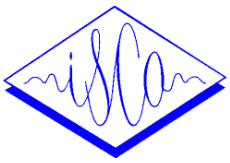

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

EVALUATING RADIO NEWS INTONATION AUTOSEGMENTAL VERSUS SUPERPOSITIONAL MODELLING

Maria Wolters

Universität Bonn
Poppelsdorfer Allee 47
D-53115 Bonn
wolters@ikp.uni-bonn.de

Hansjörg Mixdorff

TU Dresden
Mommsenstr. 13
D- 01062 Dresden
mixdorff@tfh-berlin.de

ABSTRACT

This study examines prosodic correlates of the givenness of discourse entities in German radio news speech. The material comes from the Stuttgart Radio News Corpus. Both GToBI intonation labels and a Fujisaki-style parametrization of the intonation contour were examined. We find strong word-class specific accentuation defaults; the influence of entity status is rather small and varies with word class. However, there are strong influences of newness on phrasing. The results of autosegmental and superpositional approaches complement each other nicely.

1 INTRODUCTION

In this study, we examine prosodic correlates of entity status in German radio news with respect to two intonation models, autosegmental–metrical and superpositional. The paper is structured as follows: In Section 2, we introduce the concept of entity status and briefly review the two intonation models on which our results are based. Next, in Section 3, we describe the corpus and the annotations used in this study. The results presented in Section 4 largely confirm those of [18]. We conclude in Section 5 that, at least in radio news data, which is widely used for speech synthesis, entity status is marked by phrasing rather than by accentuation. What radio news speakers do mark quite consistently is the overall structure of the discourse. We also find that superpositional models are viable alternatives to autosegmental approaches not only for speech synthesis, but also for linguistically motivated intonation research.

2 BACKGROUND

In order to synthesize the F0 contour of an utterance, we need a model of the F0 contour that both describes a complex contour by a small number of parameters and allows to generate all linguistically relevant F0 movements with the appropriate size and alignment. One possible solution is superpositional modelling. A well-known model of this type is the Fujisaki model [3], which has been adapted to German by e.g. Mixdorff [7]. The Fujisaki model describes an F0 contour in the log F0 domain by superimposing two basic components: the phrase component, which covers long-term changes in the pitch contour that are associated with intonational phrases, and accent components, which cover the short-term changes associated with pitch accents. These components are added to a speaker-specific base F0. New phrases and accents are triggered by phrase and accent commands, respectively, and the size and timing of these commands are the most important parameters of the model.

Unlike the Fujisaki model, the autosegmental–metrical approach to intonation does not use gradient information. Instead, F0 contours are modelled as sequences of abstract high tones (H) and low tones (L). Simple pitch accents consist of only one tones, complex accents consist of a sequence of tones. In each accent, the tone which is linked to the accented syllables is marked with a star (*). Phrase boundaries are signalled by phrase tones and boundary tones. A system of break indices is used to code prosodic boundary strength, ranging from 0 (clitic) to 4 (major intonational phrase). Recently, there has been much interest in the meaning of certain tone contours. A classic example of such research is [11].

In this paper, we focus on one potential function of intonation: signalling whether a referring expression specifies a “given” or a “new” discourse entity. Discourse entities are “conceptual coathooks” (Woods, cited after [17]) for the information that a hearer gets from a speaker during discourse [15, 17]. They form the basis on which hearers can construct a model of ongoing discourse. The status of an entity contains information about the role that the entity plays in the discourse, and about ways of accessing that entity [19]. If the entity still needs to be constructed, its status tells the hearer how he can build an initial description of it. The detailed information in the entity status variable can be summarized in several different ways. In this paper, we discuss four possible taxonomies (c.f. Table 1): discourse old/new (DISC, [12]), hearer old/new (HEAR, [12]), new/mediated/old (STAT3, [16]) and active/accessible/unused/new (STAT4, derived from [5]).

3 CORPUS AND METHOD

The corpus examined is a subset of the Stuttgart Radio News Corpus [13]. The subcorpus contains German radio news (23 minutes, 3285 words, 938 referring expressions, 2116 words in referring expressions) read by a single male speaker on two separate days. The data comes from the Deutschlandfunk, a nation-wide highly regarded radio station which focusses on news, reports, and art. All figures given in this paper refer to this subcorpus. The corpus has been annotated with the Stuttgart version of GToBI [6]. A word is accented if it carries at least one pitch accent. Under this criterion, 50.3% of all words are accented. We do not regard words with an L* as unaccented, because although there may be no large pitch excursion, words with L* accents are still perceived to be stressed. The most frequent accent labels in the corpus are H*L, a fall, and L*H, a rise (c.f. Table 2). For this corpus, we computed the Fujisaki model parameters of the intonation contour automatically using a novel, fully automatic approach [8]. We will mainly be concerned with accent command amplitude (Aa) and phrase command magnitude (Ap) here. A word is accented if Aa > 0 on one of its syllables. A companion paper [9] compares the results of the Fujisaki modelling to the ToBI labels as produced by human labellers. Figure 1 shows the fitted F0 contour together with the original ToBI labels, the phrase commands, and the accent commands.

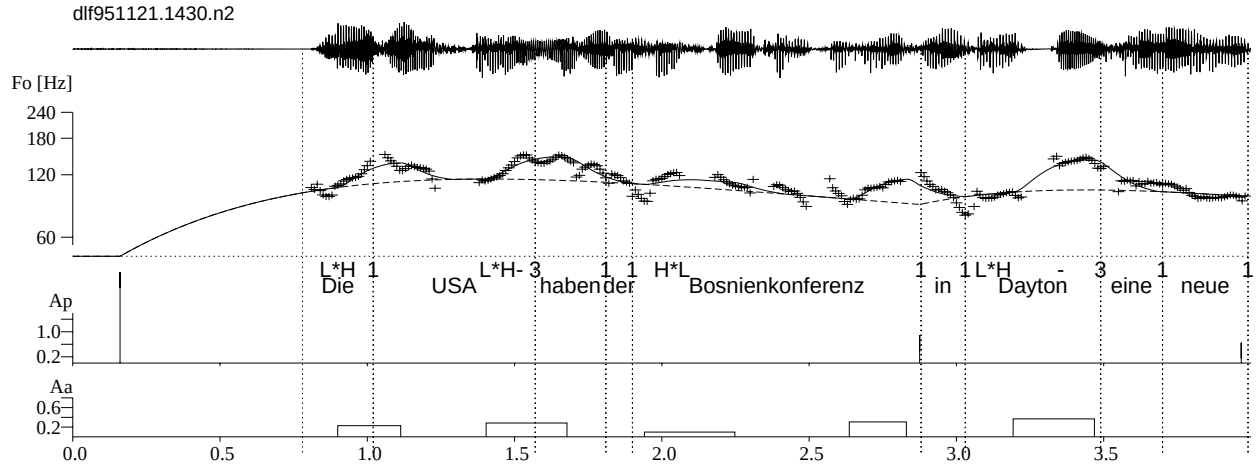

Figure 1: Sample parametrization of F0. From top to bottom: speech waveform, extracted (+-signs) and estimated F0 contours, original ToBI labels, phrase commands, and accent commands.

DISC	HEAR	STAT3	STAT4	Description	Sample
old	old	old	active	previously mentioned in the discourse	ref.exp. 3: refers to ICSLP
new	old	old	unused	known to hearer, but not previously mentioned in the discourse	ref.exp. 1: the readers of these proceedings know this conference
new	new	med	accessible	link to previously mentioned discourse entity	ref.exp. 4: refers to the speakers of ICSLP 2000 (ref.exp. 1)
new	new	new	new	unknown to hearer, no link to any previously mentioned discourse entity	ref.exp. 5: refers to existing person unknown to most readers

Sample discourse: [The ICSLP]$_1$ takes place in [Beijing]$_2$. [It]$_3$ has [many speakers]$_4$. It is not yet clear whether [Ken Mixdorff]$_5$ will also attend.

Table 1: Taxonomies of entity status. Examples refer to the sample discourse given below the table.

Accent	L*H	H*L	L*H%	H*	H*L%	L*HL
Freq.	28.4	15.8	6.8	6.2	3.9	3.0

Table 2: Frequency of all contours that occur on more than 30 words in our subcorpus; percentages relative to all accented words.

In the texts, all referring expressions were labelled manually with syntactic and semantic information as well as with aspects of entity status by a single labeller, the first author. The labels were repeatedly checked for consistency. The four taxonomies described in Table 1 were derived automatically from a much richer taxonomy. Part-of-speech (POS) labels were taken from the original corpus. The POS tagset is STTS, the standard tagset for German corpora [14]. The guidelines for coding referring expressions as well as a more detailed coding manual can be found in [19]. The categories for syntactic functions correspond to classical terminology, which makes the annotations relatively theory-independent. We will only consider the categories subject, object, prepositional object, genitive adjunct, and prepositional adjunct here.

4 RESULTS

Our preliminary hypothesis is that referring expressions which specify new discourse entities have to be accented, while those that specify old entities need not be accented [1, 10]. Table 3 shows strong accentability defaults at work: nouns are accented, pronouns, which almost always specify discourse–old entities, are not. There are significantly more pronouns with accent commands than with ToBI accents, but the median amplitude of these accent commands is rather low (0.15, 0.18) compared to that for nouns. Closer analysis of these instances reveals that the Fujisaki model detects a considerable number of weaker accents which slipped the attention of the human labeller. Proper names are even more likely to be accented than common nouns. This suggests that the influence of entity status on the accentability of the two word classes should be analysed separately for each class.

The next question is: to what degree does entity status influence the accentuation of proper names and nouns? For American English, Cahn [2] implemented an algorithm based on [11] where “given” information is marked with an accent with a starred low tone, and “new” information by an accent with a starred high tone. Thus, nouns which specify given discourse entities need not be unaccented, but they should show a marked preference for L*-type accents.

	total	ToBI	Aa	
NN	(common noun)	823	78.8	*85.7 (0.25)*
NE	(proper name)	181	87.7	89.8 (0.29)
PPER	(pers. pronoun)	20	5.0	*15.0 (0.15)*
PPOSAT	(attributive PPER)	34	2.9	*14.7 (0.18)*

Table 3: Accentability of nouns and pronouns in referring expressions. ToBI: % with ToBI accent, Aa: % with accent command, median command amplitude. *italics:* difference of more than 10%

	DISC		STAT4			
	new	old	active	acc.	new	unused
Ap > 0	43.1	25.0	25.0	35.7	46.7	45.6
Ap (mean)	0.41	0.25	0.25	0.35	0.42	0.43
BI > 2	54.1	34.1	34.1	46.5	55.3	59.4

Table 4: Frequency of phrase commands / break indices at the end of NPs. Ap: phrase command amplitude, BI: ToBI break index

For German, Kohler suggests in [4] that early F0 peaks signal established facts, middle peaks new information, and late peaks emphasis and contrast. An early peak roughly corresponds to LH*L, and a mid-to-late peak to L*H in the GToBI system. Therefore, L*H might signal new information. In order to test these hypotheses, we analysed not only the presence of accents, but also the type of the intonation contour on the word. As we observed over 80 varieties of tone contours on words in our subcorpus, including phrase- and boundary tones, we restricted ourselves to the two most frequent and basic ones, L*H and H*L. L*H and H*L contours are more likely to occur on words without secondary lexical accents, which introduces a certain lexical bias. An analysis of more complex contours and phrase–level contours is subject of future work.

The data in Table 6 tell a somewhat more complex story than the literature would lead us to expect. A series of Fisher tests for the four taxonomies DISC, HEAR, STAT3, and STAT4 reveal only 6 significant associations ($p<0.05$). In general, nouns in referring expressions that specify discourse-old are significantly more likely to carry a L*H contour. For common nouns, there is a significant influence of STAT3 and STAT4 on the presence of both ToBI accents and accent commands. This is due to a tendency to accent common nouns when the discourse entity is both discourse and hearer new. For proper names, accent command amplitude correlates with both DISC and STAT4 (Kruskal-Wallis test). Discourse-new expressions are made more prominent than discourse-old ones. For proper names, we also find a tendency to accent hearer-old names, not hearer-new ones, although that tendency is not significant because of the small number of hearer-old proper names in the corpus. Closer inspection of the data reveals that most of these accents are due to hearer–old, discourse–new entities.

There are also strong influences of syntactic function. For example, nouns in genitive adjuncts are less likely to be accented (64.6% for nouns, 55.0% for proper names). These adjuncts are frequently used to link new discourse entities to either the hearer's world knowledge or to a discourse–old entity. Therefore, they mostly refer to entities that are either discourse or hearer old. There is also a close syntactic link between a preposed genitive adjunct and its head, which leads to a deaccentuation of the adjunct [7]. In the Fujisaki model, this effect is not reflected in the frequency with which these words carry accent commands, but in the mean amplitude of these commands.

Phrasing is closely related to text structure. The phrase command at the beginning of a news story exhibits a median magnitude Ap of 2.15. All sentence boundaries (median magnitude: 1.43) and 74.8% of all commas (median magnitude: 0.73) are associated with a phrase command. 76.4% of all phrase commands appear at the end of a referring expression. Boundaries are more likely to occur after noun phrases that belong to first mentions, which tend to consist of more words and carry more modifiers [19] ($p<0.001$, taxonomies DISC, STAT4, Fisher test). For more details, see Table 4. Interestingly, these phrasing results are much more stable than the accentuation results reported in the previous paragraph. This suggests that the news reader used phrasing much more consistently for signalling (linguistic) structure than accentuation. Given the complexity of the radio news texts in which most sentences contain highly complex long NPs and semantically empty verbs, this strategy makes perfect sense [19].

5 DISCUSSION

We have seen that the prosodic correlates of entity status are not as straightforward as theory would lead us to expect. The reason for this is clear: Since prosody has a heavy functional load, intonation contours are highly polysemous. One and the same contour may signal thematicity as well as, for instance, non–finality. We clearly need to examine more closely how entity status interacts with other factors, such as syntactic function. We expect that again, phrasing will be the main intonational correlate of entity status. The main reason for the differences between our results and results from controlled experiments is that radio news texts are far more complex than experimentally elicited sentences. Another area of future work are words and phrases with multiple accents. Collecting the tones on a word or a phrase into a contour is not as straightforward as it seems, especially if the results are to be amenable to a statistical analysis. We are also interested in ways of summarizing the accent commands on a word or phrase into a set of variables which characterize the accentuation pattern of that sentence. In general, phrasing was used more consistently than accentuation.

Our findings suggest that results of gradient and categorial research into functions of intonation are complementary. What surfaces on one level as the presence versus absence of categories can surface on the other level as a lower amplitude of e.g. accent commands. Therefore, studies which compare gradient acoustic measures with phonological categories should not be discarded outright. The question remains to what degree the gradient effects from which one tends to abstract by phonological approaches are important for generating natural F0–contours.

Acknowledgements: We thank IMS Stuttgart for kindly supplying the Stuttgart Radio News corpus. Hansjörg Mixdorff was funded by the Deutsche Forschungsgemeinschaft, grant No. MI 625/4-1.

Model		POS	Syntactic Function									
			subj.		(in)dir. obj.		prep. obj.		gen. adj.		prep. adj.	
ToBI	% accented	common noun	77.1		80.4		82.5		64.6		83.9	
		proper name	95.2		100.0		100.0		55.0		88.5	
	% H*L / % L*H	common noun	13.0	28.5	10.2	25.4	15.9	19.0	8.1	13.1	17.4	20.1
		proper name	11.3	30.6	0.0	33.3	14.3	14.3	0.0	15.0	10.1	16.9
Fujisaki	% Aa >0 / median Aa	common noun	84.9	0.24	86.0	0.25	87.3	0.19	81.8	0.22	87.8	0.22
		proper name	93.6	0.29	100.0	0.41	85.7	0.34	85.0	0.19	87.4	0.24

Table 5: Prosodic marking of nouns in referring expressions – influence of syntactic functions

	common noun								**proper name**		
	DISC	HEAR	STAT3 / STAT4						DISC	HEAR	
	old / new	old / new	new	med	old	active	unused	total	old / new	old / new	total
total	729 / 94	250 / 573	187	386	250	94	156	823	27 / 154	136 / 45	181
ToBI	78.7 / 78.7	77.6 / 79.2	*82.6*	*72.2*	*74.6*	*78.7*	*76.9*	*78.7*	88.9 / 87.7	90.4 / 80.0	87.8
H*L	12.8 / 13.7	15.6 / 18.7	16.1	10.2	12.4	12.8	12.2	12.6	11.1 / 9.1	16.9 / 8.9	10.7
L*H	*35.1 / 21.3*	31.6 / 27.4	21.8	20.3	26.4	35.1	21.2	22.8	*37.0 / 18.8*	21.3 / 22.2	24.5
Aa >0	81.9 / 86.4	85.6 / 86.0	*88.6*	*80.7*	*85.6*	*81.9*	*87.8*	*85.7*	85.2 / 90.3	91.2 / 84.4	89.5
median Aa	0.20 / 0.24	0.22 / 0.24	0.24	0.23	0.22	*0.20*	0.23	0.25	*0.19 / 0.30*	0.25 / 0.25	0.25

Table 6: Prosodic marking of entity status. In STAT4, the category "old" of STAT3 is split into active and unused; "med" corresponds to "accessible", and "new" to "new". italics: significant at the $p<0.05$ level

6 REFERENCES

[1] G. Brown. Prosodic structure and the Given/New distinction. In A. Cutler and D. R. Ladd, editors, *Prosody: Models and Measurements*, Berlin etc., 1983. Springer.

[2] J. Cahn. *A Computational Memory and Processing Model for Prosody*. PhD thesis, MIT Media Lab, 1998.

[3] H. Fujisaki and K. Hirose. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan (E)*, 5(4):233—241, 1984.

[4] K. Kohler. Terminal accent patterns in single accent utterances of German: Phonetics, phonology, and semantics. in K. Kohler, editor, *Studies on German Intonation* AIPUK 25, Kiel, pages 115–186, 1991.

[5] K. Lambrecht. *Information Structure and Sentence Form*. Cambridge University Press, Cambridge, 1994.

[6] J. Mayer. *Intonation und Bedeutung*. PhD thesis, Lehrstuhl für Experimentelle Phonetik am Institut für maschinelle Sprachverarbeitung, Stuttgart, 1997.

[7] H. Mixdorff. *Intonation Patterns of German - Quantitative Analysis and Synthesis of F_0 Contours*. PhD thesis, TU Dresden, 1998. WWW: http://www.tfh-berlin.de/∼mixdorff/thesis.htm

[8] H. Mixdorff. A novel approach to the fully automatic extraction of fujisaki model parameters. In *Proceedings ICASSP 2000, vol. 3*, pages 1281—1284, Istanbul, Turkey, 2000.

[9] H. Mixdorff and H. Fujisaki. A Quantitative Description of German Prosody Offering Symbolic Labels as a By-Product In *Proceedings ICSLP 2000*, Beijing, China.

[10] S. Nooteboom and J. Kruyt. Accents, focus distribution, and the perceived distribution of given and new information: an experiment. *JASA*, 82:1512–1524, 1987.

[11] J. Pierrehumbert and J. Hirschberg. The meaning of intonation contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M.E. Pollack, editors, *Intentions in Communication*, pages 271–311, Cambridge, MA, 1990. MIT Press.

[12] E. F. Prince. The ZPG letter: Subjects, definiteness, and information-status. In W.C. Mann and S.A. Thompson, editors, *Discourse Description. Diverse Linguistic Analyses of a Fund-Raising Text*, pages 295–325. John Benjamins, Amsterdam, 1992.

[13] S. Rapp. *Automatisierte Erstellung von Korpora für die Prosodieforschung*. PhD thesis, Lehrstuhl für Experimentelle Phonetik am Institut für maschinelle Sprachverarbeitung, Stuttgart, 1998.

[14] A. Schiller, S. Teufel, and C. Thielen. Guidelines für das Tagging deutscher Textcorpora mit STTS. Technical report, IMS Stuttgart/Seminar f. Sprachwiss. Tübingen, 1995.

[15] C. L. Sidner. Focusing in the comprehension of definite anaphora. In M. Brady and R. C. Berwick, editors, *Computational Models of Discourse*, pages 267–330. MIT Press, Cambridge, MA, 1983.

[16] M. Strube. Never look back: An alternative to centering. In *Proceedings of 17th COLING / 36th ACL*, Montréal, Québec, Canada, 10–14 August 1998, volume 2, pages 1251–1257, 1998.

[17] B.L. Webber. So what can we talk about now? In M. Brady and R. C. Berwick, editors, *Computational Models of Discourse*, pages 331–371. MIT Press, Cambridge, MA, 1983.

[18] M. Wolters. Prosodic correlates of referent status. In *Proceedings XVIth ICPHS*, San Francisco, CA, 1999.

[19] M. Wolters. *Towards Entity Status*. PhD thesis, Institut für Kommunikationsforschung und Phonetik, Universität Bonn, 2000.