Published in: Proceedings, Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99), Budapest, Hungary, September 5-9, 1999. ISCA Archive, http://www.isca-speech.org/archive/eurospeech_1999. - Baixas: ISCA, 1999. pp. 21-24.

THE RELATIONSHIP BETWEEN UTTERANCE TYPE AND F0 CONTOUR IN GERMAN

Caren Brinckmann & Ralf Benzmüller

ABSTRACT

In this study we investigate the intonational characteristics of the four utterance types statement, wh-question, yes/no-question and declarative question. Readings of two German scripted dialogues were examined to ascertain characteristic features of the F0 contour for each utterance type. Final boundary tone, nuclear pitch accent, F0 offset, F0 onset, F0 range, and the slopes of a topline and a bottomline were determined for each utterance and compared for the four utterance types. Results show that for an average speaker, the final boundary tone, the F0 range, and the slope of the topline can be used to distinguish between the four utterance types. However, speakers may deviate from this pattern and exploit other intonational means to distinguish certain utterance types or choose not to mark a syntactic difference at all.

1. INTRODUCTION

In speech synthesis the realisation of an appropriate F0 contour is one of the most important tasks to solve. In a text-to-speech system only the text is available to determine the prosodic properties of the utterance to be synthesised. Often the intonational modelling of speech synthesis systems is based on averaged data across various speakers. Then specific speaker characteristics and speaking strategies tend to disappear. Therefore we also need to see whether all speakers fit the average data or whether we can identify different alternatives of tonal realisation.

Haan et al. [1] examined the tonal properties of Dutch declaratives and three types of interrogatives (wh-questions, yes/no-questions, and declarative questions) with two accents per utterance. While overall F0 span (cf. [2]) of accents (without final rise) is constant for all utterance types, the slopes of the upper and lower all points regression lines (without final rise) differ significantly. F0 offset (= last value before final rise), F0 maxima (without final rise) are also significantly different for all utterance types. F0 minima distinguish three groups of utterance types: 1. statements, 2. wh-questions, 3. declarative questions and yes/no questions. For wh-questions moreover a narrowed register was found. F0 onset distinguished two groups: high onset in wh-questions and yes/no-questions vs. low onset in statements and declarative questions. The results suggest that the slope of the declination lines is an important feature. These findings are in line with the results of Gooskens & van Heuven [3]. They found that in Dutch and Danish sentences the slope of declination depends on utterance type. In Danish questions there is no declination, whereas it is quite steep in statements.

In his model for synthesising German intonation Kohler [4] treats the global downward trend as downstep of subsequent accents, rejecting a time dependent model. Downstep can be suppressed by reinforcement or it can be reset. Möhler [5] and Adriaens [6] use declination lines to model intonation in speech synthesis. Whereas Möhler allows the topline and the baseline to decline in different but fixed slopes, Adriaens calculates the upper three declination lines relative to the baseline.

For German, Oppenrieder [7] investigated the declination properties of German declaratives, two types of interrogatives (yes/no and declarative questions), and commands in sentences with two accents. He takes into account F0 onset, F0 offset and local minima and maxima of accented words. He also calculates regression lines through the maxima and minima of the whole utterance (unlike Haan, who leaves out the final rise). The F0 offsets and the maxima and minima belong to the same regression line, which means that almost all differences he found can be explained by high ending questions and low ending statements. But his finding that declarative questions and yes/no-questions have a rising minima line is not based on the F0 offset. Some of his speakers deviated from this pattern, which points towards different speaker strategies. Oppenrieder concludes that it is possible to distinguish questions and statements by means of declination, but not different question types.

In our study we investigate the tonal properties of declaratives and three types of interrogatives. The analysis comprises nuclear pitch accent, final boundary tone, height of F0 onset and F0 offset, overall F0 range, and some parameters of upper and lower regression lines. Besides an analysis of an "average speaker", we try to identify alternate speaker specific strategies.

2. METHOD

Two female (SJ and JH) and two male (KW and MK) native Northern German speakers each performed two scripted dialogues[1] with the first author. The four speakers' role in the dialogues contained 27 statements and 69 questions. The questions were further distinguished by syntactic properties into the following utterance types: 23 wh-questions (marked by wh-word), 23 yes/no-questions (marked by verb inversion), and 23 declarative questions (syntactically identical to statements). The utterances contained between three and six possible positions for pitch accents.[2] In order to minimise microprosodic effects as far as possible, within each utterance the vowels of accentable syllables were either only [i: , y: , u:] or only [e: , ø:] (cf. [8]). In addition, the pitch-accentable syllables consisted only of sonorants and voiced fricatives. In order to minimise effects of syntactic structure, complex noun phrases were avoided, and the adjunct phrases in each utterance were attached at sentence level.

From this corpus of 96 utterances for each speaker, those utterances were selected which were both realised as one intonation phrase and had no contrastive focus. This resulted in a final corpus of 95 statements, 76 wh-questions, 87 yes/no-questions, and 89 declarative questions altogether. This final corpus was transcribed with five of the original six GToBI pitch accents (L*, H*, L*+H, L+H*, and H+L*) and all GToBI boundary tones (L-L%, L-H%, H-L%, H-H%) [9].

First, the F0 values of the targets corresponding to H and L tones within pitch accents were measured on a semitone scale[3]: for each H tone the corresponding maximal F0 value, and for each L tone the corresponding minimal F0 value was recorded. The targets for leading and trailing tones of complex pitch accents were located in the vicinity of the accented syllables. In cases of an H tone before an H- boundary tone, it was sometimes not possible to distinguish the high target for the pitch accent from the H- peak. Unless the target of the nuclear pitch accent was clearly identifiable, the F0 value for that H tone was not recorded.

We also measured the F0 value at the beginning of each utterance ('F0 onset') and the F0 value at the final boundary ('F0 offset'). Where glottalisation prevented a reliable F0 measurement, we recorded a value as close to the boundary as possible. Finally, the total speaking time for each utterance was measured.

Second, for each utterance a linear F0-with-time regression line was fitted through the F0 values corresponding to H tones of pitch accents.[4] A separate regression line was also calculated for the L tones. The slope coefficient for each line was then multiplied with the total speaking time of the utterance, yielding a measurement for the topline slope and the bottomline slope in semitones per utterance[5]. In addition, the regression lines were described by their y-axis interception, i.e. their starting value in semitones.

Third, the F0 range for the pitch accents in each utterance was calculated by subtracting the minimal F0 value from the maximal F0 value within the utterance. F0 values for boundary tone targets were not considered here, just as in the calculation of the regression lines.

The GToBI transcriptions, the F0 measurements, and the calculations explained above gave us the following parameters to characterise each utterance:

- final boundary tone and nuclear pitch accent
- F0 offset and F0 onset
- F0 range
- topline and bottomline slope and starting value.

These parameters were then subjected to analyses of variance with 'utterance type' as independent variable.

3. RESULTS

3.1 Final boundary tone and nuclear pitch accent

Across all four speakers, the final boundary tone distinguished statements and wh-questions from yes/no-questions and declarative questions. Whereas the former always ended on the low boundary tone L-L%, the latter generally ended on a high boundary tone.

	low	high		
	L-L%	**L-H%**	**H-L%**	**H-H%**
state	100	-	-	-
wh-q	100	-	-	-
y/n-q	-	-	7	93
decl-q	2	2	8	88

Table 1: Percentage of final boundary tones per utterance type

The nuclear accent in statements was predominantly H*, whereas in wh-questions L+H* was preferred. Together with the low final boundary tones, this resulted in a falling contour at the end of the utterance (with an expanded movement for wh-questions).

	H+L*	**L***	**L*+H**	**L+H***	**H***
state	7	5	-	28	60
wh-q	6	3	8	64	19
y/n-q	-	16	74	10	-
decl-q	-	17	67	16	-

Table 2: Percentage of nuclear pitch accents per utterance type

[1] The complete dialogues as well as realisations of selected utterances will be available at the time of the conference via http://www.coli.uni-sb.de/~cabr/Eurospeech99/

[2] Number of accents and distance between them were not considered in this study.

[3] Since semitones can only be used to compare two F0 values, we arbitrarily chose 1 Hz as reference value. ERB values were also calculated and led to the same results.

[4] Since at least two values are needed to calculate a (regression) line, those cases with only one F0 value present were excluded from these calculations.

[5] Semitones per second were also calculated and led to the same results.

Yes/no-questions as well as declarative questions were mostly realised with L* or L*+H as nuclear pitch accent, resulting in a rising contour.

3.2 F0 offset and F0 onset

One might expect that a comparison of the F0 offsets will show the same significant differences as the final boundary tones. This expectation was confirmed. Moreover, F0 offset of statements is lower than that of wh-questions ($p \leq 0.001$). Carrying out an ANOVA for each speaker's data separately, however, we found that this significance holds for only two of them (JH and KW).

	all	SJ	JH	KW	MK
state	84.2	89.2	89.6	79.9	78.9
wh-q	85.4	89.6	91.5	81.8	78.9
y/n-q	97.5	99.6	103.6	93.7	93.2
decl-q	97.7	98.0	104.7	94.5	93.3

Table 3: Mean F0 offset in semitones for the whole corpus ('all') and for each speaker separately

Comparing the F0 onset, we found a significant difference between declarative questions and all other utterance types: the mean F0 value measured at the beginning of declarative questions was significantly lower than the one measured at the beginning of the other utterance types ($p<0.05$). The mean values of both F0 onset and F0 offset are depicted with separate dots for each utterance type in Figure 1.

3.3 F0 range

The F0 range of the pitch accent targets in semitones grouped statements together with yes/no-questions (smaller F0 range) and wh-questions together with declarative questions (larger F0 range, $p<0.05$), the declarative questions having the largest F0 range. This general tendency could be confirmed for each speaker. However, the difference between statements and wh-questions was not significant for JH, and MK's wh-questions had a larger F0 range than his declarative questions.

	all	SJ	JH	KW	MK
state	9.0	5.8	9.5	9.8	10.6
wh-q	10.1	7.2	11.0	10.5	11.5
y/n-q	8.9	6.8	9.2	9.2	10.1
decl-q	10.5	8.3	11.9	10.7	10.8

Table 4: Mean F0 range in semitones for the whole corpus ('all') and for each speaker separately

3.4 Topline and bottomline

Considering all the data, the topline declined for all utterances. There was a significant difference between statements and questions, the latter having shallower declination ($p<0.01$). Within the questions, the wh-questions had the least steeply declining topline, but the difference was not significant due to large variation. Therefore, the topline slope could be used to distinguish between statements and wh-questions, which were found to have the same final boundary tone.

Analysing each speaker separately, exactly those topline slope characteristics could be found only for one speaker (SJ). For the other female speaker (JH), we found a significant difference between statements and wh-questions, where the topline even showed inclination ($p<0.05$); but the other two question types patterned with the statements and not with the wh-questions. For the two male speakers (KW and MK) there was no significant difference between the topline slopes of statements and wh-questions. For MK statements differed significantly only from yes/no-questions and declarative questions ($p<0.05$). For KW no significant difference regarding the topline slope could be found at all.

	all	SJ	JH	KW	MK
state	-3.62	-4.07	-0.52	-4.35	-5.30
wh-q	-0.52	-0.21	3.20	-2.09	-2.99
y/n-q	-1.00	-1.59	-0.67	-0.95	-0.89
decl-q	-1.08	-1.40	-2.23	-1.08	0.37

Table 5: Means of topline slope in semitones per utterance.

For the bottomline slope of the whole corpus, we found that the bottomline of yes/no-questions declines more steeply than the bottomline of all other utterance types ($p<0.05$). This tendency is also true for each separate speaker, but no difference was found to be significant here, except for SJ's yes/no-questions and wh-questions ($p<0.05$).

	all	SJ	JH	KW	MK
state	-1.4	-2.2	-1.2	-0.8	-1.5
wh-q	-1.6	-0.8	-1.5	-2.6	-1.3
y/n-q	-2.7	-2.6	-2.8	-2.7	-2.9
decl-q	-1.5	-1.4	-1.7	-1.8	-1.3

Table 6: Means of bottomline slope in semitones per utterance

Besides the mean slope coefficients, we determined the mean starting value for the topline and for the bottomline. The starting value for the topline was found to be higher for statements and declarative questions than for wh-questions and yes/no-questions ($p<0.05$, except for the difference between declarative questions and wh-questions, which was not significant). The starting value for the bottomline was lower for declarative questions than for all other utterance types ($p<0.05$).

Figure 1 sums up the effects of 'utterance type' on the course of topline and bottomline, the F0 onset and F0 offset, and (implicitly) the F0 range. For statements, topline and bottomline typically converge over the course of the utterance. This contrasts with all question types, which have either diverging (yes/no-questions more than wh-questions) or almost parallel top- and bottomlines (declarative questions). It also shows that the F0 onset is typically 1 semitone above the

starting value of the bottomline.

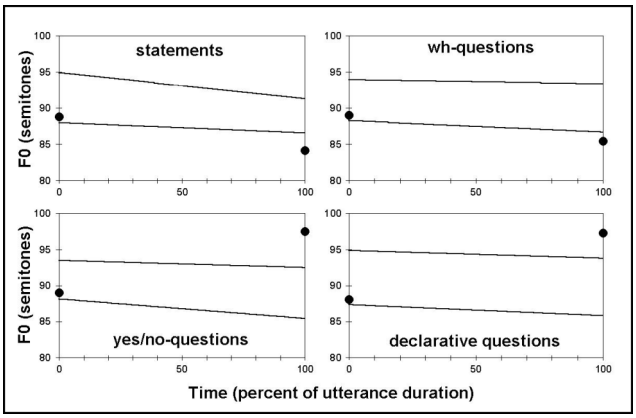

Figure 1: Topline, bottomline, F0 onset and F0 offset for the four different utterance types (means for all speakers)

4. DISCUSSION

Compared with the findings for Dutch [1], our results for German show that there is no narrowed register for wh-questions, there are clear differences in F0 range for different utterance types, and there are no inclining regression lines. Furthermore, we could not confirm Oppenrieder's finding of a rising bottomline for yes/no-questions.

Whereas Gooskens & van Heuven [3] showed that declination is absent in Danish questions, it is not clear from our data whether the shallower topline slope of interrogatives is caused by the absence of downstep. The results for topline slope support such an interpretation. On the other hand the fact that yes/no-questions have the steepest slope of the bottomline points in the opposite direction. Additional research is needed to clarify this issue.

The significant differences we found for the whole corpus between the four utterance types statement, wh-question, yes/no-question, and declarative question, are summarised in Table 7. For an average speaker, it is possible to distinguish intonationally between the four utterance types by means of final boundary tone and F0 range. However, different speakers were shown to deviate from this pattern. For speaker JH, there was no significant difference in F0 range between statements and wh-questions, but topline slope distinguished wh-questions from all other utterance types. For speaker SJ, topline slope distinguishes statements from all question types. Therefore, for both female speakers all four utterance types can be distinguished by topline slope, final boundary tone and F0 range. For the two male speakers (KW and MK), topline slope failed to distinguish statements and wh-questions. For speaker KW, offset height could be used as distinguishing factor instead. This means that some speakers use other or additional cues for signalling the utterance type compared to the "average speaker". But for speaker MK according to our measures there are no significant differences between statements and questions at all.

	state	**wh-q**	**y/n-q**	**decl-q**
final boundary tone	low	low	high	high
nuclear pitch accent	(L+)H*	(L+)H*	L*(+H)	L*(+H)
F0 offset	lowest	low	high	high
F0 onset	high	high	high	low
F0 range	small	large	small	large
topline slope	steep	shallow	shallow	shallow
bottomline slope	shallow	shallow	steep	shallow
register shape	converg	diverg	diverg	diverg

Table 7: Summary of significant differences (average speaker)

The results of our study also suggest that considering the differences in the toplines and in F0 range across the different utterance types might contribute towards the naturalness of synthetic speech. None of the intonation modules cited in the introduction account for diverging declination lines. And even a more flexible model is needed to account for speaker characteristics.

REFERENCES

[1] Haan, J., van Heuven V., Pacilly, J. & van Bezooijen, R. (1997): Intonational Characteristics of Declarativity and Interrogativity in Dutch: A Comparison. *Proc. ESCA Workshop on Intonation: Theory, Models and Applications*, Athens, pp. 173-176.

[2] Ladd, D. R. (1996): *Intonational Phonology*. Cambridge, Cambridge University Press.

[3] Gooskens, C. & van Heuven, V. (1995): Declination in Dutch and Danish: Global versus local pitch movements in the perceptual characterisation of sentence types. In: *Proc ICPhS 95*, Stockholm, Vol.2 pp. 374-377.

[4] Kohler, K. (1996): Parametric control of prosodic varables by symbolic input in TtS synthesis. In: van Santen et al. (eds.) *Progress in speech synthesis*, pp. 459-475.

[5] Möhler, G. (1998): *Theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese*. PhD, Stuttgart.

[6] Adriaens, L. (1991): *Ein Modell deutscher Intonation*. PhD. Eindhoven.

[7] Oppenrieder, W. (1989): Deklination und Satzmodus. In: Altmann et al. (eds.) *Zur Intonation von Modus und Fokus im Deutschen*. Tübingen, Niemeyer, pp. 245-266.

[8] Möbius, B., Zimmermann, A. & Hess, W. (1987): Untersuchungen zu mikroprosodischen Grundfrequenzvariationen im Deutschen. In: Tillmann, H.G. & Willée, G.: *Analyse und Synthese gesprochener Sprache*. Hildesheim, Zürich, New York, Olms, pp. 102-110.

[9] Grice, M., Reyelt, M., Benzmüller, R., Mayer, J. & Batliner, A. (1996): Consistency in Transcription and Labelling of German Intonation with GToBI. *Proc Fourth International Conference on Spoken Language Processing*, Philadelphia, pp. 1716-1719.