From G. Görz, *KONVENS 92*, Informatik aktuell, Springer, 1994, pp. 29-38.

Zur prosodischen Kennzeichnung von spontaner und gelesener Sprache

A. Batliner*, B. Johne*, A. Kießling+, E. Nöth+

*: Institut für deutsche Philologie, Ludwig-Maximilians-Universität München
e-mail: anton.batliner@phonetik.uni-muenchen.de

+: Lehrstuhl für Informatik 5 (Mustererkennung), Friedrich-Alexander-Universität Erlangen-Nürnberg, e-mail: kiessl@informatik.uni-erlangen.de

Kurzfassung: Spontane Sprache naiver Sprecher gilt i.a. als weniger regelhaft als gelesene Sprache und wurde im Hinblick auf sprachverstehende Systeme bislang noch kaum untersucht, obwohl ihre Erkennung das letztendliche Ziel solcher Systeme darstellt. In diesem Beitrag werden erste Auswertungen einer deutschen Datenbasis mit identischen spontan gesprochenen und gelesenen Äußerungen vorgestellt. Dabei wurde untersucht, inwieweit sich systematische Unterschiede in der Prosodie feststellen lassen. Den gemessenen prosodischen Kennwerten der Gesamtäußerung (wie $F0$-Offset, $F0$-Mittelwert, $F0$-Range, $F0$-Regression, Gesamtdauer etc.) wurden die Spontaneitätsurteile eines Hörtests gegenübergestellt; mit Hilfe von Diskriminanzanalysen wurde die Relevanz der Kennwerte untersucht. Es zeigt sich, daß sich Spontan- und Lesesprache systematisch in ihrer Prosodie unterscheiden, die Markierung des Unterschieds aber zum Teil eher sprecherspezifisch ist.

Abstract: Spontaneous speech of naive speakers is generally seen as less regular than read speech. Although its recognition is the ultimate aim of speech understanding systems, spontaneous speech has rarely been investigated so far. In this article first analyses on a German database containing identical utterances of spontaneous and read speech are presented. We examined what kind of systematic differences in prosody can be observed. Prosodic features computed for the whole utterance (like $F0$-offset, average $F0$, $F0$-range, $F0$-regression, total duration etc.) were put into contrast with the judgements (concerning spontaneity) obtained from a perception test. Discriminant analyses were used to judge the relevance of the investigated features. The results show that there is a systematic difference in prosody between spontaneous and read speech which however is partially speaker dependent.

1 Einleitung

Sprachverstehende Systeme mit geschriebener Sprache als Eingabe arbeiten normalerweise mit "sauberem" Input, also mit syntaktisch wohlgeformten, ganzen Sätzen in korrekter Orthographie. Pendant dazu ist bei der gesprochenen Sprache (und damit auch bei spracherkennenden Systemen) der von einem kompetenten Sprecher (Phonetiker, Schauspieler, Radiosprecher) gelesene Text, der ebenfalls wohlgeformt ist, d.h. syntaktisch vollständig und phonetisch ausgeformt (Lentostil). Spontane Sprache nicht geschulter (naiver) Sprecher wurde bis jetzt eher selten untersucht, da sie als weniger regelhaft und "gestörter" gilt. Letztlich ist man dabei aber auf Vermutungen angewiesen – es ist genauso möglich und sogar wahrscheinlich, daß auch bei spontaner Sprache Regularitäten existieren, aber – zumindest teilweise – andere als bei gelesener Sprache. Endziel für ein integriertes spracherkennendes und -verstehendes System ist aber auch die Beherrschung der spontanen Sprache – besser gesagt, die Fähigkeit, unterschiedliche Register der Spontansprache erfolgreich als Eingabe verarbeiten zu können. Natürlich hegt man dabei die Hoffnung,

daß man die an gelesener Sprache gewonnenen Ergebnisse auf Spontansprache übertragen kann – sicher zu einem großen Teil zurecht. Diese Hoffnung muß aber konkret überprüft werden. In diesem Beitrag soll nun in einem ersten Schritt untersucht werden, ob sich bei parallelisierter Spontan- und Lesesprache für das Deutsche systematische Unterschiede in der Prosodie ergeben.

2 Material und experimentelles Design

Sprecher waren vier Studenten (3 weibliche: C, X und A, ein männlicher: F), die jeweils paarweise an einer Sitzung teilnahmen; die Sprecher innerhalb eines Paares (C und X bzw. A und F) waren miteinander befreundet. Die zwei Sprecher saßen sich ohne Blickkontakt in einem Versuchsraum des Psychologischen Instituts in München gegenüber und gaben sich gegenseitig Anweisungen, was der Partner mit auf dem Tisch aufgebauten Klötzchen machen sollte. Die Sitzungen dauerten, inkl. Pausen, ca. zwei Stunden. Die Aufgaben waren so angelegt, daß sich kurze Klärungsdialoge mit häufigem Sprecherwechsel ergaben, nicht längere, raisonierende Passagen o.ä. Im Gegensatz etwa zu einem erzählenden Monolog oder zu einem freien Vortrag mit längeren Planungspausen ergab sich dabei eine "echt" spontane, lebhafte Unterhaltung, ohne daß den Versuchspersonen bewußt war, daß ihre Sprache und nicht, wie ihnen gesagt wurde, ihr kooperatives Verhalten untersucht wurde.

Damit bei den später durchgeführten Hörtests (vgl. weiter unten) die Hörer bei der Spontaneitätsbeurteilung nicht der einfachen Strategie *"Mit Häsitation = spontan, ohne Häsitation = nicht spontan"* folgen konnten, wurden aus den Redebeiträgen nur "komplette", satzwertige Äußerungen (Fragen, Imperative und Aussagen) ohne Häsitationen und Abbrüche ausgewählt, deren Signalqualität ausreichend gut war, also z.B. nicht überlagert war von Stuhlrücken, Partnereinschüben etc. Die Äußerungen waren teils satzförmig, teils elliptisch.

Nach ca. 9 Monaten lasen die Sprecher in einer Einzelsitzung die so ausgewählten Äußerungen, und zwar sowohl die eigenen als auch die des jeweiligen Partners. Die Äußerungen waren in einen genügend großen Kontext eingebettet und wurden in schriftlicher Form vorgelegt. Sie wurden allerdings nicht in der orthographisch korrekten, "kanonischen" Form vorgegeben, sondern – ohne Satzzeichen – in gemäßigter Umgangssprache, angepaßt an das spontansprachliche Pendant, wie etwa *"also was ham ma jetz für Steine"* statt *"also was haben wir jetzt für Steine"* oder gar *"also welche Steine haben wir jetzt"*. Damit war gewährleistet, daß die Leseäußerung der Spontanäußerung sehr nahekommt (gleiche Segmente, gleiche Silbenanzahl) und somit direkt mit ihr verglichen werden kann. In diesem Leseregister fehlt also das Umsetzen der kanonischen Schriftsprache, aber es fehlen natürlich nicht die anderen lesetypischen Planungs- und Umsetzungsprozesse. Das folgende Beispiel zeigt eine der Vorlagen für das Leseregister; ausgewählt wurden in diesem Fall die beiden Äußerungen von C:

X: *das steht ungefähr auf Höhe von dem hinteren Fenster*
C: *also praktisch steht des jetz bißchen weiter vorne*
X: *ja genau*
C: *und wieviel Zentimeter daneben*
X: *ja – vier bis fünf – fünf eher*

Ein erster Höreindruck von den Sprechern war folgender: C und X produzierten sehr spontan, auch beim Lesen. Sprecherin A war relativ spontan beim spontanen Produzieren,

hatte aber einen auffällig geänderten Lesestil, so daß manche Hörer meinten, sie hätten im Hörtest zwei unterschiedliche Sprecherinnen bewertet. Sprecher F war weniger spontan und eher zögerlich beim Produzieren, sowohl in der Spontansitzung als auch beim Lesen.

Die Aufnahmebedingungen entsprachen in beiden Sitzungen denen einer ruhigen Büroumgebung. Die Äußerungen wurden mit 12 Bit Auflösung und einer Abtastfrequenz von 10 kHz digitalisiert. Bei den spontanen Produktionen und den gelesenen Äquivalenten zu den eigenen Äußerungen (die gelesenen Partner-Äquivalente bleiben in den hier referierten Auswertungen noch unberücksichtigt) handelt es sich insgesamt um 886 Äußerungen (ca. 18 Minuten Sprachmaterial, Zahl der Äußerungen: C: 362, X: 236, A: 132, F: 156). Von der weiteren signalphonetischen, phonetischen, linguistischen und experimentellen Verarbeitung seien hier nur die Schritte erwähnt, mit denen die im weiteren zugrundegelegten Merkmale extrahiert wurden: Eine berechnete stimmhaft(SH)/stimmlos-Entscheidung wurde manuell anhand des Zeitsignals, gegebenenfalls auditiv unterstützt, korrigiert. Mit Hilfe dreier automatischer $F0$-Verfahren[1] (vgl. [Bat91], [Hes91], [Ree89], [Kie92]) wurden $F0$-Konturen ermittelt, aus denen eine Referenzkontur (je Frame, d.h. alle 12.8 ms ein Hertzwert) erstellt und auditiv unterstützt handkorrigiert wurde. In kritischen Bereichen wurde der $F0$-Wert periodengenau am Zeitsignal ermittelt. Irreguläre Passagen (laryngalisierte Bereiche, vgl. [Bat91]) wurden gehörsadäquat interpoliert, $F0$-Sprünge, die sich als Artefakte der Verfahren herausstellten, geglättet. Die Äußerungen wurden klassifiziert nach Satztypen (Feinkategorien nach dem Satzmodussystem von [Alt87], Fragesätze und Nicht- Fragesätze als Grobkategorien) und nach syntaktischer Vollständigkeit (410 Ellipsen vs. 476 Nicht-Ellipsen). In einem Hörtest beurteilten jeweils 10 Versuchspersonen (Studenten der Phonetik bzw. der Germanistik) die Äußerungen des gesamten Korpus hinsichtlich des Grades der Spontaneität. Dazu wurden jeweils die digitalisierten und segmentierten Äußerungen der beiden Parallelkorpora (spontane und gelesene Daten) eines Sprechers in randomisierter Reihenfolge auf ein Tonband überspielt und den Versuchspersonen dargeboten, die auf Tischvorlagen mitlasen und ihr Urteil auf einer Bewertungsskala mit vier Stufen (1:"sehr spontan" – 2:"spontan" – 3:"wenig spontan" – 4:"nicht spontan") abgaben.

Merkmal	**Beschreibung**
Onset/Offset	$F0$-Wert des ersten/letzten SH Frames
Maximum/Minimum	maximaler/minimaler $F0$-Wert
Range	Betragsdifferenz von Maximum und Minimum
Mittelwert	Mittelwert der $F0$-Werte aller SH Frames
Streuung	Streuung der $F0$-Werte aller SH Frames
Regressionskoeffizient	Koeff. der Reg.geraden durch $F0$-Werte aller SH Frames in Hz/sec
Gesamtdauer	Dauer der Gesamtäußerung in ms
Spontaneität	Mittelwert der Spontaneitätsurteile ($\in [1;4]$)

Tab. 1: Extrahierte und berechnete prosodische Merkmale

3 Fragestellung

In diesem Beitrag werden erste Ergebnisse des Vergleichs von Spontansprache mit Lesesprache vorgestellt. Wir benutzen aus der Mustererkennung bekannte statistische Klassi-

[1] An dieser Stelle sei Herrn Hess, Bonn, und Herrn Reetz, Nimwegen, für die Überlassung ihrer Programme gedankt.

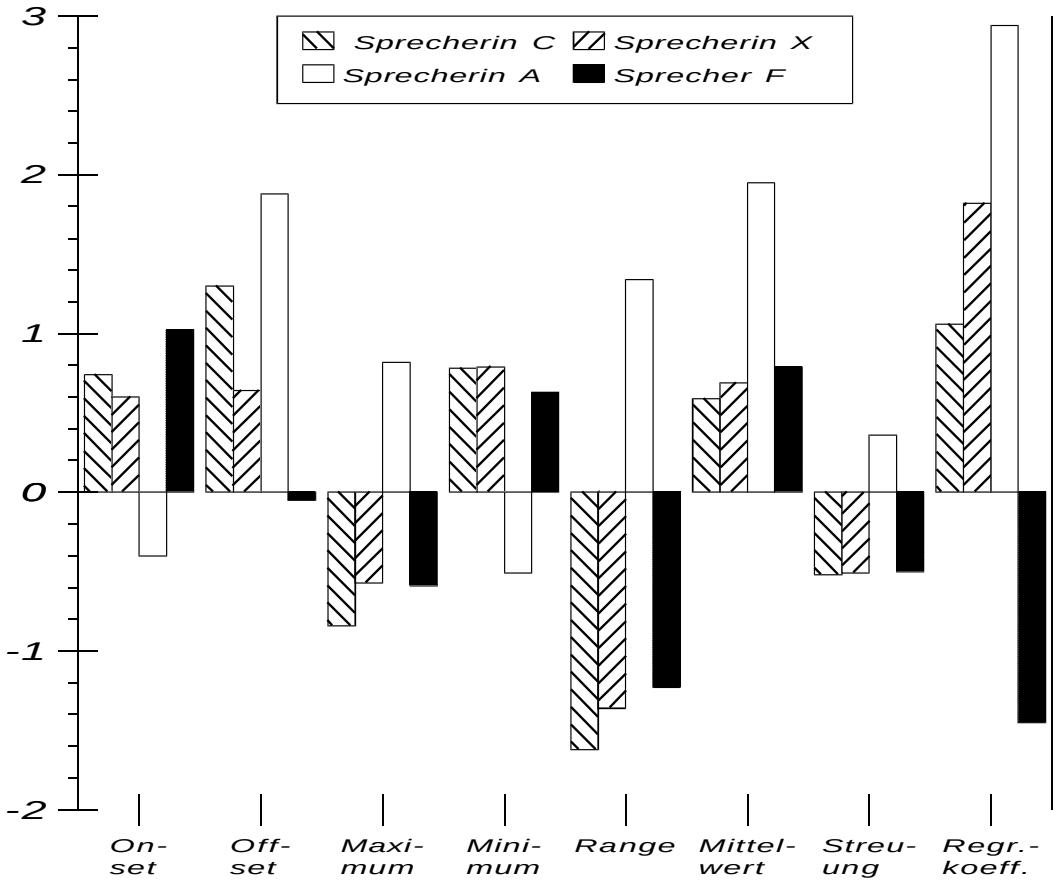

Abb. 1: Mittelwertdifferenzen der $F0$-Merkmale in Halbtönen
(Regressionskoeffizient in Ht/sec)

fikationsverfahren zur Überprüfung unserer Hypothesen, die Erkennungsraten geben aber lediglich Tendenzen wieder. Es wurde noch nicht versucht, die Merkmalsauswahl zu optimieren. Ebenso wurde bei der Klassifikation keine Trennung von Lern- und Teststichprobe vorgenommen. Die fehlende Optimierung dürfte etwas zu niedrige, die Überadaptation bei Lern=Test dagegen etwas zu hohe Erkennungsraten zur Folge haben.

Im folgenden werden nur Kennwerte für die **gesamten** Äußerungen betrachtet, nicht für einzelne Konstituenten. Da beim Lesen exakt dieselben Äußerungen mit der gleichen Silbenzahl vorgegeben wurden, ist auch ein Vergleich z.B. der Gesamtdauerwerte zulässig. Es wurden pro Äußerung die in Tab. 1 aufgeführten Merkmale extrahiert bzw. berechnet. Hertz-Werte wurden in Halbtonwerte, relativ zum Basiswert $1\,Hz$, transformiert[2] und diese wiederum zum Mittelwert normiert, d.h. der Mittelwert der jeweiligen Äußerung wurde subtrahiert. Damit können die unterschiedlichen Tonlagen, insb. die von Männern und von Frauen, angeglichen werden. Es wurde noch nicht versucht, diese Normierung zu optimieren; andere Normierungen, etwa in Hertz zum Mittelwert oder in Hertz bzw. Halbtönen zum Deklinationsverlauf, sind ebenfalls vorstellbar und sollen in einem späteren Schritt auf ihre Güte hin untersucht werden.

[2] $Ht(F0) = 12 \cdot ld(F0), \quad F0 = F0$-Wert in Hz

4 Ergebnisse und Diskussion

Im folgenden steht SPONTAN abkürzend für die nicht-gelesenen *"spontanen Äußerungen"* bzw. *"Spontanregister"*, und ebenso GELESEN für die *"gelesenen Äußerungen"* bzw. *"Leseregister"*. Abb. 1 zeigt pro Sprecher die Mittelwertdifferenzen der Halbtonwerte (GELESEN-Werte subtrahiert von SPONTAN-Werten) und Abb. 2a die der Gesamtdauer; die Werte sind positiv bei einer größeren Ausprägung in SPONTAN und negativ bei einer größeren in GELESEN. Tab. 2 zeigt die den Mittelwertdifferenzen zugrundeliegenden Werte sowie ihre Standardabweichungen. C und X verhalten sich sehr ähnlich: bei SPONTAN sind Onset, Offset, Minimum und Mittelwert höher, Maximum niedriger und der Range (als Folge des Unterschiedes zwischen Minimum und Maximum) kleiner als bei GELESEN. Die Dauer ist kürzer. A und F zeigen untereinander und gegenüber C und X deutliche Unterschiede; Bei A sind Offset und Mittelwert bei SPONTAN deutlich höher als bei den anderen drei Sprechern, der Range und die Streuung sind größer als bei GELESEN – umgekehrt wie bei den anderen drei Sprechern. Bei F unterscheidet sich der Offset von SPONTAN und GELESEN nicht, die SPONTAN-Dauer ist deutlich kürzer als bei den anderen Sprechern. Bei C, X und A ist der Regressionskoeffizient bei SPONTAN höher als bei GELESEN, nur bei F ist es umgekehrt. Man beachte, daß in diese Mittelwertdifferenz des Regressionskoeffizienten sowohl positive Werte (=steigender $F0$-Verlauf) als auch negative Werte (=fallender $F0$-Verlauf) eingehen. Trennt man diese beiden Klassen, so ist bei positiven Werten der Verlauf für C, X und F bei SPONTAN weniger steigend als bei GELESEN, für A ist es umgekehrt. Bei negativen Werten ist der Verlauf für C und X bei SPONTAN weniger fallend als bei GELESEN, für A und F ist es umgekehrt. Es soll später untersucht werden, inwiefern eine Trennung dieser beiden Klassen (oder anderer, z.B. von Fragen vs. Nicht-Fragen) eine bessere Klassifikation von SPONTAN vs. GELESEN ermöglicht.

		Sp C		*Sp X*		*Sp A*		*Sp F*	
Merkmal		MW	StA	MW	StA	MW	StA	MW	StA
Onset	SPONTAN	0.5	2.2	-0.1	2.5	0.0	3.4	0.8	3.0
	GELESEN	-0.2	2.6	-0.7	2.4	0.4	1.6	-0.2	2.5
Offset	SPONTAN	-0.2	3.6	-0.6	4.0	0.7	3.6	-1.5	3.4
	GELESEN	-1.4	3.6	-1.2	4.1	-1.2	3.0	-1.5	4.0
Maximum	SPONTAN	4.3	2.0	4.4	2.2	4.5	2.0	4.5	2.1
	GELESEN	5.1	2.0	5.0	1.9	3.7	1.5	5.1	1.7
Minimum	SPONTAN	-3.3	1.5	-3.7	1.6	-4.0	2.0	-4.1	1.7
	GELESEN	-4.1	1.5	-4.5	1.6	-3.5	1.1	-4.8	1.6
Range	SPONTAN	7.6	3.1	8.1	3.3	8.5	3.5	8.7	3.2
	GELESEN	9.2	2.7	9.5	3.0	7.1	2.0	9.9	2.8
Mittelwert	SPONTAN	96.7	2.1	97.0	1.9	96.9	2.3	81.7	2.3
	GELESEN	96.1	2.0	96.3	1.8	94.9	1.2	80.9	2.3
Streuung	SPONTAN	2.2	1.0	2.2	0.9	2.4	1.7	2.3	0.9
	GELESEN	2.7	0.9	2.7	0.9	2.0	0.6	2.8	0.8
Reg.Koeff.	SPONTAN	-1.0	10.1	0.6	11.0	1.5	8.3	-2.2	9.0
	GELESEN	-2.0	13.2	-1.2	9.4	-1.5	7.0	-0.7	12.0
Gesamtdauer	SPONTAN	1231.5	881.0	1182.9	702.8	984.2	403.8	1105.9	658.5
	GELESEN	1347.8	920.8	1289.4	688.9	1075.8	465.6	1305.4	835.9
Spont.Urteil	SPONTAN	2.2	0.6	2.2	0.5	1.9	0.6	2.3	0.7
	GELESEN	2.3	0.5	2.6	0.5	2.7	0.5	2.9	0.6

Tab. 2: Mittelwerte (*MW*) und Standardabweichungen (*StA*) der prosodischen Merkmale von Abb. 1, Abb. 2a-b und Abb. 3a-e, getrennt nach Registertyp und Sprecher

Auch wenn die Korrelation vieler Merkmale untereinander sehr klein ($< |0.1|$) ist, so sind

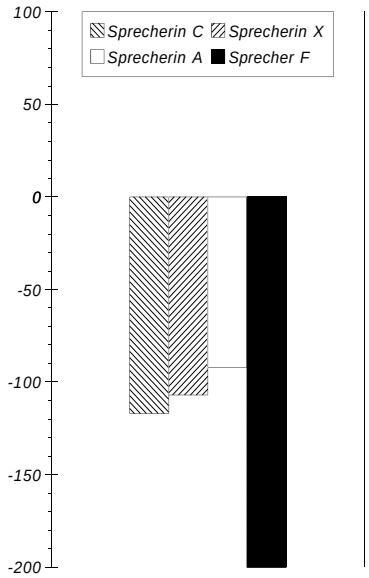

Abb. 2a: Gesamtdauer, Mittelwertdifferenzen in ms

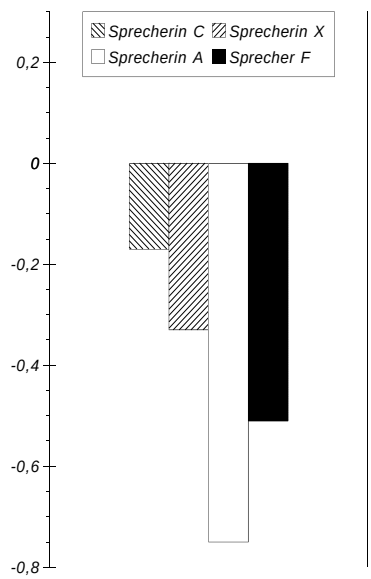

Abb. 2b: Spontaneitätsurteil, Mittelwertdifferenzen

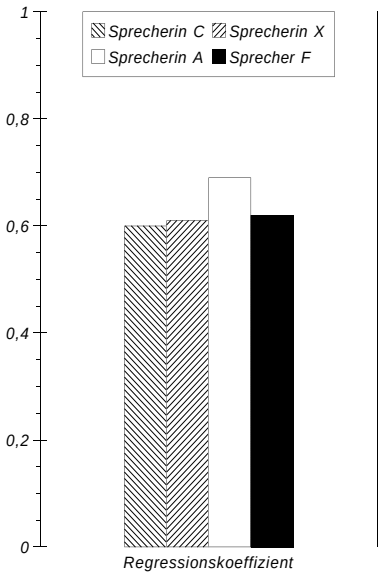

Abb. 2c: Spontaneitätsurteil, alle prosodischen Merkmale unabhängig, Regressionsanalyse

doch nicht alle unabhängig voneinander; diese Tatsache verdeutlicht die Korrelationsmatrix in Tab. 3 (Werte $> |0.4|$ sind dort fett gedruckt). Trivial, da durch die Berechnung vorgegeben, ist die hohe Korrelation von Range mit Maximum und Minimum, vgl. Tab. 1., ebenso die von Streuung mit Maximum, Minimum und Range: Die Streuung wird größer bei einem tieferen Minimum und/oder bei einem höheren Maximum. Ein tiefer Onset und/oder ein hoher Offset korrelieren positiv mit einem positiven/höheren Regressionskoeffizienten, und vice versa. Zur Korrelation von Gesamtdauer mit Spontaneitätsurteil vgl. unten.

	Onset	Offset	Maxi.	Mini.	Range	Mittelw.	Streu.	Reg.Koe.	Dauer
Offset	-0.12	—							
Maximum	-0.00	0.14	—						
Minimum	0.30	0.15	**-0.46**	—					
Range	-0.16	0.01	**0.88**	**-0.81**	—				
Mittelw.	-0.18	0.11	-0.02	-0.00	-0.01	—			
Streuung	-0.10	0.02	**0.78**	**-0.72**	**0.88**	0.04	—		
Reg.Koe.	**-0.45**	**0.49**	-0.00	-0.01	0.01	0.12	0.00	—	
Ges.Dauer	-0.14	-0.19	0.29	-0.36	0.37	0.03	0.13	-0.04	—
Spont.Urt.	-0.03	-0.17	-0.01	-0.07	0.03	-0.25	-0.11	-0.09	**0.44**

Tab. 3: Korrelationsmatrix der prosodischen Merkmale (Werte $> |0.4|$ sind fett gedruckt)

Global läßt sich der Unterschied zwischen den beiden Registern also wie folgt charakterisieren: Bei SPONTAN sind Onset, Offset und Minimum höher sowie das Maximum tiefer als bei GELESEN; damit einher gehen bei SPONTAN ein höherer Regressionskoeffizient, ein höherer Mittelwert, ein geringerer Range (Differenz von Maximum und Minimum) sowie eine geringere Streuung. Die Dauer ist immer kürzer als bei GELESEN.

Abb. 2b zeigt die Mittelwertsdifferenzen der Hörer-Spontaneitätsurteile pro Sprecher. Bei C ist der Unterschied sehr gering, bei A am größten. Diese Unterschiede bestätigen den

oben erwähnten Höreindruck. Hingegen ist bei einer Regressionsanalyse mit allen prosodischen Merkmalen als unabhängige Variablen und der abhängigen Variablen *Spontaneitätsurteil* in Abb. 2c der Unterschied zwischen den Sprechern recht gering. Der Zusammenhang zwischen der Spontaneitätsbewertung und allen prosodischen Merkmalen zusammen ist positiv, aber nicht sehr ausgeprägt (R^2, die sog. *"erklärte Varianz"*, liegt zwischen 0.35 und 0.48).

Abb. 3 zeigt für alle Sprecher zusammen (Abb. 3a) sowie für jeden einzelnen Sprecher (Abb. 3b-e) das Ergebnis von Diskriminanzanalysen mit Lern=Test, mit denen die Äußerungen als SPONTAN bzw. als GELESEN klassifiziert wurden. Aufgetragen ist auf der Y-Achse der prozentuale Anteil korrekt klassifizierter Fälle; der Erwartungswert (Zufallsbereich) ist 50 %, die Y-Achse zeigt zur Verdeutlichung nur den Bereich von 40–80 %, in dem alle Werte liegen. Zunächst sind die Ergebnisse für eine univariate Analyse aufgetragen, mit jeweils einem Merkmal als Prädiktorvariablen. Dann folgt das Ergebnis einer multivariaten Analyse (in Abb. 3a-e durch *Alle* gekennzeichnet), bei der alle Merkmale zusammen die Prädiktorvariablen bildeten. Der letzte, schwarze Balken gibt das Ergebnis für die Analyse mit dem Mittelwert der Spontaneitätsbeurteilung als Prädiktorvariablen.

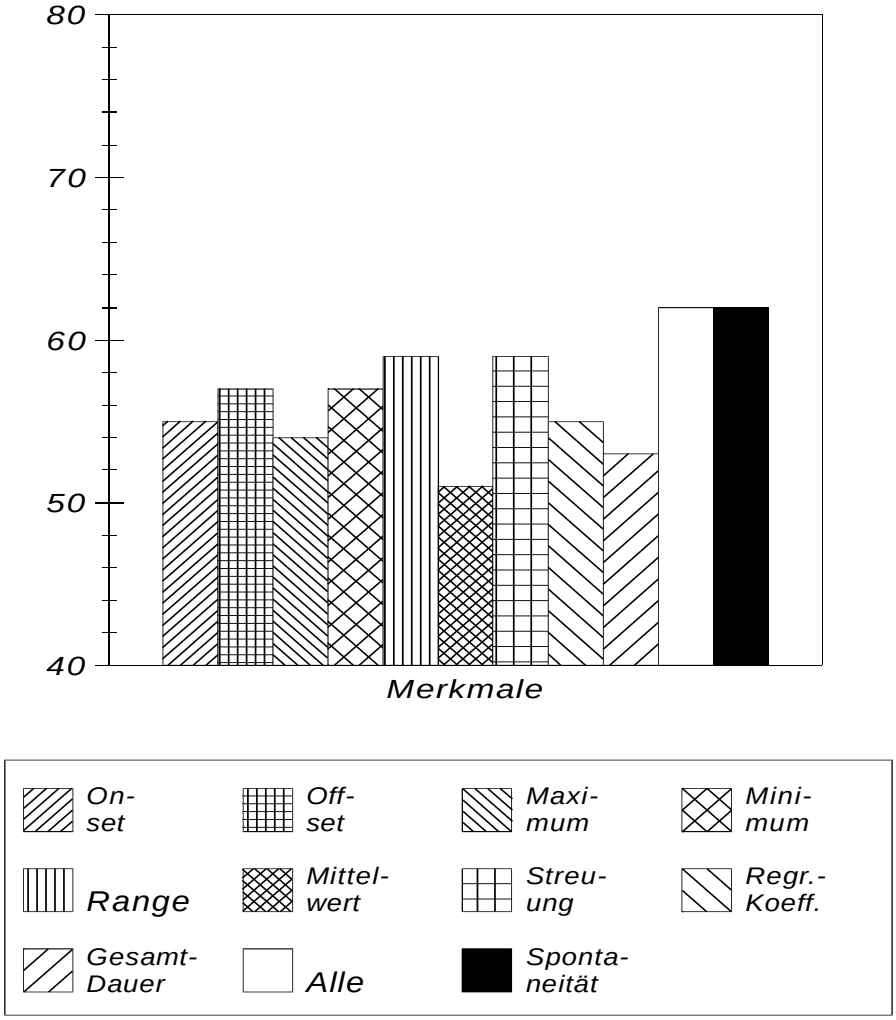

Abb. 3a: Prozent korrekt klassifiziert, alle Sprecher

Betrachtet man alle Sprecher zusammen (Abb. 3a), so ist die Klassifikation mit einzelnen Merkmalen eher im Zufallsbereich, alle Merkmale zusammen ergeben 62 %. Die Klassifikation mit der Spontaneitätsbeurteilung als Prädiktor ist ebenso niedrig (62 %). Ursache

für diese schlechte Klassifikation ist nicht eine fehlende Relevanz der Merkmale, sondern eine starke Sprecherabhängigkeit: Bei getrennt klassifizierten Sprechern sind die einzelnen Merkmale ebenfalls eher wenig relevant, mit Ausnahme von A, wo der Mittelwert allein 74 % ergibt (vgl. Abb. 3b-e). Alle Merkmale zusammen ergeben aber eine deutlich bessere Klassifikation, zwischen 73 % bei C und 76 % bei A. Damit bestätigt sich die Annahme, daß sich Spontansprache und Lesesprache systematisch in ihrer Prosodie unterscheiden; die Markierung des Unterschiedes ist aber zum Teil eher sprecherspezifisch. Mit Ausnahme des Mittelwertes bei A ist offensichtlich das Zusammenspiel der einzelnen Merkmale charakteristischer für das jeweilige Register als einzelne, herausgehobene Merkmale. Prosodische Merkmale allein genügen aber nicht zur Registerkennzeichnung, sonst könnte man ja bei der Klassifikation mit allen Merkmalen ein Ergebnis deutlich über 80 % erwarten. Dies läßt sich anhand der Gesamtdauer illustrieren: als prosodisches, globales Merkmal ist sie für die Klassifizierung irrelevant, auch wenn sie sich bei allen Sprechern systematisch unterscheidet. Sie korreliert auch positiv mit dem Spontaneitätsurteil: je länger die Äußerung ist, desto weniger wird sie als spontan bewertet, vgl. Tab. 3. Die Dauer ist aber zum größten Teil kein prosodisches Merkmal "an sich", sondern Produkt segmentaler Prozesse wie etwa Vokal- bzw. Silbenkürzung und -elision; typischerweise korreliert mit diesen Prozessen ist die Vokalzentralisierung (Änderung im spektralen Bereich). Es ist vorgesehen, anhand der für das Material vorliegenden expliziten Segmentierung zu untersuchen, inwieweit diese Prozesse – allein und zusammen mit den prosodischen Merkmalen – zur Registerkennzeichnung beitragen. Möglicherweise findet sich dann auch eine Erklärung für die Diskrepanz zwischen dem großen Unterschied von A auf der einen und C bzw. X auf der anderen Seite bei der Differenz der Spontaneitätsbeurteilung (-0,75 vs. -0,17 bzw. -0,33; vgl. Abb. 2b) und dem eher geringen Unterschied in der Klassifikation mit allen Merkmalen (76 % vs. 73 % bzw. 75 %; vgl. Abb. 3b,c,d). Man muß allerdings auch beachten, daß die Hörer bei der Spontaneitätsbeurteilung das **Ausmaß** der Spontaneität bewerten und nicht **direkt** nach SPONTAN-GELESEN klassifizieren.

5 Vergleich mit anderen Studien

Die Ergebnisse der wenigen vergleichbaren Studien zum Registerunterschied "SPONTAN-GELESEN" stimmen nur zum Teil mit den unseren überein: [Tro85] z.B. berichtet für das Deutsche bei SPONTAN einen niedrigeren Range (für C, X und F übereinstimmend) sowie niedrigere Minima (für C, X und F nicht übereinstimmend). [Bla91] findet für das Niederländische bei SPONTAN einen niedrigeren Mittelwert (nicht übereinstimmend) sowie niedrigere Streuung und niedrigeren Range (für C, X und F übereinstimmend). Ebenfalls für das Niederländische berichtet [Bei90] einen niedrigeren $F0$-Median und keine Dauerunterschiede (für C, X und F nicht übereinstimmend) sowie einen niedrigeren Range (für C, X und F übereinstimmend). Beim jetztigen Kenntnisstand läßt es sich nicht entscheiden, ob diese Gemeinsamkeiten bzw. Unterschiede sprach-, sprecher-, design- oder registerspezifisch sind, da – aus verständlichen Gründen – die Zahl der Sprecher immer recht gering ist, und da sich die Studien auch sonst unterscheiden; so gab etwa [Tro85] Akzentverteilung und finalen Tonverlauf bei GELESEN via Instruktion vor, das Material von [Bei90] ist aus einem erzählenden Monolog, nicht aus einem Dialog, etc. Ein Vergleich mit anderen Studien, die andere Merkmale zugrundelegen, wie etwa [How91], wo die Pausensetzung und Position der Hauptakzente untersucht werden, oder [Bru91] (Vergleich von Ton-Sequenzen), ist erst zu einem späteren Zeitpunkt möglich.

6 Schlußbemerkungen

Registerunterscheidung ist an sich natürlich ein interessantes Thema; anwendungs- und kommunikationsrelevant wird sie letztlich aber erst dann, wenn man untersucht, ob in den unterschiedlichen Registern die Prosodie auch unterschiedlich eingesetzt wird. Erste Ergebnisse zur Frage/Nicht-Frageunterscheidung bei elliptischen vs. nicht-elliptischen sowie spontanen vs. gelesenen Äußerungen deuten auf einen gewissen Unterschied hin. Wir nehmen an, daß die funktionale Belastung und damit die Ausprägung der prosodischen Merkmale bei elliptischen Äußerungen größer ist als bei vollständigen. Ebenfalls sollte bei gelesener Sprache wegen des fehlenden situativen Kontextes mit seinen Disambiguierungsmöglichkeiten die Ausprägung der prosodischen Merkmale größer sein als bei spontaner Sprache. Diese Annahmen sollen in einem nächsten Schritt überprüft werden.

Die diesem Bericht zugrundeliegenden Untersuchungen wurden mit Mitteln der Deutschen Forschungsgemeinschaft (Al 173/4) sowie des Bundesministers für Forschung und Technologie unter den Förderkennzeichen 01IV102F4 und 01IV102H0 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren. Für wichtige Hinweise danken wir einem der Gutachter der KONVENS-92.

Literatur

[Alt87] H. Altmann: *Zur Problematik der Konstitution von Satzmodi als Formtypen*, in J. Meibauer (Hrsg.): *Satzmodus zwischen Grammatik und Pragmatik*, Niemeyer Verlag, Tübingen, 1987, S. 22–56.

[Bat91] A. Batliner, A. Kießling, R. Kompe, E. Nöth: *"Irregularitäten" spontaner Sprache und ihre Verarbeitung mit automatischen Grundfrequenzverfahren*, in *Fortschritte der Akustik–DAGA'91, Teil B*, DPG–GmbH, Bad Honnef, 1991, S. 993–996.

[Bei90] F. K. van Beinum: *Spectro-temporal reduction and expansion in spontaneous speech and read text: The role of focus words*, in *Int. Conf. on Spoken Language Processing*, Kobe, 1990, S. 1.6.1–1.6.4.

[Bla91] E. Blaauw: *Phonetic characteristics of spontaneous and read-aloud speech*, in *Proceedings of the ESCA Workshop. Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication*, Barcelona, 1991, S. 12–1 – 12–5.

[Bru91] G. Bruce, P. Touati: *On the analysis of prosody in spontaneous speech with exemplification from Swedish and French*, in *Proceedings of the ESCA Workshop. Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication*, Barcelona, 1991, S. 13–1 – 13–5.

[Hes91] W. Hess: *Persönliche Mitteilung*, 1991, Institut für Kommunikationsforschung und Phonetik, Universität Bonn.

[How91] P. Howell, K. Kadi-Hanifi: *Comparison of prosodic properties between read and spontaneous speech material, Speech Communication*, 10 1991, S. 163–169.

[Kie92] A. Kießling, R. Kompe, H. Niemann, E. Nöth, A. Batliner: *DP-Based Determination of F0 Contours From Speech Signals*, in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Bd. 2, San Francisco, 1992, S. II–17–II–20.

[Ree89] H. Reetz: *A Fast Expert Program for Pitch Extraction*, in *Proc. European Conf. on Speech Communication and Technology*, Bd. 2, Paris, 1989, S. 476–479.

[Tro85] H. Tropf: *Zur Intonation spontan gesprochener und laut gelesener W-Fragen*, in W. Kürschner, R. Vogt (Hrsg.): *Grammatik, Semantik, Textlinguistik. Akten des 19. Linguistischen Kolloquiums Vechta 1984, Band I*, Niemeyer Verlag, Tübingen, 1985, S. 49–60.

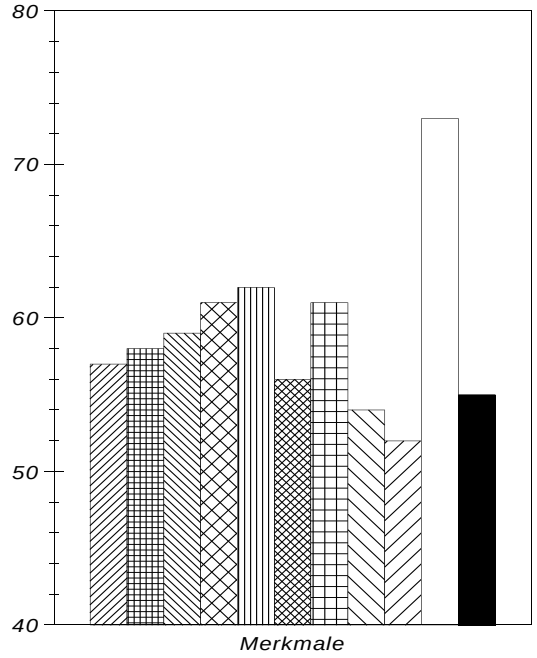

Abb. 3b: Prozent korrekt klassifiziert, Sprecherin C

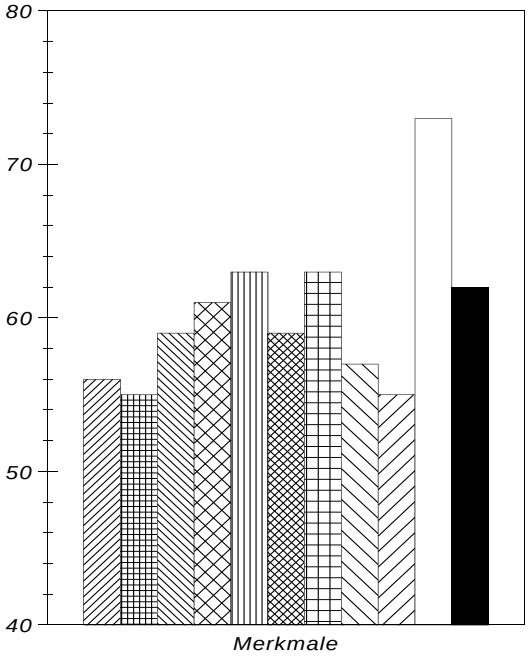

Abb. 3c: Prozent korrekt klassifiziert, Sprecherin X

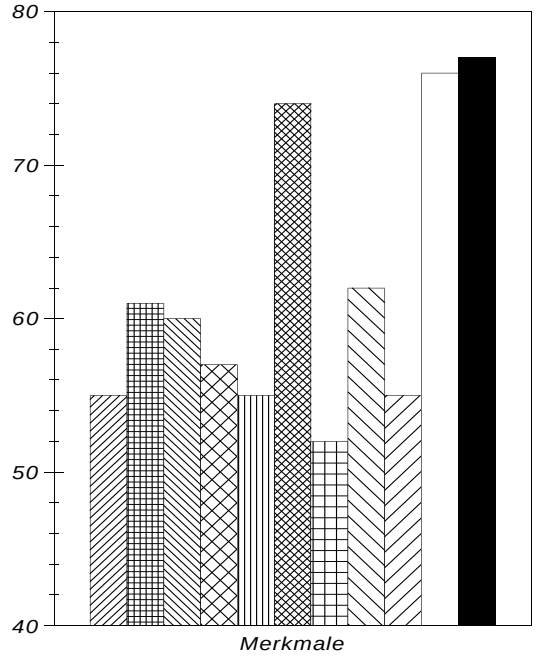

Abb. 3d: Prozent korrekt klassifiziert, Sprecherin A

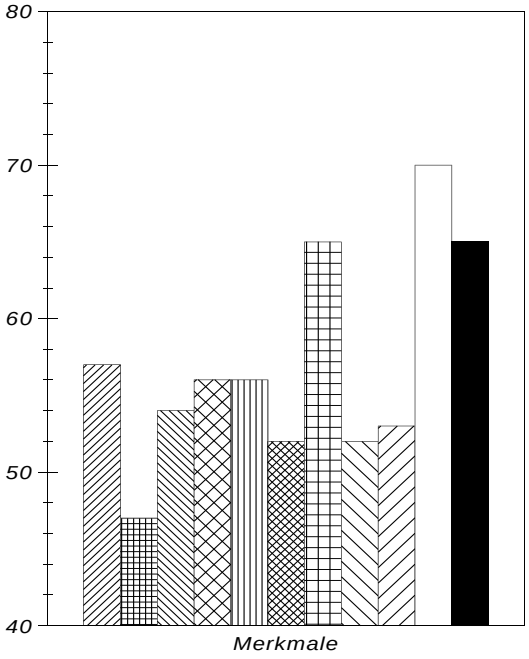

Abb. 3e: Prozent korrekt klassifiziert, Sprecher F