THE INFLUENCE OF FOCAL CONDITION, SENTENCE MODE AND PHRASE BOUNDARY LOCATION ON SYLLABLE DURATION AND THE F0 CONTOUR IN GERMAN

Hansjörg Mixdorff and Hiroya Fujisaki
Dresden University of Technology and Science University of Tokyo

ABSTRACT

The present study examines the influence of focal condition, sentence mode and phrase boundary location on the most important prosodic features of German, syllable duration and F0 contour. A small corpus uttered by a single speaker was auditorily segmented and the F0 contours modeled using a quantitative model decomposing the contour into a sequence of tone switches. The correlation of these tone switches with changes of the accent syllable durations was examined. It was found that narrow focusing of an item boosts the tone switch at its word accent syllable whereas significant lengthening of this syllable was only observed in phrase-medial, but not in phrase-final position. In the latter case, instead of noticeable lenghtening of the word accent syllable, preceding items are being compressed . This indicates that the durational organisation of the phrase behaves in a different manner than the intonational one. Pre-phrase boundary syllables are lengtened for an average 60 % compared with phrase medial ones.

1. INTRODUCTION

In earlier studies by the authors a model of German intonation was developed which uses the quantitative Fujisaki-model [1] for parametrizing a given F0 contour. The contour is described as a sequence of tone switches, major rises and falls, which are modeled by onsets and offsets of accent commands connected to accented syllables. Prosodic phrases correspond to the portion of the F0 contour between subsequent phrase commands [2, 3]. The model was integrated into a German TTS and proved to produce a high naturalness compared with other approaches. Perception experiments, however, revealed inadequacies in the duration control of the TTS-system that put limitations to the prosodic naturalness achieved [4,5]. This calls for a refined duration model. Typically, current duration models are based on the statistical evaluation of large data bases. This provides for a good coverage of possible segmental environments. The influence of focal conditions, sentence mode and boundary location on syllable duration, however, is blurred, because the segmental contexts in which they occur vary throughout the data base. Besides, the relationship with the F0 contour is usually neglected in this kind of analysis.
The current study closely examines data where the segmental context is kept constant, but the underlying linguistic information varies.

2. SPEECH MATERIAL AND METHOD OF ANALYSIS

The idea behind the corpus design was to cover a number of different linguistic functions of prosodic cues with a small number of sentences. The target utterance should contain mostly voiced sounds ensuring a continuous F0 contour. Hence, the German sentence 'Wir nehmen die U-Bahn nach Ruhleben' - *'We take the subway to Ruhleben'* (Ruhleben is a district of Berlin) was embedded into 16 different contexts. These include (1) a broad focus condition, (2) narrow focus on 'U-Bahn' or (3) narrow focus on 'Ruhleben'. The target sentence was uttered in question and statement mode as a single-phrase or part of a two-phrase utterance. Besides, a variant with a phrase boundary after 'U-Bahn' was produced. Table 1 gives an overview of all variants examined. Contexts 10 to 12 represent a sequence of two echo-questions.

Table 1: List of all contexts examined.

No.	phrasal condition	sentence mode	focus
1	single-phrase	statement	broad
2	single-phrase	statement	narrow on ‘U-Bahn’
3	single-phrase	statement	narrow on ‘Ruhleben’
4	single-phrase	echo-question	broad
5	single-phrase	echo-question	narrow on ‘U-Bahn’
6	single-phrase	echo-question	narrow on ‘Ruhleben’
7	two-phrase-initial	continuation	broad
8	two-phrase-initial	continuation	narrow on ‘U-Bahn’
9	two-phrase-initial	continuation	narrow on ‘Ruhleben’
10	two-phrase-initial	echo-question	broad
11	two-phrase-initial	echo-question	narrow on ‘U-Bahn’
12	two-phrase-initial	echo-question	narrow on ‘Ruhleben’
13	phrase boundary after ‘U-Bahn’	statement	broad
14	two-phrase-final	statement	broad
15	two-phrase-final	statement	narrow on ‘U-Bahn’
16	two-phrase-final	statement	narrow on ‘Ruhleben’

For illustration, we give examples for contexts 8, 13 and 15 in which the target phrase is part of a two-phrase utterance. Narrowly focused items are set in bold face.

“Wir fahren mit der **U-Bahn** nach Ruhleben, weil es mit dem Bus zu lange dauert.”-“*We take the **subway** to Ruhleben, because the bus takes too long.”*
“Wir fahren mit der U-Bahn, nach Ruhleben müssen wir nicht umsteigen.”-“*We take the subway, to Ruhleben we don’t need to*

change trains."
"Wir müssen keinen Parkplatz suchen, denn wir fahren mit der **U-Bahn** nach Ruhleben."-"*We don't need to search for parking space, since we take the **subway** to Ruhleben."*

All variants were uttered by a native speaker of German five times at an average speed of 6.5 syllables per second.
The speech data were directly sampled at 16kHz/16 bit using a PC soundcard. The F0 contours of all utterances were analyzed with the Fujisaki-model using the Analysis-by-Synthesis method aiming at reducing the mean square error in the log F domain. Segment boundaries were marked auditorily on the speech waveform.

3. RESULTS OF ANALYSIS

Figure 1 to Figure 3 show results of analysis for samples of contexts 1, 2 and 3. The figures display from top to bottom: the speech waveform, the extracted (plus-signs) and model-based F0 contours (solid line), the syllable-based z-score (defined as $(t_{syl} - t_\mu) / t_\sigma$ for log durations), and the underlying accent commands. The vertical lines denote syllable boundaries.

Table 2: Mean word accent syllable durations and tone switch intervals for the potentially accented syllables 'neh', 'U' and 'Ruh'.

	'neh'		'U'		'Ruh'	
No.	t_μ [ms]	dAa_μ	t_μ [ms]	dAa_μ	t_μ [ms]	dAa_μ
1	200	.48	138	.20	141	.-29
2	207	.00	**154**	**-.60**	135	.00
3	198	.00	115	.12	**140**	**-.71**
4	200	.15	141	.17	143	.41
5	206	.00	**161**	**.69**	139	.13
6	200	.00	126	.12	**159**	**.40**
7	210	.27	143	.36	155	-.47
8	206	.00	**164**	**.83**	152	-.15
9	191	.04	139	.05	**139**	**-.69**
10	194	.13	124	.21	152	.45
11	177	.35	**168**	**.52**	153	.23
12	205	.00	119	.12	**162**	**.71**
13	169	.00	134	.44	126	.48
14	191	.20	134	.29	129	-.56
15	187	.05	**155**	**-.68**	145	.00
16	168	.00	133	.08	**133**	**-.65**

The Influence of the Focal Condition

The evaluation first concentrates on the effect of focus shift on the word accent syllable durations and tone switch intervals at the potentially accented items ' 'neh-men ', ' 'U-Bahn' and ' 'Ruh-le-ben'. Table 2 gives mean syllable durations t_μ and tone intervals (in terms of changes of *Aa*, denoted as dAa_μ) for all 16 contexts. Cases where a respective item is narrowly focused are marked by bold face.
If we examine the results for the item 'U-Bahn' we find, that a narrow focus results in a significant boost of the tone switch, as can be seen from the value of dAa comparing contexts 2 (Figure 2) with contexts 1 (Figure 1) and 3 (Figure 3), for instance.

The duration of the accented syllable 'U' varies by an average 27 % between de-accented (narrow focus on 'Ruhleben') and narrow focus conditions. Independent-samples T-Test shows that these results are all highly significant ($p < .001$).
Under broad focus condition, on the average, the syllable 'U' is 8% longer than in the de-accented version, but this result does not prove to be significant ($p < .080$). In other words, for the syllable 'U' duration and tone switch interval are highly correlated ($\rho = .69$).

The tone switch interval for the item 'Ruhleben' is influenced in a similar way as the one on 'U-Bahn' since it is boosted when 'Ruhleben' becomes narrowly focused. However, the accent syllable 'Ruh' does not exhibit any significant duration change ($p > .336$).

The tone switch at the item 'nehmen' is clearly reduced under all narrow focus conditions, whereas the duration of the accent syllable 'neh' remains largely anaffected.

Sentence Mode

Since it can be expected that the sentence mode mostly influences the final part of a phrase we examined the duration of the syllable 'ben' under all contexts and found that it is significantly longer in statement-final than in question-final position (199 vs. 153 ms, $p < .001$).
In context 13 where it is the third syllable of the second phrase it is further compressed to an average of 123 ms.
Statements are generally marked by a negative tone switch at the last accented item in the phrase, whereas questions exhibit a positive tone switch at this item and a question-final rise on the last syllable 'ben'.

Phrase Boundary Location

If we compare the results from context 13 with those from context 7, both of which represent two-phrase statements under broad focus, the vastest difference found is the lengthening of the syllable 'Bahn' in context 13 (an average 318 ms against 200 ms, $p < .001$). The tone switch on the pre-boundary accent syllable 'U' in context 13 is only slightly higher than that in context 7 (.44 vs. .36).
In order to summarize the results on syllable durations, Figure 4 gives averaged syllable-based z-scores for all different focal conditions. In addition, the results from context 13 are displayed. Whereas the first four syllables "Wir nehmen die" are largely unaffected by focus shift, narrow focus on 'U-Bahn' leads to a considerable lengthening of the accent syllable 'U'. In contrast, narrow focus on 'Ruhleben' compresses the pre-focal syllables 'U', 'Bahn' und 'nach'. Comparison with context 13 shows that the phrase boundary after 'U-Bahn' stretches the phrase-final syllable 'Bahn', while compressing the second phrase-initial syllables 'nach', 'Ruh' and 'le'. This indicates a certain compensatory effect.

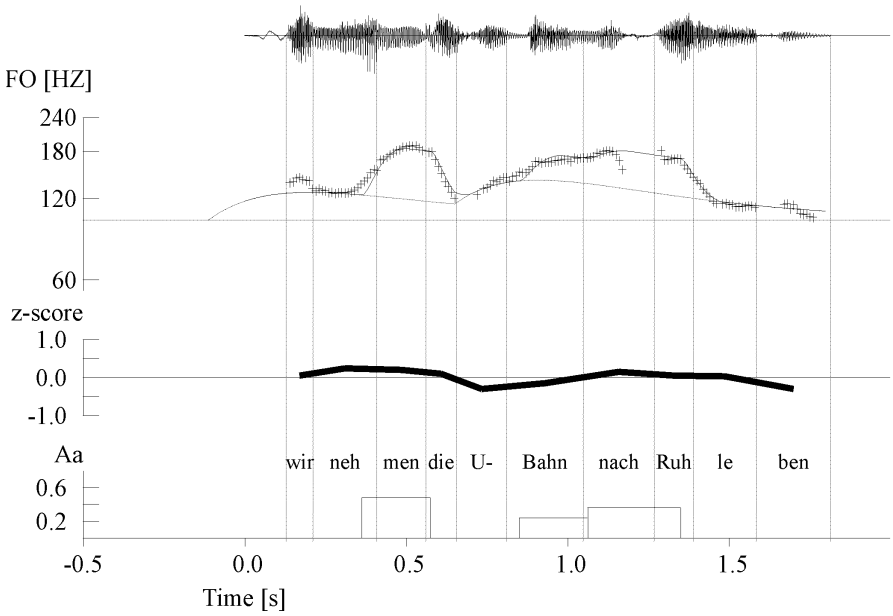

Figure 1. Example of analysis for a sample of context 1. The figure displays from top to bottom: the speech waveform, the extracted (plus-signs) and model-based F0 contours (solid line) F0, the syllable-based z-score, and the underlying accent commands. The vertical lines denote syllable boundaries.

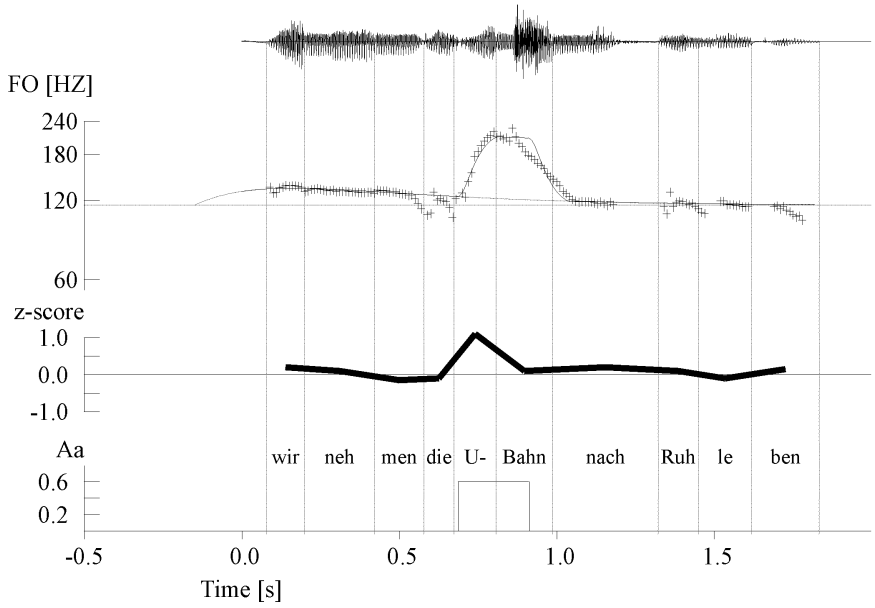

Figure 2. Example of analysis for a sample of context 2 (narrow focus on 'U-Bahn').

4. DISCUSSION AND CONCLUSIONS

It must be stated that the material examined is very limited and at this stage results cannot be generalized without further investigation. On the data presented here, it was observed that focus shift influences the F0 contour more strongly and in a more uniform way than syllable durations. This might be explained by the greater freedom in the use of tone switches for coding prominence information, since they can be completely deleted, whereas syllables cannot be compressed below a certain length. Furthermore, it is interesting to note that focus shift influences phrase-medial items in a different way than phrase-final ones as the latter do not show any significant durational changes.

As far as the indication of phrase boundaries is concerned, syllable duration obviously plays a more imminent role than modifications of the pre-phrase boundary tone switch.

We are aware of the fact that perceptual experiments are needed to verify our experimental results' validity. These will be subject of future research.

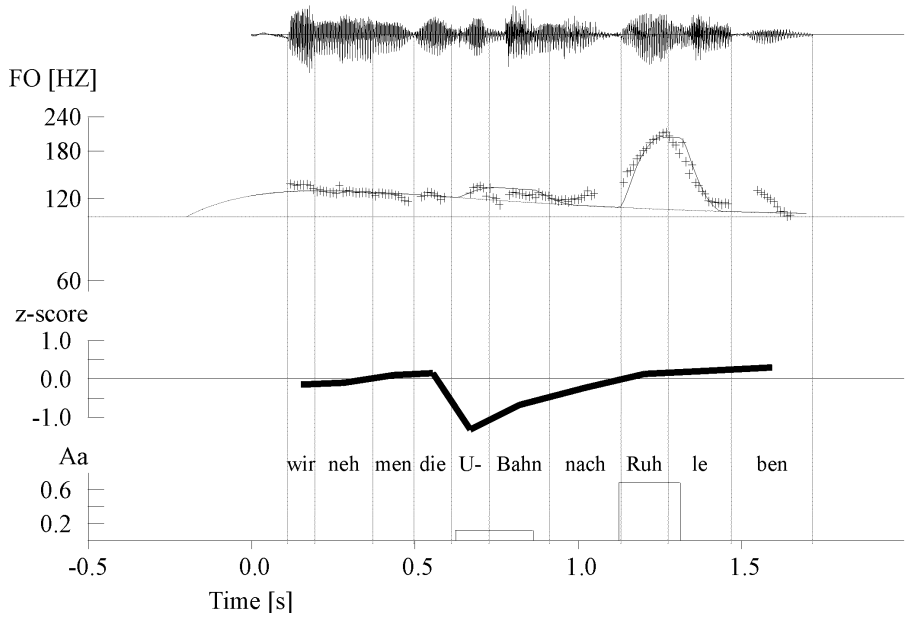

Figure 3. An example of analysis from context 3 (narrow focus on 'Ruhleben').

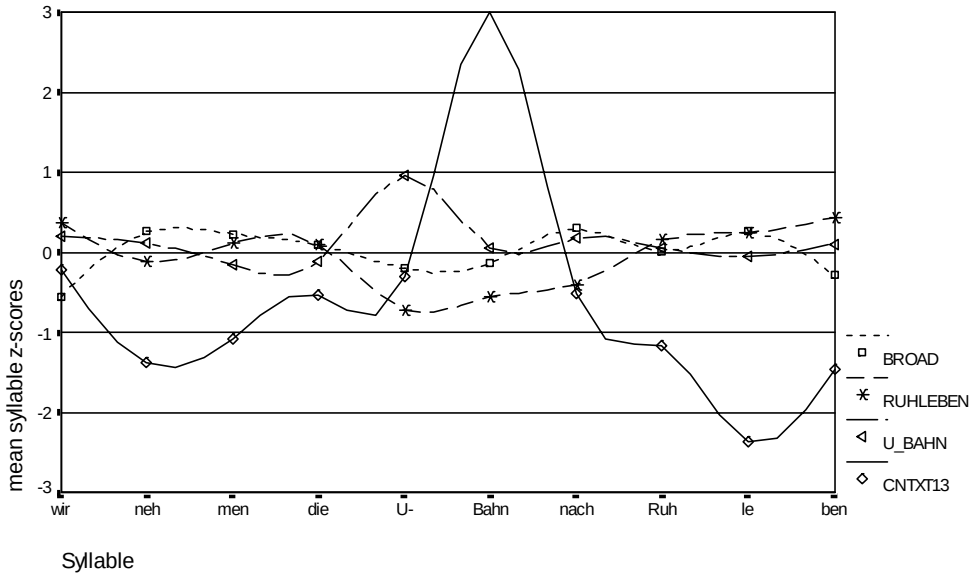

Figure 4. Syllable z-scores averaged over the three different focal condition BROAD, narrow focus on 'U-Bahn' and 'Ruhleben', respectively. CNTXT13 denotes the context in which a syllable boundary occurs after 'U-Bahn'.

REFERENCES

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

[2] Mixdorff, H., Fujisaki, H. 1994. Analysis of Voice Fundamental frequency Contours of German Utterances Using a Quantitative Model. In: *Proceedings of the ICSLP `94*, Yokohama, Bd. 4, S. 2231-2234.

[3] Mixdorff, H. 1998. *Intonation Patterns of German - Model-based Quantitative Analysis and Synthesis of F0-Contours*. PdD thesis submitted to TU Dresden.

[4] Mixdorff, H., Mehnert, D. 1998. Vergleichende Untersuchung zur Natürlichkeit synthetischer F0-Konturen. In the *Tagungsband der 9. Konferenz Elektronische Sprachsignalverarbeitung*. Dresden.

[5] Mixdorff, H., Mehnert, D. 1999. Comparing the Naturalness of Several Approaches for Generating F0 contours in German Text-to-Speech Systems. To be presented at the 137th regular meeting of the ASA/25th German Acoustics DAGA Conference. Berlin.