BMBF

Verbmobil

Verbundvorhaben

Untersuchungen zur Konsistenz prosodischer Etikettierungen

Matthias Reyelt

TU Braunschweig

Vm

Report 77
Juni 1995

Juni 1995

Matthias Reyelt

Institut für Nachrichtentechnik
Technische Universität Braunschweig
Schleinitzstraße 22
38092 Braunschweig

Tel.: (0531) 391 - 2479
Fax: (0531) 391 - 8218
e-mail: m.reyelt@tu-bs.de

Gehört zum Antragsabschnitt: 14.6 Prosodische Etikettierung

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BM-BF) unter dem Förderkennzeichen 01 IV 101 N0 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

Zusammenfassung

Um prosodische Information in einem automatischen Spracherkennungssystem analysieren zu können, muß prosodisch etikettiertes Trainings- und Testmaterial in großem Umfang zur Verfügung stehen. Dafür muß zunächst ein geeignetes Inventar prosodischer Etiketten entworfen werden, mit dem auch von wenig geschulten Transkribenten mit hoher Konsistenz prosodische Etikettierungen durchgeführt werden können. In einem ersten Experiment wurde ein Korpus von ca. 500 Sätzen parallel von nicht geschulten Testpersonen mit einem einfachen Etiketteninventar prosodisch etikettiert. Die Konsistenz dieser prosodischen Transkriptionen wurde untersucht.

Abstract

For the automatic analysis of prosody in speech recognition systems prosodically labeled speech data are required in large amounts for test and training purposes. To achieve this, appropriate prosodic labels have to be defined that are not only capable of describing prosody exactly but can be identified consistently even by transcribers without a profound theoretical knowledge of prosody. In a first experiment a corpus of nearly 500 utterances was labeled in parallel by untrained transcribers using a simple and basic inventory of prosodic labels. The consistency of the resulting prosodic transcriptions was investigated.

1 Einleitung

Automatische Spracherkennungssysteme beschränken sich zur Zeit meist auf die Analyse der im Sprachsignal enthaltenen segmentalen Information, während die suprasegmentale Information unberücksichtigt bleibt. In der Mensch–Mensch–Kommunikation erfüllt letztere Information allerdings wichtige Funktionen und ist für eine vollständige Dialoganalyse praktisch unverzichtbar (vgl. [3]). Sollen automatische Spracherkennungssysteme sinnvoll als Dialogsysteme eingesetzt werden, müssen diese Systeme gleichfalls eine Prosodieanalyse durchführen. Es gibt Ansätze dazu [4] [2] [15], jedoch sind die dabei erzielten Erkennungsraten deutlich geringer als die für die automatische Erkennung der segmentalen Information.

Vorbedingung für die Integration einer Prosodiekomponente in ein automatisches Spracherkennungssystem ist das Vorhandensein von verläßlich etikettiertem Sprachmaterial in ausreichenden Mengen für Training und Test. Während dies für die segmentale Information eher ein Zeitproblem ist, gibt es für prosodische Etikettierungen noch einige prinzipielle Schwierigkeiten. Je nachdem, was unter Prosodie verstanden wird, sind für prosodische Etikettierungen diverse Möglichkeiten denkbar. Im allgemeinsten Fall kann Prosodie als *das Suprasegmentale*, also "alles oberhalb der Lautebene" verstanden werden, hier kann eine prosodische Etikettierung alles vom Markieren syntaktischer Einheiten bis zur Beschreibung von Grundfrequenzkonturen umfassen. Im engen Sinne wird unter prosodischer Etikettierung analog der (breiten) phonetischen Etikettierung meist eine Markierung "prosodischer Phoneme" – also *Prosodeme* oder bei den meisten Systemen *Toneme* – verstanden. Anders als für die phonetische Etikettierung gibt es allerdings für das Deutsche zur Zeit kein Toneminventar, das allgemeine Akzeptanz gefunden hat.

Dies heißt nicht, daß die Intonation des Deutschen nicht untersucht worden wäre; es gibt viele Arbeiten zur Intonation des Deutschen. Allerdings lassen sich viele Beschreibungssysteme nur bedingt zur Analyse (und speziell zur Etikettierung) der Prosodie einsetzen.

Ein System prosodischer Kategorien, das zur Etikettierung größerer Mengen von Sprachdaten geeignet sein soll, muß besondere Anforderungen erfüllen. Es muß in der Lage sein, realisierte Prosodie zu beschreiben, und auch Abweichungen von "Normal-"Realisierungen berücksichtigen. Weiterhin muß die Bedeutung der Etiketten transparent sein, um ein einheitliches Verständnis der Transkriptionen zu gewährleisten. Die Etikettierungen müssen auch ohne größeren Qualitätsverlust von unterschiedlichen Transkribenten durchgeführt werden können, die nicht über fundierte theoretische Kenntnisse der Prosodie des Deutschen verfügen. Daher muß die "Etikettierbarkeit" ein von Anfang an mit zu berücksichtigendes Kriterium für ein prosodisches Beschreibungssystem sein. Das Etiketteninventar muß für Transkribenten relativ leicht erlernbar sein, um intersubjektive Inkonsistenzen, die besonders auftreten können, wenn Material an unterschiedlichen Orten etikettiert wird, soweit wie möglich zu vermeiden.

Grundsätzlich stellt sich bei prosodischen Etikettensystemen das Problem, daß die Definitionen ohne passende Sprachbeispiele nur schwer für andere nachvollziehbar sind. Bei existierenden Arbeiten zur Prosodieanalyse und prosodischen Transkription gibt es zur Lösung dieses Problems unterschiedliche Vorgehensweisen. Bei den Arbeiten von Batliner et.al. zu Satzmodus und –fokus (vgl. [1], [5]) wird auf eine formale Kategorisierung der Intonation bewußt verzichtet. Von der linguistischen Funktion wird direkt auf die prosodischen Parameter des Sprachsignals übergegangen. Dadurch ist eine genaue Beschreibung der prosodischen Realisierung von Modus und Fokus gegeben, die relativ gut nachvollziehbar ist. Andererseits enthält eine prosodisch–phonologische Etikettierung Informationen über menschliche Wahrnehmung, die auditiv wahrgenommen und nur schwer aus Merkmalen des Sprachsignals (etwa dem Verlauf der Grundfrequenz) ablesbar sind. Auch sagt eine rein funktionale Etikettierung wenig über die prosodische Realisierung dieser Funktion aus. Normalerweise werden Modus und Fokus nur selten rein prosodisch markiert. Andere Autoren (wie [11] und [6]) betonen daher die Wichtigkeit einer prosodischen Phonologie.

Für das Deutsche gibt es mehrere prosodische Beschreibungen, die speziell für die prosodische Etikettierung von Sprachdaten entworfen und verwendet wurden. Das in [8], [18] und [19] beschriebene Transkriptionssystem wurde speziell zur Beschreibung von Spontandialogen entwickelt und angewandt. In [12] wird das Kieler Intonationsmodell [10] für die prosodische Etikettierung angepaßt. Andere Beschreibungen (so z.B. [23], [21] und [7]) verwenden den Tonsequenzansatz, der schon Anfang der achtziger Jahre für das Amerikanische entwickelt wurde [13]. Unter Verwendung dieses Ansatzes wurde mittlerweile unter dem Namen ToBI (Tone and Break Indices) ein Standard für die prosodische Etikettierung des Englischen entwickelt [20]. Bei der Entwicklung dieses Standards wurden immer auch Konsistenzuntersuchungen durchgeführt, deren Ergebnisse in die (Re-)Definition der Etiketten mit einflossen.

Diese Transkriptionssysteme unterscheiden sich in der Form, in der Akzente bzw. Phrasengrenzen markiert werden; gemeinsam ist allen drei Systemen die Markierung der Lage von Akzenten und Phrasengrenzen. Alle drei Etikettierungen beruhen hauptsächlich auf akustischer Wahrnehmung, jedoch wird die optische Anzeige des Grundfrequenzverlaufs als Hilfsmittel herangezogen. Dabei besteht prinzipiell die Möglichkeit von Widersprüchen zwischen akustischer Wahrnehmung und optischem F_0-Verlauf, zwischen denen die Transkribenten entscheiden müssen. Bei allen drei Systemen wird der akustischen Perzeption eindeutig der Vorzug gegeben.

Bei den meisten Transkriptionssystemen wird der Aspekt der "Etikettierbarkeit" nicht oder zumindest nicht explizit berücksichtigt[1]. Die an dieser Stelle beschriebenen Untersuchungen sollten dazu dienen, ein eher einfaches Inventar prosodischer Etiketten so zu definieren, daß "naive" Transkribenten mit wenigen, möglichst präzisen Anweisungen ohne Training mit hinreichender Konsistenz prosodisch etikettieren können. In Abschnitt 2 werden prosodische Etikettierungen beschrieben, die von fünf untrainierten Studenten parallel an einem größeren Korpus durchgeführt wurden. Weiterhin etikettierten zwei der Studenten einen

[1] Eine Ausnahme ist das ToBI–System, hier sind Konsistenz der Etikettierungen und leichte Erlernbarkeit wichtige Kriterien für den Entwurf des Systems (vgl. [20]).

Teil des Materials zusätzlich bei verzerrter Sprache (Abschnitt 4). Hier sollte die Abhängigkeit von prosodischer Etikettierung und gesprochenem Text untersucht werden.

2 Durchgeführte Untersuchungen

In dem Experiment sollten die Testpersonen grundlegende prosodische Einheiten etikettieren, und zwar *akzentuierte Silben* und *Grenzen zwischen Intonationseinheiten (Phrasengrenzen)*. Um sicher zu gewährleisten, daß diese Etikettierungen rein auf der akustischen Perzeption beruhen, wurde auf die optische Darstellung der Grundfrequenz verzichtet. Um die Testpersonen nicht zu überfordern, wurde allerdings keine detaillierte Etikettierung kategorial verschiedener Akzente und Phrasengrenzen durchgeführt; es sollte nur das Vorhandensein eines Akzents oder einer Phrasengrenze etikettiert werden. Diese Etiketten sollten allerdings so beschrieben werden, daß sie auch für Testpersonen, die über wenig bis gar kein theoretisches Wissen über Prosodie verfügen, intuitiv verstanden werden können.

Die Definition der prosodischen Etiketten und der Etikettieranweisungen wurde daher auf Grundlage von Voruntersuchungen durchgeführt, die in [14] und [16] beschrieben wurden. In diesen Untersuchungen etikettierten verschiedene Testpersonen das gleiche Material nach unterschiedlichen Anweisungen. Obwohl im Prinzip jedesmal dieselben Einheiten etikettiert wurden, zeigte sich dort, daß schon minimal unterschiedliche Beschreibungen dieser Einheiten zu völlig verschiedenen Resultaten führen. Nach den Erfahrungen dieser Voruntersuchungen wurden das Etiketteninventar und die Anweisungen zur Etikettierung entwickelt.

Den fünf Testpersonen, die die Etikettierungen durchführten, wurden die folgenden Anweisungen gegeben:

Phrasengrenze: Markiere *Einschnitte* im Redefluß

Phrasenakzent: Etikettiere die am stärksten akzentuierte Silbe in jeder Phrase als *Phrasenakzent*

Nebenakzent: Etikettiere weitere akzentuierte Silben in der Phrase als *Nebenakzent*

Emphase: Falls der Phrasenakzent außergewöhnlich stark akzentuiert ist, etikettiere ihn als *Emphase*

Diese Anweisungen waren die einzige Information zum Etiketteninventar, die den Testpersonen gegeben wurde. Das Sprachmaterial wurde ihnen auf einer

Sprach-Arbeitsstation angezeigt. Das Sprachsignal wurde auf dem Bildschirm angezeigt und die Testpersonen konnten sich die komplette Äußerung oder beliebige Ausschnitte anhören, sooft sie wollten. Keine zusätzlichen visuellen Hilfsmittel wie z.B. Anzeige des F_0-Verlaufs wurden gegeben.

Als Sprachmaterial wurden 60 einzelne Äußerungen verwendet, die von acht verschiedenen Sprechern gelesen wurden. Insgesamt umfaßte das Korpus also 480 Sätze.

3 Ergebnisse

Die Ergebnisse zeigen, daß die Testpersonen keine prinzipiellen Schwierigkeiten mit dem Etiketteninventar hatten. Bild 1 zeigt die Anzahl vergebener Etiketten für die fünf Testpersonen. Drei von ihnen liegen sehr dicht zusammen bei Phrasengrenzen und Phrasenakzenten[2]. Die beiden anderen Testpersonen weichen merklich ab. Für das Etikett *Nebenakzent* sind die Zahlen relativ uneinheitlich, die "Schwelle" für die Wahrnehmung *akzentuiert* scheint doch recht subjektiv zu sein.

In Bild 2 ist die Übereinstimmung zwischen den Testpersonen gezeigt. Die Werte für die einzelnen Sprecher sind Mittelwerte über die Einzelkorrespondenzen zwischen zwei Testpersonen. Die Korrespondenz zwischen zwei Testpersonen wurde nach Gleichung 1 berechnet:

$$corr_{1,2,label} = \frac{n_{corr(1,2),label}}{(n_{1,label} + n_{2,label})/2} \tag{1}$$

Dabei ist $n_{corr(1,2),label}$ die Anzahl der übereinstimmend von Testperson 1 und 2 etikettierten Silben für ein spezifisches Etikett. $n_{1,label}$ gibt das Vorkommen dieses Etiketts in der Transkription der Testperson 1, $n_{2,label}$ das Vorkommen in der Transkription von Testperson 2 an. Bild 2 zeigt, daß nach der Etikettierung der ersten Sprecherin (AWE) die Übereinstimmung für Phrasenakzent und Phrasengrenze ansteigt und relativ konstant auf über 70 % bleibt. Eine Ausnahme ist der Sprecher HPT, für den die Übereinstimmung schlechter ist, was aber wahrscheinlich an dessen abweichender Intonation liegt.

Für Nebenakzent ist die Übereinstimmung wesentlich schlechter; offensichtlich ist die Unterscheidung *akzentuiert/nicht akzentuiert* schwierig ohne zusätzliche

[2]Natürlich sind beide Etiketten miteinander verkoppelt, da für jede Phrase genau ein Phrasenakzent etikettiert werden sollte. Da Anfang und Ende der Äußerungen nicht als Phrasengrenzen mitgezählt wurden, liegt also pro Satz die Anzahl der Phrasenakzente um eins höher als die Anzahl der Phrasengrenzen.

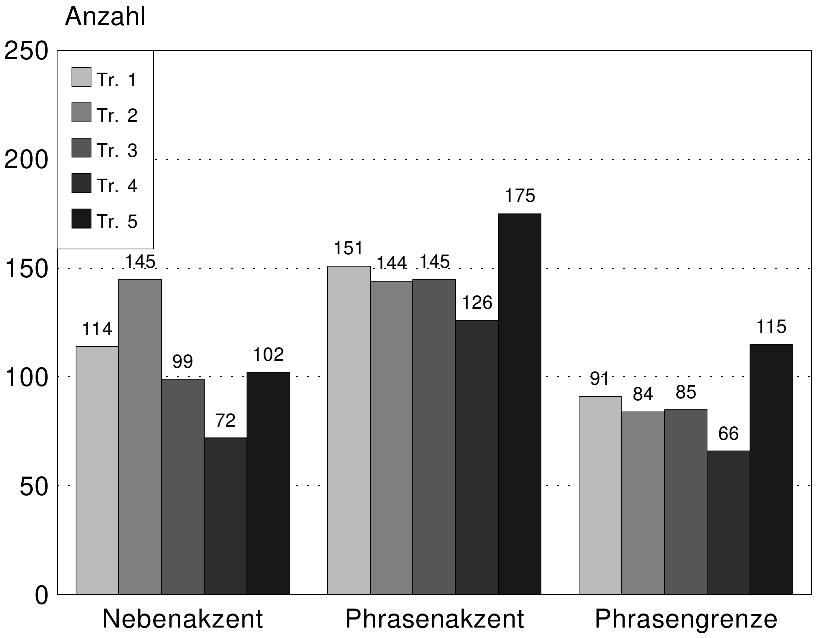

Abbildung 1: Häufigkeiten der Etiketten in den Transkriptionen der fünf Testpersonen, gemittelt über alle Sprecher (Gesamtzahl der Silben des Korpus war ca. 950)

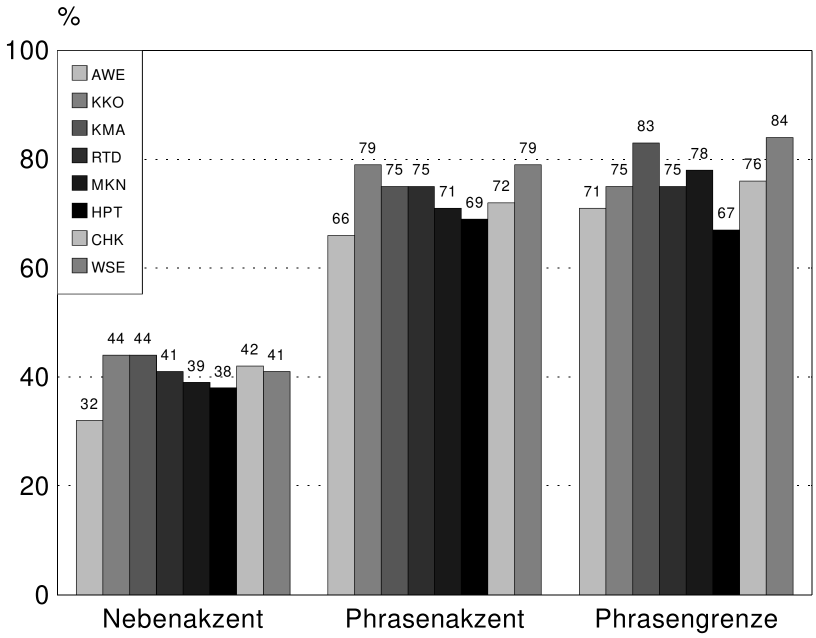

Abbildung 2: Übereinstimmungen für unterschiedliche Sprecher im Mittel über alle Testpersonen

Kriterien. Allerdings handelt es sich bei dem Sprachmaterial auch um gelesene Einzelsätze ohne Kontext. Bei Spontandialogen ist eine wesentlich ausgeprägtere prosodische Markierung der Sprache zu erwarten.

4 Etikettierung verzerrter Sprache

Um den Einfluß der linguistischen Struktur der Äußerung auf die prosodische Etikettierung zu untersuchen, wurden die 60 Äußerungen einer Sprecherin verzerrt. Dadurch sollte die segmentale Struktur zerstört und nur die suprasegmentale so gut wie möglich beibehalten werden.

Um dies zu erreichen, wurden die Kurzzeitspektren berechnet und deren Amplituden auf eine feste Schwelle begrenzt. Diese Spektren wurden mit dem Langzeitspektrum der Sprecherin multipliziert und auf die originale Lautheit normiert. Auf diese Art war der gesprochene Text unverständlich, die Prosodie blieb jedoch weitgehend erhalten. Diese Methode wurde gegenüber der in [22] angeführten Rotation des Spektrums bevorzugt, da die resultierende Sprache weniger unerträglich klingt. Dies Experiment wurde von zwei Testpersonen durchgeführt, und zwar nach genau denselben Anweisungen wie das Hauptexperiment. Die Reihenfolge der Sätze wurde geändert, so daß die Testpersonen keine Möglichkeit hatten, die Sätze zu identifizieren. Die Silbengrenzen wurden auf dem Bildschirm zusammen mit einer Nummer angezeigt, der die Akzente zugeordnet werden sollten.

Beide Testpersonen gaben nach dem Experiment an, daß die Äußerungen trotz der Verzerrung "natürlich klangen". Daher kann angenommen werden, daß zumindest die Intonation erhalten blieb. Allerdings hatten die Testpersonen große Schwierigkeiten, einen wahrgenommenen Akzent oder eine Grenze einer bestimmten Silbe zuzuordnen.

In Tabelle 1 ist die Anzahl vergebener Etiketten für verzerrte und unverzerrte Sprache verglichen. Die Anzahl realisierter Phrasen ist für verzerrte Sprache niedriger als für nicht verzerrte, was hauptsächlich dadurch zu erklären ist, daß Einleitungen wie *"Ja, ich..."* oder *"Guten Tag, ..."* wegen ihrer Kürze nicht als eigenständige Phrase etikettiert werden, wenn die Sprache verzerrt ist. In diesem Fall ist ein Einfluß der syntaktischen Struktur feststellbar. Ansonsten sind die Abweichungen eher gering. In Tabelle 2 sind die Übereinstimmungen zwischen den Transkriptionen verzerrter und unverzerrter Sprache angegeben. Werden die Übereinstimmungen mit denen für unverzerrte Sprache verglichen (vgl. Bild 2), so zeigt sich allgemein, daß die Werte bei verzerrter Sprache deutlich niedriger sind. Wird allerdings bei den Akzenten auf die Unterscheidung Phrasenakzent/Nebenakzent verzichtet, so liegt die Übereinstimmung auch für verzerrte

	Nebenakzent	Phrasenakzent	Phrasengrenze
Testperson 1	107/122	146/151	86/91
Testperson 2	116/115	154/182	94/122

Tabelle 1: Anzahl vergebener Etiketten der beiden Testpersonen für verzerrte/unverzerrte Sprache

	Nebenakzent	Phrasenakzent	Phrasengrenze
Testperson 1	45 %	62 %	48 %
Testperson 2	20 %	49 %	52 %

Tabelle 2: Übereinstimmung bei beiden Testpersonen zwischen verzerrter und unverzerrter Sprache

Sprache bei ca. 80 %. Es scheint daher, daß die Identifikation des Phrasenakzents teilweise durch die grammatische Wortkategorie beeinflußt wird, während die Unterscheidung akzentuiert/nicht akzentuiert stärker auf der Prosodie beruht.

Die Lage der Phrasengrenzen variierte stark. Wenn diese nicht durch eine Sprechpause markiert wurden, tendierten die Transkribenten dazu, Phrasengrenzen direkt vor einer akzentuierten Silben zu etikettieren. Wie schon in [9] gezeigt wird, werden Phrasengrenzen zumindest teilweise durch Dehnung realisiert. Diese Information geht natürlich zum Teil bei der Verzerrung der Sprache verloren. Weiterhin ist die Etikettierung der Phrasengrenzenposition vermutlich auch durch die syntaktische Oberflächenstruktur der Sätze beeinflußt, während bei verzerrter Sprache eventuell eher rhythmische Strukturen etikettiert werden.

5 Referenztranskription

Aus den resultierenden fünf parallelen Transkriptionen wurde eine prosodische Referenztranskription des Korpus generiert. Diese wurde durch Mehrheitsentscheidungen (3 von 5) ermittelt. Für etwa 3 % der Silben konnten keine klaren Mehrheiten für Akzente gefunden werden; in diesen Fällen wurde nach "korrekten" Phrasen (mit nur einem Phrasenakzent) entschieden.

Die Äußerungen waren identisch für alle Sprecher. Daher können auch Eigenheiten in ihrer Prosodie untersucht werden.

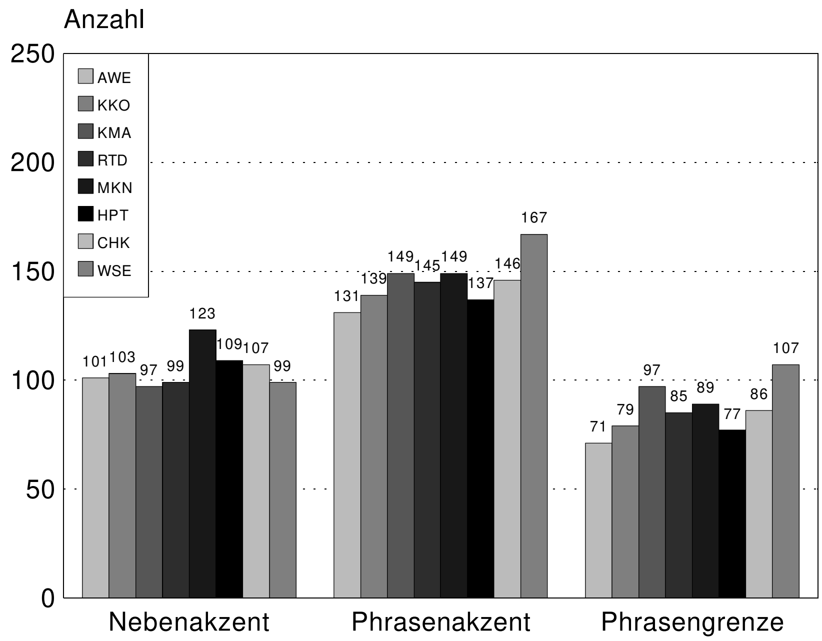

Abbildung 3: Häufigkeit prosodischer Etiketten im Vergleich der acht Sprecher (Basierend auf der Referenzetikettierung)

Es gibt einige Beispiele für variierenden lexikalischen Wortakzent in dem Material. Die Wörter *"überhaupt"* und *"direkt"*, die normal auf der letzten Silbe betont sind, werden von einigen Sprechern auf der ersten Silbe betont. Für das Wort *"Intercity"* verschiebt sich die Betonung bei zwei Sprechern von der dritten auf die erste Silbe in dem kontrastiven Kontext "...INtercity oder EUrocity ...".

In Bild 3 ist die Anzahl etikettierter Akzente für die acht Sprecher angegeben. Der Unterschied zwischen WSE und AWE beträgt immerhin 36; d.h. in mehr als der Hälfte der Äußerungen hat WSE eine Phrase mehr realisiert als AWE. Das zeigt, daß die prosodische Struktur stark sprecherabhängig ist und nur rudimentär vorhergesagt werden kann.

6 Transkriptionsfehler

Analysen der Transkriptionen zeigen, daß die Intonation, auch wenn sie nicht explizit etikettiert wurde, die Etikettierungen beeinflußt und für manche fehlerhaften Etikettierungen der nicht geschulten Transkribenten verantwortlich ist. In Bild 4 und 5 sind die F_0-Verläufe für das Ende der Frage *"nach Koblenz kommen"* gezeigt. Beide Sprecher haben einen Phrasenakzent auf *Koblenz*, jedoch führt die Realisierung als Gipfel bei Sprecher HPT zu einer starken F_0-Bewegung

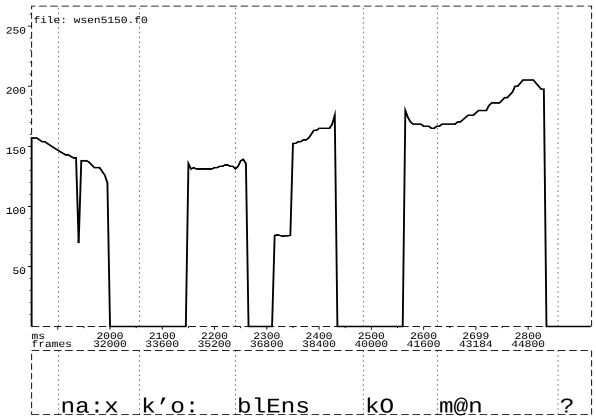

Abbildung 4: F_0-Verlauf für satzfinalen Teil *"nach Koblenz kommen?"* Sprecher WSE

am Satzende, um die Frageintonation zu erreichen. An dieser Stelle wurde von der Mehrheit der Testpersonen ein zusätzlicher Nebenakzent etikettiert. Dies ließe sich vermeiden, wenn das Inventar phonologische Kategorien für Phrasengrenzen und Akzente enthielte. Diese Fehler zeigen die Grenzen solch "naiver" prosodischer Etikettierung und die Notwendigkeit eines erweiterten Etiketteninventars.

7 Zusammenfassung

In dem hier beschriebenen Experiment wird eine Form der prosodischen Etikettierung gesucht, die so einfach ist, daß sie auch von nicht geschulten Transkribenten mit verhältnismäßig guter Konsistenz durchgeführt werden kann. Daß dies prinzipiell möglich ist, zeigen die Ergebnisse der Untersuchungen. Die Konsistenz der Transkriptionen liegt bei etwa 70–80% für Phrasengrenzen und -akzente, aber mit ca. 40% deutlich schlechter für Nebenakzente.

Allerdings kann das Experiment nur als erster Schritt zur prosodischen Etikettierung bezeichnet werden. Das verwendete Inventar prosodischer Etiketten ist nicht reichhaltig genug, um alle Ebenen der Prosodie zu beschreiben; insbesondere fehlt noch völlig die Etikettierung der Intonation. Dadurch können die etikettierten Daten nur eingeschränkt zur Prosodieanalyse eingesetzt werden.

Andererseits ist es gerade bei der prosodischen Etikettierung notwendig, zunächst die Etikettierbarkeit eines einfachen Inventars zu untersuchen, bevor von einem detailliert ausgearbeiteten Modell der deutschen Prosodie ausgehend eine

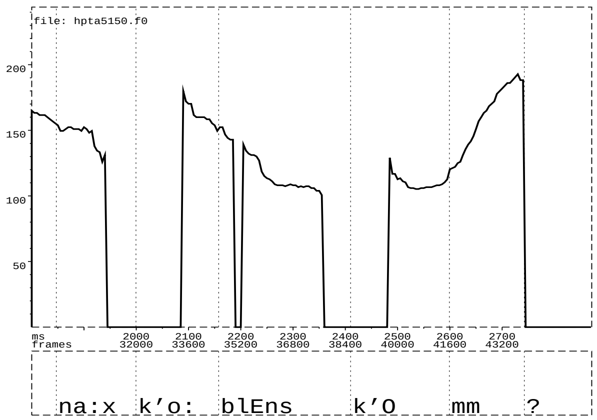

Abbildung 5: F_0-Verlauf für satzfinalen Teil *"nach Koblenz kommen?"* Sprecher HPT

vollständige Etikettierung durchgeführt werden kann.

Es besteht sonst die Gefahr, daß sich die Schulung der Transkribenten zu stark an der Theorie orientiert und zuwenig an deren Wahrnehmungsvermögen. Wenn zusätzlich noch die Konsistenz der Etikettierungen durch den Einsatz audiovisueller Hilfsmittel (von der optischen Anzeige der Grundfrequenz bis zu automatisch generierten Vorschlägen für die Etikettierung) verbessert werden soll, kann es vorkommen, daß die Transkribenten im Zweifelsfall eher gegen ihre auditive Wahrnehmung entscheiden. Dies Experiment zeigt, daß eine einigermaßen verläßliche prosodische Etikettierung auch ohne Hilfsmittel rein auditiv durchgeführt werden kann.

Als nächster Schritt wird eine Erweiterung des Etiketteninventars erfolgen. Vor allem sollen Erfahrungen zur Etikettierung der Intonation gesammelt werden. Dazu soll ein Schulungsprogramm für Transkribenten entwickelt werden. Die Erfahrungen sollen in die Entwicklung einer Arbeitsstation zur prosodischen Etikettierung einfließen, an der zur Zeit gearbeitet wird [17]. Auch soll verstärkt spontane Sprache etikettiert werden, die mittlerweile in großem Umfang verfügbar ist.

Literatur

[1] Hans Altmann, Anton Batliner, & Wilhelm Oppenrieder. *Zur Intonation von Modus und Fokus im Deutschen.* Linguistische Arbeiten 234. Niemeyer, Tübingen, 1989.

[2] Paul C. Bagshaw. An investigation of acoustic events related to sentential stress and pitch accents, in English. *Speech Communication*, 13:333–342, 1993.

[3] A. Batliner, A. Kießling, R. Kompe, E. Nöth, & B. Raithel. Wann geht der Sonderzug nach Pankow? (Uhrzeitangaben und ihre prosodische Markierung in der Mensch-Mensch– und in der Mensch-Maschine–kommunikation). In *Fortschritte der Akustik - DAGA92*, S. 541–544, Bad Honnef, 1992. DPG–GmbH.

[4] A. Batliner, R. Kompe, A. Kießling, E. Nöth, H. Niemann, & U. Kilian. The prosodic marking of accents and phrase boundaries: Expectations and results. In A.J. Rubio, Hrsg., *NATO ASI: New Advances and Trends in Speech Recognition and Coding*, S. 89–92, Bubion (Granada), June-July 1993.

[5] A. Batliner, W. Oppenrieder, E. Nöth, & G. Stallwitz. The intonational marking of focal structure: wishful thinking or hard fact? In *Proceedings ICPhS*, S. 278–281, Aix-en-Provence, 1991.

[6] Anne Cutler & D. Robert Ladd. *Prosody: Models and Measurements.* Springer, Berlin/New York, 1983.

[7] Caroline Féry. *German Intonational Patterns.* Linguistische Arbeiten 285. Niemeyer, Tübingen, 1993.

[8] Dafydd Gibbon & Margret Selting. Intonation und die Strukturierung eines Diskurses. *Zeitschrift für Literaturwissenschaft und Linguistik*, 49:53–73, 1983.

[9] Klaus J. Kohler. Prosodic boundary signals in german. *Phonetica*, 40:89–134, 1983.

[10] Klaus J. Kohler. Form and function of intonation peaks in german. In K.J. Kohler, Hrsg., *Arbeitsberichte Nr. 25 des Instituts für Phonetik und Digitale Sprachsignalverarbeitung*, S. 11–28. Inst. für Phonetik, Universität Kiel, 1991.

[11] Klaus J. Kohler. A model of German intonation. In K.J. Kohler, Hrsg., *Arbeitsberichte Nr. 25 des Instituts für Phonetik und Digitale Sprachsignalverarbeitung*, S. 295–360. Inst. für Phonetik, Universität Kiel, 1991.

[12] Klaus J. Kohler. Prosodisches Transkriptionssystem für die Etikettierung von Sprachsignaldaten. In K.J. Kohler, Hrsg., *Arbeitsberichte Nr. 26 des Instituts für Phonetik und Digitale Sprachsignalverarbeitung*, S. 239–250. Inst. für Phonetik, Universität Kiel, 1992.

[13] Janet Pierrehumbert. *The Phonology and Phonetics of English Intonation.* PhD thesis, M.I.T., 1980.

[14] M. Reyelt. Untersuchungen zur Etikettierung prosodischer Einheiten. In *Fortschritte der Akustik - DAGA 93*, S. 964–967, Bad Honnef, 1993. DPG-GmbH.

[15] Matthias Reyelt. Automatische Extraktion prosodischer Merkmale aus den Verläufen von Sprachgrundfrequenz und Lautheit. In G. Görz, Hrsg., *KONVENS 92*, S. 385–389, Berlin, 1992. Springer.

[16] Matthias Reyelt. Experimental investigation on the perceptual consistency and the automatic recognition of prosodic units in spoken German. In *Working papers*, Vol. 41, S. 238–241, Lund University, 1993. Dept. of Linguistics.

[17] Matthias Reyelt. Ein flexibles Programmpaket zur Visualisierung von Sprachdaten. In K. Fellbaum, Hrsg., *Tagungsband Elektronische Sprachsignalverarbeitung*, Berlin, 1994. Erscheint demnächst.

[18] Margret Selting. Descriptive categories for the auditive analysis of intonation in conversation. *Journal of Pragmatics*, 11:777–791, 1987.

[19] Margret Selting. The role of intonation in the organization of repair and problem handling sequences in conversation. *Journal of Pragmatics*, 12:293–322, 1988.

[20] Kim Silverman, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, & Julia Hirschberg. Tobi: A standard for labeling english prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, S. 867–870, 1992.

[21] Susanne Uhmann. *Fokusphonologie: eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie.* Linguistische Arbeiten 252. Niemeyer, Tübingen, 1991.

[22] Alex Waibel. *Prosody and Speech Recognition.* Pitman, 1988.

[23] Dieter Wunderlich. Der Ton macht die Melodie – Zur Phonologie der Intonation des Deutschen. In Hans Altmann, Hrsg., *Intonationsforschungen*, Linguistische Arbeiten 200, S. 1–40. Niemeyer, Tübingen, 1988.

Das diesem Bericht zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministers für Forschung und Technologie unter dem Förderkennzeichen 01 IV 101 N0 gefördert. Die Verantwortung für den Inhalt liegt beim Autor.