely*" in order to claim that she fully agrees, but does so independently (cf. Heritage 2002).[13]

[12] One may note in passing that this DC turns out to have been 'misfitted' yet again, albeit only in retrospect. What Mark has just said is really *not* "*exACTly*" what Deena said to her husband.

[13] A similar point may be raised for fragment (5), in which Gordon can be seen to move stepwise from talk about troubles to other matters by inquiring into ancillary matters first (*s:o you gonna reapPLY now;*) and, subsequent to Susan's response, stabilizing them topically with the DC (cf. Jefferson 1984).

Interestingly, when considering the rhythmic structure of this bit of talk, a similar pattern as in the previous example is observable.

(4a) Assets (Holt:May88:2:4, rhythmic notation)

			Interval Duration (Rhythmic tempo)	Interval variability relative to previous interval
Dee:	/ˈTHAT is ex	/	0.56	
	/ˈACTly what	/	0.54	- 4%
	/ˈWE said;	i /	0.55	+ 2%
	/ˌsAid to			

The accents on *THAT*, *-ACT-*, and *WE* can be heard to establish a rhythmic *gestalt*, thus forming rhythmic beats within it.[14] This rhythmic *gestalt* is again preserved across syntactic and intonational (note the final pitch movement falling-to-mid) boundaries. Like in the previous example, the next accent falls on the predicate of the subsequent TCU (*sAid*), and the subject of this TCU (*i*) 'occupies' the potential transition space. Once more, a close rhythmic analysis of the TCUs in question, in combination with other, previously ambivalent design features of Deena's talk, suggests *her* orientation or perhaps 'geared-ness' towards continuation. In a way, Deena can be shown to anticipate that the 'interactional space' for the subsequent TCU will be hers.

It should be stressed that the 'anticipation of interactional space' does not necessarily coincide with lexico-syntactic pre-plannedness or preformulation. This is vividly exemplified by the following example, which is otherwise rather similar to fragment (1b) in terms of the DC being used.

(5) Hostess (SBL:2:2:3:R)

(Chloe is complaining to Claire about a recent development for the hostesses to serve too much food at their bridge parties, which takes away too much playing time for her. Apparently, Claire is going to be the host of one of the next bridge parties.)

```
01   Chl:   but ↑I: thInk-=
02          =if we jUst (0.2) BRING our ˇsAn'which,=
03          =an‘ then a <<l>u:hm;>
04          some (.) hOstess serves COffee.
05          an‘ then you !S:IT! down an‘ eat your sAndwich.=
06          =<<l>if you WANT it;>=
07          =and then about THREE thIrty,
08          °hh the hostess !↑JUMPS! up and sE:rves
            some desSE:RT.
```

[14] It is also noteworthy that Deena chooses to use the full form of the copular verb over the contracted form here (i.e. *THAT is* rather than *THAT'S*). This is the only time a speaker uses a full form of *be* in this construction in the entire corpus.

```
   09          (1.2)
   10   Cla:   [WE:LL-]
   11   Chl:   [↑RIGHT] while you're PLAYing.
-> 12   Cla:   °h well THAT'S what ↑I'M going to dO;
-> 13          i'm just gOing to: °hh
   14          (0.6)
   15   Chl:    'MEMbe[r? a?]
   16   Cla:   <<h, f>[i:'m ] just gonna hAve either
                 CHEE:SEcla:ke?
   17          u? CHEE:Seca:ke?>
   18          or a SUNdae:-
```

Here, too, Claire's DC in segment 12 could well stand for itself syntactically, pragmatically (if her hosting plans were to match up with what Chloe has just suggested), and intonationally. Alternatively, it could project an elaboration of what food Claire plans to serve by virtue of exploiting a semantic-pragmatic underspecification of the claimed congruence of her hosting plans with Chloe's suggestion. Again, note also the narrow focus accentuation on the subject of the cleft clause, which can index a shift in perspective or a contrast with the undesirable hosting practices Chloe has just been complaining about prior to this fragment.

A rhythmic analysis of Claire's TCUs in segments 12-13 reveals the following:

(5a) Hostess (SBL:2:2:3:R, rhythmic notation)

			Interval Duration (Rhythmic tempo)	**Interval variability relative to previous interval**
Cla:	well	/ˈTHAT'S what /	0.41	
		/ˈ↑I'M going to /	0.58	+41%[15]
		/ˌdO; i'm just /	0.59	+ 2%
		/ˌgOing to: /	0.60	+ 2%
		/ˌ°hh		

Claire establishes a rhythmic *gestalt* with the DC through regularly placing accents on *THAT'S*, *↑I'M*, and *dO*, which therefore form beats in this local rhythmic metric. Once more, this metric is preserved across the syntactic boundary, and the next beat falls firmly within the next TCU (again on its predicate, *gOing*). Here, too, Claire 'occupies' the transition space with the beginning of this next TCU (*i'm just*). Again, the rhythmic structure integrates the DC and the following syntactic unit and creates a bind between these two TCUs which suggests Claire's orientation towards turn continuation.

[15] This interval variability seems to border on the values identified for perceptual anisochrony by Auer/ Couper-Kuhlen/ Müller (1999) (see footnote 9). Perceptively, it belongs to the rhythmic *gestalt*, though.

What this example shows beyond the general argument is that a speaker's 'anticipation of turn space' for a subsequent continuation is evidently independent from lexico-syntactic planning of this continuation (see the hitches in segments 13-14). That is, a speaker may well be oriented to the local projection of a next TCU, without necessarily having (entirely) 'planned' what this TCU consists of lexico-syntactically (note though that even the inbreath as a hesitation phenomenon is timed with regard to the local rhythmic metric, see again Auer/ Couper-Kuhlen/ Müller 1999).

What is more, a retrospective orientation towards having secured the right to talk is visible in Claire's subsequent behavior. When Chloe, after the 0.6 second hitch in Claire's elaborating TCU (segment 14), launches a return to her complaint (arguably by exemplifying the better times back in the day with '*MEMber?*, segment 15), Claire treats this as an illegitimate incoming by continuing her (projected) elaborating TCU in higher pitch and volume (cf. French/ Local 1983). This conduct further strengthens the claim that Claire treats the turn-space for the elaborating TCU following her DC as legitimately *hers*. Thus, there seems to be converging evidence for an analysis of Claire having set up a *projection* of a further discourse unit with her DC.

A final case will show that indeed the same kind of rhythmic integration of a DC with a subsequent TCU *can* actually co-occur with other continuative practices. In fragment (8), Debbie and Sarah are talking about a friend of Sarah's called Bryn, who has moved to Puerto Rico recently. Sarah has just returned from paying Bryn a visit in Puerto Rico.

(6) Happy (CFEngn6239)

```
   01   Deb:   is she HAppy thE:re?
   02   Sar:   ↑YAH;
   03          she LOVES it;
   04   Deb:   dOes she: hAs she made FRIENDS,
-> 05   Sar:   oh THAT'S why she ↓mOved there,
   06          'cause shE has a lO:t of frIends from MA:disO:n?
   07   Deb:   m_hm,
   08   Sar:   that're puerto RIcan;
```

After Debbie's first general inquiry whether Bryn is happy in Puerto Rico in segment 01, which is answered positively by Sarah in segments 02-03 (*↑YAH; she LOVES it;*), Debbie asks a follow-up question with a *yes/no*-interrogative in segment 04 (*dOes she: hAs she made FRIENDS,*). Sarah's response treats Debbie's question as problematic in several ways. Firstly, by not beginning with either *yes* or *no*, Sarah's response is non-type-conforming (cf. Raymond 2003), and secondly, her response is *oh*-prefaced, which serves as an indexical marker of the respondent having a problem with the question or treating it as inapposite (cf. Heritage 1998). As it turns out from the DC and the following elaborating clauses, the question is inapposite in terms of the presupposition it makes: That Bryn knows nobody in Puerto Rico and has to form new friendships.

Sarah doubly challenges this presupposition in segments 05-06 and 08 by stating that Bryn already knew people in Puerto Rico, whom she befriended during her time in Madison, and that these friendships have even been part of her motivation to move there herself. However, it should be noted that the DC itself (*THAT'S why she ↓mOved there,*) would not suffice to challenge the presupposition in this way. The DC by itself essentially confirms the presupposition in Debbie's question (i.e. it would yield the reading 'Bryn moved to Puerto Rico to make new friends'). Sarah's DC is therefore, upon its completion, contextually 'misfitted' and could therefore possibly project a further sequential unpackaging of Sarah's problem with Debbie's question, which had been indexed by the *oh*-preface (cf. Heritage 1998: 304). Again, there is the observation of rhythmic integration of the DC with the subsequent TCU.

(6a) Happy, rhythmic notation

		Interval Duration (Rhythmic tempo)		Interval variability relative to previous interval
Sar:	oh /ˈTHAT'S why she /	0.45		
	/ˌ↓mOved there, cause /	0.57		+26%
	/ˈshE has a /	0.32	double time	0%
	/ˌlO:t of /	0.32	double time	+9%
	/ˌfrIends from /	0.35	double time	+6%
	/ˈMA:di /	0.37	double time	
	/ˌsO:n?			

Sarah establishes a *gestalt*-like group with accents on *THAT'S* and *↓mOved*. By placing the next accent firmly within the next TCU, namely on the subject of the following subordinate clause (*shE*), Sarah creates a regular rhythmic *gestalt* across clause boundaries. The three accented syllables form beats within this *gestalt*. Subsequently, Sarah increases the rhythmic tempo by moving to double time, which shall, however, not be of further concern here. Most remarkably, Sarah rhythmically integrates her DC tightly with the (arguably) projected next clause across syntactic boundaries.

In this instance, the rhythmic integration co-occurs with a final pitch movement rising to mid, which indexes turn continuation rather than turn completion.[16] Thus, this example shows that rhythmic integration across clause boundaries *can* indeed co-occur with other continuative practices, providing converging evidence for participant orientation towards continuation. This fact might in turn be taken as indi-rect evidence for a possibly more systematic connection between 'projection' as a phenomenon on the one hand, and 'rhythmicity' as a co-occurring feature on the other (see the discussion below). However, as was shown with the exam-

[16] According to Szczepek Reed (2004), however, a final pitch movement rising-to-mid could equally well signal turn completion, which in a way warrants the additional rhythmic analysis conducted here.

ples before, tight rhythmic integration across clause boundaries *need* not co-occur with other continuative practices and still suggests a speakers' orientation towards turn continuation.

6. Summary, discussion, implications, and a caveat

6.1. Summary, discussion, and implications

This paper presented an exploratory attempt at using rhythmic analyses of conversational speech to gather additional evidence for participant orientation to locally projected next TCUs. The procedure was exemplified by considering a small collection of turn-initial demonstrative clefts (DCs), which were intuitively characterizable as projecting more-to-come, but for which other evidence of participant orientations towards their projective force was equivocal. By taking the rhythmic structure of the DCs and the subsequent TCUs into account, it was observable that the DC-speakers 'blocked' what would otherwise have been the transition space. This is regularly achieved through filling the transition space as well as tight rhythmic integration of the DC and the next TCU across syntactic and intonational boundaries (note that the 'integrating beat' typically fell on the predicate of the subsequent TCU). Such a rhythmic integration suggests the DC-speaker's orientation (or perhaps 'geared-ness') towards turn continuation.

This is coherent with findings from Couper-Kuhlen (2009a: 267), who has shown that components of *projected* multi-unit turns are "[…] held together in part through the fact that they are timed regularly with respect to one another", whereas turns, which *become* multi-unit turns *contingently*, "[…] bear the mark of their contingent construction […,] precisely because they incorporate a break in timing". In her paper, she used negated TCUs, which typically project another, elaborating TCU as examples of projected multi-unit turns and showed that these TCUs were regularly timed with respect to one another (see also Couper-Kuhlen 2009b). The procedure in this paper reversed the analytic pathway by considering the timing of adjacent TCUs as additional evidence for their 'projectedness'.

The reasoning behind this partly relies on previous findings on speech rhythm. For one, it strongly rests on the general rhythmic regularities that have been found to hold at turn-transitions (cf. Auer/ Couper-Kuhlen/ Müller 1999, chapter 3). It is by reference to these regularities that one can meaningfully say that the DC-speaker 'occupies' the transition space with the unaccented material beginning the projected TCU (the *i'm gonna* in (1b), the *i* in (4), the *i'm just* in (5)), instead of leaving a (usually untranscribed) pause to enable a timed response from the recipient. By their very nature these features are opaque to the participants as well as the analyst and only become visible through close rhythmic analysis.

Of course, these findings raise the question whether there are more general and systematic relations between 'projection' on the one hand, and 'rhythmicity' on the other. *Prima facie*, this seems plausible, given that 'projection' is enabled by a sedimentation of recurrent sequences of events, actions, and the like (cf. Auer 2002: 1-3). Generally, recurrent sequences (of events, actions, or the like) are prone to becoming rhythmicized, precisely because of the 'patterned-ness' that

results from their recurrence (just think of the assembly line worker who repeats the same sequence of actions over and over again).[17] Moreover, one of the primordial projective relations in talk-in-interaction, namely that between first-pair part (FPP) and second-pair part (SPP) in an adjacency pair, has been found to have an underlying rhythmic basis with respect to its timing (Auer/ Couper-Kuhlen/ Müller 1999, Couper-Kuhlen 2009a). Regarding unit-projection, it would thus be interesting to reconsider other constructions which have been found to function as projector phrases (e.g. pseudoclefts, *it*-extrapositions, N *be that* constructions) and see whether similar rhythmic regularities hold for them. If that turns out to be the case, one might begin to see a much grander theme: For instance, a 'rhythmicalization' of more or less formulaic (see Calude 2009 for DCs) constructions that parallels their grammaticalization as projector constructions (cf. Günthner 2011).

Furthermore, if, indeed, more systematic relations between rhythmicity and unit projection happen to be discovered in future studies, one may start considering rhythmic analyses as a valuable, additional tool at the analyst's disposal. The potential prospects to be gained from such studies are as yet uncertain, but, following Jefferson's (1996: 9) take on explorative work, they may well be worth being pursued:

> Seems to me it makes sense to push the stuff, keep pushing at it, see how far it might go. You can always pull back to a more cautious, reasonable, sensible position. But when you're doing this explorative work, go ahead and push.

After all, this could contribute to what Sidnell (2013: 82) has called a "central project of CA" (and methodically related disciplines), namely to "[discover] new forms of appropriate (i.e. data-*internal*) evidence".

6.2. A caveat

Besides the caveat introduced at the beginning of section 5 – that I do *not* wish to claim that the rhythmic structure of the talk *contributes in situ* to the projection of a next unit – there is another one that should be mentioned. It is utterly important to stress that the procedure is *not* self-contained. That is to say that simply because a TCU is rhythmically integrated with a next one, this does not *ipso facto* or in itself mean that the second has been *projected* by the first. It was stressed several times that the rhythmic analyses merely served as a *post-hoc* complementary analytic step to gather additional evidence. Detailed analyses of the sequential environment of a TCU as well as its turn design remain the primary analytic concern.

[17] In this regard, it is also interesting that metrical regularity has been found to facilitate speech planning and production (Tilsen 2011). In a similar vein, rhythmic regularity might as well co-occur with projectable patterns which reduce the exigencies of dealing with interactional contingencies.

7. References

Auer, Peter (1996): On the Prosody and Syntax of Turn-Continuations. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation. Cambridge: CUP, 57-100.

Auer, Peter (2002): Projection in Interaction and Projection in Grammar. In: InLiSt 33, 1-39 (http://www.inlist.uni-bayreuth.de/).

Auer, Peter (2007): Syntax als Prozess. In: Hausendorf, Heiko (ed.): Gespräch als Prozess – Linguistische Aspekte der Zeitlichkeit verbaler Interaktion. Tübingen: Narr, 95-142.

Auer, Peter (2009): On-line syntax: Thoughts on the temporality of spoken language. In: Language Sciences 31, 1-13.

Auer, Peter/ Couper-Kuhlen, Elizabeth/ Müller, Frank (1999): Language in Time – The Rhythm and Tempo of Spoken Interaction. Oxford: OUP.

Barth-Weingarten, Dagmar (2008): Interactional Linguistics. In: Antos, Gerd/ Ventola, Eija (eds.): Handbook of Applied Linguistics Vol. 2: Handbook of Interpersonal Communication. Berlin: DeGruyter,77-105.

Barth-Weingarten, Dagmar (2009): Contrasting and Turn-Transition: Prosodic Projection with the Parallel-Opposition Construction. In: Journal of Pragmatics 41, 2271-2294.

Biber, Douglas et al. (1999): Longman Grammar of Spoken and Written English. Harlow: Longman.

Calude, Andreaa S. (2009): Formulaic Tendencies of Demonstrative Clefts in Spoken English. In: Corrigan, Roberta et al. (eds.): Formulaic Language: Volume 1 – Distribution and Historical Change. Amsterdam: Benjamins, 55-76.

Couper-Kuhlen, Elizabeth (1993): English Speech Rhythm. Amsterdam: Benjamins.

Couper-Kuhlen, Elizabeth (2009a): Relatedness and Timing in Talk-in-Interaction. In: Barth-Weingarten, Dagmar/ Dehé, Nicole/ Wichmann, Anne (eds.): Where Prosody Meets Pragmatics. Bingley: Emerald, 257-276.

Couper-Kuhlen, Elizabeth (2009b): On Combining Clauses and Actions in Interaction. Finland Distinguished Professor Project Inaugural Lecture May 8, 2009, University of Helsinki. http://www.ojs.tsv.fi/index.php/virittaja/article/download/4206/3923 (last access: April 11, 2014).

Couper-Kuhlen, Elizabeth/ Barth-Weingarten, Dagmar (2011): A System for Transcribing Talk-in-Interaction: GAT2. English translation and adaptation of Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 12, 1-51 (www.gespraechsforschung-ozs.de).

Couper-Kuhlen, Elizabeth/ Ono, Tsuyoshi (2007): 'Incrementing' in Conversation. A Comparison of Practices in English, German and Japanese. In: Pragmatics 17(4), 513-552.

Couper-Kuhlen, Elizabeth/ Selting, Margret (2001): Introducing Interactional Linguistics. In: Selting, Margret/ Couper-Kuhlen, Elizabeth (eds.): Studies in Interactional Linguistics. Amsterdam: Benjamins, 1-22.

Ford, Cecilia E./ Thompson, Sandra A. (1996): Interactional Units in Conversation: Syntactic, Intonational, and Pragmatic Resources for the Management of Turns. In: Ochs, Elinor/ Schegloff, Emanuel A./ Thompson, Sandra A. (eds.): Interaction and Grammar. Cambridge: CUP, 134-184.

French, Peter/ Local, John (1983): Turn-Competitive Incomings. In: Journal of Pragmatics 7, 17-31.

Günthner, Susanne (2011): *N be that*-constructions in Everyday German Conversation: A reanalysis of "die Sache ist/das Ding ist" ('the thing is')-clauses as *projector phrases*. In: Laury, Ritva/ Suzuki, Ryoko (eds.): Subordination in Conversation. Amsterdam: Benjamins, 11-36.

Heritage, John (1998): *Oh*-Prefaced Responses to Inquiry. In: Language in Society 27, 291-334.

Heritage, John (2002): *Oh*-Prefaced Responses to Assessments – A Method of Modifying Agreement/ Disagreement. In: Ford, Cecilia E./ Fox, Barbara A./ Thompson Sandra A. (eds.): The Language of Turn and Sequence. Oxford: OUP, 196-224.

Heritage, John/ Raymond, Geoffrey (2005): The Terms of Agreement: Indexing Epistemic Authority and Subordination in Talk-in-Interaction. In: Social Psychology Quarterly 68(1), 15-38.

Hopper, Paul J. (2008): Die Bedeutsamkeit der mündlichen Interaktion für die Linguistik: Die Pseudocleft-Konstruktion im Englischen. In: Stefanowitsch, Anatol/ Fischer, Kerstin (eds.): Konstruktionsgrammatik II – Von der Konstruktion zur Grammatik. Tübingen: Stauffenburg Verlag, 179-188.

Hopper, Paul J./ Thompson, Sandra A. (2008): Projectability and Clause Combining in Interaction. In: Laury, Ritva (ed.): Crosslinguistic Studies of Clause Combining: The Multifunctionality of Conjunctions. Amsterdam: Benjamins, 99-123.

Hutchby, Ian/ Wooffitt, Robin (1998): Conversation Analysis: Principles, Practices, and Applications. Cambridge: Polity Press.

Jefferson, Gail (1983a): Two Explorations of the Organization of Overlapping Talk in Conversation – Notes on Some Orderliness of Overlap Onset. In: Tilburg Papers in Language and Literature 28(1), 1-28.

Jefferson, Gail (1983b): Caveat Speaker: Preliminary Notes on Recipient Topic-Shift Implicature. In: Tilburg Papers in Language and Literature 30(1), 1-25.

Jefferson, Gail (1983c): Notes on a Systematic Deployment of the Acknowledgment Tokens "Yeah" and "Mm hm". In: Tilburg Papers in Language and Literature 30(2), 1-18.

Jefferson, Gail (1984): On Stepwise Transition from Talk about Trouble to Inappropriately Next-Positioned Matters. In: Atkinson, John Maxwell/ Heritage, John (eds.): Structures of Social Action. Cambridge: CUP, 191-222.

Jefferson, Gail (1988): On the Sequential Organization of Troubles Talk in Ordinary Conversation. In: Social Problems 35(4), 418-442.

Jefferson, Gail (1996): On the Poetics of Ordinary Talk. In: Text and Performance Quarterly 16(1), 1-61.

Kupetz, Maxi (2014): Empathy Displays as Interactional Achievements – Multimodal and Sequential Aspects. In: Journal of Pragmatics 61, 4-34.

Küttner, Uwe-A. (in prep.): *That's*-initial Turns in Conversation (Working-title of PhD thesis).

Labov, William/ Fanshel, David (1977): Therapeutic Discourse: Psychotherapy as Conversation. New York: Academic Press.

Laver, John (1994): Principles of Phonetics. Cambridge: CUP.

Levinson, Stephen C. (1983): Pragmatics. Cambridge: CUP.

MacWhinney, Brian (2007): The TalkBank Project. In: Beal, Joan C./ Corrigan, Karen P./ Moisl, Hermann L. (eds.): Creating and Digitizing Language Corpora: Synchronic Databases, Vol. 1. Houndmills: Palgrave-Macmillan.

Pomerantz, Anita (1984a): Agreeing and Disagreeing with Assessments: Some Features of Preferred/ Dispreferred Turn Shapes. In: Atkinson, John Maxwell/ Heritage, John (eds.): Structures of Social Action. Cambridge: CUP, 57-101.

Pomerantz, Anita (1984b): Pursuing a Response. In: Atkinson, John Maxwell/ Heritage, John (eds.): Structures of Social Action. Cambridge: CUP, 152-163.

Raymond, Geoffrey (2003): Grammar and Social Organization: Yes/ No Interrogatives and the Structure of Responding. In: American Sociological Review 68(3), 939-967.

Sacks, Harvey/ Schegloff, Emanuel A./ Jefferson, Gail (1974): A Simplest Systematics for the Organization of Turn-Taking in Conversation. In: Language 50(4), 696-735.

Schegloff, Emanuel A. (1972): Sequencing in Conversational Openings. In: Gumperz, John J./ Hymes, Dell H. (eds.): Directions in Sociolinguistics – The Ethnography of Communication. New York: Holt, Rinehart and Winston, 346-380.

Schegloff, Emanuel A. (1980): Preliminaries to Preliminaries: 'Can I Ask You A Question'. In: Sociological Inquiry 50(3-4), 104-152.

Schegloff, Emanuel A. (1982): Discourse as an Interactional Achievement: Some Uses of 'Uh huh' and Other Things That Come Between Sentences. In: Tannen, Deborah (ed.): Analyzing Discourse: Text and Talk. Washington D.C.: Georgetown University Press, 71-93.

Schegloff, Emanuel A. (1988): Discourse as an Interactional Achievement II: An Exercise in Conversation Analysis. In: Tannen, Deborah (ed.): Linguistics in Context: Connecting Observations and Understanding. Norwood: Ablex, 135-158.

Schegloff, Emanuel A. (1996): Turn Organization: One Intersection of Grammar and Interaction. In: Ochs, Elinor/ Schegloff, Emanuel A./ Thompson, Sandra A. (eds.): Interaction and Grammar. Cambridge: CUP, 52-133.

Schegloff, Emanuel A. (2007): Sequence Organization in Interaction. Cambridge: CUP.

Schegloff, Emanuel A./ Lerner, Gene H. (2009): Beginning to Respond: *Well*-Prefaced Responses to *Wh*-Questions. In: Research on Language and Social Interaction 42(2), 91-115.

Selting, Margret (1998): TCUs and TRPs: The Construction of 'Units' in Conversational Talk. In: InLiSt 4, 1-48 (http://www.inlist.uni-bayreuth.de/).

Sidnell, Jack (2010): Conversation Analysis – An Introduction. Malden, Massachussets: Wiley-Blackwell.

Sidnell, Jack (2013): Basic Conversation Analytic Methods. In: Sidnell, Jack/ Stivers, Tanya (eds.): The Handbook of Conversation Analysis. Chichester, West Sussex: Wiley-Blackwell, 77-99.

Sidnell, Jack/ Stivers, Tanya (2013): The Handbook of Conversation Analysis. Chichester, West Sussex: Wiley-Blackwell.

Streeck, Jürgen (1995): On Projection. In: Goody, Esther N. (ed.): Social Intelligence and Interaction: Expressions and Implications of the Social Bias in Human Intelligence. Cambridge: CUP, 87-110.

Szczepek Reed, Beatrice (2004): Turn-final Intonation in English. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (eds.): Sound Patterns in Interaction – Cross-linguistic Studies from Conversation. Amsterdam: Benjamins, 97-118.

Szczepek Reed, Beatrice (2011): Analysing Conversation – An Introduction to Prosody. Basingstoke: Palgrave Macmillan.

Tilsen, Sam (2011): Metrical Regularity Facilitates Speech Planning and Production. In: Laboratory Phonology 2(1), 185-218.

Uhmann, Susanne (1996): On Rhythm in Everyday German Conversation: Beat Clashes in Assessment Utterances. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation. Cambridge: CUP, 303-365.

Walker, Gareth (2010): The Phonetic Constitution of a Turn-Holding Practice – Rush-Throughs in English Talk-in-Interaction. In: Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (eds.): Prosody in Interaction. Amsterdam: Benjamins, 51-72.

Wells, John C. (2007): English Intonation: An Introduction. Reprint. Cambridge: CUP.

Uwe-A. Küttner
Englische Sprache der Gegenwart
Institut für Anglistik und Amerikanistik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
Germany
ukuettne [at] uni-potsdam.de

On (or not on) the ‘upgrading–downgrading continuum’: The case of ‘prosodic marking’ in self-repair

Leendert Plug

Abstract
A number of recent studies of sound patterns in interaction label some stretches of talk ‘upgraded’ or ‘downgraded’ relative to other stretches of talk. These labels cover multiple phonetic parameters at once and appear to be relevant in a number of interactional contexts. In this paper I review references to these labels and discuss their theoretical status and empirical scope. I demonstrate that the question of whether a particular sound pattern can be described using these labels is not always a straightforward one to answer. I do this with reference to what psycholinguists have called ‘prosodic marking’ in instances of self-initiated self-repair. Available descriptions of ‘prosodic marking’ resemble those of phonetic ‘upgrading’, but are the two notions equivalent? I will suggest that the most informative answer might be ‘yes and no’, and conclude that the relationship between ‘prosodic marking’ in self-repair and ‘upgrading’ in other contexts illustrates the value of taking an open-minded multi-parametric approach to examining sound patterns in interaction.

phonetic upgrading and downgrading – clusters of phonetic features – prosodic marking in self-repair – Dutch – self-initiated self-repair

Eine Reihe aktueller Studien bedienen sich beim Vergleich von prosodisch-phonetischen Mustern in der Interaktion der Bezeichnungen ‚upgraded‘ und ‚downgraded‘. Diese Begriffe bezeichnen dabei eine Vielzahl phonetischer Parameter gleichzeitig und scheinen in einer Reihe interaktionaler Kontexte relevant zu sein. Der vorliegende Beitrag betrachtet den Gebrauch der Begriffe und diskutiert ihren theoretischen Status sowie ihre empirische Reichweite. Er zeigt, dass die Frage, ob sie genutzt werden können, nicht immer einfach zu beantworten ist. Im Beitrag werden sie vor dem Hintergrund des psycholinguistischen Begriffs ‘prosodic marking’ in selbst-initiierten Selbstreparaturen behandelt. Beschreibungen dieses ‘prosodic marking’ weisen Ähnlichkeiten mit phonetischem ‘upgrading’ auf, aber sind die beiden Begriffe äquivalent? Ich schlage vor, dass die informativste Antwort ‘ja und nein’ ist und schlussfolgere, dass die Beziehung zwischen ‘prosodic marking’ in Selbstreparaturen und ‘upgrading’ in anderen Kontexten den Wert eines unvoreingenommenen, multi-parametrischen Herangehens an die Untersuchung von prosodisch-phonetischen Mustern in der Interaktion demonstriert.

phonetisches upgrading und downgrading – Bündel phonetischer Merkmale – prosodische Markierung von Selbstreparaturen – Niederländisch – selbst-initiierte Selbstreparaturen

1. Introduction*

A distinctive feature of work on sound patterns in interaction in the tradition of French/ Local (1983), Local/ Wells/ Sebba (1985), Local/ Kelly (1986) and Local/ Kelly/ Wells (1986), sometimes called the *phonology for conversation* approach (see Selting 2010), is its insistence that phonetic analysis should cover multiple parameters, and be conducted without preconceptions as to how patterns along these parameters might relate to each other. In their "methodological imperatives", Local/ Walker (2005: 122) write that because we "simply do not know, from the outset, which details might be of relevance to the participants and might have a communicative function", we should "endeavour to be as open-minded as we can, and consider as many parameters as possible as potentially relevant". As Ogden (2012a: 9) points out, this approach results in descriptions of associations between particular interactional contexts – or linguistic units used in particular interactional contexts – and "patterns of phonetic variability which relate to clusters of phonetic parameters"(see also Local 1992, Curl 2005, Local/ Walker 2005, Walker 2011, Local/ Auer/ Drew 2010, Zellers/ Ogden 2013).

While many studies in the *phonology for conversation* tradition describe relevant clusters of phonetic features in straightforward multi-parametric terms – that is, by listing what pitch patterns are involved, and what intensity patterns, and what voice quality patterns, and so on – without recourse to summative, cross-parametric labels such as 'accented', 'emphasized' and the like, a number of recent studies label stretches 'upgraded' or 'downgraded' relative to other stretches. These labels cover multiple phonetic parameters at once, and appear to be relevant in a number of interactional contexts, including other-initiated self-repair and assessment sequences. They are not used widely outside of the *phonology for conversation* literature, but are increasingly cited in handbooks of conversation analysis, pragmatics and phonetics: see Couper-Kuhlen (2011: 500), who refers to 'upgrading' and 'downgrading' in a list of "new ways of describing prosodic and phonetic relationships", as well as Lindström/ Sorjonen (2012), Levinson (2012) and Smith (2013). What do these labels mean exactly, and what is their theoretical status and empirical scope?

In this paper I review references to what I will call the 'upgrading–downgrading continuum' in studies of sound patterns in interaction, and illustrate that deciding whether a particular pattern fits on the continuum, and therefore warrants the use of the associated descriptive labels, is far from trivial. I do this by examining the clustering of phonetic features in the context of self-initiated self-repair. The starting point in examining this context is the regular reference in psycholinguistic studies of self-repair to 'prosodic marking' – highlighting a repair's salience by manipulating its prosody. Is what the psycholinguists call 'prosodic marking' the same as 'upgrading'? I will suggest that the answer to this question, best summarized as 'yes and no', illustrates the value of taking an open-minded, multi-parametric approach in examining sound patterns in interaction.

* This work was supported by ESRC grant RES-061-25-0417 'Prosodic marking revisited: The phonetics of self-initiated self-repair in Dutch'. I would like to thank Paul Carter and Christina Englert for their contributions to the research reported here, and Richard Ogden, Traci Walker and the editors for helpful comments on a first draft of this paper.

2. The upgrading–downgrading continuum

As indicated above, handbook chapters on sound patterns in interaction increasingly refer to the labels 'upgrading' and 'downgrading', treating the discovery of variation along the relevant phonetic parameters as a major finding of the *phonology for conversation* approach. For example, Smith's (2013: 231) list of best exemplars of interactionally conditioned variation described in the *phonology for conversation* literature contains Local's (1992) observations on plosive release, Local/ Walker's (2004) description of transitions in multi-unit turns, and Ogden's (2006) finding that "agreement or disagreement can be indicated by phonetic *upgrading* or *downgrading* of pitch, intensity and articulatory clarity".

2.1. A brief history

As most readers will know, the concepts of 'upgrading' and 'downgrading' feature prominently, albeit in relation to lexical rather than phonetic features, in Pomerantz's (1984) account of agreement and disagreement in assessment pairs. We will return to this point below. As far as I can see, the first to refer to 'upgrading' in work on sound patterns in interaction is Curl (2002), in her PhD dissertation on the phonetics of repetition in other-initiated repair sequences. Curl investigates sequences in which one participant initiates repair on another participant's turn using an open-class repair initiator such as *huh?* or *pardon?*, and the other participant repeats all or part of their previous turn. The repeated turn or turn part constitutes a 'repetition repair'. Curl (2002: 6) writes:

> Examination of the phonetic characteristics of the collection of repetition repairs reveals two patterns: one group of repairs is done more loudly than the trouble source turn, with expanded pitch ranges, longer durations, and long-domain changes in articulation ... Another group of repairs is not louder than the trouble source turn, has non-expanded pitch ranges ..., does not have longer durations, and has similar articulations. These phonetic patterns mark the repetition repair as, respectively, 'upgraded' or 'non-upgraded' from the trouble source turn. The terms 'upgraded' and 'non-upgraded' are used as *aides memoire* for the two phonetic patterns described.

Curl goes on to show that the choice for or against upgrading is interactionally constrained: subtypes of repetition repair can be established based on the wider sequential environment in which the repair is embedded, and particular subtypes favour or disfavour upgrading. The labels 'upgraded' and 'non-upgraded' also come up in a series of subsequent papers by Curl herself (2004, 2005, Walker 2010), and a small follow-up study on Dutch repetition repairs by myself, which confirms the generalizability of Curl's account (Plug 2005).

Ogden (2006, 2012b) uses the labels 'upgraded' and 'downgraded' in his analysis of the phonetic design of second assessments – for example, *it's great* following a co-participant's *it's nice*. The use of the labels is fitting because this context is exactly that analysed by Pomerantz (1984). In her study, Pomerantz distinguishes three types of agreeing second assessment based on their lexical design: 'upgraded', 'same' and 'downgraded'. Our example of *it's nice – it's great* constitutes an upgraded second assessment, with a stronger evaluation term than that of the first assessment. *It's nice – it's nice* would constitute a same second assess-

ment, and *it's great – it's nice* a downgraded one. Ogden (2006) analyses the phonetic characteristics of these and other subtypes of second assessment, and finds that upgrading and downgrading are "achieved both lexically and phonetically simultaneously" (: 1726). The features he lists are very similar to those listed by Curl: phonetically 'upgraded' second assessments are different from preceding first assessments in having "an expanded pitch span over the whole turn", "more dynamic pitch contours on accented items", "pitch higher in the speaker's range", "slower tempo" and "closer, tenser articulations" (: 1726). Phonetically 'downgraded' second assessments are characterized by "a narrower pitch span, and a concomitant lack of dynamic pitch movement", "faster tempo", "articulations with more open stricture" and "a decrease in loudness" (: 1769).

The same labels have appeared more recently in studies by Local/ Auer/ Drew (2010), Couper-Kuhlen (2012, and see Couper-Kuhlen this volume) and Zellers/ Ogden (2013), possibly among others. These studies have in common that they do not provide as comprehensive a description of the labels' phonetic correlates as Curl and Ogden do. Local et al. investigate practices for relaunching a turn that is interrupted or not responded to on first production. They distinguish three types of relaunching according to sequential position: 'retrieving' at a next possible transition relevance place, 'redoing' after an intervening turn has been responded to, and 'resuscitating' after an intervening sequence has run its course. The first two are associated with consistent phonetic designs, such that 'retrieving' goes with phonetic downgrading relative to the first attempt at the turn, while 'redoing' typically goes together with "some kind of upgrading" (Local/ Auer/ Drew 2010: 144). The authors present measurements of pitch, intensity and duration, showing feature clustering along the lines of that described by Curl (2002) and Ogden (2006), but do not comment in detail on segmental articulations or overall articulatory settings.

Couper-Kuhlen (2012) investigates how interactants display affiliation or non-affiliation in responses to complaints embedded in longer complaint narratives. One of the practices available for displaying affiliation is offering an agreeing second assessment of the type described by Ogden (2006), so it is perhaps not surprising that Couper-Kuhlen refers to 'upgrading' in describing the phonetic characteristics of affiliative responses. Moreover, Couper-Kuhlen points out that non-affiliative responses such as factual follow-up questions or minimal acknowledgements of interest or empathy are typically phonetically 'downgraded' relative to the turns they respond to. However, while Couper-Kuhlen describes the pitch and intensity characteristics of the turns involved in detail, again showing feature clustering that closely matches that described by Curl (2002) and Ogden (2006), she does not cover temporal or articulatory aspects. In fact, Couper-Kuhlen (this volume) explicitly restricts the scope of 'upgrading' and 'downgrading' by adding the modifiers 'pitch', 'loudness', or 'prosodic': for example, 'prosodic upgrading' means upgrading of both pitch and loudness, with no claims about other phonetic parameters.

Finally, Zellers/ Ogden (2013) explore the use of statistical cluster analysis in studying the phonetics of speakers' productions of 'contrasts' such as *that's one thing* followed closely by *and then the other thing*. They conclude that 'upgrading' and 'downgrading' are appropriate labels for describing the phonetics of distinct subtypes of contrast, defined in terms of the contrast's function in the wider

interactional context. For example, they suggest that "in order for contrastive turns to have 'upgraded' global prosody compared to the surrounding talk, the contrastive structure must be used for aligning purposes". They run their clustering algorithms on measurements of pitch, intensity and duration, and, like Local/ Auer/ Drew (2010) and Couper-Kuhlen (2012), do not comment in detail on segmental articulations or articulatory settings. I will return to this point below, in the context of some critical notes on the 'upgrading–downgrading continuum'.

2.2. Critical appraisal

As pointed out in the introduction to this paper, many studies in the *phonology for conversation* tradition describe relevant clusters of phonetic features in straightforward multi-parametric terms without recourse to summative, cross-parametric labels. In fact, avoidance of such terminology can be seen as an important corollary of an insistence on open-minded multi-parametric analysis. Selting (2010: 26) refers to cross-parametric descriptive labels such as 'accented' as "traditional and holistic categories", and suggests that the reluctance to adopt these derives from a belief that "[o]n the descriptive level, only those phonetic parameters should be taken into account that can be registered objectively". This may be true, but it is equally rooted in a fundamental assumption guiding parametric phonetic analysis: namely that phonetic parameters can cluster in interesting and unpredictable ways (see Kelly/ Local 1989, Local 2003). This means that a pattern on one parameter – say, pitch – cannot be taken as a firm predictor of a pattern on another – say, loudness. More relevantly, patterns on two or three parameters – say, pitch and loudness, or pitch, loudness and voice quality – should not be taken as predictors of a pattern on a third or fourth parameter – say, tempo or articulatory setting. This kind of cross-parametric prediction is exactly what holistic categories provide.

Why, one may wonder, have researchers working in a tradition that guards against recourse to cross-parametric descriptive terms introduced and propagated the cross-parametric descriptive terms 'upgraded', 'non-upgraded' and 'downgraded'? Two downsides immediately present themselves. First, the terms allow for the kind of cross-parametric prediction that I have just referred to. As indicated above, the major studies that have propagated references to the 'upgrading–downgrading continuum' since Curl (2002) and Ogden (2006) do not explore all of the parameters that Curl and Ogden do. Local/ Auer/ Drew (2010), Couper-Kuhlen (2012) and Zellers/ Ogden (2013) all describe pitch and intensity patterns in detail, and Local/ Auer/ Drew and Zellers/ Ogden provide temporal measurements too. However, none provide details on segmental articulations or general articulatory settings in the stretches under consideration. While the temporal expansion and compression associated with upgrading and downgrading respectively are likely to go along with tenser and laxer articulatory settings, respectively, this is not necessarily the case: experiments have shown that speakers *can* speed up without an increase in segmental reduction (Van Son/ Pols 1990, 1992), and temporal compression without noticeable additional changes in segmental articulations is a characteristic feature of 'doubles' as described by Curl/ Local/ Walker (2006). This raises the question as to whether Local/ Auer/ Drew's, Couper-Kuhlen's and Zellers/ Ogden's 'upgrading' and 'downgrading' are phonetically

really the same as the 'upgrading' and 'downgrading' of Curl (2002) and Ogden (2006). Only detailed phonetic work will provide an answer.

A second downside of the use of the cross-parametric descriptive terms 'upgraded' and 'downgraded' is that clusters of phonetic features that do not fit these labels somehow seem less significant than clusters that do, even if they are similarly systematically deployed in particular interactional contexts. The former include the clusters associated with 'doubles' (Curl/ Local/ Walker 2006) and 'abrupt-joins' (Local/ Walker 2004), 'pivots' (Walker 2007), components of dispreferred response turns (Plug 2009) and several types of complaints (Ogden 2010), to name just a few. Is variation along the 'upgrading–downgrading continuum' really significant enough in comparison with these other clusters to warrant its own holistic categories?

I would suggest that the answer lies in the distribution of relevant clusters across interactional contexts. When compared with other clusters, those that can be captured in terms of 'upgrading' and 'downgrading' appear to have a particularly wide distribution. Moreover, in reviewing the studies discussed in this section, one may be struck by the coherence of the set of contexts in which these clusters are observed. While second assessments (Ogden 2006) are not particularly closely related to repetition repairs (Curl 2002), when we add turn relaunchings (Local/ Auer/ Drew 2010) it becomes clear that in all three environments, participants are, roughly speaking, repeating an action: repeating word-by-word what they have said before, or repeating the action of assessing, or repeating the launching of a turn. The contexts covered by Couper-Kuhlen (2012) and Zellers/ Ogden (2013) are somewhat different, but in both cases, affiliation, alignment and their opposites are relevant functional parameters. The use by the authors of a common set of holistic labels to describe relevant clusters of phonetic features opens the door to fruitful speculation as to why this particular set of contexts should be associated with these particular clusters of phonetic features, and generates hypotheses as to possible additional contexts in which these clusters might be relevant. These are clearly useful outcomes.

2.3. An upshot

To summarize the points just raised, I would suggest that the use of the cross-parametric labels associated with the 'upgrading–downgrading continuum' may well be worthwhile, provided that careful multi-parametric phonetic work is at its basis, and discovering the composition of relevant clusters of phonetic features remains close to the top of the empirical agenda. To put it differently, the question as to whether some observed cluster of features in some particular interactional context represents another example of variation along an 'upgrading–downgrading continuum' should be considered a hard one to answer. In the second half of the paper, I illustrate this point by addressing the question in relation to what some have described as 'prosodic marking' in the context of self-repair.

3. Prosodic marking in self-repair

I have been interested for some time in the phonetics of self-initiated self-repair, and have most recently focused on the correlates of 'prosodic marking' in this context. The concept of prosodic marking in repair was first introduced by Cutler (1983), who describes a collection of spontaneous, self-initiated, same-turn self-repairs such as *I saw her on Wednesd- uh Thursday* and *there's a har- large hole in the wall*. She describes an 'unmarked' repair as one in which the pitch, intensity and speaking rate of the repair component – *Thursday* and *large* in our examples – are not noticeably different from those of the reparandum – *Wednesd-* and *har-*. A 'marked' repair, on the other hand, "is distinguished by a quite different prosodic shape from that of the original utterance" (Cutler 1983: 81). By leaving a repair unmarked, the speaker "minimizes the disruptive effect of the error on the utterance as a whole", while marking assigns "salience" to the correction (Cutler 1983: 80).

The distinction between 'marked' and 'unmarked' repairs is used in other psycholinguistic work (see Levelt/ Cutler 1983, Levelt 1989, Howell/ Young 1990, Shattuck-Hufnagel/ Cutler 1999), but this remains rather vague on the phonetic details of 'prosodic marking'. Cutler (1983: 80–81) states without further illustration that "typically", a 'prosodically marked' repair "is uttered on a higher pitch and with greater intensity than the erroneous material". This clearly suggests a degree of overlap between the phonetic features of 'prosodic marking' in self-repair and those of 'upgrading' as described by Curl (2002), Ogden (2006) and others. Given that in addition, the context of self-initiated self-repair is closely related to that of other-initiated self-repair (Curl 2002), and self-correction can be considered as a type of 'contrastive structure' (Zellers/ Ogden 2012), it is tempting to conclude that what the psycholinguists call 'prosodic marking' is equivalent to 'upgrading', and self-initiated self-repair can be added to the list of interactional contexts in which variation along the 'upgrading–downgrading continuum' can be observed.

However, a closer reading of the sources cited above suggests that this conclusion would be premature. In their discussion of 'prosodic marking', Levelt/ Cutler (1983) acknowledge the influence of Goffman's (1981) ethno-methodological study of radio announcers' talk, in which he distinguishes between 'strident' and 'flat' self-corrections. Goffman's description of this distinction is as brief and generic as Cutler's, but appears somewhat narrower in scope. Goffman describes flat repairs as maximally fluent ones, and this fits well with Cutler's description of unmarked repairs. However, in Goffman's (1981: 216) words, in a 'strident' repair the speaker

> gives the impression of suddenly stopping in midstream because of being struck by what he has just heard himself say. Voice is raised and tempo increased. He then seems to redirect his attention to the single-minded task of establishing a corrected statement, as if this could (done quickly and forcefully enough) somehow grind the error into the ground, erase it, obliterate it, and substitute a corrected version.

Goffman's mention of "voice raising" would seem consistent with Cutler's reference to increased pitch and intensity in the repair, and with the notion of 'upgrading'. However, an increase in speaking tempo is *not* consistent with 'upgrading': Ogden (2006) includes "faster tempo" among the characteristics of 'down-

graded' second assessments. In other words, Goffman's brief description of what others have called 'prosodic marking' in self-repair hints at the possibility that 'prosodic marking' typically combines two features of 'upgrading' – a relative rise in pitch and intensity between the trouble source and the repair – with one feature of 'downgrading' – a relative rise in articulation rate. This clearly warrants a more detailed study of the phonetic features that make instances of self-initiated self-repair sound 'marked'.

3.1. Data and method

I have inspected a collection of 584 instances of lexical and phonological repair – in other words, instances in which a word choice is changed, as in *I saw her on Wednesd- uh Thursday*, or a mispronunciation is corrected, as in *there's a har- large hole in the wall*. All instances were taken from recordings of spontaneous talk-in-interaction contained in the Spoken Dutch Corpus (Oostdijk 2002). Representative examples are shown in (1) for lexical repair, and (2) for phonological repair.

(1) Lexical repair

a. *met de au- met de bus* ('by ca- by bus')
b. *als er met tekst gebrui- gewerkt wordt* ('when one use- works with text')
c. *de koelka- koelcel* ('the refrigera- cold store')
d. *die drie da- of die twee dagen* ('those three day- or those two days')
e. *een leuke k- een mooie keuken* ('a nice k- a beautiful kitchen')

(2) Phonological repair

a. [b]*aarbij* – [ʋ]*aarbij* ('with which')
b. *v*[r]*uch- v*[l]*uchtmiddel* ('means of escape')
c. *vana*[l] – *vana*[f] ('from')
d. *een b*[a]*nkj- een b*[ɑ]*nkje* ('a bench')
e. [w]*is ik z-* [m]*is ik zeker* ('I miss certainly')

The instances were analysed using the methodology described in detail by Plug/ Carter (2013); in that paper, we report on a study of the pitch and intensity characteristics of the lexical repairs. Exactly the same method was applied to the phonological repairs. All instances were classified for 'prosodic marking' by two native speakers of Dutch – myself and a linguist with particular interests in pragmatics and no detailed a priori knowledge of the phonetics of self-repair. Our instructions said that we should identify instances in which the repair sounds particularly prominent, not because of its semantic content, but because of the way it is delivered phonetically. We independently classified each instance as 'marked', 'unmarked' or 'possibly marked', with inter-rater agreement in the region of 85%. We reconsidered instances on which we disagreed initially, and in most cases reached a consensus classification. A few instances were removed from the col-

lection because we could not agree on their classification. For the purpose of the present analysis I conflated the subsets of 'possibly marked' and 'unmarked' instances; I will refer to all of these as 'unmarked' in what follows. This should isolate the phonetic features of clearly 'prosodically marked' instances of self-repair.

All instances in the collection were also subjected to a range of acoustic measurements. I focus here on three: the difference in pitch maximum between the reparandum and repair components, the difference in intensity maximum, and the difference in articulation rate. The latter was expressed using a proportional measure (see Plug 2011): for each instance, the articulation rate of the repair was divided by that of the reparandum, so that a value above one reflects speeding up between the two, and a value below one slowing down. The reason for focusing on pitch and intensity maxima is that the analysis reported in Plug/ Carter (2013) clearly shows that for both pitch and intensity, maximum measurements are stronger predictors of prosodic marking judgments than mean, median, minimum, range or standard deviation measurements, even though most of these measurements are closely correlated with each other. There is no reason to think this should be substantially different for phonological repairs, so focusing on pitch and intensity maxima should capture any trends in the larger collection of repairs effectively.

3.2. Main findings

In the analysis reported in Plug/ Carter (2013), we found that for the lexical repairs, the measures of pitch and intensity difference are strongly correlated with each other, and are significant predictors of the prosodic marking judgments: the higher the difference, the more likely a repair was classified as prosodically marked. There was no clear evidence of speakers manipulating the parameters of pitch and intensity in any other way in designing their repairs as 'prosodically marked'.

Combining the lexical and phonological repairs in a single collection does not change this picture: 'prosodic marking' mostly involves making the repair both louder and higher in pitch than the reparandum. Figure 1 illustrates that the measures of pitch and intensity difference (*F0 delta* and *Intensity delta*) are positively correlated (Pearson's $r=0.27$, $p<0.001$). Moreover, there is a heavy concentration of white data points in the top right quarter of the graph. This reflects that most instances of repair that were classified as 'prosodically marked' have a repair component that has both a higher pitch maximum and a higher intensity maximum than the reparandum. Instances that were classified as 'unmarked' (represented by the grey data points in Figure 1) show a wider dispersal, and a sizeable minority of these have a repair component that has both a lower pitch maximum than the reparandum, and a lower intensity maximum. On the other hand, there is no evidence of sizeable subsets of 'prosodically marked' instances having negative values for *Intensity delta* or *F0 delta*: instances that our raters classified as 'marked' are consistent in featuring a rise in pitch and intensity maximum between the reparandum and repair.

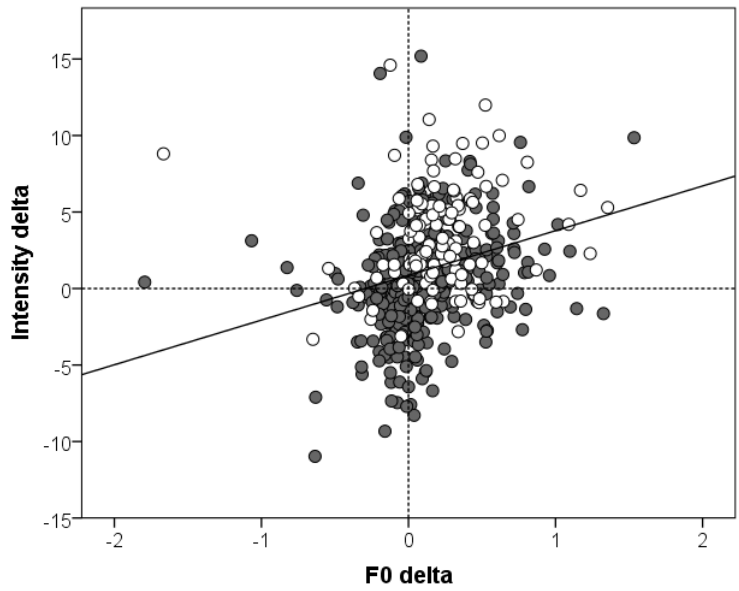


Figure 1.Correlation between measures of pitch and intensity maximum differences (*F0 delta* and *Intensity delta*) between reparandum and repair. For both measures, a value above zero (marked by a dotted line) means the repair has a higher maximum than the reparandum. White data points represent instances of repair that were auditorily judged 'prosodically marked'; grey data points represent 'unmarked' instances. The solid line is a simple regression slope.

Further statistical analysis confirms these observations. When *Intensity delta* and *F0 delta* are entered as factors in a binary logistic regression model to predict the prosodic marking judgments, *Intensity delta* turns out to be the best predictor: a model with only *Intensity delta* allows for 81.2% of instances to be correctly classified as marked or not marked, which is significantly better than chance (-2 log likelihood=521.57, χ^2=57.87, df=1, p<0.001). A model with *Pitch delta* added yields a non-significant improvement of fit (-2 log likelihood=518.23, χ^2=3.34, df=1, p=0.07) and exactly the same proportion of correctly classified instances. This is to be expected given the strong correlation between *Intensity delta* and *F0 delta*: once one is in the model, the other can add little predictive power.

Turning now to articulation rate, Figure 2 plots the measures of intensity difference and articulation rate difference (*Intensity delta* and *Proportional rate*) against each other. We now see a concentration of white data points in the top half of the graph, which reflects again that 'prosodic marking' is mostly associated with a positive difference in intensity maximum between repair and reparandum – in other words, a relatively loud repair. The intensity and articulation rate measures are clearly not strongly correlated with each other (Pearson's r=-0.06, p=0.14), and it is difficult to discern any clustering in relation to articulation rate. Still, it is just about visible that the white data points are clustered a little more heavily towards the left of the figure than the grey data points: for example, out of the ten data points with the lowest values for *Proportional rate*, six were classified as 'prosodic marking', while out of the ten data points with the highest values, only two were. Further analysis again confirms this observation: there is a relationship between the articulation rate figures and the prosodic marking judgments such that instances classified as 'marked' have a lower mean for *Proportional rate* (1.12, SD=0.39) than instances classified as 'unmarked' (1.22,

SD=0.37). This means that 'prosodically marked' instances show less speeding up between reparandum and repair than 'unmarked' instances. Taken on its own, this difference is large enough to reach statistical significance (T=-2.55, df=582, p=0.01).

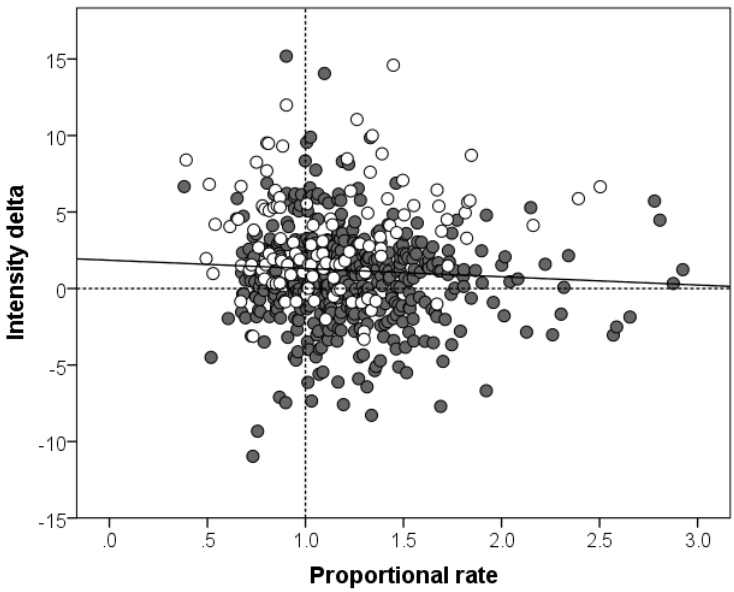


Figure 2.Correlation between measures of articulation rate and intensity maximum differences (*Proportional rate* and *Intensity delta*) between reparandum and repair. For *Proportional rate*, a value above one (marked by the dotted line on the horizontal axis) means the repair has a higher articulation rate than the reparandum. For *Intensity delta*, a value above zero (marked by a dotted line) means the repair has a higher maximum than the reparandum. White data points represent instances of repair that were auditorily judged 'prosodically marked'; grey data points represent 'unmarked' instances. The solid line is a simple regression slope.

At this point we might be tempted to conclude that 'prosodic marking' *is* like the 'upgrading' described by Curl (2002) and Ogden (2006): 'prosodically marked' repairs in our data set have repair components produced with relatively high pitch and high intensity maxima, and at a relatively low articulation rate. However, it would not be appropriate to label a 'prosodically marked' repair 'upgraded' instead. First of all, it is worth pointing out that while there appears to be a significant relationship between the articulation rate figures and the prosodic marking judgments, in comparison with the combined effect of *Intensity delta* and *F0 delta* the effect of *Proportional rate* is very small. When *Proportional rate* is added to our binary logistic regression model predicting the prosodic marking judgments, the proportion of correctly classified instances goes up only marginally, from 81.2% to 81.5%, and the improvement of model fit is on the threshold of significance (-2 log likelihood=514.38, χ^2=3.85, df=1, p=0.05). In the resulting model, *Proportional rate* is not a significant predictor (B=-0.61, SE=0.319, Wald=3.61, df=1, p=0.58). This means that while we can have considerable confidence in reporting that 'prosodic marking' involves making the repair both louder and higher in pitch than the reparandum, we should be much more cautious in reporting that 'marking' also involves a relative slowing down of the repair.

More importantly, it is crucial to observe that both 'prosodically marked' and 'unmarked' instances have a mean value for *Proportional rate* above one. The

proportion of instances with a proportional rate below one – that is, with a decrease in tempo at the repair – is greater for instances classified as 'prosodically marked' (42% vs 27%), and this difference is statistically significant (χ^2=9.77, df=1, p=0.002). Still, for both 'prosodically marked' and 'unmarked' repairs, the norm is for the repair component to be produced at a higher articulation rate than the reparandum: most repairs involve an increase in speaking tempo after the reparandum, but 'prosodically marked' ones involve less of an increase than 'unmarked' ones. This means that our findings provide no clear support for suggestions that 'prosodic marking' can be achieved through a substantial increase or substantial decrease in tempo (Cutler 1983, Levelt/ Cutler 1983). Rather, our findings confirm Goffman's (1981) description of 'prosodically marked' repairs as generally involving an increase in speaking tempo – but it must be added that 'unmarked' repairs do as well, and 'marked' repairs involve *less* of an increase on average.

3.3. So is 'marking' 'upgrading'?

As indicated above, Cutler (1983) suggests that 'prosodic marking' in self-repair can be achieved in a variety of ways, but is typically done through high pitch and intensity on the repair component. Goffman's (1981) description of what he calls 'strident' repairs suggests that high pitch and intensity are accompanied by a high speaking tempo. The current analysis supports Goffman's description, with the proviso that a relative rise in articulation rate is not one that distinguishes 'marked' and 'unmarked' instances of repair from each other. In fact, 'prosodically marked' instances show lower rises than 'unmarked' instances, and are more likely to show a fall. As such, the phonetic features of 'prosodic marking' in self-repair *are* very similar to those of phonetic 'upgrading' as characterized by Curl (2002) and Ogden (2006): they involve a rise in pitch and intensity between the reparandum and the repair, and slowly articulated repairs relative to 'unmarked' instances. However, it should be clear that it would be misleading to label all 'marked' repairs 'upgraded': in a majority of cases they still show an increase in articulation rate between reparandum and repair, and therefore, more often than not they are phonetically distinct from the 'upgraded' repetition repairs of Curl (2002), the 'upgraded' second assessments of Ogden (2006), and so on.

Why should 'prosodic marking' in self-repair be distinct from 'upgrading' in this way? In a previous study of tempo in self-repair (Plug 2011), I have suggested that a prevalence for speeding up at the repair tallies well with the observation in conversation-analytic work that while other-repair – in which listeners notice an error and invite the speaker to repair it – is frequently delayed if not avoided altogether, speakers tend to initiate self-repair without delay (Schegloff/ Jefferson/ Sacks 1977, Schegloff 1979, Fox/ Jasperson 1995, Jasperson 2002, Rieger 2003). Seen in this light, speeding up at the repair is consistent with a general drive by speakers to enable a soonest possible resumption of post-repair talk (see Jasperson 2002: 278). As I have suggested before, an interpretation in terms of face-saving seems plausible: in self-repair, the speaker moves the discourse away from the error and its correction as quickly as possible, while in other-repair the listener gives the speaker as much opportunity as possible to do the same by delaying the initiation of repair. Goffman's (1981: 215-216) appeals to the degree to which a

speaker feels "embarrassed" about a verbal error, which he considers an important factor in accounting for the shape of the subsequent repair, are arguably consistent with the notion that face-saving considerations play a role in repair design.

If a drive to allow for a soonest possible resumption of post-repair talk informs the design of most instances of self-repair, including those that speakers do not aim to attach particular salience to, then it is not surprising that the temporal characteristics of instances that *are* marked as salient are distinct from the temporal characteristics of stretches of speech marked as salient in other contexts, such as assessment sequences. The upshot of this line of argument is that 'prosodic marking' in self-repair *can* be seen as a form of 'upgrading', but its result is not an 'upgraded' repair, as the latter's characteristic of slow, deliberate articulation is moderated by a general drive to get the repair done and return to trouble-free talk as quickly as possible.

4. Conclusion

In this paper I have reviewed references to variation along an 'upgrading–downgrading continuum' in the literature on sound patterns in interaction, and investigated the extent to which 'prosodic marking' in self-repair fits this continuum. If my account is along the right lines, 'prosodic marking' in self-repair can be seen as a form of 'upgrading', but one that does not result in 'prosodically marked' repairs having all of the phonetic characteristics that phonetically 'upgraded' repetition repairs, second assessments, turn relaunchings and so on share. The proposed reason for the difference in output, despite the similarity in process, is an interfering constraint on the temporal organisation of self-repair.

Interestingly, my account of prosodic marking is reminiscent of Local/ Walker's (2004) account of 'abrupt-joins' – a practice for talking through a transition relevance place into a turn constructional unit that implements a topic shift – both in terms of the phonetic features involved, and in terms of their plausible interpretation. Local/ Walker list among the main phonetic features of 'abrupt-joins' a noticeable increase in articulation rate at the end of the first turn constructional unit – with a quick return to normal tempo in the second – and a noticeable upstep in pitch and loudness at the start of the second turn constructional unit. This cluster of phonetic features is clearly quite similar to that of 'prosodically marked' self-repairs. In terms of interpretation, it makes sense to suggest that through the increase in tempo, a speaker producing an abrupt-join is holding the turn despite the transition relevance. Through the upstep in pitch and loudness, the speaker is simultaneously marking a "new beginning" (Couper-Kuhlen 2004) in the topical trajectory of their talk. Thus, speakers may manipulate one phonetic parameter orienting to one level of interactional organisation, and another parameter orienting to another level. Similarly, in self-repair pitch and loudness can be said to be manipulated for prominence marking, while tempo is manipulated to allow for maximum progressivity in turn construction. Seen in this way, 'prosodic marking' in self-repair is a compromise between emphasising the corrected talk and getting the job of correcting done as quickly as possible, in the same way that 'abrupt-joins' involve a compromise between holding the turn and marking a break in the topical trajectory.

Whether or not my account is on the right lines, the analysis I have presented demonstrates that the question "Is this another example of 'upgrading'?" is not necessarily an easy one to address. In the case at hand, less thorough analysis might well have led to a simple conclusion that 'prosodically marked' instances of self-repair could just as well be called phonetically 'upgraded', because their pitch and intensity characteristics clearly fit this label, and the context clearly bears some similarity to those in which 'upgrading' and 'downgrading' have been identified in previous work. But their temporal characteristics do not quite fit, and I have suggested that they do not quite fit in a way that is highly informative: the proposed notion of a drive to allow for a soonest possible resumption of post-repair talk provides general insight into how speakers do repair work, or at least generates hypotheses that can be tested in further studies. This is as it should be: the availability of cross-parametric descriptive labels does not obviate the need for open-minded, multi-parametric phonetic analysis in examining sound patterns in interaction – however useful the labels may be in establishing connections between clusters observed in different interactional contexts.

5. References

Couper-Kuhlen, Elizabeth (2004): Prosody and sequence organization: The case of new beginnings. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (eds.): Sound patterns in interaction. Amsterdam: John Benjamins, 335-376.

Couper-Kuhlen, Elizabeth (2011): Pragmatics and prosody: Prosody as social action. In: Bublitz, Wolfram/ Norrick, Neal R. (eds.): Foundations of pragmatics. Berlin: Mouton de Gruyter, 491-510.

Couper-Kuhlen, Elizabeth (2012): Exploring affiliation in the reception of conversational complaint stories. In: Peräkylä, Annsi/ Sorjonen, Marja-Leena (eds.): Emotion in interaction. Oxford: Oxford University Press, 113-146.

Curl, Traci S. (2002): The phonetics of sequence organization: An investigation of lexical repetition in other-initiated repair sequences in American English. PhD thesis, University of Colorado. Published 2009, Saarbrücken: VDM Verlag Dr. Mueller.

Curl, Traci S. (2004): 'Repetition repairs': The relationship of phonetic structure and sequence organisation. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (eds.): Sound patterns in interaction. Amsterdam: John Benjamins, 273-298.

Curl, Traci S. (2005): Practices in other-initiated repair resolution: The phonetic differentiation of 'repetitions'. In: Discourse Processes 39, 1-43.

Curl, Traci S./ Local, John/ Walker, Gareth (2006): Repetition and the prosody-pragmatics interface. In: Journal of Pragmatics 38, 1721-1751.

Cutler, Anne (1983): Speakers' conceptions of the function of prosody. In: Cutler, Anne/ Ladd, D. Robert (eds.): Prosody: Models and measurements. Heidelberg: Springer, 79-91.

Fox, Barbara A./ Jasperson, Robert (1995): A syntactic exploration of repair in English conversation. In: Davis, Philip W. (ed.): Alternative linguistics: Descriptive and theoretical models. Amsterdam: John Benjamins, 77-134.

French, Peter/ Local, John (1983): Turn-competitive incomings. In: Journal of Pragmatics 7, 701-715.

Goffman, Ervin (1981): Radio talk. In: Goffmann, Ervin (ed.): Forms of talk. Oxford: Blackwell, 197-330.

Howell, Peter/ Young, Keith (1991): The use of prosody in highlighting alterations in repairs from unrestricted speech. In: The Quarterly Journal of Experimental Psychology 43A, 733-758.

Jasperson, Robert (2002): Some linguistic aspects of closure cut-off. In: Ford, Cecilia E./ Fox, Barbara A./ Thompson, Sandra A. (eds.): The language of turn and sequence. Oxford: Oxford University Press, 257-286.

Kelly, John/ Local, John (1989): Doing phonology. Manchester: Manchester University Press.

Levelt, Willem J. M. (1989): Speaking: From intention to articulation. Cambridge, Mass.: The MIT Press.

Levelt, Willem J. M./ Cutler, Ann (1983): Prosodic marking in speech repair. In: Journal of Semantics 2, 205-217.

Levinson, Stephen C. (2012): Action formation and ascription. In: Sidnell, Jack/ Stivers, Tanya (eds.): The handbook of conversation analysis. Oxford: Wiley, 103-130.

Lindström, Anna/ Sorjonen, Marja-Leena (2012): Affiliation in conversation. In: Sidnell, Jack/ Stivers, Tanya (eds.): The handbook of conversation analysis. Oxford: Wiley, 350-369.

Local, John (1992): Continuing and restarting. In: Auer, Peter/ di Luzio, A. (eds.): The contextualization of language. Amsterdam: John Benjamins, 273-296.

Local, John (2003): Variable domains and variable relevance: Interpreting phonetic exponents. In: Journal of Phonetics 31, 321-339.

Local, John/ Kelly, John (1986): Projection and 'silences': Notes on phonetic and conversational structure. In: Human Studies 9, 185-204.

Local, John/ Walker, Gareth (2004): Abrupt-joins as a resource for the production of multi-unit, multi-action turns. In: Journal of Pragmatics 36, 1375-1403.

Local, John/ Walker, Gareth (2005): Methodological imperatives for investigating the phonetic organisation and phonological structures of spontaneous speech. In: Phonetica 62, 120-130.

Local, John/ Walker, Gareth (2012): How phonetic features project more talk. In: Journal of the International Phonetic Association 42, 255-280.

Local, John/ Auer, Peter/ Drew, Paul (2010): Retrieving, redoing and resuscitating turns in conversation. In: Barth-Weingarten, Dagmar/ Reber, Elizabeth/ Selting, Margret (eds.): Prosody in interaction. Amsterdam: John Benjamins, 131-159.

Local, John/ Kelly, John/ Wells, Bill (1986): Towards a phonology of conversation: turn-taking in urban Tyneside speech. In: Journal of Linguistics 22, 411-437.

Local, John/ Wells, Bill/ Sebba, Mark (1985): Phonology for conversation: Phonetic aspects of turn delimitation in London Jamaican. Journal of Pragmatics 9: 309-330.

Ogden, Richard (2006): Phonetics and social action in agreements and disagreements. In: Journal of Pragmatics 38, 1752-1775.

Ogden, Richard (2010): Prosodic constructions in making complaints. In: Barth-Weingarten, Dagmar/ Reber, Elizabeth/ Selting, Margret (eds.): Prosody in interaction. Amsterdam: John Benjamins, 81-103.

Ogden, Richard (2012a): The phonetics of talk in interaction – introduction to the special issue. In: Language and Speech 55, 3-11.

Ogden, Richard (2012b): Prosodies in conversation. In: Niebuhr, Oliver (ed.): Understanding prosody: The role of context, function and communication. Berlin: Mouton De Gruyter, 201-218.

Oostdijk, Nelleke (2002): The design of the Spoken Dutch Corpus. In: Peters, Pam/ Collins, Peter/ Smith, Adam (eds.): New frontiers of corpus research. Amsterdam: Rodopi, 105-113.

Plug, Leendert (2005): The phonetics of repetition in other-initiated repair in Dutch: A comparative exercise. York Papers in Linguistics Series 2 (3): 145-171.

Plug, Leendert (2009): On tempo in dispreferred turns: A recurrent pattern in a Dutch corpus. In: Barth-Weingarten, Dagmar/ Dehé, Nicole/ Wichmann, Anne (eds.): Where prosody meets pragmatics. Bingley: Emerald, 227-258.

Plug, Leendert (2011): Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. In: Journal of Phonetics 39, 289-297.

Plug, Leendert/ Carter, Paul (2013): Prosodic marking, pitch and intensity in spontaneous lexical self-repair in Dutch. In: Phonetica 70, 155-181.

Pomerantz, Anita (1984): Agreeing and disagreeing with assessments: Some features of preferred/ dispreferred turn shapes. In: Atkinson, John M./ Heritage, John (eds.): Structures of social action. Cambridge: Cambridge University Press, 57-101.

Rieger, Caroline L. (2003): Repetitions as self-repair strategies in English and German conversations. In: Journal of Pragmatics 35, 47-69.

Schegloff, Emanuel (1979): The relevance of repair to syntax-for-conversation. In: Syntax and Semantics 12, 261-286.

Schegloff, Emanuel/ Jefferson, Gail/ Sacks, Harvey (1977): The preference for self-correction in the organization of repair in conversation. In: Language 53, 361-382.

Selting, Margret (2010): Prosody in interaction: State of the art. In: Barth-Weingarten, Dagmar/ Reber, Elizabeth/ Selting, Margret (eds.): Prosody in interaction. Amsterdam: John Benjamins, 3-40.

Shattuck-Hufnagel, Stefanie/ Cutler, Anne (1999): The prosody of speech error corrections revisited. In: Proceedings of the Fourteenth International Congress of Phonetic Sciences (ICPhS 1999). Vol. 2. San Francisco, 1483-1486.

Smith, Rachel (2013): New directions in speech perception. In: Jones, Mark J./ Knight, Rachael-Anne: Bloomsbury companion to phonetics. London: Bloomsbury, 227-241.

Van Son, Rob J.J.H./ Pols, Louis C. W. (1990): Formant frequencies of Dutch vowels in a text, read at normal and fast rate. In: Journal of the Acoustical Society of America 88, 1683-1693.

Van Son, Rob J.J.H., Rob/ Pols, Louis C. W. (1992): Formant movements of Dutch vowels in a text, read at normal and fast rate. In: Journal of the Acoustical Society of America 92, 121-127.

Walker, Gareth (2007): On the design and use of pivots in everyday conversation. In: Journal of Pragmatics 39, 2217-2243.

Walker, Traci (2010): Repetition and contrast across action sequences. In: Barth-Weingarten, Dagmar/ Reber, Elizabeth/ Selting, Margret (eds.): Prosody in interaction. Amsterdam: John Benjamins 263–268.

Zellers, Margaret/ Ogden, Richard (2013): Exploring interactional features with prosodic patterns. In: Language and Speech. Online advance access, doi: 10.1177//002383091 3504568.

Leendert Plug
Linguistics and Phonetics, School of Modern Languages and Cultures
University of Leeds
Leeds LS2 9JT
United Kingdom
L.Plug [at] leeds.ac.uk

'Mitfühlend sprechen': Zur Rolle der Prosodie in Empathiedarstellungen[1]

Maxi Kupetz

Abstract
Im Zentrum dieser interaktional-linguistischen Untersuchung steht die Frage, welchen Beitrag die prosodisch-phonetische Gestaltung einer Äußerung zu ihrer Interpretation als Darstellung von Empathie leistet. Dem liegt eine interaktionale Auffassung von Empathie zu Grunde: Empathie wird nicht als psychoemotionaler Zustand konzeptualisiert, sondern als Darstellung von Verstehen/ Verständnis/ Mitgefühl/ Mitleid in sozialer Interaktion. Gegenstand der Untersuchung ist die interaktive Bearbeitung persönlicher Erlebnisse in deutschen Alltagsgesprächen. Es wird gezeigt, dass – an spezifischen sequentiellen Positionen – Kombinationen folgender prosodisch-phonetischer Ressourcen potenziell relevant sind, um Äußerungen als empathisch, hier i.S.v. 'mitfühlend', interpretierbar zu machen: geringere und/ oder abnehmende Lautstärke/ Diminuendo, tiefes Tonhöhenregister, flach auslaufende Kontur, Behauchung, Knarrstimme, geringe Sprechspannung/ weiche Artikulation, Silbendehnung, Lippenrundung, rhythmische Integration/ Legato-Rhythmus.

Empathie – Affektivität – Verstehen/ Verständnis – Konversationsanalyse – Interaktionale Linguistik – Alltagsgespräche

This interactional-linguistic study aims at revealing which prosodic-phonetic resources can be used to make an utterance interpretable as a display of empathy when personal experiences are negotiated in German everyday interaction. Empathy refers to displays of understanding and/ or sympathy and/ or compassion for the emotional situation of the other. It will be shown that, in specific sequential positions, (clusters of) the following prosodic-phonetic resources are possibly relevant to make an utterance interpretable as affect-oriented and 'sympathetic': low and/ or decreasing volume/ diminuendo, low pitch register, flat contours, breathiness, creak, soft articulation, lengthening, lip rounding, rhythmic integration, legato.

[1] Mein herzlicher Dank gilt Margret Selting, Ingmar Rothe und den Herausgeberinnen des Bandes für ihre detaillierten Kommentare zu früheren Versionen des Beitrags.

1. Einleitung

Der Zusammenhang zwischen Prosodie und dem Ausdruck und der Identifikation von Emotionen ist lange anerkannt und bereits in einer Vielzahl emotionspsychologischer und sprechwissenschaftlicher Untersuchungen beschrieben worden (vgl. u.a. Tischer 1993, Banse/ Scherer 1996, Kehrein 2002, Kranich 2003, Paeschke 2003, Wendt 2007). Zumeist beruhen diese auf experimentellen Sprachaufnahmen, Hörexperimenten und detaillierten akustischen Analysen phonetisch-prosodischer Parameter schauspielerischer Emotionsausdrücke. Auch konversationsanalytische bzw. interaktional-linguistische Studien haben, basierend auf Aufnahmen natürlicher Gespräche, deutlich die Rolle der Prosodie für verschiedene Typen von Affektdarstellungen zeigen können (vgl. u.a. Selting 1994, 2010a, Local/ Walker 2008, Couper-Kuhlen 2009, Ford/ Fox 2010). Für die Darstellung von Empathie und/ oder Mitleid im Speziellen gibt es bereits einige interaktionsorientierte Untersuchungen, die prosodische Verfahren in den Blick nehmen (vgl. u.a. Hepburn/ Potter 2007, 2012, Reber 2009, 2012, Couper-Kuhlen 2012a, Weiste/ Peräkylä 2014). Diese Studien beziehen sich jedoch auf englischsprachige und finnischsprachige Daten. In bisherigen interaktionalen Untersuchungen zu Affektivität im Deutschen wurden prosodische Ressourcen zur Darstellung von Mitgefühl/ Mitleid/ Empathie eher 'am Rande' beschrieben (vgl. u.a. Schwitalla 1995: 230ff., Günthner 2000: 251ff., Selting 2010a: 254, Kupetz 2013: 89ff.). Selting macht einige Forschungsdesiderata der interaktionalen Prosodie-Forschung explizit: "What role do phonetic-prosodic parameters and bundles of features play in the signaling of emotive involvement, affect and stance in interaction? Which phonetic-prosodic parameters are used to signal which affects?" (2010b: 23).

Im Fokus des vorliegenden interaktional-linguistischen Beitrags steht die Beschäftigung mit der Rolle der Prosodie für Empathiedarstellungen in deutschen Alltagsgesprächen. Der Untersuchung liegt eine interaktionale Auffassung von Empathie zu Grunde: Empathie wird nicht als psycho-emotionaler Zustand konzeptualisiert, sondern als Darstellung des Verstehens der emotionalen Situation einer betroffenen Person bzw. als Darstellung von Verständnis oder Mitgefühl für die emotionale Situation einer betroffenen Person (vgl. z.B. Hepburn/ Potter 2007: 99ff., Couper-Kuhlen 2012a: 122f.). In Anlehnung an Maynard (1997: 94) werden die Personen als 'Betroffene' (*consequential figure)* bezeichnet, die von einem Ereignis betroffen sind und deren persönliche Erfahrungen und (emotionale) Situationen, die im Zusammenhang mit diesem Ereignis stehen, in der Interaktion verhandelt werden. Diese 'Betroffenen' können, müssen aber nicht im Gespräch anwesend ein.

Gegenstand der vorliegenden Betrachtung sind Reaktionen auf affektgeladene Erzählungen persönlicher Erlebnisse bzw. die interaktionale Aus- und Behandlung emotionalen Erlebens von anwesenden Betroffenen. Im Fokus steht affektives Erleben negativer Valenz. Folgenden Fragen wird dabei nachgegangen: Welchen Beitrag leistet die prosodische Realisierung zur Interpretation einer Handlung in ihrer holistischen Gestalt als Darstellung von Empathie bzw. Mitgefühl bzw. Mitleid? In welchem Verhältnis stehen prosodisch-phonetische und andere Ressourcen? Wie kann 'mitfühlendes Sprechen' (als Teilnehmerkategorie) analysiert werden?

Im Folgenden werden zuerst bisherige Forschungsergebnisse interaktionaler Arbeiten zu Prosodie und Empathiedarstellungen zusammengefasst und darin verwendete Begrifflichkeiten erörtert (Abschnitt 2). Anschließend werden kurz das verwendete Korpus und die methodischen Überlegungen, die der Untersuchung zu Grunde liegen, beschrieben (Abschnitt 3). Dem folgen ausführliche Analysen von fünf Beispielen aus deutschen Alltagsgesprächen, anhand derer veranschaulicht wird, welche prosodisch-phonetischen Parameter 'mitfühlendes Sprechen' bzw. affektorientierte Empathiedarstellungen ausmachen und welche Funktionen diese in ihrem jeweiligen sequentiellen Kontext haben (Abschnitt 4). Abschließend werden die Beobachtungen zusammengefasst und kurz mit Ergebnissen einschlägiger Untersuchungen zum Englischen verglichen (Abschnitt 5).

2. Prosodie und Mitgefühl/ Mitleid: Forschungsstand und Begriffserläuterung

Eine der wenigen deutschsprachigen Arbeiten, in der die Rolle der Prosodie für den Ausdruck von Mitgefühl und/ oder Mitleid deutlich wird, stammt von Schwitalla (1995). Er beschreibt für die deutsche Alltagssprache ein "Repertoire standardisierter Mittel" zur Darstellung von Mitleid, das u.a. Stimmfärbungen (als Ausdruck von Trauer, Sorge und Mitleid), Interjektionen (z.B. *och mann*, *ach gott*), wiederholte Intonationskonturen und Mitleidslaute/ expressive Laute (z.B. gedehntes *m::* und *ou:* oder auch Seufzen) umfasst (ebd.: 229ff.). Mitgefühl bzw. Mitleid wird in seinen Daten auch durch 'bekümmertes' oder 'klagendes' Sprechen nahe gelegt (ebd.: 230). Auf die prosodischen Parameter, die diese Sprechweisen und Stimmfärbungen ausmachen, wird nicht näher eingegangen.

In der deutschsprachigen Literatur (z.B. Fiehler 1990, Schwitalla 1995) werden Untersuchungsgegenstände aus dem Phänomenbereich Mitleid/ Mitgefühl/ Anteilnahme von den Autoren nicht zwangsläufig explizit mit dem Begriff Empathiedarstellung zusammengebracht. In der englischsprachigen Literatur argumentieren die Autoren z.T. unterschiedlich hinsichtlich dessen, was die vermeintlichen Teilnehmerkategorien *empathy* (hier: Empathie, Einfühlungsvermögen, Mitgefühl) und *sympathy* (hier: Anteilnahme, Bedauern, Trost, Mitleid, Mitgefühl) ausmacht. Gerade im Bereich 'Mitgefühl' ist sowohl eine adäquate Abgrenzung der Kategorie als auch eine adäquate Übersetzung vom Englischen ins Deutsche oftmals schwierig, wie die Vielfalt der verwendeten Begriffe zur Beschreibung zum Teil sehr ähnlicher kommunikativer Phänomene zeigt.

Reber (2009) fokussiert in ihrer Untersuchung englischer Alltags- und Radiogespräche mit Anruferbeteiligung die Lautobjekte <<t>[ɑ:].> und <<t>[hɑ:].> als Mittel zur Darstellung 'konkordanter Affektivität' bzw. Mitgefühl und beschreibt hierfür als potentiell relevante prosodisch-phonetische Parameter tief fallende, flach auslaufende Konturen, rhythmische Integration, Knarrstimme und/ oder ähnliche Lautstärke wie im vorangegangenen Redebeitrag desselben Sprechers. In dieser prosodisch-phonetischen Form kann 'ah'[2] je nach sequentieller Position zur Darstellung von *empathy* (hier: Mitgefühl) oder auch zur Darstellung von *sympa-*

[2] Die Darstellung der Lautobjekte in diesem Beitrag entspricht der Darstellung der Lautobjekte der Autorin in den jeweiligen Publikationen.

thy (hier: Mitleid, Bedauern, Anteilnahme) verwendet werden (Reber 2012). In Problemgesprächen wird 'ah' verwendet, um die vorherige Äußerung über ein Problem als vollständig zu behandeln und gleichzeitig erhöhte emotionale Beteiligung zu signalisieren, was in Anlehnung an Jefferson (1988) als Darstellung von *empathy* (hier: Mitgefühl) bezeichnet wird (Reber 2012: 179ff.). In Reaktionen auf Ankündigungen schlechter Neuigkeiten durch die *consequential figure* selbst wird vom Rezipienten das Lautobjekt 'ah in der oben beschriebenen prosodisch-phonetischen Form eingesetzt und in der Regel im nachfolgenden Turn durch idiomatische Ausdrücke wie z.B. [*poor* + noun]-Konstruktionen oder *I'm so sorry* als Darstellung von *sympathy* (hier: Mitleid, Bedauern, Anteilnahme) explizit gemacht (ebd.: 198ff.). Durch diese Verwendungsweisen in unterschiedlichen sequentiellen Kontexten wird nahe gelegt, dass es sich bei 'ah' zur Darstellung von *empathy* und 'ah' zur Darstellung von *sympathy* um zwei Teilnehmerkategorien handelt.

Couper-Kuhlen (2012a) betont, dass in affektgeladenen Beschwerdeerzählungen in englischen Alltagsgesprächen die Anpassung der Rezipientenreaktion an die vorangegangene Äußerung hinsichtlich der zeitlichen Realisierung (ohne Verzögerung) und hinsichtlich der generellen prosodischen Gestaltung (Tonhöhenverlauf, Akzentstärke u. ä.) wichtig ist, damit die Rezipientenreaktion als affiliativ oder gar empathisch interpretiert werden kann. Weitere Untersuchungen haben gezeigt, dass im sequentiellen Kontext von Anfragen und Ablehnungen an dritter Position Äußerungen mit 'gedämpfter' prosodischer Gestalt (*subdued prosody*) verwendet werden, um Enttäuschung interpretierbar zu machen (Couper-Kuhlen 2009, 2011, 2012b). Die relevantesten prosodischen Parameter 'gedämpfter' Äußerungen sind – im Vergleich zur Normallage des jeweiligen Sprechers – niedrigere Lautstärke, niedrigeres Tonhöhenregister, Behauchung und schwache Artikulation. Diese Beobachtungen sind deshalb für den vorliegenden Beitrag von Bedeutung, weil eine ähnliche prosodische Gestalt in einem anderen sequentiellen Kontext auch der Darstellung von Mitleid (*sympathy*) dienen könnte, wie Couper-Kuhlen deutlich hervorhebt (2009: 107).

Ähnliche Parameter finden Hepburn/ Potter (2007, 2012) in Reaktionen auf Weinen in Gesprächen eines Kinderschutznotrufs in Großbritannien. Sie unterscheiden in ihrer Untersuchung empathische Reaktionen (*empathic receipts*) von ausschließlich vokal realisierten mitfühlenden Reaktionen (*sympathic receipts*). Für Erstere ist ein propositionaler Gehalt konstitutiv, der das Verstehen des Zustands des Gegenübers deutlich macht: "ERs [empathic receipts, MK] are made up of two elements: 1. A formulation of the crying party's mental state. 2. A marker of the contingency or source of the mental state formulation" (2007: 109). Letztere machen die Autoren am prosodischen Format fest:

> usually stretched, sometimes with elevated or rising-falling pitch and/ or creaky delivery, sometimes explicitly involving some kind of token such as 'oh' or 'aw', sometimes with softened volume and increased 'breathiness' or aspiration (2012: 201).

Diese *sympathic receipts* werden für Handlungen wie Mitfühlen oder Mitleiden (*sympathizing*) und/ oder Trösten und Beruhigen (*soothing*) eingesetzt (ebd.).

Hepburn/ Potter (2007, 2012) und Reber (2012) unterscheiden, mit unterschiedlichen Argumenten begründet, zwischen Mitgefühl (*empathy*) und Mitleid

(*sympathy*) und beschreiben diese als unterschiedliche Teilnehmerkategorien.[3] Eine rein formbasierte Funktionszuschreibung, die besagt, ein *empathic receipt* konstituiere sich auf Grund eines bestimmten propositionalen Gehalts und ein *sympathic receipt* aus bestimmten prosodischen Parametern (Hepburn/ Potter 2007, 2012), scheint jedoch aus zwei Gründen diskussionswürdig: Erstens kann anhand der sequentiellen Behandlung der vermeintlichen Empathie- oder Mitleidsdarstellungen durch die Teilnehmer selbst nicht eindeutig gezeigt werden, dass es sich um zwei distinkte Teilnehmerkategorien handelt. Zweitens indiziert eine Darstellung von Mitgefühl immer auch eine Interpretation des bisherigen interaktionalen Geschehens, hier also ein bestimmtes Verstehen der (emotionalen) Situation des Anderen, und macht dieses Verstehen verhandelbar.[4] Aus diesem Grund plädiere ich für einen großzügigen Umgang mit dem Begriff der Empathiedarstellung, der Darstellungen von Verstehen/ Verständnis, von Mitgefühl und/ oder von Mitleid als Spielarten von Empathie einschließt. Eine solche Konzeption von Empathiedarstellungen eröffnet den Raum für die Betrachtung von Handlungen, die nicht eindeutig verstehens- oder affektorientiert sind. Vielmehr lassen bestimmte Ressourcen, wie z.B. Lautobjekte, eben genau offen, ob sie sich auf kognitive oder affektive Zustände beziehen (Reber 2009: 198ff.). Teilnehmer können sich im Verlauf eines Gesprächs, in dem emotionales Erleben verhandelt wird, genau dieser Vagheit bedienen und Reaktionen nuancieren, sodass eine trennscharfe Unterscheidung der Kategorien wenig hilfreich erscheint (vgl. Kupetz 2014 zur Veränderung der Verfahren im sequentiellen Verlauf einer längeren Affektbehandlung).

3. Methodische Überlegungen und Korpusbeschreibung

Aus der Zusammenfassung bisheriger Untersuchungen wird bereits deutlich, dass im Englischen prosodisch-phonetische Parameter wie Knarrstimme, Behauchung, Silbendehnung, geringe Lautstärke und tiefes Tonhöhenregister eine wichtige Rolle für affektorientierte Empathiedarstellungen bzw. Darstellungen von Mitgefühl/ Mitleid zu spielen scheinen. In dieser Untersuchung geht es darum, die prosodisch-phonetischen Parameter mitfühlenden Sprechens in deutschen Alltagsgesprächen herauszuarbeiten und gegebenenfalls das Verhältnis prosodischer Parameter zu anderen, z.B. lexiko-semantischen, Ressourcen in Darstellungen von

[3] Auch Ruusuvuori (2005) unterscheidet zwischen *empathy* als Orientierung der Teilnehmer am Verstehen des Gegenübers und *sympathy* als Orientierung der Teilnehmer am Mitfühlen/ Teilen emotionaler Erfahrungen. Ihre Argumentation basiert jedoch auf detaillierten Sequenzanalysen, die zwar lexiko-semantischen und z.T. syntaktischen Ressourcen Beachtung schenkt. Prosodische Ressourcen stehen jedoch nicht im Fokus der Untersuchung.

[4] In anderen Disziplinen (u.a. (Sozial-)Psychologie, Kognitionswissenschaft und Neurowissenschaft) ist die Relation zwischen Kognition und Emotion seit langem Forschungs- und Diskussionsgegenstand (vgl. allgemein z.B. Scherer/ Ekman 1984; im Hinblick auf Empathie z.B. Duan/ Hill 1996). Aus Sicht eines Interaktionsforschers ist zu vermuten, dass im Moment der Darstellung von Mitgefühl immer auch eine Art von Verstehen oder Verständnis im Sinne einer Anerkennung der Situation des Gegenübers nahe gelegt wird. Selbst wenn sich im folgenden sequentiellen Verlauf herausstellt, dass es sich bei der Darstellung von Mitgefühl um eine inadäquate Einschätzung der Situation des Gegenübers handelt, z.B. wenn die Darstellung mit *nee, so schlimm fand ich das gar nicht* zurückgewiesen wird, so ist dies nur möglich, da vorher eine bestimmte Interpretation und entsprechende affektive Haltung (Mitgefühl) überhaupt relevant und somit verhandelbar gemacht wurde.

Mitgefühl näher zu beleuchten. Es wird dabei davon ausgegangen, dass Prosodie immer nur im Zusammenspiel mit anderen verbalen und kinetischen[5] Ressourcen in einem bestimmten sequentiellen Kontext interpretierbar wird (vgl. u.a. Couper-Kuhlen/ Selting 1996, Selting 2013). Betrachtet werden Sequenzen, in denen (a) Affektivität entweder thematisiert und/ oder durch multimodale Ressourcen kontextualisiert wird, z.B. in Erzählungen persönlicher Erfahrungen, und in denen (b) in Reaktion darauf Verständnis/ Mitgefühl/ Mitleid für die betroffene Person zum Ausdruck gebracht wird (vgl. u.a. Fiehler 1990, Reber 2009).

Im Rahmen einer konversationsanalytischen Herangehensweise zur Erforschung von Affektivität sind, wie Couper-Kuhlen (2009) ausführt, einige grundlegende Prämissen zu beachten:

> CA *can* make a specific contribution to the study of affect and emotion if we adopt the following as axiomatic:
>
> - Affect and emotion are performed as *displays* […] in interaction.
> - These displays are realized as *embodied practices*.
> - The practices are *situated* at specific sequential positions within interaction.
> - The practices are interpreted in a *context-sensitive* fashion.
>
> (Couper-Kuhlen 2009: 96, Hervorhebg. im Original)

Im ersten Teil des empirischen Kapitels (4.1) werden zwei Fälle vorgestellt, in denen die Interaktionsteilnehmerinnen selbst durch die Verwendung lexikosemantischer Ressourcen ihre affektive Haltung als 'mitfühlend' oder 'mitleidend' kategorisieren. Diese expliziten Beispiele eignen sich, um herauszuarbeiten, welche prosodisch-phonetischen Ressourcen grundsätzlich überhaupt relevant sein können, um 'mitfühlendes' Sprechen interpretierbar zu machen. Konversationsanalytischen Maximen entsprechend (siehe z.B. Sidnell 2013) soll somit sichergestellt werden, dass es sich bei 'mitfühlendem Sprechen' nicht um ein bloßes analytisches Konstrukt handelt, sondern dass TeilnehmerInnen diese spezifische prosodische Gestaltung von Äußerungen tatsächlich als affektorientiert interpretieren. Dies entspricht der Forderung von Local/ Walker (2008: 704f.) nach einer ganzheitlichen Betrachtung interaktionsorganisatorischer und sequentieller sowie lexikalischer, syntaktischer und phonetischer Details, die die Teilnehmerorientierung an der dargestellten Affektivität in den Vordergrund stellt, anstatt aus einer Analytikerperspektive heraus einzelnen Äußerungen affektive Attribute zuzuschreiben.

In einem zweiten Schritt (4.2) wird der Fall einer Pseudodarstellung behandelt, da er Rückschlüsse auf das metakommunikative Wissen von TeilnehmerInnen über die relevanten prosodisch-phonetischen Ressourcen für Mitgefühlsdarstellungen zulässt (Reber 2009).

In einem dritten Schritt (4.3) werden implizite Darstellungen von Mitgefühl behandelt. Dadurch soll verdeutlicht werden, dass die zuvor beschriebenen proso-

[5] Der Begriff ‚Kinesik' umfasst die Beschreibung aller Handlungen, die durch Körperbewegungen entstehen können. Darunter fallen u.a. Mimik, Gestik, Körperhaltung, Proxemik und Blickverhalten (Bußmann 2002: 344). Die Beschreibung dieser Ressourcenebene als ‚kinetisch' wird hier dem Begriff ‚visuell' (vgl. u.a. Deppermann 2013, Selting 2013) vorgezogen, um alle semiotischen Ressourcen gleichermaßen produktionsseitig zu beschreiben: verbal, vokal und kinetisch/ kinesisch. Der Begriff der ‚visuellen' Ressource ist dementsprechend als Subkategorie kinetischer Ressourcen für das Blickverhalten reserviert.

disch-phonetischen Merkmalsbündel eingesetzt werden können, um Empathie durch 'mitfühlendes Sprechen' darzustellen, auch wenn die Lexiko-Semantik der Äußerung allein dies nicht unbedingt nahe legt.

Der Analyse liegt ein interaktionales Verständnis von Prosodie zu Grunde, d.h. alle hörbaren Parameter, "all audible aspects in and of speech that are produced by the human vocal apparatus" (Ford/ Couper-Kuhlen 2004: 3), werden betrachtet, sofern sie potentiell interaktional relevant sind, d.h. als ‚kommunikative Signale' (Selting 2010b: 5) verwendet werden. Im Speziellen geht es darum, herauszuarbeiten, welche prosodisch-phonetischen Parameter 'mitfühlendes Sprechen' kontextualisieren. Von der Betrachtung ausgeschlossen werden Artikulationsweisen und Stimmqualitäten, die auf physiologische Besonderheiten der Sprecher zurückzuführen sind, z.B. Flüsterstimme auf Grund einer Erkältung (Szczepek Reed 2011: 13). Ausschlaggebend für die Entscheidung, ob es sich um ein interaktional relevantes oder physiologisch bedingtes Phänomen handelt, sind detaillierte Sequenzanalysen und ggf. Vergleiche von spezifischen Äußerungen mit Äußerungen in Normallage des Sprechers/ der Sprecherin (zur Problematik der Beschreibung prosodischer Kategorien als Teilnehmerkategorien siehe Selting 1995: 20ff.). Zusätzlich wird der Frage nachgegangen, welche Rolle die prosodische Anpassung einer Äußerung an die prosodische Form der Vorgängeräußerung spielt, um 'mitfühlend' zu wirken (Couper-Kuhlen 2012a).[6] Prosodische Anpassung (*prosodic matching*) ist eine Form der prosodischen Orientierung und kann auf unterschiedlichen Ebenen erfolgen (Szczepek Reed 2006): Tonhöhenverlauf, Tonhöhenregister, Tonhöhensprünge, Lautstärke, Sprechgeschwindigkeit, Stimmqualität, Lautproduktion, Rhythmus (z.B. Auer/ Couper-Kuhlen 1994). Bei der prosodischen Anpassung kann das prosodische Muster der Vorgängeräußerung entweder 'heruntergespielt', 'verstärkt' oder 'komplementiert' werden (z.B. Ogden 2006, Szczepek Reed 2006; siehe auch Maynard/ Freese 2012 zur prosodischen Orientierung beim Erzählen und Aufnehmen guter und schlechter Nachrichten im Englischen).

Das Korpus der vorliegenden Untersuchung umfasst 12,5 Stunden deutscher Alltagsgespräche unter Freunden, die mit mindestens zwei Kameras und einem separaten Audioaufnahmegerät aufgezeichnet wurden. Gut neun Stunden Videomaterial wurden unter Leitung von Margret Selting und Elizabeth Couper-Kuhlen im Rahmen des Projekts "Emotive involvement in conversational storytelling" im Exzellenzcluster "Languages of Emotion" der Freien Universität Berlin erhoben (Kennzeichnung LoE). Zusätzlich wurden drei weitere Stunden aufgezeichnet, wobei den InteraktionsteilnehmerInnen in diesen Gesprächen vorgegeben wurde, unter anderem über 'Momente des Glücks und Momente des Verlusts' zu sprechen (Kennzeichnung AGmT, Überlegungen zu dieser Art von Daten siehe Kupetz 2014). Die Daten wurden nach den Konventionen des Gesprächsanalytischen Transkriptionssystems GAT 2 (Selting et al. 2009) transkribiert. Anmerkungen zur Erweiterung der Konventionen befinden sich im Anhang dieses Beitrags. Die Namen der Sprecher und Sprecherinnen wurden geändert.

[6] Andere Verfahren, z.B. der Vergleich der prosodischen Gestalt einer empathischen/ mitfühlenden Äußerung mit der 'Normallage' des Sprechers, wurden nur in uneindeutigen Fällen hinzugezogen. Somit wird der konversationsanalytischen Prämisse Rechnung getragen, dass die Interaktionsteilnehmer kontextsensitiv die Äußerungen des Gegenübers interpretieren und im Hinblick darauf lokal ihre Äußerungen gestalten (vgl. z.B. Bergmann 2001).

4. Empirische Analysen

4.1. Prosodische Verfahren in expliziten Mitgefühlsdarstellungen

In diesem Teilkapitel werden Fälle behandelt, in denen Mitgefühl/ Mitleid von den TeilnehmerInnen explizit gemacht wird, d.h. Mitgefühl/ Mitleid wird auf lexiko-semantischer Ebene nahe gelegt. Diese Äußerungen signalisieren in der Regel nicht nur lexiko-semantisch, sondern auch prosodisch 'Mitfühlen'. Die Gestalt dieser Äußerungen wird also 'zerlegt' (Selting/ Couper-Kuhlen 2000: 84), um potentiell relevante prosodisch-phonetische Merkmalsbündel herauszuarbeiten, die den Höreindruck 'mitfühlenden Sprechens' ausmachen. Indem aus Teilnehmerperspektive Evidenz für die Interpretation einer kommunikativen Handlung als 'mitfühlend' oder 'mitleidend' geliefert wird, soll einer voreiligen Zuschreibung affektiver Zustände zu prosodisch-phonetischen Formen entgegengewirkt werden (Local/ Walker 2008).

In Beispiel (1) wird vergangenes Erleben rekonstruiert. Eine Teilnehmerin bringt gegenüber der anwesenden 'Betroffenen' ihre affektive Haltung hinsichtlich einer vergangenen Situation explizit zum Ausdruck. Sie kategorisiert diese affektive Haltung selbst als 'Mitgefühl'. Heike (Hei), Hanna (Han) und Birte (Bir) sprechen darüber, dass sie in letzter Zeit wenig weggegangen bzw. beim Weggehen schnell müde geworden sind und immer wenig Alkohol getrunken haben:

(1) LoE_VG_01_Kater (10:52)

```
   18   Han:   und wIr ham (.) s nIch ma eine
               FLAsche ge[schafft zu zweit;]
   19   Bir:             [A:H NEE:;        ]
   20          [man SCHAFFT auch wIrklich nIchts] mehr;=
   21   Hei:   [<<pp> (ja;) man wird A:LT dU;>  ]
   22   Bir:   [|=wie dU ↑!AUF!stehen konntest;       ]nE,
                |((zu Hei blickend))
   23   Han:   [<<pp, behaucht> (s gEht ja GAR nich.)>]
   24   Bir:   ich l |am DONnersta:g,
   25   Hei:         |((zieht die Schultern kurz hoch und nickt mehrmals,
                       lacht leise))
   26   Bir:   °h ich bin AUFgewacht und |dAchte,
                                         |((sich die Hand kurz vor
                                          den Mund haltend))
   27          es war hAlb ELF oder so;
-> 28   Bir:   <<all> (weil m) [ich] DACHte->=
   29   Hei:                  |[JA;]
                              |((leicht nickend))
-> 30   Bir:   =oh gOtt <<creaky, dim> ↓die |A:Rme hEi:`_ke.>
                                            |((sich die Hand vor
                                             den Mund haltend))
```

```
-> 31   Bir:   |<<soft, h> ich hAb so <<smile voice> mIt dir
               | ( (die Hand am Kinn haltend) )
   32   Hei:   | ( (sieht Bir an, nickt mehrmals leicht) )
        Bir:   ge[↑`-FÜH:LT->>]
   33   Hei:     [°hh         ]
   34          bei mIr war noch das GEILste-
   35          ich steh AUF,
   36          mach mich FERtig,
   37          geh ARbeiten,
   38          komm WIEder,
   39          frank GRINST mIch an-
   40          na warn gUter A:bend; ne,
```

Birte bewertet ihren veränderten Alkoholkonsum mit: *man SCHAFFT auch wIrklich nIchts mehr;=* (Segment 20) in Überlappung mit Heike, die scherzhaft das Alter als Grund einführt (Segment 21), und mit Hanna, die leise, abschließend bewertet (Segment 23). Durch schnellen Anschluss, Herstellung von direktem Blickkontakt mit Heike und den Einsatz stärkerer Tonhöhenbewegungen im Vergleich zur vorangegangenen Äußerung initiiert Birte eine neue, themenverwandte Episode (Segment 22) und präzisiert, auf welche Begebenheit sie anspielt (Segment 24).[7] Heike zeigt durch Nicken an (Segment 25), dass sie die Anspielung versteht und dass Birte fortfahren kann. Diese schildert, wie sie an dem Donnerstagmorgen aufwachte, und leitet eine Gedankenwiedergabe ein (Segment 26). Durch den Einsatz des verbum sentiendi *dachte* (Günthner 1997: 266ff.) wird deutlich, dass es sich um die Wiedergabe eines eigenen vergangenen inneren Monologs handelt. Diese wird jedoch durch eine Parenthese hinausgezögert (Segment 27), die im Sinne einer 'Deskription situativer Details' (Bergmann 1987: 147ff., Christmann/ Günthner 1996: 13ff.) der Nennung des konkreten Zeitpunkts dient. Birte fährt mit einer kurzen, schwer verständlichen Passage fort (Segment 28), die von Heikes Rezeptionssignal *JA;* (Segment 29) überlappt wird. Letztendlich wird in Segment 28 sowohl durch die Lexiko-Semantik als auch durch die prosodische Form (schnell, in mittlerem Tonhöhenregister und mit gleichbleibender Tonhöhenbewegung) eine Gedankenwiedergabe projiziert, die in Segment 30 eingelöst wird: *oh gOtt <<creaky, dim> ↓die A:Rme hEi:`ke.>.*

Evidenz dafür, dass die Darstellung der affektiven Haltung in Segment 30 als 'Mitgefühl' oder 'Mitleid' bezeichnet werden kann, liefert die Teilnehmerin selbst im darauffolgenden Segment, indem sie explizit macht: *<<soft, hoch> ich hAb so <<smile voice> mIt dir ge↑`-FÜH:LT->>* (Segment 31).

Bereits in Segment 30 selbst wird durch den Einsatz verschiedener semiotischer Ressourcen Mitgefühl kontextualisiert. Der lexiko-semantische Gehalt der Äußerung legt augenscheinlich eine Darstellung von Mitgefühl/ Mitleid nahe. In der diachronen germanistischen Sprachwissenschaft gilt *Oh Gott* als syntagmatische, sekundäre Interjektion, bestehend aus primärinterjektionalem Material (*oh*) und lexikalischem Material (*Gott*), die dem Ausdruck emotionaler Involviertheit

[7] Da Hanna und Birte zusammen wohnen, kann davon ausgegangen werden, dass Hanna darüber informiert ist, dass Birte am Donnerstagabend bei Heike zu Besuch war. Der Verlauf des Abends bzw. des darauffolgenden Morgens scheint ihr jedoch unbekannt zu sein.

dient. Ihre Funktion liegt laut Nübling nicht in der Referenz auf den benannten Gegenstand, sondern im Ausdruck einer emotiven Haltung (2001: 32f.). Diese ist jedoch nicht unbedingt spezifisch für eine bestimmte Affektart. So wäre es vorstellbar, dass *Oh Gott* auch zum Ausdruck von Überraschung o.ä. verwendet wird. Die Verwendung von [*arme* + X] ist jedoch eine routinisierte Form für die Darstellung von Mitleid. Schwitalla beschreibt die 'Anredeform' *du arme* mit fallender Intonation als standardisiertes Mittel zum Ausdruck für Mitleid (Schwitalla 1995: 230ff.; vgl. für das Englische Reber 2009, 2012). Im vorliegenden Fall (Segment 30) handelt es sich jedoch nicht um eine Anrede, sondern eher um ein Epitheton, ein Attribut, das hier einerseits der Charakterisierung von Heike in der Erzählwelt und andererseits der Bewertung ihrer Situation *in-situ* dient.

Auch auf kinetischer Ebene wird Betroffenheit nahe gelegt. Hier ist zu beobachten, dass Birte Heike im Verlauf der Äußerung durchweg ansieht und zum Ende der Äußerung die Hand horizontal vor Kinn und Mund hält.

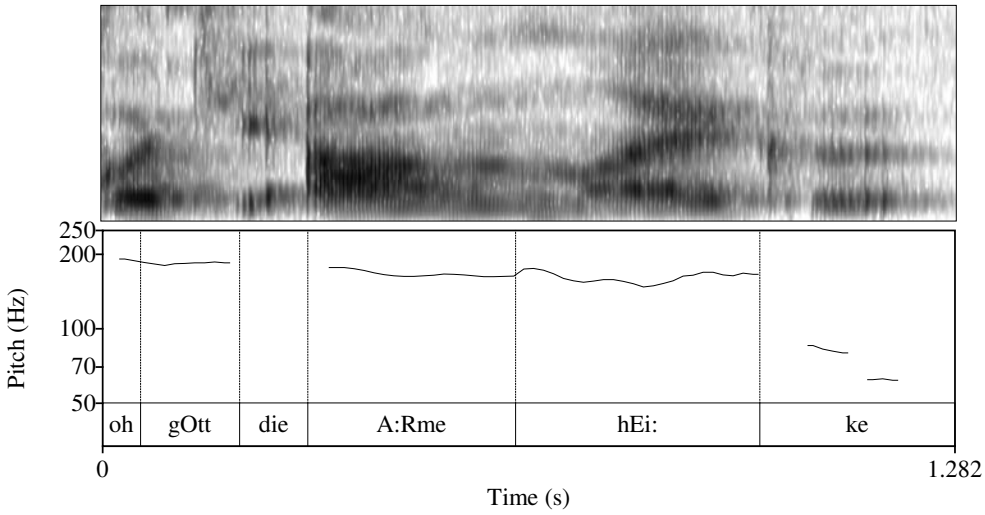


Abb. 1: Spektrogramm (0-5000 Hz) und Grundfrequenz (50-250 Hz) von LoE_VG_01_Kater_Segment 30

Die lexiko-semantischen und kinetischen Verfahren sind also bereits Indikatoren für den Ausdruck von Mitgefühl. Auf prosodisch-phonetischer Ebene sind nun folgende Parameter zu finden: leicht abnehmende Lautstärke, abnehmende Tonhöhe und am Ende eine tief fallende, flach auslaufende Kontur. Eine Messung der Grundfrequenz[8] soll den Höreindruck bestätigen (vgl. Abb. 1).

Die Unterbrechung bzw. der vermeintliche Tonhöhensprung in der Grundfrequenzmessung wird durch die starke Glottalisierung der dritten und der letzten Silbe erklärbar. Knarrstimme am Ende des Segments macht den Höreindruck einer tiefen finalen Tonhöhenbewegung aus. Das prosodisch-phonetische Merkmalsbündel in Segment 30 steht im starken Kontrast zur prosodischen Form von Segment 28, das, wie oben beschrieben, der Einleitung der Gedankenwiedergabe

[8] Da die Relation zwischen Tonhöhe und Grundfrequenz nicht linear ist, werden Abbildungen der Veränderung der Grundfrequenz f0 immer 'logarithmisch' (in Hertz) abgebildet, um dem Höreindruck gerecht zu werden (vgl. z.B. Ogden 2009: 44f.).

dient und schnell, in mittlerem Tonhöhenregister und mit gleichbleibender Tonhöhenbewegung realisiert ist.

Die Teilnehmerin Heike wird also in der Erzählwelt als 'bemitleidenswert' kontextualisiert (Segment 30). Birtes emotionale Haltung gegenüber Heike wird durch das Zusammenspiel lexiko-semantischer, kinetischer und prosodisch-phonetischer Ressourcen zum Ausdruck gebracht und als 'Mitfühlen' kategorisiert (Segment 31).[9]

Auf lexiko-semantischer Ebene ist die Äußerung *<<soft, hoch> ich hAb so <<smile voice> mIt dir ge↑ `¯FÜH:LT->>* eine Beschreibung einer vergangenen emotionalen Haltung, im Sinne einer Thematisierung von Erleben und Emotionen (Fiehler 1990). Hier verschwimmt jedoch die Trennung zwischen Affektdarstellung in der Erzählwelt und Affektdarstellung *in-situ*, ähnlich wie bei

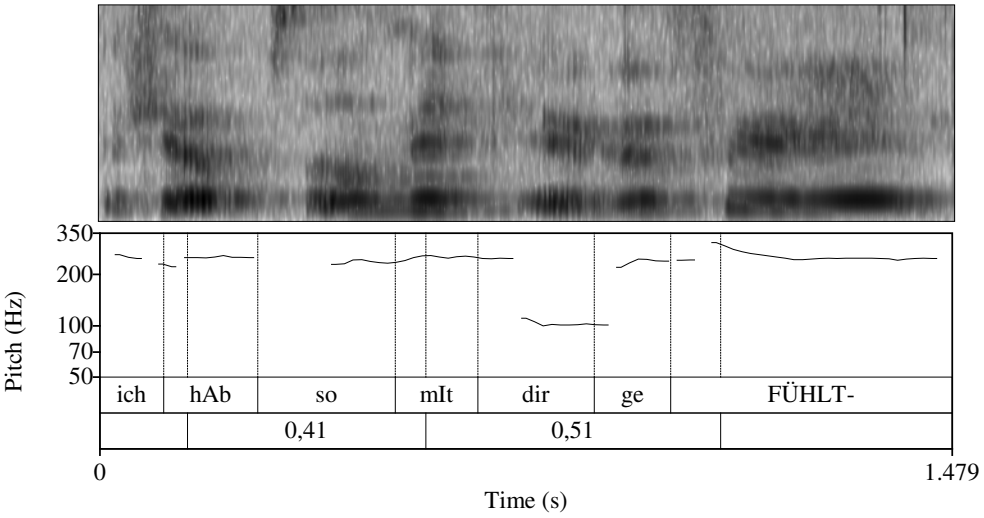


Abb. 2: Spektrogramm (0-5000 Hz), Grundfrequenz (50-350 Hz) und Rhythmus von LoE_VG_01_Kater_Segment 31

'sich überlagernden Stimmen' (z.B. Christmann/ Günthner 1996, Günthner 1997): Es überlagern sich Birtes Problemorientierung und Mitgefühl aus der Erzählwelt und Birtes Rahmung des Erzählten als 'amüsant' im *hic-et-nunc*: Affektivität mit positiver Valenz wird durch das erhöhte Tonhöhenregister und die zunehmende Lächelstimme nahe gelegt (vgl. für das Englische Maynard/ Freese 2012). Gleichzeitig sind ähnliche prosodisch-phonetische Verfahren zu beobachten wie in Segment 30, das Mitgefühl interpretierbar gemacht hat: flach auslaufende Kontur und Knarrstimme auf *dir*, die in Praat als Tonhöhensprung dargestellt wird. Dazu kommen hier Silbendehnung und weiche Artikulation/ geringe Sprechspannung. Auch eine rhythmische Realisierung ist hörbar (vgl. Abb. 2).

Die erste Zeile im Praat-Textgrid enthält den Wortlaut, die zweite Zeile die Messung des Zeitintervalls zwischen den Schlägen; gemessen wurde hier jeweils ab Vokalanfang der betonten Silbe. Die Abweichung der Längen der Füße beträgt

[9] Hier wird auch deutlich, dass eine starke Abgrenzung der Kategorien 'Mitgefühl' und 'Mitleid' aus Teilnehmersicht unnötig ist. Während [*arme* + X]-Konstruktionen alltagsweltlich als Darstellung von Mitleid angesehen werden, wird in diesem Beispiel im Folgeturn genau diese rekonstruierte Verwendungsweise als 'Mitgefühl' bezeichnet.

20% und liegt somit im Toleranzbereich des als isochron Wahrgenommenen.[10] Die rhythmische Realisierung wird im Folgenden noch einmal in Kadenzschreibweise dargestellt. Diese Schreibweise verdeutlicht im Zusammenspiel mit der Angabe der Zeitmessung, dass die Füße isochron, d.h. mit annähernd gleicher Dauer, realisiert werden (vgl. u.a. Auer/ Couper-Kuhlen 1994, Selting et al. 2009). Ein Fuß umfasst jeweils eine akzentuierte Silbe und die ihr folgenden unakzentuierten Silben bis zur nächsten akzentuierten Silbe, diese ausgeschlossen. Gemessen wird jeweils ab Vokalanfang der akzentuierten Silbe:

```
Bir:   ich
       /hAb so       /0.41s
       /mIt dir ge   /0.51s
       /↑`¯FÜH:LT-
```

Die Rhythmisierung klingt durch die geringen Tonhöhenbewegungen und die geringe Lautstärkeveränderung auf den akzentuierten Silben 'weich' und 'fließend' und kann aus diesem Grund als Legato-Rhythmus bezeichnet werden (Stock 1996: 69). Sie geht mit geringer Sprechspannung einher und ist wohl im Zusammenspiel mit dem 'gleitenden' Tonhöhenverlauf maßgeblich für den Höreindruck der Äußerung als 'affektorientiert' und 'mitfühlend' verantwortlich (vgl. hierzu die Parameter des *motherese* bei Bose 2003: 250ff.).

Zusammenfassend ist festzuhalten, dass in Segment 30 bereits auf Grund des lexiko-semantischen Gehalts (idiomatische Wendung [*arme* + X]) eine Darstellung von Mitleid – in der Erzählwelt – nahe gelegt wird. In Segment 31 wird diese Darstellung verbal explizit als 'Mitfühlen' kategorisiert. Hier wird ein Übergang von der Erzählwelt zum Hier und Jetzt geschaffen, da sich Thematisierung einer vergangenen affektiven Haltung (Fiehler 1990) gegenüber der anwesenden *consequential figure* und *in-situ* Affektdarstellung überlagern. Eine Reaktion der *consequential figure* wird somit unmittelbar relevant gemacht: Heike bestätigt diese doppelte Orientierung, einerseits an der Problemhaftigkeit der vergangenen Situation, andererseits an dem Unterhaltungswert im Moment, indem sie fortwährend Birte ansieht, leicht nickt, leicht lächelt (Segment 32) und erzählt, wie der von Birte angesprochene Vormittag für sie weiter verlief (Segmente 33ff.).

In Beispiel (1) sind im Zusammenspiel mit einer expliziten Mitgefühlsdarstellung die folgenden prosodisch-phonetischen Parameter als potentiell relevant für 'mitfühlendes Sprechen' zu beobachten: abnehmende Lautstärke, tiefes Tonhöhenregister, final fallende/ flach auslaufende Kontur, Silbendehnung, Knarrstimme, Legato-Rhythmus, geringe Sprechspannung/ weiche Artikulation.

In Beispiel (2) wird eine explizite Mitleidsdarstellung in einem veränderten *participation framework* (vgl. Goffman 1981) betrachtet: Hier ist die *consequential figure* nicht anwesend. Auch hier sind ähnliche prosodisch-phonetische Merkmalsbündel zu finden. Der Ausschnitt stammt aus demselben Gespräch wie Beispiel (1). Hanna erzählt von ihrer Bekannten, die ihr kürzlich gesagt hat, dass sie einen Tumor im Rücken hat. Vor Beginn des gezeigten Ausschnitts hat Hanna

[10] "Abweichungen in der Dauer der Intervalle bis zu ungefähr 20%, bei vorhandener Intonationsgrenze bzw. Sprecherwechsel bis zu ungefähr 30% [werden] in der Regel nicht als nachteilig für perzeptuelle Isochronie empfunden" (Auer/ Couper-Kuhlen 1994: 91f.).

bereits ausführlich dargestellt, dass die Bekannte schon seit längerem über Schmerzen klagt, ihr jedoch nicht geglaubt wurde und sie nun die Bestätigung einer tatsächlichen Erkrankung erhalten hat. Sie schränkt ein, dass es ein gutartiger Tumor ist, der dennoch operiert werden muss. Bisher hat sie auffällig wenige Rezeptionssignale von Heike und Birte erhalten und führt nun aus:

(2) LoE_VG_01_Tumor im Rücken (52:34)

```
   40   Han:   =aber die operatiOn is halt AUCH nich ohne,
   41           [weil da kann] AUCH viel schIef gehen;
   42   Hei:   |[<<t> hmhm,> ]
               |((leicht nickend))
   43   Han:   |((sieht Hei an; verzieht kurz den Mund))
   44   Hei:   |GLAUB [ich;         ]
-> 45   Bir:          [<<p> o:::h.> ]
   46   Hei:   [grAde so im RÜCKen, ]
   47   Han:   [deswEgen is dAs nich]|GUT-=
                                     |((Bir ansehend))
-> 48   Han:   =dA TA:T |se |mir <<behaucht> Auch>
                         |((ins Leere sehend))
                 voll ˍlEid.=
   49   Hei:                 |((zieht die Mundwinkel nach unten))
   50   Han:   =[also-   ]
-> 51   Bir:   =[JA naTÜR]lich-
-> 52           <<dim> das [is ja KLA:R.>  ]
   53   Han:               [(und vor Allem)]hat sie jetzt
               ja dAuernd SCHMERzen;
```

Nachdem Hanna betont, dass ein operativer Eingriff am Rücken problematisch ist (Segment 40), erhält sie in Überlappung mit ihrer Begründung (Segment 41) eine Rezipientenreaktion von Heike, durch die die Anerkennung der Problemhaftigkeit des Geschilderten deutlich wird. Anschließend verzieht Hanna kurz das Gesicht, indem sie die Mundwinkel nach hinten zieht (Segment 43), was als Zeichen einer 'heiklen Angelegenheit' interpretiert werden kann. Heike macht in Überlappung hiermit ihre Anerkennung des Problems explizit: *GLAUB ich;* (Segment 44). In Überlappung mit der letzten Silbe von Segment 44 äußert Birte ein leises *<<p> o:::h.>* (Segment 45). Dabei sieht sie die Erzählerin an, diese hat jedoch ihren Blick auf Heike gerichtet. Wenn auch leise realisiert, ähnelt das Lautobjekt sehr dem Englischen [ɑː], das zur Darstellung von Mitgefühl/ Mitleid in Problemgesprächen verwendet werden kann (Reber 2009). Hier sind folgende prosodisch-phonetischen Parameter zu beobachten: Silbendehnung und tief fallende, flach auslaufende Kontur, Lippenrundung (hier im Zusammenhang mit der Vokalqualität). Auf Grund der sequentiellen Position, im Anschluss an Hannas dramatisierte Problemerläuterung (Segmente 40f.), kann das Lautobjekt als affektive Aufnahme des dargestellten Problems interpretiert werden. Die andere Rezipientin der Problemerzählung, Heike, schließt ihrer Verstehensbehauptung in Segment 44 eine

Verstehensbekundung[11] an: *grAde so im RÜCKen,* (Segment 46). In Überlappung damit bewertet Hanna die Situation der nicht-anwesenden Person (Segment 47) und realisiert die Äußerung mit gleichbleibender Tonhöhenbewegung, um unmittelbar anzuschließen: *=dA TA:T se mir <<behaucht> Auch> `voll ˌlEid.=* (Segment 48).

Auf lexiko-semantischer Ebene macht Hanna in diesem Segment anhand der 'leidtun'-Konstruktion ihre affektive Haltung, die alltagssprachlich als 'Mitleid' oder 'Mitgefühl' beschrieben werden kann, explizit. Auch hier sind die prosodisch-phonetischen Parameter aus Beispiel (1) zu beobachten: geringe

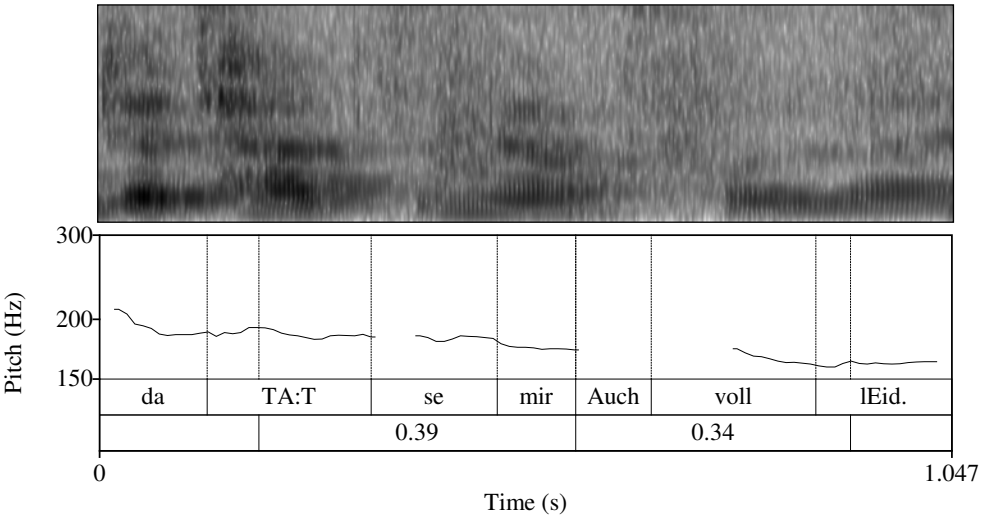


Abb. 3: Spektrogramm (0-5000 Hz), Grundfrequenz (150-300 Hz) und Rhythmus von LoE_VG_01_Tumor_Segment 48

Sprechspannung, Vokaldehnung, teilweise Behauchung und eine tief fallende, flach auslaufende Kontur. Letztere ist auch anhand der Messung des Tonhöhenverlaufs verifizierbar (vgl. Abb. 3). Zudem kann, ähnlich wie in Beispiel (1), auch hier von einer Legato-Rhythmisierung gesprochen werden, d.h. die akzentuierten Silben sind nicht sehr viel prominenter als die unakzentuierten Silben, dennoch ist die Äußerung hörbar rhythmisch produziert. Dies belegt die Messung der Intervalllänge zwischen den Schlägen:

```
Han:   =dA
       /TA:T se mir /      0.39s
       /Auch `voll  /      0.34s
       /ˌlEId.=
```

[11] Sacks unterscheidet zwischen dem Behaupten von Verstehen (*claiming understanding*) und dem Bekunden von Verstehen (*exhibiting understanding*) (1995: 252ff.). Ersteres kann durch Aussagen wie 'Ich verstehe, was du meinst' geschehen, letzteres impliziert, dass eine Anwendung des Gesagten vollzogen wurde, sodass das Verstehen 'bewiesen' werden kann, z.B. indem aus der Aussage 'Ich bleibe bis Montag' die Schlussfolgerung 'Ah, noch eine Woche' gezogen wird.

Ähnlich wie in Beispiel (1) ist die Äußerung nicht wie 'neutrales', 'distanziertes' Sprechen über vergangene Erfahrungen gerahmt, sondern macht *in situ* eine erhöhte emotionale Involviertheit der Teilnehmerin Hanna interpretierbar. Trotz der Präteritum-Konstruktion legt die Verwendung spezifischer prosodisch-phonetischer und mimischer Ressourcen nahe, dass die Grenzen zwischen Thematisierung eines vergangenen Affekts und einer *in situ* Affektdarstellung verschwimmen.

Birte nimmt Hannas Mitleidsdarstellung durch mehrmaliges Nicken auf und behandelt diese als selbstverständlich: *=JA naTÜRlich-* (Segment 51). Koordiniert mit der zweiten akzentuierten Silbe des Segments legt Birte ihren Kopf schräg nach hinten. Neben den semantischen tragen also auch die kinetischen Ressourcen zum Ausdruck von Verständnis bei, ebenso die folgende Äußerung *<<dim> das is ja KLA:R.>* (Segment 52), die noch weniger stark artikuliert, legato, diminuendo, mit finaler Dehnung und final tief fallend realisiert ist. Durch die Bewertung ('natürlich', 'klar') behandelt Birte Hannas Mitleidsdarstellung als legitim und erwidert gleichzeitig durch eine ähnliche prosodische Rahmung die Affektorientierung der Vorgängeräußerung. In den Segmenten 53ff. führt Hanna weitere Aspekte der problematischen Lage ihrer Freundin an. Diese Segmente sind im Vergleich zu Segment 48 in mittlerem Tonhöhenregister, ohne Veränderung der Stimmqualität, ohne hörbaren Rhythmus produziert, also eine Rückkehr zur 'Normallage'. Dies ist ein weiteres Indiz für die Relevanz des prosodisch-phonetischen Merkmalsbündels in Segment 48, das auf die starke Affektorientierung der Teilnehmerin schließen lässt.

In diesem Abschnitt wurden Beispiele analysiert, in denen die dargestellten affektiven Haltungen der Interaktionsteilnehmerinnen von ihnen selbst auf lexikosemantischer Ebene als 'Mitgefühl' bzw. 'Mitleid' explizit gemacht wurden und die gleichzeitig als 'mitfühlendes Sprechen' hörbar waren. Es galt, von diesen Beispielen ausgehend, die prosodisch-phonetischen Parameter zu ermitteln, die für den Höreindruck 'mitfühlenden Sprechens' potentiell relevant sind. Die Analyse dieser expliziten Fälle, in denen Mitgefühl bzw. Mitleid als Teilnehmerkategorie relevant gemacht wird, sollte einer voreiligen Zuschreibung eines prosodisch-phonetischen Merkmalsbündels als 'mitfühlend' entgegenwirken (Local/ Walker 2008).

4.2. Prosodische Verfahren in einer Pseudodarstellung

Weitere Evidenz für die Relevanz des Zusammenspiels spezifischer prosodisch-phonetischer Ressourcen für den Höreindruck 'mitfühlenden' Sprechens soll das folgende Beispiel einer 'Pseudodarstellung' liefern. Reber (2009) beschreibt in ihrer Untersuchung englischer Lautobjekte Pseudodarstellungen als spielerische Inszenierungen und macht deutlich, dass diese aus Analytikerperspektive in zweierlei Hinsicht aufschlussreich sind:

> Die spielerische Parodie des Lautobjekts und damit des so dargestellten Affekttyps liefert Evidenz dafür, dass die Darstellung von Affektivität mit dieser sprachlich-formalen Ressource […] eine Teilnehmerkategorie und nicht ein analytisches Konstrukt bildet. Außerdem zeigt die Nachahmung, dass die Sprecherin über ein kom-

> munikatives Metawissen sowohl bezüglich der Form als auch der Funktion des Lautobjekts verfügen muss, das es ihr erlaubt, das Lautobjekt gezielt einzusetzen.
> (Reber 2009: 210)

Ähnlich wie in Rebers Untersuchung geht es in Beispiel (3) darum, (a) anhand einer Pseudodarstellung Rückschlüsse auf das Wissen der Teilnehmerinnen über eine mögliche (holistische) Gestalt von Mitgefühls-/ Mitleidsdarstellungen zu ziehen und (b) anhand dieser Gestalt weitere Evidenz für die potentiell relevanten prosodisch-phonetischen Verfahren für 'mitfühlendes Sprechen' zu gewinnen.

Die Schwestern Sandra (San) und Emma (Emm) sitzen gemeinsam beim Brunch. Emma erzählt, wie sie in Abwesenheit ihres Mannes ihren Kindern Pancakes, eine Art Eierkuchen, zubereitet hat, die sie dann selbst nicht essen konnte, weil sie schnell das Haus verlassen musste:

(3) LoE_VG_04_pancakes (00:42)

```
   73   Emm:   =<<all> |und da hab ich die PANcakes für die
                       |((San ansehend))
               kInder gemacht;>=
   74          =<<all> ↓dann kam HARry wieder musst ich
               ganz schnell lOs weil ich ja spät drAn war;>
   75   San:   [hm,]
   76   Emm:   [°h  ]<<dim, rall> |und dEswegen hatte ich diese
                                  |((Gesicht verziehend))
               |PANcakes gemacht->=
               |((auf offene Hand blickend))
-> 77          |=<<dim, creaky> kOnnt se aber gar nich ESSen;>
               |((offene Hand schnell anhebend und fallen lassend))
-> 78   Emm:   (0.3)       ((verzieht die Mundwinkel nach unten))
-> 79   San:   |<<creaky> `_[ɒ:::].>
               |((die Augenbrauen zusammenziehend))
   80          |<<creaky> m:[:.>                    ]
               |((Blickkontakt lösend))
   81   Emm:   |((wendet Blick
                 zum Brötchen))
   82   Emm:                [<<pp> ja; |(´gen`AU;)>]
                                       |((Brötchen aufschneidend))
   83   San:   °h na wir hAm ja wir ham (.) !Ü!belst spät
               gefrÜhstückt heute,
```

Nach einer kurzen chronologischen Beschreibung der Begebenheiten (Segmente 73f.) fasst Emma die Pointe ihrer kurzen Erzählung zusammen: *°h <<dim, rall> und dEswegen hatte ich diese PANcakes gemacht->=* (Segment 76), *=<<dim, creaky> kOnnt se aber gar nich ESSen;>* (Segment 77). Neben dem lexikosemantischen Gehalt der Äußerung wird auch durch Mimik ('Gesicht verziehen' in Segment 76) und Gestik (verstärkende Handgeste in Segment 77) sowie durch die Veränderung der Stimmqualität Emmas Haltung hinsichtlich des Erlebten deut-

lich: Sie kontextualisiert diesen Zustand als bedauerlich bzw. sich als 'bemitleidenswert'. Höhepunkt dieser Darstellung ist Emmas Mimik in Segment 78:

Abb. 4: Mimik LoE_VG_04_pancakes_Segment 78

Der verwendete Gesichtsausdruck lässt Emmas Darstellung übertrieben wirken und weist auf eine spielerische Inszenierung ihrer 'bemitleidenswerten' Situation hin. Sandra reagiert hierauf mit dem Lautobjekt <<*creaky*> `*[ɒ:::]*.> (Segment 79)[12], das sie mit zusammengezogenen Augenbrauen realisiert:

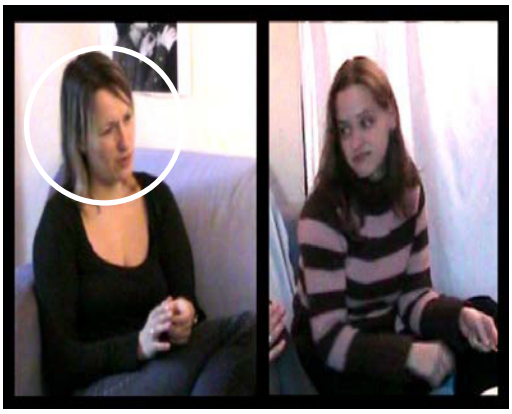


Abb. 5: Mimik LoE_VG_04_pancakes_Mitte Segment 79

Die Verwendung dieses Lautobjekts mag in Koordination mit den mimischen Mitteln (zusammengezogene Augenbrauen) und in der sequentiellen Position (nachdem sich Emma als bemitleidenswert dargestellt hat) auf den ersten Blick eine prototypische Darstellung von Mitleid/ Mitgefühl nahe legen (vgl. Golato 2012: 259f.). Auch ein Vergleich mit Form und Funktion des Englischen Lautobjekts 'ah' (Reber 2012) stützt diese Interpretation. Eine detaillierte sequentielle Analyse lässt jedoch erkennen, dass es sich in diesem Beispiel um eine Pseudodarstellung (vgl. Reber 2009), also eine spielerische Inszenierung von Mitleid/ Mitgefühl, handelt.

[12] Es wurde hier das IPA-Zeichen des gerundeten offenen Hinterzungenvokals gewählt, da es dem Höreindruck nach Sandras Äußerung am ehesten entspricht. Zum Nachhören: Ladefoged (o.A.): http://www.phonetics.ucla.edu/course/chapter1/vowels.html, Zugriff am 24.07.2013.

Ein erstes Indiz für eine solche Interpretation ist die Tatsache, dass Emma ihre Erzählung im Vorfeld nicht als problematisch einführt (Sacks 1971: 310f., Mandelbaum 2013: 495ff.). Auch explizit negative Bewertungen, wie sie in einer tatsächlichen Problemerzählung erwartbar wären, fehlen (für Bewertungen in Beschwerdegeschichten siehe z.B. Selting 2010a, 2012). Ein weiteres Indiz für eine Pseudodarstellung ist der sequentielle Verlauf nach dem Lautobjekt (Segmente 80ff.). Auffällig ist, dass beide Gesprächsteilnehmerinnen unmittelbar im Anschluss an Pointe und Mitleidsdarstellung andere Aktivitäten beginnen. Sandra leitet durch den geknarrten Nasal *m::.* (Segment 80) und das Abschwenken ihres Blicks den Übergang zu einer anderen Aktivität ein. Emma äußert in Überlappung mit Sandras Knarrlaut sehr leise *<<pp> ja; ´gen`AU;>* (Segment 82) und bewertet somit Sandras vorherige Reaktion als folgerichtig und angemessen und ihre Erfahrung als nicht weiter behandlungswürdig. Sandra erzählt anschließend von ihrem eigenen Frühstück (Segmente 83ff.). Zweite Erzählungen (*second stories*) sind eine typische Alltagspraktik, um Verstehen und Zustimmung zu einem Aspekt einer vorherigen (ersten) Erzählung darzustellen (vgl. u.a. Sacks 1995, Arminen 2004, Selting 2012). Sandra macht genau das: Sie nimmt zwar das Thema 'Frühstück' aus Emmas Erzählung auf, schildert ihr eigenes Frühstück jedoch als erfreuliches Ereignis und geht somit nicht auf die – vermeintlich negative – Valenz der vorherigen Erzählung von Emma ein. Auch Emma selbst wendet sich dem Aufschneiden eines Brötchens zu, fokussiert also eine neue Aktivität. Demnach finden weder eine Dramatisierung von Emmas 'negativem' Erleben noch eine explizite Problembehandlung statt, wie es in anderen Beispielen der Fall ist (vgl. z.B. Beispiel (4)). Das Fehlen einer weiteren Thematisierung des Problems und explizit negativer Bewertungen im unmittelbaren sequentiellen Umfeld lassen also darauf schließen, dass es sich bei Sandras *<<creaky> `_[ɒ:::].>* in Segment 79 um eine Pseudodarstellung von Mitleid handelt: Durch diese spielerische Inszenierung von Affektivität zeigt sie Problemresistenz (Reber 2009: 207f.). Sandra spiegelt Emmas spielerisch-übertriebene Probleminszenierung zwar wider, zeigt jedoch durch den direkten Übergang zur nächsten Erzählung, dass es sich offensichtlich um ein Problem handelt, das nur spielerisch zu behandeln ist und keiner weiteren ernsthaften Thematisierung bedarf. Emmas kurze Aufnahme (Segment 82) bestätigt dies. Sie ratifiziert Sandras Reaktion als angemessen und macht keine weitere Problembehandlung relevant.

Eine Sequenzanalyse ergibt also, dass es sich in Segment 79 um eine Pseudodarstellung von Mitgefühl handelt. Zu dieser spielerischen Inszenierung von Mitgefühl werden auch wieder einige der bereits bekannten prosodisch-phonetischen Ressourcen verwendet: Vokaldehnung, Knarrstimme und eine tief fallende, flach auslaufende Tonhöhenbewegung am Einheitenende (vgl. Abb. 6). Festzuhalten ist zudem, dass die Äußerung hinsichtlich der Stimmqualität (Knarrstimme) und hinsichtlich des Rhythmus eine prosodische Orientierung an der Vorgängeräußerung aufweist. Letzteres bestätigt die Intervallmessung der Füße:

```
Emm:   /kOnnt se aber gar nich  /      0.48s
       /ESSen; (0.3)            /      0.49s
San:   /`_[ɒ:::].
```

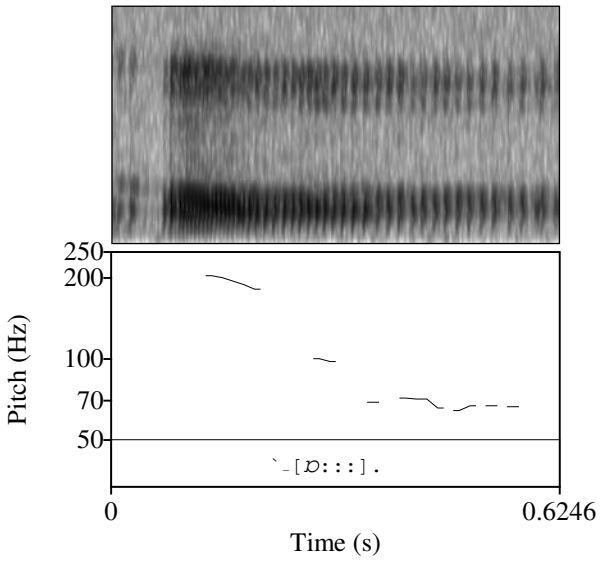


Abb. 6: Spektogramm (0-5000 Hz) und Grundfrequenz (50-250 Hz) von LoE_VG_04_pancakes_Segment 79

Die Bedeutung der prosodischen Orientierung an der Vorgängeräußerung, vor allem des Timings, für die Interpretation von responsiven Handlungen als affiliativ wird bereits von Couper-Kuhlen (2012a) deutlich hervorgehoben. Die prosodische Orientierung an der Vorgängeräußerung hinsichtlich Stimmqualität und Rhythmus kann ebenfalls als potentieller Bestandteil des prosodischen Merkmalsbündels für 'mitfühlendes Sprechen' festgehalten werden.

Beispiel (3) verdeutlichte also zwei Aspekte: GesprächsteilnehmerInnen können anhand bestimmter Ressourcen Affekte darstellen und ebenso für alle Beteiligten interpretierbare Pseudodarstellungen inszenieren. Diese Fähigkeit weist auf ihr kommunikatives Metawissen über den Einsatz spezifischer (prosodisch-phonetischer) Ressourcen zur Affektdarstellung in verschiedenen Interaktionsmodalitäten, z.B. ernsthaft oder spielerisch, hin (Reber 2009: 210ff.). Dass das prosodisch-phonetische Merkmalsbündel für 'mitfühlendes Sprechen' auch in dieser spielerischen Pseudodarstellung verwendet wurde, ist also ein weiteres Indiz für die generelle Relevanz dieses Merkmalsbündels für die Interpretation 'mitfühlenden Sprechens'. Weiterhin wurde deutlich, dass das Lautobjekt [ɑ:::] eine vokale Ressource ist, deren Bedeutung nicht inhärent ist, sondern die kontextsensitiv zu interpretieren ist. In einer bestimmten prosodisch-phonetischen Gestalt, d.h. mit Vokaldehnung, mit Knarrstimme, mit tief fallender, flach auslaufender Kontur und mit rhythmischer Integration, kann es, muss es aber nicht – je nach sequentieller Position – als Darstellung von Mitgefühl interpretiert werden.

4.3. Prosodische Verfahren in impliziten Empathiedarstellungen

Die Beispiele (1) bis (3) wiesen regelmäßig markante prosodisch-phonetische Merkmalsbündel auf, die als 'mitfühlendes Sprechen' interpretiert werden konnten, zum einen auf Grund ihres Zusammenspiels mit expliziter Lexik (Abschnitt

4.1.), zum anderen auf Grund ihrer offensichtlich spielerischen Verwendung als Pseudodarstellung in Reaktion auf eine affektgeladene Erzählung persönlichen Erlebens (Abschnitt 4.2.). Anhand der nächsten beiden Beispiele soll veranschaulicht werden, dass in Alltagsgesprächen diese prosodisch-phonetischen Merkmalsbündel auch in anderen Situationen verwendet werden, wo sie – ohne entsprechende Lexiko-Semantik, aber an spezifischen sequentiellen Positionen – 'mitfühlendes Sprechen' interpretierbar machen können. Zunächst ein Beispiel für ambige Lexiko-Semantik.

Sina (Sin) und Iso sind befreundet und frühstücken zusammen. Sina erzählt von ihrem Cousin, der als Jugendlicher bei einem Autounfall ums Leben kam, und vom Tod ihres Opas, der sich teilweise die Schuld an dem Autounfall gab und genau ein Jahr nach dem Tod des Cousins verstarb:

(4) AGmT_01_Verluste (11:23)

```
-> 58   Sin:   |und mein opa s_fAst (.) |geNAU auf_n
               | ((löst Blickkontakt))
   59   Iso:                            | ((Blick zum Tisch))
        Sin:   tAg n jahr <<dim> spÄter
               |gestOrben;>
   60   Iso:   | ((beginnt, Saftglas abzustellen, runzelt die Stirn))
   61          (0.65) ((Iso stellt Glas ab, Blick zum Tisch))
-> 62   Iso:   |<<p, behaucht> ↑ACH du |schEi:ße;>
               | ((Blick nach unten))          | ((zieht Augenbrauen mit
                                                  den Fingern nach))
-> 63   Sin:   |<<p> ?hmhm,>
               | ((nickend, ins Leere blickend, Zunge unter Unterlippe schiebend))
   64          (1.2) ((Iso sieht nach unten, zieht Augenbrauen mit den
                      Fingern nach))
-> 65   Sin:   |aber dAs fand ich SCHLIMmer;
               | ((ins Leere blickend))
   66   Iso:   | ((reorientiert sich zu Iso))
   67   Sin:   ich weiß nich waRUM;=
   68          =aber an meinem Opa hing ich unglaublich DOLle;=
```

Sinas Äußerung über den Tod ihres Großvaters ist ohne Blickkontakt und zum Ende der Einheit diminuendo und im tiefen Tonhöhenregister realisiert. Obwohl an dieser Stelle noch keine explizite Bewertung der Erzählerin hinsichtlich der Umstände erfolgt ist, wird die Äußerung durch das Zusammenspiel von Kinesik, Prosodie und Lexiko-Semantik als hochproblematische Pointe interpretierbar, die eine affektorientierte Aufnahme erwartbar macht. Dies wird auch in der folgenden mittellangen Pause deutlich: Die Erzählerin hat weiterhin den Blick ins Leere gerichtet und führt ihre Darstellung nicht unmittelbar weiter aus. Der Interpretation von Isos Pause als lang, und damit als potenzielle Projektion einer dispräferierten Äußerung, kann entgegengesetzt werden, dass ihr Blick und ihr körperliches Verhalten auf das Absetzen des Saftglases auf den Tisch konzentriert sind. Das zeitgleiche Stirnrunzeln legt bereits nahe, dass Iso die Tragweite und Implikationen

von Sinas Äußerung erfasst hat. Dies wird in folgender Reaktion noch deutlicher: *<<p, behaucht> ↑ACH du schEi:ße;>* (Segment 62).

Dass Sina Isos Reaktion als affektorientiert interpretiert, wird im nächsten Segment deutlich: *<<p> ʔhmhm,>* (Segment 63). Durch Nicken, den ins Leere gerichteten Blick und die unter die Unterlippe geschobene Zunge legt sie ihre Betroffenheit nahe. In den Segmenten 65 bis 68 macht Sina dann ihre emotionale Involviertheit explizit, indem sie ihr persönliches Erleben beim Ableben ihres Großvaters thematisch weiter ausführt. Daran wird deutlich, dass sie Isos Reaktion nicht nur als Erkenntnisprozessmarker + Bewertung, sondern durchaus als affektorientiert und 'mitfühlend' interpretiert hat und diese Affektorientierung auch als 'angemessen' ratifiziert.

Ähnlich wie in den Beispielen (1) bis (3) sind folgende prosodisch-phonetischen Parameter in Segment 62 zu beobachten: geringe Lautstärke, Vokaldehnung auf der vorletzten Silbe, Behauchung. Im Vergleich zur Vorgängeräußerung, die im tiefen Tonhöhenregister endet, entsteht der Eindruck eines kleinen Tonhöhensprungs nach oben auf der ersten Silbe. Eine rhythmische Anpassung an die Vorgängeräußerung ist hör- und auch messbar:

```
Sin:   ge
       /NAU auf_n     /  0.33 sek
       /tAg n jahr    /  0.39 sek
       /spÄter ge     /  0.36 sek
       /stOrben;      /  0.93 sek[13] =   0.31 sek
       /(   )         /                   0.31 sek[14]
       /(   )         /                   0.31 sek[15]
Iso:   /↑ACH du      /  0.42 sek
       /schEi:ße;
```

Zum Ende der Äußerung von Sina ist ein isochroner Rhythmus etabliert, d.h. zwischen den akzentuierten Silben liegen sehr ähnliche Zeitintervalle (zwischen 0.33 und 0.39 sek). Das Intervall zwischen der letzten akzentuierten Silbe von Sinas Äußerung und der ersten akzentuierten Silbe von Isos Äußerung ist vergleichsweise lang und beträgt 0.93 sek. Dennoch ist die Äußerung auditiv rhythmisch integriert, wenn man *silent beats* annimmt. Auer/ Couper-Kuhlen beschreiben das Phänomen des 'leeren Schlags' wie folgt: "[E]in rhythmischer Schlag [kann] in einer schon etablierten Gestalt gelegentlich leer bleiben (d.h., auf der dafür vorgesehenen Stelle findet kein Ereignis statt)" (1994: 87). Dem Höreindruck entsprechend, kann eine Messung der Intervalllänge eine rhythmische Integration von Segment 62 bestätigen: Das vergleichsweise lange Intervall von 0.93 sek überbrückt zwei 'leere Schläge' (die jeweils ein Intervall von ca. 0.31 sek haben), nach denen Iso ihre Äußerung 'passend' anschließt. Von 0.31 sek als kürzestem und 0.42 sek als längstem Intervall ausgehend, ergibt sich eine Abweichung der Intervalllängen von 26%, was bei Sprecherwechsel im Rahmen des isochron Wahrgenommenen liegt. Auch hier ist also, neben dem Vorkommen der anderen proso-

[13] Bis zum Vokalanfang der nächsten akzentuierten Silbe inklusive Pause von 0.65s.
[14] 1. silent beat
[15] 2. silent beat

disch-phonetischen Merkmale, das wichtige Kriterium des 'passenden Timings' für die Interpretation der Äußerung als affiliativ erfüllt (Couper-Kuhlen 2012a).

Im letzten Beispiel wird ein Merkmalsbündel aus den bisher beschriebenen prosodisch-phonetischen Ressourcen im Zusammenspiel mit lexiko-semantischen Ressourcen verwendet, die an sich kein Mitgefühl oder affektgeladene Bewertungen nahe legen. Der Ausschnitt stammt auch aus dem Gespräch zwischen Iso und Sina. Sina erzählt, dass sie in der Woche zuvor das Ende ihrer Lieblingsserie gesehen hat und 'unglaublich traurig' war, was sie jedoch nicht nur auf das Staffelende, sondern auch auf ihre anderen Erlebnisse der letzten Woche zurückführt:

(5) AGmT_02_Zeugnis (18:10)

```
    8   Sin:   <<dim> weil (glaub ich) Alle letzte wOche
               auch zu MIR <<creaky> gemein wAren;>>
    9          (--) ((Sin und Iso greifen jede zu einem Brötchen))
   10   Sin:   <<creaky, pp> [(so;)>]
-> 11   Iso:                [wie   ]SO ↓-de:nn:-
```

((Sina erzählt daraufhin ausführlich, dass sie ihr Master-Zeugnis nicht erhalten hat, weil ihr eine Seminarleistung nicht anerkannt wurde. Das stellt für Sina ein großes Problem dar, da das Zeugnis die Voraussetzung für den Beginn des bereits geplanten Referendariats ist.))

Sinas Äußerung in Segment 8 ist diminuendo, mit mittelfallender Tonhöhenbewegung, Rachenenge und Knarrstimme am Einheitenende realisiert. Die akzentuierten Silben fallen nicht durch hohe Tonhöhenbewegungen oder hohe Lautstärke auf, sind aber hörbar rhythmisch produziert. Dadurch wird der Fokusakzent auf 'mir' erklärbar. Die prosodische Gestalt der Äußerung legt 'Weinerlichkeit' nahe. Im Zusammenspiel mit den lexiko-semantischen und kinetischen Ressourcen (Verwendung des kindlichen Ausdrucks 'gemein sein', Körperkrümmung nach vorn, heruntergezogene Mundwinkel) stellt sich Sina als 'bemitleidenswert' dar. In der folgenden kurzen Pause findet keine Orientierung an dieser Äußerung statt, beide Interaktionsteilnehmerinnen greifen zu einem Brötchen, zeigen also ihre kurzzeitige Orientierung an der Parallelaktivität Frühstücken. Anschließend setzen beide Sprecherinnen in Überlappung ein: Sina mit einem kurzen, leisen, knarrenden *so;*, das eine stilisierte 'Trotzigkeit' interpretierbar und eine Behandlung ihrer bemitleidenswerten Situation relevant macht (Segment 10). In Überlappung damit fragt Iso nach: *wieSO ↓de:nn:-* (Segment 11) und nimmt die von Sina angeschlagene Interaktionsmodalität auf. Im Sinne eines 'Empathieangebots' (Pfänder/ Gülich 2013) schafft Iso durch ihre Frage einen interaktionalen Raum zur Behandlung des emotionalen Erlebens von Sina, der von der Betroffenen im weiteren Verlauf des Gesprächs auch als solcher genutzt wird.

Auch in diesem Segment sind die bereits bekannten prosodisch-phonetischen Parameter zu beobachten: Vokal- und Konsonantendehnung auf der letzten Silbe, flach auslaufende Kontur, verringerte Sprechspannung und leichte Lippenrundung (vgl. hierzu *motherese* in Bose 2003: 223ff., 250ff.). Zudem ist eine prosodische Orientierung an der Vorgängeräußerung auf der Ebene des Rhythmus zu beobachten:

```
Sin:   weil (glaub ich)
       /ALLe letzte   /  0.58 sek
       /wOche auch zu /  0.5 sek
       /MIR gemein    /  0.5 sek
       /wAren; (-)    /  0.41 sek
Sin:   /(-)[(so;)]    /  (0.41 sek)[16]
Iso:       [wie  ]
       /SO ↓-de:nn:-
```

Ähnlich wie in Beispiel (4) erscheint das Intervall zwischen der letzten akzentuierten Silbe der Sprecherin Sina und der ersten akzentuierten Silbe der Sprecherin Iso mit 0.82 sek vergleichsweise lang. Die Äußerung ist jedoch hörbar rhythmisch an die Vorgängeräußerung angepasst. Auch hier kann man das Phänomen des 'leeren Schlags' (Auer/ Couper-Kuhlen 1994: 87) zur Erklärung heranziehen und messtechnisch bestätigen: Durch zwei geteilt entspricht die Intervalllänge dem in der Vorgängeräußerung etablierten isochronen Rhythmus. Die Diskrepanz aller Intervalllängen, den 'leeren Schlag' eingeschlossen, beträgt maximal 29%, was im Toleranzbereich des als isochron Wahrgenommen liegt (ebd.: 82). Isos Äußerung ist also rhythmisch der Vorgängeräußerung von Sina angepasst, ihre Frage ist auf Sinas Problemdarstellung 'eingestimmt'.

Diese prosodisch-phonetischen Merkmale können hier also zugewandtes, 'mitfühlendes Sprechen' auch ohne Rückgriff auf einschlägige Lexiko-Syntax kontextualisieren. Das Aufzeigen des Anerkennens der emotionalen Situation der Betroffenen und der Bereitschaft zur weiteren Thematisierung wird von der Betroffenen selbst als Darstellung von Verständnis bzw. Mitgefühl interpretiert, was daraus geschlossen werden kann, dass sie auf das Angebot eingeht und ihr persönliches Erleben im Verlauf des Gesprächs weiter ausführt.

5. Zusammenfassung

Vor dem Hintergrund der gezeigten Beispiele sind zusammenfassend eine Reihe prosodisch-phonetischer Ressourcen festzuhalten, die potenziell relevant sind, um die Interpretation einer Äußerung als empathisch, hier i.S.v. 'mitfühlend', nahe zu legen: geringere bzw. abnehmende Lautstärke/ Diminuendo, tiefes Tonhöhenregister, flach auslaufende Kontur, rhythmische Integration/ Legato-Rhythmus, Behauchung, Knarrstimme, geringe Sprechspannung/ weiche Artikulation, Silbendehnung, Lippenrundung. Nicht alle Ressourcen müssen in jedem Fall gleichzeitig eingesetzt werden, es handelt sich jedoch immer um ein Merkmalsbündel aus mehreren dieser Ressourcen, das im Zusammenspiel mit anderen verbalen und/ oder kinetischen Ressourcen und auf Grund seiner sequentiellen Position als 'mitfühlend' interpretierbar wird. In Reaktionen auf unmittelbar zuvor erfolgte Affektdarstellungen der *consequential figure* konnte eine prosodische Orientierung an der Vorgängeräußerung vor allem auf der Ebene des (isochronen) Rhythmus festgestellt werden.

[16]silent beat

Die Beobachtungen decken sich weitestgehend mit den Beschreibungen prosodisch-phonetischer Merkmale zum Ausdruck 'mitfühlenden Sprechens' im Englischen und Finnischen, vor allem im Hinblick auf die Parameter Tonhöhenverlauf, Lautstärke, rhythmische Integration, Silbendehnung, Artikulation und Stimmqualitäten (vgl. u.a. Couper-Kuhlen 2009, 2012a, Hepburn/ Potter 2012, Reber 2012, Weiste/ Peräkylä 2014). Darüber hinaus wurden Legato-Rhythmus und Lippenrundung als weitere mögliche Parameter gefunden; Parameter, die auch im Zusammenhang mit *motherese* im Deutschen beschrieben werden (Bose 2003).

Einer interaktional-linguistischen Untersuchung liegt die Prämisse zu Grunde, dass Äußerungen von den InteraktionsteilnehmerInnen kontextsensitiv interpretiert werden und auch vom Analytiker als solche zu interpretieren sind. Insbesondere prosodisch-phonetische Verfahren gehören zu den indexikalischen Ressourcen, die nur in bestimmten sequentiellen Positionen als Ausdruck eines bestimmten Affekts interpretierbar sind. Die in diesem Beitrag beschriebenen Ressourcen, die Äußerungen als 'mitfühlend' kontextualisieren, sind besonders häufig in folgenden sequentiellen Umgebungen zu finden: in lexiko-syntaktisch expliziten Mitleidsbekundungen, in Thematisierungen affektiver Haltungen gegenüber nichtanwesenden Betroffenen oder zu Beginn von Erzählungen, in denen persönliches Erleben mit hoher emotionaler Involviertheit verhandelt wird. Anhand der gezeigten Beispiele wurde deutlich, dass es einer starken Kontextualisierung einer *consequential figure* als 'bemitleidenswert' bedarf, um eine Darstellung von Mitgefühl/ Mitleid/ Verständnis relevant zu machen. Ähnliches beschreibt Reber für die Lautobjekte [ɑː] und [hɑː] im Englischen, die an sequentiellen Positionen eingesetzt werden, "in denen durch die vorherige Handlung des Problemproduzenten eine Darstellung von erhöhter affektiver Beteiligung und Affiliation relevant gemacht wurde" (2009: 206). Die Kontextualisierung der *consequential figure* als bemitleidenswert kann durch diese selbst erfolgen (Bsp. (3), (4), (5)) oder durch eine andere Person (Bsp. (1), (2)). In allen Beispielen führte diese Kontextualisierung zu einer starken affektorientierten Reaktion.

Alle Beispiele weisen ähnliche prosodisch-phonetische Merkmalsbündel auf, die regelmäßig mit der Interpretation von 'mitfühlendem Sprechen' kookkurieren. In den Beispielen (1) und (2) wurde die affektive Haltung der Teilnehmerinnen mit Hilfe lexiko-semantischer Ressourcen als Mitgefühl bzw. als Mitleid explizit gemacht, in den Beispielen (4) und (5) wurde diese Affektorientierung auf Grund des Zusammenspiels des spezifischen prosodisch-phonetischen Merkmalsbündels mit anderen kinetischen Ressourcen und auf Grund der sequentiellen Position interpretierbar. Der Einsatz des Merkmalsbündels zur spielerischen Inszenierung einer Äußerung als 'mitfühlend' (Bsp. (3)) bietet zusätzliche Evidenz dafür, dass es sich bei 'mitfühlendem Sprechen' nicht nur um eine analytische Zuschreibung einer affektiven Haltung zu einer prosodisch-phonetischen Gestalt handelt, sondern dass es tatsächlich eine alltagsweltliche Kategorie zu sein scheint, die die TeilnehmerInnen aktiv verhandeln (vgl. Local/ Walker 2008).

Es bleibt zu vermuten, dass eine Relation zwischen dem Grad emotionaler Involviertheit und der Anzahl verwendeter multimodaler Ressourcen zur Darstellung eben dieser Involviertheit besteht, d.h. prosodisch-phonetische Verfahren 'mitfühlenden Sprechens' wirken im Zusammenhang mit lexiko-semantischen Ressourcen, die 'Mitfühlen' oder 'Mitleiden' nahe legen, möglicherweise 'stärker

involviert' als 'mitfühlend gesprochene' Äußerungen anderen lexiko-semantischen Gehalts. Wenig weist jedoch darauf hin, dass es sich bei Mitleid und Mitgefühl um distinkte Teilnehmerkategorien handelt (vgl. Reber 2009). Empathiedarstellungen sind vielfältig. Sie können eine Affektorientierung und/ oder eine Verstehensorientierung aufzeigen (vgl. Kupetz 2014). Handlungen, die auf Grund ihrer prosodisch-phonetischen Form als 'mitfühlend' interpretiert werden, sind in der Dimension der Affektorientierung zu verorten.

6. Literatur

Arminen, Ilkka (2004): Second stories: The salience of interpersonal communication for mutual help in Alcoholics Anonymous. In: Journal of Pragmatics 36, 319-347.

Auer, Peter/ Couper-Kuhlen, Elizabeth (1994): Rhythmus und Tempo konversationeller Alltagssprache. In: Zeitschrift für Literaturwissenschaft und Linguistik 96, 78-106.

Banse, Rainer/ Scherer, Klaus R. (1996): Acoustic profiles in vocal emotion expression. In: Journal of Personality and Social Psychology 70(3), 614-636.

Bergmann, Jörg R. (1987): Klatsch – Zur Sozialform der diskreten Indiskretion. Berlin: Walter de Gruyter.

Bergmann, Jörg R. (2001): Das Konzept der Konversationsanalyse. In: Brinker, Klaus et al. (Hrsg.): Text- und Gesprächslinguistik. Band 2: Gesprächslinguistik. HSK 16. Berlin: Walter de Gruyter, 919-927.

Bose, Ines (2003): Dóch da sin ja' nur mûster: Kindlicher Sprechausdruck im sozialen Rollenspiel. Frankfurt a.M.: Peter Lang.

Bußmann, Hadumod (Hrsg.) (2002): Lexikon der Sprachwissenschaft. 3., aktual. und erw. Auflage. Stuttgart: Kröner.

Christmann, Gabriela B./ Günthner, Susanne (1996): Sprache und Affekt – Die Inszenierung von Entrüstungen im Gespräch. In: Deutsche Sprache 24, 1-33.

Couper-Kuhlen, Elizabeth (2009): A sequential approach to affect: The case of 'disappointment'. In: Haakana, Markku/ Laakso, Minna/ Lindström, Jan (Hrsg.): Talk in Interaction – Comparative dimensions. Helsinki: Finnish Literature Society, 94-123.

Couper-Kuhlen, Elizabeth (2011): Affectivity in cross-linguistic and cross-cultural perspective. In: Stehl, Thomas (Hrsg.): Sprachen in mobilisierten Kulturen: Aspekte der Migrationslinguistik. Potsdam: Universitätsverlag, 231-257.

Couper-Kuhlen, Elizabeth (2012a): Exploring affiliation in the reception of conversational complaint stories. In: Peräkylä, Anssi/ Sorjonen, Marja-Leena (Hrsg.): Emotion in Interaction. Oxford: OUP, 113-146.

Couper-Kuhlen, Elizabeth (2012b): On affectivity and preference in responses to rejection. In: Text & Talk 32(4), 453-475.

Couper-Kuhlen, Elizabeth/ Selting, Margret (1996): Towards an interactional perspective on prosody and a prosodic perspective on interaction. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (Hrsg.): Prosody in conversation. Interactional studies. Cambridge: CUP, 11-56.

Deppermann, Arnulf (2013): Introduction: Multimodal interaction from a conversation analytic perspective. In: Journal of Pragmatics 46(1), 1-7.

Duan, Changming/ Hill, Clara E. (1996): The current state of empathy research. In: Journal of Counseling Psychology 43(3), 261-274.

Fiehler, Reinhard (1990): Kommunikation und Emotion. Berlin: de Gruyter.

Ford, Cecilia E./ Couper-Kuhlen, Elizabeth (2004): Conversation and phonetics: Essential connections. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (Hrsg.): Sound Patterns in Interaction – Cross-linguistic studies from conversation. Amsterdam: John Benjamins, 3-25.

Ford, Cecilia E./ Fox, Barbara A. (2010): Multiple practices for constructing laughables. In: Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (Hrsg.): Prosody in Interaction. Amsterdam/Philadelphia: Benjamins, 339-368.

Goffman, Erving (1981): Forms of Talk. Philadelphia: University of Pennsylvania Press.

Golato, Andrea (2012): German *oh*: Marking an Emotional Change of State. In: Research on Language and Social Interaction 45(3), 245-268.

Günthner, Susanne (1997): The contextualization of affect in reported dialogues. In: Niemeier, Susanne/ Dirven, René (Hrsg.): The Language of Emotions. Amsterdam: John Benjamins, 247-275.

Günthner, Susanne (2000): Vorwurfsaktivitäten in der Alltagsinteraktion: Grammatische, prosodische, rhetorisch-stilistische und interaktive Verfahren bei der Konstitution kommunikativer Muster und Gattungen. Tübingen: Max Niemeyer.

Hepburn, Alexa/ Potter, Jonathan (2007): Crying receipts: Time, empathy, and institutional practice. In: Research on Language and Social Interaction 40(1), 89-116.

Hepburn, Alexa/ Potter, Jonathan (2012): Crying and crying responses. In: Peräkylä, Anssi/ Sorjonen, Marja-Leena (Hrsg.): Emotion in Interaction. Oxford: OUP, 195-211.

Jefferson, Gail (1988): On the sequential organization of troubles talk in ordinary conversation. In: Social Problems 35(4), 418-442.

Kehrein, Roland (2002): Prosodie und Emotionen. Tübingen: Max Niemeyer.

Kranich, Wieland (2003): Phonetische Untersuchungen zur Prosodie emotionaler Sprechausdrucksweise. Frankfurt a.M.: Peter Lang.

Kupetz, Maxi (2013): Verstehensdokumentationen in Reaktionen auf Affektdarstellungen am Beispiel von ‚das glaub ich'. In: Deutsche Sprache 13 (1), Themenheft "Interaktionale Linguistik des Verstehens" hrsg. von Arnulf Deppermann, 72-96.

Kupetz, Maxi (2014): Empathy displays as interactional achievements – multimodal and sequential aspects. In: Journal of Pragmatics 61, 4-34.

Ladefoged, Peter (o.A.): A course in phonetics. 5. Ausgabe. (http://www.phonetics.ucla.edu/course/chapter1/vowels.html, letzter Zugriff am 24.07.2013)

Local, John/ Walker, Gareth (2008): Stance and affect in conversation: On the interplay of sequential and phonetic resources. In: Text & Talk 28(6), 723-747.

Mandelbaum, Jenny (2013): Storytelling in conversation. In: Stivers, Tanya/ Sidnell, Jack (Hrsg.): The Handbook of Conversation Analysis. Chichester: Wiley-Blackwell, 492-508.

Maynard, Douglas W./ Freese, Jeremy (2012): Good news, bad news, and affect – Practical and temporal 'emotion work' in everyday life. In: Peräkylä, Anssi/

Sorjonen, Marja-Leena (Hrsg.): Emotion in Interaction. Oxford: Oxford University Press, 92-112.

Nübling, Damaris (2001): Von 'oh mein Jesus!' zu 'oje!' – Der Interjektionalisierungspfad von der sekundären zur primären Interjektion. In: Deutsche Sprache 29, 20-45.

Ogden, Richard (2006): Phonetics and social action in agreements and disagreements. In: Journal of Pragmatics 38, 1752–1775.

Paeschke, Astrid (2003): Prosodische Analyse emotionaler Sprechweise. Berlin: Logos Verlag.

Pfänder, Stefan/ Gülich, Elisabeth (2013): Zur interaktiven Konstitution von Empathie im Gesprächsverlauf. Ein Beitrag aus der Sicht der linguistischen Gesprächsforschung. In: Breyer, Thiemo (Hrsg.): Grenzen der Empathie – Philosophische, psychologische und anthropologische Perspektiven. Paderborn: Wilhelm Fink, 433-458.

Reber, Elisabeth (2009): Zur Affektivität in englischen Alltagsgesprächen. In: Buss, Mareike et al. (Hrsg.): Theatralität des sprachlichen Handelns. Eine Metaphorik zwischen Linguistik und Kulturwissenschaften. München: Fink, 193-215.

Reber, Elisabeth (2012): Affectivity in Interaction: Sound objects in English. Amsterdam: John Benjamins.

Ruusuvuori, Johanna (2005): "Empathy" and "sympathy" in action: Attending to patients' troubles in Finnish homeopathic and general practice consultations. In: Social Psychology Quarterly 68(3), 204-222.

Sacks, Harvey (1971): Das Erzählen von Geschichten innerhalb von Unterhaltungen. Übersetzt von Ursula Christiansen. In: Kjolseth, Rolf/ Sack, Fritz (Hrsg.): Zur Soziologie der Sprache: ausgewählte Beiträge vom 7. Weltkongreß der Soziologie. Opladen: Westdeutscher Verlag, 307-314.

Sacks, Harvey (1995): Lectures on Conversation: volume II. Hrsg. von Gail Jefferson, 1. Publikation in einem Band. Oxford: Blackwell.

Scherer, Klaus R./ Ekman, Paul (Hrsg.) (1984): Approaches to Emotion. Hillsdale NJ: Lawrence Erlbaum.

Schwitalla, Johannes (1995): Kommunikation in der Stadt. Teil 4: Kommunikative Stilistik zweier sozialer Welten in Mannheim-Vogelstang. Berlin: de Gruyter.

Selting, Margret (1994): Emphatic speech style – with special focus on the prosodic signalling of heightened emotive involvement in conversation. In: Journal of Pragmatics 22, 375-408.

Selting, Margret (1995): Prosodie im Gespräch – Aspekte einer interaktionalen Phonologie der Konversation. Tübingen: Max Niemeyer.

Selting, Margret (2010a): Affectivity in conversational storytelling: An analysis of displays of anger or indignation in complaint stories. In: Pragmatics 20(2), 229-277.

Selting, Margret (2010b): Prosody in interaction: State of the art. In: Barth-Weingarten, Dagmar; Elisabeth Reber; Margret Selting (Hrsg.): Prosody in Interaction. Amsterdam/Philadelphia: John Benjamins, 3-40.

Selting, Margret (2012): Complaint stories and subsequent complaint stories with affect displays. In: Journal of Pragmatics 44, 387-415.

Selting, Margret (2013): Verbal, vocal, and visual practices in conversational interaction. In: Müller, Cornelia et al. (Hrsg.): Body – Language – Communication: An international handbook on multimodality in human interaction. Handbücher zur Sprach- und Kommunikationswissenschaft. Band 38(1). Berlin u.a.: de Gruyter Mouton, 589-609.

Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 10, 353-402.

Sidnell, Jack (2013): Basic conversation analytic methods. In: Stivers, Tanya/ Sidnell, Jack (Hrsg.): The Handbook of Conversation Analysis. Chichester: Wiley-Blackwell, 77-99.

Stock, Eberhard (1996): Deutsche Intonation. Berlin: Langenscheidt.

Szczepek Reed, Beatrice (2006): Prosodic Orientation in English Conversation. Basingstoke: Palgrave Macmillan.

Szczepek Reed, Beatrice (2011): Analysing Conversation – An introduction to prosody. Basingstoke: Palgrave Macmillan.

Tischer, Bernd (1993): Die vokale Kommunikation von Gefühlen. Weinheim: Psychologie-Verlags-Union.

Weiste, Elina; Anssi Peräkylä (2014): Prosody and empathic communication in psychotherapy interaction. In: Psychotherapy Research, (http://dx.doi.org/10.1080/10503307.2013.879619, letzter Zugriff am 04.09.2014)

Wendt, Beate (2007): Analysen emotionaler Prosodie. Frankfurt am Main: Peter Lang.

7. Anhang

Um den Höreindruck des Tonhöhenverlaufs, der als ‚(tief-)fallend, flach auslaufend' zu beschreiben ist, im Transkript abbilden zu können, wird die übliche GAT 2-Konvention (Selting et al. 2009) um folgende ikonische Zeichenkombination erweitert: ‚`_' für ‚tief fallend, flach auslaufend' bzw. ‚`¯' für ‚fallend, flach auslaufend'. Diese ‚flach auslaufenden' Konturen können sich über mehrere betonte und unbetonte Silben erstrecken.

Kinetische Handlungen werden in Doppelklammern in eigener Zeile und durch ‚|' mit den verbalen und vokalen Handlungen aligniert aufgeführt.

Maxi Kupetz
Universität Potsdam
Institut für Germanistik
Am Neuen Palais 10
14469 Potsdam
Germany
kupetz [at] uni-potsdam.de

Towards unravelling prosodic characteristics of speaker-overlapping laughing in conversational speech corpora

Jürgen Trouvain and Khiet P. Truong

Abstract
It is often reported that in spontaneous discourse the laughing of interlocutors overlaps with each other. Although transcripts of interactional linguistics do consider features of overlap, the main prosodic information regarding pitch, intensity, duration and interactional timing of laughter so far remained untracked. This paper aims to show that acoustic-phonetic studies can help to unravel the prosodic characteristics of speaker-overlapping laughs. For this purpose the annotated laughs of four corpora of conversational speech were analysed. The inspection of several thousands of laughs reveals that overlapping laughs are significantly longer, with a higher intensity, a higher fundamental frequency and are more voiced than non-overlapping laughs. Moreover, the frequency of overlapping laughs seems to be determined by the number of participants. Finally, a closer analysis of the interactional timing in one dialogue corpus revealed that the scheme "speaker invites recipient to join in to laugh" occurs more often than "recipient laughs before speaker".

laughter – corpora – conversational speech – overlapping vocalisation - prosody

Es wird häufig berichtet, dass sich im spontanen Diskurs die Gesprächspartner beim Lachen überlappen. Obwohl Transkripte in der Gesprächsforschung Merkmale der Überlappung beachten, blieb bisher die primäre prosodische Information zu Tonhöhe, Intensität, Dauer und Timing in der Interaktion unbeachtet. Akustisch-phonetische Studien können dabei helfen, die prosodischen Charakteristika sprecher-überlappenden Lachens aufzudecken. Zu diesem Zwecke wurden mehrere Tausend annotierter Lachereignisse in vier Gesprächskorpora analysiert. Es zeigte sich, dass überlappende Lachereignisse signifikant länger, mit höherer Intensität, einer höheren Grundfrequenz und mehr stimmhaften Anteile produziert wurden als nicht-überlappende Lacher. Die meisten Lacher in Dialog-Korpora waren allerdings nicht-überlappend, im Gegensatz zu einem Korpus mit Mehr-Parteien-Gesprächen. Zudem zeigte eine genauere Analyse des interaktionellen Timings in einem Dialog-Korpus, dass das Schema "Sprecher lädt Rezipient zum Lachen ein" häufiger vorkommt als "Rezipient lacht vor dem Sprecher".

Lachen – Korpora – Gesprächsdaten – überlappende Vokalisierungen – Prosodie

1. Introduction

In contrast to most read speech we can frequently find in conversational speech nonverbal vocalisations such as laughter, coughing or clearing the throat. When we exclude so-called fillers like *uh* and *uhm* as well as feedback expressions such as *hm*, *u-huh* and *yeah* from the nonverbal vocalisations, laughter seems to be the

most frequent vocalisation after audible breathing sounds (Trouvain/ Truong 2012). Although laughing is so immanently present in every-day speech, research of laughter in conversations can so far be considered as rather limited from a phonetic/ prosodic as well as from a perspective of discourse and conversation analysis (exceptions are the collections in Trouvain/ Campbell 2007 and Wagner/ Vöge 2010). The aim of this paper is to present some findings of the prosodic characteristics of laughter in interaction which can be of interest for discourse and conversation analysis.

1.1. Transcribing laughter

In discourse and conversation analytical transcription systems such as the "Gesprächsanalytische Transkriptionssystem 2" (GAT 2) (Selting et al. 2009: 367), it is recommended to transcribe laughter either as *hahaha, hehe, hihi* when it is a shorter or "syllabic" laugh corresponding to the number of pulses or laugh syllables, or as "((laughs))" between double rounded brackets as a description of laughter. Speech-laughs or laughed speech should be transcribed as "<<laughing> word >" whereas the convention for smile voice is "<< :-) > word>".

In their example transcript taken from a real conversation, Selting et al. (2009: 394) depict laughter at various places. The transcriptions of various laughing events differ from event to event: In line 28 the laughs start with *phhh*, continued by *hohoho*, in lines 38 and 39 *hm*, in lines 40 and 40 and 41 it is *(lacht verstohlen, ca. 1.2 Sek lang)*, in lines 62, 63 and 65 it is *hehehe* or *he he*, and in line 66 *he ho ho*. (Interlinear translations of the text were inserted below the respective lines in the transcript.)

(1) Excerpt from Selting et al. (2009)

```
27   S2:   und sich mit  den NACHbarn   ange[legt,=ne,  ](1.2)
           and PRON with the neighbours argued     PTCL
           and argued with the neighbours

28   S1:                                     [phhh hohoho]

((...))

36   S2:   und ham wir immer  gesagt HIT[ler;=ne,]
           and have we always said   Hitler   PTCL
           and we always said Hitler

37   S1:                                 [HITler; ]

38   S1:   h[m, ]

39   S2:    [hm,]
40   S1:   [((lacht verstohlen,        ca. 1.2 Sek. lang))]
           ((laughs in a stealthy way, ca. 1.2 sec.))

41   S2:   [((lacht verstohlen,        ca. 1.2 Sek. lang))]

((...))
```

```
61   S2:   jetzt wohnt er nämlich mit seiner SCHW:ESter
           now  lives he actually with  his  sister

           zusa[mmen. ]
           together
           now he actually lives together with his sister

62   S1:       [hehehe]

63   S1:   he [he ]

64   S2:      [und] die LAUfen: (.) RUM  wie  n  URaltes
               and  they run      around like an old-age

           ehe[paar.]=ne,
           couple    PTCL
           and they behave like an old-age couple

65   S1:      [he he]

66   S1:   he ho [ho ]

67   S2:         [OUH] mann. (.) heh
                 PTCL man
                 Oh boy!
```

The diversity of these transcriptions show various interpretations of various laughter events, be it a laugh with a "syllabic" structure ("hohoho"), laughs with different vowel qualities ("he" vs. "ho"), laughing without any vowel articulation ("hm"), with a more literal description ("verstohlen" in German or "stealthy" in English) or with temporal information ("ca. 1.2 Sek. lang"). Especially with the interpretative comments, it is unclear whether a second transcriber would come up with the same or similar transcription of the laughing events. While it is not the aim of this paper to criticise this kind of diversity in transcription, it shows two things: 1) each transcription is always an interpretation. This holds in particular for non-words without standard spelling, and 2) the diversity of the ways to transcribe laughter mirrors the diversity of laughter in general. The current paper would like to take up this issue by comparing the prosodic features of speaker-overlapping laughter with those of non-overlapping laughter.

1.2. Overlapping laughter

The transcripts above show a further important feature of laughter in interaction: all laughs in this example transcript were overlapping, either with speech of the co-participant (e.g. lines 28-65) or with laughter of the co-participant (e.g. lines 38-41). It is not unusual that speakers overlap with their vocalisations. The conversational principle "one speaker at a time" (Sacks et al. 1974, Stivers et al. 2009) does not always hold (see e.g. Liddicoat 2007). This restriction is supported by corpora of conversational speech where a considerable amount of "cross-talk" was observed (e.g. Campbell 2007b or Heldner/ Edlund 2010). Laughter in particular has a tendency to overlap with laughter as could be shown by Laskowsi/ Burger (2007), Truong/ Trouvain (2012b) and also Smoski/ Bachorowski (2003).

Laughter seems to serve as the optimal opportunity for a joint vocalisation. In lines 40 and 41 of the transcript above we see synchronous laughter of both interlocutors whereas the laughter in lines 28, 62, 63 and 65 overlaps with speech of the interlocutor. Interestingly, the overlaps in the latter cases are transcribed as aligned on a syllabic basis, e.g. "he [he]" in line 63 aligns with "[und]" in line 64.

We define overlapping laughter (henceforth OL) here as laughter of speaker A that overlaps with laughter of speaker B. We consider all other instances of laughter as non-overlapping laughter (henceforth NOL), i.e. the laughter in lines 28, 62, 63 and 65 would be NOL. So far it seems unclear whether OL can be seen as a frequent phenomenon in conversational speech. But in any case it would be interesting to find out more about the mechanisms of how interlocutors manage to produce an OL in talk-in-interaction. One plausible challenging explanation is the "invitation-acceptance scheme" proposed by Jefferson (1979):

> Laughter can be managed as a sequence in which speaker of an utterance invites recipient to laugh and recipient accepts that invitation. One technique for inviting laughter is the placement, by speaker, of a laugh just at completion of an utterance, and one technique for accepting that invitation is the placement, by recipient, of a laugh just after the onset of speaker's laughter.

1.3. Research questions

This scheme provides an elegant explanation for how partners in conversation accomplish OL, which might be a common form in social interaction. Nevertheless, there are various questions left unanswered:

- Is there a difference in the phonetic substance between OL and NOL?
- Do interlocutors show proportionally more or fewer OL than NOL?
- Among the OL, how many times does the invitation-acceptance scheme apply?
- How does this scheme work in time? Do laugh invitees wait to be invited?

This paper will take up such questions. Before we present some of our findings we provide some background on the phonetics of laughter and overlapping vocalisations.

2. Phonetic background on laughter and overlapping vocalisations

2.1. Phonetic sciences on laughter

Laughter typically occurs in social interaction, especially in conversations and not so much in read speech or other forms of monologues. Since phonetic research consists to a substantial degree of investigations of controlled speech, conversations are only infrequently the object of this branch of research. In addition, in phonetics typically single words and sentences of read speech are investigated, thus ignoring nonverbal vocalisations at all.

The limited amount of research on laughing performed by phoneticians brought about the phonetics of laughing as a multi-disciplinary research matter. Among them are psychological disciplines with focus on developmental, evolutionary or affective perspectives but also speech technology and other speech-based disciplines such as conversation analysis and interactional linguistics.

2.2. Defining laughter

Although there is no accepted standard definition of laughter, there seems to be an agreement that laughter is a social and/ or affective signal expressed with an acoustic and a visual display in interaction with others. There are various reasons why humans laugh. In the literature, factors such as cooperation, social bonding, affiliation, creating a pleasant atmosphere, feedback or back-channelling, topic termination or humour are mentioned among others (see e.g. Bachorowski et al. 2001, Smoski/ Bachorwoski 2003, Chafe 2007, Holt 2010). This list suggests that laughter is connected to many factors at rather diverse levels.

A question still unsolved is whether laughing and smiling belong to the same category or whether they are two distinct categories. Depending on the definition, smile voice could theoretically also belong to laughter, however, we consider it here as different from laughter. Smiling per se happens in a silent way whereas laughter per se happens in a non-silent way. The acoustic transmission of smiling needs some speech as a "carrier", whereas laughter is usually produced as an autonomous vocalisation. However, laughing can also be produced during articulation of speech (Nwokah et al. 1999, Trouvain 2001). These speech-laughs or laughed speech is again distinct from smiled speech. Often a speech-laugh turns into a laugh vocalisation.

When we look more closely into laugh vocalisations we see rather quickly that there is not only *one* type of laughter (e.g. often transcribed with "hahaha" or "hehehe") but that there is a bundle of various phonetic forms and combinations of these forms (see also Trouvain 2003). The following parameters seem to be relevant:

- voicing,
- duration,
- the number of "syllables" or "calls" (cf. Bachorowski et al. 2001) or "pulses" (Chafe 2007),
- onset noise before and an offset of inhalation and silence after the laughed "syllables" (cf. Chafe 2007),
- intensity,
- fundamental frequency (can be often extremely increased).

It should be noted that there is a huge variability regarding the listed parameters between individuals but also intra-individually (e.g. Vettin/ Todt 2004). This large phonetic variation of laughs also includes a considerable amount of laughs that show a rather mild intensity with a comparably short duration – phonetic conditions that make these mild laughs susceptible to be overheard, particularly when

the listener is focused on speech or when the laugh events occur in overlapping vocalisations.

2.3. Overlapping vocalisations in conversations

Laughter seems to represent an optimal opportunity for joint vocalisation. Various studies show that there is a strong tendency of laughter to overlap with vocalisations (either speech or nonverbal vocalisations) of the interlocutor (Laskowsi/ Burger 2007, Truong/ Trouvain 2012b and also Smoski/ Bachorowski 2003). Joint vocalisations are not uncommon. Think of "collaborative completions" as continuations by the conversational partner with matching prosodic features (e.g. Lerner 1991, Local 2005). Other joint vocalisations that require across-speaker coordination are, for instance, synchronous reading aloud (Cummins 2007), singing in a choir and a common prayer in church.

Fig. 1 illustrates the often observed overlap of dialogue partners when laughing with an example taken from the German Lindenstraße corpus (IPDS 2006) transcribed in (2).

(2) Lindenstraße Corpus (IPDS 2006, Dialogue 4, sec. 164-173)

```
164   L: is halt son   skelett    im flur    und sie meint dann so
         is just such a skeleton in corridor and she  says then so

166      mai, was <<interjection> pfn>
         well  what INTERJ

167      <<laughing>machen die leut   [da>   (---)    °h]
                    make   the people there
         There is a skeleton in the corridor and she goes like "well, what are the people doing
         there?"

168   S:                               [hehehehehehehe °h]
169       echt   nee hier wars  so,  dass wir ne kurze szene wo
          really no  here  was it so that we  a  short scene where

          sie  in der straße entlanggehen und dann °h
          they in the street  go along    and then
          Really? No, here we had a short scene where she walked along the street and then
          …
```

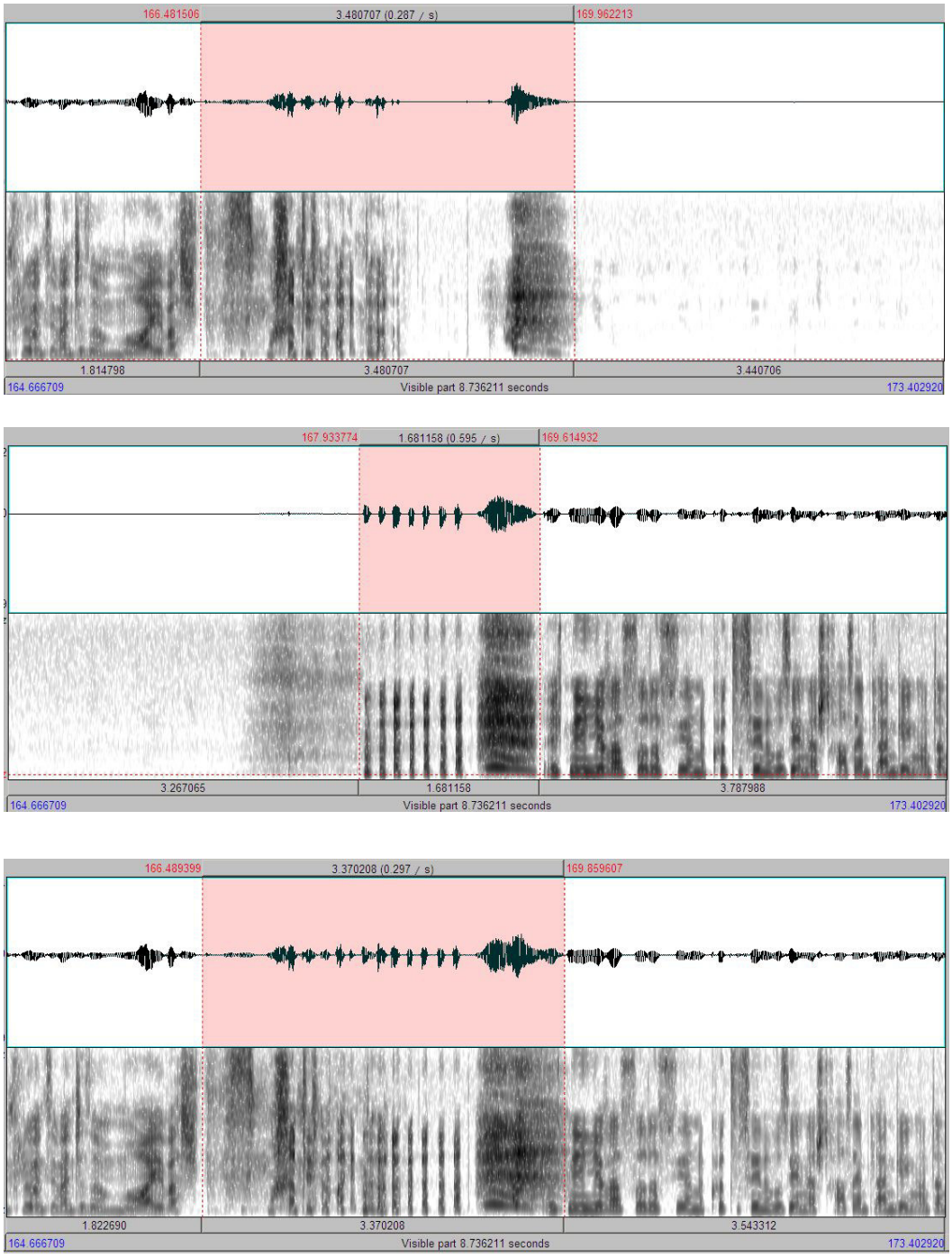


Fig. 1: Example of laughter at turn-taking (from Lindenstraße Corpus (IPDS 2006), Dialogue 4, from 164 to 173 sec) represented as wave form and spectrogram (0-8 kHz). Top: left channel speaker. Mid: right channel speaker. Bottom: mixed speech signal. Marked with colour: the laugh production.

Since each speaker was recorded with a different microphone, their vocalisations can be represented separately. Such a recording with separate channels allows us to determine the *exact timing* of the laughter in the dialogue flow and the *exact phonetic-prosodic realisations* of the laughter by each speaker. In recordings without channel separation (which seems to be the case in most dialogue corpora, e.g. the audio file that corresponds to the example transcript in Selting et al.

2009), these details of prosody are masked. In addition, some less intense vocalisations may not even be noticeable when masked in a recording without channel separation.

3. Method

We report here selected aspects from previous studies on laughing based on already annotated corpora of conversational speech (Truong/ Trouvain 2012a, b and Trouvain/ Truong 2012, 2013). In contrast to purely auditory approaches, which are usual in conversation analytic research, we are interested in the acoustic properties of various prosodic parameters including the proportion of voicing, the fundamental frequency, the intensity and the duration of laugh tokens. These acoustic characteristics have the advantage that they can be compared across corpora and independently of the interpretation of transcribers. A further advantage is that *large* amounts of data can be analysed by automatic procedures.

The size of data we are working with goes beyond the data set usually considered in qualitative studies as used in conversation and interactional analysis. We would like to stress that this difference is not meant to disqualify qualitative studies. Rather, quantitative studies are considered a way to *complement* qualitative studies in order to strengthen the findings and theoretical statements therein.

There are two important differences between qualitative and quantitative approaches of the phonetic/ prosodic analysis of vocalisations. First, for a quantitative analysis the data will be treated automatically and not manually as it is usual for a qualitative analysis. There are some prerequisites for an automatic processing: the electronic availability of annotations, the time stamps of the start and the end of the annotated laugh token, and the corresponding speech signal files for all speakers of the conversation. As stated before, the channels of both speakers must be separated in order to clearly recognise the overlap, which is important to perform an acoustic-phonetic analysis.

Second, employing a quantitative analysis of already existing corpora means that the annotations were made by others than the current researchers. This fact can sometimes lead to some disagreement between the existing annotations of certain laugh tokens and how the researchers would have annotated the same token (for more details on these issues see Truong/ Trouvain 2012a). When working with smaller sample sizes the mentioned disagreements could affect the findings. For our rather big sample size (more than 18,000 laugh tokens, see also Table 1) we consider this risk marginal.

3.1. Corpora used

Four English-speaking corpora of conversational speech were selected for analysis.

1. The AMI Meeting Corpus (Carletta 2007) contains audiovisual recordings (100 hours) of elicited design meetings, in which a team of 4 persons are discussing the design of a new remote control. In these meetings, the par-

ticipants were put together in a room and were assigned certain roles to play (project manager, marketing expert, industrial designer, and user interface designer).

2. The ICSI Meeting Corpus (Janin et al. 2003) contains audio recordings of 75 natural research meetings that were actually held at the International Computer Science Institute (ICSI) over a period of 3 years. The number of participants per meeting varies between 3 and 11, yielding an average of 6 participants per meeting. The topics of the meetings concern, among others, natural language processing, the development of the ICSI Meeting Corpus, and methods to compensate for noise for automatic speech recognition.
3. The Diapix Lucid Corpus (Baker/ Hazan 2011) contains audio recordings of unscripted task-based dyadic interactions. Each participant is given a different version of a cartoon picture and is seated in a separate room. The two participants communicate via headsets to locate the twelve differences between the two pictures.
4. The HCRC Map Task Corpus (Anderson et al. 1991) contains audio recordings of unscripted task-based dyadic interactions (with and without eye contact). Each participant is given a different version of a map and a role, that of "instruction giver" or "instruction follower". The instruction giver's map contains a route that should be reproduced on the instruction follower's map with as few deviations as possible.

As shown in Table 1 the four corpora differ severely with respect to number of recorded speakers, number of recorded conversations and duration of each conversation. Consequently, the number of annotated laughs differs as well. Differences in the annotation practice and on the technical level (more details in Truong/ Trouvain 2012a) led to a reduction of the number of laughs used in the current analysis. The main differences concern:

- speech-laughs which are sometimes ignored and sometimes inconsistently labelled (for this analysis we ignored them altogether).
- the definition of what counts as a laugh: sometimes the annotated laugh is in reality composed of two or more laughs, and vice versa, two annotated laughs are in reality one laugh. It also happens that the annotated laugh is only partially a laugh or sometimes it is unclear whether it was a laugh or not.
- technical errors: sometimes the annotated laughs show incorrect time stamps for beginning and/ or end.

Despite the listed drawbacks of annotation we consider the existing corpora as usable – conceding, though, that we are not considering completely correct data. A very time-consuming re-annotation would be necessary to obtain more homogeneous laughter annotations across corpora, that in turn will lead to more consistent and reliable research results.

On a qualitative level the corpora are not homogenous either. AMI and ICSI are multi-party conversations whereas Map Task and DiaPix are dialogues. The level of acquaintance is rather different, with the exchanges including friends, colleagues and also strangers. The sensual modalities also differ: two corpora (AMI

and ICSI) used face-to-face conversations, one corpus (DiaPix) used only the acoustic channel and one corpus (Map Task) used both conditions. There are also differences regarding the communicative settings in which the recordings were situated with the ICSI corpus containing authentic conversations and the other three corpora with naturalistic but task-based speech.

	AMI	**ICSI**	**Map Task**	**DiaPix**
no. of annotated laughs	16477	12574	1002	582
no. of used laughs	8587	8268	966	575
no. of speakers	679	494	250	114
no. of conversations	165	74	95	52
no. of speakers per convers.	4	3-11	2	2
mean duration of convers. (in mins, standard deviation in parentheses)	35.1 (13.5)	55.0 (15.9)	6.8 (3.1)	7.7 (2.3)
total durat. of all convers. (in hours)	94.5	67.9	11.4	6.8
visual contact	yes	yes	yes/ no	no
relationship between speakers	mostly strangers	colleagues	friends, strangers	friends

Table 1: Descriptive features of the four inspected corpora.

Apart from these differences, we can assume that *single* tokens of laughter from these (and other) corpora may heavily depend on conditions like visual contact, social distance, and communicative task. There is no doubt that a *qualitative* analysis has a high level of explanatory power – however, it is restricted to an *extremely limited* number of tokens (sometimes n=1). When we are interested in generalisations, for instance to check a certain hypothesis based on a qualitative analysis, then certain qualitative prosodic features of laughs should occur across various corpora and the inherent differences described above. Corpora with annotations with hundreds and thousands of tokens can represent a solid base for such an approach.

3.2. Analysis

We carried out a semi-automatic acoustic analysis of the prosody of corpora of conversational speech which are already annotated. Ideally, the annotated labels "laughter" are aligned with the corresponding section in the speech signal. These sections of speech can be automatically analysed with self-authored Praat scripts for prosodic parameters such as duration, fundamental frequency (F0), intensity and also the amount of voiced frames.

The temporal alignment was important to see whether speakers overlap or not. Please note that two laughs which overlap are always *partially* overlapping. Of course for some laughs one can have the auditory impression that both speakers start laughing at the same time. However, with the temporal granularity of milliseconds which is usual for acoustic-phonetic studies the delay of time is always nicely visible. Schematically such overlaps can be pictured like in Fig. 2, where

the aforementioned "invitation-acceptance-pattern" by Jefferson (1979) is illustrated.

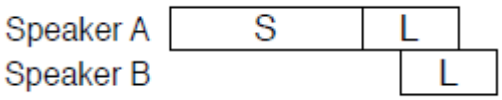


Fig. 2: Schematic illustration of two overlapping laughs (L). The current speaker ends the phase of articulated speech (S) with a laugh. The speaker in the role of the recipient joins in with a laugh after the speaker's laugh.

Transcript 3 from Jefferson (1979: 81) can serve as an example where the speaker laughs at the completion of an utterance and that the recipient accepts this invitation by laughing after the onset of the speaker's laugh.

(3) Excerpt from Jefferson (1979: 81)

```
Ellen:   He s'd well he'd said I am cheap he said, 'hh
         about the big things. he says but not the liddle
         things, hhhHA HA [HA HA HA
Bill:                     [heh heh heh
```

The advantage of taking already annotated data is that the annotation of thousands of laughs is already available (cf. Table 1). The already mentioned disadvantage is that many various labellers performed the annotation in possibly different ways. As a laugh we consider a span of vocalisation of one speaker annotated as "laugh". This means that an overlap of laughing as illustrated in Fig. 2 counts as two laughs (one laugh from speaker A and one laugh from speaker B). It is important to note that speech-laughs were not considered here because in some corpora they were not regarded at all and in others they were not annotated in a consistent way.

The frequencies of occurrences of overlapping (OL) and non-overlapping laughs (NOL) were determined in all four corpora with the help of Praat scripts. OL are laughs like those in Fig. 2. Please note that NOL are laughs that do not overlap with the laugh of the interlocutor but they can overlap with the speech of the other. For each laugh annotated and used, the mean values of the fundamental frequency (F0), the intensity, the duration and the number of voiced frames were automatically extracted, again using Praat scripts. Thereafter the values for F0 and intensity were normalised (using a z-transformation as the standard procedure of normalisation).

A closer analysis of the timing schemes was performed with the data of the Map Task corpus only. For this purpose we first created plots of "speech and laugh activities" like the one in Fig. 3.

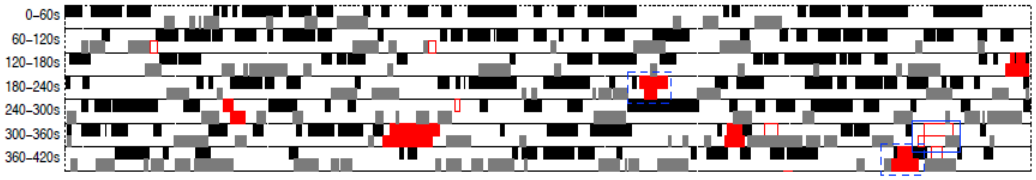


Fig. 3: Speech and laugh activity of the first seven minutes of the dialogue q1nc2 from the Map Task corpus. Each line represents 60 seconds of speech with filled dark bars for the speech activity of the speaker with the instructor role and filled grey bars for the speech activity of the speaker with the receiver role. Empty sections represent silences. Filled red bars stand for overlapping laughs, empty red bars for non-overlapping laughs. Blue rectangle with solid border marks an example of a clear laughter pair in which the laughs do not overlap. Blue rectangles with dashed borders mark examples of Jefferson's invitation-acceptance laughter pairs.

Theoretically, two overlapping combinations are possible: either the current speaker starts laughing (pattern 1, see Fig. 2 and Table 2) or the co-participant (pattern 2). Likewise, two non-overlapping combinations are theoretically possible when either the current speaker shows some "solo laughter" (pattern 3) or his/ her interlocutor (pattern 4).

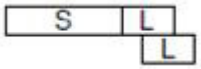
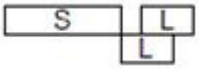
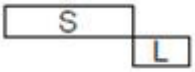
1 speaker A ends the turn with a laugh, speaker B starts laughing *after* A's onset of the laugh (= the "invitation-acceptance scheme" proposed by Jefferson (1979));	
`A: und sich mit den NACHbarn angelegt, he[he]` `B:  [phhh hohoho]`	
2 speaker A ends the turn and speaker B starts laughing *before* A's onset of the laugh;	
`A: und sich mit den NACHbarn angelegt,  [hehe]` `B:  phhh [hohoho]`	
3 speaker A ends the turn with a laugh, no laugh of speaker B overlaps or follows;	
`A: und sich mit den NACHbarn angelegt, hehe` `B:`	
4 speaker A ends the turn without a laugh, but speaker B laughs.	
`S2: und sich mit den NACHbarn angelegt,` `S1:  phhh hohoho`	

Table 2: The four theoretically possible patterns are illustrated with the help of the first example in the transcript from the introductory section.

4. Results

4.1. Frequency of occurrence

Fig. 4 shows the absolute number of occurrences of OL and NOL in the four corpora. In one corpus, ICSI, the majority of laughs are OL (65%). For the three other corpora the picture is reversed with OL between 35% (AMI), 38% (Map Task) and 41% (Diapix).

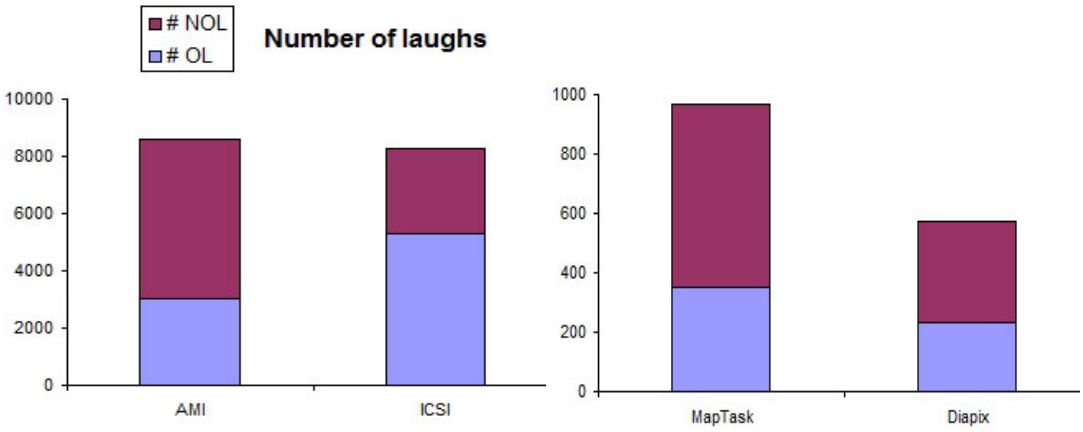


Fig. 4: Absolute number of laughs of the four corpora divided into overlapping (OL) and non-overlapping laughs (NOL). The larger corpora (with more than 8000 tokens) are on the left, the smaller corpora (with more than 500 tokens) on the right.

A look at the relative amount of laughter (related to the total speaking time) reveals differences between the four corpora (Fig. 5). The multi-party conversations show fewer laughs per minute per speaker than the dialogues.

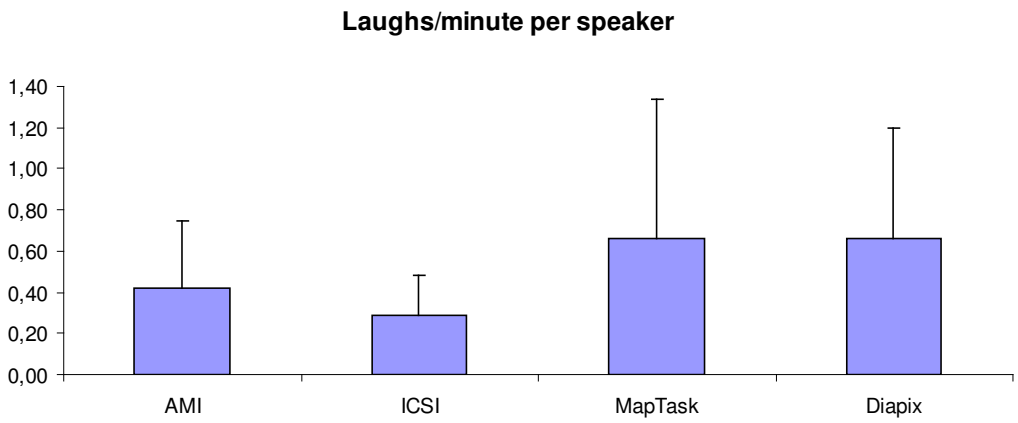


Fig. 5: Laughing rate as the relative number of laughs of the four corpora expressed as an average number of laughs per minute per speaker. Illustrated are the means and standard deviation.

4.2. Prosodic characteristics of overlapping and non-overlapping laughter

The results for the mean duration of OL and NOL are illustrated in Fig. 6. OL are longer in all corpora. This difference in duration is statistically significant based on t-tests for each corpus (for the statistical details the reader is referred to Truong/ Trouvain 2012b). There are also statistically significant differences between the different corpora with ICSI showing the longest durations followed by AMI and both dialogue corpora with the shortest durations.

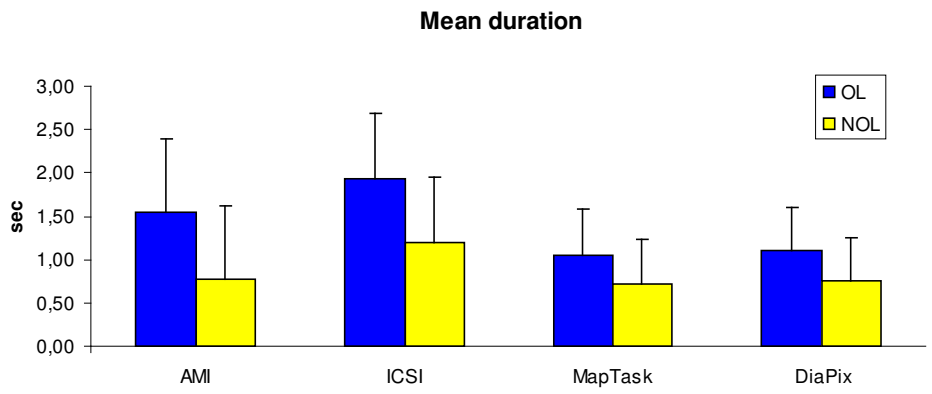


Fig. 6: Values for mean and standard deviation of the duration in the four corpora for overlapping (OL) and non-overlapping laughs (NOL).

The analysis of the fundamental frequency reveals that OL is higher than NOL. The same relationship can be stated for the mean and the maximal intensity as well as for the amount of voiced portions. All differences proved to be statistically significant for all corpora.

When we consider completely unvoiced tokens the number is higher for NOL (see Fig. 7). This finding holds again for all corpora with statistical significance.

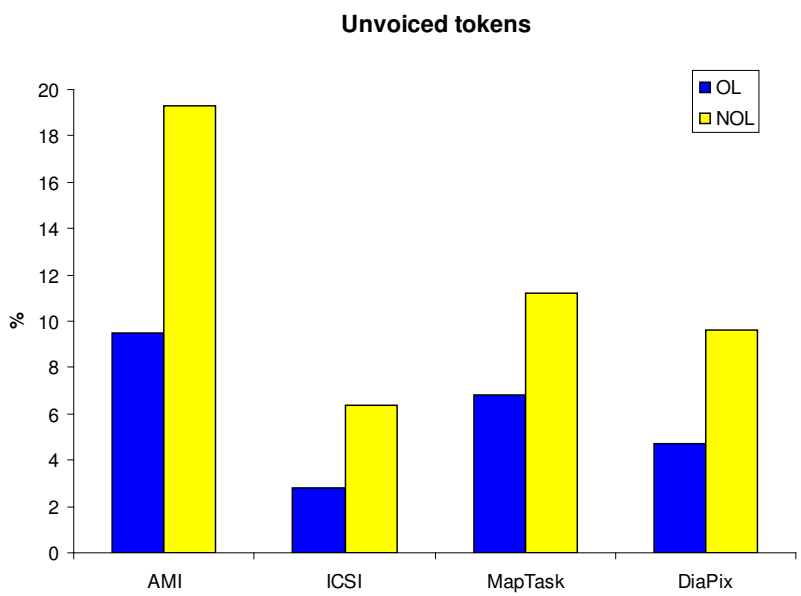


Fig. 7: Percentage of completely unvoiced tokens in the four corpora for overlapping (OL) and non-overlapping laughs (NOL).

To summarise the results, the production of an OL in comparison to a NOL is characterised by a longer duration, a higher fundamental frequency, a higher intensity, a higher amount of voicing and a reduced tendency for complete devoicing. Interestingly, there is statistical evidence for all four acoustic parameters investigated for all four corpora investigated. For this reason we can consider the results as very stable across the different corpora.

4.3. Overlapping laughter as "invitation followed by acceptance"

Table 3 summarises the results regarding the analysis of the timing of laughs with respect to speech in the Map Task corpus. As already seen before, the laughs of the type OL represent the minority of all laughs (38.6%). The two combinations assumed for the analysis of OL (see the first two left-hand rows in Table 3) account for 24.8% of all laughs. Type 1 (speaker invites recipient to laugh) occurs twice as often as type 2 (recipient does not wait with laughing).

Both NOL combinations amount to half of the data (50.8%). Type 3 (speaker laughs without response) and type 4 (only recipient laughs) are balanced.

It must be noted that about one quarter of all laughs could not be assigned unambiguously to one of the four assumed categories. Either the annotated laughs were part of one larger and complex laugh (e.g. with inhalation phases in between which were not annotated as part of the laugh), or there was no speech or other vocalisation immediately preceding, which could be seen as a further category (see Table 3). In eight percent of the cases we were simply unsure as to how to analyse the laughs regarding their timing to speech because we did not see a clear fit in the proposed patterns. All these laughs require a qualitative analysis for further classification.

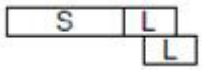
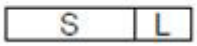
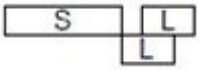
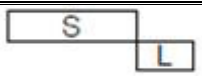
Type	OL	Type	NOL
	16.1%		24.2%
	8.7%		26.6%
'unsure' 'no speech before' 'complex'	3.9% 2.4% 7.5%	'unsure' 'no speech before' 'complex'	3.3% 5.8% 1.5%
total	38.6%	total	61.4%

Table 3: Frequency of speech-laugh patterns observed in the HCRC Map Task corpus (shown in percentages of the total number of laughs). OL= overlapping laugh, NOL= non-overlapping laugh, S=speech, L=laugh (data taken from Trouvain/ Truong 2013).

5. Discussion and conclusions

5.1. Laugh-in-interaction: laughing alone vs. laughing together

The analysis of the annotated laugh tokens in four different corpora shows that a substantial amount of laughter occurs at a time when other interlocutors are laughing as well. The amount of overlapping laughs ranges from one third of all laughs to two thirds. This finding supports the idea that laughter is a joint vocalisation par excellence. The principle "one speaker at a time" which is often assumed to be valid in conversation does obviously not hold for laughing in conversations. Although this is not a new finding, we now have a notion that this is not an infrequent event and can therefore not be considered marginal. Moreover, the results show that types of conversation can heavily vary in their number of OL. This fact has to be taken into account for investigations of laughter in general.

The ICSI corpus, with up to 11 participants in a conversation, shows a remarkably high number of overlapping laughs compared to the other corpora. In this corpus the mean duration of a laugh is substantially longer than in the other corpora. However, each speaker does not show as many laughs per minute as speakers in the other corpora. These observations may reflect the fact that people in a larger communicative community behave differently in situations in which only two persons are engaged. Experiments with the same speakers in dyadic vs. multiparty conversations could shed more light on aspects of how individuals change their laughing behaviour in terms of frequency of occurrence and prosodic parameters.

Overall, the majority of laughs were produced as NOL. One possible explanation for a NOL is that it is a rejected invitation. This would be covered by type 3 of our analysis of the Map Task corpus, which amounts to one third of the entire data. However, a qualitative analysis is needed to decide whether each laugh not responded to by a recipient was really a rejection. The strengths of a quantitative analysis are i) to see how frequent or marginal certain types of laughs occur, ii) what their general acoustic patterns are and iii) to show with the visualisation techniques presented where in the conversation relevant laughs were produced. Thus it can help to find interesting events faster and to gain an overall picture of conversations, in our case regarding laughter.

5.2. Phonetic-prosodic distinction of overlapping and non-overlapping laughter

OL substantially differ from NOL in terms of their prosodic make-up. The distinction between OL and NOL regarding their productions is clear-cut and based on a solid fundament of data. This finding does not exclude that NOL and OL are two separate categories. Moreover, there can be a lot of variation between individuals, text types (or communicative settings) and cultures. However, at least for the English data the *general* pattern of the phonetic-prosodic substance of OL and NOL seems to be clear: The prosodic signaling system of OL compared to NOL uses longer duration, a higher fundamental frequency, a higher intensity and more

voicing. All these acoustic features probably contribute to attract more attention from the listener. In contrast, NOL don't seem to be primarily produced to get attention from the listeners. It thus might be that some laughs occur below our perceptual threshold, especially when overlapped by *speech* of the other – a circumstance which could also affect transcribers of conversational data.

One explanation for this listener-attraction of OL is that laughter is used as a positive social signal to display affiliation and social bonding. OL as a mainly voiced vocalisation with longer duration, higher intensity and higher pitch requires more physiological effort than the virtual default case. In line with this, there are findings that dialogues with voiceless laughs were associated with feelings of social exclusion. Laughing for the purpose of social integration ("laughing *with* somebody") is accomplished best with voiced and prosodically prominent forms of laughter typical of OL. Laughing for the purpose of social segregation ("laughing *at* somebody") seems to work better with unvoiced laughs with a lower physiological effort often found in NOL (Cirillo/ Todt 2005).

5.3. "Laugh to be invited"

The pattern of the "invited laugh" as described by Jefferson (1979) seems to occur substantially more often than the case of the "anticipated laugh", where the recipient laughs before the onset of the "inviting" speaker. Obviously there is a majority of occasions where people show a tendency to "wait" to be invited to a shared laugh rather than anticipating an overlapping laugh. This result could be seen as a type of convention – similar to the widespread convention that somebody entering a room greets first, answered by the person/s already in the room.

An open question remains how laughter is invited. Does in those "invited" cases the laugh of the "inviter" start with the prosodic characteristics of a typical OL, which then signals the recipient to join in or not?

As mentioned above the majority of laughs are non-overlapping. Future studies must reveal the possible communicative functions of these laughs, including rejections of invitations, self-comments (cf. Trouvain 2001) or comments as part of a feedback expression (cf. Ward/ Tsukahara 2000).

5.4. Prosody of laughing beyond transcriptions

We have seen at the beginning of this paper what kind of information is usually recorded in transcripts of conversational analysis. Often these transcripts are based on recordings of only one channel (which can be mono or stereo). Working with signals without a separation of the different speakers carries the risk that acoustic signals in overlapping phases are masked, making the job of human perception of laughs of different persons very hard and sometimes impossible.

A further advantage of separate channels is the possibility to determine the *exact* timing of overlapping vocalisations. This granularity is presumably not relevant for most applications of these transcripts. However, it can play a role for instance in learning more about vocal alignment patterns like the laugh-invitation and who was really first.

The prosodic characteristics of fundamental frequency and intensity have not been studied in conversation analysis so far. Duration is either indicated by the number of laugh "syllables" like [hehehe] or a time span like "1.2 sec". It remains unclear which benefit the indication of the vowel quality bears, laughing with a high front vowel might be different for the interpretation of a communicative situation than laughing vowels that resemble low back vowels. However, acoustic analyses of (many) laugh events show a general tendency that laughter has no clear vocal tract configurations (see e.g. Bachorowski et al. 2001). Instead it could be more helpful for more detailed acoustic analysis to give some indications of the perceived pitch, loudness, duration and also whether the laugh was voiced or voiceless.

5.5. Laughter and interactional prosody

Conversation analysis and interactional linguistics are important research approaches to study the prosodic nature of talk-in-interaction, often ignored by phonetic sciences and speech processing. The analysis of fine phonetic details, be it on an auditive or on an acoustic level, unravel phenomena that are of great interest for phonetic and prosodic research. These phenomena definitively include laughter and other nonverbal vocalisations (see Ogden 2013 for a recent example). Analysis of single cases can help to uncover conversational mechanisms how laughing works in conversations; see, for instance, studies showing that shared laughter is often associated with topic termination (Holt 2010, Shaw et al. 2013). It is extremely important to describe the link between certain pragmatic functions, e.g. topic termination, and prosodic characteristics.

Going beyond singular examples, on the other hand, can help to understand how often and how stable certain features and patterns occur. The latter is what we did with our corpus studies. Another example is the work by Bonin et al. (in press) to investigate in large corpora the aforementioned link between topic changes and laughter. In our view the findings and their interpretations can be useful for interactional phonetics in general and in particular for the prosody of laughing as one of the most frequent nonverbal vocalisations.

6. Acknowledgements

The authors would like to thank the editors of this volume for valuable feedback on an earlier draft of this paper. This work was partly funded by the EU's 7th Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet).

7. References

Anderson, Anne H. et al. (1991): The HCRC Map Task Corpus. In: Language and Speech 34, 351-366.

Bachorowski, Jo-Anne/ Smoski, Moria J./ Owren, Michael (2001): The acoustic features of human laughter. In: Journal of the Acoustical Society of America 110, 1581-1597.

Baker, Rachel/ Hazan, Valerie (2011): DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. In: Behavior Research Methods 43(3), 761-770.

Bonin, Francesca/ Campbell, Nick/ Vogel, Carl (in press). Time for laughter. In: Knowledge-Based Systems. DOI: 10.1016/j.knosys.2014.04.031.

Campbell, Nick (2007a): Whom we laugh with affects how we laugh. In: Proceedings of the Interdisciplinary Workshop on the Phonetics of Laughter Saarbrücken, 61-65.

Campbell, Nick (2007b): Approaches to conversational speech rhythm: speech activity in two-person telephone dialogues. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 343-348.

Carletta, Jean (2007): Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. In: Language Resources and Evaluation 41, 181-190.

Chafe, Wallace (2007): The Importance of Not Being Earnest. Amsterdam: Benjamins.

Cirillo, Jasmin/ Todt, Dietmar (2005): Perception and judgement of whispered vocalisations. In: Behaviour 142, 98-125.

Cummins, Fred (2007): Speech synchronization: Investigating the links between perception and action in speech production. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 529-532.

Heldner, Matthias/ Edlund, Jens (2010): Pauses, gaps and overlaps in conversations. In: Journal of Phonetics 38, 555-568.

Holt, Elizabeth (2010): The last laugh: shared laughter and topic termination. In: Journal of Pragmatics 42 (6), 1513-1525.

IPDS (2006): Video Task Scenario: Lindenstraße – The Kiel Corpus of Spontaneous Speech, Volume 4, DVD, Institut für Phonetik und Digitale Sprachsignalverarbeitung Universität Kiel.

Janin, Adam et al. (2003): The ICSI meeting corpus. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 364-367.

Jefferson, Gail (1979): A technique for inviting laughter and its subsequent acceptance/ declination. In: Psathas, George (ed.): Everyday Language: Studies in Ethnomethodology. New York: Irvington, 79-96.

Laskowski, Kornel/ Burger, Susanne (2007): Analysis of the occurrence of laughter in meetings. In: Proceedings of the Interspeech Conference, Antwerp, 1258-1261.

Lerner, Gene H. (1991): On the syntax of sentences in progress. In: Language in Society 20, 441-458.

Liddicoat, Antony J. (2007): An Introduction to Conversation Analysis. New York: Continuum.

Local, John (2005): On the interactional and phonetic design of collaborative completions. In: Hardcastle, William/ Beck, Janet MacKenzie (eds.): A Figure of Speech: a Festschrift for John Laver. New Jersey: Lawrence Erlbaum, 263-282.

Nwokah, Eva E. et al. (1999): The integration of laughter and speech in vocal communication: a dynamic systems perspective. In: Journal of Speech, Language & Hearing Research 42, 880-894.

Ogden, Richard (2013): Forms and functions of clicks in English conversation. In: Journal of the International Phonetic Association 43(3), 299-320.

Sacks, Harvey/ Schegloff, Emanuel A./ Jefferson, Gail (1974): A simplest systematics for the organisation of turn-taking for conversation. In: Language 50, 696-735.

Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 10, 353-402.

Shaw, Chloe/ Hepburn, Alexa/ Potter, Jonathan (2013): Having the last laugh: On post-completion laughter particles. In: Glenn, Phil/ Holt, Elizabeth (eds.): Studies of Laughter in Interaction. London: Bloomsbury, 91-106.

Smoski, Moria J./ Bachorowski, Jo-Anne (2003): Antiphonal laughter in developing friendships. In: Annuals of the New York Academy of Sciences 1000, 300-303.

Stivers, Tanja et al. (2009): Universals and cultural variation in turn-taking in conversation. In: Proceedings of the National Academy of Sciences of the United States of America (PNAS) 106(26), 10587-10592.

Trouvain, Jürgen (2001): Phonetic aspects of "speech-laughs". In: Proceedings of the Conference on Orality & Gestuality (Orage), Aix-en-Provence, 634-639.

Trouvain, Jürgen (2003). Segmenting phonetic units in laughter. In: Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS), Barcelona, 2793-2796.

Trouvain, Jürgen/ Campbell, Nick (eds.) (2007): Proceedings of the Workshop on the Phonetics of Laughter, Saarbrücken (http://www.coli.uni-saarland.de/conf/laughter-07).

Trouvain, Jürgen/ Truong, Khiet P. (2012): Comparing nonverbal vocalisations in conversational speech corpora. In: Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals, Istanbul, 36-39.

Trouvain, Jürgen/ Truong, Khiet P. (2013): Exploring sequences of speech and laughter activity using visualisations of conversations. In: Proceedings of the Workshop on Affective Social Speech Signals (WASSS 2013), Grenoble, 5 pages.

Truong, Khiet P./ Trouvain, Jürgen (2012a): Laughter annotations in conversational speech corpora possibilities and limitations for phonetic analysis. In: Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals, Istanbul, 20-24.

Truong, Khiet P./ Trouvain, Jürgen (2012b): On the acoustics of overlapping laughter in conversational speech. In: Proceedings of the Interspeech Conference, Portland, 4 pages.

Vettin, Julia/ Todt, Dietmar (2004): Laughter in conversation: features of occurrence and acoustic structure. In: Journal of Nonverbal Behaviour 28(2), 93-115.

Wagner, Johannes/ Vöge, Monika (eds.) (2010): Laughter in Interaction. Special Issue of the Journal of Pragmatics 42 (6), 1469-1764.

Ward, Nigel/ Tsukahara, Wataru (2000): Prosodic features which cue back-channel responses in English and Japanese. In: Journal of Pragmatics 32, 1177-1207.

Jürgen Trouvain
Saarland University
Department of Computational Linguistics and Phonetics
Campus C7.2
66123 Saarbrücken
Germany
Trouvain [at] coli.uni-saarland.de

Khiet P. Truong
University of Twente
Human Media Interaction
P.O. Box 217
7500 AE Enschede
The Netherlands
k.p.truong [at] utwente.nl

Stimmlich-artikulatorischer Ausdruck in der Kita: Erzieherinnen betrachten mit zwei- und vierjährigen Kindern ein Bilderbuch

Ines Bose und Stephanie Kurtenbach

Abstract
In einer Untersuchung zum Gesprächsverhalten von frühpädagogischen Fachkräften im Gespräch mit Kindern im Kindergartenalltag wurde besonders der Umgang mit dem stimmlich-artikulatorischen Ausdruck betrachtet. Material sind authentische Videodaten von 11 Erzieherinnen aus fünf Kindertagesstätten im Gespräch mit zwei- und vierjährigen Kindern beim gemeinsamen Betrachten eines Bilderbuchs (ausführlich vgl. Bose/ Kurtenbach/ Nixdorf 2013).
Im Beitrag wird anhand von Gesprächsbeispielen zweier Erzieherinnen gezeigt, dass ein sensitiv-responsiver Umgang mit dem stimmlich-artikulatorischen Ausdruck in der Kita gelingen kann, aber offensichtlich sehr anspruchsvoll ist: Während es der einen Erzieherin hervorragend gelingt, sich an den unterschiedlichen Kommunikationsbedürfnissen und Sprachentwicklungsständen von zwei- und vierjährigen Kindern zu orientieren und darauf sensitiv-responsiv zu reagieren, verwendet eine andere Erzieherin den stimmlich-artikulatorischen Ausdruck vergleichsweise starr und schablonenhaft und geht am Interesse und an den kommunikativen Fähigkeiten der Kinder vorbei.
Sensitiv-responsiv eingesetzte stimmlich-artikulatorische Ausdrucksweisen können die Erzieherinnen nutzen, um das Interesse der Kinder beim gemeinsamen Ansehen des Bilderbuches zu wecken, eine vertrauensvolle Gesprächssituation zu schaffen und das Gespräch zu strukturieren. Erzieherinnen können Figuren aus dem Buch animieren und Ereignisse demonstrieren oder illustrieren. Durch gegenseitiges Imitieren oder Modifizieren expressiver Sprechausdrucksgestaltungen können sie emotionale Nähe und Gemeinsamkeit erzeugen sowie eine Gesprächsatmosphäre etablieren, in der die Kinder zum Sprechen und Reflektieren über das Buch sowie zum Spielen mit sprachlichen, para- und nichtsprachlichen Formen ermuntert werden.

Mündliche Kommunikation in Kindertagesstätten – sensitiv-responsives Verhalten von Erzieherinnen – stimmlich-artikulatorischer Ausdruck

This article presents a study on the communicative behavior of early education personnel conversing with children in everyday situations in the kindergarten. In particular it focuses on the way the teachers use vocal-articulatory resources. The corpus consists of authentic video data of 11 kindergarten teachers from five day-care centers, having conversations with two- and four-year-olds while looking at a picture book together (for more detail see Bose/ Kurtenbach/ Nixdorf 2013).

By means of the examples of two kindergarten teachers' conversations, it will be demonstrated that a sensitive-responsive use of vocal-articulatory resources can be used successfully in day-care centers, but is obviously quite challenging: While one kindergarten teacher is particularly successful in taking into account the children's different communicative needs and stages of language development

and reacting accordingly in a sensitive-responsive way, the other teacher uses her vocal-articulatory resources in a rather schematic manner and thus fails to meet the interests and communicative competence of the children.

Early education personnel can employ sensitive-responsive vocal-articulatory behavior in order to attract the children's interest to look at picture books together, to create a secure conversational situation and to structure the conversation. They can bring the book's characters to life, demonstrate and illustrate what is happening. Through mutual imitation or modification of expressive ways of speaking, kindergarten teachers are also able to create emotional closeness and a feel of similarity and establish a conversational atmosphere, in which children are encouraged to talk about and reflect upon the book as well as to play with linguistic, para- and non-linguistic forms.

oral communication in kindergarten – vocal-articulatory resources – sensitive-responsive use

1. Kommunikation in der Kindertagesstätte

Die folgende Analyse ist eingebettet in aktuelle sprechwissenschaftliche Forschungen und Praxisprojekte zur Kommunikation zwischen frühpädagogischen Fachkräften und Kindern in Kindertagesstätten (Kitas) am Seminar für Sprechwissenschaft und Phonetik der Martin-Luther-Universität Halle-Wittenberg (vgl. Kurtenbach/ Bose (Hrsg.) 2013). Der Fokus dieser Forschungen liegt sowohl auf der Analyse des Gesprächsverhaltens von Pädagoginnen und Kindern als auch auf der Stärkung von Sprachbildung und -förderung im Kita-Alltag. Letztere bilden einen wichtigen Aufgabenbereich der Erzieherinnen. Den analytischen Zugang der Gesprächsforschung (mit sprechwissenschaftlicher Schwerpunktsetzung) kombinieren wir mit sprach- und sprechtherapeutischer bzw. -beratender Expertise (Schwerpunkt auf kindlicher Sprachentwicklung). In den genannten Projekten interessieren uns vor allem folgende Fragen:

- Wie verlaufen Gespräche zwischen Erzieherinnen und Kindern in Kindertageseinrichtungen?
- Worin zeigt sich Gesprächskompetenz von Erzieherinnen?
- Gibt es typische Verhaltensweisen der Erzieherinnen und wenn ja, welchen Einfluss haben sie auf das Gesprächsverhalten der Kinder?
- Welche Rahmenbedingungen beeinflussen die Kommunikation zwischen Erzieherinnen und Kindern?
- Wie können aus den Forschungserkenntnissen praxistaugliche Schulungsinhalte abgeleitet und im Sinne einer alltagsintegrierten Sprachbildung eingesetzt werden?

1.1. Gesprächskompetenz von Erzieherinnen

Aufgrund der internationalen Leistungsvergleiche seit der ersten PISA-Studie im Jahr 2000 spielen Bildung und Förderung der Teilkompetenz Sprache in den Kitas eine besondere Rolle, und die Einrichtungen stehen diesbezüglich unter großem

Leistungsdruck. Sie sind angehalten, die frühpädagogischen Fachkräfte zu professionalisieren, damit diese zur Sprachbildung und -förderung befähigt werden. Der Dialog zwischen Erzieherinnen und Kindern sowie zwischen den Kindern untereinander gilt als Motor für den Spracherwerb. Dabei sollen die Erzieherinnen „bedeutungsvolle Gesprächsanlässe" im Sinne der kindlichen Erfahrung schaffen und den Kindern besondere Aufmerksamkeit widmen (Albers 2009, 2011). Erzieherinnen haben die Aufgabe, Gesprächssituationen zu moderieren, die Gespräche behutsam zu leiten und zu begleiten, möglichst alle Kinder zu eigenen Äußerungen zu ermutigen und ihnen – gemäß ihrem individuellen Entwicklungsstand – Impulse und gegebenenfalls Hilfestellung zu geben (Albers 2009, Jampert et al. 2009).

Während Eltern intuitiv mit ihren Kindern interagieren und sich durch den ständigen intensiven Kontakt sensibel an deren individuellen Entwicklungsstand anpassen (vgl. z.B. Papoušek 1994), muss im Rahmen pädagogischer Kontexte persönliche Intuition ergänzt werden durch pädagogische Professionalität (vgl. z.B. Merkel 2005, Bredel et al. 2008, Nentwing-Gesemann et al. 2011). Dazu gehört die Fähigkeit, das kommunikationsförderliche Potenzial in alltäglichen Interaktionssituationen zu erkennen und auszuschöpfen. Sprachliche Bildungswirkungen stellen sich also im Kita-Alltag nicht selbstverständlich ein, sondern es bedarf einer entsprechenden Sensibilisierung oder auch Schulung der Fachkräfte. Erzieherinnen müssen eine sensitiv-responsive (Beller/ Beller et al. 2007, Rempsberger 2013) sprach- und kommunikationsbildende wie -fördernde Grundhaltung erwerben und sie im Kita-Alltag professionell umsetzen; also „die Signale des Kindes bemerken und sich auf die Signale des Kindes hin angemessen verhalten" (Rempsberger 2013: 9). Gefordert ist also eine professionelle Gesprächskompetenz der frühpädagogischen Fachkräfte (zum Begriff der Gesprächskompetenz vgl. u.a. Lepschy 2002, Deppermann 2004, Meer/ Spiegel 2009).

Nach eigenen Beobachtungen zeichnen sich sensitiv-responsiv agierende Erzieherinnen in der Kita dadurch aus, dass sie kindliches Verhalten genau beobachten, das kindliche Kommunikationsbedürfnis erkennen, nach selbst geäußerten Fragen feinfühlig abwarten, kindlichen Äußerungen zuhören, sich den Kindern sprachlich, para- und nichtsprachlich freundlich-aufmerksam zuwenden, sich mit den Kindern auf Blickhöhe begeben, ihre Impulse unterstützen u.v.m.. Sprachlich greifen sie kindliche Kommunikationsimpulse auf, passen eigene Äußerungen an, imitieren und bestärken kindliche Sprachäußerungen, wiederholen und erweitern sie, greifen mit Fragen das Interesse des Kindes auf und motivieren zu Antworten, führen die Themen der Kinder fort (Kurtenbach 2011). Im Folgenden geht es vor allem darum zu zeigen, wie Erzieherinnen sensitiv-responsiv auch mit dem stimmlich-artikulatorischen Ausdruck umgehen.

1.2. Stimmlich-artikulatorischer Ausdruck

„Stimmlich-artikulatorischer Ausdruck" wird in der Sprechwissenschaft als Oberbegriff für situations- und stimmungsadäquat konventionalisierte Gestaltungsweisen im Sprechschall verwendet (zur Begriffsbestimmung vgl. z.B. Krech et al. 1991, Gutenberg 2001, Bose 2003 und 2010, Bose et al. 2013). Es handelt sich um einen Merkmalskomplex aus stimmlich-melodischen, aber auch aus tempora-

len, dynamischen und artikulatorischen Parametern, die an auditiv wahrnehmbaren physiologisch-akustischen Grundlagen der Stimmproduktion und Artikulation orientiert sind. Über den stimmlich-artikulatorischen Ausdruck wird die Haltung des Sprechers zum geäußerten Inhalt eines Gesprächsbeitrags markiert und vom Partner interpretiert (Rollen-, Gruppen-, Handlungsmuster). Stimmlich-artikulatorischer Ausdruck verweist damit nicht nur auf das Individuum, das seinen Empfindungen willkürlich oder unwillkürlich Ausdruck gibt, sondern auch auf den Adressaten, bei dem Empfindungen und Assoziationen angeregt werden, und damit auf den Interaktionsprozess.

Eltern behandeln ihre kleinen Kinder intuitiv zumeist als aktive Gesprächspartner: Sie hören den kindlichen (vorsprachlichen und sprachlichen) Äußerungen aufmerksam zu, beantworten sie unmittelbar, wiederholen und variieren sie stimmlich-artikulatorisch auf spielerische Weise. Sie passen sich damit intuitiv an die sich entwickelnden kommunikativen Bedürfnisse und Fähigkeiten der Kinder an (Papoušek 1994, Tomasello 1999). Generell verwenden die Eltern vereinfachte, prototypische stimmlich-artikulatorische Grundmuster (z.B. das Motherese; vgl. stellvertretend Bruner 1987, Papoušek 1994, Ritterfeld/ Siegert 2000), wiederholen sie häufig, kontrastieren sie und setzen sie ritualisiert als Antwort auf kindliche Äußerungen ein. Damit erleichtern sie den Kindern die Strukturierung der Kommunikation; darüber hinaus vermitteln sie Sprache und Wissen über Sprache. Die Kinder erwerben stimmlich-artikulatorische Ausdruckskonventionen im Verlauf ihrer Kommunikationsentwicklung sehr früh und probieren sie expressiv und differenziert aus (Konopczynski/ Tessier 1994, Wells 2010), zum Beispiel in ihren Rollenspielen. Von den elterlichen Ausdrucksmustern sind für sie offensichtlich besonders das Motherese und das komplementäre, autoritative Ausdrucksmuster sehr eindrucksvoll. Zumindest benutzen sie diese beiden Muster selbst häufig, wenn sie im Rollenspiel in der Spielidentität eines ‚Erwachsenen' zu einem ‚kleinen Kind' sprechen. Dabei verwenden sie weniger die sprachstrukturellen Merkmale, sondern vor allem die stimmlich-artikulatorischen Ausdrucksformen in sehr treffender Weise (vgl. die Untersuchungen von Bose 2003).

Im vorliegenden Beitrag wird anhand von zwei Erzieherinnen gezeigt, dass ein sensitiv-responsiver Umgang mit dem stimmlich-artikulatorischen Ausdruck auch in der Kita gelingen kann, aber offensichtlich sehr anspruchsvoll ist: Während es der einen Erzieherin hervorragend gelingt, sich an den unterschiedlichen Kommunikationsbedürfnissen und Sprachentwicklungsständen von zwei- und vierjährigen Kindern zu orientieren und darauf sensitiv-responsiv zu reagieren (Beispiele 1 und 2), verwendet die andere Erzieherin den stimmlich-artikulatorischen Ausdruck vergleichsweise starr und schablonenhaft und geht am Interesse und an den kommunikativen Fähigkeiten der Kinder vorbei (Beispiel 3).

2. Datenbasis

2.1. Videokorpus

Für die Analyse des stimmlich-artikulatorischen Ausdrucks von Erzieherinnen in der Kita haben wir Videodaten benutzt, die aus dem Landesmodellprojekt Sach-

sen „Sprache fördern – Erprobung und Multiplikation von Methoden der Sprachförderung“ (2007-2011) stammen (vgl. Abschlussbericht des BBW 2011). In diesem Projekt sollten Effekte von Schulungsmaßnahmen für die Erzieherinnen zur Sprachförderung erfasst werden. Dazu wurden in ausgewählten Kitas die Sprachleistungen der Kinder (mittels standardisierter Sprachtests) wie auch das Gesprächsverhalten der Pädagoginnen (vor allem der Einsatz der vermittelten Sprachlehrstrategien und die Komplexität des grammatischen Angebots) in einem Prä-Postvergleich erhoben. Für die Betrachtung des stimmlich-artikulatorischen Ausdrucks in diesem Artikel ist dieses Design irrelevant, da die Schulungsmaßnahmen keine parasprachlichen Hinweise oder Empfehlungen beinhalteten.

Eine externe Beobachterin filmte in fünf Kitas 44 Kleingruppengespräche von 11 frühpädagogischen Fachkräften (Zeitraum 2009-2010, Gesprächsdauer zwischen sieben und dreißig Minuten – vgl. ausführlich zum Korpus Kurtenbach/ Bose/ Thieme 2013: 26 ff.). Als zu filmende Situation wurde das gemeinsame Betrachten eines Bilderbuchs gewählt – jeweils mit einer Kleingruppe von zweijährigen und einer von vierjährigen Kindern. Es handelt sich hierbei um eine alltägliche Gesprächssituation in der Kita, die vielerlei Möglichkeiten bietet, Kinder in ihrer Sprachentwicklung zu unterstützen und zu fördern (vgl. stellvertretend Grimm 2003, Braun 2007, Tarplee 1996). Die Kinder konnten wählen, wie lange sie sich am Buchanschauen beteiligen wollten. Die Pädagoginnen sollten gemäß den Schulungsschwerpunkten den Kindern Gelegenheit zu sprachlichen Äußerungen geben, kindliche Äußerungen aufgreifen, bestätigen und erweitern (ausführlich hierzu vgl. Kurtenbach/ Bose/ Thieme 2013).

2.2. Gesprächssituation ‚Gemeinsam ein Bilderbuch betrachten‘

Im Landesmodellprojekt war für die Videoaufnahmen die Betrachtung des Frühlingswimmelbuches von Rotraud Susanne Berner (2008) vorgegeben. Wie alle Wimmelbücher enthält dieses Buch keinen Text, sondern besteht ausschließlich aus detailreichen Situationsbildern, auf denen es viel zu entdecken gibt. Einige Figuren und Gegenstände sind am unteren Buchrand prominent positioniert und tauchen im Verlauf des Buches immer wieder auf. Der Betrachter des Buches kann diese Figuren und Gegenstände in verschiedenen Episoden während eines Frühlingstages verfolgen. Da das Buch keinen Text enthält, wurde in den hier videografierten Situationen keine Geschichte vorgelesen oder nacherzählt. Stattdessen mussten sich die Gesprächspartner auf einen Gesprächsgegenstand einigen und interaktive Verfahren entwickeln, um auf das Bilderbuch Bezug zu nehmen. Meist haben sie auf den detailreich bebilderten Buchseiten zusammengehörende Personen und Gegenstände identifiziert und benannt. Gelegentlich haben sie darüber hinaus aufgrund des wiederholten Auftauchens der Personen und Gegenstände im Buch begonnen, eine mögliche Geschichte zu entwerfen. Manchmal gingen die Reflexionen auch über das Buch hinaus und es wurde – angeregt durch die Bilder im Buch – über eigene Erlebnisse und Erfahrungen gesprochen (ausführlicher hierzu vgl. Kurtenbach/ Bose/ Thieme 2013). Während die vierjährigen Kinder mit der Bilderfülle meist gut zurechtkamen und Zusammenhänge zwischen den vielen Detailbildern (im Sinne einer Geschichte) erkennen und formulieren

konnten, waren die zweijährigen Kinder in der Regel damit noch überfordert und auf die Strukturierungshilfe der Erzieherinnen angewiesen.

3. Untersuchungsmethode

Wir haben das Videokorpus aus dem genannten Landesmodellprojekt über die dort verfolgten Forschungsziele hinaus gesprächsanalytisch ausgewertet. Dabei haben wir neben sprachlichen auch nichtsprachliche, vor allem aber parasprachliche Merkmale im Gesprächsverhalten der Erzieherinnen und Kinder betrachtet (vgl. Kurtenbach/ Bose (Hrsg.) 2013). Als Grundlage für die detaillierte Analyse dienten Transkripte in ELAN nach GAT 2 (Selting et al. 2009), sie wurden durch weitere Eintragungen ergänzt (vgl. die Legende der Notationszeichen am Ende des Beitrags). An der Durchsicht des Korpus waren mehrere Expertinnen beteiligt, so dass eine intersubjektive Übereinstimmung bei der Analyse garantiert ist.

Um die parasprachlichen Merkmale zu betrachten, haben wir in den Gesprächen zunächst nach systematisch wiederkehrenden stimmlich-artikulatorischen Ausdrucksgestaltungen von Erzieherinnen und Kindern gesucht (zum Vorgehen vgl. Bose et al. 2012, Bose/ Kurtenbach/ Nixdorf 2013). Ausgewählt wurden insbesondere solche Beispiele, in denen Pädagoginnen und Kinder bestimmte Ausdrucksgestaltungen mit erkennbarer Funktion und Wirkung verwendet haben. Zum Beispiel rahmen die Erzieherinnen zu Beginn von Bilderbuchbetrachtungen durch sehr expressive Ausdrucksgestaltungen (u.a. Tonhöhensprünge, starke Dehnungen der Vokale, auffällige Akzentgestaltung) die Situation als eine besondere und schaffen damit eine Gemeinsamkeit: Sie lenken die Aufmerksamkeit der Kinder auf das Bilderbuch und wecken deren gespannte Erwartung. Wichtig war in unseren Untersuchungen der enge Bezug von stimmlich-artikulatorischen Ausdrucksgestaltungen vor allem zur sprachlichen Ebene des Gesprächs, aber auch zu nichtsprachlichen Handlungen. Selbstverständlich sind *eindeutige* Form-Funktions-Zuschreibungen nicht möglich, in den Gesprächen überlagern sich verschiedene Funktionen und Wirkungen. Aber es lassen sich Sequenzen finden, in denen eine bestimmte Funktion bzw. Wirkung für bestimmte stimmlich-artikulatorische Ausdrucksmuster im Vordergrund zu stehen scheint.

Die identifizierten Musterverwendungen wurden dann hinsichtlich folgender Merkmale aus dem Katalog von Bose (2003) auditiv-phonetisch analysiert: Stimmklang, Sprechstimmlage, Melodieumfang und -verlauf, Lautheit, Gliederung/ Pausierung, Sprechgeschwindigkeit, Akzentuierung, Phonationsart, Sprechrhythmus und Sprechspannung. In den hier präsentierten Transkriptausschnitten ist der stimmlich-artikulatorische Ausdruck überwiegend anhand von Komplexeindrücken (Gestalten) vermerkt, in den paraphrasierenden Erläuterungen und Interpretationen wird er detailliert beschrieben.

4. Exemplarische Analysen

Für den vorliegenden Artikel haben wir drei exemplarische Analysen ausgewählt, die für das Korpus typische stimmlich-artikulatorische Ausdrucksgestaltungen zeigen. Zunächst werden zwei Ausschnitte aus Gesprächen einer Erzieherin

(ERZ1) analysiert, zum einen mit drei zweijährigen, zum anderen mit drei vierjährigen Kindern. Grund für diese Auswahl ist ein inhaltlicher Bezug: In beiden Aufnahmen gibt es eine längere Gesprächspassage über dieselben Bilder im Frühlingswimmelbuch, nämlich über die Tätigkeit und Funktion von drei Windrädern auf einem Feld (Seite 2 links oben). Für alle Kinder war das Buch neu.

Darüber hinaus wurde ein drittes Beispiel ausgewählt: Es handelt sich um einen Ausschnitt aus einem Gespräch einer anderen Erzieherin (ERZ2) mit drei Vierjährigen, in dem zahlreiche abgebildete Gegenstände und Personen auf zwei Doppelseiten des Buchs benannt werden. Hier kennen die Kinder das Buch offensichtlich, für die Erzieherin ist es dagegen neu.

4.1 Beispiel 1 (Erzieherin 1 mit Zweijährigen): *und die WINDräder die !^DREH::!en sich;/*

Die Erzieherin sitzt mit den drei etwa zweijährigen Kindern auf einer Matte (s. Standbilder 1_1 und 1_2: rechts die Erzieherin, auf ihrem Schoß Ken, in der Mitte Ron und links außen Mia; vgl. Audiobeispiel 1[1]; Gesamtdauer ca. 15 min). Mia und Ron haben hörbare artikulatorische Schwierigkeiten. Ken ist in seiner Sprachentwicklung verzögert; er verwendet in der gesamten Aufnahme lediglich ein Protowort (*toto*), wenn er auf die Bilder im Buch zeigt. Er beteiligt sich nicht verbal am Gespräch, trotz mehrfacher Versuche der Erzieherin, ihn zum Sprechen zu ermuntern. Allerdings verfolgt er das Gespräch aufmerksam.

Zu Beginn der Situation hatte sich die Erzieherin Zeit genommen, um die Situation als besondere zu rahmen und die Aufmerksamkeit der Kinder auf das Bilderbuch zu richten: Sie hat expressiv-freundlich gewürdigt, dass die Kamerafrau *EXtra [...] aus der ´GRO:ßen stadt ^LEIPzig [...] ein !TOL!les !^BU:CH!* mitgebracht hat. Sie hat die Kinder wertschätzend als *die !^GRO:!ßen* bezeichnet, die gemeinsam mit ihr in einer gemütlichen Ecke des extra hergerichteten Raumes dieses neue Bilderbuch anschauen dürfen. Ausführlich hat sie mit den Kindern die Umschlagseite mit den zahlreichen Bildern betrachtet und kommentiert.

Schließlich schlägt Ken die erste Doppel-Innenseite auf. Die Kinder entdecken die abgebildeten Gegenstände und Personen, zeigen darauf und benennen sie (z.B. einen Bagger, einen Mann mit Bart). Die Erzieherin greift diese Benennungen auf und erweitert sie für gewöhnlich sprachlich-inhaltlich (z.B. dass mit dem Bagger vielleicht eine Straße gebaut wird). Der hier diskutierte Ausschnitt beginnt bei Minute 4:11: Ken zeigt auf die drei abgebildeten Windräder und es entspinnt sich eine längere Erörterung über deren Tätigkeit.

[1] Aus Datenschutzgründen kann hier nur die Tonaufnahme zugänglich gemacht werden.

Bild 1_1: 04:31	Bild 1_2: 04:44

1) Windräder (PO_T1_S_2_04:11-04:46)[2]

```
450   KEN:    ((zeigt auf Windräder-Bild)
451   ALLE:   ((schauen auf das Bild))
452   ERZ1:   ((schaut Ken an) <<erstaunt>°hh und was ist
                DA:S?//>)
453   KEN:    ((zeigt erneut auf Windräder-Bild))
454   ERZ1:   <<lächelnd flüsternd> (-) SAG_s mal;//>
455   RON:    ((zeigt auch auf Bild) [<<ff, skand> OH (das
                                                   ist)>])
456   MIA:    ((zeigt auch auf Bild) [<<f> DA:S?//>         ])
457   RON:    ((schaut ERZ an) (was_s DAS;//))
458   ERZ1:   <<all> WEIß ich ni was IST_n das;//>
459           [           was IST_n das,//]
460   RON:    [<<flüsternd> (...) RÄder;//)>]
461   ERZ1:   <<behaucht> `RÄ:↑`de:r `RICH´tig,//>
462           (.) [<<flüsternd> WIN:Drä:de:r;//>       ]
463   MIA:        [((zeigt auf Windräder-Bild) DA:S?//)]
464   KEN:    ((zeigt erneut auf Bild))
465   ERZ1:   und die WINDräder die !^DREH::!en sich;/
466           da kann man;/
467           (.)<<flüsternd> da muss !^GA:NZ! viel WIND kommen
              und DANN>
468   MIA:    <<lachend> HE,//>
469   ERZ1:   ((macht Kreisbewegung der Windräder mit linker
              Hand) <<p> ^DREH:en die sich wenn der wind
              kommt;//>)
470   MIA:    hehehe
471   KEN:    [((zeigt wieder auf Bild))]
```

[2] Aus Datenschutzgründen kann hier nur die Tonaufnahme zugänglich gemacht werden.

Ken zeigt auf das Bild von den drei Windrädern (Z. 450) und die anderen beiden Kinder wenden ihren Blick hin zu diesem Bild (Z. 451). Die Erzieherin reagiert prompt auf Kens Impuls:

- Sie blickt Ken an.
- Sie signalisiert ein Erstaunen (mit hörbarem Einatmen): *°hh.*
- Sie fragt ihn mit markierter Fragemelodie (steigend): *und was ist DA:S?//* (Z. 452).

Ken reagiert nonverbal, indem er wieder auf die Windräder zeigt. Die Erzieherin ermuntert ihn nun liebevoll lächelnd im Flüsterton: *SAG_s mal;//* (Z. 454). Ken antwortet jedoch nicht. Dafür reagieren die anderen beiden Kinder:

- Ron beugt sich vor, zeigt auf das Bild und ruft sehr laut und skandierend: *OH (das ist)* (Z. 455).
- Mia zeigt ebenfalls auf das Bild und greift den akzentuierten Sinnkern der Frage der Erzieherin aus Zeile 452 auf: *DA:S?//* (Z. 456).
- Auch Ron wiederholt die Frage der Erzieherin: *(was_s DAS;//)* (Z. 457).

Die Erzieherin gibt vor, es nicht zu wissen: *WEIß ich ni was IST_n das;//* (schnell, wie „nebenbei" gesprochen; Z. 458) und wiederholt die Frage noch einmal (Z. 459). Ihre „Verzögerungsstrategie" ist erfolgreich, denn parallel zur Fragewiederholung antwortet Ron flüsternd: *[(… RÄder;//)]* (Z. 460).

Die Erzieherin bestätigt Rons Antwort, indem sie sie prosodisch verstärkt wiederholt (behauchter Stimmklang, steigend-fallende, dann fallend-steigende Melodiebewegung mit großem Umfang, Lautdehnung,): *`RÄ:↑^de:r `RICH´tig,//* (Z. 461f.). Danach spezifiziert sie die Räder, indem sie ein Fachwort einführt, das zweijährige Kinder vermutlich noch nicht kennen, und verwendet dafür einen staunenden Flüsterton mit mehrfachen Lautdehnungen: *WIN:Drä:de:r;//* (Z. 462).

Nachdem die Kinder erneut auf das Bild gezeigt haben (Z. 463f.) und also klar ist, dass die Erzieherin ihre Aufmerksamkeit besitzt, erläutert sie die Funktion der Windräder:

- Zuerst erzählt sie, was die Windräder machen: *und die WINDräder die !^DREH::!en sich;/* (Z. 465) – sie wiederholt das Fachwort noch einmal in akzentuierter Form und malt das Verb *drehen* stimmlich-artikulatorisch geradezu aus (starke Lautdehnung, auffällige Melodie, sehr starke Akzentuierung).
- Dann erklärt sie, wodurch die Windräder sich drehen: *da kann man;/ da muss !^GA:NZ! viel WIND kommen und DANN* (Z. 466f.) – dabei malt sie wiederum stimmlich-artikulatorisch die Steigerung *ganz* aus (ebenfalls Lautdehnung, auffällige Melodie und Akzentuierung).
- Schließlich bringt sie beide Äußerungen in einen Zusammenhang: *^DREH:en die sich wenn der wind kommt;//* (Z. 469) – dabei hebt sie prosodisch wieder das Verb *drehen* hervor. Außerdem unterstützt sie diese Äußerung gestisch, indem sie mit der linken Hand die Kreisbewegung der Windräder nachahmt (vgl. Bild 1_1).

Die Kinder zeigen auf das Bild von den Windrädern, Mia kommentiert die Äußerung der Erzieherin außerdem mit einem lauten Lachen (Z. 468, 470).

Nun schlägt die Erzieherin eine gemeinsame Aktion vor:

```
472   ERZ1:   [<<all> wollen wir mal ^WIND machen> wollen wir
              mal `´WIND machen?//      ]
473           °h wir PUSten mal.//
474           <<flüsternd> ACHtung es geht mal los wir PUSten
              mal.//>
475           ((guckt nacheinander jedes Kind an))
476           °hh((pustet 2s auf die Windräder im Buch, schaut
              abwechselnd auf die Windräder und die Kinder))
477           <<flüsternd> mit GANZ viel WIND;//>
478           nochMA:L-/
479   ALLE:   ((pusten 4s auf die Windräder im Buch))
```

Die erste Frage (Z. 469) formuliert sie sehr rasch, in der Wiederholung spricht sie eindringlicher (langsamer und mit deutlich markierter Fragemelodie). Sie fordert zum Pusten auf:

- Sie holt tief Luft, um das Pusten vorzubereiten: *°h* (Z. 473).
- Die folgende verbale Aufforderung flüstert sie, gestaltet sie also stimmlich auffällig („geheimnnisvoll-spannend"): *wir PUSten mal.// ACHtung es geht mal los wir PUSten mal.//* (Z. 473f.).
- Darüber hinaus animiert sie die Kinder auch nonverbal zum Mitmachen, indem sie während ihrer Äußerung alle nacheinander auffordernd anschaut (Z. 475).

Dann pustet sie selbst ausgiebig (2s lang) auf die Windräder im Buch und schaut dabei abwechselnd auf die Windräder und die Kinder (Z. 476). Die Kinder sind vollkommen auf das Geschehen konzentriert:

- Mia schaut ihr beim Pusten zu.
- Ron schaut gebannt auf die Windräder im Buch.
- Ken blickt zunächst auf den Mund der Erzieherin, lächelt und schaut dann ebenfalls die Windräder an.

Die Erzieherin fordert die Kinder noch einmal zum Pusten auf: *mit GANZ viel WIND;// nochMA:L-/* (Z. 477f.), und nun pusten die Kinder gemeinsam mit ihr etwa 4 s lang auf die Windräder im Buch (Z. 479; vgl. Bild 1_2).

Dieser kurze Ausschnitt enthält viele für das Gesamtkorpus typische stimmlich-artikulatorische und körperliche Ausdrucksformen mit erkennbarer Funktion. Generell kennzeichnend für die Gespräche mit den Zweijährigen ist neben häufigem Blick- und Körperkontakt ein expressiver stimmlich-artikulatorischer Ausdruck im Sinne einer malenden Prozedur (Ehlich 1991: 138f., Redder 1994: 240): ein sehr freundliches Muster mit überwiegend hellem Stimmklang, starke Hervorhebung von Akzentsilben, legatohaft-melodisches Sprechen mit großen Interval-

len, starke und häufige Lautdehnungen, sehr geringes Sprechtempo. Dieses Muster hat Ähnlichkeit mit dem elterlichen Motherese.

Die Erzieherin rahmt in dieser Situation mittels stimmlich-artikulatorischer Expressivität die Bilderbuchbetrachtung als besondere Situation und fokussiert die Aufmerksamkeit der Kinder:

- Zu Beginn der hier detailliert betrachteten Sequenz reagiert sie auf einen kindlichen Impuls (Ken hatte auf die Windräder gezeigt), indem sie durch den Flüsterton und das hörbare Einatmen ein Erstaunen demonstriert. Dadurch erzeugt sie eine Spannung und richtet die Aufmerksamkeit der Kinder auf bestimmte Bilder, so dass sich eine Geschichte zu diesen Bildern entwickeln kann.
- Mittels Flüsterton, behauchtem und lächelnd-freundlichem Stimmklang ermuntert sie die Kinder, sich am Gespräch zu beteiligen.
- Zwar gelingt es der Erzieherin im betrachteten Ausschnitt trotz mehrfacher freundlicher Aufforderung nicht, den sprachentwicklungsverzögerten Ken zu einer verbalen Äußerung zu ermuntern, aber sie erreicht es, dass er durchgängig auf das Gespräch konzentriert ist und nonverbal kohärent agiert (umblättern, zeigen, pusten).

Mit expressivem stimmlich-artikulatorischem Ausdruck und übertreibenden prosodischen Grundmustern hebt die Erzieherin inhaltstragende Wörter hervor. Auf diese Weise verbindet sie sprachlich Markantes oder Relevantes mit prosodischer Markierung. Sie verwendet häufig melodisch-dynamisch-temporal besonders auffällige Akzente, ausgeprägte Melodieverläufe mit großen Intervallen und starke Lautdehnungen, um

- kindliche Impulse expressiv zu bestätigen,
- Sprachäußerungen der Kinder zu wiederholen (mitunter in stimmlich-artikulatorischer Verstärkung des kindlichen Musters),
- kindliche Impulse weiterzuführen und in einen möglichen Erzählkontext zu stellen,
- eigene Sprachäußerungen zu wiederholen (ebenfalls oft stimmlich-artikulatorisch verstärkt),
- Wörter einzuführen, die für die Kinder vermutlich neu sind (hier: lexikalische Spezifizierung von Rädern zu Windrädern und Benennung ihrer Tätigkeit: sich drehen).

Diese Strategie der Erzieherin ist erfolgreich, die Kinder werden zu eigenen Produktionen angeregt: Mia und Ron wiederholen mehrfach die expressiv hervorgehobenen inhaltstragenden Wörter mit ähnlichem stimmlich-artikulatorischem Ausdruck. Doch nicht nur stimmlich-artikulatorisch agiert die Erzieherin in diesem Ausschnitt sensitiv-responsiv, auch nonverbal macht sie ein den Kindern altersangemessenes Angebot: Die Kinder verstehen zwar die Erläuterungen der Erzieherin, sind selbst aber noch nicht in der Lage, sich verbal über Windräder und deren Tätigkeit und Funktion zu äußern. Um die Kinder dennoch am Geschehen

aktiv zu beteiligen, regt die Erzieherin an, die abgebildeten Windräder zu animieren, und illustriert diese Anregung para- und nonverbal:

- Mit einer drehenden Handbewegung ahmt sie das Drehen der Windräder nach.
- Mittels Pusten erzeugt sie Wind, um die Windräder zum Drehen zu bringen.

Pusten können die Kinder, und sie übernehmen diese Animation begeistert – das Angebot der Erzieherin zur aktiven Beteiligung ist erfolgreich und es entsteht Gemeinsamkeit. Die Erzieherin macht den Kindern durchgängig nonverbale Angebote, sich am Geschehen zu beteiligen:

- Sie blickt die Kinder häufig freundlich-ermunternd an.
- In anderen Passagen dieses Gesprächs berührt sie die Kinder, wenn sie mit ihnen spricht (z.B. streichelt über ihren Kopf und Rücken und nimmt sie in den Arm).

4.2 Beispiel 2 (Erzieherin 1 mit Vierjährigen): *dann !´DREHN! die sich und dann kann man !^DA:!von !^STRO:M! bekommen.//*

Auch die vierjährigen Kinder sitzen mit derselben Erzieherin auf der Matte (rechts Paul, dann die Erzieherin, links neben ihr Sandra und links außen Kati; vgl. Standbilder 2_1 und 2_2; Audiobeispiel 2; Gesamtdauer ca. 23 Minuten). Kati kann manche Laute noch nicht korrekt verwenden (z.B. Unsicherheiten bei /k/ und /t/). Sie beteiligt sich sprachlich seltener als die anderen beiden Kinder am Gespräch, verfolgt aber das Geschehen sehr aufmerksam.

Eine Analyse der Anfangsphase des Gespräches (0:00-4:30) findet sich in Bose/ Kurtenbach/ Nixdorf (2013): Auch vor dieser Bilderbuchbesprechung findet über etwa drei Minuten ein themenreiches, ausgedehntes Vorgespräch statt, in dem die Erzieherin begeistert auf das neue Buch hinweist und die bevorstehende Bilderbuchbetrachtung aufwändig zelebriert, also auch hier erwartungsvolle Spannung und Vorfreude erzeugt.

Der hier diskutierte Ausschnitt beginnt bei Minute 10:16. Die Erzieherin hat das Buch auf dem Schoß und hält es mit der rechten Hand. Auch Sandra hält es fest. Alle gucken ins Buch und betrachten die zweite Doppel-Innenseite. Unmittelbar vor diesem Ausschnitt hatte Paul anhand eines Hochstand-Bildes rechts oben vom Jäger und dessen Berufsaufgaben erzählt. Die Erzieherin hatte Fragen dazu gestellt und Paul für seine Äußerungen gelobt. Nun gibt sie das Rederecht explizit an Kati. Das Mädchen hatte schon wiederholt angesetzt, über die im Buch abgebildeten Windräder (es nennt sie *Mühlen*) zu sprechen.

Bild 2_1: 10:35

Bild 2_2: 10:38

(2) Strom (PO_T1_S_4_10:16-11:11)

```
ERZ1:   SO.//
        und JETZT hören wir mal der !^KA!ti zu;/
        =die kommt [<<all,p> gar ni zu WORT,//>]
KATI:              [      das sind das         ] sin
        MÜhlen.//
ERZ1:   <<hauch,len> ge[<<p>^NAU:;//>> ]
KATI:                  [°h der Opa hat ] AUCH solche
        mü(.) MÜHlen.//
ERZ1:   =solche !^WIN:D!mühlen sind das./ NE:,//
        wenn der !´WIN:D!(-)kommt dann-/
        ((macht Kreisbewegung mit Hand über Windräder-
        Bild) was MACHen die denn dann;/)
        [        `´DANN?//                 ]
ALLE:   [((gucken auf das Windräder-Bild)) ]
PAUL:   ((imitiert Drehbewegung mit Arm über seinem Kopf)
        (.) dann [!DREH!en die sich.//)]
SANDRA:          [ DREHen die sich.//  ]
KATI:            [ DREHen die sich.//  ]
ERZ1:   <<len> ge!^NAU::!.//>
        dann !´DREHN! die sich und dann kann man
        !^DA:!von !^STRO:M! bekommen.//
        (-) no:r,//
        die(.)[GEben uns dann !^STRO::M! ab.//]
SANDRA:       [ aber WENN aber wenn SONne SO  ]
        wenn SONne is dann KOMM se NE.//
ERZ1:   <<len> DANN ((Drehbewegung mit linker Hand)
        !^DREH:N!) die sich ja nich NOR,//
        die WARten auf den WIN:D;/ das sind
        !^WIN:D!räder.//>
PAUL:   !^WIN:D!räder;/
        [ge^NAU: und- ]
ERZ1:   [ge^NAU./     ] das sind die ^WINDräder.//
```

Die Erzieherin erteilt Kati das Rederecht und begründet das: *SO.// und JETZT hören wir mal der !KA!ti zu;/ =die kommt [gar ni zu WORT,//]* (Z. 600ff.). Gegen Ende spricht sie schneller und leiser, während Kati bereits anfängt zu reden. Kati benennt zuerst das Bild von den Windrädern: *[das sind das] sin MÜHlen.//* (Z. 603). Die Erzieherin korrigiert Katis ungenaue Bezeichnung (Mühlen statt Windräder) zunächst nicht, sondern bestätigt die Feststellung mit einer Bestätigungsformel in einem prosodisch markierten Muster (langsam, behaucht, steigend-leicht fallender Melodieverlauf in großem Intervall, Lautdehnung): *ge^NAU:;//* (Z. 604). Auch hier nimmt sie ihre Lautstärke zum Ende zurück. Währenddessen hat Kati nach hörbarer Einatmung eine weitere Äußerung begonnen. Kati stellt eine Verbindung zwischen dem Bild im Buch und ihrer Familie her: *[°h der Opa hat] AUCH solche mü(.) MÜHlen.//* (Z. 605) und schaut dabei vom Buch weg in den Raum.

Die Erzieherin geht nicht auf Katis Erwähnung des *Opas* und seiner *Mühlen* ein, sondern spezifiziert nun doch die *Mühlen* zu *Windmühlen* mit stimmlich-artikulatorischer Markierung (steigend-fallende Melodie in großem Intervall, starke Lautdehnung, sehr starke Akzentuierung) – damit hebt sie die Spezifizierung besonders hervor: *=solche !^WIN:D!mühlen sind das./ NE:,//* (Z. 606). Dann fokussiert sie die Windräder im Buch, indem sie nach deren Funktion fragt: *wenn der !´WIN:D! (-) kommt dann-/ was MACHen die denn dann;/* (Z. 607ff.).

- Wiederum malt sie das Wort Wind stimmlich-artikulatorisch aus, wenngleich in geringerem Ausmaß als im Gespräch mit den Zweijährigen (Melodieanstieg, Lautdehnung, danach mittellange Pause) und hebt es dadurch erneut hervor.
- Während der Frage deutet sie über dem Bild mit der linken Hand die Kreisbewegung der Windräder an und lenkt damit den Blick der Kinder auf dieses Bild (Z. 609f.).
- Sie endet mit einer unvollständigen Äußerung: *`´DANN?//* (Z. 609) und fordert die Kinder dadurch zur Vervollständigung auf.

Paul übernimmt die Drehbewegung in ausgeprägterer Form (er benutzt den ganzen linken Arm dafür und vollführt größere Kreisbewegungen) und vervollständigt die Äußerung der Erzieherin: *(.) dann [!DREH!en die sich.//]* (Z. 611); ebenso reagieren Sandra und Kati *[DREHen die sich.//]* (Z. 612f.).

Die Erzieherin bestätigt die chorische Vervollständigung der Kinder wiederum sprachlich knapp, aber stimmlich-artikulatorisch ausgeprägt (langsam, steigend-(leicht) fallender Melodieverlauf mit großem Intervall, Lautdehnung): *ge^NAU:;//* (Z. 614). Anschließend wiederholt sie die Äußerung der Kinder und führt sie inhaltlich fort, indem sie den Nutzen der Windräder thematisiert: *dann !´DREHN! die sich und dann kann man !^DA:!von !^STRO:M! bekommen.//* (Z. 615ff.). Die inhaltstragenden Wörter *drehen, davon* und *Strom* hebt sie stimmlich-artikulatorisch besonders hervor (temporal, dynamisch, melodisch). Durch eine mittellange Pause gibt die Erzieherin den Kindern Gelegenheit, die Sprecherrolle zu übernehmen (Z. 616). Als das nicht passiert, fügt sie eine sächsisch-regiolektale Vergewisserungspartikel an: *(-) no:r,//* (Z. 616), wartet dann aber nicht weiter ab, sondern paraphrasiert ihre eigene Äußerung über den Nutzen der Windräder: *die (.) [GEben uns dann !^STRO::M! ab.//]* (Z. 617) und hebt das

Wort *Strom* wiederum stimmlich-artikulatorisch auffällig hervor (melodisch, dynamisch, temporal).

Währenddessen formuliert Sandra eine Einschränkung für die Nutzung der Windkraft mit einer Wenn-Dann-Konstruktion: *[aber WENN aber wenn SONne SO] wenn SONne is dann KOMM se NE.//* (Z. 618). Obwohl diese Einschränkung inhaltlich nicht richtig ist (es ist ja nicht die Sonne, die das Drehen der Windräder verhindert, sondern der fehlende Wind), korrigiert die Erzieherin Sandras Argument zunächst nicht:

- Stattdessen paraphrasiert sie lediglich sehr freundlich die Folge: *DANN !^DREH:N! die sich ja nich* (Z. 620) und endet Bestätigung heischend mit der Partikel: *NOR,//* (Z. 621).
- Wiederum hebt die Erzieherin die Wörter *dann, drehen* und *nor* stimmlich-artikulatorisch hervor (melodisch, dynamisch, temporal).
- Während des Wortes drehen deutet sie mit der linken Hand erneut eine Drehbewegung der Windräder an.

Ohne Pause schließt die Erzieherin nun eine inhaltlich zutreffendere Erklärung dafür an, dass die Windräder sich nicht drehen: *die WARten auf den WIN:D;/* und verwendet dann zum ersten Mal das korrekte Fachwort: *das sind !^WIN:D!räder.//* (Z. 621):

- Durch die Position am Ende der Sprecheinheit wird es zum besonders stark akzentuierten Sinnkern.
- Diese Akzentsilbe (wie auch die der anderen inhaltstragenden Wörter Wind, warten) macht die Erzieherin stimmlich-artikulatorisch besonders auffällig (melodisch, dynamisch, temporal).

Paul übernimmt von der Erzieherin das Fachwort identisch in Formulierung und stimmlich-artikulatorischem Ausdruck; außerdem schließt er die Bestätigungsformel an, die die Erzieherin auch oft verwendet: *!^WIN:D!räder;/ [ge^NAU: und-]* (Z. 622f.). Chorisch überlappend wiederholt die Erzieherin: *[ge^NAU./]* und führt fort: *das sind die ^WINDräder.//* (Z. 624).

```
625   SANDRA: UND und die ´SOMmerräder die sin wo`ANders.//
626   ERZ1:   <<lachend> und die [s=SOMmerräder<<f>sind
                                                 woANders.//>]
627   PAUL:                       [((stupst ERZ an)Aber(.)aber
                                                        aber) ]
628   ERZ1:   ((sich zu Sandra beugend) und die haha
              ´SOMmerräder DIE müssen wir bald auf die ^AUtos
              DRAUFbaun WA:,//)
629           [geNAU,/ ]
630   PAUL:   [geNAU.//]
631   ERZ1:   <<lachend> das ist WIEder was ANderes.//>
632            ((zu KF)haha
```

Sandra bringt einen neuen Aspekt ins Gespräch ein – sie weist auf den Kontrast zwischen *Wind*rädern und *Sommer*rädern hin: *UND und die ´SOMmerräder die sin wo`ANders.//* (Z. 625). Dieser Kontrast erscheint thematisch ziemlich unvermittelt – möglicherweise kommt Sandra durch die Klangähnlichkeit der ersten Silben der Wörter *Wind* und *Winter* auf diese Idee. Bei der Akzentuierung der Wörter *Sommer* und *woanders* übernimmt auch sie wie zuvor schon Paul das auffällige stimmlich-artikulatorische Muster, das die Erzieherin mehrfach für die Markierung von Akzentsilben verwendet hatte.

Die Erzieherin ist offensichtlich von diesem Einfall überrascht, sie muss lachen. Dann bestätigt sie Sandras Äußerung, indem sie sie wiederholt: *[und die s=SOMmerräder sind woANders.//]* (Z. 626). Währenddessen stupst Paul die Erzieherin an und versucht, eine Widerspruchsäußerung anzubringen: *[(Aber (.) aber aber)]* (Z. 627). Die Erzieherin geht nicht auf Pauls Widerspruch ein, sondern besteht weiterhin auf ihrem Rederecht, indem sie die Lautstärke ihrer Äußerung in Z. 626 steigert, und wendet sich zu Sandra. Sie führt den Gedanken um die Sommerräder lachend weiter aus, indem sie einen passenden Kontext entwirft: *und die haha ´SOMmerräder DIE müssen wir bald auf die ^AUtos DRAUFbaun WA:,//* (Z. 628). Sie fügt die von ihr oft gebrauchte Bestätigungsformel *[geNAU,/]* an (Z. 629), gleichzeitig gebraucht auch Paul erneut diese Formel *[geNAU.//]* (Z. 630). Nun spricht die Erzieherin lachend die vor der Gruppe sitzende Kamerafrau an: *das ist WIEder was ANderes.// haha* (Z. 631f.). Durch den Blickkontakt zur Kamerafrau wechselt die Erzieherin kurzzeitig den Gesprächsrahmen und stellt damit gleichsam nonverbal ein Einverständnis zwischen den beiden Erwachsenen über die unerwartete und ungewöhnliche Idee von Sandra her.

Auch dieser kurze Ausschnitt enthält viele für das Gesamtkorpus typische stimmlich-artikulatorische und körperliche Ausdrucksformen mit erkennbarer Funktion. Einige Formen sind denen sehr ähnlich, die bereits für das Gespräch mit den Zweijährigen herausgestellt wurden:

- Die Erzieherin bestätigt die kindlichen Äußerungen nach wie vor – meist sprachlich knapp (mit Bestätigungsformeln; vgl. z.B. genau in Z. 604, 614, 624, 629), aber stimmlich-artikulatorisch sehr ausgeprägt – dadurch drückt sie ihre Wertschätzung aus, ohne die Kinder, die gerade ein Thema entwickeln, unterbrechen zu müssen.
  - Auch hier verwendet die Erzieherin (malende) stimmlich-artikulatorisch ausgeprägte Muster und prosodische Übertreibungen mit unterschiedlicher Wirkung und Funktion (vgl. z.B. Z. 606-607: *solche !^WIN:D!mühlen sind das./ NE:,// wenn der !´WIN:D! (-) kommt dann-/* oder Z. 615-617: *dann !´DREHN! die sich und dann kann man !^DA:!von !^STRO:M! bekommen.// (-) no:r,// die (.) GEben uns dann !^STRO::M! ab.//):*
  - Sie fokussiert auf das Geschehen im Buch.
  - Sie animiert Gegenstände (hier: das Drehen der Windräder).
  - Sie spezifiziert kindliche und eigene Äußerungen (vgl. z.B. *MÜHlen* von Kati aus Z. 603 und 605 zu *!^WIN:Dmühlen* in Z. 606).
  - Sie hebt Argumente und Kontrastierungen sowie vermutlich neue Begriffe hervor, indem sie sie am Ende von Äußerungen positioniert, so dass sie

den Sinnkern-Akzent erhalten und damit prosodisch auffallen (vgl. z.B. *!^STRO:M!* in Z. 615 bzw. *!^STRO::M!* in Z. 617).

- Sie erreicht, dass die Kinder neue Wörter wiederholen, und fördert auf diese Weise den Erwerb (vgl. z.B. Paul in Z. 622: *!^WIN:D!räder;/*).

- Sie setzt den Blickkontakt sehr intensiv ein: Sie wendet sich immer demjenigen Kind zu, mit dem sie spricht, und schaut es freundlich-auffordernd an. Oft erreicht sie damit, dass dieses Kind zu sprechen anfängt.
- Die Erzieherin setzt an turnübergangsrelevanten Stellen Pausen, um den Kindern Gelegenheit zu geben, die Sprecherrolle zu übernehmen (vgl. z.B. Z. 616).
  - Kinder und Erzieherin übernehmen gegenseitig stimmlich-artikulatorische und nonverbale Ausdrucksformen sowie sprachliche Formulierungen und agieren konzertierend (vgl. Schwitalla 1993), aber häufiger und in ausgeprägterer Form, als es im Gespräch mit den Zweijährigen der Fall war (vgl. z.B. alle Kinder gleichzeitig in Z. 611-614: *DREHen die sich.//* oder Erzieherin und Paul gleichzeitig in Z. 629-630: *geNAU.//*). Im Gespräch mit den Zweijährigen war es vor allem die Erzieherin, die die Äußerungen der Kinder stimmlich-prosodisch aufgegriffen hat (vgl. Beispiel 1, Z. 461f.).

Andere Ausdrucksformen dagegen finden sich nur im Gespräch mit den Vierjährigen, wie die folgenden Beobachtungen zeigen:

- Die prosodischen Phrasen sind länger – sowohl diejenigen der Erzieherin als auch die der Kinder. Das Sprechtempo ist deutlich höher, wenngleich immer noch ziemlich langsam.
- Melodisch-klanglich „malt“ die Erzieherin nicht so stark wie im Gespräch mit den Zweijährigen:
  - Die Intervalle sind deutlich kleiner.
  - Akzentsilben werden stärker dynamisch markiert.
  - Es fehlt der aufmerksamkeitsheischende Flüsterton.

Dennoch sind die Kinder über die gesamte Situation hinweg sehr konzentriert und bringen eigene Ideen ins Gespräch ein.

- Sobald die Kinder anfangen zu sprechen, nimmt die Erzieherin häufig ihre Äußerungen zugunsten der kindlichen Gesprächsimpulse zurück (sie bricht ab oder wird leiser; vgl. z.B. Z. 602 und 604 zugunsten von Kati).
- Gelegentlich wird die Erzieherin aber auch lauter, wenn die Kinder anfangen zu sprechen. Sie sichert sich dadurch das Rederecht, vor allem dann, wenn sie bestimmte Impulse eines Kindes zugunsten anderer Impulse fokussiert:
  - Als Paul einen Widerspruch vorbringt (*[Aber(.)aber aber)];* Z. 627), besteht sie (lauter werdend) auf ihrem Rederecht: *und die [s=SOMmerräder<<f>sind woANders.//>]* (Z. 626) und geht nicht auf Paul ein. Es könnte der Eindruck entstehen, als ob sie unsensibel mit ihm

umgeht. Wenn man aber das Gesamtgespräch berücksichtigt, dann zeigt sich, dass ihre Aufmerksamkeit auch über längere Passagen mehrfach auf Paul gerichtet ist, dass sie seine Impulse aufgreift und fortführt (z.B. unmittelbar vor dem hier präsentierten Ausschnitt).

- Diese Strategie durchzieht das gesamte Gespräch: Um die Aufmerksamkeit ganz intensiv auf ein bestimmtes Kind zu richten, ignoriert die Erzieherin (vorübergehend) die Impulse der anderen oder fördert sie zumindest nicht so wie die des gerade besonders beachteten Kindes.
- Das gilt insbesondere dann, wenn es der Erzieherin gelungen ist, ein Kind zum Sprechen zu ermuntern, das aufgrund von verzögerter Sprachentwicklung generell sehr wenig spricht.

Im Vergleich zu den zweijährigen Kindern beteiligen sich die Vierjähigen sehr viel intensiver verbal-sprachlich am Gespräch und bringen eigene Argumente oder Widerspruch vor:

- Das Thema Windräder wird hier ausgiebiger und deutlich differenzierter diskutiert.
  - Im Gespräch mit den Zweijährigen erwähnt die Erzieherin lediglich, dass Windräder sich bei Wind drehen, und erzeugt diesen Wind gemeinsam mit den Kindern nonverbal durch Pusten.
  - Mit den Vierjährigen spricht sie dagegen über mehrere Turns hinweg auch über die Funktion der Windräder, Strom zu erzeugen. Die Kinder bringen sich mit eigenen Ideen ein (kein Drehen bei Sonne; Winterräder – Sommerräder), die Erzieherin greift diese Impulse auf und führt sie thematisch fort.

Beide Ausschnitte der Erzieherin 1 zeigen, dass sie nicht nur stimmlich-artikulatorisch sensitiv-responsiv agiert, sondern ebenso sprachlich und nichtsprachlich. Durch dieses kindbezogene Agieren gelingt es dieser Erzieherin, einen Spannungsbogen aufzubauen, die Aufmerksamkeit der Kinder darauf zu fokussieren und sie zu Impulsen für den Fortgang eines Themas anzuregen. Dieses Beispiel zeigt, dass auch in Kindertagesstätten gelingen kann, was Wieler (1997) für gelingendes Vorlesen im familialen Umfeld beschreibt: ein steter Wechsel zwischen Äußerungen zum Bilderbuch und daraus folgenden oder darüber hinausgehenden Reflexionen. Durch die Gestaltung einer kooperativen Atmosphäre regt die Erzieherin die Kinder zu solchen Reflexionen an. Beständig greift sie kindliche Initiativen auf und ermuntert mit dieser offenen, feinfühligen Haltung die Kinder zu kreativen Beiträgen.

4.3 Beispiel 3 (Erzieherin 2 mit Vierjährigen): *ein ↑↑ROL`ler.// … ein MÄDchen auf dem ↑↑ROL`ler.// … WIEder ein ↑ROL`ler;// … ein CI`ty`roller./*

Das abschließend vorgestellte kurze Beispiel stammt aus einer Detailanalyse eines größeren Ausschnittes dieses Gesprächs in Bose/Kurtenbach/Nixdorf (2013: 86 ff.). Hier wird nur auf einzelne Aspekte eingegangen, die zeigen, dass es dieser Erzieherin weit weniger gelingt, sich sensitiv-responsiv auf die Impulse und die

kommunikativen Fähigkeiten dieser Kinder einzustellen. Sowohl stimmlich-artikulatorisch als auch sprachlich agiert diese Erzieherin (ERZ2) vergleichsweise starr-schablonenhaft; das Gespräch geht über das Benennen der abgebildeten Gegenstände kaum hinaus.

Auch ERZ2 sitzt mit drei Vierjährigen auf der Matte (rechts außen im gelben Pullover Mandy, dann die Erzieherin, links neben ihr im grauen Shirt Sina und links außen Melanie; vgl. Standbilder 3_1-2, Gesamtdauer ca. 7 Minuten).

Bild 3_1: 01.42

Bild 3_2: 06.25

(3) Roller (PO_T2_P_4_01.42-02.33)

```
80   ERZ2:   ↑↑!GUCK!,/
81           ein ↑↑MÜLL`au`to.//
82           da ↑↑WAR `ges`tern eins beim ↑↑KIN`der`garten das
             hat ↑AUCH `den `müll abgeholt das sah GANZ
             ge↑↑NAU`so `aus.//
83   SINA:   und da [        (waren) dann ZWEI/       ]
84   MANDY:         [<<ruf> GUCK mal ein ´STRU`:MPF//>]
85   SINA:   ((macht große Gesten mit beiden Armen)
             hier [!GUCK! (...)//)         ]
86                [ein STRUMPF. //         ]
87   ERZ2:        [<<h> ein> STRU`::MPF.// ]
88            hier ist ein [↑STRU:`MPF.//]
89                         [ DA          ]
90            << f> waren (ZWEIne/) hier(.)hier VORne,//>
91   ERZ2:    <<bestätigend> hä::,//>
92   SINA:   (... und STEHN und ä hängen,/)
93   ERZ2:   [ <<bestätigend> hä=ä:.//>]
94   SINA:   [und WENN und HIER//      ]
95   MANDY:  ((wischt mit der Hand über das Bild vom Haus)
             DIS hier mach_ma bloß;/
96           DIS hier mach_ma bloß.//)
97   ERZ2:   <<sehr hoch> DIS hier mach_ma blo:ß;/
98           no,/>
99           <<all> mir KÖNnen ja mal HIER.//
```

```
100          WAS hattest du ge↑SEHN `me`lanie.//>
101   MELA:  [<<pp> DA:S HIER.//> ((zeigt auf Rollerbild))]
102   MANDY: [hier WARte mal.//                           ]
103   ERZ2:  [ ein   ] ↑↑ROL`ler.//
104   MANDY: [hier.//] ((blättert Seite um))
105   ERZ2:  ein MÄDchen auf [dem ↑↑ROL`ler.//]
106   MANDY:                 [DAS hier.//     ]
107          mÜSS_ma [erst.// ]
108   SINA:          [WIEder_n] MÜLLeimer.//
109   ERZ2:  schon ´WIE`:der ein `MÜLL:[`auto:.//  ]
110   MANDY:                          [DAS müss_ma] hier erst
             ma MACHen.//
111   ERZ2:  ja[WOLL.//]
112   SINA:    [een    ] ROLler wieder,//
113   ERZ2:  WIEder ein ↑ROL`ler;//
114          ge↑↑NAU`:./
115          ein CI`ty`roller./
116          [NE,//]
117   SINA:  [SCHON] WIEder des hier./
118          GUCK;//
119          WIEder die (...)//
120   ERZ2:  [schon ↑↑WIE`der  ]
121   MELA:  [schon WIEder.//  ]
122   ERZ2:  die FRAU.//
123          und WIEder ein ↑ROL`ler;/
124          ge↑↑NAU;//
```

Diese Erzieherin verwendet fast durchgängig ein bestimmtes Melodie- und Akzentmuster: Auf den Akzentsilben macht sie auffallend hohe Tonhöhensprünge und wechselt dabei ins Kopfregister, in den Nachakzentsilben fällt die Melodie stufenweise ab. Es entsteht ein gleichförmiger Sprechgesang, vgl. zum Beispiel Z. 103: *[ein] ↑↑ROL`ler.//*; Z. 105: *ein MÄDchen auf [dem ↑↑ROL`ler.//]* oder Z. 113: *WIEder ein ↑ROL`ler;//*. Außerdem spricht sie fast durchgängig hell und hoch bis sehr hoch (bis eine Oktave über ihrer physiologischen Sprechstimmlage, die sie z.B. in Z. 97-98 verwendet). Die Stimme klingt zwar freundlich, aber aufgrund der stark überhöhten Stimmlage sehr künstlich und angestrengt. Diese stimmlich-artikulatorische Gestaltung der Erzieherin weist Merkmale auf, die auch für das Motherese typisch sind (vgl. u.a. Bose 2003). Eltern benutzen dieses Muster, wenn sie freundlich-fürsorglich mit ihren kleinen Kindern sprechen. Aber im Gegensatz zu dem von den Eltern intuitiv sensibel und flexibel gebrauchten Motherese wirkt die Verwendung hier schematisch und starr und dem Alter der Kinder nicht angemessen (unempathisch, unauthentisch).

Diese Starrheit zeigt sich auch auf der sprachlichen Ebene: Obwohl die Erzieherin zu Beginn des Ausschnitts selbst einen Bezug zwischen dem Bilderbuchgeschehen und der gemeinsamen Lebenswelt im Kindergarten herstellt, indem sie auf das Müllauto verweist, das gestern im Kindergarten den Müll abgeholt hat und genauso aussah wie das im Buch (vgl. Z. 82), verfolgt sie diesen Einfall nicht wei-

ter und etabliert ihn nicht als Gesprächsthema. Stattdessen greift sie im Wesentlichen nur diejenigen Äußerungen der Kinder auf, in denen sie Einzelbilder benennen. Sie wiederholt die Benennungen und führt sie gelegentlich knapp inhaltlich fort. Andere (komplexe und steuernde) Impulse der Kinder ignoriert sie.

So wischt Mandy in Z. 95-96 mit einer Hand über das Bild vom großen Haus und will offensichtlich das Gespräch auf das große Haus lenken: *DIS hier mach_ma bloß;/ DIS hier mach_ma bloß.//*. Die Erzieherin wiederholt zwar die Äußerung des Kindes und stimmt formal zu (Z. 97-99), aber sie ignoriert dennoch seinen Impuls. Denn sie wendet sich mit einer neuen Frage an ein anderes Kind: *WAS hattest du ge↑SEHN `me`lanie.//* (Z. 100). Melanie ist nicht mehr in das Gespräch eingebunden, sondern dreht sich immer wieder vom Buch weg, sie sitzt ganz außen und sieht offensichtlich wenig von den Abbildungen im Buch. Womöglich versucht die Erzieherin durch ihre Frage, das Mädchen in die Gesprächssituation zu integrieren.

Melanie zeigt der Aufforderung folgend wohl auf das Bild eines Mädchens mit einem Roller: *DA:S HIER.//* (Z. 101; im Video nicht zu erkennen). Die Erzieherin greift diese Äußerung auf und benennt das Bild selbst*: ein ↑↑ROL`ler.// ... ein MÄDchen auf dem ↑↑ROL`ler.//* (Z. 103, 105).

Mandy hat inzwischen die Seite umgeblättert und insistiert – sie versucht noch einmal, ihren Impuls zur Chronologie des Betrachtens durchzusetzen: *DAS hier.// mÜSSn_ma erst.//* (Z. 106-107). Sina hingegen führt mit einer Bildbenennung wieder auf das Müll-Thema zurück: *WIEder _n MÜLLeimer.//* (Z. 108).

Die Erzieherin reagiert zwar nicht auf Mandys Versuch, einen inhaltlichen Fokus und eine Chronologie der Bildbetrachtung durchzusetzen (Z. 106-107), aber sie greift den Impuls von Sina aus Z. 108 insofern auf, als sie korrigierend wiederholt: *schon ´WIE`:der ein `MÜLL:`auto:.//* (Z. 109). Sie führt diesen Impuls aber nicht im Sinne einer Geschichte weiter. Mandy versucht wiederum, ihren Impuls zur Chronologie des Betrachtens durchzusetzen: *DAS müss_ma hier erst ma MACHen.//* (Z. 110). Die Erzieherin stimmt wiederum formal zu: *jaWOLL.//* (Z. 111), aber sie setzt Mandys Vorschlag nicht um. Stattdessen geht sie noch einmal auf die Benennung des Rollers durch Sina ein (Z. 112: *een ROLler wieder.//*). Sie wiederholt die kindliche Äußerung bestätigend und spezifizierend: *WIEder ein ↑ROL`ler;// ge↑↑NAU:./ ein CI`ty`roller./ NE,//* (Z. 113-116).

Insgesamt wird das Gespräch in diesem Ausschnitt bestimmt vom Benennen der Bilder durch die Kinder und der bloßen, meist stimmlich-artikulatorisch imitierenden Wiederholung oder sehr knappen lexikalisch-syntaktischen Erweiterung durch die Erzieherin, ohne dass sie diese Benennungen in eine kohärente Beschreibung oder Erzählung überführen oder auf den szenischen Zusammenhang im Buch hinweisen würde. Es gelingt dieser Erzieherin also nicht, einen gemeinsamen Gesprächsgegenstand zu fokussieren. Alle Beteiligten sprechen zwar häufig überlappend und gleichzeitig, aber über verschiedene Gegenstände, also nicht im Sinne eines thematisch fokussierten konzertierenden Sprechens, das für konsensuelle gesellige Gespräche üblich ist (vgl. stellvertretend Schwitalla 1993, Bose 2003) und auch in den Beispielen 1 und 2 zu finden war.

Die Kinder, die das Buch offensichtlich kennen, weisen die Erzieherin mehrfach auf den Erzählstrang einer Geschichte hin. Das zeigten bereits die wiederholten vergeblichen Versuche von Mandy zur Einhaltung einer Chronologie der Bildbetrachtung (Z. 95-97, Z. 106-107; weitere Belege in Bose et al. 2013). Aber

die Erzieherin greift diese Impulse nicht auf, sondern reagiert nur dann, wenn die Kinder auf ein bestimmtes Bild zeigen und es benennen. Sie wiederholt diese Benennungen bestätigend, aber sie verwendet überwiegend den unbestimmten Artikel. Es scheint so, als verstünde sie nicht, dass es sich um die ständig wiederkehrenden Protagonisten des Buches handelt, die über die Buchseiten hinweg Verschiedenes erleben bzw. mit denen Verschiedenes passiert.

In der Folge verfallen die vierjährigen Kinder immer stärker in den Benenn-Modus, oft in Form von Einwort-Äußerungen. Damit jedoch sind sie erheblich unterfordert und verlieren die Lust am Gespräch. Sie blättern viel schneller durch das Buch als in anderen Aufnahmen, dabei überblättern sie auch Seiten. Sie benennen anscheinend wahllos Bilder, die ihnen beim flüchtigen Draufblicken ins Auge fallen, und weisen nicht mehr auf einen Zusammenhang zwischen den Bildern hin (vgl. Bose et al. 2013: 90 ff.). Mandy bringt ihr Unbehagen einige Minuten später schließlich explizit sprachlich und nichtsprachlich zum Ausdruck (Z. 371, 373):

```
371   MANDY:  <<nörgelnd> wann sin_mer denn FERtig.//>
372   ERZ2:   MAGST nich mehr SCHAUen.//
373   MANDY:  ((schüttelt den Kopf))
```

Kurz nach Z. 373 wird das Betrachten des Bilderbuchs vorzeitig (mitten im Buch) abgebrochen; damit ist diese Aufnahme die kürzeste im gesamten Korpus.

Auch in diesem dritten Beispiel ist ein Zusammenhang zwischen den sprachlichen und parasprachlichen Verhaltensweisen der Erzieherin zu beobachten. Die Unflexibilität der Erzieherin zeigt sich sprachlich wie auch stimmlich-artikulatorisch und nichtsprachlich: Das fast durchgängig verwendete starre Melodie- und Akzentmuster ist gepaart mit eindimensionalen sprachlichen Reaktionen auf die zahlreichen kindlichen Impulse. Offensichtlich ist die Erzieherin hier, im Kontrast zu den Beispielen 1 und 2, mit der Situation überfordert, am Interesse und an den kommunikativen Fähigkeiten der Kinder geht sie vorbei.

Gesprächshemmendes Verhalten einer Erzieherin in so drastischer Ausprägung wie im Beispiel 3 kommt im Korpus nur selten vor. Allerdings sind ähnliche Verhaltensweisen in abgeschwächter Form auch in den Aufnahmen von anderen Erzieherinnen zu finden. Wie die Beispiele 1 und 2 zeigen, kann ein sensitiv-responsiver Umgang mit dem stimmlich-artikulatorischen Ausdruck in der Kita zwar gelingen, aber offensichtlich ist er nicht selbstverständlich.

5. Schluss

Aus den Analysen des Gesamtkorpus wie aus den hier ausgewählten Ausschnitten geht hervor, dass stimmlich-artikulatorischer Ausdruck und sprachliches sowie nichtsprachliches Agieren eng miteinander verknüpft sind:

- Erzieherinnen, die mit dem stimmlich-artikulatorischen Ausdruck eher starr und schematisch umgehen, agieren meist auch sprachlich und nichtsprachlich unflexibel. Sie fragen zum Beispiel kaum, sondern wiederholen nur die Benennungen der Kinder (oft in einem schablonenhaft gebrauchten Motherese).

Andere stellen dagegen viele Fragen (sog. Quizfragen), lassen sich dabei aber nicht vom Interesse und den Initiativen der Kinder leiten, sondern nutzen die Fragen eher zur Überprüfung kindlichen Wissens. Die Unflexibilität kann sich auch darin zeigen, dass Erzieherinnen selbst viel reden, sich dabei mehrfach wiederholen (inhaltlich oft nicht weiterführend) und dadurch verhindern, dass die Kinder zu Wort kommen können. In allen Fällen orientieren sich die Erzieherinnen dabei zu wenig an den Impulsen der Kinder.

- Erzieherinnen, die mit dem stimmlich-artikulatorischen Ausdruck dagegen sensitiv-responsiv umgehen, agieren meist ebenso mit Sprache und Körperausdruck. Es gelingt ihnen, einen Spannungsbogen aufzubauen, die Aufmerksamkeit der Kinder darauf zu fokussieren und sie zu Impulsen für den Fortgang einer Geschichte anzuregen. Dabei agieren sie sprachlich sehr variantenreich, sie erweitern die Äußerungen der Kinder, sie stellen Fragen, die sich am aktuellen Interesse und an den Erfahrungen der Kinder orientieren, sie führen die Kinder zu neuen Themen, motivieren zum Reflektieren. Vor allem lassen sie sich auf kindliche Initiativen ein und führen diese fort, gerade auch dann, wenn sie mit dem eigentlichen Buchinhalt (anscheinend) nichts zu tun haben. Auf diese Weise fordern sie die Kinder heraus, eigene Ideen zu entwickeln und miteinander darüber zu kommunizieren.
- Die Erzieherinnen agieren dabei oft mit sehr expressiven stimmlich-artikulatorischen Ausdrucksgestaltungen, die auf eine externe Beobachterin übertrieben wirken können. Die Analysen (vgl. Bose et al. 2013) zeigen, dass gerade solche Gestaltungen – punktuell, flexibel verwendet und auf die konkrete Gesprächssituation bezogen – ganz besonders die Aufmerksamkeit der Kinder wecken und sie vor allem zu eigenen Produktionen anregen (vgl. hierzu auch Braun 2007). Unsere Untersuchungen konnten zeigen, dass für ein gelingendes Bilderbuch-Gespräch nicht nur verbal-sprachliche Aspekte der Erzieherin eine Rolle spielen, sondern ein flexibler, variantenreicher, situations- und partnerbezogener Umgang mit dem stimmlich-artikulatorischem Ausdruck ebenso bedeutungsvoll ist.

6. Literatur

Abschlussbericht des Landesmodellprojekts „Sprache fördern“ – Erprobung und Multiplikation von Methoden der Sprachförderung in Kindertagesstätten (2011). Berufsbildungswerk Leipzig für Hör- und Sprachgeschädigte gGmbH. <http://www.bbw-leipzig.de/fileadmin/sprache_foerdern/Abschlussbericht/Abschlussbericht_e-book.pdf> (15.2.2013).

Albers, Timm (2009): Sprache und Interaktion im Kindergarten. Eine quantitativ-qualitative Analyse der sprachlichen und kommunikativen Kompetenzen von drei- bis sechsjährigen Kindern. Bad Heilbrunn: Julius Klinkhardt.

Albers, Timm (2011): Sag mal! Krippe, Kindergarten und Familie: Sprachförderung im Alltag. Weinheim, Basel: Beltz.

Beller, Kuno/ Beller, Simone/ Mertens, Hans/ Preissing, Christa (2007): Abschlussbericht des Projekts. Erzieherqualifikation zur Erhöhung des sprachli-

chen Anregungsniveaus in Tageseinrichtungen für Kinder. Eine Interventionsstudie. Berlin. <http://www.beller-und-beller.de/ESIA-Abschlussbericht-05-2007-2.pdf> (14.02.2011)

Berner, Rotraut Susanne (2008): Frühlingswimmelbuch. Hildesheim: Gerstenberg.

Bose, Ines (2003): *dóch da sín ja ' nur mûster //* - Kindlicher Sprechausdruck im sozialen Rollenspiel. (Hallesche Schriften zur Sprechwissenschaft und Phonetik 9). Frankfurt a. M. u.a.: Peter Lang.

Bose, Ines (2010): Stimmlich-artikulatorischer Ausdruck und Sprache. In: Deppermann, Arnulf/ Linke, Angelika (Hrsg.): Sprache intermedial: Stimme und Schrift, Bild und Ton. (= Jahrbuch des IDS 2009). Berlin / New York: de Gryuter, 29-68.

Bose, Ines/ Bößhenz, Sylvia/ Pietschmann, Judith/ Rothe, Ingmar (2012)*:"°hh hh° also von KUNdenfreundlich halt ich da nIcht viel bei ihnen;"* – Analyse und Optimierung von Callcenterkommunikation am Beispiel von telefonischen Reklamationsgesprächen. In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion (ISSN 1617-1837). Ausgabe 13, 143-195 (www.gespraechsforschung-ozs.de).

Bose, Ines/ Kurtenbach, Stephanie/ Nixdorf, Sophie (2013): Formen und Funktionen des Sprechausdrucks in Gesprächen zwischen Erzieherinnen und Kindern. In: Kurtenbach, Stephanie/ Bose, Ines (Hrsg.): Gespräche zwischen Erzieherinnen und Kindern. Beobachtung, Analyse, Förderung. (=HSSP 47). Frankfurt a. M. u.a.: Peter Lang, 67-100.

Braun, Barbara (2007): Gemeinsam ein Bilderbuch lesen – Vermitteln und Aneignen in der Kommunikation von Mutter und Kind. In: Meng, Katharina/ Rehbein, Jochen (Hrsg.): Kindliche Kommunikation – einsprachig und mehrsprachig. Münster u.a.: Waxmann, 127-154.

Bredel, Ursula/ Ehlich, Konrad/ Falk, Simone/ Guckelsberger, Susanne/ Kemp, Robert F./ Komor, Anna/ Reich, Hans H./ Trautmann, Caroline (2008): Desiderate der Forschung zur kindlichen Sprachaneignung. In: Ehlich, Konrad/ Bredel, Ursula/ Reich, Hans H. (Hrsg.): Referenzrahmen zur altersspezifischen Sprachaneignung. (= Bildungsforschung, Band 29/II; hgg. vom BMBF). Bonn, Berlin, 255-270.

Bruner, Jerome S. (1987): Wie das Kind sprechen lernt. Bern: Huber Verlag.

Deppermann, Arnulf (2004): ‚Gesprächskompetenz' – Probleme und Herausforderungen eines möglichen Begriffs. In: Becker-Mrotzek, Michael/ Brünner, Gisela (Hrsg.): Analyse und Vermittlung von Gesprächskompetenz. Frankfurt a. M. u.a.: Peter Lang, 15-27.

Ehlich, Konrad (1991): Funktional-pragmatische Kommunikationsanalyse. Ziele und Verfahren. In: Flader, Dieter (Hrsg.): Verbale Interaktion: Studien zur Empirie und Methodologie der Pragmatik. Stuttgart: Metzler, 127-143.

Grimm, Hannelore (2003): Sprachscreening für das Vorschulalter. (unter Mitarbeit von M. Aktas und U. Kießig). SSV. Göttingen: Hogrefe.

Konopcynski, Gabrielle/ Tessier, Sophie (1994): Structuration intonative du langage emergent. In: Halford, B. K./ Pilch, H. (eds.): Intonation. Tübingen: Narr, 157 - 192.

Kurtenbach, Stephanie (2011): Einblick in die Qualifizierungsmaßnahme des Landesmodellprojekts Sprache fördern, Sachsen: Methoden der Sprachförderung. In: Sievert, Ulrike/ Voigt-Zimmermann, Susanne (Hrsg.): Klinische Sprechwissenschaft. Aktuelle Beiträge aus Wissenschaft, Forschung und Praxis. (HSSP 35). Frankfurt a. M. u.a.: Peter Lang, 127-133.

Kurtenbach, Stephanie/ Bose, Ines (Hrsg.) (2013): Gespräche zwischen Erzieherinnen und Kindern. Beobachtung, Analyse, Förderung. (= HSSP 47). Frankfurt a. M. u.a.: Peter Lang.

Kurtenbach, Stephanie/ Bose, Ines/ Thieme, Tabitha (2013): Gemeinsam ein Bilderbuch anschauen. Untersuchung zum Gesprächsverhalten von Erzieherinnen. In: Kurtenbach, Stephanie/ Bose, Ines (Hrsg.) (2013): Gespräche zwischen Erzieherinnen und Kindern. Beobachtung, Analyse, Förderung. (= HSSP 47). Frankfurt a. M. u.a.: Peter Lang, 23-50.

Lepschy, Annette (2002): Lehr- und Lernmethoden zur Entwicklung von Gesprächsfähigkeit. In: Brünner, Gisela/ Fiehler, Reinhard/ Kindt, Walther (Hrsg.): Angewandte Diskursforschung. Bd. 2: Methoden und Arbeitsbereiche. Radolfzell: Verlag für Gesprächsforschung, 50-71.

Meer, Dorothee/ Spiegel, Carmen (Hrsg.) (2009): Kommunikationstrainings im Beruf. Erfahrungen mit gesprächsanalytisch fundierten Fortbildungskonzepten. Radolfzell: Verlag für Gesprächsforschung.

Merkel, Johannes (2005): Warum das Pferd von hinten aufzäumen? Grundsätze zur Sprachförderung im Elementarbereich, insbesondere von Kindern mit anderer Muttersprache. In: Textor, Martin R. (Hrsg.): Kindergartenpädagogik. <http://www.kindergartenpaedagogik.de/1296.html> (09.08.2012)

Nentwing-Gesemann, Iris/ Fröhlich-Gildhoff, Klaus/ Harms, Henriette/ Richter, Sandra (2011): Professionelle Haltung – Identität der Fachkraft für die Arbeit mit Kindern in den ersten drei Lebensjahren. Expertise des Deutschen Jugendinstitutes im Rahmen der WiFF-Weiterbildungsinitiative Frühpädagogische Fachkräfte. Frankfurt a. M.: Heinrich Druck und Medien GmbH.

Papoušek, Mechthild (1994): Vom ersten Schrei zum ersten Wort: Anfänge der Sprachentwicklung in der vorsprachlichen Kommunikation. Bern: Huber.

Redder, Angelika (1994): ‚Bergungsunternehmen' – Prozeduren des Malfeldes beim Erzählen. In: Brünner, Gisela/ Graefen, Gabriele (Hrsg.): Texte und Diskurse. Methoden und Forschungsergebnisse der funktionalen Pragmatik. Opladen: Westdeutscher Verlag, 238-264.

Remsperger, Regina (2013): Das Konzept der Sensitiven Responsivität. Ein Ansatz zur Analyse des pädagogischen Antwortverhaltens in der ErzieherInnen-Kind-Interaktion. In: Frühe Bildung 1. Göttingen: Hogrefe, 12-19.

Ritterfeld, Ute/ Siegert, Susanne (2000): Die Bedeutung naiver Sprachlehrstrategien in Erwachsenen-Kind-Dyaden. In: Logos Interdisziplinär. 8. Jg./ 1. Köln, 37-43.

Schwitalla, Johannes (1993): Über einige Weisen des gemeinsamen Sprechens. Ein Beitrag zur Theorie der Beteiligungsrollen im Gespräch. In: Zeitschrift für Sprachwissenschaft 11/ 1, 68-98.

Selting, Margret et al. (Hrsg.) (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2): In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion. Ausgabe 10 (2009), 353-402. <www.gespraechsforschung-ozs.de> (25.05.2011).

Tarplee, Clare (1996): Working on young children's utterances: prosodic aspects of repetition during picture labelling. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation: Interactional studies. Cambridge: CUP: 406-435.

Tomasello, Michael (1999): The Cultural Origins of Human Cognition. Cambridge (Mass.)/ London: Harvard Univ. Press.

Wieler, Petra (1997): Vorlesen in der Familie: Fallstudien zur literarisch-kulturellen Sozialisation von Vierjährigen. Weinheim u. a.: Juventa.

7. Legende der Notationszeichen

Die Gesprächsausschnitte sind mit einer Kodierung versehen, die Angaben zu Ort, Erzieherin und Alter der Kinder enthält. Die Notate lehnen sich stark an das Feintranskript von GAT 2 (Selting et al. 2009) an, sie enthalten aber wenige zusätzliche Angaben zur Sprechgliederung.

Siglen

MARC, RON, KATI usw.	Kinder (Namen verfremdet)
ERZ	Erzieherin
KF	Kamerafrau

Sprechgliederung

Aussprüche: prosodisch abgeschlossene Sprecheinheiten; Teilaussprüche: prosodisch nicht nichtabgeschlossene Sprecheinheiten)

//	Ausspruchsgrenze
/	Teilausspruchsgrenze

Akzentuierung (im Vergleich zu GAT 2 vereinfacht)

NIEder	Akzentsilbe in Sprecheinheiten

Angaben zum stimmlich-artikulatorischen Ausdruck

<<erstaunt, bestätigend>>	Ausdrucksmuster
<<behaucht, flüsternd>>	Ausdrucksmerkmale

Prof. Dr. phil. habil. Ines Bose
Diplom-Sprechwissenschaftlerin

Dr. phil. Stephanie Kurtenbach
Klinische Sprechwissenschaftlerin (Dipl.)

Seminar für Sprechwissenschaft und Phonetik
Institut für Slavistik und Sprechwissenschaft
Philosophische Fakultät II
Martin-Luther-Universität Halle-Wittenberg
Advokatenweg 37
06114 Halle (Saale)
Germany
Ines.bose [at] sprechwiss.uni-halle.de
Stephanie.kurtenbach [at] sprechwiss.uni-halle.de

Prosodic, lexical and sequential cues for assessments with German *süß*: Assemblages for action and public commitment

Beatrice Szczepek Reed

Abstract
This paper shows how conversationalists assemble lexical, prosodic, and sequential cues to accomplish social actions and commitment to a stance. The assemblage [*süß* + prosodic stylization + freestanding turn design] is used as a practice to implement positive and highly affiliative assessments of child-like agents. Participants use the same form of assessment to mitigate complaints about adult referents, and to assign non-membership to referents from different membership categories. While the assemblage in its general form remains constant, the specifics of the prosodic stylization show a considerable degree of flexibility. Consequently, the paper argues that the interactionally most relevant aspect of prosodic stylization is that it is *notably different* from surrounding talk, rather than *how precisely* it is different. The analysis shows that social actions are not accomplished or contextualized by individual cues, but by assemblages, which emerge online and have to be defined with the necessary flexibility to fit the nature of spontaneous talk.

Assemblage - assessments - prosodic stylization - German - Interactional Linguistics

Dieser Artikel zeigt, wie GesprächsteilnehmerInnen lexikalische, prosodische und sequentielle Merkmale zusammenfügen, um soziale Handlungen auszuführen und sich einer Haltung zu verpflichten. Die *assemblage* [*süß* + prosodische Stilisierung + freistehender Turn] wird als Praktik verwendet, um positive und stark affiliierende Bewertungen kindlicher Referenzobjekte durchzuführen. TeilnehmerInnen gebrauchen die gleiche Form von Bewertung, um Beschwerden über erwachsene Referenzobjektezu entschärfen und um Referenzobjekten anderer *membership categories* Nicht-Mitgliedschaft zuzuweisen. Während die *assemblage* als solche in ihrer Form konstant bleibt, weisen die spezifischen Merkmale der prosodischen Stilisierung erhebliche Flexibilität auf. Somit argumentiert der Artikel, dass der interaktional relevanteste Aspekt von prosodischer Stilisierung eher ihre *beträchtliche Abweichung* vom sie umgebenden Gesprächsdesign ist als *die präzise Form* der Abweichung. Die Analyse zeigt, dass soziale Handlungen nicht durch einzelne Gesprächsmerkmale ausgeführt oder kontextualisiert werden, sondern durch *assemblages*, die in einem *online*-Prozess hervor- und zusammentreten, und die mit der für spontane Gesprächsdaten notwendigen Flexibilität definiert werden müssen.

Assemblage - Bewertungen - prosodische Stilisierung - Deutsch - Interaktionale Linguistik

1. Introduction

There is growing understanding in Interactional Linguistics and beyond that the linguistic behavior of a single speaker cannot be considered in isolation from other behavior, from other users of language, and from its local context of use. Human interaction relies on bundles of linguistic and non-linguistic cues for collaborative sense-making. This means that linguistic and non-linguistic domains, such as syntax, turn construction, prosody, lexis, and embodied behavior have to be considered together for the analysis of talk-in-interaction. This study shows how a cluster of turn constructional resources (lexical, prosodic, sequential) is systematically employed to implement a definable social action. The cluster of interactionally relevant features that participants assemble for that purpose is referred to as an 'assemblage', with loose reference to the term as it is used in cultural studies (Deleuze/ Guattari 1987, Marcus/ Saka 2006). In that field, as in Interactional Linguistics, the aspect of assembling as an emergent activity is highly significant. For an analysis of natural talk the term 'assemblage', rather than, for example, 'pattern', emphasizes participants' active moment-by-moment building of interactional practices from a variety of available component parts for the accomplishment of social actions, with changes in the assembled cues resulting in changes to the action (to-be-)accomplished.

We know that small changes to the delivery of turns, such as variations in word choice (e.g. Raymond/ Heritage 2006) or prosodic delivery (e.g. Local 1992) can have a substantial impact on the social work performed by participants through talk. However, we also know that natural interaction involves a considerable amount of redundancy (e.g. Kaimaki 2011). For example, participants can be observed to orient routinely and systematically to a variety of behavioral patterns for turn completion; among them cues from syntax, lexis, articulation, pitch, voice quality, loudness, syllable duration, tempo, torso movement, gesture, and gaze (cf. French/ Local 1983, Local/ Wells/ Sebba 1985, Local/ Kelly/ Wells 1986, Selting 1996, Wells/ Peppé 1996, Wells/ Macfarlane 1998, Local/ Walker 2004, Ogden 2004, Stivers/ Rossano 2012, Li 2013, 2014). Not all of these patterns are called upon on each occasion, and at times directly contrasting cues may be used at different turn completion points (Szczepek Reed 2004). With respect to sound patterns, previous work has also shown that in certain contexts speakers relate the prosodic design of their own contributions to that of other speakers. This is observably the case for turns in responding position, where participants frequently make use of prosodic matching (Szczepek Reed 2006, 2009a, b, 2010, 2012. See also Goldberg, 1978, Couper-Kuhlen 1996, this volume, Müller 1996, Tarplee 1996, Wells/ Corrin 2004, Wells 2010, Gorisch/ Wells/ Brown 2012) in order to present their talk as following on from a previous speaker's contribution (Szczepek Reed 2012). In these circumstances the main interactional impact of the prosodic delivery is its backward-orienting function, achieved by the repetition of the way a previous speaker delivered their talk prosodically. What counts most in such cases is that a participant is doing prosodically *the same* as a previous speaker, rather than *what exactly* they are doing. The analysis reported here, in contrast, shows a context in which the opposite is the case: what is important in the excerpts below is that the TCU in question is prosodically *different*, and in-

deed markedly so. The change itself, rather than the precise nature of how it is achieved, is what occasions the turn's main interactional function at a specific sequential position. In the case reported here, the issue at hand is the assemblage [*süß* + prosodic stylization + freestanding turn design], for example:

```
<<high, tense, held pitch>ganz SÜ::ß ->.
```

This assemblage is found to perform highly affiliative and at times mitigating assessments of third parties in first position.

In the following, methods and data are briefly introduced before analyses of three basic types of assessments with *süß* are presented. Subsequently, the extract which provided the first noticing for this study is analyzed in depth and finally a number of concluding observations are offered.

2. Methodology, data, and transcription

The methodology adopted here follows an interactional-linguistic approach, in that language is considered to be a collection of resources for social action and interaction (Selting/ Couper-Kuhlen 2001). The analysis has been conducted on recordings of six hours of naturally occurring German conversations around the dinner table. Each recorded interaction is roughly one hour long and involves two or three participants. The dinner host (Linda), a speaker of Swabian German, is present in all interactions as her apartment was the venue for the recordings. The participants have given their consent for the data to be used for research purposes in anonymized form; however, in two of the six recordings participants did not give permission to be video taped. As a result, those interactions are available as audio recordings only.

The recordings have been transcribed according to an adapted version of GAT2 (Couper-Kuhlen/ Barth-Weingarten 2011, see Appendix). An English translation, or at times a gloss, is provided for each line of transcript, although this should only be regarded as a rough guide to the original. The German word *süß* itself does not always have a straightforward English equivalent, except where it refers to taste as *sweet* (extracts (2) and (3)). At other times *cute* or *(aw)bless* may be more appropriate. For this reason, Betz/ Golato's (2008) practice of maintaining the original German word, rather than an attempt at translation, is followed here wherever the referent of *süß* is not food.

All prosodic and embodied information is provided in the German transcript line. Only one line of translation is provided, instead of the interlinear gloss required by GAT2 notations, for a more accessible transcript, especially in cases where a large amount of prosodic and embodied information is provided. Consider, for example:

```
2562   M:   <<all>und dann hAt er so_n>verBEULten BEsen,
            and then he's got this beat-up broom

2563        damit macht er dann SAU<<breathy>ber.>
            and that's what he cleans up with
```

As Ashmore/ Reed (2000) have noted, audio recordings and transcripts are 'analytical objects' rather than original data (see also Kelly/ Local 1989). Further, both the visual detail provided by the video recordings and the necessity to translate from another language introduce additional layers of complexity and additional opportunities for aspects of the original event to be missed. Therefore it is important to keep in mind Ten Have's (1997) observation that 'transcripts are unavoidably incomplete, selective renderings of the recordings'. All audio recordings are available from the author upon request, but can be shared for research purposes only. For a single example of the assemblage [*süß* + prosodic stylization + freestanding turn design], please see LIS1.

3. Analysis

This study began with a first noticing of *süß* in the following dinner table conversation, where it is used with highly remarkable prosody and facial expressions. The extract comes from a longer sequence, which will be considered again in more detail later. Linda has just told Julia that their mutual friend Mark is going traveling with an admirer'.

(1a) Linda and Julia 3, 11.31LJ3-LJ7

```
   529   J:   [<<breathy, frowning, holding gaze, torso
              forward>hat-] hat_n verEHRer?>
              (he) has has an admirer

   530   L:   <<smiling, nodding>m[_hm:::,>]

   531   J:                       [der ihn ] AN[wirbt?]
                                   who is romancing him

   532   L:                 <<smiling, nodding>[     m]_hm:::,>

-> 533        [<<tense, nodding, eyes closed, smiling>GAN:Z>
                                                       really

-> 534        <<high, tense, vibrato, eyes half closed, smiling,
              frowning SÜ]::::ß - >
```

The exaggerated prosodic and facial design of the assessment item (*ganz*) *süß* (line 533) involves tense phonation, high pitch register, held pitch with vibrato, vowel lengthening, (half-)closed eyes, and a smiling-frowning facial expression. This highly marked turn design seems to combine with the positively valenced lexical item *süß* to display strong commitment to high regard for the referent (Mark). However, the turn is a surprising choice for an adult referent, whom the German lexical items *'n Verehrer … der* (*an admirer…who,* masculine gender) reveals to be gay. It was therefore considered helpful to investigate other assessments with *süß* in order to understand the role of the lexical choice and the prosodically stylized turn design. In the following, three basic uses of *süß* will be presented.

3.1. Prosodically unmarked *süß*

In its basic sense, German *süß* can be used to assess the taste of food as *sweet*. In the 3 occurrences of this use in a collection of 28 cases, *süß* is embedded into a simple clause, such as a copula construction, rather than being freestanding; and occurs with unmarked prosodic or facial attributes. For example, in the following extract Barbara assesses the taste of an Indian sauce, which another participant had feared to be too hot.

(2) Linda, Barbara, Sophie, 14.35 LBS1

```
   923   L:   DAN[ke.]
                thank you

   924   B:      [die] IS nich SCHARF;
                 this isn't hot

-> 925          die is ↑SÜ:ß. (.)
                it is sweet

   926   L:   ECHT?
                really
```

Barbara's *süß* marks a contrast with the description of the sauce in the previous TCU (*die is nich scharf*, line 924), and *süß* therefore receives a degree of prominence typically found with such contrastive statements. The prominence is realised by means of syllable lengthening and a pitch step-up. Barbara's facial expression remains relaxed with the exception of the lip rounding and lip protrusion necessary for the articulation of [y;]. A similarly unmarked combination of cues occurs in the following assessment of a piece of dragon fruit that one of the participants is eating. Linda's articulation of *süß* again involves the necessary degree of lip rounding and protrusion, but otherwise shows relaxed facial expressions. Her prosodic design is appropriate for the turn-final location of *süß*, that is, it involves a degree of final lengthening and prominence typical for a focus accent.

(3) Linda and Julia Part 2, 00.25 LJ2

```
   32    J:   zum in_d GOSCH neischlage.
              makes you want to hit him

   33         (1.1)

-> 34    L:   ((giggles))

   35         <<pointing>dIE sind SÜ:ß.>
              these are sweet

   36         (.)

   37    J:   <<chewing>mhm,>

   38    L:   <<pointing>dIE> sind REIF.>
              these are ripe
```

Both instances of *süß* are first-position assessments of currently and interactionally available material objects (food on the table). In this use of *süß* the item is embedded into a simple syntactic construction. Regarding the prosodic and articulatory production of *süß*, both instances exhibit a degree of syllable lengthening that is consistent with their turn-final location and focus accent, and in (2), with contrastive stress.

Default, unmarked turn design can also be observed when *süß* is embedded into longer syntactic constructions that assess third parties, rather than immediately accessible tastes. Examples from the corpus include:

(4) Linda, Frauke, Jens, 22.05 LFJ3

```
F:   sie war so SÜß katrin;
```
she was so cute Katrin (was)

(5) Linda, Frauke, Jens, 17.55 LFJ2

```
F:   und die ham so SÜße KIDS auch;
```
and they've got such cute kids as well

(6) Linda and Mark, 19.32 LM4

```
L:   desch aber SÜß dass er_s WAHRnimmt.
```
but it's cute that he realises it

Of the 28 cases of *süß* in the corpus, 14 occur as part of a longer construction. 12 of these cases show a prosodic delivery that is unmarked except for what is appropriate for locally relevant accents, contrasts, and turn-internal positions. It appears that the extremely marked prosodic and visual design seen in extract (1a) is typically found in syntactically and sequentially freestanding position, as the following sections will show.

3.2. [*süß* + prosodic stylization + freestanding turn design]as a practice for affiliative assessments of 'little' referents

Süß can be used in a freestanding format similar to other German assessment tokens (*klasse, super*) in first responding position. In this format and location, *süß* may be prosodically unmarked; however, in the corpus for this study it is very often extremely marked (13 out of 14 cases). The following extract is a typical example. Here, Sibylle has been telling Ines and Linda about a map of Germany in her parents' road atlas on which she drew pencil lines as a teenager to measure the distance between her town and the hometown of a boy she had a crush on.

(7) Linda, Ines, Sibylle, 1.05.20 LIS1

```
1025   S:   ich hatte mich auf_m kirchentag in berlIn -
```
at the church congress in Berlin I had

```
   1026          in so_n TYP aus cuxHAven verknallt.
                 developed a crush on this guy from Cuxhaven

   1027    I:    mhm,


   1028    S:    da war ich FÜNFzehn er war schon SECHzehn.
                 I was fifteen he was already sixteen

   1029          [GANZ toll.]
                 really great

   1030    I:    [<<high>mm:::?>]

   1031          mhM,

   1032    S:    und da hab ich mir zuhause AUSgemessen wie WEIT das
                 is.=
                 and so at home I measured how far it is

   1033          =und lag meinen ELtern in [Ohrn;=
                 and kept on and on at my parents

   1034    I:                              [<<high, smiling>o:::?>]1
                                            aw

   1035    S:                                   =und jEtzt sind da]
                 so LInien immer [noch drin;
                 and now there are these lines still in there

-> 1036    I:                   [<<high, tense, smiling,
                                 portamento>O:::;
                                aw

-> 1037          SÜ:::ß;>]

   1038    S:    ECHT so ] WITZig;
                 really (it's) so funny

   1039          (.)

   1040    I:    KLASse.
                 excellent
```

Ines's response to Sibylle's telling is *o süß* [see Audio Szczep_01_O-süß_LIS1], which translates into English along the lines of *aw bless*. Both syllables are heavily lengthened, the speaker uses a high pitch register, and her phonation is tense throughout. The pitch contour for both items is a portamento movement, that is, a very gradually declining pitch contour across *o*, and again on *süß*. Although the recording is audio only, the articulation can clearly be heard to involve lip spreading, i.e. smiling. Such extreme prosodic marking, which goes considerably be-

[1] Please note the use of *o* rather than *oh* in the German transcript line to represent the German equivalent of *aw*, rather than German *oh*.

yond the routine prosody for turn taking, sequence organization, and contrastive emphasis, is referred to here as 'prosodic stylization' (Szczepek Reed 2006).[2]

Ines's turn is sequentially in first responding position, directly following the assessable (line 1035), and displays commitment to a highly affiliative stance. Earlier in the sequence, Ines has already received parts of Sibylle's telling with prosodically stylized continuers (lines 1030, 1034), thus aligning with a storytelling activity that makes affective involvement relevant. Following the *süß*-assessment, Ines continues to display affiliation with a positively valenced third-position assessment *klasse* (line 1040). This extract is the only one in the collection in which a second assessment is offered. All other cases of freestanding *süß* are the only assessment following the assessable turn and are typically sequence-closing. In the above instance, rights to assess (Heritage/ Raymond 2005) are being handled carefully by participants. As the assessable is Sibylle's younger self, her second position assessment is constrained by preferences regarding self-praise. She neither upgrades nor downgrades, but instead offers *echt so witzig* (*so funny,* line 1038), which conveys a non-committal stance.

The lexical content of *o süß*, the extreme prosodic marking, and the freestanding position combine to implement a distinctly positive assessment, showing strong public commitment to an affiliative stance and affective involvement (Goodwin/ Goodwin 1987). This is supported by the smiling delivery and the additional positive assessment in third position. The assessable is Sibylle's behavior as a teenager. What is being assessed with *o süß* is the 'sweet' and innocently romantic behavior of a young person. Later examples show that this intrinsic connection between a 'sweet' referent and the reference form 'sweet' cannot be taken for granted.

In the following extract a similar pattern can be observed. Eleven-year-old Sophie is telling Linda how she fed raccoons during a recent visit to the zoo.

(8) Linda, Barbara, Sophie, 28.02 LBS4, LBS5

```
   1657   S:   ICH hatte hier (.) WASCHbärfutter?
               I had the food for the raccoons here

   1658        <<acc>und DIE ham - >=
               and they

   1659        =mach mal DEIne hand so hin? (.)
               put your hand like this

   1660        °h und die haben mit IHRen händen,
               and they with their hands

   1661        (0.7)

   1662        äh des so RAUS genomm[en,                    ]>
               took it like this

-> 1663   L:                         [<<smiling>S:Ü[:::ß - >]
```

[2] See Szczepek Reed (2006) for an overview of the use of the term 'stylisation' in the literature, where it is often restricted to 'sing-song' intonation.

```
   1664   S:                                        [und HAM ]
                des dann geGESSen.
                and then they ate it

-> 1665   L:   <<high, loud, smiling>↑↓SÜ::ß.>

   1666   B:   wie ÄFF[che:n.
                like little monkeys

   1667   L:          [wO WAR denn des sophIE;
                        where was this Sophie
```

Once again, the assemblage [*süß* + prosodic stylization + freestanding turn design] implements a highly positive assessment of a sweet and innocent third party. This time the item is repeated after an initial placement in overlap with ongoing talk (lines 1663, 1665). As in the previous case, strong commitment to affiliation is apparent in the smiling delivery of both turns. However, while both instances of the assemblage involve notable prosodic stylization, the specifics of the prosodic design differ both from the previous example and from each other. While (4) shows tense phonation, both instances in (5) are produced with modal voice quality. There is also a difference in pitch contour, as neither instance in (5) shows the portamento contour seen in (4). Instead, line 1663 shows a held level pitch throughout the entire item. This format stands out from surrounding talk due to its lack of pitch movement, and due to the rarity of level pitch in turn-final position (Szczepek Reed 2004). In contrast, line 1665 shows a wide rising-falling movement. Finally, while (4) is delivered with high pitch register, only the second instance in (5) (line 1665) shows the same feature.

All three examples show extensive vowel lengthening, and also lip spreading. The latter may be related to an increase in vowel fronted-ness observable in (5) (Fagel 2010).[3] At line 1665, Linda's F_2 formant in [y;] has a mean value of 1999 Hz, which is considerably higher than in her articulation of unmarked *süß* at other times in the corpus, such as:

(9) Linda and Mark, 19.09 LM3

```
L:   dEs find ich schonmal SÜ:ß
     that I find rather cute
```

Mean F_2 for [y;]:1876 Hz

(10) Linda and Mark, 19.22 LM4

```
L:   desch aber SÜß dass er_s WAHRnimmt
     but it's cute that he realises it
```

Mean F2 for [y;]: 1856 Hz

[3] The formant responsible for perception of vowel fronted-ness is F_2. Where possible formant analyses have been carried out using Praat 5.3.82 (Available at http://www.fon.hum.uva.nl/praat/). However, a definite increase in F_2 can only be claimed for certain for Linda, for whom there is considerably more data available due to her presence in all recordings, and whose formant values can therefore be compared with other instances of *süß*. As formant frequencies depend very much on phonemic context and a speaker's physiology no definite claims regarding vowel quality will be made with regard to other speakers.

Fagel (2010) shows that lip spreading can have a raising effect on F_2 formants in certain vowels, including /i:/ (his study does not include /y:/). However, later cases (1b, lines 533, 536) show that smiling does not necessarily result in an increase in vowel fronted-ness in our data.

Once again, [*süß* + prosodic stylization + freestanding turn design] implements a first-position assessment, and this time no second assessments are offered by other speakers. Instead, Linda closes the storytelling sequence with *süß* before she goes on to initiate a request for information (line 1667). With regard to the referent of *süß* here, too, a 'little' agent co-occurs with the reference form 'sweet' in addition to the child-adult nature of the interaction itself.

The following extract shows a similar occurrence of *süß*, this time referencing (the giving of) a chocolate Easter bunny. Linda enquires of eleven-year-old Sophie whether she received Easter presents (a 'nest') from her grandmother.

(11) Linda, Barbara, Sophie, 24.26 LBS3

```
   1452   L:   GAB_s jetzt n nestle oder NICH.
               so was there a little nest or not

   1453   S:   ja ne so ne (.) kleine dOse mit SÜßichkeiten,
               yes a like a          small box with sweets

   1454        und so_n HA:se mit ZWANzich EUro hinten [drin. ]
               and like a bunny with twenty euros in its back

   1455   L:                                            [o:a::;]
                                                         wow

   1456        des_isch ja_n GOLDhase;
               that was a gold bunny

   1457        im WAHRschten sinne des WO:Rtes
               <<smiling>GWE:s[n::.>]
               quite literally

   1458   S:                  [ja.  ]
                               yes

-> 1459   L:   <<smiling>S::Ü::ß;>

   1460        und da hatse sich WAHNsinnich gfreut;
               and she was really happy

   1461        über dein selber gmachtes NESCHtle;=
               to get the little nest you made

   1462        =ODer?
               was she
```

Linda's prosodic delivery of *süß* involves considerable lengthening, this time both on the first and the second segment, [s] and [y;]. Once again there is an increase in vowel fronted-ness, as F_2 for [y;] has a mean value of 2074 Hz. The pitch contour is falling minimally, while other prosodic features remain unmarked (phonation, pitch register, loudness). As in the above extract the assemblage [*süß* + prosodic

stylization + freestanding turn design] accomplishes a positive and highly committed assessment of a (behavior involving) a 'little' referent in sequence-closing position. Affiliation is displayed by Linda's preceding turn (lines 1455-1457), where she receives the news of the Easter present with an extreme case formulation (*Goldhase,* line 1456) and the beginning of smiling delivery (line 1457), which continues into the *süß*-assessment. The choice of referent continues the list of 'little' things or agents with which this assemblage co-occurs.

A final example in this line of argument is the following one, in which a three year-old child is assessed as *süß*. Here the assemblage is incorporated into a longer turn, where it functions linguistically as an added adverbial phrase and sequentially as a freestanding assessment TCU. Linda is telling Michael that she needs to call her niece, who left a message on her answer phone.

(12) Linda and Michael, 1.00.27 LMi1

```
   3059   L:   wEIl die geht sonscht in_s BE:TT.=
                because she'll go to bed otherwise

   3060        =und die hat mir geschtern auf_sBAND geSPROchen ->
                and she left me message on the answerphone yesterday

-> 3061        <<high, tense, held pitch>ganz SÜ::ß ->
                really süß

   3062        und die WA:Rtet jetzt schon immer;
                and now she always waits

   3063        dass die PAtentante zurückrUft;
                for her godmother to call her back
```

Again the prosodic stylization involves both similarities and differences with previous cases. It includes vowel lengthening, high pitch register, tense voice quality and held pitch across the entire phrase *ganz süß*. The articulation of [y;] is further front than for Linda's unmarked items, with 1990 Hz as a mean F_2 value. What is substantially different here is that [*süß* + prosodic stylization + freestanding turn design] is not a first-position assessment by a different speaker, but is produced by the speaker who has produced the assessable (line 3060). This practice of assessing one's own talk is shown to be used as an interactional strategy in the following section. The stylized prosodic delivery hearably sets the assemblage apart from preceding and following talk and shows it to be designed as a freestanding assessment TCU.

The four extracts above have shown that an assemblage of lexical, sequential, and prosodic cues is used to perform highly committed affiliative first-position assessments of small or otherwise 'sweet' assessables. The assessments refer to children, small animals or animal-like objects, and most of them are also addressed at children. The assemblage has been represented as

[*süß* + prosodic stylization + freestanding turn design].

It is argued here, first, that prosodic markedness can be achieved by a range of prosodic cues, rather than a defined set of specific changes; and second, that prosodic markedness as such is a defining factor for the accomplishment of a highly committed affiliative assessment with *süß*. In order for the second argument to hold, it would also need to be shown that a modification of the assemblage to [*süß* + freestanding turn design] without prosodic stylization does not accomplish the same action. The corpus indeed contains one such case.

(13) Linda, Barbara, Sophie, 36.40 LBS6, LBS7

```
   2140   L:   du hasch ne SCHÖne (.) KETTe.
                you've got a lovely necklace

   2141        hasch die SELber gmacht?
                did you make it yourself

   2142        (0.7)

   2143   S:   hmʔm.

   2144        hat mAma mir aus itALien mitgebracht.
                mum brought it back from Italy for me

   2145   L:   die_sch TOLL.
                it's lovely

   2146        (2.6)

   2147   B:   von liMO:ne.
                from Limone

-> 2148   L:   [SÜß;]

   2149   B:   [ am ] LA:go die GARda.
                at Lake Garda

   2150        (1.1)

-> 2151   L:   SÜß sieht die aus.
                it looks cute

   2152        (3.4)

   2153        JA mädels;
                well girls
```

In this instance, the assessment accomplished by the modified assemblage implements a compliment paid to the recipient, rather than an assessment of a non-present third party. In this function, public commitment to affiliation is not made in relation to a stance previously expressed by the recipient, but in relation to locally available attributes and possessions of the recipient herself. The absence of prosodic stylization thus seems to result in an assemblage of cues that is employed for a different interactional function. While prosodic stylization thus appears to have a contrastive function in highly affiliative assessments with *süß*, its individ-

ual form seems less relevant. Regarding the specifics of the prosodic delivery of *süß* throughout the collection, only vowel lengthening can be observed in all cases. Other features, such as increased fronted-ness in vowel articulation, high pitch register, and a change in voice quality occur in some instances, but not others. With regard to pitch contour, there seem to be two main movements, portamento and level tone, with a number of exceptions. These observations suggest that there is a core set of features, vowel lengthening and possibly a held or portamento contour, which speakers combine with other prosodic features to design turns as notably separate from the surrounding run of talk.

In the following section the same assemblage is shown to be used as an interactional practice for referring to and mitigating the complainable behavior of adults.

3.3. [*süß* + prosodic stylization + freestanding turn design]as a practice for mitigating the behavior of adult referents

Out of the 13 cases of freestanding, prosodically stylized *süß*, 6 either refer to children, and/ or are addressed at them. The remaining 7 refer to, and are addressed at, adults. In the following it is argued that where the assemblage [*süß* + prosodic stylization + freestanding turn design] is used to assess adult referents, this is done to mitigate potential criticism of their complainable behavior. This is made possible because the use of the assemblage allows participants to treat adult referents interactionally as child-like and thus not fully responsible for their behavior. See, for example, the extract below, in which Linda and Mark are eating cake bought for them by Gaby, a non-present friend, who is currently visiting from out of town.

(14) Linda and Mark, 10.22 LM1

```
   593   L:   die sehn TOLL aus.
              they look amazing

   594   M:   JA.
              yes

   595        hm:: so RICHtich zum REIN[beißen.]
              mm really yummy

   596   L:                              [aber  L]EIder weiß gAby
              nimmer wo die konditoREI isch;=
              but unfortunately Gaby can't remember where the pastry shop is

-> 597        =<<creaky, quiet>SÜ::ß - >

   598        M:E[CHT?]
              really

   599   L:       [des ] ISCH halt so bei Ortsfremden.
                   that's what it's like with people who are not locals
```

```
600          die LATSCHen[halt R]AUS,
             they'll wander out

601    M:                [ja:.  ]
                          yes

602    L:    irgendwo HIN,
             somewhere


603          und finden wieder ZRÜCK;
             and find their way back

604          [aber] beSCHREIben können sie_s dir dann NICH.
             but they can't describe it to you

605    M:    [klar.]
             sure
```

Similar to (12), this assessment is produced by the same speaker who produces the assessable turn. The assemblage [*süß* + prosodic stylization + freestanding turn design] at line 597 is latched directly onto the previous TCU without a break and is designed with many of the features found in other cases. The vowel is lengthened, pitch is held level, and there is a change in voice quality, this time to creak. Vowel fronted-ness is again increased compared to Linda's default values with a mean F_2 for [y;] of 2039 Hz. This assemblage is also quieter than surrounding talk.

The referent of *süß* is of a different type than in earlier cases: while previous assessments with *süß* referred to 'little' things or agents and their 'sweet' behavior, this referent is an adult whose behavior is neither recognizably sweet nor innocent. Instead, the assessable is presented as a complaint, and the referent's behavior is made relevant as problematic: Gaby cannot remember where she bought the cakes. That Linda's turn is indeed designed as a complaint is observable from its initiation with *aber leider* (*but unfortunately*). It is this complainable behavior that is assessed with the positively valenced item *süß* in the directly following TCU.

By latching onto her complaint an affiliative and positive assessment, Linda is able to mitigate without delay any criticism inherent in her complaint. Further, by using an assessment assemblage that is routinely found with little and innocent agents who cannot be held responsible for their actions, Linda retrospectively contextualizes the referent and her behavior as minimally guilty of any complainable action. She thereby manages to avert any potential second position complaints by co-participants that may otherwise relevantly follow (Drew and Walker 2008, Ogden 2010). This analysis is confirmed by Linda's subsequent talk (lines 599-604), which maintains a defending stance towards the complainable issue.

Another example of this strategy is the following, in which Mark tells Linda about his new partner (cf. ex. (1a, b)), a senior academic, whose apartment he has recently visited for the first time. He is in the process of describing various features, such as the lack of furniture and his partner's apparent lack of interest in his surroundings.

(15) Linda and Mark, 46.34 LM5

```
   2562   M:   <<all>und dann hAt er so_n>verBEULten BEsen,
                and then he's got this beat-up broom

   2563        damit macht er dann SAU<<breathy>ber.>
                and that's what he cleans with

   2564        °hhhh he

   2565        [<<laughing>ach JA,              ]>
                oh well

-> 2566   L:   [<<tense, quiet, smiling>S:Ü::ß ->]

   2567   M:   stAUbsauger gibt_s AUCH net -
                there's no vacuum either

   2568   L:   ja=aber GUCK,
                yes but look

   2569        wEnigschtens MACHT er sAUber.
                at least he does clean up

   2570   M:   geNAU.
                exactly
```

As in the previous example, the assessable in this instance is neither little nor young. Instead, the assessed behavior is presented in a complaint about substandard cleaning equipment (line 2562-2563). The resulting implication of a lack of cleanliness is later mitigated by Linda (2568-2569). As in the previous case, the complaint is immediately followed by a relativizing item produced by the same speaker who makes the complaint (*ach ja*, line 2565). In overlap Linda offers her assessment with prosodically stylized, freestanding *süß*. The stylization involves sound and syllable lengthening, spread lips, tense phonation, held pitch and a decrease in loudness. Once again, the assemblage [*süß* + prosodic stylization + freestanding turn design] is used to mitigate potential criticism of an adult's behavior that intrinsically possesses none of the features denoted by the lexical item *süß*. By using an assemblage often found to assess small, and by implication innocent, referents, a not-so-innocent referent is treated with minimal attribution of responsibility. This once again has a sequential implication in that further complaints are averted (lines 2568-70).

3.4. [*süß* + prosodic stylization + freestanding turn design]as a practice for assigning non-membership

The above observations provide insight into the multi-layered referential possibilities afforded by freestanding, prosodically stylized *süß*. They also allow for a deeper understanding of the sequence that inspired this analysis (see ex. (1a)). In assessing adult referents and their behavior with the assemblage described above, the implication arises that those referents are not only treated as little, but that

their behavior is potentially complainable and must be mitigated. It is this combination of highly committed benevolence, as if to a child or to a friend who has misbehaved, directed at the entirely un-complainable behavior of two adults, that accounts for the notable effect of the sequence in question. The following transcript provides a detailed notation of the entire sequence, during which several *süß*-assessments are made. The sequence is part of a longer conversation in which Linda and Julia are discussing who to invite to a bicycle ride that Linda is trying to organize. Julia suggests asking Mark, one of Linda's best friends.

(1b) Linda und Julia, 11.20 LJ3-LJ7

```
516   J:   aber DEN kannst du [doch FRAgen;     ]
           but you could ask him couldn't you

517   L:                      [<<high>AH: NEE:;>]
                               ah no

518   J:   [((             ))]

519   L:   [der isch der geht] mit nem::;
           he is he's going with an

520        (0.2)((hand gesture, raised eyebrows, slow nod, gaze
           at J))

521        <<slow, smiling, holding gaze>verEH:Rer?>
           admirer

522        <<slow, smiling, holding gaze, head nod>nach
           itA:Lien.>
           to Italy

523   J:   ((hand gesture, raised eyebrows, slow nod, smiles,
           holds gaze with L))

524        <<low, creaky, smiling, holding gaze>NEI:::N;>
           no

525   L:   [<<smiling, holding gaze, nodding>m_hm::: ->]

526   J:   [<<smiling, holding gaze, leaning forward>nach
           it]Alien;>
           to Italy

527   L:   <<smiling, holding gaze, nodding>JA,
           yes

528        [weil       ] des_n italiener ISCH,>
           because this is an Italian

529   J:   [<<breathy, frowning, holding gaze, torso
           forward>hat-] hat_n verEHRer?>
           (he) has has an admirer
```

```
   530   L:   <<smiling, nodding>m[_hm:::,>]

   531   J:                        [der ihn ] AN[wirbt?]
                                    who is romancing him

   532   L:                  <<smiling, nodding>[     m]_hm:::,>

-> 533         [<<tense, nodding, eyes closed, smiling>GAN:Z>
                really
->             <<high, tense, vibrato, eyes half closed, smiling,
               frowning SÜ]::::ß - >
```

Still 1: Julia (recipient, left), Linda (speaker, right), line 533

```
   534   J:   [seit<<creaky, smiling, frowning>WANN,>]
               since when

   535   L:   °hh<<smiling>du des ging schon vor der KUR los
               anSCHEInend.>
               it started even before he went to the resort apparently

-> 536         <<high, tense, held pitch, eyes half closed,
               smiling, frowning>so:: Sü:::ß ->
               so süß

   537         auch bei so_m porTAL halt.
               through one of those portals again

   538         (1.0)

-> 539         <<giggling, extra high, breathy, eyes half closed,
               smiling, frowning>gan:z SÜ::ß ->
               really süß

   540         wenn_er von DEM [erZÄH:]LT ->
               when he talks about him

   541   J:                    [((  ))].
```

```
    542   L:   <<high, giggling, smiling, frowning>da wird er ganz
               WEICH so::;>
               he gets really soft like

    543        und GANZ,°hh
               and really

    544        (1.0)

    545        emoTIOna:l,
               emotional

    546        und es tut ihm einfach so [GUT.]
               and it's just so good for him

    547   J:         <<smiling, frowning>[och ][COO:]:L;>
                                           aw     cool

    548   L:                                 [und ] dessen
               BRUder,
               and his (that person's) brother

    549        hat ne FErienwohnung,
               has a holiday appartment

    550        irgendwo bei GEnuA:?
               somewhere near Genua

    551        ((L smiles and holds gaze with J))

    552   J:   <<becoming higher, tense, creaky, smiling,
               frowning>und dann wi- will er ihn MITneh::m,>
               and so he wants to take him

    553   L:   ((nods))

-> 554   J:   <<falsetto, loud, smiling, frowning>oh: is das
               SÜ::ß;>=
               aw that's so süß

    555        =<<high, loud, smiling>und die ham sich auch schon
               ge↑TROFfen und [sO::.>  ]
               and they've already met and so on

    556   L:        <<smiling>[JA↑↓::.>]
                    yes

    557   J:   ((voiced gasp, mouth wide open, raised eyebrows,
               held gaze))°hhh

    558        <<creaky, held gaze>och TOLL::;>
               aw great

    559   L:   <<smiling, held gaze>und der kleine italIEner,
               and the little Italian

    560        will immer schon gern MEHR?>
               already wants more
```

```
   561   J:   <<breathy, held gaze>oh ach↑SO::;
              oh I see

   562   L:   und ER sagt immer<<hand gesture, swaying torso
              movement>
              NEE,>
              and he keeps saying no

   563        so weit bin_ich NICH,=
              I'm not ready

   564        =und du mussch mir jetzt ZEIT lassn.
              And you need to give me time

   565        (0.2)((L smiles))

-> 566   J:   <<high, tense, loud, portamento, smiling,
              frowning>↑o ↑SÜ[:::ß;>   ]
              aw süß

   567   L:                [<<smiling with lips pressed together,
              frowning, nodding>m_hm::,]

   568        (2.2)((L and J hold gaze, frowning and smiling))

   569   L:   ((high shoulder shrug, giggles))
```

During this sequence Mark, his male partner, and their behavior as a couple are assessed with prosodically stylized *süß* on five occasions (lines 533, 536, 539, 554, 566). The assemblage at line 554 does not involve freestanding position; instead, it is produced in the form of an exclamatory formulaic phrase, in which the news receipt marker *oh* is followed by an inverted sentence structure found only with this type of strongly valenced expressions (*is das süß*).This instance is one of two cases where prosodically stylized *süß* is not in freestanding position; the other one is the similarly routine expression '*is ja süß*' by the same speaker in a different part of the same conversation. It may therefore be the case that formulaic phrases fill similar sequential slots (Raymond 2013) as do freestanding items.

The first three *süß*-assessments are done by Linda in response to her own talk. Interestingly, on none of these occasions (lines 533, 536, 539) is an actual assessable presented explicitly. Instead, what is being assessed remains implicit to a degree, with tacit reference to the news of Mark's 'admirer' and their travel plans (lines517-532).All three assessments display some of the features previously described for the assemblage, with the prosodic stylization including changes in pitch register and phonation, held pitch, and vowel lengthening. There are also additional features, including considerable vibrato on the vowel at line 533 and giggling at line 539. Interestingly, vowel frontedness only occurs in Linda's third use of the assemblage (line 539), where the mean F_2 value is 2032 Hz. The first two instances (lines 533, 536) show an even lower mean value than some of Linda's un-stylized uses of *süß*: [y;] in line 533 has a mean F_2 value of 1758 Hz, while line 536 has a mean value of 1726 Hz. Thus vowel fronted-ness, which seems to occur fairly frequently throughout the collection, cannot be considered a systematic feature, but an optional variation. Further, while vowel fronted-ness

may have been interpreted as resulting from lip spreading, these instances show that this is not necessarily the case. In addition to these prosodic features, Linda's assessment is accompanied by an exaggerated smiling-frowning facial expression with half-closed eyes, most strongly displayed in the first assessment at line 533 (Still 1).

Julia's first assessment with stylized *süß* (line 554) is offered directly in response to news of the couple's travel arrangements, while her second assessment (line 566) responds to the news that Mark's partner is more sexually interested than Mark himself, and Mark is asking for more 'time' (line 564). Both assessments are produced with strong prosodic stylization. The first one (line 554) includes a change from Julia's prior talk to falsetto voice, an increase in loudness as well as vowel lengthening, and the smiling-frowning facial expression seen in Linda's earlier uses of the assemblage. Line 566 involves the same facial expression, high pitch register, increased loudness, tense phonation, vowel lengthening, and two portamento movements on *o* and *süß*.

All five assessments are extreme variations of the already exaggerated designs seen in earlier examples. All of the assessables describe the behavior of an adult gay couple who have recently met. As extract (7) shows, romantic behavior may indeed be assessed with prosodically stylized, freestanding *süß*. However, while (7) refers to the behavior of a 15 year-old teenager, the referents in this extract are grown men, and their behavior (plans to travel together) less romantically innocent than that of young Sybille (drawing lines on a map).Nevertheless, it would be difficult to identify these assessments in themselves as necessarily related to the difference in membership category (in contrast to Mark and his partner, Linda and Julia are heterosexuals). No explicit references are made to homosexuality or gay stereotypes. Without such references the analyst is left to imagine whether the same exaggeratedly amused reaction would have been displayed if, for example, Linda had spoken of a female friend who told her boyfriend she needed more time before she became sexually intimate. However, the sequence does contain an explicit reference to a different membership category attributed to Mark's partner, who is not German. At line 559 Linda refers to him as 'the little Italian'. The use of 'little' underpins an analysis of *süß* at other places in the sequence as at least in part belittling of third parties who are not of the same membership category (nationality, sexual orientation).That this is done entirely without (displayed) malice can be seen, for example, in the frequent use of straightforwardly positive assessments, such as *und es tut ihm einfach so gut* ('and it's just so good for him', line 546), *och cool* ('aw cool', line 547), and *och toll* ('aw great', line 558).

While the assessments in this sequence display high, even exaggerated commitment to a positive stance towards their referent, the use of the assemblage resonates with its more frequent uses: its reference to children and other small agents, its occurrence in child-adult interactions, and its mitigating function assessing the complainable behavior of adult referents. It is these resonances that introduce a potentially belittling stance towards non-members of the same categories, even if the explicit talk itself displays high levels of commitment to friendship and social intimacy. The assemblage thus not only performs positive assessments of third party referents, but it assigns the referents to a non-

membership status by treating them interactionally as either child-like or deviant when in fact neither attribute applies.

4. Concluding observations

This paper has presented an analysis of a single assemblage of lexical, prosodic and sequential cues. An analytical decision was made not to approach the phenomenon as a word (*süß*) that is accompanied by certain prosodic or sequential features (stylization, freestanding position). Instead, the whole bundle of features was presented together as equally significant. In doing so the analysis committed itself to that bundle, but not to other uses of a) *süß,* b) prosodic stylization, or c) freestanding turn design. The term 'assemblage' was chosen to refer to such bundles because it carries the notion of an emergent, active assembling of cues, which is well-fitted to the analysis of spontaneous talk.

A claim was made that the precise nature of the prosodic stylization is not as interactionally relevant as the presence of stylization itself. This conclusion was reached after a detailed prosodic analysis of stylized, first-position assessments with *süß* showed only a single stable feature, which was vowel lengthening. With the exception of vowel lengthening, all other features showed wide variation, including unmarked patterns. Pitch contour, where considerable variation would be possible in theory, showed two main formats, level and portamento, with a number of exceptions. Table 1 shows prosodic variations of all 15 cases of stylized *süß*. 13 are freestanding, two are part of a formulaic construction (LJ6, LJ8).

Ex.	Duration	Voice quality	Contour	Pitch register	Loudness	Lips	F_2
LM1	vowel	creaky	level	default	quiet	default	2039Hz
LM2	sound	breathy	level	default	quiet	spread	N/A
LM5	sound + vowel	tense	level	default	quiet	spread	N/A
LJ1	sound + vowel	modal	level	default	default	rounded	1921Hz
LJ3	sound + vowel	tense	level	high	default	spread	1758Hz
LJ4	vowel	tense	level	high	default	spread	1726Hz
LJ5	vowel	breathy	level	high	default	spread	2032Hz
LJ6	vowel	falsetto	rising	high	loud	spread	N/A
LJ7	vowel	tense	portamento	high	loud	spread	N/A
LJ8	sound + vowel	modal	fall-to-mid	low	default	default	N/A
LMi1	vowel	modal	level	high	default	default	1990Hz
LIS1	vowel	tense	portamento	high	default	spread	N/A
LBS3	sound + vowel	modal	fall-to-mid	default	default	spread	2074Hz
LBS4	vowel	modal	level	default	default	spread	N/A
LBS5	vowel	modal	rising-falling	high	default	spread	1999Hz

Table 1: Prosodic variation of stylized *süß*

All items have in common that they implement the same category of social action: affiliative, positive assessments in first position. What defines them prosodically is that they stand out, often prominently, from the surrounding run of talk. They do not continue whatever prosodic design was used before, and the talk following them does not continue their prosodically stylized design. Their most fundamental characteristic is that they are different from what came before. Precisely how they are different, however, does not seem to be a major factor with regard to the social action and commitment they accomplish. Similar observations have been reported with respect to pitch by Kaimaki (2011), who finds that for non-valenced news receipts rising and falling pitch contours are in free variation. This is not to say that individual and specific changes in prosodic design are meaning- or function-less in specific interactional environments. There may indeed be links between, for example, certain prosodic designs and degrees of affective involvement (see Kupetz, this volume). However, with regard to the overall accomplishments of talk what matters most seems to be how a prosodic design relates to its prosodic surroundings. For the public display of strong affiliation with *süß* what appears to be most relevant is an exaggerated break from previous prosodic designs. For other social actions, the opposite may be the case. For example, in the case of return greetings, continuing a previously established prosodic pattern is the defining factor (Szczepek Reed 2006, 2009a, b).

Thus, two ways of approaching prosody in interaction present themselves to the analyst: the worthwhile detailing of specific prosodic features and their individual interactional functions; and the exploration of prosodic features more broadly as a resource for implementing categories of actions and sequential structure. This chapter has presented a contribution to the latter.

5. References

Ashmore, Malcolm/ Reed, Darren (2000): Innocence and nostalgia in conversation analysis: The dynamic relations of tape and transcript. In: Forum: Qualitative Research 1(3). (http://www.qualitativeresearch.net/index.php/fqs/article/view/1020, last access: 29-09-2014).

Couper-Kuhlen, Elizabeth (1996): The prosody of repetition: On quoting and mimicry. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation. Cambridge: CUP, 366-405.

Couper-Kuhlen, Elizabeth/ Barth-Weingarten, Dagmar (2011): A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 12: 1-51 (www.gespraechsforschung-ozs.de).

Betz, Emma/ Golato, Andrea (2008): Remembering relevant information and withholding relevant next actions: The German token achja. In: Research on Language and Social Interaction (41) 1, 58-98.

Deleuze, Gilles/ Guattari, Félix (1987): A Thousand Plateaus. Minneapolis: University of Minneapolis Press.

Drew, Paul/ Walker, Traci (2008): Going too far: Complaining, escalating and disaffiliations. In: Journal of Pragmatics 41 (12), 2400-2414.

Fagel, Sasha (2010): Effects of smiling on articulation: Lips, Larynx and Acoustics. In: Development of Multimodal Interfaces: Active Listening and Synchrony 5967, 294-303.

French, Peter/ Local, John (1983): Turn-competitive incomings. In: Journal of Pragmatics 7, 701-715.

Goldberg, Jo Ann (1978): Amplitude shift: A mechanism for the affiliation of utterances in conversational interaction. In: Schenkein, Jim (ed.): Studies in the Organization of Conversational Interaction. New York: New York Academic Press, 199-218.

Goodwin, Charles/ Goodwin, Marjorie H. (1987): Concurrent operations on talk: Notes on the interactive organization of assessments. In: IPRA Papers in Pragmatics 1 (1), 1-54.

Gorish, Jan/ Wells, Bill/ Brown, G.uy J. (2012): Pitch contour matching and interactional alignment across turns: an acoustic investigation. In: Language and Speech 55 (1), 57-76.

Heritage, John/ Raymond, Geoffrey (2005): The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. In: Social Psychology Quarterly 68 (1), 15-38.

Kaimaki, Marianna (2011): Sequentially determined function of pitch contours: the case of English news receipts. In: York Papers in Linguistics, Series 2, 11.

Kelly, John/ Local, John (1989): Doing Phonology. Manchester: Manchester UP.

Li, Xiaoting (2013): Language and the body in the construction of units in Mandarin face-to-face interaction. In: Szczepek Reed, Beatrice/ Raymond, Geoffrey (eds.): Units of Talk – Units of Action. Amsterdam: Benjamins, 343-375.

Li, Xiaoting (2014): Multimodality, Interaction and Turn-taking in Mandarin Conversation. Amsterdam: Benjamins.

Local, John (1992): Continuing and restarting. In: Auer, Peter/ di Luzio, Aldo (eds.): The Contextualization of Language. Amsterdam: Benjamins, 273-296.

Local, John/ Walker, Gareth (2004): Abrupt-joins as a resource for the production of multi-unit, multi-action turns. In: Journal of Pragmatics 36 (8), 1375-1404.

Local, John/ Kelly, John/ Wells, Bill (1986): Towards a phonology of conversation: turn-taking in Tyneside English. In: Journal of Linguistics 22 (2), 411-437.

Local, John/ Wells, Bill/ Sebba, Mark (1985): Phonology for conversation: phonetic aspects of turn delimitation in London Jamaican. In: Journal of Pragmatics 9, 309-330.

Marcus, George E./ Saka, Erkan (2006): Assemblage. In: Theory, Culture & Society 23 (2-3), 101-106.

Müller, Frank E. (1996): Affiliating and disaffiliating with continuers: prosodic aspects of recipiency. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation. Cambridge: CUP, 131-170.

Ogden, Richard (2004): Non-modal voice quality and turn-taking in Finnish. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (eds.): Sound Patterns in Interaction. Amsterdam: Benjamins, 29 – 62.

Ogden, Richard (2010): Prosodic constructions in making complaints. In: Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (eds.): Prosody in Interaction. Amsterdam: Benjamins, 81-103.

Raymond, Geoffrey (2013): At the intersection of turn and sequence organization: On the relevance of "slots" in type-conforming responses to polar interrogatives. In: Szczepek Reed, Beatrice/ Raymond, Geoffrey (eds.): Units of Talk – Units of Action. Amsterdam: Benjamins, 169-206.

Raymond, Geoffrey/ Heritage, John (2006): The epistemics of social relations: owning grandchildren. In: Language in Society 35, 677-705.

Selting, Margret (1996): On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. In: Pragmatics 6 (3), 371-388.

Selting, Margret/ Couper-Kuhlen, Elizabeth (eds.) (2001): Studies in Interactional Linguistics. Amsterdam: Benjamins.

Stivers, Tanya/ Rossano, Frederico (2012): Mobilizing response in interaction: A compositional view of questions. In: De Ruiter, Jan P. (ed.): Questions: Formal, functional and interactional perspectives. New York: CUP, 55-80.

Szczepek Reed, Beatrice (2004): Turn-final intonation in English. In: Couper-Kuhlen,

Elizabeth/ Ford, Cecilia E. (eds.): Sound Patterns in Interaction. Amsterdam: Benjamins, 97-117.

Szczepek Reed, Beatrice (2006): Prosodic Orientation in English Conversation. London: Palgrave Macmillan.

Szczepek Reed, Beatrice (2009a): Prosodic orientation: A practice for sequence organization in broadcast telephone openings. In: Journal of Pragmatics 41(6), 1223-1247.

Szczepek Reed, Beatrice (2009b): FIRST or SECOND: Establishing sequential roles through prosody. In: Barth-Weingarten, Dagmar/ Dehé, Nicole/ Wichmann, Anne (2009) (eds.): Where Prosody Meets Pragmatics. Bingley: Emerald, 205-222.

Szczepek Reed, Beatrice (2010): Prosody and alignment: A sequential perspective. In: Cultural Studies of Science Education 5 (4), 859-867.

Szczepek Reed, Beatrice (2012): Beyond the particular: Prosody and the coordination of actions. In: Language and Speech 55 (1), 13-34.

Tarplee, Clare (1996): Working on young children's utterances: Prosodic aspects of repetition during picture labeling. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in conversation. Cambridge: CUP, 406-435.

Ten Have, Paul (1997): In the presence of data: Conversation Analysis as 'empirical philosophy'. Paper presented at Ethnomethodology, an Improbable Sociology?, Cerisy-la-Salles, France, June 1997. (www.paultenhave.nl/presence.htm, last access: 29-09-2014)

Wells, Bill (2010): Tonal repetition and tonal contrast in English carer-child interaction. In Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (eds.): Prosody in interaction. Amsterdam: Benjamins, 243-262.

Wells, Bill/ Macfarlane, Sarah (1998): Prosody as an interactional resource: Turn-projection and overlap. In: Language and Speech 41 (3-4), 265-294.

Wells, Bill/ Peppé, Sue (1996): Ending up in Ulster: prosody and turn-taking in English dialects. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody and Conversation. Cambridge: CUP, 101-130.

Wells, Bill/ Corrin, Juliette (2004): Prosodic resources, turn-taking and overlap in children's talk-in-interaction. In: Couper-Kuhlen, Elizabeth/ Ford, Cecilia E. (eds.): Sound Patterns in Interaction. Amsterdam: Benjamins, 119-144.

6. Appendix

Adaptations from Couper-Kuhlen and Barth-Weingarten (2011)
Suprasegmental and embodied information

<<low>>	low pitch register
<<high>>	high pitch register
<<loud>>	increased loudness
<<quiet>>	decreased loudness
<<nodding>>	
<<smiling>>	
<<frowning>>	

Dr Beatrice Szczepek Reed
Department of Education
The University of York
York
YO10 5DD
United Kingdom
beatrice.szczepek.reed [at] york.ac.uk

Kooperative Gesten und Prosodie als theatrale Gestaltungsmittel

Beatrix Schönherr

Abstract
Prosodie gehört zu den elementaren theatralen Gestaltungsmitteln. Und nicht anders als bei der Analyse spontansprachlicher Daten ist es sinnvoll, ihre Funktion im Zusammenwirken mit weiteren Gestaltungsmitteln zu untersuchen. In diesem Beitrag wird die Rolle der Prosodie im Zusammenhang mit einem bestimmten Gestentyp untersucht, der in Theaterinszenierungen relativ häufig vorkommt und bislang gesprächslinguistisch wenig erforscht wurde. Das konstitutive Merkmal dieses Gestentyps besteht darin, von zwei oder mehr GesprächsteilnehmerInnen gemeinsam (symmetrisch oder komplementär) durchgeführt zu werden. Dafür schlage ich den Terminus *kooperative Gesten* vor. Kooperative Gesten sind nicht nur ein wichtiger Bestandteil der Theatergestik, sondern spielen auch außerhalb des Theaters in verschiedenen kommunikativen Situationen eine bedeutsame Rolle. Ihre Analyse erfordert die detaillierte Untersuchung, wie diese Gesten mit anderen Gestaltungsmitteln (vor allem lexiko-semantischen, prosodischen und blickbezogenen) in der Interaktion zusammenwirken. Das geschieht in diesem Beitrag methodisch einerseits im Rahmen der Gestenforschung, insbesondere der funktionalen Klassifikationen der Gesten, andererseits im Rahmen der interaktionalen Linguistik. Dabei erweist sich, dass kooperative Gesten im Theater ein geeignetes Mittel darstellen, um dramaturgisch wichtige Positionen und Entwicklungen zu fokussieren. Die Prosodie ist eng mit den kooperativen Gesten verbunden und an der Hervorbringung der genannten Wirkung entscheidend beteiligt.

Theaterprosodie – Theatergestik – kooperative Gesten – performative Gesten – Kookkurrenz Prosodie – Gestik

Prosody and gestures are basic means of expression in the theatre and should be investigated jointly together with further devices/resources of expression such as lexico-semantics, syntax, proxemics, and gaze. In this setting, the article deals with a very special type of gestures that occurs on stage quite frequently and has not been a common subject of research in the field of conversation and discourse analysis so far. The basic feature of this type of gestures is that it is performed collaboratively – either symmetrically or complementarily – by two or more participants in the interaction. For this phenomenon, which can just as well be observed in everyday conversations, I propose the term 'cooperative gestures'. In the scenes that are analyzed in this article cooperative gestures occur in situations when a turn of events takes place or a conflict arises. The cooperative gestures then bring these circumstances into focus. As in many other cases, prosody is tightly connected with the cooperative gestures and participates in producing the above-mentioned communicative effects.

Theatre – prosody – gestures – cooperative gestures – performative gestures – co-occurrence prosody – gestures

1. Einleitung

Mein Forschungsinteresse gilt seit längerem der Frage, wie die Prosodie mit verschiedenen syntaktischen und semantischen Strukturen auf der einen Seite und mit Gestik und Blickverhalten auf der anderen Seite zusammenhängt. Um die Ergebnisse aus der Analyse spontansprachlicher Daten besser auf ihre Reichweite abschätzen zu können, habe ich zuletzt zum Vergleich Theaterdaten herangezogen (vgl. Schönherr 2014). Dabei interessiert mich die Frage, ob die festgestellten Kookkurrenzen auch in der im Hinblick auf Produktion und Rezeption gänzlich anders gearteten Situation einer (verfilmten) Theateraufführung zu beobachten sind. Die Untersuchung prosodischer Aspekte gehört ebenso wie die Analyse der Gesten zu den zentralen Aufgaben einer umfassenden theatersemiotischen Aufführungsanalyse (vgl. Fischer-Lichte 1994: 28). Theaterprosodie und Theatergesten waren demgegenüber bislang kaum Gegenstand gesprächslinguistischer Forschung (vgl. aber Schönherr 1999 und Schönherr 2014). Das ist einerseits verständlich: Obwohl Theaterszenen an der Oberfläche oft (beinahe) wie Alltagsgespräche aussehen, sind sie von den Produktions- und Rezeptionsbedingungen her natürlich ganz anders zu beurteilen als die im gesprächslinguistischen Kontext üblicherweise geforderten spontansprachlichen Daten. Auf der anderen Seite kann wie oben erwähnt gerade dieser Unterschied auch für Untersuchungen fruchtbar gemacht werden. Dabei gewonnene Erkenntnisse sind natürlich für die Erforschung des Theaters relevant, andererseits können sie aber auch Anregungen für die linguistische Untersuchung spontansprachlicher Daten liefern.

Bei den folgenden Ausführungen steht die Rolle der Prosodie im Zusammenspiel mit einer besonderen Art von Gesten im Mittelpunkt. Dieser Gestentyp spielte bisher in der Gesprächsforschung keine große Rolle. Es handelt sich um Gesten, bei denen der sonst allgegenwärtige Wechsel von Sprecher- und Hörerrolle, von (vorwiegend) Produktion und (vorwiegend) Rezeption des multimodalen Interaktionsgeschehens für Sekunden aufgehoben scheint. Damit meine ich nicht den Umstand, dass auch die Aufgabe des Zuhörens und Zusehens im Gespräch keineswegs passiv ist, sondern vielfältige Aktivitäten wie z.B. verbale Rezeptionssignale, Hörergesten, Blickzu- und -abwendung etc. umfasst. Ich meine vielmehr eine weitgehende Gemeinsamkeit und Gleichberechtigung bei der Durchführung einer Geste durch zwei oder mehr GesprächsteilnehmerInnen, und zwar in der Art, dass dieses Kooperieren konstitutiv für die betreffende Geste ist. Das ist, wie im folgenden Abschnitt 2. noch näher ausgeführt wird, ein eher ungewöhnlicher Gestentyp. Umso mehr interessiert die Frage, ob sich die häufig beschriebene Synchronisation und das funktionale Zusammenwirken zwischen Prosodie und Gestik auch bei diesen besonderen Gesten feststellen lassen.

Solche Gesten sind keineswegs auf die Theaterbühne beschränkt, sie sind auch im sozialen Alltagsleben weit verbreitet. Allerdings dürften sie aus Gründen, die sich noch zeigen werden, in einer Theaterinszenierung in größerer Dichte zu beobachten sein. Eine größere Belegzahl erleichtert das Erkennen von Mustern. Außerdem werden diese Gesten im Theater geradezu 'vorgeführt'. Sowohl im Hinblick auf die Sichtbarkeit als auch auf andere Aspekte der Nachvollziehbarkeit wird den Bedürfnissen des Publikums Rechnung getragen. Das kommt auch der analysierenden Linguistin zu Gute. Außerdem sind die von mir untersuchten Fälle für den Fortgang der Handlung von entscheidender Bedeutung, zeigen einen Kon-

flikt besonders deutlich auf oder stehen an einem Wendepunkt. Durch diese exponierte Position kann die Funktion dieses Gestentyps relativ gut analysiert werden. Damit dürfte die Verwendung von Theaterdaten sowohl für die Erforschung der Theaterprosodie und der Theatergestik sinnvoll sein als auch Anregungen geben, dem hier im Mittelpunkt stehenden Gestentyp und seinen Bezügen zur Prosodie sowie allgemein seiner Einbettung in ein multimodales Interaktionsgeschehen auch in spontanen Gesprächen Aufmerksamkeit zu schenken.

2. Ein besonderer Gestentyp

In seiner klassischen Darstellung der Gesten schildert Efron (1941/ 1972: 89f.), dass ostjüdische Einwanderer in New York bei ihren Gesprächen häufig den Körper des Gesprächspartners berühren, um seine Aufmerksamkeit zu erlangen oder ihn zu unterbrechen. Oft ergreifen sie dabei das Handgelenk ihres Gegenübers und halten es fest. In einem Fall konnte Efron sogar beobachten, dass einer der Gesprächspartner in der Hitze der Auseinandersetzung das Handgelenk des andern festhielt und damit seinerseits gestikulierte. Das ist ein seltenes Beispiel in der Forschungsliteratur, bei dem eine Person die gestikulierenden Hände des Gesprächspartners berührt und damit Einfluss auf die Fortsetzung der Gestik und indirekt auch des Gesprächs nimmt. Daraus erhebt sich die Frage: Ist es tatsächlich so selten, dass in einem Gespräch die ZuhörerInnen durch Berührungen Einfluss auf die Gestik der SprecherInnen nehmen oder sogar an dieser Gestik teilhaben?

Hier ist nicht der Platz, das der Gestikforschung zu Grunde liegende Interaktionsmodell ausführlich zu diskutieren – allein schon deshalb nicht, weil diese Gestikforschung in sehr unterschiedlichen Disziplinen stattfindet und in sich heterogen ist. Überlegungen, die Herstellung von Sinn und die Konstitution von Handlungen als interaktiven Prozess zu erfassen, sind in Forschungsarbeiten zum nonverbalen Verhalten durchaus zu finden (vgl. z.B. Scherer 1980: 15-16), wenn auch im Allgemeinen nicht so prononciert und so programmatisch wie in der Forschungstradition der ethnomethodologischen Konversationsanalyse (z.B. Garfinkel 1967; vgl. auch die Ausführungen zum "Konstitutionsmodell der Kommunikation" bei Brinker/ Sager 2010: 122-123; vgl. auch die Grundlagen des Programms der interaktionalen Linguistik, Selting/ Couper-Kuhlen 2001). Diese ethnomethodologische Forschungshaltung ist meiner Überzeugung nach als Perspektive mit vielen Konzepten der Gestikforschung gut vereinbar. In diesem Beitrag gehe ich von einem Konstitutionsmodell der Kommunikation (Brinker/ Sager 2010: 122) aus und möchte die folgenden Aussagen – z.B. zur Rolle von SprecherIn und HörerIn – im Rahmen eines solchen – multimodal interpretierten – Modells verstanden wissen.

Gestik wird gemeinhin mit der Rolle des Sprechers/ der Sprecherin verknüpft (vgl. McNeill 1992: 23 et passim; vgl. auch Kendon 1980). Das ist nur folgerichtig, wenn man wie McNeill von der Theorie des *growth point* ausgeht, wonach Sprache und Gestik nur zwei verschiedene Seiten ein und derselben Medaille und bereits in der Äußerungsplanung untrennbar verknüpft sind. Dann spricht eben nur eine Partei, während bis zum Sprecherwechsel die andere zuhört und zusieht und sich auf Rückmeldesignale beschränkt. Jedenfalls gilt das für die von Müller

(1998: 103-107) als 'freie Gesten' bezeichneten Körperbewegungen, denen sie die Selbstberührungen als typische Hörergesten gegenüberstellt.

Gestik ist außerdem, grob gesprochen, zum Anschauen da, zur Rezeption durch den Gesichtssinn gedacht, und nicht, wie etwa Berührungen, um haptisch rezipiert zu werden.

Gestik im Sinne von 'freien Gesten' wird also typischerweise a) im Gespräch nur von der Person in der Sprecherrolle eingesetzt und b) weitgehend ohne Berührung des Gesprächspartners/ der Gesprächspartnerin durchgeführt. Eine solche Auffassung von redebegleitenden Gesten wird bislang in der Forschung gemeinhin vertreten – und in sehr vielen Gesprächssituationen dürften auch tatsächlich die meisten vorkommenden freien Gesten die genannten Eigenschaften aufweisen. Auch das oben erwähnte Beispiel von Efron widerspricht dieser Aussage nicht grundsätzlich: Das Berühren der gestikulierenden Hände des Gesprächspartners stellt eine Art gestischen Unterbrechungsversuchs dar, die Behinderung der Gestik des anderen ist vergleichbar mit lautem überlappendem Sprechen mit dem Ziel der Turn-Übernahme vor Erreichen eines TRPs (*transition relevance place*) (vgl. Sacks/ Schegloff/ Jefferson 1974). Damit handelt es sich hier um einen Regelverstoß, der die dahinterstehende Regel bestätigt.

Eine Ausnahme von den beiden Regeln stellen allerdings jene weitgehend konventionalisierten Gesten dar, die zu ihrer korrekten und vollständigen Durchführung die Mitwirkung des/ der anderen benötigen. Ein Beispiel hierfür wären verschiedenste Formen des (gestischen oder multimodalen) Grüßens. Betrachten wir die in unserem Kulturkreis übliche Form des Händeschüttelns. Wer die rechte Hand zum Handschlag vorstreckt, erwartet, dass das Gegenüber die Hand mit seiner eigenen ergreift, worauf beide Hände (mehr oder weniger) Druck ausüben ('Händedruck') und sich gemeinsam auf und ab bewegen ('Händeschütteln'). Das Angebot zum Händedruck macht – vergleichbar mit dem ersten Teil einer Paarsequenz – die Erwiderung bzw. Kooperation konditionell relevant. Michaels (2010: 100f.) plädiert am Beispiel von hinduistischen Grußgesten dafür, diese Erwartungshaltungen im Rahmen einer übergreifenden Kommunikationstheorie zu sehen.

> Ganz gleich, was man einem anderen gibt – Gaben, (Gruß-)Worte, Einladungen, Gesten oder Blicke –, man setzt sich dem Nehmer bzw. Empfänger aus, indem man ihm die Wahlmöglichkeit gibt zu einer kulturell und sozial akzeptierten oder inakzeptablen Erwiderung. (Michaels 2010: 100)

Allerdings gibt es auch im Hinblick auf den Paarsequenz-Charakter einen wichtigen Unterschied zwischen der verbalen und der gestischen Grußsequenz. Während der verbale Gruß aufgrund des linearen Charakters der Sprache tatsächlich in einer zeitlichen Abfolge von erstem Schritt und zweitem Schritt ausgeführt wird, wird der gestische Gruß vielleicht einseitig initiiert, dann aber simultan durchgeführt (Händeschütteln). Dass ursprünglich ein Interaktionsteilnehmer/ eine Interaktionsteilnehmerin die Initiative ergriffen hat, wird nur bei Abweichungen deutlich, insbesondere wenn der Gruß nicht erwidert wird, wenn also das Angebot der ausgestreckten Hand verweigert wird. Dann zeigt sich, dass auch hinter diesem gleichzeitig durchgeführten Grußverhalten ein erster Schritt einen zweiten bzw. das Kooperieren konditionell relevant macht. Ich möchte für diese Gruppe von Gesten im Folgenden die Bezeichnung *kooperative Gesten* verwenden.

Überraschenderweise werden diese kooperativen Gesten in den verschiedensten funktionalen Gestenklassifikationen (vgl. Schönherr im Druck) nicht als eigene Gruppe erwähnt. Sie lassen sich vielmehr je nach Gesichtspunkt verschiedenen Gestengruppen zuweisen. Einerseits besitzen sie oft die Eigenschaft von Emblemen. Damit sind jene Gesten gemeint, die auch unabhängig von einer etwaigen begleitenden verbalen Äußerung eine eindeutige, kulturell festgelegte Bedeutung haben (vgl. Efron 1941/ 1972: 96 sowie Ekman/ Friesen 1969). Ein Beispiel wäre das Kopfnicken als Ausdruck der Zustimmung.

Ein anderer Aspekt der meisten kooperativen Gesten wird durch den von Müller (1998) verwendeten Terminus *performative Gesten* erfasst: "Mit performativen Gesten werden Sprechhandlungen dargestellt und gestisch vollzogen" (:111). Müller stellt die performativen Gesten den referentiellen Gesten gegenüber, die auf Gegenstände, Eigenschaften, Verhalten etc. referieren.

Auch Frickes Gliederung der Gesten geht aus semiotischer Perspektive von diesem grundsätzlichen Unterschied zwischen performativen und referentiellen Gesten aus: "Performative Gesten sind dadurch gekennzeichnet, dass ihr jeweiliges Bezeichnungspotential für die Äußerung als Sprechhandlung nicht konstitutiv ist. So wird mit der emblematischen Geste des Schwörens ein Schwur vollzogen, nicht jedoch auf diesen referiert." (Fricke 2012: 25; vgl. auch Fricke 2007: 220). Die performativen Gesten sind bei Fricke eine Untergruppe der 'nichtreferierenden Gesten'. Weitere nichtreferierende Gesten wären taktstockartige Gesten und ideographische Gesten. (Letztere bilden den Gedankengang ab.) Referierende Gesten wären dagegen z.B. deiktische Gesten (Zeigegesten) und ikonische Gesten. Ich kann mir theoretisch auch kooperative Gesten vorstellen, die zu den referierenden Gesten zählen, etwa gemeinsam durchgeführte Kinetographen (diese bilden einen Bewegungsablauf ab), doch dürften solche Gesten sehr selten sein. In meinen weiteren Ausführungen behandle ich nur performative kooperative Gesten.

Viele kooperative Gesten dürften schließlich auch Eigenschaften der rituellen Gebärden im Sinne von Wundt (1911: 158f.) aufweisen. Einzelne Beispiele, die er nennt, vor allem jene aus dem Bereich der 'rituellen Rechtssymbole', entsprechen durchaus dem, was ich in diesem Beitrag behandle, z.B. "Handreichung bei der Vertragsschließung" oder "das die Besitznahme andeutende Ergreifen eines Gegenstands" (:159). Auch die "langsame, feierliche Form" (:159) ist bei den hier untersuchten kooperativen Gesten zu beobachten. Nur die Einbettung in den größeren Rahmen eines feierlichen Zeremoniells (vgl. Wundt 1911: 158) fehlt weitgehend.

Kooperative Gesten sind also eine Teilgruppe der performativen Gesten, oft zugleich den Emblemen zuzurechnen und/ oder in Verbindung mit Ritualen verwendet, deren definierendes Merkmal darin besteht, zu ihrer korrekten Durchführung der Mitwirkung von (zumindest) zwei InteraktionsteilnehmerInnen zu bedürfen. Dabei kann die Form der Geste bei beiden InteraktionsteilnehmerInnen weitgehend symmetrisch sein wie etwa bei einem Handschlag zur Besiegelung einer Vereinbarung oder bei einem Begrüßungs-Wangenkuss. Sie kann aber auch komplementär sein wie der höfische Handkuss oder die Einheit von Anbieten und Übernahme eines Gegenstands. Das sollte als Arbeitsdefinition für die folgenden Ausführungen eine ausreichende Basis darstellen.

Wie oben ausgeführt, kann eine Verbindung der kooperativen Gesten zu den Paarsequenzen im Sinne der ethnomethodologischen Konversationsanalyse hergestellt werden, wobei im Gegensatz zu Paarsequenzen bei kooperativen Gesten die Abfolge von erstem Zug und Reaktion darauf allerdings nicht das vollständige Muster darstellt, sondern lediglich die Einleitung der eigentlichen Geste.

Mir ist nur eine Arbeit bekannt, die auf den ersten Blick ein vergleichbares Phänomen beschreibt. Furuyama (2000: 105) definiert folgendermaßen einen Gestentyp:

> Collaborative gestures are those which interact with the gestures of the communicative partner. They are basically not complete in themselves in terms of what they represent without the existence of the gesture of the communicative partner.

Sowohl die Bezeichnung als auch die Beschreibung erinnern sehr an die hier vorgestellten kooperativen Gesten. Bei näherer Betrachtung stellt sich jedoch heraus, dass etwas anderes gemeint ist. Das Untersuchungskorpus besteht aus Gesprächen zwischen jeweils zwei Personen, von denen die eine der anderen das Falten eines Origami-Modells (Origami = japanische Kunst des Papierfaltens) erklären soll, und zwar als 'Trockenübung', ohne Verwendung von Papier. Die Person, die die Instruktionen gibt, verwendet dabei wie zu erwarten viele ikonische und deiktische Gesten. Die lernende Person imitiert dann häufig diese Gesten. Wenn sie dabei auf die Geste des Instrukteurs zeigt (ein Fall von Deixis am Phantasma im Sinne von Bühler 1934/ 1982) oder das von der Geste des Instrukteurs hervorgerufene virtuelle Objekt irgendwie manipuliert, spricht Furuyama von *collaborative gesture*. Weil sie sich auf die Gesten des Gesprächspartners/ der Gesprächspartnerin beziehen, sind *collaborative gestures* 'meta-gestural' (2000: 106). Sowohl diese Eigenschaft als auch die Zugehörigkeit zu den ikonischen und deiktischen Gesten unterscheidet die *collaborative gestures* von den hier beschriebenen kooperativen Gesten. Ein weiterer gewichtiger Unterschied besteht in der sequentiellen Abfolge: *Collaborative gestures* folgen der Geste, auf die sie sich beziehen, nach, und erfüllen eine andere Funktion als diese (z.B. Nachfrage). Kooperative Gesten weisen zwar in der Einleitungsphase ebenfalls eine Abfolge von Initiierung und Akzeptieren (bzw. Ablehnung) auf, werden in der Kernphase aber überwiegend gleichzeitig und gemeinsam durchgeführt und erfüllen eine gemeinsame Funktion (wenngleich diese für die Beteiligten an der Durchführung nicht in allen Aspekten identisch sein muss).

Obwohl kooperative Gesten keinen prominenten Platz in der Gestenforschung einnehmen, werden einige von ihnen in anderen Disziplinen sehr ausgiebig behandelt. Zu den verschiedenen gestischen Formen des Grüßens gibt es z.B. viele ethologische, ethnologische, kulturvergleichende und -historische etc. Studien (vgl. die Auflistung bei Bergmann 1994: 195).

Kooperative Gesten, mit oder ohne damit verbundene(n) Berührungen, haben eine starke Wirkung und werden deshalb in konventionalisierter Form gerne benutzt, um die persönliche Beziehung auszudrücken, herzustellen oder zu bestärken. Viele Gruß- und Abschiedskonventionen sind wie erwähnt mit Gesten und Berührungen verbunden (vgl. Eibl-Eibesfeldt 1970: 193-221). Auch in anderen Situationen, die das Verhältnis der InteraktionspartnerInnen zueinander betreffen, spielen Gesten und gestische Berührungen eine wichtige Rolle, zum Ausdruck von Liebe und Freundschaft, zur Bezeugung von Harmonie oder zur Beschwichti-

gung (vgl. Montagu 1971 und Hall 1966; vgl. auch den Überblick über die zum Teil widersprüchlichen Forschungsergebnisse zum Berührungsverhalten bei Dreischer 2001). Ein anderer Bereich umfasst die Herstellung neuer sozialer Strukturen, das Treffen von Vereinbarungen, die Annahme von Angeboten und Ähnliches.

Vor diesem Hintergrund ist es interessant zu beobachten, wie es um den Einsatz dieser spezifischen Gestenform auf der Theaterbühne bestellt ist. Prinzipiell spiegelt aus theatersemiotischer Sicht der Einsatz der Gestik die Verwendung der Gesten im außertheatralen Alltag wider – wenngleich je nach Theatertradition auch ganz spezifische, streng geregelte gestische Codes vorliegen können (vgl. Fischer-Lichte 1994: 82). Allerdings bestehen einige wichtige Unterschiede zwischen Alltagsgesten und Theatergesten. Nach Fischer-Lichte (1994: 28) haben Gesten im Theater eine etwas komplizierte semiotische Struktur: Sie verweisen ikonisch auf Gesten außerhalb der Bühne und erhalten erst indirekt über diesen Verweis eine Bedeutung. Außerdem werden sie (im Gegensatz zu Gesten im Alltag) grundsätzlich mit kommunikativer Absicht eingesetzt: Auch ein Adaptor (z.B. Kopfkratzen oder Augenreiben) wird auf der Bühne kommunikativ eingesetzt, um zum Beispiel eine bestimmte psychische Verfassung der Figur zu vermitteln. Hintergrund für die Interpretation dieser Geste ist natürlich auch dabei der entsprechende Informationsgehalt des betreffenden Verhaltens im Alltag, das dort üblicherweise zwar nicht mit kommunikativer Absicht gezeigt wird, wohl aber von den InteraktionsteilnehmerInnen interpretiert und in manchen Fällen auch thematisiert bzw. interaktional relevant gemacht werden kann.

Eine Eigenschaft, die Fischer-Lichte (2010) den theatralen Gesten zuspricht – womit sie allerdings nicht alle im Theater vorkommenden Gesten meint –, könnte nicht viel anders auch den kooperativen Gesten zugeschrieben werden. Mit 'theatralen Gesten' sind solche Gesten gemeint – sei es im Theater oder im sozialen Leben –, die eine 'transformative Kraft' aufweisen: "Beide verwandeln den, der sie ausführt, in den Augen derer, die sie wahrnehmen. Beide vermögen zu einer neuen Selbst-, Fremd- und Weltwahrnehmung zu führen." (Fischer-Lichte 2010: 222). Genau genommen liegt hier allerdings allgemeiner die Parallele zu den performativen Gesten nahe – von gemeinsamer Durchführung ist bei den theatralen Gesten nicht die Rede. Ein wichtiger Unterschied zwischen der transformativen Kraft der theatralen Gesten im Sinne von Fischer-Lichte und der Wirkung der performativen Gesten darf freilich nicht übersehen werden: Letztere beruht auf Konventionen und Ritualen, was für die transformative Kraft nicht gelten dürfte.

Im Folgenden wird das Zusammenspiel von Prosodie und Gestik in ausgewählten Theaterszenen, in deren Zentrum jeweils kooperative Gesten stehen, untersucht. Neben dem Bezug zur Prosodie werden weitere Analyseaspekte berücksichtigt. Kooperative Gesten sollen einerseits im Kontext der anderen vorkommenden Gesten analysiert werden, andererseits wird aber auch ihr Zusammenwirken mit weiteren Ausdrucksmitteln (vor allem lexiko-semantischen und syntaktischen sowie proxemischen und blickbezogenen) untersucht. Daraus soll sich zum einen eine Einschätzung ergeben, welche Funktion die kooperativen Gesten im Rahmen der jeweiligen Inszenierung erfüllen. Zum anderen können Erkenntnisse darüber gewonnen werden, ob die oft beobachtete enge Verbindung von Gestik und Prosodie auch im Zusammenhang mit diesem besonderen Gestentyp auftritt.

Diese Ergebnisse sind einerseits theaterspezifisch, lassen andererseits aber durch die erwähnte Rückbindung der Theatergesten an die Gesten im sozialen Zusammenleben der Menschen im Alltag auch Schlüsse für die Analyse von Interaktionen im Allgemeinen zu.

3. Untersuchungsmaterial

Das Korpus für die vorliegende Untersuchung besteht aus ausgewählten Szenen aus vier verschiedenen aufgezeichneten Theaterinszenierungen. Dabei handelt es sich um jeweils zwei (auch und besonders in Bezug auf den Inszenierungsstil) unterschiedliche Aufführungen der Stücke "Don Carlos" von Friedrich Schiller und "Emilia Galotti" von Gotthold Ephraim Lessing. Im Folgenden die Eckdaten der Inszenierungen (nur jene Darstellerinnen und Darsteller sind angeführt, die in den analysierten Beispielszenen mitwirken):

Don Carlos

a) Inszenierung Franz Peter Wirth (Fernsehspiel D 1984), Rolf Boysen (König Philipp), Marita Marschall (Königin Elisabeth), Jacques Breuer (Don Carlos), Robert Atzorn (Marquis Posa)

b) Inszenierung Andrea Breth (Burgtheater 2005), Sven-Eric Bechtolf (König Philipp), Johanna Wokalek (Königin Elisabeth), Philipp Hauß (Don Carlos), Denis Petkovic (Marquis Posa)

Emilia Galotti

a) Inszenierung Thomas Langhoff (Münchner Kammerspiele 1984), Rolf Boysen (Odoardo Galotti), Margit Carstensen (Gräfin Orsina)

b) Inszenierung Michael Thalheimer (Deutsches Theater 2002), Peter Pagel (Odoardo Galotti), Nina Hoss (Gräfin Orsina)

Das Korpus umfasst also zwei Mal jeweils eine etwas ältere und eine jüngere Inszenierung desselben Stücks. Der Grund für diese Auswahl war weniger die Erfassung der (zweifellos vorhandenen) historischen Entwicklung der Aufführungs- und Regie-Stile. Vielmehr sollen die unterschiedlichen Inszenierungsstile die Untersuchung ermöglichen, ob an vergleichbaren – von der Textbasis her mehr oder weniger identischen – Stellen jeweils dieselben oder ähnliche kooperative Gesten beobachtet werden können, ungeachtet sonstiger Unterschiede in der Interpretation der jeweiligen Szenen.

Aus methodischer Sicht muss ich noch eine wichtige Einschränkung bzw. Ergänzung vornehmen: Ich habe die betreffenden Inszenierungen nicht live gesehen, sondern arbeite mit Verfilmungen dieser Inszenierungen. Dadurch ist meine Wahrnehmungsmöglichkeit durch die Kameraführung, vor allem durch die Wahl des Bildausschnitts, durch die Einstellungsgröße und die Kameraperspektive, beeinflusst. Umso wichtiger war es deshalb, für die Analyse solche Stellen auszuwählen, bei denen die zu untersuchenden Gesten ausreichend sichtbar waren, und zwischen tatsächlich wahrnehmbaren und nur vermuteten Bewegungen streng zu unterscheiden. Ob die Aufführungen für die Verfilmung auch stilistisch an die

Sprache des Films angeglichen sind, indem z.B. prosodische und gestische Mittel sparsamer eingesetzt werden, kann ich mangels direkten Vergleichs nicht mit Sicherheit beurteilen. Ich kann aber sagen, dass die von mir verwendeten Verfilmungen von Inszenierungen auf mich durchaus 'theatral' wirken.

4. Zur Transkription

Die Notation des verbalen und prosodischen Anteils des Verhaltens erfolgt gemäß den GAT 2-Konventionen (Selting et al. 2009). Für die Notation von Gestik und Blickverhalten werden dagegen eigene Zeilen verwendet. Pro Verhaltens-'Kanal' wird eine eigene Zeile angelegt. Nur die Zeilen mit dem verbalen Anteil (bzw. mit der Notation einer Pause, falls keine simultane verbale Äußerung stattfindet) werden durchnummeriert, die darunterstehenden Zeilen sind jeweils wie bei Partiturschreibung simultan zur obersten Zeile zu lesen. Um den Bezug auf den verbalen Anteil möglichst präzis zu machen (was besonders bei der gestischen Unterstützung von prosodischen Akzenten und bei der gestischen Grenzmarkierung wichtig ist), werden die Zeichen * für Anfangs- und Endpunkt eines länger dauernden Verhaltens und ◊ für den Zeitpunkt eines kurzen Ereignisses verwendet. Falls das Verhalten bzw. eine erreichte Position über mehrere Zeilen andauert, wird der Beginn mit * notiert und am Ende der Zeile das Zeichen ⇒ eingefügt. Im Falle von ◊ und bei Platzmangel auch bei * wird das zu diesem Zeitpunkt stattfindende Verhalten in einer weiteren Zeile direkt unterhalb dieser Zeichen erläutert.

Obwohl diese Schreibweise im Ansatz einer Partiturschreibung vergleichbar ist, werden überlappende Verhaltensweisen nicht als Partitur dargestellt, sondern wie im GAT üblich durch Einklammern der jeweils überlappenden Teile mit eckigen Klammern. Besonders hinweisen möchte ich auf Fälle wie die Zeilen 6 und 7 im Transkript (1) (vgl. Abschnitt 5.1).

```
6   P:   [(1.0)]
         steht auf
7   C:   [(1.0)]
         lässt li Hand von Ps Schulter sinken
```

Diese sind so zu lesen: Während eines Schweigens von einer Sekunde Dauer steht P auf und C lässt gleichzeitig (!) seine Hand von der Schulter sinken. Die eckigen Klammern für Gleichzeitigkeit werden also einheitlich nur in den nummerierten Zeilen für das verbale Verhalten gesetzt, und sollte dort nur eine Pause notiert sein, dann wird konsequenterweise diese Pause bei beiden Interaktionsteilnehmer-Innen notiert.

Das Blickverhalten ist nicht durchgehend notiert, sondern nur dann, wenn es in der Interpretation berücksichtigt wird. In den Gestikzeilen verwendete Abkürzungen sind im Anhang angeführt.

5. Analyse von ausgewählten Szenen

Vor der Analyse noch ein Hinweis, wie ich die Darstellenden bzw. die Charaktere bezeichne. Streng genommen müsste ich jeweils formulieren "Der Darsteller des Don Carlos ergreift die Hand" und nicht "Don Carlos ergreift die Hand". Das ist jedoch auf die Dauer etwas umständlich. Deshalb verwende ich die letztere Form, quasi als metonymische Abkürzung.

5.1. Ein Handschlag zur Besiegelung des Freundschaftsbunds

Der folgende Szenenausschnitt aus "Don Carlos" stammt aus dem 1. Akt, 9. Auftritt. Don Carlos, der spanische Infant, hat seinen Freund aus Kindertagen, den Marquis von Posa, wiedergetroffen und bittet ihn, ungeachtet des Standesunterschieds zwischen dem zukünftigen Monarchen und seinem Untertan, um brüderliche Freundschaft. Der Marquis gibt zu bedenken, dass er in der Situation eines Abhängigen keine echte Freundschaft empfinden könne. Carlos argumentiert dagegen, dass von Posa auf Grund seiner Verachtung für die Vergünstigungen des Hofes nie in die Position des Abhängigen kommen könnte. Der Transkriptausschnitt beginnt mit dem Abschluss dieser Argumentation.

(1) Hier meine Hand (Wirth 1984)

(Der Marquis von Posa sitzt auf einer Bank, die rechte Hand auf dem Schoß liegend, die linke neben sich auf der Bank. Er blickt nach rechts, Don Carlos ins Gesicht. Dieser steht schräg vor ihm, seine linke Hand in die Hüfte gestützt, mit der rechten gestikulierend.)

```
1   C:   du  `bIst  ein  `REICHrer  `Untertan;=
         *    geht mit zwei großen Schritten zu P hin    *
         *legt li Hand P auf Schulter**li Hand bleibt liegen      ⇒
              ◊       ◊      ◊
            kleine Tn mit re Hand, Handfläche nach oben

2        =als  `ich  ein  `kÖnig je `SEIN  werde.
                            ◊    ◊    ◊
                          rhythmisches Senken des li Ellbogens
                          gleichzeitig stufenweise Drehung Kopf nach li
         *führt re Hand zur Brust ** hält Fingerspitzen auf Brust*

3        °hh `wEr von uns wird der ´GLÄUbiger des andern
               ◊                        ◊
            T mit re Hand              wegwerfende Handbewegung re

4        und `wEr der `SCHULDner sein.
               ◊          ◊
                 Tn mit re Hand

5        (1.0)
         C und P sehen einander regungslos an
```

```
 6   P:   [(1.0)]
          steht auf

 7   C:   [(1.0)]
          lässt li Hand von Ps Schulter sinken

 8   P:   wohl´An ich `WEIche.
          *  hebt rechte Hand        *

 9        `hIEr meine `HAND.
          streckt re Hand aus und bietet sie zum Handschlag

10   C:   (1.0)
          blickt die Hand von P an, ergreift sie dann und blickt ihm wieder in die
          Augen

11        °hh der ´MEInige?

12        (--)

13   P:   auf `Ewig.

14        (--)

15   C:   so `TREU wie `hEUte dem in`fAnten;=

16        =auch `dErmal`EInst dem ´KÖnig zugetan?

17        (--)

18   P:   ´dAs `SCHWÖR ich ihnen.

19        (1.0)

20   C:   (2.0)
          umarmt P stürmisch
```

Bereits die Analyse der ersten Transkriptzeilen – die letzten Worte einer längeren Argumentation von Don Carlos – zeigt interessante prosodische und gestische Ausdrucksformen. Wie oft beschrieben (vgl. Schönherr 1997: 45-52 und die dort zitierte Literatur) kookkurrieren hier Positionsveränderungen mit prosodischen und syntaktischen Strukturen einerseits sowie mit Änderungen auf der Handlungsebene andererseits. Hatte Carlos zuvor rhetorische Fragen formuliert, zieht er nun aus der (impliziten) Beantwortung dieser Fragen den Schluss, mit dem der Transkriptausschnitt beginnt. Zugleich verstärkt er seinen Appell an den Marquis von Posa: Er legt seine linke Hand auf die Schulter des Marquis, wobei diese Bewegung wohl nicht zufällig synchron mit dem stärksten Akzent der Intonations-

phrase ihren Abschluss findet. Diese Berührung stellt Nähe her und verstärkt zugleich die Eindringlichkeit der Aussagen – zumal sich die anderen Bewegungen von Carlos zeitweilig so auf seine linke Hand übertragen, dass er den Marquis leicht "schüttelt" (was im Transkript aus Gründen der Übersichtlichkeit nicht vermerkt ist). In Zeile 2 vollführt er mit der freien Hand eine selbst-deiktische Geste (Lausberg o.J.), wobei er die Hand mit gekrümmten Fingern zur Brust führt, bis er sie zeitgleich mit der akzentuierten Silbe *`kÖ(nig)* erreicht (*stroke*, vgl. Kendon 1980), s. Abb. 1. Dort bleiben die Fingerspitzen dann bis zum Ende der Intonationsphrase.

Abb. 1: Selbstdeiktische Geste (Don Carlos – Infant von Spanien. R.: Franz Peter Wirth. D 1984. Jacques Breuer (Don Carlos), Robert Atzorn (Marquis Posa))

Obwohl nun beide Hände "ruhiggestellt" sind, führt Carlos trotzdem Taktzeichen bzw. *batons* (Efron 1941/ 1972) oder *beats* (McNeill 1992) durch. Die Silben *`kÖ(nig)*, *je* und *`SEIN* – also zwei betonte und eine von ihnen umschlossene unbetonte Silbe – werden jeweils sowohl durch ein rhythmisches Senken des linken Ellenbogens (soweit die Fixierung der Hand auf der Schulter des Gesprächspartners das erlaubt) als auch durch schrittweise Drehbewegungen von Kopf und Oberkörper hervorgehoben. Damit verbunden ändert sich die Blickrichtung: Am Beginn von Zeile 2 war der Blick leicht zur rechten Seite, am Marquis vorbei, gerichtet, am Ende der Intonationsphrase blickt Carlos dem Marquis wieder ins Gesicht. Beide Arten von Taktzeichen – sowohl das Schwingen der Ellbogen bei fixierten Händen als auch rhythmische Kopfbewegungen – habe ich bei einer anderen kommunikativen Gattung, nämlich bei Nachrichtenmoderationen im Fernsehen (vgl. Schönherr 2005) schon beobachtet. Auch dort war die Möglichkeit, die Worte mit Taktzeichen zu begleiten, eingeschränkt – wenn auch aus anderen,

nämlich gattungsimmanenten Gründen. In den Zeilen 3 und 4 sehen wir weitere Taktzeichen, wobei eines, die Hervorhebung von *´GLÄU(biger),* zugleich als performatives Emblem eine zusätzliche Bedeutung einbringt. (Die Verwendung dieses Emblems könnte so umschrieben werden, dass damit etwas – z.B. ein vorgebrachter Standpunkt – dezidiert zurückgewiesen wird. Das passt zur gleichzeitigen verbalen Äußerung, die der Widerlegung eines Arguments dient, auch wenn der exakte Bezug der Geste nicht eindeutig ist.)

Zusammengefasst sehen wir in den Zeilen 1-4: Gestik und/ oder Positionsveränderungen zur Grenzmarkierung und zur Hervorhebung, beides in Kookkurrenz mit der Prosodie, teilweise auch mit der Syntax – und natürlich mit dem lexikosemantischen Teil der Äußerung; Berührung zur Herstellung von Nähe und Intensivierung der Aussage; Gestik zur Selbstreferenz und zum Vollzug performativer Handlungen.

So weit unterscheidet sich dieses multimodale Geschehen in einer Theateraufführung nicht von Alltagsgesprächen, wenn man vielleicht einmal von der 'publikumsfreundlichen' Positionierung absieht: Obwohl das Gespräch als sehr eindringlich dargestellt wird, stehen/ sitzen sich die beiden Männer nicht gegenüber, sondern etwa im rechten Winkel zueinander, beide halb seitlich von vorne zu sehen. Diese Position wäre in einem außertheatralen Gespräch möglicherweise als Offenheit der Interaktanten für weitere TeilnehmerInnen zu deuten, was im Kontext des Stückes aber weitgehend ausgeschlossen werden kann.

Ab Z. 6 können wir die Reaktion des Marquis von Posa sehen: Er steht zuerst auf. Damit übernimmt er einerseits die Sprecherrolle, indem er eine aktivere Position einnimmt. Andererseits deutet sich hier auch eine Änderung seines 'Standpunkts' im metaphorischen Sinn an. Erst nach der Positionsänderung beginnt er zu sprechen, signalisiert verbal seine Zustimmung und bietet seine Hand zur Besiegelung des Versprechens an. Dabei können zwei Phasen unterschieden werden, jeweils inhaltlich, syntaktisch, prosodisch und gestisch: In der ersten Phase zieht er für sich das Resümee aus den Argumenten von Carlos (*wohl´An ich `WEIche.*). Prosodisch fällt dabei die Aufeinanderfolge von steigendem und fallendem Akzent auf, oft ein Signal dafür, dass zwei Einheiten irgendwie inhaltlich zueinander in Beziehung gesetzt werden (hier etwa: ‚wenn das so ist – gebe ich nach‘). Die Handbewegung, die er in dieser ersten Phase macht, ergibt erst zusammen mit dem zweiten Teil einen Sinn. Sie ist nämlich eine *preparation* (Kendon 1980: 212) bzw. ein *onset* (Wallbott 1981: 290) in Bezug auf die folgende Geste des Anbietens der Hand, die auch verbal kommentiert wird (Z. 9). Hier, in der zweiten Phase, liegt der zentrale Teil der Interaktion bei der Geste – sie könnte auch ohne den verbalen Teil durchgeführt werden, während die verbale Äußerung ohne die begleitende Geste sinnlos bliebe. Diese Geste erzeugt aus den oben erwähnten Gründen immer Aufmerksamkeit, weil sie auf Erwiderung bzw. Mitwirkung ausgelegt ist. Hier wird sie noch zusätzlich dadurch an Bedeutung aufgeladen, dass die Hand eine Sekunde in der Luft stehen bleibt (s. Z. 10). Carlos reagiert dadurch, dass er die ausgestreckte Hand anblickt (s. Abb. 2). Eine solche fokussierte Aufmerksamkeit wird redebegleitenden Gesten nur selten gewidmet.

Carlos ergreift die Hand mit seiner Rechten und erhebt den Blick wieder. Damit hat er den erwarteten zweiten Schritt getan und die durch den Handschlag besiegelte Abmachung wird gültig. In den folgenden Frage- und Antwortsequen-

Abb. 2: "Hier meine Hand" (Don Carlos – Infant von Spanien. R.: Franz Peter Wirth. D 1984. Jacques Breuer (Don Carlos), Robert Atzorn (Marquis Posa))

zen wird der Inhalt des Bundes noch weiter spezifiziert. Die abschließende Umarmung beschließt diesen Teil des Gesprächs. Es ist wegen des gewählten Bildausschnitts nicht feststellbar, wie lange die beiden Hände in der Handschlagposition verharren. Da die beiden Männer nahe voreinander stehen bleiben, ist die Annahme jedoch naheliegend, dass während der erwähnten spezifizierenden Sequenzen (Z. 11-18) die Position der Hände beibehalten wird. Dieser so geschlossene Bund wird noch durch die intimere Geste der Umarmung unterstrichen, die zugleich Carlos' Freude über diese Entwicklung zeigt.

Auch in der zweiten von mir untersuchten Inszenierung des "Don Carlos", in der Version des Burgtheaters unter der Regie von Andrea Breth, wird die Geste des Handschlags in dieser Szene optisch ins Zentrum gerückt, allerdings auf andere Weise. Jedenfalls gilt das für die mir zur Verfügung stehende Verfilmung, die hier besonders stark die filmischen Mittel nützt und damit Wirkungen erzeugt, die von der Rezeptionssituation im Zuschauerraum deutlich abweichen. Die Gestik der beiden Protagonisten bleibt durch Großaufnahme der Gesichter über weite Strecken des Dialogs unsichtbar. Nur in der entscheidenden Szene wechselt der Bildausschnitt und es werden nur die beiden Hände gezeigt.

(2) Hier meine Hand (Breth 2005)

```
2   P:   wohl´AN? (-)

3        ich ´`WEIche; (---)
```

```
4          ´hIEr [(.) meine `HAND.]
           *  streckt re Hand zu C   *

5    C:          [ (--]            ]
                 *nähert seine Hand der Ps*

4    C:    (1.0) der `MEInige.
           Hände im Off

5    P:    (-) auf `Ewig. (---)
           Hände im Off

6          und in des `wOr:ts (.) ver`WEgenster bedeutung
           *    Hände im Off       * *              * *          ⇒
                                   re Hand zurückgezogen  ergreift Ps Hand, drückt sie
```

Abb. 3: Rückzug der angebotenen Hand (Don Carlos, Infant von Spanien. Inszenierung Andrea Breth. A/D 2005. Philipp Hauß (Don Carlos), Denis Petkovic (Marquis Posa))

Bei den Worten, mit denen der Marquis seine Hand bietet (Z. 2-4), fällt im Bereich der Prosodie eine große Ähnlichkeit zwischen den beiden Inszenierungen auf, was Akzentstellen und -typen sowie finale Tonhöhenverläufe betrifft. Damit entsteht auch in der Breth-Inszenierung eine Kookkurrenz zwischen prosodischer, inhaltlicher und syntaktischer Gliederung. Und wiederum deckt sich diese Gliederung zeitlich mit dem gestischen Ablauf (wobei eine etwaige *preparation* für die Geste des Handreichens in dieser Inszenierung aus Gründen der Kameraführung unsichtbar bleibt).

Dass der Handschlag wie oben erwähnt optisch deutlich hervorgehoben wird, ist eine Parallele zur zuerst besprochenen Szene in der Wirth-Inszenierung. Allerdings gibt es auch einen Unterschied: Der Marquis von Posa bietet auch hier zuerst seine Hand an, simultan zur entsprechenden verbalen Ankündigung, und Carlos reagiert darauf, indem er seine Hand der angebotenen nähert. Anstelle des Blicks von Carlos bedingt hier nun aber wie erwähnt der Bildausschnitt die Fokussierung auf die Geste. Dann wechselt der Bildausschnitt wieder zu den Großaufnahmen der Gesichter. Deshalb kann nicht mitverfolgt werden, ob der Händedruck erfolgt oder nicht. Als die Hände das nächste Mal sichtbar werden (ab *ver`WEgenster*), hat Marquis Posa (s. Abb. 3, links im Bild) seine Hand wieder zurückgezogen, um erst zum Ende seiner Äußerung hin Carlos' Hand zum Händedruck zu umfassen. Das legt nahe, dass er während der Formulierung der näheren Bedingungen des Bundes, insbesondere während der Spezifikation *und in des `wOr:ts* (.) *ver`WEgenster bedeutung* (diese Stelle wurde in der Wirth-Inszenierung gestrichen), sein Angebot noch einmal kurz aussetzt. So ist er es hier, der zuerst die Initiative ergreift, auf Carlos' Reaktion und Zustimmung wartet, und erst, als die Art des Angebots noch näher erläutert worden ist, mit seinem Händedruck die Geste abschließt. Das könnte so gedeutet werden, dass er stark darauf achtet, die Kontrolle über die Vorgänge zu behalten.

In beiden Inszenierungen jedoch ist der Handschlag als kooperative Geste klar performativ. Durch ihren Vollzug wird eine bestimmte Abmachung besiegelt oder anders formuliert: Wer die Geste initiiert, bietet eine Abmachung an. Wer auf diese Initiative eingeht und an der gemeinsamen Durchführung der Geste mitwirkt, nimmt damit das Angebot an. Im soeben besprochenen Beispiel werden sowohl die Initiierung der Geste verbal begleitet als auch der Inhalt der Abmachung parallel zur Durchführung der kooperativen Geste verbal thematisiert. Die genannten Phasen der kooperativen Geste haben zudem auch eine Entsprechung in der prosodischen Gliederung der begleitenden Äußerungen.

5.2. Eine Einladung wird akzeptiert

Die Eigenschaft von kooperativen Gesten, eine Übereinkunft auszudrücken, kann auch verstecktere Botschaften übermitteln. Betrachten wir dazu das folgende Beispiel, wieder aus "Don Carlos". Es handelt sich um eine Szene aus dem 1. Akt, 6. Auftritt. König Philipp hat seine Frau, Elisabeth von Valois, im Garten ohne Hofdamen angetroffen, was ihn zu schweren Vorwürfen an die Königin und zur ungnädigen Entlassung einer Hofdame veranlasst hat. Dabei war sein ganzer Hofstaat zugegen. Die Königin hat sich verteidigt ("Bloß Zwang bewacht die Frauen Spaniens? Schützt sie ein Zeuge mehr als ihre Tugend?") und die Vorwürfe zurückgewiesen. Sie ist schon vorher im Stück als freiheits- und menschenliebende Antagonistin des Königs, der ein gnadenloses Terrorregime mit Hilfe der Inquisition und seiner Generäle aufrechterhält, charakterisiert worden. Der König beendet die Szene, indem er eine kurze Ansprache an sein Gefolge richtet. Darin kündigt er die Rückkehr vom Sommersitz in Aranjuez nach Madrid an und nennt den Grund: die Hinrichtung von "Ketzern". Hier setzt das Transkript ein.

(3) Blutgericht (Wirth 1984)

(König Philipp spricht durchgehend eher laut, mit leicht erhobener Stimme. Er zeigt – soweit sichtbar – keine Gestik. Die Königin steht schräg hinter ihm und ist mit Körper und Blick zur Seite gewandt.)

```
 1   P:   ein ´schAUerndes ex`EMpel.=

 2        =soll die `Irrenden be`KEHren. [(1.5)]

 3   E:                                  [(1.5)]
                                             ◊
                                   wendet Blick und Körper zu P

 4   P:   ´drEIhundert `OPfer. (-)

 5        sind ´rEIf zum `TOD. [(1.3)]

 6   E:                       [(1.3)]
                                  ◊
                          schlägt Augen nieder

 7   P:   der ´rAUch von ihren [`FLAMmen] verkündige dem

 8   E:                       [(1.0)]
                                ◊
                            senkt den Kopf

 9   P:   `drEImal `hEIligen ¯gOtt- (--)

10        wie ↑`GLORreich [philipp seine `SCHULD entrichtet.
                                                     (1.5)]

11   E:                  [(3.0)                           ]
                        *blickt zu P                      ⇒

12   P:   dies ´blUtgericht soll ohne `BEIspiel sein. [(3.0)]
                                                        ◊
                                   dreht Kopf und teilweise Oberkörper nach links
13   E:                                                [(3.0)]
                                                        ◊
                     wendet Blick, dann auch Kopf und Oberkörper nach li von P weg
14   P:   mein `gAnzer `HOF.

15        (--) ist `fEIerlich ge`LAden.
          *hebt li Arm*

16        (4.0)
```

```
17   P:   [(10.0)]
           *    *
```

blickt ausgestreckten Arm an

```
18   E:   [(10.0)]
           *   *
```

blickt ausgestreckten Arm an, legt dann langsam ihre re Hand auf li Hand von P

Der König spricht mit unbewegter Miene und – so weit der Bildausschnitt diese Aussage zulässt – ohne Gestik. Prosodisch ist sein Sprechstil durch Tonhöhe und Lautstärke als offiziell gekennzeichnet. Die gegenüber der Normallage des Königs etwas erhöhte globale Tonhöhe und die größere Lautstärke sind geeignet, die Stimme weithin hörbar zu machen, und kontextualisieren seine Worte (zusammen mit der Körperorientierung) als Ansprache an eine größere Gruppe. Auch die vergleichsweise kurzen intonatorischen Phrasen und die hohe Akzentdichte, die langen Pausen und die mehrfache Abfolge des Akzentmusters "steigender Akzent gefolgt von fallendem Akzent" erhöhen den Eindruck einer maßvollen, bewusst gestalteten Ansprache. Im Kontrast zu dieser zur Schau getragenen Emotionslosigkeit stehen die Reaktionen der Königin. Sie fährt erschrocken auf, senkt den Blick, lässt dann den Kopf sinken, blickt noch einmal zum König und wendet sich dann mit gesenktem Kopf ab. Am Ende seiner Rede zeigt Philipp, dass ihm ihre Gegenwart, obwohl sie hinter ihm steht, bewusst ist. Er blickt aus dem Augenwinkel zur Seite, dann streckt er seinen linken Arm waagrecht, mit der Handfläche nach unten, schräg nach vorne und wartet. Hier liegt wieder eine Geste vor, die ein Angebot bedeutet. Der König bietet auf höfische Art seiner Frau den Arm. Diese feierliche Geste ist durch seine letzten Worte (*mein `gAnzer `HOF. (--) ist `fEIerlich ge`LAden.*), aber auch durch die prosodische Gestaltung der Ansprache vorbereitet. Diese letzten Worte, die die kurze Rede mit einer Einladung abschließen, sind durch die geänderte Körperorientierung (s. Ende von Z. 12) vermehrt auch an die Königin gerichtet. Prosodisch wird dieser letzte Satz als zwei intonatorische Phrasen realisiert. In der kurzen Pause, die die beiden Phrasen trennt, hebt der König wie beschrieben seinen Arm an. Damit fällt diese Geste mit einer prosodischen Zäsur zusammen. Andererseits besteht auch eine Gemeinsamkeit mit der syntaktischen Gestaltung: *mein `gAnzer `HOF.* projiziert eine Fortsetzung der Äußerung, und in gleicher Weise verlangt das Angebot des Arms als Initiierung einer kooperativen Geste eine Fortsetzung bzw. einen – allerdings gemeinsamen – Abschluss. Wie oben im Beispiel des Handschlags wird auch diese Geste dadurch fokussiert, dass in einer nun folgenden Pause sowohl der König als auch die Königin den ihr angebotenen Arm anblicken (s. Abb. 4). Elisabeths Mienenspiel und ihr langes Zögern zeigen, dass es ihr schwer fällt, die von ihr erwartete Reaktion zu zeigen. Schließlich legt sie ihre rechte Hand auf die Hand des Königs und lässt sich wegführen. Die Bedeutsamkeit dieses Augenblicks wird in der Inszenierung dadurch noch weiter erhöht, dass kurz gezeigt wird, wie die Mitglieder des Hofstaates einander 'beredte Blicke' zuwerfen. Somit wird die Zustimmung zu dieser Geste mehrfach mit Bedeutung 'aufgeladen'. An der Oberfläche stimmt die Königin nur zu, dem König dorthin zu folgen, wo er sie hinführen wird. Im Kon

Abb. 4: Der König bietet seinen Arm (Don Carlos – Infant von Spanien. R.: Franz Peter Wirth. D 1984. Rolf Boysen (König Philipp), Marita Marschall (Königin Elisabeth))

text der vorangehenden Ansprache des Königs jedoch stimmt sie implizit auch zu, den grausamen Exekutionen beizuwohnen bzw. sich nicht öffentlich dagegen aufzulehnen. Damit ist dieser Gestentyp hervorragend dazu geeignet, ihren inneren Konflikt und ihre erzwungene Unterwerfung anzudeuten.

Während die Inszenierung von Franz Peter Wirth durch die historischen Kostüme, die Dekoration und die Gestik der Personen das Stück tatsächlich im 16. Jahrhundert spielen lässt, verlegt Andrea Breth die Handlung in die Gegenwart. In derselben Szene tragen die anwesenden Männer weiße Bademäntel (Abb. 5) – man ist quasi im Wellnessurlaub – und geben sich zumindest vordergründig eher entspannt. Aus Platzgründen kann ich die entsprechende Stelle nicht ausführlich vorstellen. Interessant ist, dass auch hier die abschließende Einladung *mein ganzer ´hOf,* (-) *ist f´EIerlich ge↑LAden.* mit einer kooperativen Geste verknüpft wird: Während der kleinen Ansprache haben alle Männer Gläser mit Sekt erhalten, nun heben sie gemeinsam mit Philipp bei der intonatorischen Phrase *mein ganzer ´hOf,* den Arm etwas an und heben daraufhin bei der zweiten intonatorischen Phrase gemeinsam das Glas, wobei der Stroke der Bewegung (wenig überraschend) mit der stark akzentuierten Silbe (*ge*)*↑LA*(*den*) zusammenfällt. In beiden Bewegungsphasen des Glashebens setzt Philipp einen Sekundenbruchteil früher mit der Bewegung ein als sein Gefolge, initiiert also die kooperative Geste. Die in der Durchführung der Geste erkennbare Abfolge von zwei Schritten weist in der prosodischen Gestaltung eine Entsprechung auf: Der einfache Satz wird als zwei selbständige intonatorische Phrasen realisiert, wovon die erste einen weiterweisenden finalen Tonhöhenverlauf aufweist, während die zweite durch den stark

Abb. 5: "Mein ganzer Hof ist feierlich geladen" (Don Carlos, Infant von Spanien. Inszenierung Andrea Breth. A/D 2005. Sven-Eric Bechtolf (König Philipp), Nicholas Ofczarek (Herzog Alba), Bernd Birkhahn (Graf von Lerma) u.a.)

fallenden finalen Tonhöhenverlauf einen Schlusspunkt setzt. Mit dieser prosodischen Gliederung korrespondieren der erste Teil der Geste als vorbereitende Bewegung und der zweite Teil als eigentliche Durchführung des Glashebens und als Abschluss. Eine besondere Engführung erfährt diese Korrespondenz von Prosodie und Gestik im Stroke der Bewegung, der wie bereits erwähnt mit der prosodisch prominentesten Silbe (*ge*)↑*LA*(*den*) zusammenfällt.

Die Königin steht abseits von dieser Männerrunde und betrachtet sie niedergeschlagen. Durch die gemeinsame Durchführung dieser kooperativen Geste signalisieren die Männer des Hofstaats, dass sie die Einladung annehmen, und implizit wohl auch, dass sie mit den Hinrichtungen einverstanden sind. Von der zeitlichen Abfolge her gesehen ist interessant, dass gestisch hier gleichzeitig zur verbal formulierten Einladung die Zustimmung der Eingeladenen abgerufen und – nur einen Sekundenbruchteil später – erteilt wird. Bei genauerem Hinsehen sind außerdem Unterschiede in der Präzision und Deutlichkeit der Bewegung zu erkennen, die mit der inneren Übereinstimmung der Männer mit der Position des Königs zu tun haben, wie im weiteren Verlauf der Handlung ersichtlich wird. Das Angebot des Königs zur Mitwirkung an der Geste wird somit allgemein angenommen und bestärkt so die gemeinsame Haltung und den Gehorsam gegenüber Philipp; gleichzeitig lässt sich durch die Art, wie dies geschieht, auch zwischen rückhaltloser und eher halbherziger Zustimmung unterscheiden.

5.3. Ein höfischer Handkuss und die Bitte um Versöhnung

Ein letztes Beispiel aus "Don Carlos" zeigt einen Fall, in dem ein gestisches Angebot abgelehnt wird, wodurch die gemeinsame Durchführung scheitert. Es handelt sich um einen Ausschnitt aus dem 2. Akt, 1. Auftritt. Don Carlos erzwingt eine Aussprache mit König Philipp, seinem Vater. Das Verhältnis der beiden ist äußerst angespannt. Carlos hat seinen Vater als hartherzig und kalt erlebt, Philipp hält seinen Sohn für einen Schwächling und misstraut ihm. Carlos' Ziel in diesem Gespräch ist es, den Oberbefehl über die Truppen zu erhalten, die in das aufständische Flandern geschickt werden.

(4) Versöhnung! (Wirth 1984)

(König Philipp steht vor seinem Schreibtisch, Don Carlos einige Meter von ihm entfernt in der Nähe der Tür.)

```
1   C:   wa`RUM von ihrem `hErzen mich `SO lange
            *           *   ◊                ◊
         dreht Arme nach außen, HF oben    kleine Tn durch Senken ⇒

2        ver`STOßen vater.
                ◊     ◊

3        [°hh hh° °hh hh°]
         im Off

4   P:   [(2.5)]
         sieht C bewegungslos an ⇒

5   C:   °h <<f> ´wAs hab ich ge`TAN. > h° °hh
         im Off

6        (1.0)

7        °h ´jEtzt oder `NIE <<verhaucht> mein vater;
         *       geht mit großen Schritten zu P              ⇒

8        °hh h° °hh
                  *

9        °hh ver`SÖHnung. > °h     (--)
         *               *          ◊
         kniet nieder und ergreift Ps re Hand   küsst Ps Hand

10  P:   °h <<t> `lAss mich und [steh `AUF.]= >
                   ◊
                 will durch Zurücklehnen festgehaltene Hand C entziehen
```

```
11   C:                                [1.0]
                                       hebt Gesicht, schüttelt Kopf

12        <<h, f> =ver¯SÖHnung; >
          senkt Gesicht wieder zur Hand

13   P:   zu `KÜHN wird mir <<gepresst> dies `GAUkel`spIEl; >
                                        ◊
                                        entreißt C Hand und wendet sich ab

14   C:   °h <<f> zu ´KÜHN?
          * ergreift Hand Ps wieder, sieht zu ihm auf ⇒

15        die `LIEbe deines ´KINdes? > °h

16   P:   vollends `TRÄnen.

17        `UN<<gepresst>würdger `ANblick. >
                        ◊
                        entreißt C seine Hand, geht dann weg
```

Der höfische Handkuss kommt im Laufe des Stückes immer wieder vor. Die Hand zum Kusse zu reichen, ist gemäß dem spanischen Hofzeremoniell einerseits ein Zeichen der Gnade des Herrschers, gleichzeitig die Darstellung seines Herrschaftsanspruches. Die andere Partei kniet zum Kusse nieder und zeigt dadurch ihre Unterwerfung, erlebt die dargebotene Hand aber wohl auch als Huldbeweis. So reicht in dieser Inszenierung zum Beispiel die Königin Marquis Posa die Hand zum Kuss, der König seinen Generälen, der Großinquisitor gar dem König. In allen Fällen wird zuerst die Hand zum Kuss dargeboten, dann erst folgt als zweiter Schritt das Niederknien und der Kuss. Das ist in dieser Szene zwischen Carlos und seinem Vater anders. Carlos nähert sich dem König unaufgefordert, wirft sich auf die Knie, ergreift die Hand des Königs und zieht sie an seine Lippen (Abb. 6). Dadurch verletzt er die vorgeschriebene Abfolge. Dieses Verhalten entspricht auch seinen Worten (*j´Etzt oder `NIE*) und den prosodischen Eigenschaften seiner Äußerung: er atmet hörbar heftig aus und ein, spricht streckenweise mit verhauchter Stimme, teilweise laut und hoch und mit vielen starken Akzenten. Alles kann als Anzeichen für seine Emotionalität, Impulsivität und Entschlossenheit gedeutet werden. Die Geste misslingt jedoch. Philipp, der sie nicht initiiert hat, entzieht Carlos seine Hand und wertet dessen Verhalten auch verbal ab (*dies `gAUkelspiel*). Er ist prosodisch deutlich weniger emphatisch: Er spricht leiser und tiefer als Carlos, mit geringerer Akzentdichte und ohne das hörbare heftige Atmen und den verhauchten Stimmklang, die die Sprechweise von Carlos prägen. Stattdessen sind Philipps jeweilige Anstrengungen, seinem Sohn seine Hand zu entziehen, durch gepresste Sprechweise hörbar.

Abb. 6: "Versöhnung" (Don Carlos – Infant von Spanien. R.: Franz Peter Wirth. D 1984. Rolf Boysen (König Philipp), Jacques Breuer (Don Carlos))

Durch ihr Misslingen wird die Geste des Handkusses mit anderen Mitteln als in den vorherigen Beispielen fokussiert und aufgeladen. Carlos ist überschwänglich und hält sich nicht an die höfischen Regeln, Philipp ist gefühlskalt und misstrauisch und kann mit der unkonventionellen Annäherung seines Sohns nicht umgehen. In dieser Geste wird also im Kleinen das tragische Verhältnis der beiden abgebildet. Vorbereitet und verstärkt wird diese Wirkung durch die prosodische Gestaltung der Szene, die denselben Gegensatz zwischen dem jugendlichen Gefühlsüberschwang von Carlos und Philipps kalter, etwas verkrampfter Reaktion darauf abbildet.

In der Inszenierung von Andrea Breth, die an dieser Stelle eine umfangreichere Textfassung verwendet, lässt Philipp zuerst jovial einen Handkuss zu, dann aber lehnt er sich abweisend weg, als Carlos ihn mit dem Ausruf *ver`SÖHnung* umarmt. Ich kann hier nicht die Unterschiede zwischen den beiden Inszenierungen herausarbeiten. Eine Parallele besteht jedenfalls darin, dass Philipp seine Zurückweisung auch dadurch zum Ausdruck bringt, dass er das Mitwirken an einer kooperativen Geste verweigert.

5.4. Eine Waffe wird angeboten

Das letzte Beispiel stammt aus Lessings "Emilia Galotti", 4. Akt, 7. Auftritt. Die Szene spielt im Lustschloss des Prinzen von Guastalla. Die Gräfin Orsina, früher die Geliebte des Prinzen, hat soeben einsehen müssen, dass ihre Versuche, die Gunst des Prinzen zurückzugewinnen, gescheitert sind. Ihr steht der bürgerliche

Odoardo Galotti gegenüber. Er ist gerade in das Schloss gekommen auf der Suche nach seiner Tochter Emilia, deren Kutsche in der Nähe überfallen worden ist. Orsina informiert nun Galotti, dass der Überfall im Dienste des Prinzen geschehen ist, weil dieser in Emilia verliebt ist und ihre Eheschließung im letzten Moment verhindern wollte. Dabei ist der Bräutigam Emilias getötet worden und Emilia ist in der Gewalt des Prinzen. Galotti ist schockiert und wütend.

(5) Nehmen Sie! (Langhoff 1984)

(Galotti steht mitten im Raum, sucht mit beiden Händen etwas in seinen Taschen und sieht um sich. Die Gräfing Orsina sitzt, ihm halb den Rücken zugewendet, auf einem Sofa. Sie sieht ihm kurz zu, dann blickt sie gerade vor sich hin.)

1 O: `dAmit kann ich `AUShelfen.

2 [$_1$(4.0)]
sucht in einer Tasche, zieht ein Messer hervor und hält es mit re Hand nach oben

3 G: [$_1$(4.0)]
sieht ihr bewegungslos zu

4 O: [$_2$(11.0)]
hält das Messer regungslos nach oben, steht nach einigen Sekunden auf, Blick immer auf G gerichtet

5 G: [$_2$(11.0)]
sieht starr das Messer an, geht, als O steht, langsam zu ihr, löst erst ganz zum Schluss den Blick vom Messer und sieht O in die Augen

6 O: <<flüsternd> DA. > (1.2)

7 <<p> `NEHmen sie. >

8 G: [(3.5)]
senkt Blick wieder auf Messer, zieht langsam re Hand aus Tasche und umfasst damit re Faust (mit Messer) von O, dann schneller auch mit der li Hand, zieht sie dabei näher zu sich, blickt ihr in die Augen

9 O: [(3.5)]
bleibt ruhig stehen, bis G ihre Hand umfasst, legt dann rasch auch ihre zweite Hand um die Gs, lächelt, steht ganz nahe vor G

10 hh° (-) °h (---) h° °hh
blickt kurz auf das Messer, dann G in die Augen, wieder auf das Messer, dann in seine Augen

```
11        <<p> `nEhmen sie ge`SCHWIND. > [(1.5)]
                                          ◊
                                     schiebt Kinn vor

12   G:                                [(1.5)]
                                     *schüttelt leicht Kopf ⇒

13   O:   ? `Eh uns `jEmand ¯SIEHT-

14        (2.5)

15   G:   <<p, verhaucht> wer wieder `sAgt;
                          ◊
                     hebt Kinn leicht
                                 *schließt Hände enger*

16        dass du eine `NÄRrin bist. (1.0)
                                     ◊
                                    nickt

17        `dEr hat es mit `MIR zu tun. >
           *   kleine Nickbewegungen    *

18        [(7.0)]
          legt Messer zur Seite, ohne den Blick von O zu wenden, und schließt wieder
          die Hände um Os Hände

19   O:   [(7.0)]
          atmet erregt, seufzt auf, als sich ihre Hände wieder umfassen

20        <<p, verhaucht> `mIr wird die gelegenheit ver`SAGT;=

21        =ge`BRAUCH davon zu machen; (1.2)

22        `Ihnen wird sie nicht `FEHlen. [(1.0)]
23   G:                                  [(1.0)]
                                          ◊
                                rasches Kopfschütteln ⇒ (inkl. Z. 24)

24   O:   `dIEse ge`LEgenheit. (1.0)

25        un:d sie: werden sie er`GREIfen. >[(--)]

26   G:                                      [(--)]
                                         *heftiges Kopfnicken ⇒

27   O:   <<flüsternd> [die ERSte die BESte > (1.0)]

28   G:                [(2.2)                      ]
                        *       heftiges Kopfnicken        ⇒
```

```
29   O:   [wenn sie ein `MANN sind. °hh (--)]

30   G:   [(1.8)                            ]
          *      heftiges Kopfnicken        ⇒

31   O:   ´ICH, (1.1)

32        ich bin nur ein `WEIB. [(1.1)]

33   G:                          [(1.1)]
                                  *   ⇒
                          langsames Kopfschütteln

34   O:   `ABER. (1.0)

35        ´sO kam ich `HER. °hh

36        ´FEST ent`SCHLOSsen; [°hh h° °h]

37   G:                        [(1.5)]
                                *  *
                             Kopfnicken

38        [(22.5)]
          *      *
          senkt den Blick auf die umschlungenen Hände, löst sie dann ganz langsam von
          ihren, bis zuletzt den Kontakt der Hände mit ausgestreckten Armen
          aufrechterhaltend, hebt dann die Hände über den Kopf und wendet sich um

39   O:   [(22.5)]
          *      *
         folgt seinem Blick und sieht auch die Hände an, löst die Umschlingung mit ihm
         langsam, blickt weiterhin ihre Hände an, dreht sich weg und dann wieder zurück

40        `wIr.

41        `kÖnnen uns `ALles vertrauen.
```

Die Übergabe des Messers gibt der Darstellerin und dem Darsteller die Gelegenheit zu einer "großen Szene". Nach einer kurzen verbalen Ankündigung (Z. 1) wird das Angebot vorerst gestisch unterbreitet: durch das Präsentieren des Messers, das mehr als 10 Sekunden lang – also eine Ewigkeit in den Dimensionen eines Gesprächs – angeboten wird. Während dieser ganzen Zeit fällt kein Wort, Orsina erhöht nur ihre Präsenz, indem sie aufsteht. Galotti ist von der Waffe wie hypnotisiert, er wendet keinen Blick von ihr (s. Abb. 7), kommt näher und steht schließlich vor der Gräfin, blickt ihr in die Augen.

Abb. 7: Gräfin Orsina bietet eine Waffe an (Emilia Galotti. Inszenierung: Thomas Langhoff. D 1984. Rolf Boysen (Odoardo Galotti), Margit Carstensen (Gräfin Orsina))

Da kommt von ihr auch verbal das Angebot*: <<flüsternd> DA.* > (1.2) *<<p> `NEHmen sie.* > Prosodisch ist diese Äußerung markiert – zuerst flüsternd, dann immer noch leise. Dadurch wird die Situation als vertraulich und intim kontexualisiert. Galotti nimmt das Angebot an – er ergreift das Messer. Zugleich übernimmt er gestisch auch die weitere, prosodisch ausgedrückte Definition der Situation durch Orsina. Er hält seine Hände so vor seiner Brust – und ihre mit dem Messer in den seinen –, dass die Position der beiden an ein Liebespaar im vertrauten, zärtlichen Gespräch erinnert (s. Abb. 8). Orsina beteiligt sich aktiv daran, diese vertraute Position einzunehmen und erneuert ihr Angebot, noch immer leise sprechend (Z. 11-13). Galotti übernimmt ihren intimen Tonfall und versichert sie seiner Wertschätzung (Z. 15-17). (Als "Närrin" ist sie ihm vom Kammerherrn Marinelli vorgestellt worden, der die beiden allein lassen musste und vorbeugend verhindern wollte, dass Galotti Orsinas zu erwartende Anschuldigungen ernst nimmt.) Dann legt Galotti das Messer zur Seite, die eigentliche Übergabe ist vollzogen. Er behält aber die intime Haltung bei – so wie Orsina ihren Tonfall beibehält. Ihre Sprechweise ist nun auch verhaucht, unterbrochen von gut hörbaren tiefen Atemzügen – so als wollte sie ihn zur Rache nicht auffordern, sondern verführen. Die beiden blicken sich dabei ständig in die Augen, Galotti begleitet ihre Äußerungen mit emphatischem Kopfschütteln bzw. Nicken.

Da die Hände fixiert sind, können sie kaum gestikulieren, teilweise treten stattdessen Kopfbewegungen auf: Zwischen Z. 11 und 13 erzeugt Orsina mit einem Vorschieben des Kinns eine Grenze, Galotti signalisiert in Z. 15 mit einem leichten Anheben des Kinns die Übernahme der Sprecherrolle. Insgesamt wirkt

Abb. 8: Inniges Einverständnis an (Emilia Galotti. Inszenierung: Thomas Langhoff. D 1984. Rolf Boysen (Odoardo Galotti), Margit Carstensen (Gräfin Orsina))

die Position verbunden mit der leisen, verhauchten, von langen Pausen unterbrochenen Sprechweise Orsinas aber fast zeitlos, wie ein Verweilen in einer intimen Traumwelt außerhalb der sozialen Schranken und der aktuellen Bedrängnis. Dann blickt Galotti auf ihre verschlungenen Hände, ihr Blick folgt in dieselbe Richtung, und als würde ihnen durch den Blick bewusst, was sie tun, tauchen sie aus der Entrückung auf und kehren allmählich in die Realität zurück, lösen sich ganz langsam voneinander. Galotti hält die Hände über dem Kopf, als wolle er erneuten Kontakt vermeiden, sie blickt ihre Hände an, als sei sie noch im Bann der Begegnung.

Die kooperative Geste "ein Angebot durch Übergabe eines Gegenstandes machen und dieses Angebot akzeptieren" wurde in dieser Szene einerseits mit Mitteln hervorgehoben und aufgeladen, die bereits in den oben analysierten Szenen zu beobachten waren, vor allem durch (hier extrem langen) Stillstand im Gesprächsablauf und durch das Blickverhalten. Prosodisch deutet besonders Orsina das an sich martialische Angebot und die damit verbundenen Wünsche, Galotti möge (auch für sie) Rache nehmen, als Liebesgeflüster um, als große gegenseitige Verführung. Diese intime Stimmung wurde schon bei der Geste der Übernahme des Messers durch Galotti erzeugt, wie Abb. 8 zeigt (die Klinge ist direkt vor Galottis Kinn zu sehen).

Und ein letztes Mal in groben Umrissen der Vergleich mit einer weiteren Inszenierung, nämlich mit der ganz anders angelegten von Michael Thalheimer. Auch hier findet die Übergabe einer Waffe statt, einer Pistole, die auf Grund des vorherigen Geschehens am Boden liegt. Und auch hier reagiert die Gräfin auf Galotti, allerdings nicht auf sein Suchen in den Taschen, sondern auf seinen Blick. Er sieht die Pistole am Boden an, sie folgt seinem Blick (Blickverhalten als Ersatz

für eine deiktische Geste) und sagt dann mehrmals *`NEHmen sie*. Beide gehen zur Pistole, aber während sie sich langsam bückt, um die Waffe am Lauf zu ergreifen und ihm während des Sprechens den Griff hinzuhalten, sieht er die Pistole nicht mehr an, sondern nur noch sie. Er greift zwar mit einer Hand nach der Waffe, folgt aber mit der anderen mit einem kleinen Abstand behutsam den Umrissen ihres Körpers und küsst sie schließlich stürmisch. Erst als sie sich losgerissen hat und geht, sieht man, dass er in einer Hand tatsächlich die Pistole hält, während er sich mit der anderen den Mund wischt. Das Blickverhalten Galottis erweckt hier den Eindruck, dass es nicht so sehr das Angebot der Waffe ist, das ihn beschäftigt, sondern die erotische Ausstrahlung der Gräfin, der er schnell erliegt. In der Langhoff-Inszenierung ging dagegen, wie der Blick Galottis zeigte, die Faszination zuerst vom Messer aus. Erst aus der Annahme dieses Angebots entstand die auch erotisch konnotierte Nähe zur Gräfin. Dieser Unterschied passt sehr gut in das Gesamtkonzept der Thalheimer-Inszenierung: Dort werden die Männer generell so dargestellt, dass sie in erster Linie von erotischen Bedürfnissen (und Machtgelüsten) getrieben sind und den Frauen ziemlich ratlos gegenüberstehen. Das kann, wie wir gesehen haben, auch durch die Umdeutung einer kooperativen Geste ausgedrückt werden.

Abb. 9: "Nehmen Sie!" (Emilia Galotti. Inszenierung: Michael Thalheimer. D 2002. Peter Pagel (Odoardo Galotti), Nina Hoss (Gräfin Orsina))

6. Zusammenfassung: Kooperative Gesten und ihre Analyse in einem multimodalen/multikodalen Beschreibungsansatz

Ich habe den Terminus *kooperative Geste* vorgeschlagen, um eine Art von performativen Gesten zu beschreiben, für die die gemeinsame Durchführung durch

(mindestens) zwei InteraktionsteilnehmerInnen konstitutiv ist. Die hier analysierten kooperativen Gesten in verfilmten Theaterinszenierungen kommen jeweils in dramaturgisch wichtigen Szenen vor. Sie können ohne gleichzeitige verbale Äußerung auftreten, aber auch eingebettet sein in ein 'multimodales' Verhalten. Ihnen wird in der Interaktion erhöhte Aufmerksamkeit geschenkt, z.B. durch Pausen im Interaktionsablauf oder durch Blick auf die gestenausführenden Körperteile (meist Hand/ Arm). Eine weitere Hochstufung der Bedeutsamkeit erfahren diese Sequenzen durch prosodisch markierte Äußerungen. Dadurch werden diese Gesten, die durch ihren performativen Charakter (z.B. Besiegelung einer Übereinkunft) an sich schon wichtig sind, zusätzlich mit Bedeutung 'aufgeladen'. Kooperative Gesten eignen sich folglich sehr gut zur augenfälligen Darstellung von dramatischen Entwicklungen, Positionen und Konflikten. Darin werden sie in den hier untersuchten Beispielen durch prosodische Mittel unterstützt.

Die für die Kernphase der kooperativen Gesten typische Gleichzeitigkeit der Durchführung durch die TeilnehmerInnen ist in der initiierenden Phase so nicht unbedingt vorhanden. Vielmehr kommt häufig von einer Seite das Angebot zu einer solchen Geste, das – in dieser Hinsicht durchaus vergleichbar mit der Situation in verbalen Paarsequenzen – die Annahme dieses Angebots durch die andere Seite konditionell relevant macht. Und nicht anders als bei den meisten Typen von verbalen Paarsequenzen ist die Annahme die bevorzugte Variante, nicht die Verweigerung. Letztere ist stark markiert und deutet auf massive Nichtübereinstimmung in der Definition der Situation und/ oder der Beziehung der InteraktionsteilnehmerInnen hin. Auch diese Eigenschaft von kooperativen Gesten macht sie ideal geeignet für die Bühne. Sie dürften somit zum zentralen Repertoire von Theatergesten gehören.

Die Bedeutung von kooperativen Gesten im Theater kann nur dann sinnvoll erfasst werden, wenn ihr Zusammenwirken mit weiteren Ausdrucksmitteln – vor allem prosodischen, lexiko-semantischen, syntaktischen, proxemischen und blickbezogenen – untersucht wird. Umgekehrt ist aber auch der Schluss zu ziehen, und das kann nicht stark genug betont werden, dass sich die Bedeutung der betreffenden Szenen, in denen kooperative Gesten vorkommen, keinesfalls erschließen lässt, wenn nur der verbale Verhaltensanteil berücksichtigt wird. Deshalb wurde hier ein vielschichtiger und komplexer Beschreibungsansatz verwendet, bei dem die einzelnen Verhaltenselemente gesondert transkribiert und anschließend als Gesamtheit interpretiert werden. Dabei ist besonders wichtig, dass die verschiedenen Verhaltensmodi nicht isoliert betrachtet werden, sondern dass ihre vielfältigen Wechselwirkungen und Kookkurrenzen berücksichtigt werden. Das entstehende Gesamtkonstrukt hat Gestaltcharakter und modifiziert wiederum die Bedeutung der einzelnen Elemente.

Das Zusammenwirken von Prosodie und Gestik ist in diesem multimodalen Untersuchungsansatz ein wichtiger Detailaspekt. Wie bei anderen gestischen Phänomenen hat sich auch im Fall der kooperativen Gesten gezeigt, dass sie sowohl unter rein zeitlich-synchronem Aspekt als auch im Hinblick auf die Funktion bzw. Kontextualisierungsleistung sehr eng mit der Prosodie verknüpft sind. Prosodie befindet sich in gewisser Weise in einem Grenzbereich bzw. an einer Schnittstelle zwischen den sprachlichen und den gestischen Äußerungsanteilen.

Aus den hier vorgestellten Ergebnissen ergeben sich vielfältige Anknüpfungspunkte für die Untersuchung von kooperativen Gesten in der Alltagsinteraktion. Mögliche Fragen wären zum Beispiel: Wie häufig treten diese Gesten im Vergleich zum Theater auf? Wie sind sie in den Gesprächsablauf eingebettet? An welchen Stellen kommen sie vor? Welche Aktivitätstypen sind damit verbunden? Welche Formen des Zusammenwirkens von kooperativen Gesten mit anderen Ausdrucksmitteln (Lexiko-Semantik, Syntax, Prosodie, Blickverhalten, Proxemik) können beobachtet werden? Wie modifiziert dieses Zusammenwirken die Bedeutung/ Wirkung der kooperativen Gesten? Nicht zuletzt ist aus Sicht der Gestenforschung natürlich interessant, welche Implikationen die Dominanz der kooperativen Geste im Vergleich mit kookkurrierenden anderen Ausdrucksformen für die Theorie einer multimodalen bzw. multikodalen Interaktion hat. Umgekehrt ist aus Sicht der Prosodieforschung die Frage zu stellen, welche Bedeutung die enge Verknüpfung von Prosodie und Gestik bei diesem besonderen Gestentyp für den Stellenwert der Prosodie in der Gesamtheit der Ausdrucksmittel in der Interaktion und somit für eine theoretische Verortung der Prosodie hat.

7. Quellen

7.1. Filmverzeichnis

Don Carlos – Infant von Spanien (Friedrich Schiller). R.: Franz Peter Wirth. D: HR/BR/Taurus 1984. Fassung: VHS. 180 Minuten.

Don Carlos, Infant von Spanien (Friedrich Schiller). Inszenierung: Andrea Breth. Burgtheater Wien 2005. Fernsehregie: Peter Schönhofer. A/D: ORF/ZDF Theaterkanal/3sat 2005. Fassung: 2 DVDs. Arthaus Musik GmbH 2009. 198 Min.

Emilia Galotti (Gotthold Ephraim Lessing). Inszenierung: Thomas Langhoff. Münchner Kammerspiele 1984. Bildregie: Thomas Langhoff. D: ZDF 1984. Fassung: vom ZDF genehmigter Mitschnitt am 20.11.1984. 130Min.

Emilia Galotti (Gotthold Ephraim Lessing). Inszenierung Michael Thalheimer. Deutsches Theater 2002. Fernsehregie Hannes Rossacher. D: ZDF Theaterkanal/3sat 2002. Fassung: DVD. Theater Edition 2009. 79 Min.

7.2. Literaturverzeichnis

Bergmann, Jörg R. (1994): Kleine Lebenszeichen. Über Form, Funktion und Aktualität von Grußbotschaften im Alltag. In: Sprondel, Walter M. (Hrsg.): Die Objektivität der Ordnungen und ihre kommunikative Konstruktion. Für Thomas Luckmann. Frankfurt am Main: Suhrkamp, 192-225 (= Suhrkamp-Taschenbuch Wissenschaft 1140).

Breth, Andrea (2005): Don Carlos, Infant von Spanien. Burgtheater. ORF Produktion.

Brinker, Klaus/ Sager, Sven F. (2010): Linguistische Gesprächsanalyse. Eine Einführung. 5., neu bearbeitete Auflage. Berlin: Erich Schmidt Verlag (= Grundlagen der Germanistik 30).

Bühler, Karl (1934/ 1982): Sprachtheorie: Die Darstellungsfunktion der Sprache. Stuttgart/ New York: Fischer.

Dreischer, Anita (2001): "Sie brauchen mich nicht immer zu streicheln ...". Eine diskursanalytische Untersuchung zu den Funktionen von Berührungen in medialen Gesprächen. Frankfurt am Main: Peter Lang (= Arbeiten zur Sprachanalyse 39).

Efron, David (1941/ 1972): Gesture, Race, and Culture. Den Haag: Mouton.

Eibl-Eibesfeldt, Irenäus (1970): Liebe und Haß. Zur Naturgeschichte elementarer Verhaltensweisen. München: R. Piper.

Ekman, Paul/ Friesen, Wallace V. (1969): The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. In: Semiotica 1, 49-98.

Fischer-Lichte, Erika (1994): Semiotik des Theaters: Eine Einführung. Band 1: Das System der theatralischen Zeichen. 3. Auflage. Tübingen: Narr.

Fischer-Lichte, Erika (2010): Gesten im Theater. Zur transformativen Kraft der Geste. In: Wulf, Christoph/ Fischer-Lichte, Erika (Hrsg.): Gesten. Inszenierung – Aufführung – Praxis. München: Wilhelm Fink, 209-224.

Fricke, Ellen (2007): Origo, Geste und Raum. Lokaldeixis im Deutschen. Berlin/ New York: De Gruyter (= Linguistik – Impulse & Tendenzen 24).

Fricke, Ellen (2012): Grammatik multimodal. Wie Wörter und Gesten zusammenwirken. Berlin/ Boston Mass.: De Gruyter (= Linguistik – Impulse und Tendenzen 40).

Furuyama, Nobuhiro (2000): Gestural interaction between the instructor and the learner in origami instruction. In: McNeill, David (ed.): Language and Gesture. Cambridge: Cambridge University Press, 99-117.

Garfinkel, Harold (1967): Studies in Ethnomethodology. Englewood Cliffs, N.J.: Prentice-Hall.

Hall, Edward T. (1966): The Hidden Dimension. Garden City, New York: Doubleday.

Kendon, Adam (1980): Gesticulation and Speech: Two Aspects of the Process of Utterance. In: Key, Mary Ritchie (ed.): The Relationship of Verbal and Nonverbal Communication. The Hague: Mouton, 207-227.

Langhoff, Thomas (1984) Emilia Galotti. Münchner Kammerspiele. ZDF Produktion.

Lausberg, Hedda (o.J.). Module III. Functional Gesture Coding. Unpublished Manuscript.

McNeill, David (1992): Hand and mind. What gestures reveal about thought. Chicago/ London: The University of Chicago Press.

Michaels, Axel (2010): Göttern schüttelt man nicht die Hände. Grußgesten im hinduistischen Ritual. In: Wulf, Christoph/ Fischer-Lichte, Erika (Hrsg.): Gesten. Inszenierung – Aufführung – Praxis. München: Wilhelm Fink, 91-103.

Montagu, Ashley (1971): Touching: The Human Significance of the Skin. New York/ London: Columbia University Press.

Müller, Cornelia (1998): Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich. Berlin: Berlin-Verlag Spitz (= Körper, Zeichen, Kultur 1).

Sacks, Harvey/ Schegloff, Emanuel/ Jefferson, Gail (1974): A Simplest Systematics for the Organization of Turn-taking for Conversation. In: Language 50 (4), 696-735.

Scherer, Klaus R. (1979): Kommunikation. In: Scherer, Klaus R./ Wallbott, Harald G. (Hrsg.): Nonverbale Kommunikation. Forschungsberichte zum Interaktionsverhalten. Weinheim/ Basel: Beltz, 14-24.

Schönherr, Beatrix (1997): Syntax – Prosodie – nonverbale Kommunikation. Empirische Untersuchungen zur Interaktion sprachlicher und parasprachlicher Ausdrucksmittel im Gespräch. Tübingen: Niemeyer (= Reihe Germanistische Linguistik 182).

Schönherr, Beatrix (1999): So kann man das heute nicht mehr spielen! Über den Wandel der sprecherischen Stilideale auf der Bühne seit den 60er Jahren. In: Pümpel-Mader, Maria/ Schönherr, Beatrix (Hrsg.): Sprache – Kultur – Geschichte. Sprachhistorische Studien zum Deutschen. Innsbruck: Innsbrucker Beiträge zur Kulturwissenschaft. Germanistische Reihe, 145-170.

Schönherr, Beatrix (2005): Wie viel Gestik braucht eine Fernsehmoderatorin? Nonverbale Redegliederung und Akzentuierung in Fernsehnachrichtensendungen. In: Osnabrücker Beiträge zur Sprachtheorie 70, 171-191.

Schönherr, Beatrix (2014): Die Methode ‚Datensitzung' und ihre Nutzbarmachung für die Erforschung von Theatergestik: „Ich sehe was, was du nicht siehst ... oder doch?" In: Schwarze, Cordula/ Konzett, Carmen (Hrsg.): Interaktionsforschung: Gesprächsanalytische Fallstudien und Forschungspraxis. Berlin: Frank & Timme, 177-209.

Schönherr, Beatrix (im Druck): Categories and functions of posture, gaze, face, and body movements. In: Müller, Cornelia et al. (eds.): Body – Language – Communication/ Körper – Sprache – Kommunikation. Vol. 2. Berlin/ New York: Mouton de Gruyter, 1333-1341 (= Handbücher zur Sprach- und Kommunikationswissenschaft/ Handbooks of Linguistics and Communication Science 38/2).

Selting, Margret/ Couper-Kuhlen, Elizabeth (2001): Forschungsprogramm "Interaktionale Linguistik". In: Linguistische Berichte 187, 257-287.

Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 10, 353-402 (www.gespraechsforschung-ozs.de).

Thalheimer, Michael (2002): Emilia Galotti. Deutsches Theater. ZDF Theater Produktion.

Wallbott, Harald G. (1981): Subjektive und objektive Aspekte gestischen Verhaltens: Pilotuntersuchungen an psychiatrischen Patienten. In: Winkler, Peter (Hrsg.): Methoden der Analyse von Face-to-Face-Situationen. Stuttgart: Metzler, 285-301.

Wirth, Franz Peter (1984): Don Carlos. Fernsehspiel. ZDF Produktion.

Wundt, Wilhelm (1911): Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte. Erster Band: Die Sprache. 3., neu bearbeitete Auflage. Leipzig: Wilhelm Engelmann.

8. Abkürzungen und Zeichen

In den Zeilen zur Notation der Gestik verwendete Abkürzungen und Zeichen:

HF	Handfläche
li	links/linke(r/s)

re	rechts/rechte(r/s)
T/Tn	Taktzeichen/baton (Sg./Pl.)
◊	Zeitpunkt eines punktuellen Ereignisses (jeweils in der Zeile darunter näher erläutert)
* *	zeitliche Erstreckung eines Verhaltens (zwischen den Zeichen erläutert, bei Platzmangel in der Zeile darunter)
* ⇒	Verhalten dauert über den angezeigten Zeitraum hinaus an (2. Zeichen steht am Zeilenende)

Mag. Dr. Beatrix Schönherr
Institut für Germanistik
Universität Innsbruck
Innrain 52
6020 Innsbruck
Austria
beatrix.schoenherr [at] uibk.ac.at

Prosody as dialogic interaction[1]

Elizabeth Couper-Kuhlen

Abstract
In this paper I will be arguing that in a very real sense participants' prosodic behaviors in focused encounters are inter-calibrated with one another and that this is one important means of bringing off social interaction. I will present some new findings in this vein based on an empirical analysis of the freestanding particle *oh* in response to informings and of the freestanding particles *okay/ alright/ sure* in response to requests in everyday American English conversation. In conclusion I will touch on the methodological challenges of studying prosody dialogically and outline some of the analytic pay-offs of viewing prosody this way.

Dialogism – Particle – Informing – Request – Prosodic Orientation – Upgrade – Downgrade – Transition Timing

Dieser Beitrag argumentiert, dass das prosodische Verhalten der TeilnehmerInnen in fokussierter Interaktion in einem sehr buchstäblichen Sinne aufeinander abgestimmt ist und dass dies eine wichtige Ressource zur Herstellung sozialer Interaktion ist. Er präsentiert einschlägige neue, empirische Befunde zur Analyse der freistehenden Partikel *oh* als Antwort auf Mitteilungen (*informings*) sowie der Partikeln *okay/ alright/ sure* als Antworten auf Bitten (*requests*) in amerikanisch-englischer Konversation. Der Beitrag schließt mit einer Diskussion der methodologischen Herausforderungen, die die dialogische Untersuchung von Prosodie mit sich bringt, sowie des analytischen Nutzens solch eines Verständnisses von Prosodie.

Dialogismus – Partikeln – Mitteilungen – informings – Bitten – requests – prosodische Orientierung – upgrade – downgrade – zeitliche Koordination von Sprecherwechseln

[1] Earlier versions of this paper were presented at the annual meeting of the Verein für Gesprächsforschung (Mannheim, 2013), the CA Workshop on Responding (Singapore, 2014) and the Center of Excellence on Intersubjectivity Seminar on Prosody (Helsinki, 2014). I am grateful to the participants at these meetings for constructive comments and suggestions. My thanks also to the editors of this volume, who have been instrumental in helping me sharpen the argument. All remaining shortcomings are my own.

1. Rethinking prosody dialogically

In a recent volume with the subtitle "Interactional and contextual theories of human sense-making" Per Linell (2009) makes a strong plea for "rethinking language, mind and world dialogically", by which he means that we must acknowledge 'the other' as being constantly present whenever the individual engages in thinking or communication. This other-orientation is all-pervasive, not only in situated interactions but also in how the mind works: in short, it is an attribute of human sense-making.

What might it mean to think of *prosody* dialogically? A recent paper by Goodwin et al. (2012) hints at what is involved. They write with respect to the prosodic production of a response cry in the data fragment they are analyzing:

> It requires an analytic framework that extends beyond the voice of the actor producing the prosody to encompass the target being responded to and operated on. It is an interactive, dialogic action rather than the expression of something internal to a single individual. (Goodwin/ Cekaite/ Goodwin 2012: 23)

This highly suggestive remark holds the key to a dialogic view of prosody. The crux of the matter is that prosody, or rather its sense-making import in interaction, must be conceptualized as residing not in the voice of the individual but in the way that individual voice relates to the voice of the co-participant.

This is basically what Szczepek Reed (2006) had in mind with the term *prosodic orientation*. She describes prosodic orientation as "…the conversational activity of displaying awareness of another speaker's prosody in the prosodic design of one's own next turn" (2006: 33f). In this paper I will be building on the notion of prosodic orientation and showing how speakers orient to another's prosody, e.g., by *matching, upgrading* or *downgrading* the pitch and loudness of an interlocutor's prior turn, and/ or by adjusting the timing of their next-turn incoming so that it will be *well-timed, early,* or *late* with respect to the pacing of an interlocutor's prior turn. I will also be showing how these kinds of prosodic orientation become interpretable in specific sequential contexts.

Dialogic prosody makes itself especially noticeable in responsive position. In this sense the present study can be seen as building on earlier work by Müller (1996), Couper-Kuhlen (1996), Wells (2010), and Gorisch et al. (2012), where various types of prosodic repetition in responsive position are discussed. However, it goes one step further by bringing in a *relational* aspect. It argues that what is relevant is not just whether the prosody of a response is the same or different from that of the turn it is responding to, but also whether the response displays more or less of some prosodic feature in the prior turn and whether it is earlier or later than the beat anticipated from that turn. In this respect it views prosody as emerging from, and indeed constituting interaction with prior talk.

In the following I examine the dialogic prosody of freestanding particles when used to build preferred responses ("+" actions in Schegloff's (2007) terminology) to (1) news deliveries and informings, and (2) requests for action in American English conversation. This work arose as part of a large-scale study on the grammar of responsive actions in everyday conversational interaction (Thompson et al.

forthc.). It was in conjunction with this joint project that the need for a dialogic perspective on prosody imposed itself on me anew.[2]

2. Particle responses to informings

With an 'informing' Thompson et al. (forthc.) refer to a turn in which the speaker assumes a K+ (knowing) stance vis à vis a recipient who, accordingly, is cast in a K- (unknowing) position (Heritage/ Raymond 2005). Informings in this understanding include news deliveries as described by Maynard (1997, 2003) as well as also other sorts of tellings as described, e.g., by Jefferson (1978, 1988). The preferred, or "+" action, response to an informing is a claim by the recipient that whereas they were previously uninformed about the matter in question, they are now informed, i.e. they now know something they did not know before. As Heritage's (1984) seminal work has shown, the prime way to do this in English is with the particle *oh*. With a news-receipt *oh*, recipients propose to have undergone a change of state from not knowing (K-) to now knowing (K+). Building on earlier studies of the prosody of *oh* by Local (1996) and Reber (2012), I will argue that it is also the *dialogic* prosody of a freestanding news-receipt *oh* that is interactionally relevant, i.e. how its prosody relates to the prosody of the co-interlocutor's informing turn.

The particle *oh* is found as a response to at least two different kinds of informings in English conversation. In one case the informing is simply *volunteered*, with or without a more or less elaborate preliminary sequence or 'pre-' (Maynard 1997, Schegloff 2007), often as part of a longer telling. The news-receipt particle *oh* then appears as a response in second position. Here is a fragment where this happens twice:

(1) "Bud just left" (NB 008)[3]

(From a telephone call between Emma and her friend Nancy. Bud is Emma's husband. We join the call soon after Nancy has announced that she has had a raise. When Emma suggests that she should go shopping with her new money, Nancy replies with a smile "Well I should but you know at eight dollars a month anything I'd buy'd be using up my raise for half a year". The excerpt begins immediately after Emma agrees with "Yeah".)

```
->  1 Emm: .hhhhh[Bud just left]to play go:lf he's gotta go to Riverside=
    2 Nan:       [   °Y e  a h°]
=>  3 Nan: =[↓ O h : .  ]
    4 Emm: =[on a compan]y dea:l so, .t.h[hhhhh
    5 Nan:                               [Oh::?
    6 Emm: ↑GOD [it's be-]
    7 Nan:      [To River]side toda:y?
    8 Emm: .hhh Yeah they: they're gonna tee off at twelve it's a company
    9      dea:l so (.) the couple was supposed to come do:wn to(.)la:st
   10      ni:ght 'n you know k-Harry and Kathryn they're uh k-
-> 11      cuz Harry was gonna play k-
=> 12 Nan: Oh[:.
```

[2] Earlier explorations into dialogic prosody are reported in Auer et al. 1999; Couper-Kuhlen 1992, 1993, 1996, 2009b; Schegloff 1998; and Ogden 2006.

[3] The transcription system used in this and all following excerpts (with the exception of (10)) is Jeffersonian (see, e.g., Jefferson 2004) with standardized orthography. The single arrow refers to the informing, the double arrow to its receipt.

```
13 Emm:   ['n company and then .hhh there was a death in their fa:mily
14      so: (.) [.hhh
15 Nan:         [Aww:::.
16     (.)
17 Emm: ↑THE:Y gosh uh this has really been a wee:k ha:sn't it?=
```

The two cases in question occur in line 03, where Nancy produces an *oh* in response to Emma's out-of-the-blue informing that Bud has left to play golf in Riverside, and in line 12, where Nancy produces another *oh* in response to the volunteered information that actually Harry was going to play golf on the company deal.

But a news report or telling can also be *question-elicited*. In this case, the 'news' appears in second position and the news-receipt *oh* comes in third position. Here is a fragment where two pieces of news are presented in response to a query and each is receipted with *oh*:

(2) "Came down Friday night" (NB 007)

(From a telephone call between Emma and her sister Lottie. Lottie lives at Newport Beach, where Emma also has a vacation flat that she rents out when not using it herself. Bud is Emma's husband.)

```
    1 Emm: ...morning.
    2 Lot: Well WHERE'VE YOU BEE:N.
    3 Emm: .hhhh OH I'VE BEEN DOW:N HE:RE,
    4      (0.2)
    5 Lot: I was down there over:: Memorial Day'n you weren't
    6      the:re.
->  7 Emm: Oh I wasn't here Memorial: no buh- Bud had to WORK
    8      Fri:day.
    9      (0.4)
=> 10 Lot: .k Oh::::[:::.  ]
   11 Emm:          [And he] had to] take a]
   12 Lot:                   Go:d  ] I went] do:wn there=
   13      =I think it w-Yeah it was Fri:day.h
   14      (0.2)
   15 Emm: .hh Well I was here Sa:turday: uh: (0.5) uh let's see oh
   16      I came down: uh Friday ni:ght,
   17      (0.3)
   18 Lot: Yah:ah.
-> 19 Emm: Well Bud had to play go:lf uh Thursday.
   20      (.)
   21 Emm: So he [didn't take] Sa-uh f-] Friday o:ff s[o
=> 22 Lot:       [O h : : :] :_: : : . ]              [Yeh rode down
   23      muh my bi:cycle th[ere en:nu:h h]uh=
   24 Emm:                   [O h : : :_:? ]
   25 Lot: =was nobody wa(h)s the↑::re.
```

In line 02 Lottie asks where Emma has been and on learning that Emma has been down at her beach flat, delivers in line 05 what Pomerantz (1980) has called a 'my-side' telling: she reports that she came down to look for Emma on Memorial Day but didn't find her there. This turn is question-like in that it elicits information from Emma about where she was on Memorial Day. Emma's subsequent information, namely that Bud had to work Friday, which meant that she couldn't come down for Memorial Day, is receipted by Lottie with *oh* (line 10).

When Lottie now reiterates her my-side telling about coming down on Friday (lines 12-13), Emma furnishes more information about her situation: she specifies

when she got down to the beach (lines 15-16) and explains that Bud couldn't take Friday off because he had to play golf on Thursday (line 19). This information too is receipted by Lottie with an *oh* (line 22). Both these *oh*s (in lines 10 and 22) are produced as news receipts by the same party (Emma) who elicited the information from the co-participant (Lottie).

Observe that in (1) and (2) there are other tokens of *oh* that do not function as news receipts: see, e.g., line 05 in (1) and line 24 in (2). These *oh*s are alike in having overall rising intonation, in contrast to news-receipt *oh*s (see above), which have overall falling intonation. As Thompson et al. (forthc.) argue, rising-pitch *oh*s, do not position the speaker as someone who is in the know, but rather as someone who doubts what they have just been told.[4] In the following we focus exclusively on *oh*s with overall falling intonation that function as news receipts.

Yet, overall falling news-receipt *oh*s do not all sound alike. Some are short, others are long; some are peaked, others are flat; some end low in the speaker's voice range, others end mid; some have an overlay of breathy voice, others have creaky voice. Prosodic features such as these can be described independently of the specific context in which *oh* appears. However, I would like to argue that there are further distinguishing features in the delivery of a news-receipt *oh* that can only be appreciated if they are calibrated with the prosodic context in which the *oh* token appears. It is these 'dialogic' prosodic features (I will argue) that contribute significantly to the interactive work that news-receipt *oh* particles do.

Indeed, detailed interactional analysis shows that many news-receipt *oh*s accomplish more than merely receipting new information: they display their speaker's cognitive-affective stance towards the informing and/or the 'consequential figure' involved (see also Freese/ Maynard 1998, Wilkinson/ Kitzinger 2006, Couper-Kuhlen 2009a, Reber 2012, Golato 2012, Maynard/ Freese 2012, Koivisto forthc.).[5] In such cases, a freestanding news-receipt *oh* does not close down the sequence but instead invites more talk to deal with the stance displayed. Much as Reber (2012) has pointed out with respect to minimal responses in British English informing sequences, the type of stance displayed appears to depend on the specifics of the prior sequence.

The news-receipt *oh*s produced in response to *volunteered* informings range in cognitive-affective coloring from interest and/ or surprise to various forms of empathy or sympathy. Participants make these dimensions interpretable through the dialogic prosody of the *oh* delivery in relation to the informing (see also Freese/ Maynard 1998, 2012). Participant interpretations become analytically accessible through what happens next.

For instance, in lines 10-12 of (1), Emma volunteers the information that Harry and Kathryn were scheduled to come down and that Harry was supposed to play golf on the company deal:

---

4 An initial appreciation of this can be had by examining what happens after Nancy's rising-pitch *oh* in line 5 of (1): Emma breaks off her next turn *↑GOD [it's be-]* (line 6) to address Nancy's doubts by providing more information about Bud's trip to Riverside (lines 8-13).

5 See Maynard (1997:94) for more on the notion of consequential figure and its relevance to news deliveries. Reber (2012) also finds consequential figures relevant for the work that news-receipt *oh*s do.

(3) "Harry was gonna play" (excerpt from (1))

```
    8 Emm: .hhh Yeah they: they’re gonna tee off at twelve it's a company
    9      dea:l so (.) the couple was supposed to come do:wn to(.)la:st
   10      ni:ght ‘n you know k-Harry and Kathryn they’re uh k-
-> 11      cuz Harry was gonna play k-
=> 12 Nan: Oh[:.
   13 Emm:   ['n company and then .hhh there was a death in their fa:mily
   14      so: (.)
```

The news-receipt *oh* that Nancy produces in line 12 is slightly lengthened and has a (rising-)falling contour. But just as important is how it is calibrated with the turn it is responsive to. For one, it is on time, i.e. it occurs without gap or overlap after the unit concerning Harry's playing comes to a point of possible completion. More precisely, it is the timing of accented syllables in Emma's informing that sets up a regular beat (see, e.g., Auer et al 1999) that can be expected to continue; Nancy produces her *oh* on the next regular beat after Emma reaches a transition relevance point (TRP) in her turn. Rhythmically, this could be represented as follows, where an accent mark (‘) indicates the speaker's accented syllables and slashes aligned underneath one another indicate that the temporal intervals created by these accents are roughly equivalent in duration:[6]

(3´) Rhythm of Nancy's incoming in line 12 of (3)

```
   11 Emm:   cuz   /’Harry was gonna   /
                   /’play              /
   12 Nan:         /’oh:
```

In addition to its timing, the volume on Nancy's *oh* (line 12) is *softer* than in Emma's informing, but its overall pitch is *higher* and its pitch span *greater*. This can be appreciated from a side-by-side combination of Praat pictures showing the pitch configuration of the informing and of its response separately, each scaled relative to the speaker's individual pitch range:[7] Figure 1 shows that Nancy's pitch contour on *oh* is both higher and wider than the overall pitch configuration of Emma's informing. Whereas Emma's pitch span in this segment is 8.5 semitones, Nancy's is 11.1 semitones. We will refer to such an increase in pitch height and width as a pitch *upgrade* vis à vis the prior turn-unit.[8]

How is Nancy's well-timed but pitch-upgraded *oh* to be interpreted? To answer this, we must take into consideration the immediate sequential context in which it is produced. Let us first note that Emma does not treat the information she is providing as a surprise source (Wilkinson/ Kitzinger 2006); that is, her turn-unit is not explicitly built to elicit surprise from her interlocutor. Instead, the information

[6] See Auer et al. (1999) and Couper-Kuhlen (1993, 2009b) for more on conversational rhythm and its notation.

[7] The values used for Emma's overall pitch range were: 421-128 Hz; those for Nancy's were 487-119 Hz. These values represent an average of the three highest and lowest Hz measures for each speaker in a one-minute sample of speech.

[8] Both Curl 2005 and Ogden 2006 speak of 'phonetic' upgrading for a cluster of features including increased pitch span. See also Plug, this volume.

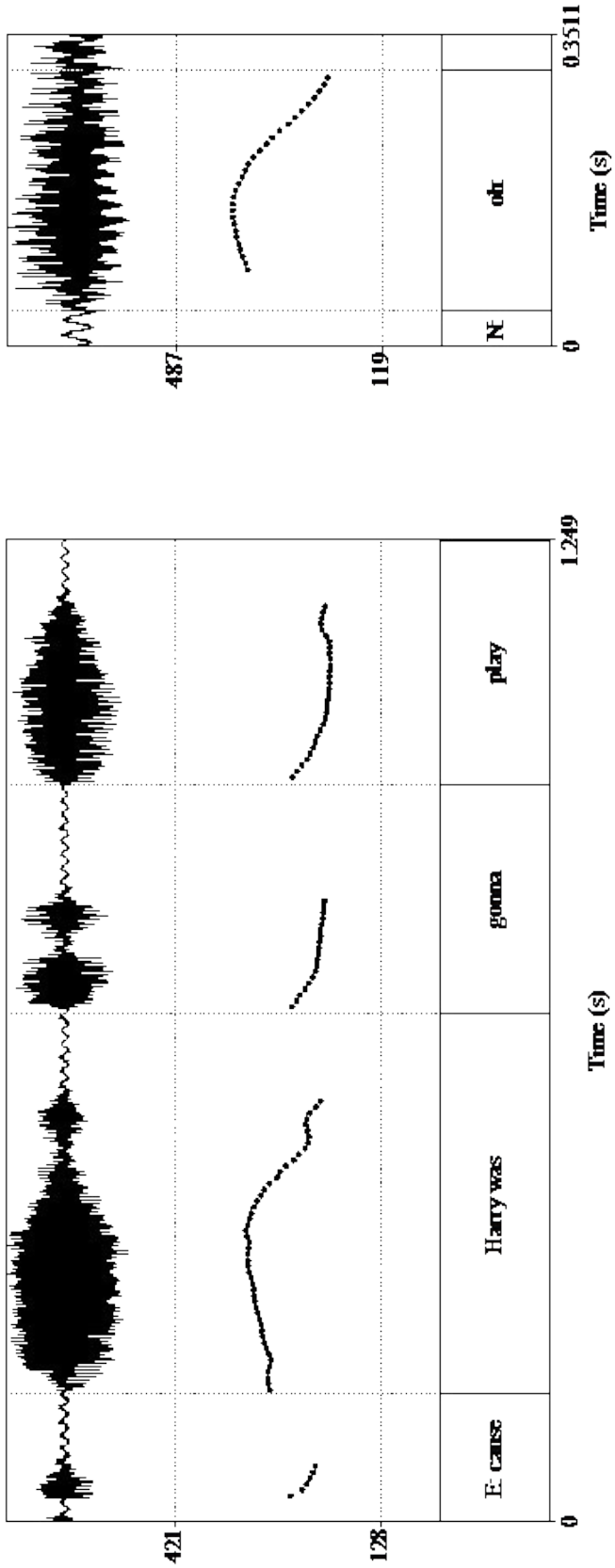


Figure 1: Individually scaled pitch contours for lines 11 and 12 in (3)

appears to be given in a by-the-way fashion. Moreover, it is relatively innocuous in the sense that it does not foreshadow positive or negative consequences for either of the co-participants. Given these considerations, one working hypothesis might be that Nancy's options for next turn include displaying interest or the absence thereof. The well-timed, pitch-upgraded *oh* produced could then be said on iconic grounds to display (mild) interest, encouraging Emma to tell more. In fact, this is precisely what Emma proceeds to do (lines 13-14). She does not break off in order to provide a repair, as she does in line 06 subsequent to Nancy's *oh* in line 05.

Consider now the falling-intonation *oh* produced earlier by Nancy in response to a different piece of news volunteered by Emma earlier in the same fragment:

(4) "Riverside" (excerpt from (1))

```
-> 1 Emm: .hhhhh [Bud just left]to play go:lf he's gotta go to Riverside=
   2 Nan:        [ °Y e  a h°  ]
=> 3 Nan: =[↓O h : .   ]
   4 Emm: =[on a compan]y dea:l so, .t.h[hhhhh
   5 Nan:                               [Oh::?
   6 Emm: ↑GOD [it's be-]
```

In response to Emma's announcement that Bud has just left to play golf in Riverside (line 01), Nancy also produces a slightly lengthened *oh*. This news receipt, however, sounds quite different from the one in line 12 and stands in a different relation to the informing it is responsive to. For one, compared to the timing of Emma's turn in line 01, Nancy's *oh* in line 03 comes in *before* the next beat:

(4´) Rhythm of Nancy's incoming in line 03 of (4)

```
   1 Emm:          /'Bud just left to play  /
                   /'golf he's gotta go to  /
                   /'Riverside==on a        /
   3 Nan:    /'oh:                                          (early)
   4 Emm:          /'company deal so
```

Here Nancy is orienting to a possible TRP at the end of line 01: Emma produces some final lengthening on *Riverside*, which also carries the main accent in the unit *he's gotta go to Riverside*. But rather than wait for the next rhythmic beat, Nancy comes in beforehand with her *oh*. In the event it ends up overlapping Emma's *on a company deal so* (line 04), which is produced as an increment to the possibly complete unit at the end of line 01.

Moreover, compared to the volume and pitch of Emma's prior turn, Nancy's *oh* is audibly softer as well as lower, and flatter: see the side-by-side Praat picture (Figure 2). As Figure 2 shows, this *oh* by Nancy begins lower than any of the pitch accents in the prior turn by Emma and its range is more compressed in comparison to that of Emma's informing.[9] We might then say that Nancy's *oh* in line 03 has a pitch *downgrade* vis à vis Emma's volunteered informing.

[9] Nancy's pitch span here is 1.7 semitones, compared to Emma's, which is 8.9 semitones.

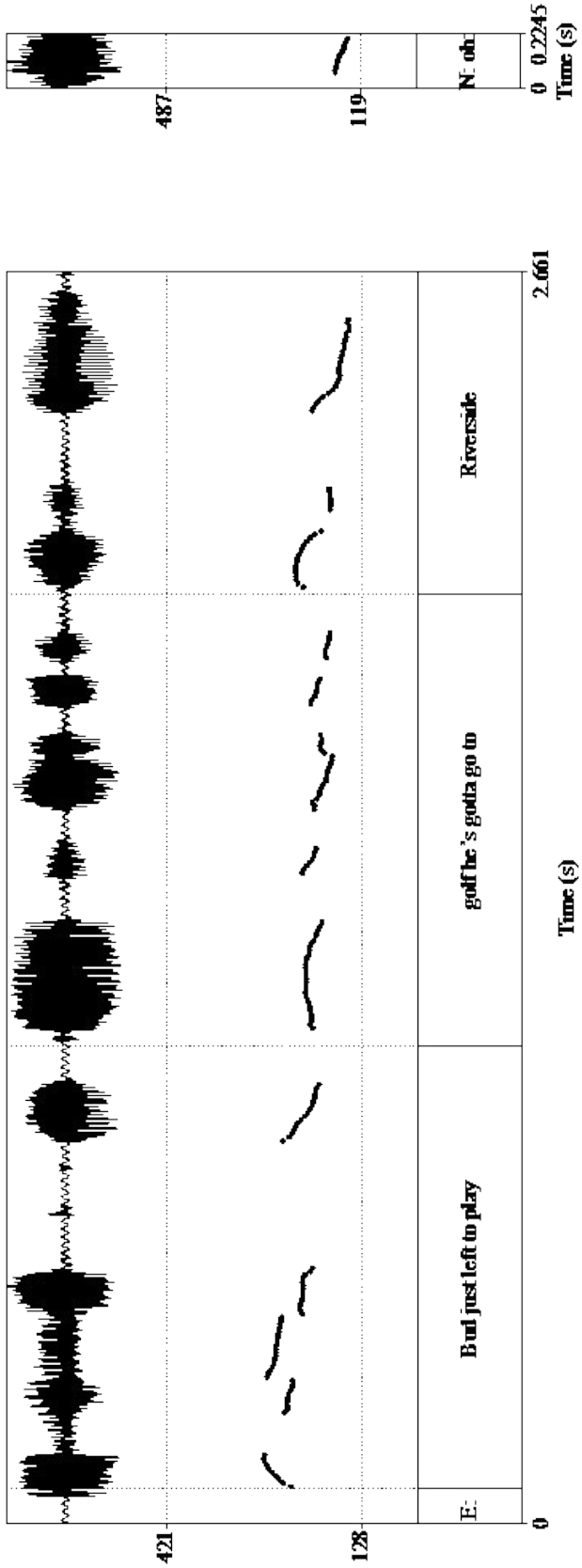


Figure 2: Individually scaled pitch contours for lines 1 and 3 in (4)

In sum, this news receipt has early timing in relation to the prior informing and its pitch and volume are downgraded. These features, we would argue, are just as 'marked' as are those we saw on the upgraded *oh* in line 12 of (1): they do not happen incidentally nor are they neutral in effect. Instead, they are *achieved* through calibration with the prosody of the prior turn and are indexically interpretable in the context created by this turn. To appreciate what this *oh* is doing with its dialogic prosody, we must appeal once again to its sequential location. Emma's turn in line 01 is delivered at so-called 'anchor position' in this conversation (Schegloff 1986), i.e. at a point after the exchange of greetings and how-are-you's where a caller's first topic, the reason for the call, can be expected. Emma's turn in line 01 is thus doing 'bringing up to date'. But at the same time it delivers information that is framed as having potential negative consequences for Emma: in her words, *he left* and *he's gotta go*. This suggests that Bud's going to play golf in Riverside is an incipient complainable. Nancy's *oh* is well-fitted to such a reading: with its soft volume and pitch downgrade, it is hearable not as a display of interest (as in (3)), but rather as a display of something akin to sympathy or empathy (see also Reber 2012). That Emma is indeed intent on doing complaining at this point will be seen from the way she continues: in overlap with Nancy's *oh* she trails off [10] (line 04) and, now in speakership mode, moves into what is arguably a further complaint with *↑God it's be* (line 06). The complaint character of this (aborted) turn can be seen by comparing Emma's resumption of the topic once the insert sequence, occasioned by Nancy's other-repair initiation in lines 5 and 7, has been completed: *gosh this has really been a wee:k ha:sn't it?,* said in a lamenting tone of voice (line 17).

What we have seen so far then is that dialogic prosodic features can create a cognitive-affective lamination on news-receipt *oh* that becomes interpretable in the particular sequential context of the informing. The point has been that in addition to effects such as, e.g., a rise-fall pitch contour, syllable lengthening, and breathy or creaky voice, it is the relation of the *oh* response to the prior informing in terms of pitch height/range, volume, and timing that is relevant for conveying stances such as interest or surprise (with *upgrading*) vs. empathy or sympathy (with *downgrading*).

Similar prosodic details are relevant for the work that news-receipt *oh* responses to *question-elicited* informings do. However, the sequential environment of question elicitation affords a slightly different set of interpretive parameters for the informing and its response. This is because questions and question-like turns reveal speakers' epistemic stance and their assumptions and expectations concerning what the answer will be (Pomerantz 1988, Stivers and Hayashi 2010, Heritage/ Raymond 2012). The informing provided as an answer to an eliciting question can meet these expectations and confirm these assumptions, or not. When it does not, the way the informing is receipted can have various kinds of affective lamination ranging from revelation or revised understanding (Koivisto forthc.) to mild disappointment (Couper-Kuhlen 2009a).

Consider, for instance, the *oh* produced by Lottie in fragment (2) in response to the information that Emma was not down at the beach on Memorial Day because

[10] For more on trail-off 'conjunctionals' see Local/ Kelly (1986).

Bud had to work that day. This information is elicited by Lottie's my-side telling that Emma was not there on Friday when she (Lottie) came by (lines 5-6).

(5) "Bud had to work Friday" (excerpt from (2))

```
    2 Lot: Well WHERE'VE YOU BEE:N.
->  3 Emm: .hhhh OH I'VE BEEN DOW:N HE:RE,
    4      (0.2)
    5 Lot: I was down there over:: Memorial Day'n you weren't
    6      the:re.
->  7 Emm: Oh I wasn't here Memorial: no buh- Bud had to WORK
    8      Fri:day.
    9      (0.4)
=> 10 Lot: .k Oh::::[:::. ]
```

The news-receipt *oh* in line 10 that Lottie produces in response to Emma's informing that Bud had to work Friday is significantly lengthened and, compared to the pacing of Emma's informing, noticeably delayed:

(5´) Rhythm of Lottie's incoming in line 10 of (2)

```
    8 Emm:   /'no buh-          /
             /'Bud had to       /
             /'work Friday
   10 Lot:   (0.4)  /'Oh:::::          (late)
```

In this context such a rhythmic delay can display cognitive processing: Lottie may be signaling that she needs some time to process the implications of Emma's informing, thereby underlining that it goes against her expectations.

In addition, Lottie's *oh* here is upgraded in terms of its pitch and loudness by comparison with Emma's prior turn-unit (see Figure 3).[11] The work that Lottie's delayed and prosodically upgraded *oh* is accomplishing here must be seen in the context of its occurrence. Lottie's initial question *well where've you been?* in line 2, delivered in a loud voice, is strikingly misplaced, coming as it does immediately after Emma's *(good) morning,* i.e. at a point where a return of Emma's greeting would be expectable next. With this delivery and positioning, Lottie's question implies that Emma has not been where Lottie expected her to be, namely at the beach (lines 5-6). Emma's reply that she was not there over Memorial Day because Bud had to work on Friday thus reveals news to Lottie that runs counter to expectation and causes her to revise her understanding of the events. In the context of Lottie's expectation that Emma would be at the beach on Friday and her possible concern about Emma's well-being upon finding that she was not there, the delayed and upgraded *oh* is thus hearable as indexing revelation: 'NOW I understand'. Moreover, it is a display of revelation associated with some emotive involvement, in context possibly relief that there is a harmless explanation for Emma's absence. The extent of Lottie's emotive reaction can be seen in what happens next. In overlap with Emma's continued telling, Lottie comes back, with an expletive, to the mistaken assumption that led to her riding her bicycle over to Emma's

[11] In this case the prosodic upgrading of Lottie's *oh* is noticeable in particular through its increased volume and raised pitch ceiling (371 Hz, or 22.7 semitones above 100 Hz) compared to that of Emma's informing (337 Hz, or 21.1 semitones).

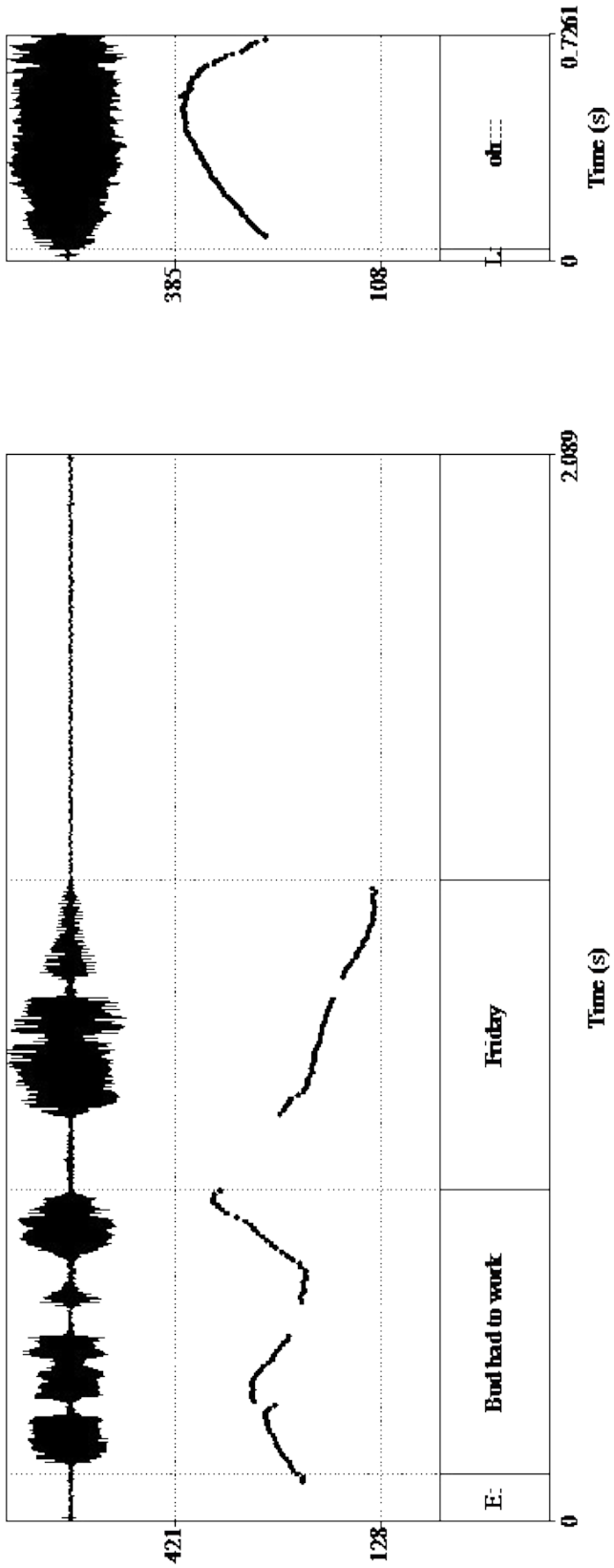


Figure 3: Individually scaled pitch contours for lines 8 and 10 in (5)

place on Friday (lines 12-13), prompting Emma to reiterate that she only arrived Friday night (lines 15-16). With question-elicited informings, prosodically upgraded *oh*-responses displaying revised understanding or revelation are not infrequent when participants are disabused of assumptions they have made in prior talk. But informings which 'correct' prior assumptions and expectations do not invariably trigger prosodic upgrading. This can be seen from examining Lottie's second *oh* in receipt of elicited information (line 22) in fragment (2):

(6) "Play golf Thursday" (excerpt from (2))

```
   12 Lot:                Go:d] I went] do:wn there=
   13      =I think it w-Yeah it was Fri:day.h
   14      (0.2)
   15 Emm: .hh Well I was here Sa:turday: uh: (0.5) uh let's see oh
   16      I came down: uh Friday ni:ght,
   17      (0.3)
   18 Lot: Yah:ah.
-> 19 Emm: Well Bud had to play go:lf uh Thursday.
   20      (.)
   21 Emm: So he [didn't take] Sa-uh f-] Friday o:ff s[o
=> 22 Lot:       [O h : : :] :_: : :.  ]              [Yeh rode down
   23      muh my bi:cycle th[ere en:nu:h h]uh=
   24 Emm:                   [O h : : :_:? ]
   25 Lot: =was nobody wa(h)s the↑::re.
```

The lengthened *oh::::::* that Lottie produces in line 22 in response to Emma's information that Bud had to play golf on Thursday (elicited through Lottie's reiterated my-side telling in lines 12-13) is produced after a slight delay:

(6´) Rhythm of Lottie's incoming in line 22 of (6)

```
   19 Emm:   well /'Bud had to play   /
                  /'golf uh           /
                  /'Thursday. (.)     /
   21 Emm:        /So he [didn't take Sa-uh
   22 Lot:              /[`Oh:::::                (late)
```

In other words, although there is an opportunity for Lottie to come in after the TRP of Emma's turn *well Bud had to play golf uh Thursday* (see line 20), she does not do so. Instead Lottie's news receipt does not come until Emma has continued her turn, spelling out the implications of her informing. Lottie's news-receipt *oh* in line 22 is thus noticeably delayed vis à vis the informing and, like her *oh* in line 10, may be displaying some cognitive processing.

Yet in contrast to the *oh* in line 10, Lottie's *oh* in line 22 is overall lower in pitch, reduced in range and impressionistically softer in volume compared to Emma's informing (see Figure 4).[12] That is, although its delay indexes some cognitive processing and thus a form of revised understanding or revelation 'NOW I understand', this downgraded news-receipt *oh* appears to lack strong emotive involvement. This may be related to the fact that Emma's second piece of information is ancillary to the first: it is an account for why Bud had to work on

[12] Lottie's pitch span here is 8.4 semitones, as compared to Emma's, which is 20.1 semitones.

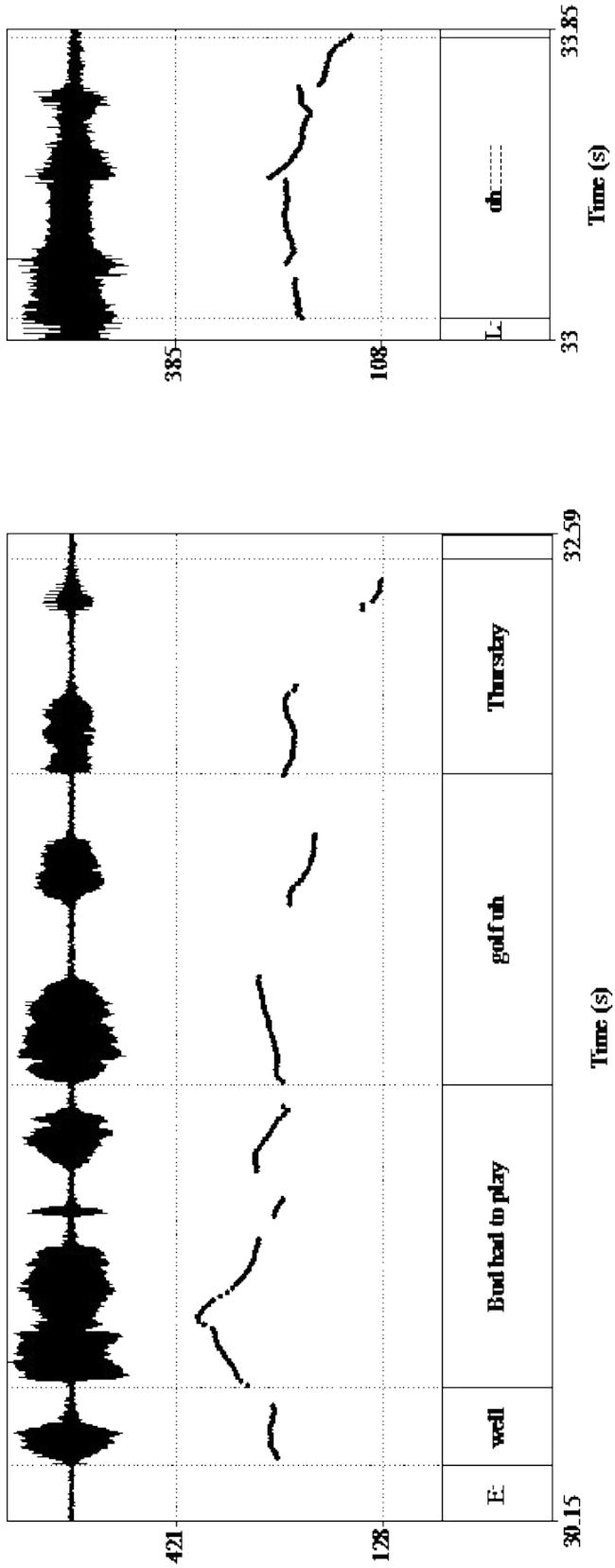


Figure 4: Individually scaled pitch contours for lines 19 and 22 in (6)

Friday, not for why she was not where Lottie expected her to be on Friday. In other words, its relation to Lottie's misconceived assumption is at best indirect. The reduced pitch and volume on Lottie's revelatory *oh* is thus commensurate with the ancillary role the informing plays in her revised understanding.

To summarize: we have reviewed a representative selection of news-receipt *oh* responses to informings, both volunteered and question-elicited. Our argument has been that the dialogic prosodic-phonetic details of delivery, in particular the relation of *oh* to the prior informing in terms of pitch, loudness, and timing, contribute to a cognitive-affective coloring of the response particle that takes on specific context-sensitive interpretations in its particular sequential environment. *Oh* is rather unique among English particle responses in having this much prosodic-phonetic potential, a property that may be attributable to its extremely minimal form. In fact, the more lexical substance a particle response to an informing has, the less varied its prosodic-phonetic shape appears to be by comparison with, e.g., *oh really* or *no kidding* (Thompson et al. forthc.).

3. Particle responses to requests for action

Dialogic prosodic formatting is also relevant for the work that another set of response particles does in a different type of sequential environment. These are the particles *okay, alright* and *sure*, all of which can be used in American English to mark compliance with a prior request for action (Thompson et al. forthc.). The survey carried out by Thompson et al. (forthc.) did not detect any significant differences between these particles when used as complying responses to a request.

The term "request for action" refers to a class of turns, typically non-instructional in nature, in which the speaker either *asks* a recipient to perform some future action or *tells* a recipient to do so. In both cases it is understood to be the requester who will benefit in one way or another from this 'service'. The request-for-action category thus includes what have been called "directives" (Craven/ Potter 2010), "pleads" (Wootton 1981), as well as "prohibitives" (Thompson et al. forthc.).

Requests for action can be implemented with a variety of morpho-syntactic forms, including (in English) imperatives (*Do X!),* polar interrogatives (*Would you do X?),* and declaratives (*I need X*). As Curl and Drew (2008) have shown, the morpho-syntactic format of a request can display varying degrees of *entitlement*, the understanding that one has the right to impose on another, and/or *contingency*, the understanding that there may be factors involved making the request difficult for the requestee to grant.

The particle responses that are used most frequently in American English conversation to comply with a request for action are *alright, okay* and *sure*. These particles tend to occur as freestanding responses when requestees are complying with requests that are done with a display of strong entitlement and low contingency (Thompson et al forthc.). That is, these particles are used to comply with requests when requesters are treating them as routine and easy to comply with, when they are not making a ‘big deal’ out of them. Example (7) shows such a case, where Mom makes a routine request of her son, who then responds with the complying particle *alright*:

(7) "Blessing" (Virginia, 5.15-12.7)
(Mom has just sat down to dinner with her family. Wesley is her grown son.]

```
-> 1 Mom: (c'n) we have the blessi-ih-buh-
-> 2      Wesley would you ask the blessi[ng please,
=> 3 Wes:                                [alright.
   4      (0.2)
   5 Wes: heavenly father give us thankful hearts (for) these and all
   6      thy blessings °amen.
   7      (.)
   8 Vir: >°amen.<
   9      (2.0)
```

As *mater familias* Mom displays high entitlement in requesting her son to ask the blessing; she also does not show any doubt about his ability to grant this request. Wesley's minimal token *alright* is thus fully sufficient as a signal of compliance for this kind of routine request.

What stances does the prosody of a compliant response convey? Interestingly, it is first and foremost the timing of a compliant particle response to a request that is relevant for the interactive work it does. In (7) Wesley's *alright*, for instance, occurs off-beat with respect to the rhythm of Mom's request. The result is an early incoming that overlaps the end of Mom's turn with its appended *please*:

(7´) Rhythm of Wesley's incoming in line 3 of (7)

```
2 Mom: but   /'Wesley would   /
             /'you ask the    /
             /'blessing       /
3 Wes: /'alright                                    (early)
  Mom:       /please
```

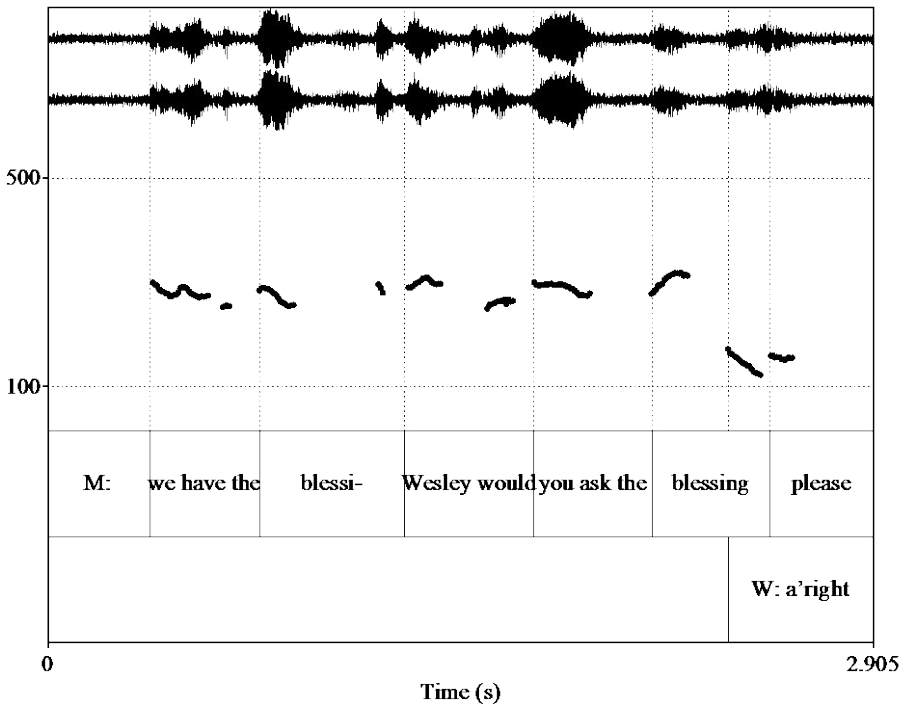


Figure 5: Early onset of line 3 in Extract (7)

The overlap in Wesley's response is not a turn-taking violation. On the contrary, it comes at a point in time when Wesley can recognize what the gist of Mom's turn will be: not only is saying the blessing a routine activity before beginning to eat but Mom has nominated him by name. Wesley's response therefore is delivered in what Jefferson has called 'recognitional overlap' (Jefferson 1984).

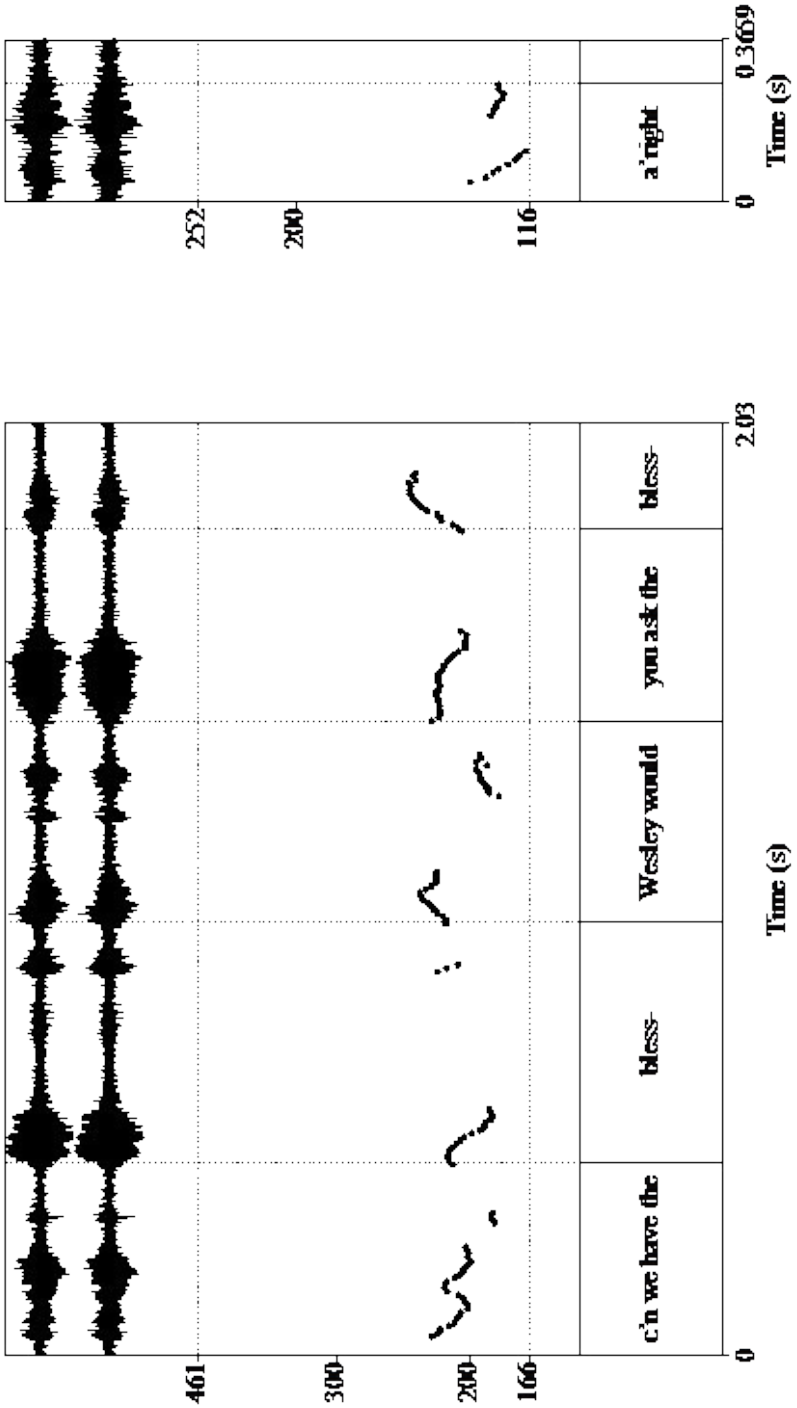


Figure 6: Individually scaled pitch contours for lines 1-3 in (7)

In terms of pitch and loudness, Wesley's *alright* is produced with low pitch and volume, essentially matching the relatively low pitch and volume of Mom's request: see Figure 6. Wesley's *alright*, with its matching pitch and volume and especially with its overlapped recognitional onset, is interpretable as conveying full willingness to comply with Mom's request. Evidence for Wesley's willingness will be seen in the promptness with which he now begins to fulfill Mom's request and ask the blessing.

Particle responses that mark full willingness to comply with a prior request are typically produced at recognizable TRPs, or even, as in (7), in recognitional overlap: in other words, they set in on the next beat or even before the next beat established by the pacing of the request (see also Thompson et al forthc.). With this timing, such particles do not require any pitch or loudness upgrading to be interpretable as fully compliant. However, if they *do* have some form of prosodic upgrading, then there is likely to be an affective lamination, which becomes interpretable in context. This is what we see happening in the following excerpt:

(8) "Condense it down" (Cutie Pie, CSAE)

(Jill is out of town visiting her friend Jen and is talking to her boyfriend Jeff long-distance. Jeff now asks Jill to tell him what she and Jen have been saying about him.)

```
    1 Jef: .ts .h well,
    2      (0.3)
    3      give me the highlights.
    4      (0.3)
    5 Jil: the highlights?
    5 Jef: yeah,
    7      (0.3)
    8 Jil: of what we: talk about you?
    9      (0.2)
   10 Jef: yeah.
   11      (0.3)
   12 Jil: .ts u::hm_
   13      (0.3)
   14      #let me think#,
   15      #let me think of some: good ones.#
-> 16 Jef: condense it down to (.) a minute.
=> 17 Jil: £↑OKAY↑£
   18      WE: said something s:o funny today.
   19      .hh and I thought when (0.2) Jen said it#,
   20      #it captured you so well?=
```

In this request sequence, Jill's *okay* is timed to set in *early* with respect to Jeff's request:

(8´) Rhythm of Jill's incoming in line 17 of (8)

```
   16 Jef: condense it
                  /'down to (.) a/
                  /'minute       /
   17 Jil: /'okay                                  (early)
```

With this timing Jill's response displays full willingness to comply. However, her *okay* is also higher and louder than Jeff's request, as we can see from the combination Praat picture in Figure 7.

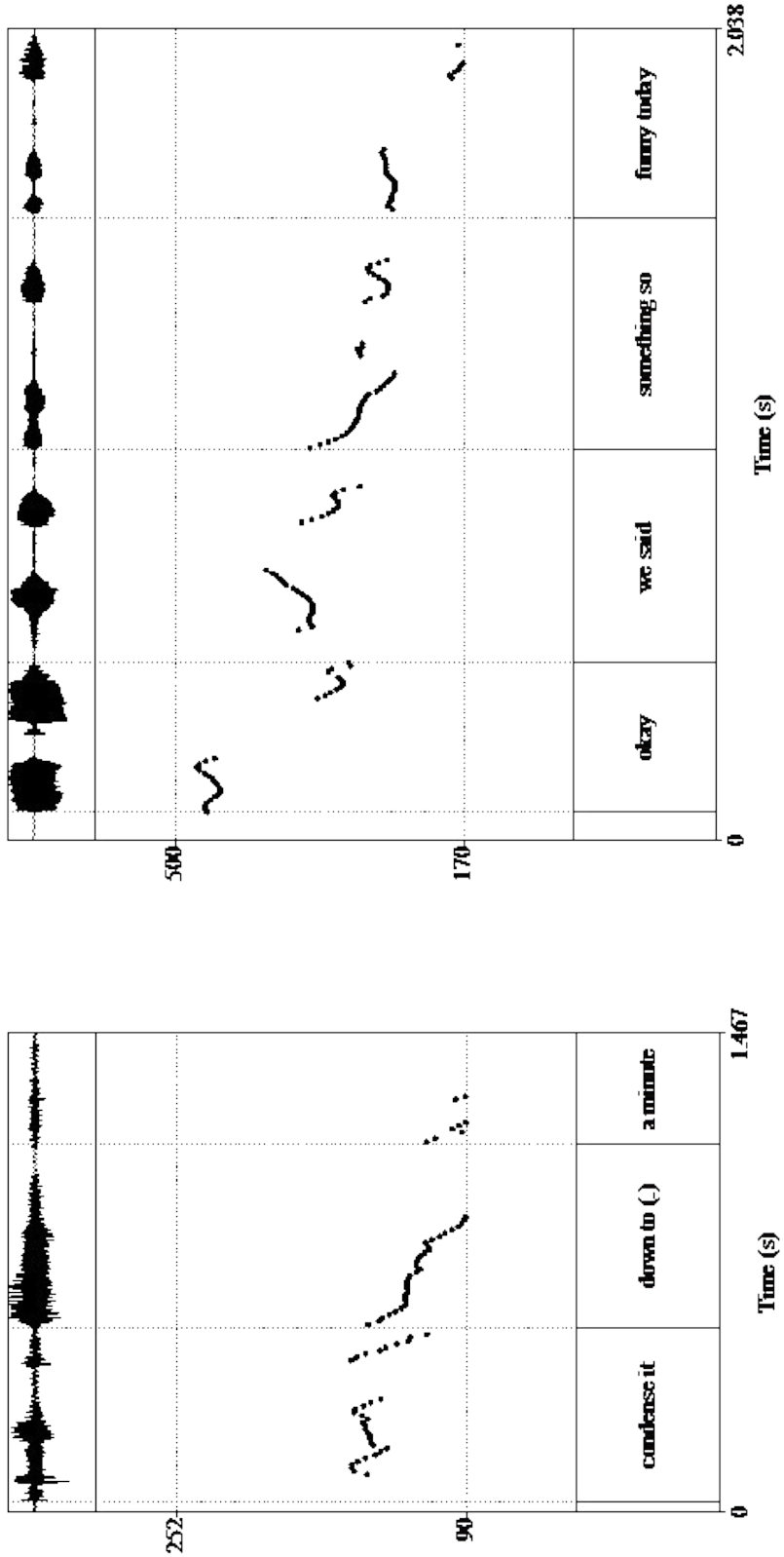


Figure 7: Individually scaled pitch contours for lines 16 and 17 in (8)

As shown in Figure 7, the pitch of Jill's *okay we said something so funny today* starts much higher and covers a much wider pitch span than does Jeff's request *condense it down to (.) a minute*.[13] The volume of Jill's onset is also significantly greater than at the end of Jeff's turn. With this prosody and in this context, Jill's response is thus interpretable as displaying not only full willingness, but also a certain amount of pleasure (it is overlaid with smile voice) and enthusiasm in complying with what Jeff has asked. Evidence for this will be seen in the fact that she rushes immediately into a funny story that she and Jen have told about Jeff.

In contrast, when complying particles are produced with *delayed* timing in relation to that of a prior request turn, these responses are open to the interpretation that recipients, although they are acquiescing, are hesitant or reluctant to do so. Here is a case where the recipient of a request indicates some reluctance in complying:

(9) "Tortilla chips" (Kamunsky 2, p. 3)

(Alan has called his friend Shawn to invite him to a surprise birthday party for a mutual friend. Shawn has responded evasively, whereupon it emerges that he is 'grounded'.)

```
    1 ALA: [Oh en if you can make it
    2      bring oh I don(’t know) a bag of
    3      potato chips or tortilla chips or something.
    4      ˙hh cuz Karen's bringing the chips'n the potato chips (_)
    5      ˙hh eh the uh dip en the potato chips
->  6      so if[you wanna bring]=
    7 SHA:      [°(W  o   :  w )]=
->  8 ALA: =some tortilla chips would be fi:ne,
=>  9 SHA: ˙hh Oka[y,
   10 ALA:        [Okay
```

Shawn's *okay* in line 09, delivered with a stylized (sing-song) pitch configuration, is slightly delayed in relation to the timing of Alan's request to bring some tortilla chips (lines 06 + 08, Fig. 8):

(9´) Rhythm of Shawn's incoming in line 09 of (9)

```
   6 ALA: so if you wanna bring some tor-
   8      /’tilla chips would be   /
          /’fine                   /
   9 SHA:     /’okay                                  (late)
```

[13] Jill's pitch range here has a span of 17.8 semitones; Jeff's, a span of 7.8 semitones.

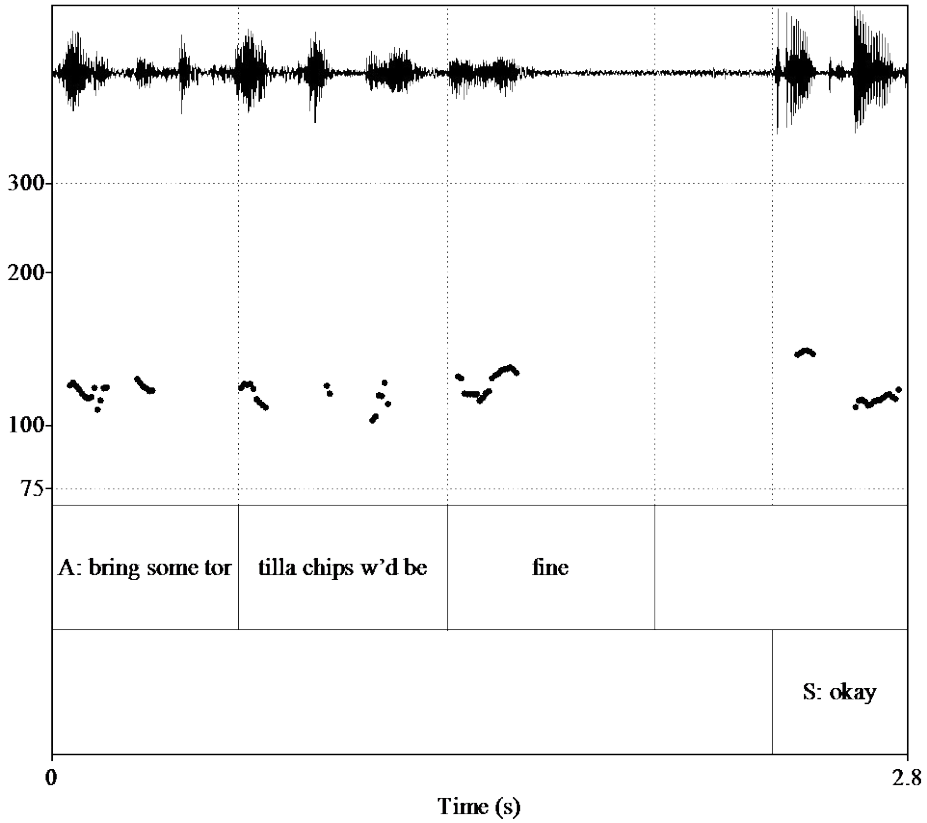


Figure 8: Late incoming of line 09 in (9)

Like Wesley's acquiescing *alright* in Ex. (7), Shawn's pitch and loudness levels on *okay* in Ex. (9) lack any noticeable upgrading or downgrading in relation to the prior request: see Figure 9. The interactional import of this prosody on Shawn's verbally compliant response must be seen in the context of what has transpired so far in this telephone conversation: Alan has invited Shawn to a surprise birthday party for a mutual friend. Although Shawn remains non-committal about being able to come, Alan nevertheless proceeds to announce when the party will start and, as seen in (9), to ask Shawn to bring tortilla chips along if he can come. Shawn's ambivalence is palpable: On the one hand, he displays interest in Alan's invitation by asking what their friend wants for his birthday (not shown here) but on the other hand, he avoids any explicit acceptance of the invitation. The hesitancy with which Shawn accedes to Alan's request to bring tortilla chips if he can come displays the same ambivalence: at once willingness to commit, via the lexical choice of *okay*, but reluctance in terms of accountability, via its delivery with delayed timing.

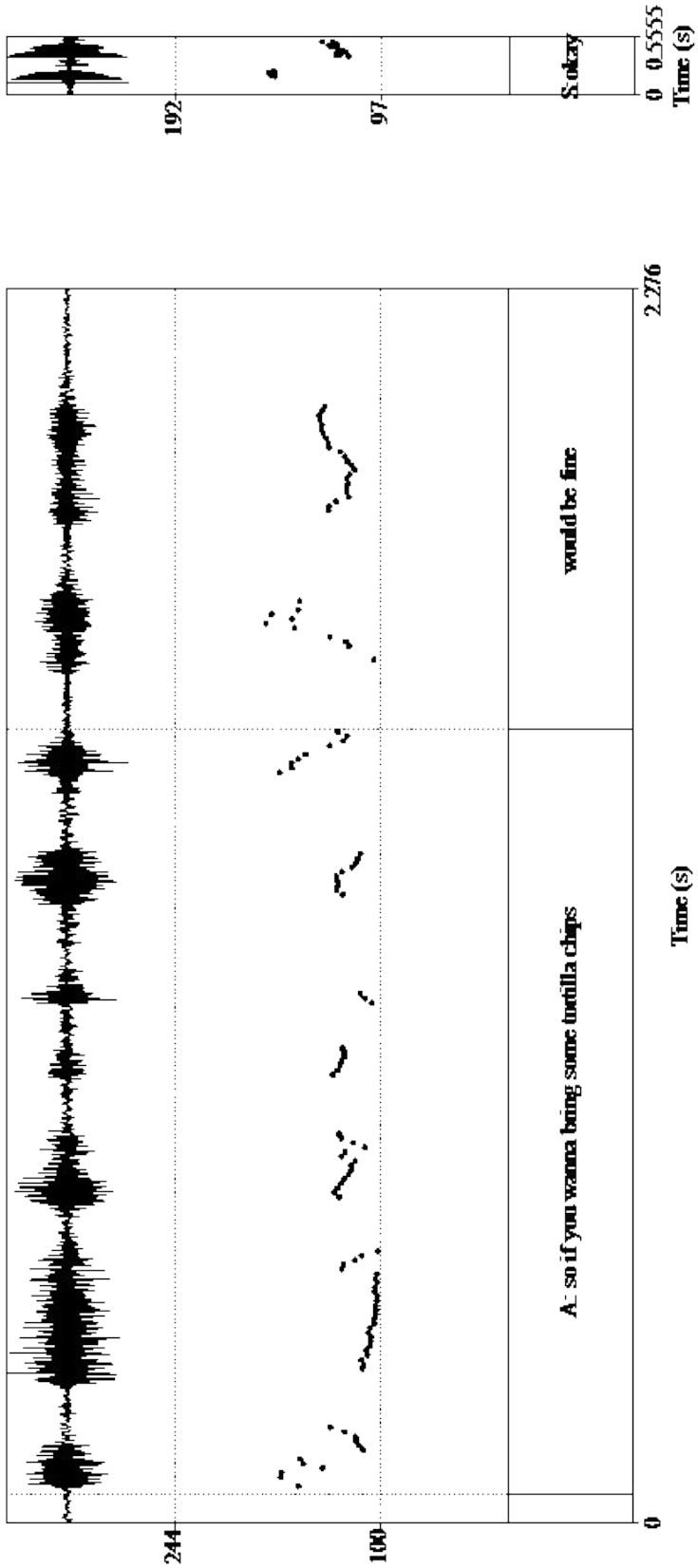


Figure 9: Individually scaled pitch contours for lines 6/8 and 9 in (9)

In the proper circumstances, a delay in timing on a lexically acquiescing particle such as *okay, alright* or *sure* can amount to a subtle form of resistance to the request. This is what happens in the following case, where the requestee's resistance is unmasked by the requester:

(10) "Is that okay" (Happy Hour)[14]
(Molly is trying to recruit her friends Susie and Oliver, who are a couple, to do some videotaping for her.)

```
    1   MOL:   can i record YOU guys?
    2          (0.5)
    3   SUS:   [dee and DEE:::][15]
    4   MOL:   [do you ha:ve- ]
    5          NO::;
    6          YOU guys.
    7   MOL:   do you have DIN:ner together:?
    8          (0.2)
    9   MOL:   do you just SIT?: and watch tee vee and TALK?
   10          (0.8)
   11   SUS:   mHM,
-> 12   MOL:   can i just- give you the CAMera and you can
               set it UP for a week?
   13          (0.2)
-> 14   SUS:   d'you CARE?

((16 lines omitted))

   31   SUS:   .hhhh
   32          [uhm:;
   33   OLI:   [mm
-> 34   SUS:   is that oKAY:?
   35          (2.0)
        ((Oliver swallowing a French fry and running left hand through
          hair))
=> 36   OLI:   S::URE::.
   37   MOL:   thAt was a: (.) hOney we'll tAlk about it LATer?
   38          and i'm gonna say NO!
   39          (1.3)
   40   SUS:   is it oKAY?
   41          (0.3)
   42   OLI:   well it's not: NOT oKAY::,
   43          BUT - you KNOW:,
   44          (0.7)
   45   SUS:   yEs or NO:::;
   46          (0.2)
   47   OLI:   S[URE:.     ]
   48   SUS:    [rIght NOW;]
   49          oKAY;
   50          (0.2)
   51   OLI:   (it'll) TAKE forever,
   52          (0.9)
   53   SUS:   i KNOW; but it's ALL right;
   54          (0.5)
```

Molly's request that her friends tape themselves for a week at home (line 12) is redirected by Susie to Oliver, whose 'caring' she appears to anticipate. However, Oliver's response is not immediately forthcoming, which leads to some jocular

[14] This excerpt uses the GAT 2 transcription system (Selting et al. 2009).
[15] A reference to 'Dungeons and Dragons', a popular computer game at the time.

teasing by his friends (not shown here). After this, Susie pursues a response from Oliver, on behalf of her friend Molly, with *is that okay?* (line 34). Oliver now responds with the complying particle *sure* (line 26), thereby formally acceding to Molly's request.

However, the embodied and prosodic delivery of Oliver's token suggests that he is anything but "sure" about fully committing to Molly's request. He first finishes chewing the French fry in his mouth and swallows it, while simultaneously running his hand through his hair. His production of *sure* is thus delayed by two seconds, and when it comes, it is hyperarticulated. The initial fricative of *sure* is lengthened, and the /r/ is strongly retroflexed. Prosodically, the token *sure* has a strong rise-fall contour; moreover, as the combined Praat picture makes clear, it is also upgraded in terms of pitch and loudness by comparison with Susie's prior turn (see Figure 10).

Oliver's pitch span is wider here than Susie's and he reaches a higher pitch peak in his range than she does in hers.[16] In other circumstances, this kind of prosodic upgrading might index heightened commitment; however, with Oliver's extremely delayed timing and his phonetic hyperarticulation, the upgrading here conveys just the opposite: an ironic reading of *okay* conveying reluctance to comply, if not resistance to Molly's request. This reading of Oliver's turn is supported by his post-completion stance marker in the form of a smile, which because of its delayed production becomes a commentary on the just-completed delivery. Thus, conflicting prosodic and phonetic cues on the acquiescing response particle lead to the interpretation that Oliver is being ironic. And indeed, this is the way Molly interprets his response (lines 37-38).

To summarize: this section has presented a representative sampling of particle responses that signal compliance with a prior request. The argument has been that here too, dialogic prosodic features are relevant for the interactional import of the response: in the case of requesting, whether the recipient is displaying full willingness to comply, or instead some reluctance, if not resistance. In contrast to particle responses to informings, where it is pitch and loudness upgrading or downgrading that is relevant for stance-related shades of meaning, here it is primarily the feature of timing: early or on-time responses signaling full willingness, delayed responses signaling hesitation and/or reluctance. In the case of particle responses to requests, pitch and loudness upgrading has only an added effect, that of marking enthusiasm over and above full willingness, or of marking ironic 'enthusiasm' in addition to reluctance and/or resistance.[17] These findings have interesting implications for dialogic prosody: they suggest that although dialogic features are always somehow relevant for stance-marking in (preferred) responses, the precise ways in which they are relevant differs from sequence to sequence. We return to this point below.

[16] Oliver's pitch range is 15.6 semitones, while Susie's is 6.9 semitones.

[17] There were no cases of well-timed but prosodically downgraded complying responses with particles in the database for this study.

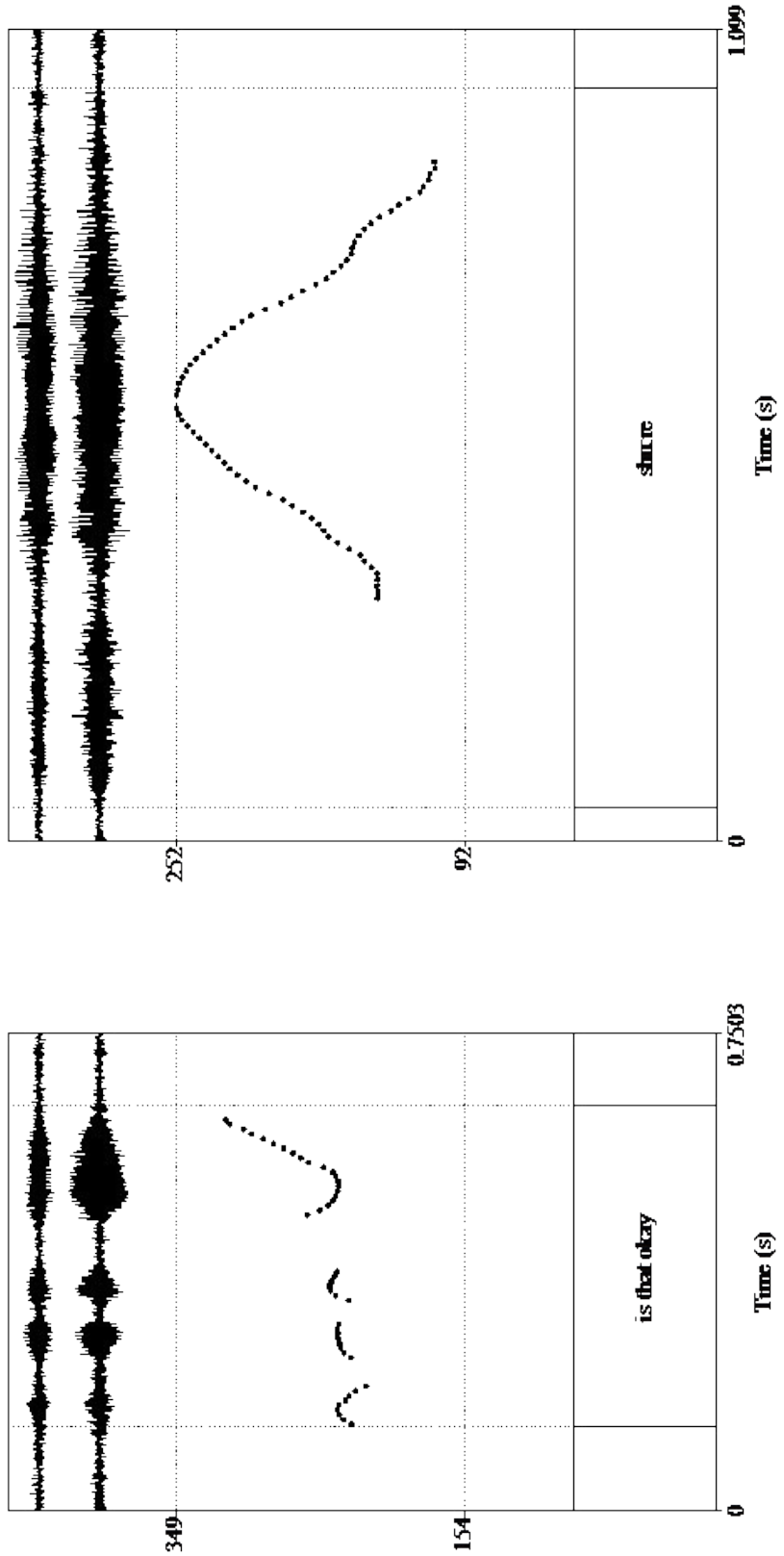


Figure 10: Individually scaled pitch contours for lines 34 and 36 in (10)

4. Methodological challenges and analytic pay-offs

The above discussion has made a (renewed) plea for an approach to prosody that treats it dialogically: in responsive actions, this means examining not just what the prosody is like with respect to the individual's voice, but how the prosody of the response relates to the prosody of the prior turn. Does it have the same parameters or different ones? Are these parameters more or less strong, more or less extensive than those of the prior turn? Or, with respect to timing, does the timing of the response 'match' the pacing of the prior turn, or not?

This is not to deny that there are methodological challenges in analyzing prosody across speakers. Since individuals have different vocal apparatuses and different natural rhythms, analysts inevitably find themselves asking questions like: What counts as 'the same', what counts as 'different'? For instance:

With respect to *pitch*: What may seem like a high pitch for a male voice may actually be lower, on an absolute scale, than the lowest pitch of a female voice. Prosodists have always known that pitch must be analyzed relatively, i.e. relative to the individual's habitual pitch range. An individual-based calibration, however, makes it difficult to compare across speakers unless the relative pitch peaks and valleys (=pitch span) of one individual's voice are put in relation to those of another. This is what the combined Praat pictures shown above have attempted to do, namely to show the relative pitch height and span on *oh* as produced by one voice in relation to the relative pitch height and span of a turn in the interlocutor's voice.

With respect to *loudness*: The situation is even more complex here, since individual speakers not only have their own habitual volume levels but can also be at different distances from the recording device. Ideally, however, one would want to put what is loud for one speaker in relation to what is loud for another. Many of the same issues that are relevant for pitch apply to loudness as well, although our methods for representing relative loudness across speakers are less well developed.

With respect to *syllable duration* and *speech rate*: Here the issue is not how fast or slow one segment of a speaker's turn is compared to another segment of that same turn, but rather how fast or slow one speaker's turn is compared to that of a prior speaker. What remains to be established is how to factor out differing habitual rates of speech.

With respect to the *timing* of turns: The status of a well-timed, early or late incoming must be determined in relation to the pacing of the prior turn. For English, a rhythmic metric has proved appropriate for capturing these distinctions (see, e.g., Auer et al 1999). But whether, and if so how, such a rhythmic metric might apply to other languages remains an open question.

With respect to *voice quality*: Factoring out different habitual voice qualities, the challenge here is to determine if and how the overlaid voice quality of one turn relates to the voice-quality overlay, if there is one, of a prior turn.

Despite these challenges, there are clear analytic payoffs in approaching prosody dialogically. Some of these will have become evident in the discussion above. For one, with dialogically calibrated features we have a handle on how participants 'inter-act' prosodically when a particular type of response format (here, a

particle format) is held constant across different sequence types. The results are revealing:

- (i) As intimated above, dialogic prosodic parameters are always relevant for particle responses, but differentially so in different sequence types. This will be seen from Table 1, which presents an overview of the particle responses and their prosodic design as discussed in the examples above.

Particle response	**Upgraded**			**Downgraded**			**Timing**	**Stance**
	Pitch peaks	*Pitch range*	*Loud-ness*	*Pitch peaks*	*Pitch range*	*Loud-ness*		
To informings								
(3) oh	√	√					On time	+ ☺[18]
(4) oh				√	√	√	Early	+ ☹[19]
(5) oh	√		√				Delayed	+ ☺
(6) oh				√	√	√	Delayed	+ 😐[20]
To requests								
(7) alright							Early	+ ☺[21]
(8) okay	√	√	√				Early	+ ☺
(9) okay							Delayed	+ ☹[22]
(10) sure	√	√	√				Delayed	+ ☹

Table 1: Prosodic design of particle responses in Informing and Request sequences[23]

From this table it becomes clear that for Informing sequences it is primarily pitch and loudness upgrading or downgrading that is relevant for the stance a particle response is conveying, whereas for Requests it is timing that is relevant for the stance of a particle response:

In Informing sequences, pitch and loudness *upgrading vs. downgrading* are systematically involved in ☺ vs. ☹/ 😐 stances, these symbols being emblematic for interest/ surprise/ emotively involved revelation vs. empathy/ sympathy/ emotively uninvolved revelation. But some ☺ stances are conveyed by on-time, others by delayed responses; some ☹/ 😐 stances are conveyed by early, others by delayed responses. That is, the timing of a particle response is of secondary importance in Informing sequences. It accomplishes in the main a display of cognitive processing, which is interactionally relevant for the distinction between responses to elicited vs. volunteered informings.

In Request sequences, by contrast, it is the *timing* of a particle response that is systematically relevant for the stance being conveyed: ☺ stances (= full willingness ± enthusiasm) are signaled by early or on-time incomings, ☹ stances (= hesitation/ reluctance/ resistance) are signaled by delayed incomings. The presence of pitch and loudness upgrading on a particle response conveys additional lamina-

[18] ☺ = interest/surprise/emotively involved revelation
[19] ☹ = empathy/sympathy
[20] 😐 = emotively uninvolved
[21] ☺ = full willingness ± enthusiasm
[22] ☹ = hesitation/reluctance/resistance
[23] Preferred ("+" action) responses only. Dialogic prosodic dimensions only.

tions optionally: for instance, it can bring in an added component of enthusiasm, which enriches the basic stance of full willingness or ironizes a stance of reluctance.

- (ii) The cognitive-affective interpretations that dialogic prosody makes relevant are also sequence-specific. Although all of the particles considered here are implementing 'preferred', or "+" action, responses, their prosodic design hints at more subtle stances being taken by their speakers over and above the simple preferred next action. For instance, in volunteered informing sequences, what is at stake are things like 'Am I interested in this or not?', 'Does this news have positive repercussions for me/you, or negative ones?' But in elicited informing sequences, what is at stake is 'Is this informing what I expected or not?', 'Does it cause me to revise my prior understanding?', 'Is it ancillary to my assumptions or not?', 'Does the revision of understanding have positive implications for you/me, or not?'

By contrast, in request sequences, what is at stake is not interest, sympathy or revelation. Instead it is degrees of willingness to comply with what is being requested. Willingness can be full or 'over the top' – shading off into enthusiasm – or it can be less than full, being variously interpretable as hesitation, reluctance or resistance. For different sequence types there is then a specific 'menu' of cognitive-affective choices, which are intimately tied up with the type of action involved. This constrains the ascription of cognitive-affective meaning to prosodic formattings and renders them more readily interpretable for interactants.

In addition to these specific insights, here with respect to the workings of prosodic design and response particle, viewing prosody dialogically has the more general advantage that, like e.g. Szczepek Reed's (2006, 2012) and Ogden's (2006) work, it puts the study of prosody on a par with the study of social action. In action analysis it is not a question of what individual action a particular turn is implementing, but rather of how that action relates to a prior action, or of what kind of subsequent action it makes relevant. We would be foolhardy to try to determine, for instance, what action some particular turn is implementing in a context-free fashion. And so it is with prosody: both its form and meaning potential must be seen in relation to what has preceded and what is expected to follow. If we can manage to take this perspective, we will have completed the dialogic 'turn' in the study of prosody as well.

5. References

Auer, Peter/ Couper-Kuhlen, Elizabeth/ Müller, Frank-Ernst (1999): Language in Time: The Rhythm and Tempo of Spoken Interaction. Oxford: OUP.

Couper-Kuhlen, Elizabeth (1992): Contextualizing discourse: The prosody of interactive repair. In: Auer, Peter/ di Luzio, Aldo (eds.): The Contextualization of Language. Amsterdam: Benjamins, 337-364.

Couper-Kuhlen, Elizabeth (1993): English Speech Rhythm: Form and function in everyday verbal interaction. Amsterdam: Benjamins.

Couper-Kuhlen, Elizabeth (1996): The prosody of repetition: On quoting and mimicry. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation. Interactional studies. Cambridge: CUP, 366-405.

Couper-Kuhlen, Elizabeth (2009a): A sequential approach to affect: The case of 'disappointment'. In: Haakana, Markku/ Laakso, Minna/ Lindström, Jan (eds.): Talk in interaction – comparative dimensions. Helsinki: Finnish Literature Society (SKS), 94-123.

Couper-Kuhlen, Elizabeth (2009b): Relatedness and timing in talk-in-interaction. In: Barth-Weingarten, Dagmar/ Dehé, Nicole/ Wichmann, Anne (eds.): Where Prosody Meets Pragmatics. Bingley: Emerald, 257-276.

Craven, Alexandra/ Potter, Jonathan (2010): Directives: Entitlement and contingency in action. In: Discourse Studies 12(4), 419-442.

Curl, Traci S. (2005): Practices in other-initiated repair resolution: The phonetic differentiation of 'repetitions'. In: Discourse Processes 39(1), 1-44.

Curl, Traci S./ Drew, Paul (2008): Contingency and action: A comparison of two forms of requesting. In: Research on Language and Social Interaction 41(2): 129-153.

Freese, Jeremy/ Maynard, Douglas W. (1998): Prosodic features of bad news and good news in conversation. In: Language in Society 27, 195-219.

Golato, Andrea (2012): German *oh*: Marking an emotional change of state. In: Research on Language and Social Interaction 45(3), 245-268.

Goodwin, Marjorie H./ Cekaite, Asta/ Goodwin, Charles (2012): Emotion as stance. In: Peräkylä, Anssi/ Sorjonen, Marja-Leena (eds.): Emotion in Interaction. Oxford: OUP, 16-41.

Gorisch, Jana/ Wells, Bill et al. (2012): Pitch contour matching and interactional alignment: An acoustic investigation. In: Language and Speech 55, 57-76.

Heritage, John (1984): A change-of-state token and aspects of its sequential placement. In: Atkinson, J. Maxwell/ Heritage, John (eds.): Structures of Social Action. Studies in Conversation Analysis. Cambridge: CUP, 299-345.

Heritage, John/ Raymond, Geoffrey (2005): The terms of agreement: Indexing epistemic authority and subordination in assessment sequences. In: Social Psychology Quarterly 68, 15-38.

Heritage, John/ Raymond, Geoffrey (2012): Navigating epistemic landscapes: Acquiescence, agency and resistance in responses to polar questions. In: de Ruiter, Jan P. (ed.): Questions: Formal, functional and interactional perspectives. Cambridge: CUP, 179-192.

Jefferson, Gail (1978): Sequential aspects of storytelling in conversation. In: Schenkein, Jim (ed.): Studies in the Organization of Conversational Interaction. New York: Academic Press, 219-248.

Jefferson, Gail (1984): Notes on some orderliness of overlap onset. In: D'Urso, Valentina/ Leonardi, Paolo (eds.): Discourse Analysis and Natural Rhetorics. Padova: Cleup Editore, 11-38.

Jefferson, Gail (1988): On the sequential organization of troubles-talk in ordinary conversation. In: Social Problems 35(4), 418-441.

Jefferson, Gail (2004): Glossary of transcript symbols with an introduction. In: Lerner, Gene H. (ed.): Conversation Analysis. Studies from the first generation. Amsterdam: Benjamins, 13-31.

Koivisto, Aino (In press): Displaying now-understanding: The Finnish change-of-state token 'aa'. In: Discourse Processes.

Linell, Per (2009): Rethinking Language, Mind, and World Dialogically: Interactional and contextual theories of human sense-making. Charlotte NC, Information Age Publishing.

Local, John (1996): Conversational phonetics: Some aspects of news receipts in everyday talk. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation. Cambridge: CUP, 177-230.

Local, John/ Kelly, John (1986): Projection and 'silences': Notes on phonetic and conversational structure. In: Human Studies 9, 185-204.

Maynard, Douglas W. (1997): The news delivery sequence: Bad news and good news in conversational interaction. In: Research on Language and Social Interaction 30, 93-130.

Maynard, Douglas W. (2003): Bad News, Good News: Conversational order in everyday talk and clinical settings. Chicago: University of Chicago Press.

Maynard, Douglas W./ Freese, Jeremy (2012): Good news, bad news, and affect: practical and temporal 'emotion work' in everyday life. In: Peräkylä, Anssi/ Sorjonen, Marja-Leena (eds.): Emotion in Interaction. Oxford: OUP.

Müller, Frank Ernst (1996): Affiliating and disaffiliating with continuers: prosodic aspects of recipiency. In: Couper-Kuhlen, Elizabeth/ Selting, Margret (eds.): Prosody in Conversation. Cambridge: CUP, 131-170.

Ogden, Richard (2006): Phonetics and social action in agreements and disagreements. In: Journal of Pragmatics 38, 1752-1775.

Pomerantz, Anita (1980): Telling my side: 'Limited access' as a 'fishing device'. In: Sociological Inquiry 50, 186-198.

Pomerantz, Anita (1988): Offering a candidate answer: An information seeking strategy. In: Communication Monographs 55(1), 360-373.

Reber, Elisabeth (2012): Affectivity in interaction. Sound objects in English. Amsterdam: Benjamins.

Schegloff, Emanuel A. (1986): The routine as achievement. In: Human Studies 9, 111-151.

Schegloff, Emanuel (1998): Reflections on studying prosody in talk-in-interaction. In: Language and Speech 41(3-4), 235 - 263.

Schegloff, Emanuel A. (2007): Sequence Organization in Interaction: A Primer in Conversation Analysis, Vol. 1. Cambridge: CUP.

Selting, Margret/ Auer, Peter/ Barth-Weingarten, Dagmar/ Bergmann, Jörg/ Bergmann, Pia/ Birkner, Karin/ Couper-Kuhlen, Elizabeth/ Deppermann, Arnulf/ Gilles, Peter/ Günthner, Susanne/ Hartung, Martin/ Kern, Friederike/ Mertzlufft, Christine/ Meyer, Christian/ Morek, Miriam/ Oberzaucher, Frank/ Peters, Jörg/ Quasthoff, Uta/ Schütte, Wilfried/ Stukenbrock, Anja/ Uhmann, Susanne (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion, 10, 353-402. (www.gespraechsforschung-ozs.de)

Stivers, Tanya/ Hayashi, Makoto (2010): Transformative answers: One way to resist a question's constraints. In: Language in Society 39, 1-25.

Szczepek Reed, Beatrice (2006): Prosodic Orientation in English Conversation. Basingstoke, Hampshire: Palgrave Macmillan.

Szczepek Reed, Beatrice (2012): Beyond the particular: Prosody and the coordination of actions. In: Language and Speech 55(1), 13-34.

Thompson, Sandra A./ Fox, Barbara/ Couper-Kuhlen, Elizabeth (in press). Grammar and everyday talk: Building responsive actions. Cambridge: CUP.

Wells, Bill (2010). Tonal repetition and tonal contrast in English carer-child interaction. In: Barth-Weingarten, Dagmar/ Reber, Elisabeth/ Selting, Margret (eds.): Prosody in interaction. Amsterdam: Benjamins, 243-262.

Wilkinson, Sue/ Kitzinger, Celia (2006): Surprise as an interactional achievement: Reaction tokens in conversation. In: Social Psychology Quarterly 69(2), 150-182.

Wootton, Anthony J. (1981): The management of grantings and rejections by parents in request sequences. In: Semiotica 37(1/2), 59-89.

Elizabeth Couper-Kuhlen
University of Helsinki
Center of Excellence for Research on Intersubjectivity in Interaction
P.O. Box 3 (Fabianinkatu 33)
00014 University of Helsinki
Finland
elizabeth.couper-kuhlen [at] helsinki.fi

Anhang/ Appendix:

Zusammenstellung der wichtigsten GAT 2-Transkriptionskonventionen

(vgl. Selting, Margret/ Peter Auer/ Dagmar Barth-Weingarten/ Jörg Bergmann/ Pia Bergmann/ Karin Birkner/ Elizabeth Couper-Kuhlen/ Arnulf Deppermann/ Peter Gilles/ Susanne Günthner/ Martin Hartung/ Friederike Kern/ Christine Mertzlufft/ Christian Meyer/ Miriam Morek/ Frank Oberzaucher/ Jörg Peters/ Uta Quasthoff/ Wilfried Schütte/ Anja Stukenbrock/ Susanne Uhmann (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion 10, 353-402 (www.gespraechsforschung-ozs.de).

Minimaltranskript

Sequenzielle Struktur/ Verlaufsstruktur

`[ ]` `[ ]`	Überlappungen und Simultansprechen

Ein- und Ausatmen

`°h / h°`	Ein- bzw. Ausatmen von ca. 0.2-0.5 Sek. Dauer
`°hh / hh°`	Ein- bzw. Ausatmen von ca. 0.5-0.8 Sek. Dauer
`°hhh / hhh°`	Ein- bzw. Ausatmen von ca. 0.8-1.0 Sek. Dauer

Pausen

`(.)`	Mikropause, geschätzt, bis ca. 0.2 Sek. Dauer
`(-)`	kurze geschätzte Pause von ca. 0.2-0.5 Sek. Dauer
`(--)`	mittlere geschätzte Pause v. ca. 0.5-0.8 Sek. Dauer
`(---)`	längere geschätzte Pause von ca. 0.8-1.0 Sek. Dauer
`(0.5)`	gemessene Pausen von ca. 0.5 bzw. 2.0 Sek. Dauer
`(2.0)`	(Angabe mit einer Stelle hinter dem Punkt)

Sonstige segmentale Konventionen

`und_äh`	Verschleifungen innerhalb von Einheiten
`äh, öh, etc.`	Verzögerungssignale, sog. "gefüllte Pausen"

Lachen und Weinen

`haha hehe hihi`	silbisches Lachen
`((lacht))((weint))`	Beschreibung des Lachens
`<<lachend> >`	Lachpartikeln in der Rede, mit Reichweite
`<<:-)> soo>`	"smile voice"

Rezeptionssignale

`hm, ja, nein, nee`	einsilbige Signale
`hm_hm, ja_a,` `nei_ein, nee_e`	zweisilbige Signale
`ʔhmʔhm,`	mit Glottalverschlüssen, meistens verneinend

Sonstige Konventionen

((hustet))	para- und außersprachliche Handlungen u. Ereignisse
<<hustend>>	sprachbegleitende para- und außersprachliche Handlungen und Ereignisse mit Reichweite
( )	unverständliche Passage ohne weitere Angaben
(xxx), (xxx xxx)	ein bzw. zwei unverständliche Silben
(solche)	vermuteter Wortlaut
(also/ alo) (solche/ welche)	mögliche Alternativen
((unverständlich, ca. 3 Sek))	unverständliche Passage mit Angabe der Dauer
((...))	Auslassung im Transkript
->	Verweis auf im Text behandelte Transkriptzeile

Basistranskript

Sequenzielle Struktur/Verlaufsstruktur

=	schneller, unmittelbarer Anschluss neuer Sprecherbeiträge oder Segmente (*latching*)

Sonstige segmentale Konventionen

:	Dehnung, Längung, um ca. 0.2-0.5 Sek.
::	Dehnung, Längung, um ca. 0.5-0.8 Sek.
:::	Dehnung, Längung, um ca. 0.8-1.0 Sek.
?	Abbruch durch Glottalverschluss

Akzentuierung

akZENT	Fokusakzent
ak!ZENT!	extra starker Akzent

Tonhöhenbewegung am Ende von Intonationsphrasen

?	hoch steigend
,	mittel steigend
–	gleichbleibend
;	mittel fallend
.	tief fallend

Sonstige Konvention

<<erstaunt> >	interpretierende Kommentare mit Reichweite

Feintranskript

Akzentuierung

akZENT	Fokusakzent
akzEnt	Nebenakzent
ak!ZENT!	extra starker Akzent

Auffällige Tonhöhensprünge

↑	kleinere Tonhöhensprünge nach oben
↓	kleinere Tonhöhensprünge nach unten
↑↑	größere Tonhöhensprünge nach oben
↓↓	größere Tonhöhensprünge nach unten

Verändertes Tonhöhenregister

`<<t> >`	tiefes Tonhöhenregister
`<<h> >`	hohes Tonhöhenregister

Intralineare Notation von Akzenttonhöhenbewegungen

`` `SO ``	fallend
`´SO`	steigend
`¯SO`	gleichbleibend
`ˆSO`	steigend-fallend
`ˇSO`	fallend-steigend
↑`	kleiner Tonhöhensprung hoch zum Gipfel der Akzentsilbe
↓´	kleiner Tonhöhensprung herunter zum Tal der Akzentsilbe
↑¯SO bzw. ↓¯SO	Tonhöhensprünge zu auffallend höheren bzw. tieferen gleichbleibenden Akzenten
↑↑`SO bzw. ↓↓´SO	auffallend hohe bzw. tiefe Tonhöhensprünge zum Gipfel bzw. Tal der Akzentsilbe

Lautstärke- und Sprechgeschwindigkeitsveränderungen, mit Extension

`<<f> >`	forte, laut
`<<ff> >`	fortissimo, sehr laut
`<<p> >`	piano, leise
`<<pp> >`	pianissimo, sehr leise
`<<all> >`	allegro, schnell
`<<len> >`	lento, langsam
`<<cresc> >`	crescendo, lauter werdend
`<<dim> >`	diminuendo, leiser werdend
`<<acc> >`	accelerando, schneller werdend
`<<rall> >`	rallentando, langsamer werdend

Veränderung der Stimmqualität und Artikulationsweise

`<<creaky> >`	glottalisiert, "Knarrstimme"
`<<flüsternd> >`	Beispiel für Veränderung der Stimmqualität, wie angegeben

Summary of the most important GAT 2 transcription conventions

(see also Couper-Kuhlen, Elizabeth and Dagmar Barth-Weingarten (2011): A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 12: 1-51 (www.gespraechsforschung-ozs.de).

Minimal transcript

Sequential structure

`[ ]` `[ ]`	overlap and simultaneous talk

In- and outbreaths

`°h / h°`	in- / outbreaths of appr. 0.2-0.5 sec. duration
`°hh / hh°`	in- / outbreaths of appr. 0.5-0.8 sec. duration
`°hhh / hhh°`	in- / outbreaths of appr. 0.8-1.0 sec. duration

Pauses

`(.)`	micro pause, estimated, up to 0.2 sec. duration appr.
`(-)`	short estimated pause of appr. 0.2-0.5 sec. duration
`(--)`	intermediary estimated pause of appr. 0.5-0.8 sec. duration
`(---)`	longer estimated pause of appr. 0.8-1.0 sec. duration
`(0.5)/ (2.0)`	measured pause of appr. 0.5/ 2.0 sec. duration (to tenth of a second)

Other segmental conventions

`and_uh`	cliticizations within units
`uh, uhm, etc.`	hesitation markers, so-called "filled pauses"

Laughter and crying

`haha, hehe, hihi`	syllabic laughter
`((laughs)), ((cries))`	description of laughter and crying
`<<laughing> >`	laughter particles accompanying speech with indication of scope
`<<:-)>so>`	smile voice

Continuers

`hm, yes, no, yeah`	monosyllabic tokens
`hm_hm, ye_es, no_o`	bi-syllabic tokens
`?hm?hm`	with glottal closure, often negating

Other conventions

`((coughs))` `<<coughing> >`	non-verbal vocal actions and events with indication of scope
`( )`	unintelligible passage
`(xxx), (xxx xxx)`	one or two unintelligible syllables
`(may i)`	assumed wording
`(may i say/let us say)`	possible alternatives
`((unintelligible, appr. 3 sec))`	unintelligible passage with indication of duration

((...))	omission in transcript
->	refers to a line of transcript relevant in the argument

Basic transcript

Sequential structure

=	fast, immediate continuation with a new turn or segment (latching)

Other segmental conventions

:	lengthening, by about 0.2-0.5 sec.
::	lengthening, by about 0.5-0.8 sec.
:::	lengthening, by about 0.8-1.0 sec.
?	cut-off by glottal closure

Accentuation

SYLlable	focus accent
!SYL!lable	extra strong accent

Final pitch movements of intonation phrases

?	rising to high
,	rising to mid
–	level
;	falling to mid
.	falling to low

Other conventions

<<surprised> >	interpretive comment with indication of scope

Fine Transcript

Accentuation

SYLlable	focus accent
sYllable	secondary accent
!SYL!lable	extra strong accent

Pitch jumps

↑	smaller pitch upstep
↓	smaller pitch downstep
↑↑	larger pitch upstep
↓↓	larger pitch downstep

Changes in pitch register

<<l> >	lower pitch register
<<h> >	higher pitch register

Intralinear notation of accent pitch movements

`SO	falling
´SO	rising
¯SO	level
ˆSO	rising-falling
ˇSO	falling-rising

↑`	small pitch upstep to the peak of the accented syllable
↓´	small pitch downstep to the valley of the accented syllable
↑¯SO bzw. ↓¯SO	pitch jumps to higher or lower level accented syllables
↑↑`SO bzw. ↓↓´SO	larger pitch upsteps or downsteps to the peak or valley of the accented syllable

Loudness und tempo changes, with scope

<<f> >	forte, loud
<<ff> >	fortissimo, very loud
<<p> >	piano, soft
<<pp> >	pianissimo, very soft
<<all> >	allegro, fast
<<len> >	lento, slow
<<cresc> >	crescendo, increasingly louder
<<dim> >	diminuendo, increasingly softer
<<acc> >	accelerando, increasingly faster
<<rall> >	rallentando, increasingly slower

Changes in voice quality and articulation, with scope

<<creaky>>	glottalized
<<whispery>>	change in voice quality as stated