FROM *PROC. OF EUROSPEECH 99, BUDAPEST, HUNGARY, SEPTEMBER 1999, (PP. 519–522).*

AUTOMATIC ANNOTATION AND CLASSIFICATION OF PHRASE ACCENTS IN SPONTANEOUS SPEECH

A. Batliner M. Nutt V. Warnke E. Nöth J. Buckow R. Huber H. Niemann*

Chair for Pattern Recognition, University of Erlangen–Nuremberg
Martensstr. 3, D–91058 Erlangen, Germany
* now with Bavarian Research Center for Knowledge Based Systems (FORWISS)
Am Weichselgarten 7, D–91058 Erlangen, Germany

e–mail: batliner@informatik.uni-erlangen.de
www: http://www5.informatik.uni-erlangen.de

ABSTRACT

During the last years, we have been working on the automatic classification of boundaries and accents in the German VERBMOBIL (VM) project (human-human communication, appointment scheduling dialogues). A sub-corpus was annotated manually with prosodic boundary and accent labels, and neural networks (NN) trained with a large set of prosodic features were used for automatic classification. The classification of boundaries could be improved markedly with a combination of the NN with a language model (LM) that was trained with manually annotated syntactic-prosodic boundary labels in a much larger sub-corpus. Here we show how a combination of NN with LM along similar lines can be used for an improvement of accent classification as well. For the training of the LM, accents are annotated automatically in the transliteration with the help of a rule–based system that uses part–of–speech (POS) as well as other linguistic/phonological information.

1. INTRODUCTION

This research has been conducted under the VM project which aims at automatic speech–to–speech translation in appointment scheduling dialogues in an end–to–end system. At present, accent information is used for disambiguation, cf. *in einem ZUG* (in a train) vs. *in EINEM Zug* (at one stroke) [8]. For the training of our LMs, large annotated databases are needed. The perceptual labelling of accents, however, is very time consuming, and the database provided in VM is therefore not too large. Moreover, accentuation is influenced by speaker idiosyncrasies, rhythm, etc. as well. For linguistic interpretation, this has to be treated rather as random noise, i.e., the higher linguistic modules syntax and semantics are not necessarily and always interested in 'perceptual' accents but rather in 'syntactic–semantic' accents.

In [2] we have shown that the classification of boundaries could be improved markedly with a combination of an NN with an LM that was trained with manually annotated syntactic-prosodic boundary labels in a much larger sub-corpus (14 hours of speech). In theory, there is one prominent, primary/phrase accent in one (accent) phrase (AP). Even if this might not always be true, it seems worth while to try and label accent position automatically on the basis of phrase boundary positions that were labelled using only the transliteration of the dialogues. The degree of freedom, so to speak, is, however, probably higher for accents than for boundaries; we should therefore not only rely on the 'syntactic–phonologically correct' position of accents but in the classification phase, we should take into account their concrete phonetic realization as well in order to use this information for semantic disambiguation, discourse analysis, and translation. This means that we have a sort of 'open' concept of phrase accents: (syntactic–prosodic) phrases and accents are mutually dependent on each other, and in a default reading *(out–of–the–blue)*, there might be a one–to–one relationship. Accents are, however, used for different purposes. Moreover, it depends on the hierarchy of phrases and accents: which level of accentuation should be attributed to which level of phrasing? In our context, it is simply a matter of the application/special task whether we want to have several or exactly one accent per phrase, cf. section 6.

Older accounts of a German accent phonology can be found, e.g., in [7], newer ones within the tone–sequence–approach are, e.g., [12] and [4]. The database in these studies consists either of introspective material or of elicited, read speech. The automatic assignment of accent position for the *read* speech material of the ERBA database along similar lines as described below turned out to be very successful [5, 6] with recognition rates of up to 96%. Up to now, there are no accounts of accentuation based on large *spontaneous* German speech databases.

In [11], the most important factors affecting pitch accent placement in (read) American English were POS, word class, break index after word, and number of syllables to the next pitch accent. A prominence–based approach for the generation of accents for German is described in [13]. The parameters used in these studies are similar to those that we will use in our approach.

2. MATERIAL AND ANNOTATION

The annotation of acccents follows a ToBI–like strategy: first, prosodic boundaries are labelled, then, within the prosodic phrase, one (or more than one) primary/phrase accent position is labelled, and then, if necessary, secondary accent positions are labelled as well [10]. Four different types of syllable–based accent labels are distinguished which can be mapped onto word–based labels denoting if a word is accentuated or not: EC: *emphatic* or *contrastive accent*, PA: *primary accent*, SA: *secondary accent*, and UA: *any other unaccentuated syllable* (default). 33 dialogues (approx. 2 h of speech) have been labelled along these lines [10]. These accent labels will serve as reference for our accent rules. They have been taken as data for the training of statistical classifiers as well, cf. [6]. For our rule–based labels, we use for PA A3, for SA A2, and for no accent A0, in analogy to our boundary labels [2]. For classification, we are at the moment only interested in the two-class problem 'accentuated word' ($A = \{EC, PA, SA\}$) vs. 'not accentuated word' ($\neg A = UA$). Note that another clustering that, e.g., assigns the intermediate label SA to $\neg A$, or a three–class problem A3, A2, and A0, would of course be possible as well depending on the application.

In the training database BS–TRAIN, there are 30 dialogues, 797 turns, and 96 minutes of speech; the corresponding figures for the test database BS–TEST are: 3 dialogues, 64 turns, and 11 minutes of speech. BS–TRAIN and BS–TEST are mostly entailed in the first CD–Rom (CD1); on all German CD–ROMs of the first phase of VM 1993-1996 (CDs–VMI), there are 793 dialogues, 13.924 turns, and 2.035 minutes of speech. We see that approx. twenty times the amount of data is available for the training of a classifier if we are not confined to the prosodic–perceptual labels; as for other disadvantages of a purely

[1]This work was funded by the German Federal Ministry of Education, Science, Research and Technology (*BMBF*) in the framework of the Verbmobil Project under Grants 01 IV 102 H/0, 01 IV 102 F/4, and 01 IV 701 K5. The responsibility for the contents of this study lies with the authors.

part of speech (POS)	cover class	main class
noun	NOUN	CW
proper name	NOUN	CW
auxiliary	AUX	FW
copulative verb	AUX	FW
verb (all other verbs)	VERB	CW
infinitive (or 1./3. pers. plur.)	VERB	CW
participle (pres./past, not infl.)	APN	CW
adjective, not infl., pred./adv.	APN	CW
adjective, infl. (attributive)	API	CW
article, pronoun	PAJ	FW
particle (adv., prep., conj.)	PAJ	FW
interjection	PAJ	FW
character (spelling mode)	NOUN	CW
fragment (of a noun, capital letter)	NOUN	CW
fragment (¬noun, small letter)	API	CW
special word SW (list)	PAJ	FW
particle verb PV (list and rule)	VERB	CW
particle verb particle PVP (list)	PAJ	FW
not accentuated (list of exceptions)	EX	CW

Table 1. parts–of–speech in the lexicon

perceptual labelling, cf. [2]. We thus wanted to develop an automatic labelling of accent position based on the syntactic–prosodic phrase. Boundaries of these phrases (so called M boundaries) were annotated for CDs–VMI using only the written transliteration. A phrase is denoted to its left either by the beginning of the turn (no M label but it can be assigned automatically) or by an M label, and to its right again by an M label or by the end of the turn. The M labels are described in detail in [2]. For our present purposes, we map the detailed classes onto M3 (main phrase boundary including minor phrase boundary M2), M0 (no phrase boundary, default), and MU (ambiguous boundary). The special treatment of MU for accent assignment is described in more detail in [9, 3].

3. PART–OF–SPEECH LABELS (POS)

For the automatic assignment of phrase accent position starting from the syntactic–prosodic M labels, we have to know the POS of each word which is therefore annotated in the lexicon. This lexicon contains all word forms that can be found in CDs–VMI. Besides, there exist special lists of words that can have different syntactic functions or semantic meanings depending on whether they are accentuated or not. Actually, the POS can only be annotated unequivocally if the syntactic context is known. For the isolated word form in the lexicon, we have to find a compromise using the following strategy: If in doubt, we rely on the transliteration, e.g., in the case of near–homographs where the initial letter (capital vs. small letter) can tell apart noun from adjective. We use probability in general and the probability in the VM scenario, we specify if possible (unequivocal morphology), and we underspecify if necessary, i.e., if we cannot tell apart different POS. Details can be found in [9, 3]. In Table 1, each POS label is described shortly and mapped onto its cover class and its main class (content word CW or function word FW). The first 15 POS are self–explaining. 'Special words' (SWs) are FWs that can have a different semantic depending on whether they are accentuated or not, cf. the examples in section 1. Particle verbs PV and their particles PVP as well as 'not accentuated' (some rare CWs) behave differently from their original class and have to be modelled especially. These cases cannot be dealt with in this paper; we refer to [9, 3]. (There, some other necessary preprocessing is described as well, e.g., the treatment of auxiliaries with the function of a full verb, or the treatment of spelled sequences of characters.) If SWs are found in an AP, both the normal accent positions and these words are given the label 'accent undefined' AU; examples are given in section 4. In a first approximation, cf. section 5, we can disregard these labels. In the future, we want to leave AU positions out from the LM training and classify them only with the NN, i.e., we will rely only on the acoustic realization.

Such an approach yields erroneous results in some cases; we believe, however, that this does not matter very much for our rules. E.g., particles that can be either a conjunction at the beginning of an AP or a local adverb somewhere in an AP might be told apart most of the time because of their position in the AP. Another example is the word *halt* which can be either an exclamation/imperative form of the verb *halten* (hold on/wait) or a modal particle (just/simply). We annotate it as a particle because this probability is very high in our corpus. In the case of the other reading, even if it was analyzed as an FW it would most certainly constitute a one-word-phrase and thus be annotated correctly with A3. Moreover, such an approach is much more usable if one has to deal not with the spoken word chain but with word hypotheses graphs where the left and right context of a word cannot be defined easily - and such a task is, after all, the 'real life' job of automatic speech processing.

4. ASSIGNMENT OF ACCENT POSITION

The 'accent grammar' described in the following is a rule based system with sequential control. Input is the surface POS sequence, especially the last four words in an AP. In German, there is a tendency towards an accentuation in the ultima, i.e., towards the right edge of a phrase; we assume that the accent is somewhere on the last four words of an AP. If an AP is longer than four words, additional default rules are used. In the rules, word$_{n-3}$ is the fourth last word *(anteantepenultima)*, word$_{n-2}$ the third last word *(antepenultima)*, word$_{n-1}$ the second last word *(penultima)*, and word$_n$ the last word *(ultima)* in an AP. If we can formulate the rules with only the two main classes, the position of the accent(s) is given in a basic version in Table 2, 'general rules'. '*' denotes A3, '+' A2. Bracketed positions are optional but of course the optional position at word$_i$ presupposes that all positions word$_{i+1}$ ... word$_{n-1}$ are not empty. This is the case for an FW in ultima position, and no CW or an alternation of CWs and FWs in the second and third last position (word$_{n-1}$ and word$_{n-2}$). It is also the case for CW in ultima together with FW in penultima position. For all other constellations with a CW clash, we need the POS N(oun) and V(erb) with their complement A (¬N,¬V). These rules are given in Table 3 in a basic version, together with the rules for PVs and PVPs. Note that for the moment, only the starred positions A3 are fixed. It is a matter of empirical evaluation whether the positions of A2 are valid or whether for some constellations, they should better be modelled with either A3 or A0. For classification, the rules given in Tables 2 and 3 were modified slightly, based on the distribution of the accents in the training database [9, 3]. There might still be some inconsistencies in the formulation of our rules but for the moment, we do not want to make them fully consistent simply because we do not know yet which formulation mirrors best the empirical distribution.

The following examples illustrate some of the rules given in 2 and 3:

(1) *kein FW:A0 Problem CW:A3 für FW:A0 mich FW:A0* (no problem for me)
(2) *wie lange CW:A2 wir FW:A0 brauchen CW:A3 werden FW:A0* (how long it'll take for us)
(3) *ich bedank' mich FW:A0 auch SW:A0:AU recht CW:A2:AU herzlich CW:A3:AU* (thank's a lot)
(4) *... alle Termine *N:A3 festgelegt +A:A2 haben FW:A0* (... have fixed all dates)
(5) *... dass das zu lange *A:A3 geht +V:A2* (... that it'll take too long)

Examples (1) and (2) illustrate general rules (Table 2) without SWs or CW clash. In example (3), if the SW *auch* is accentuated as an inclusive focus particle, it is presupposed that someone else expressed his thanks as well. If it is a modal particle and not accentuated, it does

word_{n-3}	word_{n-2}	word_{n-1}	word_n
FW in ultima, general rules			
(FW)	(FW)	(FW)	*FW
*CW	FW	FW	FW
(FW)	*CW	FW	FW
(FW)	(FW)	*CW	FW
+CW	FW	*CW	FW
CW	*CW	FW	FW
FW in ultima, specific rules			
(FW)	CW	CW	FW
CW	CW	CW	FW
CW in ultima, general rules			
(FW)	(FW)	(FW)	*CW
+CW	FW	FW	*CW
CW	+CW	FW	*CW
(FW)	+CW	FW	*CW
CW in ultima, specific rules			
(FW)	(FW)	CW	CW
(FW)	CW	CW	CW
(CW)	CW	CW	CW
CW	FW	CW	CW

Table 2. Accent rules, general, basic version

		CW	CW	FW
	word_{n-3}	word_{n-2}	word_{n-1}	word_n
word_{n-3}	word_{n-2}	word_{n-1}	word_n	
		+N	*N	
		A	*N	
		V	*N	
		*N	+A	
		+A	*A	
		V	*A	
	(*N)	*N	+V	
		*A	+V	
		*V	V	
			*CW	+PVP
		*CW	+PV	

Table 3. Accent rules, specific, basic version

not carry any specific meaning. Thus, *auch* and all other accentuated words in this phrase are labelled with AU. (In our present system, we only classify a word in 'absolute terms' without taking into account the degree of accentuation of other AU positions in the same AP. We thus leave the decision amongst words labelled with AU to the higher linguistic modules in VM.) This difference might not be considered to be too relevant but cf. *FINDE ich schon* (I'll find that for sure) vs. *finde ich SCHON* (I really do believe that). The last two examples (4) and (5) illustrate other specific rules (Table 3) with CW clash.
Note that it is easy to find cases that possibly are not covered or contradicted by our rules: The degree of freedom is rather high for different degrees of accentuation, especially for words that can carry the SA; these words can often as well have PA or no accent. This is no matter to be solved via examples but only via statistical distribution. Using introspection, native speakers can often find (slightly) different accent pattern for one and the same AP. There can always be idiomatic expressions with an accent distribution that is not covered by our rules, and there can be some erroneous POS labels or M labels. POS is not unequivocal; this causes, e.g., a wrong analysis for the constellation [+A *A FW] *auf halb A2 drei A3 einigen M3* (to agree on half past two). *einigen* is here not the pronoun 'some' but a verb. The analysis is wrong, the result of the wrong application of a rule, however, is insofar correct as a verb does not carry A3 in this context.
We do not assume a total de–accentuation of that CW in a [CW CW] sequence that is not the carrier of the PA but annotate it most of the time with SA. Two factors might be responsible for that: first, a CW clash is not automatically an accent clash if there are de–accentuated syllables in between. Second, there might be something like accent spreading from the PA onto the adjacent CW resulting in an SA on that CW. The choice between these two expla-

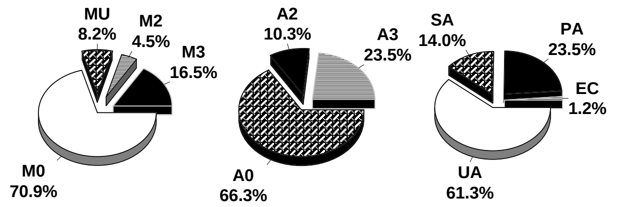

Figure 1. The three stages in our approach: syntactic–prosodic boundaries (left), rule–based accent positions (middle), acoustic–perceptual accent labels (right) for BS–TEST

train and test	λ_{NN}	λ_{LM}	$\mathcal{RR}$	$\mathcal{RR}_{\overline{C}}$
train: rule–based	0,0	1,0	86,6	86,3
(LM, NN)	0,5	0,5	87,9	87,1
test: rule–based	1,0	0,0	78,4	76,8
train: rule–based	0,0	1,0	76,3	75,3
(LM, NN)	0,6	0,4	82,5	81,9
test: acoust.–perc.	1,0	0,0	79,9	79,5
train: rule–based (LM)	0,0	1,0	76,3	75,3
acoust.–perc. (NN)	0,6	0,4	82,4	81,8
test: acoust.–perc.	1,0	0,0	80,2	79,8

Table 4. Classification results in percent

nations can depend on whether one considers accent to be a phenomenon based on the syllable (phonetic point of view) or based on the word (syntactic–semantic point of view). Actually, the underlying reason does not matter as long as the phenomenon is modelled adequately.
Figure 1 illustrates the three stages in our approach with the distribution of the pertaining labels: We started with the syntactic–prosodic M boundaries and assigned each M3 and M2 boundary one A3 accent. These accents in turn are correlated with the acoustic–perceptual accents PA/EC. Approximately half of the MU boundaries should correspond to one PA/EC label. We can see that there are some additional A2 accents but of course not one per phrase, and that there are more SA labels than A2 labels. Further details can be found in [9, 3].

5. AUTOMATIC CLASSIFICATION

As classifier, we use the freely available NN simulator SNNS [14]. SA, PA, and EC as well as A2 and A3 belong to the class 'accentuated' A. For the training of the NN with the rule–based labels, we disregard all words labelled with AU. The feature vector consists of our usual 276 prosodic features [2, 1]. As NN, we use a multi–layer perceptron. In the training phase, each class is represented with an equal number of cases which means that, if necessary, items are copied. For our experiments, we annotated CD-ROMs 1-5 with rule–based accents; all words with ambiguous accents AU were mapped onto their respective main classes A0, A2, or A3. These data were taken as training sample for a trigram-LM. As test sample, we use BS-TEST, either with acoustic–perceptual or with rule–based accent labels. In Table 4, $\mathcal{RR}$ displays the overall recognition rate, $\mathcal{RR}_{\overline{C}}$ displays the mean of the class–wise computed recognition rate.
We ran three experiments each with 10 different weighting factors for LM and NN always summing up to 1.0; displayed are the 'edges' and the weighting factor λ that yields the best results. With the first set of experiments (Table 4 above), we try to classify rule–based labels. The LM–probabilities are combined with an NN that is trained on 4000 turns with rule–based labels. With the LM alone (λ_{NN} = 0,0), an $\mathcal{RR}$ of 86,6% could be achieved which is more than 8% higher than that achieved with the NN alone (λ_{NN} = 1,0). Best results yields a combination of LM and NN that takes into account both classifiers to the same extent: $\mathcal{RR}$ = 87,9, λ_{NN} = 0,5. With the sec-

ond set of experiments (Table 4 middle), we try to classify acoustic–perceptual labels. The LM–probabilities are combined with an NN that is trained on the rule–based labels of BS-TRAIN. With the LM alone (λ_{NN} = 0,0), results are worse ($\mathcal{RR}$ = 76.3) than with the NN alone ($\mathcal{RR}$ = 79.9). Again, best results are achieved with a combination of LM and NN: $\mathcal{RR}$ = 82.5, λ_{NN} = 0,6. The last set of experiments (Table 4 below) uses the same constellation as the second but an NN that is trained with the acoustic–perceptual labels. Results are almost the same as for the second set. These results obtained so far meet our expectations that, in analogy to the classification of boundaries, (1) classification with rule-based labels is in the same range as classification with acoustic–perceptual labels or even better, (2) an LM alone yields very good results, and (3) a combination of LM and NN improves classification results even more. The classification of acoustic–perceptual labels improves by more than 2% (11% reduction of error rate) and that of rule–based labels by more than 1% (10% reduction of error rate) with a combination of LM and NN. These results corroborate our results obtained for boundaries [2] and can be taken as a confirmation of the hypothesis that prosodic events like accents and boundaries are – to a large extent – not independent phenomena: not independent from each other, and not independent from syntax. The recognition rate for boundaries [2] is better than that for accents most probably because of the factors mentioned in section 1 and 6.

6. SUMMARY AND CONCLUSION

Starting point in our approach is the 'accent phrase' AP that is delimited to its left and to its right by a syntactic prosodic boundary. All words are annotated with POS in the lexicon. For the last four words in an AP, rules for accent assignment were formulated and evaluated iteratively. The following parameters were considered to be relevant in our approach: POS, position in the AP, rhythm, and possibly accent spreading. There is a tendency towards the following hierarchy: CWs are 'stronger', i.e., more prone to be the carrier of the phrase accent, than FWs. Within the CWs, nouns are stronger than adjectives, and both are stronger than verbs. If these two rules cannot decide amongst two words, the rightmost 'wins'. Special words have to be treated in a special way, e.g., particle verbs and particles that can have different functions depending on whether they are accentuated or not. We thus end up with two different accent labels for each word: a 'rule-based' syntactic-prosodic label that can be underspecified, and a perceptual-prosodic label. Even if both types discriminate primary and secondary accent, in the classification experiments, they were mapped onto one accent category resulting in a two-class-problem [+/-accent]. In analogy to our classification experiments with boundaries, we used NN alone, LM alone and combinations of NN with LM to classify automatically both types of accents. It was shown that the combination of NN with LM improved the classification of both types of accents.
We have already pointed out above that matters are a bit more complicated for accent classification than for boundary classification. There are more degrees of freedom which can often be traced back – at least in our application – to two main functions of accentuation: first, normal, default, *(out-of-the-blue)* accentuation which best can be treated with an LM; we have seen that these accents can be predicted to a great extent (88%). Second, special, non-default accentuation which best can be treated with a combination of LM and NN but with emphasis on the NN. We have to rely on the NN if we have to decide upon which of n possible PAs are realized or not. Lists for these special words have to be adapted to the needs of those higher linguistic modules that make use of accent information. Besides, a certain amount of 'noise', i.e., random variation, has to be faced as well. This noise is partly due to sparse phenomena which *could* be analyzed linguistically but cannot possibly be modelled statistically, partly simply due to intrinsic variation.
In a text–to–speech or content–to–speech system, the generation of a prosody which corresponds to the intention of the speaker is still a difficult task. This can partly be caused by problems encountered in the semantic content analysis if the content has to be analyzed based on speech recognition output. A sort of short–cut would be to pass prosody through as is, e.g., to use the probabilities for accentuation obtained by a classifier for generation. For shallow processing and translation without any content analysis, this might be the only manageable and possibly best solution. This means that one could use *prosodic substance* as a substitute for *semantic content.* (Note that the different requirements of generation and synthesis have to be met: generation wants to put emphasis on the salient part of the utterance; synthesis wants to produce naturally sounding speech which means in turn an alternation of accentuated and unaccentuated words.)
In the near future, we want to vary systematically the number of accents per AP with different tunings of our accent grammar and of the clustering of our accent labels. These results will in turn be processed and evaluated by the higher linguistic modules in VM.

REFERENCES

[1] A. Batliner, J. Buckow, R. Huber, V. Warnke, E. Nöth, and H. Niemann. Prosodic Feature Evaluation: Brute Force or Well Designed? In *Proc. ICPhS99*, San Francisco, 1999.

[2] A. Batliner, R. Kompe, A. Kießling, M. Mast, H. Niemann, and E. Nöth. M = Syntax + Prosody: A syntactic–prosodic labelling scheme for large spontaneous speech databases. *Speech Communication*, 25(4):193–222, 1998.

[3] A. Batliner, V. Warnke, E. Nöth, J. Buckow, R. Huber, and M. Nutt. How to label accent position in spontaneous speech automatically with the help of syntactic–prosodic boundary labels. Verbmobil Report 228, 1998. (http://www.dfki.uni-sb.de/verbmobil/)

[4] C. Féry. *German Intonational Patterns.* Niemeyer, Tübingen, 1993.

[5] A. Kießling, R. Kompe, A. Batliner, H. Niemann, and E. Nöth. Automatic Labeling of Phrase Accents in German. In *Proc. ICSLP94*, volume 1, pages 115–118, Yokohama, 1994.

[6] A. Kießling. *Extraktion und Klassifikation prosodischer Merkmale in der automatischen Sprachverarbeitung.* Shaker, Aachen, 1997.

[7] P. Kiparsky. *Über den Deutschen Akzent.* Akademie–Verlag, Berlin, 1966.

[8] H. Niemann, E. Nöth, A. Kießling, R. Kompe, and A. Batliner. Prosodic Processing and its use in Verbmobil. In *Proc. ICASSP97*, volume 1, pages 75–78, München, 1997.

[9] M. Nutt. Automatische Bestimmung von Akzenten und ihr Einsatz bei der Dialogakterkennung. Diplomarbeit, Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen–Nürnberg, 1998.

[10] M. Reyelt. *Experimentelle Untersuchungen zur Festlegung und Konsistenz suprasegmentaler Einheiten für die automatische Sprachverarbeitung.* Shaker, Aachen, 1998.

[11] K. Ross, M. Ostendorf, and S. Shattuck-Hufnagel. Factors Affecting Pitch Accent Placement. In *Proc. ICSLP92*, volume 1, pages 365–368, Banff, 1992.

[12] S. Uhmann. *Fokusphonologie.* Niemeyer, Tübingen, 1991.

[13] C. Widera, T. Portele, and M. Wolters. Prediction of Word Prominence. In *Proc. EUROSPEECH97*, volume 2, pages 999–1002, Rhodes, 1997.

[14] A. Zell and et al. SNNS, Stuttgart Neural Network Simulator, User Manual, Version 4.1., 1995. Institute for parallel and distributed high performance systems, University of Stuttgart.