BMBF

Verbmobil

Verbundvorhaben

Abbildung prosodischer Etiketten auf prototypische F0-Verläufe

Jörg Reinecke

TU Braunschweig

Vm

Report 92
Juli 1995

Juli 1995

Jörg Reinecke

Institut für Nachrichtentechnik
Technische Universität Braunschweig
Schleinitzstraße 22
38092 Braunschweig

Tel.: (0531) 391 - 2479
Fax: (0531) 391 - 8218
e-mail: reinecke@ifn.ing.tu-bs.de

Gehört zum Antragsabschnitt: 14.3 Werkzeuge zur prosodischen Etikettierung

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

Zusammenfassung

Der vorliegende Artikel befaßt sich mit der Möglichkeit der Abbildung prosodischer Etiketten auf zugehörige Sprachgrundfrequenzverläufe. Diese werden im Rahmen einer Resynthese der Originaläußerung verwendet, die den Transkribenten beim prosodischen Etikettieren die Möglichkeit der auditiven Kontrolle ihrer Etikettenvergabe ermöglichen soll. Ausgehend von einer manuell erstellten prosodischen Etikettierung wird die Ausarbeitung eines Modells zur Steuerung des Umsetzungsprozesses und die Festlegung der dabei auftretenden freien Parameter beschrieben. Durch den Vergleich einer unter Zuhilfenahme der Resynthese erzeugten Transkription zweier Testäußerungen mit der zugehörigen manuell erstellten Transkription wird eine erste Güteabschätzung des Verfahrens vorgenommen.

1 Einleitung

Der sprachliche Kommunikationsprozeß zwischen Sprecher und Zuhörer wird wesentlich durch die Prosodie beeinflußt. Sie vermittelt dem Zuhörer unter anderem Informationen, die über den reinen Wortlaut der Äußerung hinausgehen. Hierbei kann es sich um so unterschiedliche Dinge wie die Intention des Sprechers, dessen Gemütslage oder dessen Alter und Geschlecht handeln. Je nach Art der Information und abhängig von der Betrachtungsweise kann von relevanten, redundanten und irrelevanten Prosodiekomponenten gesprochen werden [2].

Aus Sicht der Spracherkennung seitens eines Menschen oder einer Maschine beinhaltet der redundante Anteil Informationen, die bereits im Wortlaut der Äußerung enthalten sind und beim Erkennungsprozeß als zusätzliche Merkmale hinzugezogen werden können.

Der relevante Anteil umfaßt in diesem Zusammenhang Prosodieanteile, die über den Wortlaut hinausgehende Informationen vermitteln. Dies ist zum Beispiel die Frageintonation bei Sätzen mit Aussagestruktur oder die Wortakzentuierung bei Wörtern wie *umfahren*, deren doppelsinnige Bedeutung allein aufgrund der unterschiedlichen Wortakzentuierung aufgelöst werden kann.

Der irrelevante Anteil der Prosodie beinhaltet all die Informationen, die aus Sicht der Spracherkennung nicht benötigt werden, wie etwa Angaben über die Gemütslage des Sprechers. Diese Informationen können allerdings für andere wissenschaftliche Disziplinen von besonderer Wichtigkeit sein, so daß die hier vorgenommene Klassifizierung prosodischer Informationen immer kontextsensitiv zu sehen ist.

Im Rahmen der Forschung zur automatischen Spracherkennung wird versucht, die im Sprachsignal enthaltene redundante und relevante prosodische Information beim Erkennungsprozeß mit auszuwerten und somit die Erkennung robu-

ster, schneller bzw. weniger aufwendig zu gestalten. Insbesondere gilt dies bei der Erkennung fließend gesprochener Spontansprache, da hierbei die korrekte Artikulation der Laute nachläßt und die feste Syntax- und Grammatikstruktur der Lesesprache aufbricht.
Zum Training derartiger Erkenner wird prosodisch etikettiertes Sprachmaterial in erheblichem Umfang benötigt. Die Erstellung dieses Materials geschieht derzeit ausschließlich von Hand durch geeignet geschulte Transkribenten. Diese vergeben für jede von ihnen auditiv wahrgenommene prosodische Kategorie ein Etikett aus einem festen Etiketteninventar. Als technische Hilfsmittel stehen ihnen dabei Rechnerarbeitsplätze zur Verfügung, die die visuelle Darstellung des Sprachsignals und daraus abgeleiteter Parameterverläufe, beispielsweise der Sprachgrundfrequenz oder der Signalenergie, erlauben. Ferner bieten sie die Möglichkeit der akustischen Ausgabe des gesamten Sprachsignals oder von Teilen daraus. Die Schwierigkeit für den Transkribenten besteht in der Zuordnung eines Etiketts zu einer von ihm wahrgenommenen Kategorie. Er muß diese Zuordnung in Erinnerung an seine Schulung vornehmen und hat keine Möglichkeit, seine Auswahl in irgendeiner Form zu überprüfen.
An dieser Stelle setzt das am Institut für Nachrichtentechnik entwickelte Konzept einer Arbeitsstation zur prosodischen Etikettierung an [3]. Neben den üblichen visuellen und akustischen Darstellungsformen des Signals bietet es die Möglichkeit der Resynthese der zu untersuchenden Äußerung unter Berücksichtigung der prosodischen Etikettierung. Dies bedeutet für den Transkribenten, daß er jederzeit die von ihm vergebenen Etiketten im Kontext der gesamten Äußerung auditiv verifizieren kann. Hierdurch soll die Fehleranfälligkeit der prosodischen Etikettierung verringert und gleichzeitig deren Konsistenz erhöht werden.
Die prinzipiellen Grundlagen der Umsetzung prosodischer Etiketten und der anschließenden Resynthese wurden in [3] beschrieben. Die hier vorgestellten Arbeiten betreffen die Abbildung konkreter prosodischer Etiketten auf zugehörige prototypische F_0-Verläufe.

2 Prosodische Etikettierung

Die prosodische Etikettierung umfaßt das Auffinden prosodischer Kategorien im Sprachsignal und deren Kennzeichnung durch Etiketten aus einem festen Etiketteninventar. Für das Deutsche existiert noch kein allgemein anerkanntes prosodisches Kategoriensystem und daher auch kein damit verbundenes Etiketteninventar. Die hier vorgestellten Arbeiten zur Umsetzung prosodischer Etiketten beziehen sich auf ein im Rahmen des Verbundprojektes VERBMOBIL verwendetes Etikettiersystem, das auf dem Tobi-System für amerikanisches Englisch basiert [5]. Die Markierung prosodischer Ereignisse erfolgt dabei auf drei Ebenen:

- Markierung prosodischer Phrasen
- Markierung der Akzentuierung innerhalb jeder Phrase
- Markierung des Intonationsverlaufs an Akzent- und Phrasengrenzpositionen.

Die eingangs erwähnte auditive Kontrolle der Etikettenvergabe durch den Transkribenten bezieht sich dabei im wesentlichen auf den dritten Punkt, der Etikettierung des Intonationsverlaufs.
Die Markierung der Intonationskontur geschieht durch Abbildung des komplexen Verlaufs auf zwei Töne, *hoch* und *tief* (H und L), wobei der hohe Ton noch in bestimmten Konstellationen als *Downstep* H (!H) auftreten kann. Gekennzeichnet wird hierdurch ein Absinken der Tonhöhe zweier aufeinanderfolgender hoher Töne. Je nach Auftreten der Töne werden drei Gruppen von prosodischen Etiketten unterschieden:

- Akzenttöne
- Phrasentöne
- Grenztöne.

Akzenttöne treten im Bereich der Akzentpositionen auf und werden mit L* für einen tiefen und H* für einen hohen Ton markiert. Daneben können diese Töne auch bitonal als Übergang von hoch nach tief bzw. umgekehrt auftreten. Die Charakterisierung dieser Tonübergänge geschieht durch Hinzufügen eines zweiten Tones vor oder hinter dem eigentlichen Akzentton.

L*	tiefe Wahrnehmung im Akzent
H*	hohe Wahrnehmung im Akzent
L*+H	tiefe Wahrnehmung im Akzent mit nachfolgendem Anstieg der Intonation
H+!H*	tiefe Wahrnehmung im Akzent mit vorangegangenem Abfall der Intonation (der Systematik nach H+L*)
H*+L	hohe Wahrnehmung im Akzent mit nachfolgendem Abfall der Intonation
L+H*	hohe Wahrnehmung im Akzent mit vorangegangenem Anstieg der Intonation

Phrasentöne werden isoliert an intermediären Phrasengrenzen markiert. Dies sind schwache Grenzen innerhalb einer prosodischen Phrase. Hierzu stehen die Etiketten L- und H- zur Verfügung, je nachdem ob die Intonation an dieser Stelle tief oder hoch wahrgenommen wird.
Grenztöne werden an prosodischen Phrasengrenzen etikettiert. Auch hier stehen zwei Töne L% und H% zur Verfügung. Diese werden aber nicht isoliert vergeben, sondern stets in Verbindung mit einem vorangestellten Phrasenton L- oder H-.

Die zeitliche Lage dieses Phrasentons im Intonationsverlauf wird nicht eindeutig beschrieben. Er bezieht sich auf einen Bereich zwischen der letzten Akzentmarkierung und der Phrasengrenze. Durch informelle Begutachtung verschiedener Äußerungen wurde festgestellt, daß seine Ausprägung meist im Bereich der Silbe liegt, die der Akzentsilbe folgt.

L-L%	Intonation fällt unmittelbar nach letzter Akzentposition und bleibt tief
L-H%	Intonation fällt unmittelbar nach letzter Akzentposition und steigt zur Phrasengrenze hin an
H-L%	Intonation fällt nach letzter Akzentposition bis zur Phrasengrenze hin ab
H-H%	Intonation bleibt nach letzter Akzentposition bis zur Phrasengrenze auf hohem Niveau

3 Umsetzung prosodischer Etiketten

Die prinzipiellen Grundlagen der hier beschriebenen Umsetzung sowie die systemtechnischen Voraussetzungen sind in [4] beschrieben. Der Abbildung prosodischer Etiketten auf zugehörige F_0-Verläufe liegt ein Modell zugrunde, das in Anlehnung an ein in [1] beschriebenes Intonationsmodell entwickelt wurde. Jedem Ton (L, !H, H) ist ein eigenes diskretes Grundfrequenzniveau zugeordnet. In Halbtondarstellung bilden diese drei Niveaus zueinander parallele Geraden. Die Angabe der Niveauhöhe erfolgt relativ zum L-Niveau, so daß sich Männer- und Frauenstimmen nur in der absoluten Höhe des L-Niveaus unterscheiden. Sämtliche Grundfrequenzbewegungen beginnen bzw. enden auf einer der Niveaulinien, wobei ein Wechsel von einem Niveau zum anderen oder eine Bewegung entlang einer Niveaulinie stattfinden kann. Das Modell besitzt folgende freie Parameter, die zu bestimmen sind:

- Abstand der Niveaulinien zum Grundniveau (L-Niveau)
- Zeitliche Synchronisierung der Akzenttöne zur Akzentmarkierung
- Dauer des bitonalen Übergangs
- Zeitlicher Abstand des Phrasentons zur vorangehenden Akzentmarkierung

Zur Festlegung der freien Parameter wurden aus der prosodischen Etikettierung eines Verbmobil-Dialogs (N002K, Terminabsprache zwischen einem Mann und einer Frau) besonders charakteristische Realisierungen der unterschiedlichen Etiketten ausgewählt und hinsichtlich der Modellparameter untersucht. Hierzu wurden die Äußerungen mit unterschiedlichen Modellparametern resynthetisiert und

vom Autor mit den Originaläußerungen auditiv verglichen. Als guter Kompromiß zwischen männlicher und weiblicher Stimme erwies sich eine Niveaudifferenz zwichen L und H von 6 HT (Halbtönen) und zwischen L und !H von 3 HT.
Die zeitliche Synchronisierung von akzentuierenden F_0-Bewegungen erfolgt bei herkömmlichen Synthesesystemen meist relativ zum Beginn des akzentuierten Vokals [7]. Demgegenüber erfolgt hier die Markierung der Akzentposition im Silbenkern, wobei es letztendlich dem Transkribenten überlassen bleibt, wo genau er die Markierung setzt. Dies bedeutet für die Resynthese, daß der Vokalbeginn nicht bekannt ist und daß die Position der Etiketten hinsichtlich der Lage innerhalb der Silbe gewissen Schwankungen unterworfen ist. Trotz dieser Unsicherheit wurde die Position der Akzenttöne auf die Akzentmarkierung gesetzt. Die damit erzielten sehr guten Syntheseresultate bestätigen nachträglich diese Wahl.
Für die Dauer des bitonalen Übergangs wurden ebenfalls mehrere unterschiedliche Werte durch Resynthese und auditiven Vergleich ausprobiert. Letztendlich stellte sich ein Wert von 150ms als sehr guter Kompromiß heraus.
Wie bereits bei der Beschreibung der prosodischen Etikettierung erläutert wurde, ist die Ausprägung des Phrasentons nicht eindeutig bestimmt. Um dennoch zu praktikablen Ergebnissen zu gelangen, wurde postuliert, daß er sich im Bereich der dem Akzent folgenden Silbe befindet. Zur Festlegung des zeitlichen Abstands des Phrasentons zur vorausgehenden Akzentmarkierung wurden wiederum Resyntheseexperimente mit unterschiedlichen Werten durchgeführt. Auch hier erwies sich ein Wert von 150ms als sehr gut. Die Experimente haben jedoch gezeigt, daß dieser Wert nur gilt, wenn ein Tonwechsel zwischen letztem Akzentton und folgendem Phrasenton aufgetreten ist. Bei gleichbleibenden Ton verschmelzen beide Töne miteinander. Faktisch bedeutet dies, daß hierbei der Phrasenton nicht realisiert wird. Eine weitere Konsequenz der gleichen Werte für die Dauer des bitonalen Übergangs und des zeitlichen Abstands des Phrasentons zur vorangehenden Akzentmarkierung ist, daß bei der Resynthese zwischen Etikettenfolgen wie L*, H-H% und L*+H, H-H% nicht unterschieden werden kann.
Für jedes prosodische Etikett sind die zugehörigen Modellparameter in einer Liste abgelegt. Zum Synthesezeitpunkt wird aus der Abfolge der prosodischen Etiketten unter Berücksichtigung der Einträge in dieser Liste eine Folge diskreter Punkte im Halbton-Zeit-Bereich generiert. Die synthetische F_0-Kontur ergibt sich durch die geradlinige Verbindung dieser Punkte.

4 Untersuchungen

Der eigentliche Zweck der Umsetzung prosodischer Etiketten und daran anschließender Resynthese ist die eingangs beschriebene auditive Verifikation der Etikettenvergabe seitens des Transkribenten. Der direkte Nachweis, daß mit Hilfe

	L+H*	L+!H*	L*+H	L*+!H	!H+L*	H+L*	H+!H*	!H*	H*	L*	?*	L-L%	L-H%	H-L%	H-H%	L-	H-	?-	?%	#	SUMME
L+H*	10	.	5	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	20
L+!H*	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1
L*+H	.	.	28	2	.	.	.	.	6	1	1	.	.	.	.	.	.	.	.	.	38
L*+!H	.	.	1	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	3
!H+L*	.	.	.	.	.	.	2	.	2	1	.	.	.	.	.	.	.	.	.	.	5
H+L*	.	.	.	.	.	.	3	1	2	1	.	.	.	.	.	.	.	.	.	.	7
H+!H*	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1
!H*	.	.	1	.	.	.	.	18	9	1	1	.	.	.	.	.	.	.	.	.	30
H*	2	.	4	.	.	.	1	.	61	2	2	.	.	.	.	.	.	.	.	.	72
L*	.	.	.	.	.	.	1	2	1	11	.	.	.	.	.	.	.	.	.	.	15
?*	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1
L-L%	.	.	.	.	.	.	.	.	.	.	.	48	.	4	1	.	2	.	.	1	56
L-H%	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	2
H-L%	.	.	.	.	.	.	.	.	.	.	.	2	.	6	2	.	.	.	1	.	11
H-H%	.	.	.	.	.	.	.	.	.	.	.	.	.	1	20	.	.	.	.	.	21
L-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	3	3	.	33	62
H-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	21	1	.	13	37
?-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	2
?%	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
#	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1
SUMME	12	.	39	2	.	.	8	21	89	18	5	52	.	11	23	25	26	6	1	47	.

Tabelle 1: Korrespondenzmatrix Originaltranskription / Resynthese-Transkription

dieses Verfahrens die gewünschte Verbesserung der Transkriptionsergebnisse erzielt werden kann, ist äußerst aufwendig. Hierzu müßte ein großes Testmaterial von hinreichend vielen Transkribenten mit und ohne Unterstützung durch die Resynthese prosodisch etikettiert werden und anschließend untersucht werden, wie sich die Transkriptionen unterscheiden. Um dennoch eine Aussage über die Güte des vorgestellten Verfahrens treffen zu können wurde vom Autor eine Untersuchung durchgeführt, die klären sollte, wie sich eine herkömmlich erstellte Etikettierung verändert, wenn das gleiche Material mit dem neuen Verfahren noch einmal etikettiert wird. Hierzu wurden aus dem Verbmobil-Material die beiden Dialoge N002K und N019K ausgewählt, bei denen es sich um Terminabsprachen zwischen einem Mann und einer Frau bzw. zwischen zwei Männern handelt.
Ausgangspunkt der Untersuchung war die zugehörige prosodische Etikettierung. Die Äußerungen wurden zunächst unter Berücksichtigung dieser Originaletikettierung resynthetisiert und mit dem tatsächlichen Sprachsignal auditiv verglichen. Zeigten sich hierbei wahrnehmbare Unterschiede, wurde durch Auswahl anderer Etiketten versucht, die akustische Übereinstimmung an den betreffenden Positionen zu erhöhen. Dabei wurden nach Möglichkeit lediglich intonatorische Etiketten ausgetauscht und keine neuen Akzent- oder Phrasengrenzpositionen festgelegt. Die Ergebnisse dieser Untersuchung sind in Tabelle 1 wiedergegeben. Die abgebildete Matrix zeigt die Korrespondenzen zwischen den Etiketten der Origi-

naltranskription (horizontal) und denen der mit Hilfe der Resynthese erstellten Transkription (vertikal). Die mit ? gekennzeichneten Etiketten zeigen an, daß sich der Transkribent an der entsprechenden Stelle für kein gültiges Intonationsetikett entscheiden konnte. Die # markieren Etiketten, für die keine Korrespondenzen in der jeweils anderen Transkription zu finden waren. Der Grund hierfür liegt in Einfügungen bzw. Löschungen von Etiketten in der Resynthese-Transkription. Die hohe Zahl der L- und H- Einfügungen ist durch die Eigenheit der Resynthese zu erklären, auch zu Beginn einer Äußerung die Angabe eines Niveaus zu erwarten. In der Originaletikettierung wird aber nur das Phrasenende markiert. Dies bedeutet, daß zu Beginn jeder Äußerung ein zusätzliches L- bzw. H- vergeben werden muß, je nachdem ob die Wahrnehmung tief oder hoch ist.
Die dargestellten Resultate lassen sich auf unterschiedliche Weise interpretieren. Werden die Korrespondenzen entlang der Hauptdiagonalen betrachtet, ergeben sich die folgenden Übereinstimmungen bei den einzelnen Kategorien:

- Akzente 68 %
- Grenz- und Phrasentöne (ohne Einfügungen) 82 %

das heißt, daß 68 % aller in der Originaltranskription enthaltenen *Akzentetiketten* nach der auditiven Verifikation beibehalten wurden und andererseits 32 % der Originaletiketten ersetzt wurden. Für die Etikettierung der Intonation im Bereich der Phrasen- und Grenztöne liegt der Übernahmeanteil deutlich höher bei 82 %. Diese Diskrepanz läßt sich eventuell durch die größere Zahl der Etiketten zur Markierung der Akzentinitonation erklären, die eine feinere Differenzierung zulassen.
Zur Überprüfung dieser These wurde für die *Akzentetiketten* Tabelle 1 weiter ausgewertet und nicht nur die direkten Korrespondenzen berücksichtigt, sondern zusätzlich die restlichen Zuordnungen hinsichtlich der Parameter *gleiche Kategorie* und *andere Kategorie* ausgewertet. Die daraus resultierenden Ergebnisse sind in Tabelle 2 dargestellt.
In der Spalte *gleiche Kategorie* sind alle Ersetzungen der Originaletiketten durch Etiketten der gleichen Kategorie zusammengefaßt. Gleiche Kategorie bedeutet in diesem Zusammenhang, daß bei beiden Etiketten der Akzentton gleich ist, also entweder H* oder L*. Demgegenüber weisen die in der Spalte *andere Kategorie* dargestellten Korrespondenzen unterschiedliche Akzenttöne auf.
Bei Betrachtung der Ergebnisse der 1. Spalte fallen sofort die beiden Etiketten L*+!H und H+!H* auf, die hier gar keinen bzw. nur einen sehr geringen Grad der Übereinstimmung aufweisen. Beim L*+!H liegt dies sicherlich auch am seltenen Auftreten des Etiketts, lediglich zweimal wurde es vergeben.
Das Etikett H+!H* ist, wie eingangs erwähnt, seiner Systematik nach als H+L* zu werten. Dies wird durch die in Tabelle 1 und 2 angegebenen Werte noch einmal

Etikett	identisch		gleiche Kateg.		andere Kateg.	
L+H*	10	(83 %)	2	(17 %)		
L*+H	28	(72 %)	1	(3 %)	10	(25 %)
L*+!H			2	(100 %)		
H+!H*	1	(12 %)	6	(75 %)	1	(12 %)
!H*	18	(86 %)	2	(9 %)	1	(4 %)
H*	61	(68 %)	15	(17 %)	13	(15 %)
L*	11	(65 %)	3	(17 %)	3	(17 %)
SUMME	129	68 %)	31	(16 %)	28	(15 %)

Tabelle 2: Übereinstimmungen und Ersetzungen (nach Kategorien aufgeschlüsselt)

unterstrichen.
Werden die Werte aus Tabelle 2 für die identische Korrespondenz und die Ersetzungen der Originaletiketten durch ein Etikett der gleichen Kategorie (Spalten 1 und 2) zusammengefaßt, ergibt sich eine Gesamtübereinstimmung der beiden Etikettierungen von 84 %. Der Wert liegt in der Größenordnung des Wertes, der für die Phrasen- und Grenztonetikettierung ermittelt wurde. Dieses Resultat kann als Indiz dafür gewertet werden, daß unter Zuhilfenahme der Resynthese tatsächlich eine feinere Differenzierung der Akzenttöne möglich ist.
Der verbleibende Wert von 16 % (Akzente) bzw. 18 % (Grenzen) Zuordnungen der Etiketten aus unterschiedlichen Kategorien liegt sicherlich in den fehlerbehafteten Transkriptionen, sowohl der Original als auch der durch Resynthese erzeugten begründet. In Konsistenzuntersuchungen zur herkömmlichen prosodischen Etikettierung wurden Übereinstimmungen von maximal 70 - 80 % ermittelt [6]. Dies bedeutet auch für die hier als Referenz angenommene Originaletikettierung eine Unsicherheit bei der Etikettenvergabe von 20 - 30 %, sodaß sich die zunächst hoch anmutende Zahl von Abweichungen in Tabelle 2 relativiert.

5 Bewertung und Ausblick

Die beschriebenen Untersuchungen bestätigen die prinzipielle Durchführbarkeit der Abbildung prosodischer Etiketten auf prototypische F_0-Verläufe. Die im Kapitel *Untersuchungen* dargestellte Auswertung des Vergleichs einer rein manuell und einer mit Hilfe der Resynthese erstellten prosodischen Etikettierung läßt den Schluß zu, daß die durch Resynthese unterstützte Etikettierung mindestens genauso gut wie die manuell erstellte ist. Beide Etikettierungen weisen einen Grad an Übereinstimmung auf (82 - 84 %), wie er auch für, durch unterschiedliche Transkribenten, manuell erstellte Etikettierungen ermittelt wurde [6].
Insbesondere die durch die Werte in Tabelle 2 bestätigte Annahme einer feine-

ren und damit stärker an der Originaläußerung orientierten Kategorisierung der Akzenttöne läßt den Schluß zu, daß sich hierdurch auch eine Erhöhung der Gesamtkonsistenz der prosodischen Etikettierung erzielen läßt.
Die beschriebenen Untersuchungen sind nur als ein erster Test des Verfahrens anzusehen. Durch subjektive Vergleichstests mit mehreren Testpersonen muß die korrekte Wahl der Modellparameter noch evaluiert werden. Gegebenenfalls müssen diese dann angepaßt bzw. das Modell erweitert werden.
Ob die angestrebte Konsistenzerhöhung durch das Verfahren erreicht werden kann, muß letztendlich die Auswertung einer großen Anzahl damit erstellter prosodischer Etikettierungen zeigen.

Literatur

[1] Adriaens, L.M.H.: Ein Modell deutscher Intonation. Dissertation, Technische Universität Eindhoven 1991

[2] Paulus, E; Gerken, H.-D.; Reinecke, J; Veidt, J.: Der Nutzwert prosodischer Merkmale für die automatische Spracherkennung. Proc. Elektronische Sprachsignalverarbeitung, Berlin 1990

[3] Reinecke, J.: Konzept einer Arbeitsstation zur Segmentierung und Etikettierung prosodischer Einheiten. Fortschritte der Akustik, Tagungsband DAGA 93, Frankfurt a.M. 1993, 960-963

[4] Reinecke, J.: Ein System zur Modifikation prosodischer Eigenschaften fließend gesprochener Sprache. Studientexte zur Sprachkommunikation, Heft 11, Proc. Elektronische Sprachsignalverarbeitung, Berlin 1994, 213-220

[5] Reyelt, M; Batliner, A.: Ein Inventar prosdischer Etiketten für VERBMOBIL. Verbmobil Memo 33, Juli 1994

[6] Reyelt, M: Untersuchungen zur Konsistenz prosodischer Etikettierungen. Tagungsband KONVENS 94, Hrsg.: H. Trost, 290 - 299

[7] Wolf, H.E.: Intonationsmotive zur Prosodiesteuerung nach Regeln für die Sprachsynthese. Proc. FASE/DAGA 1982, 1031 - 1038