SYNTAX AND PROSODIC PHRASING IN NEWS READINGS

Hansjörg Mixdorff

Faculty of Computer Science, Berlin University of Applied Sciences
mixdorff@tfh-berlin.de

Abstract: The current paper presents results on the relationship between syntax and prosodic phrasing in German news readings. A prosodically labeled database was analyzed with respect to the prosodic means employed by the speaker at prosodic boundaries of varying depth. A quantitative model was used for parametrizing the *F0* contours, whereas durational variations were captured by calculating syllabic z-scores. Statistical analysis shows that BI 4 boundaries are usually marked by pauses and resets of the declination line whereas shallow boundaries are mostly signaled by durational cues. Since certain inconsistencies were found concerning the distinction between BI 3 and 4 boundaries, the latter were regrouped into intra- and inter-sentence boundaries.

1 Introduction

It is an undisputed fact that talkers employ prosodic means for chunking an utterance into meaningful entities. Among the most important features facilitating segmentation we identify - with descending degree of importance - pauses, *F0* cues and durational cues. The interaction between these various prosodic cues, however, is not yet fully understood.

In the current study we therefore investigate the relationship between quantitatively defined prosodic cues and break indices as assigned by a G-ToBI labeler of a larger speech database of German radio news readings.

We then examine the relationship between syntax and prosodic boundaries and conclude with a short discussion concerning the distribution of prosodic cues at phrase boundaries. This study is part of a research project directed towards the development of quantitative prosodic models for Text-to-Speech synthesis.

2 Speech Material and Method of Analysis

The speech database examined is part of a German corpus compiled by the Institute of Natural Language Processing, University of Stuttgart and consists of 48 minutes of Deutschlandfunk news stories read by a male speaker [1], of a total of 13151 syllables.

Thanks to their regularity, radio news represent an interesting speaking style for examining the relationship between syntax and prosodic features. Underlying texts are thematically unrestricted, syntactically complex, and highly informative. Furthermore, the data is real-life material produced by a professional speaker in a neutral manner. This speech material also appears to be a good basis for deriving prosodic features for a TTS system which in many applications serves as a reading machine.

The corpus contains boundary labels on the phone, syllable and word levels and linguistic annotations such as part-of-speech. The corpus was prosodically labeled following the Stuttgart ToBI system [2].

F0 contours were extracted at a spacing of 10 ms. Instead of a prosodic analysis on the raw *F0* contours proper, the contours were parametrized by means of the well-known Fujisaki model of the

production process of *F0* [3]. The model produces very close approximations of natural *F0* contours in the *log F* domain based on two types of input commands: Impulse-wise phrase commands and step-wise accent commands. The resulting phrase component corresponds to the slow falling movement of the *F0* contour known as declination, onto which the faster changing accent component is superimposed which is associated with accented syllables.
Earlier work by the author was dedicated to a model of German intonation which uses the model for parametrizing *F0* contours. The contour is described as a sequence of linguistically motivated tone switches, major rises and falls, which are modeled by onsets and offsets of accent commands connected to accented syllables or boundary tones. Boundary tones are defined as rising tone switches occurring at phrase-finale syllables not bearing the lexical accent.

Prosodic phrases correspond to the portion of the *F0* contour between consecutive phrase commands [4]. The model was integrated into the TU Dresden TTS system DRESS and proved to produce a high naturalness compared with other approaches [5].

The Fujisaki-parameters were extracted applying an automatic multi-stage approach [6]. The mean base frequency *Fb* and time constants *alpha* and *beta* of the current speaker were estimated to be 50.2 Hz, 0.95/s and 20.3/s, respectively.

Duration contours were determined by clustering syllables with respect to their number of phones and the property of the nuclear vowel being either schwa or non-schwa. These features were found to be the most important intrinsic factors of syllable duration. Normalized durations were then calculated as the z-score of a syllable, i.e., with respect to class duration mean and standard deviation.

3 Results of Analysis

Table 1 gives an overview of boundary definitions in the Stuttgart G-ToBI system.

BI	Description
0	Word boundaries in clitic groups
1	Phrase-medial word boundaries
2	Disjuncture (pause) without clear tonal cue
3	Intermediate phrase boundary
4	Intonation phrase boundary

Table 1 - Definition of break indices in the Stuttgart G-ToBI system

Since BI2 boundaries occur very rarely in the database (N=78), we will focus our analysis mainly on BIs 3 (N=636) and 4 (N=735). In this analysis we relate the break indices assigned by the labeler to the following prosodic features:

- Pauses
- With respect to the *F0* contour: Resets or readjustments of the declination line as marked by phrase commands, and boundary tones with rising *F0* contours as marked by accent commands assigned to phrase-final syllables
- Normalized syllabic durations

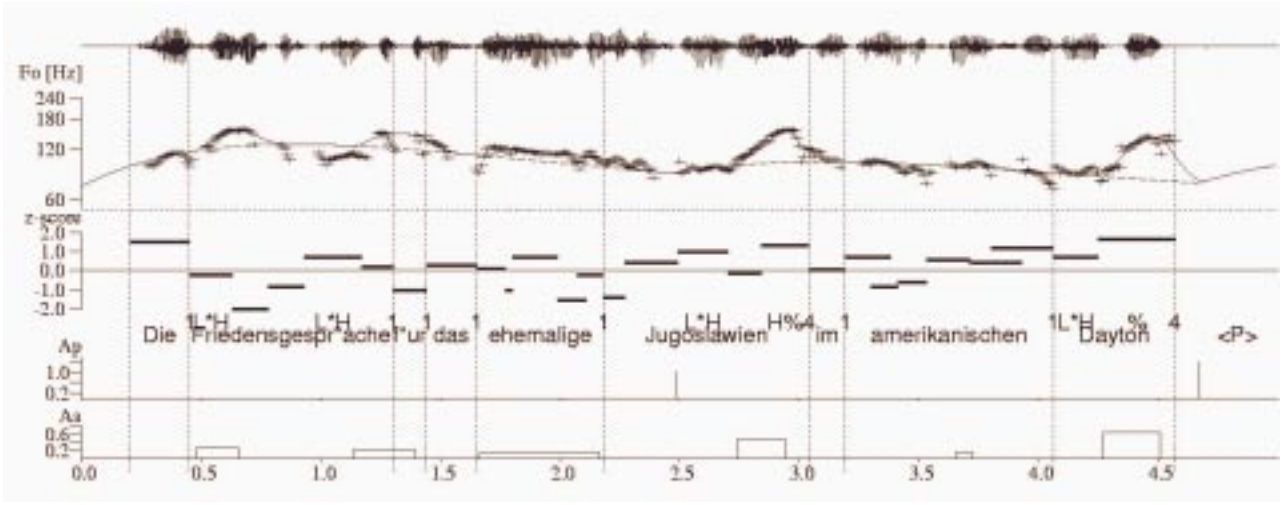


Figure 1 - Example of analysis from the database: (from top to bottom) Speech waveform, extracted and model-generated *F0* contours, syllabic z-score as indicated by horizontal lines of the length of the syllables, the ToBI tier, text of utterance, and the underlying phrase and accent commands. Text: "Die Friedensgespräche für das ehemalige Jugoslawien im amerikanischen Dayton..."-*"The peace talks for the former Yugoslavia in the American city of Dayton..."*

Figure 1 displays an example of analysis, showing from top to bottom: the speech waveform, the extracted (+ signs) and model-generated *F0* contours (solid line), the syllabic z-score as indicated by horizontal lines of the length of the syllables, the ToBI tier, the text of the utterance, and the underlying phrase and accent commands.

3.1 Locations of Prosodic Boundaries

Table 2 gives the types of boundaries with regard to the syntax and the occurrence of BI3 and 4.

type of boundary	BI 3	BI4
total	636	753
sentence boundary (full-stop)	0	360
comma	133	215
no punctuation mark, total	503	160
no punctuation mark, to the right of NPs	431	137

Table 2 - Types of boundaries assigned BI3 and 4, occurrences

We see that 80% of BI4 boundaries occur at punctuation marks in the text. About 85% of BI3 and BI4 boundaries which cannot be accounted for by punctuation marks are found to the right of nouns. This finding is easily explained by the syntactic requirements of German, since attributes are placed before, and not after the noun like often observed in French or Spanish, for instance. Therefore, NPs in German are concluded by the noun.

In Figure 1 we see two BI4 boundaries: One after 'Jugoslawien' and one after 'Dayton'. Whereas the first one is not followed by a pause, the second one is. This suggests certain inconsistencies of the human labeler assigning the break indices, as the prosodic cues are obviously not of the same strength in the two boundaries.

3.2 Analysis of Pauses

Statistical analysis of pauses yields the following picture : 83.0% of sentence boundaries are connected with a pause, with a duration of 716 ms/336 ms (mean/s.d.). Of the intra-sentence boundaries only 52.6% are accompanied by a pause of 327 ms/132 ms (mean/s.d.).

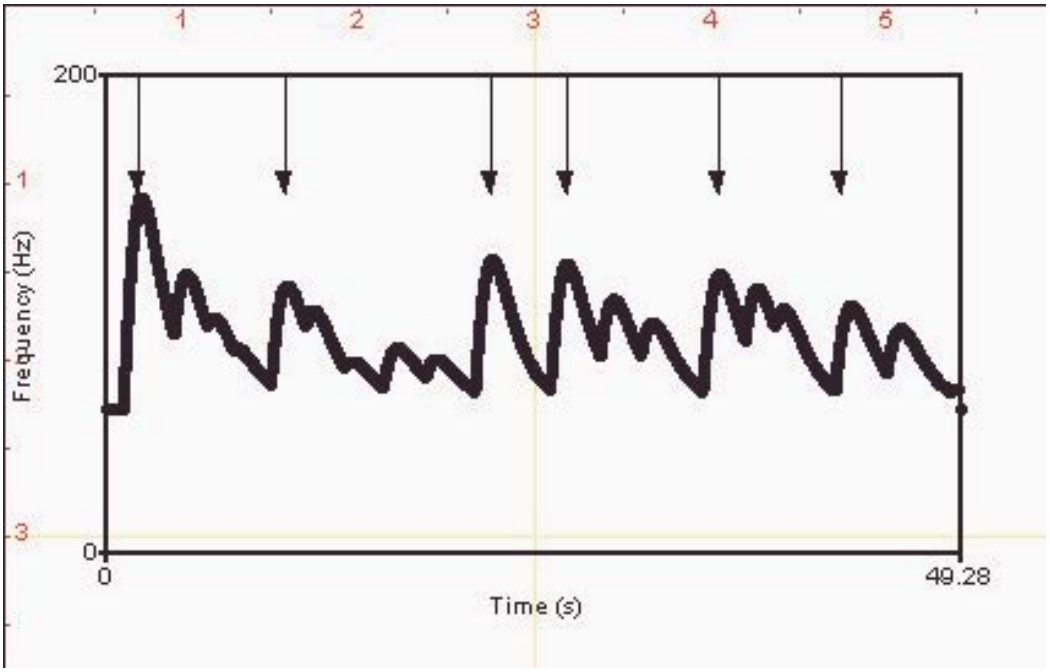
3.3 Analysis of *F0* Contours

Phrase commands. With respect to prosodic phrasing, the occurrence of a phrase command triggers a reset or an adjustment of the declination line. It is therefore located before any utterance and upcoming new phrase.
About 54.8% of break index 3- and 96.2% of break index 4-labeled-boundaries are aligned with the onset of a phrase command, with a mean phrase command magnitude *Ap* of 0.67 and 1.32, respectively.

In Figure 2, top, the phrase component extracted for a complete news paragraph is displayed where sentence onsets are marked with arrows. As can be seen from the figure, the magnitudes of the underlying phrase commands nicely reflect the phrasal structure of the paragraph. The underlying text is given in Figure 2, bottom, with the location of extracted phrase commands indicated by ↑, and break index 3 and 4 boundaries given in brackets. The example shows that most major syntactic phrase boundaries coincide with phrase commands, but in some cases of intra-sentence boundaries the phrase commands occur well into the segmental start of a prosodic phrase ("sagte der parlamentarische..."). This indicates that for intra-sentence boundaries the timing of phrase commands is more loosely related to the segmental onset of the underlying phrases. Due to their finite time constant phrase commands generally occur within an average distance of 300 ms of the segmental onset of a phrase. This is also the case in Figure 1, as the first phrase command (not visible), is located at -0.230 s.

In order to separate intra-sentence from inter-sentence boundaries more consistently, a distinction not expressed by the BI3 and 4 labels, boundaries were regrouped by the punctuation marks at which they occur, i.e. full-stops and commas/other. Subsequently we found that all inter-sentence-boundaries (N=357) are aligned with the onset of a phrase command. 68% of all intra-sentence boundaries (N=1014) exhibit a phrase command, with the figure rising to 71% for 'comma boundaries'.

The mean phrase command magnitude for intra-sentence boundaries, inter-sentence-boundaries and paragraph onsets amounts to 0.8, 1.68 , and 2.28 respectively, which shows that *Ap* is a good indicator for boundary strength. This is also reflected by the high correlation between *Ap* and the break index to the left of the first syllable in a phrase (ρ=0.696). *Ap* is also strongly correlated with the duration of inter-phrase pauses (ρ=0.571).
Since *Ap* is the parameter which describes the degree of readjustment of the declination line for the global *F0* contour, an important problem connected with *Ap* is whether the speaker adjusts the declination line depending on the length of the sentence (s)he is going to produce - starting on a higher level for longer utterances - which would be a proof for some kind of utterance pre-planning.
If we calculate the correlation coefficients for *Ap* with the current phrase duration and the duration of the preceding phrase in syllables, we see that the correlation with the duration of the preceding phrase (ρ=0.32) is clearly higher than the respective figure for the current phrase (ρ=0.11). Therefore we do not find strong evidence for an utterance pre-planning as discussed above. The degree of readjustment of the declination line is more influenced by the time elapsed since the



↑Der SPD-Fraktionsvorsitzende Scharping (3) wird↑den Bundestagsabgeordneten seiner ↑Partei (4) am Nachmittag in ↑Bonn (3) die Vertrauensfrage stellen (4). ↑ Dieser Schritt sei notwendig ↑geworden (3), da Scharping durch seine Abwahl als Parteichef (3)↑ beschädigt worden sei (4), sagte↑ der parlamentarische Geschäftsführer der SPD Fraktion↑Struck (4) der Deutschen Presseagentur (4). ↑ An der Sitzung wird auch der neue Parteivorsitzende Lafontaine teilnehmen (4). ↑ Nach Angaben der Bildzeitung (3) hat der saarländische Ministerpräsident ↑ unterdessen (3) ein Zehn-Punkte-↑ Programm (4) zum Kampf gegen die Erwerbslosigkeit vorgelegt (4). ↑ Darin sind unter anderem kürzere Arbeits↑zeiten (4), flexiblere Tarif↑verträge (3) und längere Maschinenlaufzeiten vorgesehen (4). ↑ Zudem (3) sollen Überstunden nur noch in Freizeit ↑abgegolten (4) und die Lohnnebenkosten gesenkt werden. (4)

Figure 2 - Phrasing profile of a complete news story (top) and underlying text with locations of phrase commands and break indices. Vertical arrows in the profile mark the beginning of new sentences (full stops). Rephrasing is easily identifiable by a strong readjustment of the declination line

preceding phrase command, which means that it becomes the stronger the lower the *F0* contour drops. The negative correlation between *Ap* and the index of the current phrase (ρ=-0.507) indicates that *Ap* is reduced for phrases later in the utterance which means a reduction of the *F0* range in the phrase component.

Duration of phrases. About 80% of the phrase command onsets occur within 3.4 s after the preceding phrase commands and consist of up to 13 syllables. An interesting detail is the fact that utterance-final phrases are generally longer (mean =3.24 s) than utterance-initial (mean =2.72 s) or medial ones. This may be explained by the observation that in declarative utterances the *F0* contour generally reaches its lowest point where the vocal folds relax and no need is given for phrasing and readjusting the declination line.

Boundary Tones. An additional means of marking phrase boundaries are so-called boundary tones (H% in the ToBI notation). We define boundary tones as distinct rises of *F0* at phrase-final syllables which are not lexical accent syllables. Boundary tones in the narrow sense only occur infrequently in the database, but it was observed that accent commands assigned to accented syllables preceding a prosodic boundary at a distance of between one and three

syllables are delayed by up to 100 ms. This indicates that they fulfil a function similar to a boundary tone, namely, the *F0* contour is sustained at a medium level signaling continuation. Accents preceding an intra-sentence boundary are also generally more prominent in terms of their accent command amplitude *Aa* than others (mean of *Aa* 0.34 against 0.25).

3.4 Analysis of Duration Contours

A large number of lower level prosodic boundaries are only signaled by durational means, i.e. by a lengthening of the phrase-final syllable. Compare, for instance, syllables 'en' in "Jugoslawien" and 'ton' in "Dayton" (Figure 1) which have a relatively high z-score. Statistical analysis shows that the lengthening effect mainly concerns the syllable immediately preceding the boundary (+73% compared with syllables at a distance of 2+ syllables from the boundary) and, to a much smaller degree the penultimate (+12%). This pattern prevails irrespective of whether or not a syllable is also accented.

3.5 Summary

Table 3 displays the occurrence of prosodic cues at intra- and inter-sentence boundaries. It can be seen that 31% of intra-sentence boundaries are mainly signaled by lengthening of the phrase-final syllable. More than one third of intra-sentence boundaries exhibit a readjustment of the declination line as triggered by phrase commands. The vast majority of inter-sentence boundaries (80%) exhibits a pause as well as a phrase command. This classification according to measurable prosodic features suggests that two types of higher level phrase boundaries (BI3 and 4 vs. intra-/inter-sentence) may not be sufficient for adequately describing boundaries, especially when we aim at the automatic chunking of utterances.
As far as rising boundary tones / (delayed) pre-boundary accents are concerned, these occur at most intra-sentence boundaries (i.e., continuation rises). There is, however, a minority of intra-sentence boundaries (17%) in the data preceded by a falling *F0* contour. These instances occur regularly after the introduction of a new subject, for instance, in the example displayed in Figure 2 after "Der SPD-Fraktionsvorsitzende Scharping..." Since the German default intonation would suggest a rising intonation to a medium level, this appears to be an idiosynchrasy of the speaker.

prosodic cues	intra-sentence	inter-sentence
durational	317	1
durational+pause	1	0
durational+phrase command	382	73
durational+phrase command+pause	314	283
Total	1014	357

Table 3 - Occurrence of prosodic cues at intra- and inter-sentence boundaries

4 Conclusions

The current study examined the realization of phrase boundaries in news readings. Most prosodic boundaries closely follow the syntax of the underlying text, especially punctuation marks and the grouping of noun phrases. This could be expected, as the text is at the basis of the utterances. The situation is probably different in conversational speaking styles which are

more governed by pragmatic considerations and interaction.
Prosodic cues generally employed include pauses, declination resets and boundary tones. Inter-sentence boundaries usually exhibit longer pauses, declination line resets and pre-final lengthening, whereas intra-sentence boundaries are marked to a lesser extent by declination adjustments, but by high boundary tones and lengthening. This graded use of prosodic cues suggests that two levels for deep phrase boundaries may not be sufficient.

With respect to the Fujisaki parametrization of *F0* contours it can be stated that the positions and magnitudes of phrase commands clearly reflect the phrasal structure of a discourse. Furthermore the raised accent command amplitudes *Aa* of pre-boundary accents and boundary tones indicate that these also serve as landmarks for structuring an utterance. Future research will be dedicated to the analysis of more informal speaking styles.

Acknowledgements

This research was supported by the Deutsche Forschungsgemeinschaft grant Mi 625/4-1.

References

[1] Rapp, S.: *Automatisierte Erstellung von Korpora für die Prosodieforschung*. PhD thesis Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, 1998.
[2] Mayer, J. *Transcription of German Intonation: The Stuttgart System*. Technischer Bericht, Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, 1995.
[3] Fujisaki, H. and Hirose, K.: *Analysis of voice fundamental frequency contours for declarative sentences of Japanese*. Journal of the Acoustical Society of Japan (E), 1984, 5(4): pp. 233-241.
[4] Mixdorff, H.: *Intonation Patterns of German - Model-based Quantitative Analysis and Synthesis of F0 Contours*. D.Eng. thesis TU Dresden (http://www.tfh-berlin.de/~mixdorff/thesis.htm), 1998.
[5] Mixdorff, H. and Mehnert, D.: *Exploring the Naturalness of Several German High-Quality-Text-to-Speech Systems*. Proceedings of Eurospeech '99, vol.4, Budapest, Hungary, 1999, pp.1859-1862.
[6] Mixdorff, H.: *A novel approach to the fully automatic extraction of Fujisaki model parameters"*. In *Proceedings ICASSP 2000*, vol. 3, Istanbul, Turkey, 2000, pp. 1281-1284.