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 on German radio news data. The data comes from the Stuttgart Radio News corpus (SRN, [18]). The F0 contours in the corpus have been prosodically labelled using the Stuttgart version of GToBI [10], and a superpositional Fujisaki-style description of the contour has been generated [12, 13]. This allows us to examine prosodic correlates of entity status with respect to two intonation models: the autosegmental–metrical and the 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 [26]. We conclude in Section 5 that, at least in the radio news data which is such a popular basis for speech synthesis research, entity status is not marked prosodically. What radio news speakers do mark quite consistently is the overall structure of the discourse. In general, we find that Fujisaki–style superpositional models are viable alternatives to autosegmental approaches for speech synthesis.

2 BACKGROUND

In order to synthesize the F0 contour of an utterance, we need an model of the F0 contour that both describes a complex contour by a small number of parameters and allows to generate all linguistically relevant pitch movements with the appropriate size and alignment. One possible solution is superpositional modelling. In this approach, F0 contours are described by a set of equations that correspond to long-term and short-term changes in the contour. A well-known superpositional model is the Fujisaki model (see e.g. [5]), which has been adapted to German by Möbius [14] and Mixdorff [11]. Much recent work on functions of intonation is based on the autosegmental-metrical framework [8, 16].

In this paper, we examine prosodic correlates of entity status with respect to both accentuation and phrasing. The concept of givenness is problematic. An entity can be "given" to the hearer by world knowledge, the communication situation, or the previous discourse. An entity can be "new" just because it has not been mentioned in the discourse or it can be "new" to both the discourse and the hearer. To counter these problems, the concept of entity status [27] was developed. Discourse entities are "conceptual coathooks" (Woods, cited after [25]) for the information that a hearer gets from a speaker during discourse [25, 21]. The status of an entity contains inter alia information about how the hearer can access that entity, or, if the entity still needs ot be constructed, how the hearer can build an initial description. In this paper, we discuss four straightforward taxonomies of entity status: discourse old/new (DISC, [17]), hearer old/new (HEAR, [17]), new/mediated/old (STAT3, [24]) and active/accessible/unused/brand-new (STAT4, derived from [9]). These taxonomies are summarised in Table 1.

Entity status also has strong linguistic correlates which need to be controlled for. For example, given entities tend to be pronominalized. Work on British English [1], American English [19], and German [23] shows that content words are far more likely to be accented than function words. Nouns are more likely to be accented than adjectives, and verbs more likely to be accented than most function words, such as pronouns. Therefore, we will analyze the effect of entity status on accentuation separately for nouns and pronouns. Other potential factors, which we will not consider in detail here, are syntactic function and sortal class.

3 DATA

Our corpus is a subset of the Stuttgart Radio News Corpus [18]. The subcorpus contains German radio news (23 minutes, 3285 words, 938 referring expressions) read by a single male speaker on two separate days. All numbers given in this paper refer to this subcorpus. The corpus has been annotated with the Stuttgart version of GToBI [10], which

DISC	HEAR	STAT3	STAT4	Description
old	old	old	active	previously mentioned in the discourse
new	old	old	unused	known to hearer, but not previously mentioned in the discourse
new	new	med	accessible	link to previously mentioned discourse entity
new	new	new	new	unknown to hearer, no link to any previously mentioned discourse entity

Table 1: Taxonomies of entity status.

is based on [4]. In ToBI [22], pitch accents are modelled as sequences of high tones (H) and low tones (L). The main tone gets 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). The most frequent accents in the corpus are H*L, a fall, and L*H, a rise. A word is accented if it carries at least one pitch accent.

For this corpus, we computed the Fujisaki model parameters of the intonation contour automatically [12] on the basis of hand-corrected pitch marks. A Fujisaki model consists of two types of equations: phrase equations, which cover long-term changes in the pitch contour that are associated with intonational phrases, and accent equations, which cover the short-term changes associated with pitch accents. 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 resulting model. We will mainly be concerned with accent command amplitude (AA) here. A word is accented iff AA>0 on one of its syllables. A companion paper [13] compares the results of the Fujisaki-style modelling to the ToBI labels as produced by human labellers. This paper also shows examples of fitted F0 contours. In the texts, all referring expressions were labelled manually with syntactic and semantic information and with several taxonomies of givenness 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. POS labels were taken from the original corpus. The POS tagset is STTS, the standard tagset for German corpora [20]. The guidelines for coding referring expressions as well as a more detailed coding manual can be found in [6, 27].

4 RESULTS

A first analysis of the texts shows that given entities are relatively scarce. Only 18% of all referring expressions refer to previously mentioned entities. Of these referring expressions, 33.3% are realized as pronouns, and 96% of all pronouns specify given discourse entities. From the literature [2, 15] we derive the preliminary hypothesis that referring expressions which specify new entities have to be accented, while that those that specify old entities need not be accented [15]. Table 2 shows strong accentability defaults at work: nouns are accented, pronouns are not. Proper names

POS		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. pron.	20	5.0	*15.0 (0.15)*
PPOSAT	attrib. pers. pron.	34	2.9	*14.7 (0.18)*

Table 2: 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%

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. 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. Apparently, the Fujisaki model picks up small excursions in the F0 contour that were not large enough for the human labeller to perceive an accent.

The lack of accent on pronouns is easy to explain in terms of entity status: pronouns specify active discourse entities [2]. But what about nouns, where entity status is more variable? For American English, Cahn [3] implemented an algorithm 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. For German, Kohler shows in [7] 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 with 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. Since there are over 80 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 with no 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 3 confirm neither hypothesis. A series of Fisher tests for the four taxonomies DISC, HEAR, STAT3, and STAT4 reveal only 6 significant associ-

	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
B/I > 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

ations ($p<0.05$.) 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, 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. Discourse-new expressions are made more prominent than discourse-old ones. For proper names, we 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 (65%) than elsewhere in the sentence; their average accent command amplitude is also significantly lower (0.209). Logistic regression reveals that of the three variables entity status (represented by STAT4, STAT3), sortal class, and syntactic function, syntactic function ($p<0.001$) exerts larger influence on accentuation than entity status ($p<0.01$).

Phrasing is closely related to text structure. The first phrase command of a news item is marked with a median amplitude of 2.15. All sentence boundaries (median amplitude: 1.43) and 74.8% of all commas (median amplitude: 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 after noun phrases that belong to first mentions, which tend to consist of more words and carry more modifiers [27] ($p<0.001$, 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 these radio news texts, where most sentences contain highly complex long NPs and semantically empty verbs, this strategy makes perfect sense [27].

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 high functional load, intonation contours must be highly polysemous. One and the same contour can be used for signalling thematicity or for indicating non–finality. It is not clear whether the categories in a phonological system such as ToBI manage to cover all relevant distinctions. 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 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 to phonological categories should not be discarded outright. The question is to what degree the gradient effects that are abstracted away by phonological approaches are important for generating natural F0–contours.

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

6 REFERENCES

[1] Bengt Altenberg. *Prosodic patterns in spoken English — Studies in the correlation between Prosody and Grammar for text-to-speech conversion*. Lund University Press, Lund, 1987.

[2] Gillian Brown. Prosodic structure and the Given/New distinction. In Anne Cutler and D. Robert Ladd, editors, *Prosody: Models and Measurements*, Berlin etc., 1983. Springer.

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

[4] Caroline Féry. *German Intonation Patterns*. Niemeyer, Tübingen, 1993.

[5] 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.

[6] Lynette Hirschman and Nancy Chinchor. MUC-7 coreference task definition, `http://www.muc.saic.com/proceedings/`, 1997.

	common noun								proper name	
	disc. old / new	hearer old / new	new	med	old	active	unused	total	disc. old / new	hearer old / new
total	729 / 94	250 / 573	187	386	250	94	156	823	27 / 154	136 / 45
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
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
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*	39.7 / 37.8
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
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

Table 3: Prosodic marking of entity status

[7] Klaus Kohler. Terminal accent patterns in single accent utterances of german: Phonetics, phonology, and semantics. pages 115–186.

[8] D. Robert Ladd. *Intonational phonology*. Number 79 in Cambridge Studies in Linguistics. Cambridge University Press, Cambridge, UK, 1996.

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

[10] Jörg Meyer. *Intonation und Bedeutung*. PhD thesis, Lehrstuhl für Experimentelle Phonetik am Institut für maschinelle Sprachverarbeitung, Stuttgart, 1997.

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

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

[13] Hansjörg Mixdorff and Hirose Fujisaki. In *Proc. ICSLP*, 2000.

[14] Bernd Möbius. *Ein quantitatives Modell der deutschen Intonation — Analyse und Synthese von Grundfrequenzkonturen*. Niemeyer, Tübingen, 1993.

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

[16] Janet Pierrehumbert and Julia Hirschberg. The meaning of intonation contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 271–311, Cambridge, MA, 1990. MIT Press.

[17] Ellen 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.

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

[19] K. Ross and M. Ostendorf. Prediction of abstract prosodic labels for speech synthesis. *Computer Speech and Language*, 10:155–185, 1996.

[20] Anne Schiller, Simone Teufel, and Christine Thielen. Guidelines für das Tagging deutscher Textcorpora mit STTS. Technical report, IMS Stuttgart/Seminar f. Sprachwiss. Tübingen, 1995.

[21] 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.

[22] K. Silverman, M. Beckman, J. Pitrelli, M. Osterndorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. Tobi: A standard for labeling English prosody. In *Proceedings of the International Conference on Spoken Language Processing, Banff*, pages 867–870, 1992.

[23] Eberhard Stock and C. Zacharias. *Deutsche Satzintonation*. VEB Verlag Enzyklopädie, Leipzig, 1982.

[24] Michael Strube. Never look back: An alternative to centering. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal, Québec, Canada, 10–14 August 1998, volume 2, pages 1251–1257, 1998.

[25] Bonnie Lynn 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.

[26] Maria Wolters. Prosodic correlates of referent status. In *Proc. ICPHS*, San Francisco, CA, 1999.

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