INTERSPEECH 2009 BRIGHTON

Compression and Truncation Revisited

Claudia K. Ohl *Hartmut R. Pfitzinger*

Institute of Phonetics and Digital Speech Processing (IPDS)
Christian-Albrechts-University at Kiel, Germany

ohl@ipds.uni-kiel.de hpt@ipds.uni-kiel.de

Abstract

This paper investigates the influence of varying segmental structures on the realizations of utterance-final rising and falling intonation contours. Following Grabe's study on adjustment strategies in German, i.e. truncation and compression, a similar experiment was carried out, using materials with decreasing stretches of voicing in questions, lists, and statements. However, the results presented in the present paper could not confirm the idea of such common adjustment strategies. Instead, considerable variation was found as to how the phrase-final intonation contours were adjusted to the respective amounts of voicing: the strategies varied strongly across different word groups.

Index Terms: prosody, pitch, sentence mode, adjustment strategies, compression, truncation

10.21437/Interspeech.2009-312

1. Introduction

Truncation and compression are assumed to be two basic mechanisms of how intonation contours are modified when the segmental structure contains decreasing scope for voicing either because of only few voiced phones or high speech rate.

How do intonation contours adapt to segmental strings of decreasing duration? By compressing the melodic patterns, by truncating them or by a combination of both mechanisms?

According to von Essen 1964 [14, p. 22] neither extending nor shortening the segmental chain cause a change to the underlying melody of an utterance, so that the duration of voiced segments has no influence on the perception of intonation [14, p. 22]. Similarly, Bierwisch stated that if the F0 movement has to be carried out within a single syllable (e.g. <*Mann*> [man]), or even a single accentuated vowel (e.g. <*Fisch*> [fɪʃ]), then the resulting contour is simply "an automatic phonetic consequence that can not and may not be included in the representation of intonation" [2, p. 142]. Also, prosodic features are generally said to have great inter- and intra-speaker variations [10], and therefore, stronger variation than that caused by truncation or compression mechanisms must be expected.

The first to mention *truncation* and *rate adjustment* as common adjustment strategies of intonation contours to varying voicing durations were Erikson & Alstermark 1972 [3] in their study of Swedish. Bannert & Bredvad-Jensen 1975 [1] later introduced the term *compression* for *rate adjustment* in their study on dialect-based accent realizations in Swedish. Gartenberg & Panzlaff-Reuter 1991 [4] investigated intonation contours with short stretches of voicing in German utterances (e.g. *Sie strickt*, engl. *She knits*). Hanssen, Peters & Gussenhoven 2007 [7] analyzed short Dutch phrase-final F0 contours and found a contour-dependent choice of truncation, temporal compression, and range compression. Grabe 1998 [5, 6] concludes in her contrastive study on English and German that English is a compression language whereas German compresses rises but truncates falls (for more details, see Sec. 2.1 as well as Sec. 4). Niebuhr & Ambrazaitis 2006 [11] found a strong correlation between the duration of slope movements and segmental chains, concluding that the slope's shape is determined by the syllabic structure, and also varies from speaker to speaker.

Peters & Pfitzinger 2008 [12] conducted a study on the perception of phrase-final intonation contours (in the utterance *"Ulf stickt"*), and in particular the influence of voicing duration and F0 interval variations on the perception of interrogative and declarative utterances. The authors found that a voicing duration of 50 ms is long enough to carry communicatively relevant intonation contours: F0 rises of two or more semitones lead to questions while F0 falls of one or more semitones lead to statements. Isačenko & Schädlich 1966 [8, p. 58] showed that even an F0 interval of one semitone is sufficient to yield a perceivable prominence of a syllable, and assumed that the actual F0 interval is used by the speaker for "the expressive and emotional connotations which can be conveyed by intonation." [9, p. 57]

The present study attempts to provide an answer to the question: How do phonetic realizations of phonological intonation categories in phrase-final words with only little scope for voicing differ from realizations in words that are fully or mostly voiced? Important parameters in the analyses of the data are *duration*, *F0 interval*, and *F0 slope*.

2. Method

2.1. Materials

Following Grabe's studies [5, 6], materials were gained from two different texts: a modified version of "Little Red Riding Hood" and a fictional dialogue about a lottery winner. The target words, sorted into three word groups, were:

(1)	<*Mann*>	<*Nachbar*>	
(2)	<*Schiff*>	<*Schief*>	<*Schiefer*>
(3)	<*Fisch*>	<*Fleisch*>	<*Fleischwurst*>

They were embedded into the stories in phrase-final positions and occurred in statements, questions, and lists (only word group 3 occurred in lists). Since each word had to be read in every position once, a full set included three versions of each text. In addition, the speakers were asked to repeat the sets five times and to read in a normal speaking style in order to avoid a type of reading that is used when a fairy tale is read to a child (e.g. exaggerated intonation, overly precise pronunciation).

Although the texts were derived from Grabe [5, 6], slight changes were conducted. Since the "Little Red Riding Hood" version initially used by Grabe was also designed to establish an intonation inventory of Northern Standard German, it was simply too large for the purpose of the present study and therefore was reduced to those sentences containing the target words

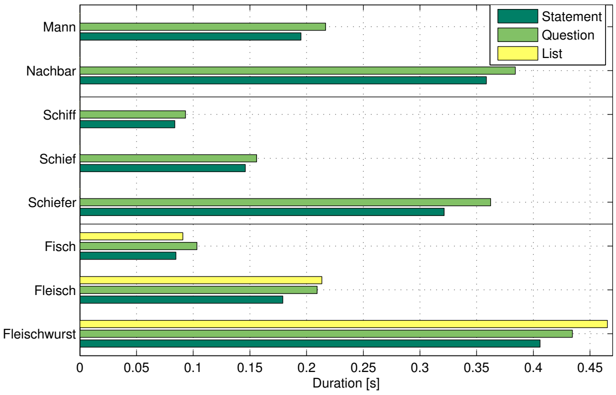

Figure 1: *Mean durations of all target words of all speakers.*

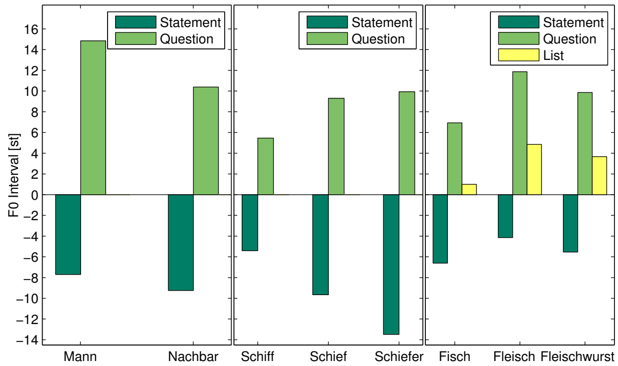

Figure 2: *Mean F0 intervals in all target words for all speakers.*

mentioned above. An introductory and a final sentence were added as well as a phrase containing a question variant, leading to a full set of sentence types (statement, question, list).

2.2. Subjects

Ten subjects from northern Germany participated in the experiment, five female and five male. All were students at the Christian-Albrechts-University Kiel and aged between 23–36.

2.3. Analyses

The recordings were made at the Institute of Phonetics and Digital Speech Processing (IPDS) in a highly sound-absorbent booth using a *Microtech Gefell M-940* large membrane condenser microphone and an *RME Fireface 800* soundcard at 24 bit amplitude resolution and 32 kHz sampling frequency.

A total number of 1650 target words was recorded (750 statement items, 750 question items, 150 list items). Acoustic parameters analyzed were duration and fundamental frequency. Phone and word segmentation was carried out manually. F0 contours were extracted by means of ESPS *get_f0* [13] in 10 ms steps, transformed to semitones in order to facilitate the comparison of female and male speakers, and then stylized by first-order polynomial regression which turned out to be more adequate than e.g. median filtering of the first and last five F0 values of a contour. All data processing were done with *Matlab* followed by statistical analyses carried out using *SPSS*. Independent variables were: *sentence type* (*ST*: statement, question, list), *gender* (male, female), *speaker*, *word*, and *repetition*. Dependent variables were: *F0 beginning*, *F0 end*, and *F0 interval*.

3. Results

3.1. Duration

Fig. 1 shows the duration differences of the words' voiced stretches within the corresponding word groups: <*Schiff*>, <*Fisch*>, and <*Mann*> are in all cases shorter than the other longer words in their word groups. <*Schief*> and <*Fleisch*> are always longer than <*Schiff*> and <*Fisch*> and at the same time shorter than <*Schiefer*>, <*Fleischwurst*>, and <*Nachbar*>.

Relation	S<Q	S>Q	S=Q	Q<L	Q>L	Q=L	S<L	S>L	S=L
Female	28	11	1	8	7	0	11	2	2
Male	36	3	1	8	6	1	13	2	1

Table 1: *Frequency of the type of duration differences between words in statements (S), questions (Q) and lists (L), grouped according to gender. ">" stands for "longer than", "<" for "shorter than", and "=" for "equal to".*

A comparison of sentence types (see Fig. 1) reveals that the mean duration of words in questions and lists is longer than in statements, i.e. duration increases when contours rise, but decreases when they fall. Among the rising contours, questions and lists, the duration differences vary.

Table 1 presents the frequencies of the type of duration differences between statements, questions, and lists, e.g., if a statement or a question is longer more frequently. Grouping the speakers according to gender reveals that in the case of statement-question-relations, the realizations of male speakers were more consistent than those of female speakers. The larger amount of statement-question-relations (40) can be accounted for by sentence type: every word occurs as a question and a statement (5 female or male subjects × 8 words), whereas comparisons for list items (15) could only be made for word group 3 (5 female or male subjects × 3 words).

3.2. F0 interval

F0 interval is the difference between the lowest and the highest reliably measurable F0 values within a word. Since the F0 contours in the target words used in this study turned out to never consist of complex tones, first-order polynomial regression was sufficient to reliably extract F0 beginning and F0 end values as well as their interval. Fig. 2 shows the F0 intervals averaged over all speakers. Table 2 shows the absolute F0 intervals in order to allow for comparing the effect of the sentence types, irrespective of changes into the negative or positive.

Obviously, the three word groups adjust differently to different word durations. While in the case of word group 1 (<*Mann*> vs. <*Nachbar*>) absolute F0 interval increases for statements and decreases for questions (see Table 2), word group 2 (<*Schiff*>, <*Schief*>, <*Schiefer*>) acts differently: with increasing word duration, F0 interval increases for ques-

Word:	Mann	Nachbar	Schiff	Schief	Schiefer
Statement:	8	9	5	10	14
Question:	15	10	6	9	10

Word:	Fisch	Fleisch	Fleischwurst
Statement:	7	4	6
Question:	7	12	10
List:	1	5	4

Table 2: *Mean F0 interval magnitudes in all target words according to sentence type.*

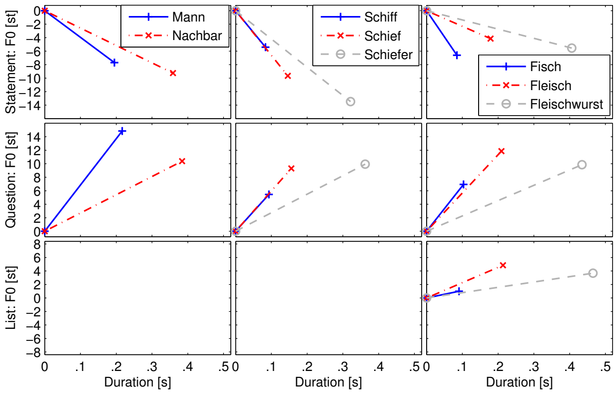

Figure 3: *Mean F0 contours in all target words for all speakers.*

tions and statements. However, word group 3 (<*Fisch*>, <*Fleisch*>, <*Fleischwurst*>) displays yet a different scheme: the shortest word has the lowest F0 interval, while, in the case of rises (questions and lists) the greatest F0 interval can be found with the word <*Fleisch*>; in between lies the longest word of this group. For the statement variants, <*Fleisch*> has the lowest F0 interval, the highest occurs in the shortest word <*Fisch*>, and the longest word lies between the two.

For the parameter *F0 interval* a univariate ANOVA with the factors *sentence type* (*here:* statement and question), *word* and *gender* was carried out. The results are given in Table 3. Of the two main effects, as expected *sentence type* has the greatest effect on F0 interval with an explained variance of 53%. The other effect, *word* together with *ST*word*, explains 7% of the variance clearly indicating a surprisingly strong word effect.

In order to include all sentence types in the statistics, another ANOVA was carried out, using the same factors as before, but additionally containing the list items and by that reducing the factor *word* to word group 3. The results are presented in Table 4. Again, *sentence type* has the greatest effect on F0 interval accounting for 69% of the explained variance. In this case *word* accounts for 5% of the observed variance. In addition, there is an *ST*gender* interaction explaining 2% of the variance.

F0 interval	df	*F*	p	expl. variance
ST	1	4571.900	<0.001	53.29%
Word	7	682.444	<0.001	2.98%
Gender	1	41.724	0.135	0.03%
ST*Word	7	894.917	<0.001	3.91%
ST*Gender	1	1093.700	<0.001	0.68%
Word*Gender	7	18.989	0.418	0.08%
ST*Word*Gender	7	221.848	<0.001	0.97%

Table 3: *ANOVA of* F0 interval *with the factors* sentence type *(*ST *here: statement, question),* word *(here: all), and* gender.

F0 interval	df	*F*	p	expl. variance
ST	2	551.178	<0.001	68.90%
Word	2	37.449	<0.001	4.68%
Gender	1	5.044	0.025	0.32%
ST*Word	4	2.475	0.044	0.62%
ST*Gender	2	15.158	<0.001	1.89%
Word*Gender	2	2.306	0.101	0.29%
ST*Word*Gender	4	4.236	0.002	1.06%

Table 4: *ANOVA of* F0 interval *with the factors* sentence type *(here: statement, question, list),* word *(here: <Fisch>, <Fleisch>, <Fleischwurst>), and* gender.

3.3. F0 slope, F0 beginning, and F0 end

F0 slopes, i.e. the degree of rising or falling of the fundamental frequency, are displayed in Fig. 3. The panels show the effect of sentence type and its interaction with different words on the stylized F0 contour, combining measured and averaged voicing durations and F0 intervals. The initial value for every word in every position was set to zero to allow for a visual comparison.

In the word groups 1 (<*Mann*>, <*Nachbar*>) and 3 (<*Fisch*>, <*Fleisch*>, <*Fleischwurst*>) all words in statements and questions have steeper F0 slopes if they have shorter voicing durations. However, in word group 2, the F0 contour of the shortest word <*Schiff*> is achieved by exact truncation of the F0 contour of the word <*Schief*> and retains its F0 slope, whereas in word group 3, and only in the case of statements, the F0 contour of the shortest word <*Fisch*> is adjusted by increasing the steepness of the F0 slope and ending on an even lower F0 level than in <*Fleisch*> and <*Fleischwurst*>. Furthermore, the F0 slopes of the words produced in list context show a rather inhomogeneous scheme. Obviously, Fig. 3 appears to suggest word-dependent choices of adjustment strategies in statements, questions, and lists.

For the parameters F0 beginning and F0 end two univariate ANOVAs with the factors *sentence type* (*here:* statement and question), *word* and *test subject* were carried out. The results are given in Table 5. At F0 beginning all factors as well as their interactions were significant. As expected, the factor *test subject* accounted for the greatest amount of variance, i.e. 35%. Remarkably, 14% of the variance could be explained by *sentence type* indicating that subjects also adjust the start point of F0 according to sentence type. For F0 end also all factors as well as their interactions were significant. The factor *sentence type* could explain 31% of the observed variance and *test subject* explained another 24%. A univariate ANOVA including all three sentence types yielded similar results.

4. Discussion

What are the consequences regarding the underlying question of how phonetic realizations in words with little scope for voicing differ from those with greater amounts of voicing?

It is self-evident that realizations of intonation contours in utterances with only little scope for voicing will include

F0 beginning	df	*F*	p	expl. variance
ST	1	2230.124	<0.001	13.79%
Word	7	226.587	<0.001	9.81%
Subj	9	625.922	<0.001	34.84%
ST*Word	7	49.762	<0.001	2.15%
ST*Subj	9	42.045	<0.001	2.34%
Word*Subj	63	3.669	<0.001	1.43%
ST*Word*Subj	63	4.747	<0.001	1.85%

F0 end	df	*F*	p	expl. variance
ST	1	4511.281	<0.001	30.84%
Word	7	49.999	<0.001	2.39%
Subj	9	395.301	<0.001	24.32%
ST*Word	7	64.141	<0.001	3.07%
ST*Subj	9	26.147	<0.001	1.61%
Word*Subj	63	4.004	<0.001	1.72%
ST*Word*Subj	63	4.982	<0.001	2.15%

Table 5: *ANOVA of* F0 beginning *(top) and* F0 end *(bottom) with the factors* sentence type *(*ST*, here: statement, question),* word *(here: all), and* test subject.

some form of temporal adjustment of the voiced segments (e.g. stretching), range adjustment or shortening of the F0 contour. However, adjustment strategies can not be generalized by rules, such as those suggested by the idea of truncation and compression, stating that the F0 contour will be cut off (i.e. truncated) in the case of falling contours, while in the case of rises the contour will naturally be compressed (see Fig. 3).

The results from the present study showed that speakers generally stretched the voiced segments when they produced rising contours on words with little scope for voicing. The data also suggests some influence of speaker-dependent choices concerning the adjustment strategy, as well as the degree to which a particular strategy is used or combined. However, another important and more obvious factor, presented in this paper is the apparent influence of the words themselves on the strategy chosen by the speakers. With regard to Grabe's results, this means that speaker-specific realization preferences can not be ruled out and must at least be taken into account and further analyzed as a possible source of variation.

Furthermore, speakers appeared to vary their choice of adjustment strategy according to word groups (see Fig. 3). A clear example is the difference between the realizations of <*Schiff*> and <*Schief*> in comparison to those of <*Fisch*> and <*Fleisch*> in statements and questions: in word group 2 (<*Schiff*>, <*Schief*>, <*Schiefer*>), the contour's duration is shortened and F0 interval is reduced, whereas in word group 3 (<*Fisch*>, <*Fleisch*>, <*Fleischwurst*>), the F0 slope of the shortest word <*Fisch*> is adjusted by ending on a lower F0 than in <*Fleisch*> and <*Fleischwurst*>. This suggests that perhaps the words as a whole influence the strategy for the realization of a particular contour (*here:* rising or falling). Why this is the case, is a matter that requires further research.

Grabe 1998 [5, 6] as well as Hanssen et al. 2007 [7] use F0 slope, i.e. F0 interval divided by voicing duration, as a correlate of truncation and compression. If the absolute F0 slope increases while the voicing duration decreases, compression is assumed. Truncation on the other hand is only assumed, if the absolute F0 slope either remains constant or decreases with decreasing voicing duration. In order to compare the results from the present study to those of Grabe's study of adjustment strategies in Northern Standard German [5, 6], compare Fig. 3 and Fig. 4. It is obvious that the new results from the current study yield a completely different figure. While Fig. 4 displays a very systematic relation between the words and suggests clear differences concerning the choice of adjustment strategy, confirming the concept of truncation and compression, Fig. 3 can not serve to support these results as was already indicated: no systematic strategy of contour modifications could be found.

In summary, the data shows that the choice of adjustment strategy is not a characteristic of an entire language system (*here:* German) or even just a variety like Northern Standard German. While the slopes of the words showed intra-speaker consistency, the choice of strategy as well as the degree to which a particular strategy is used or combined depends mostly on the word itself.

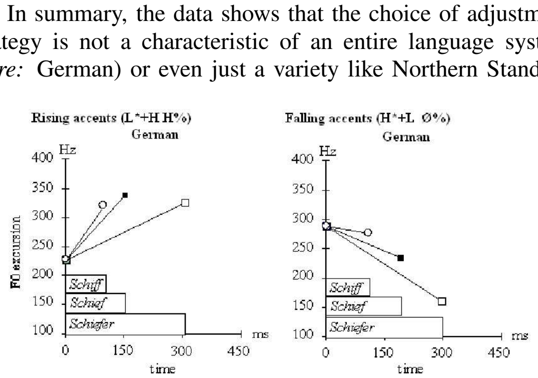

Figure 4: *Results of F0 slope for rising (left) and falling contours (right), taken from Grabe 1998 [5, ch. 5.3.2].*

5. Conclusion

The results from the present paper already indicate that the existence of two strategies used in German intonation is not as certain as it seemed. The data showed no common adjustment strategies shared by all speakers. Rather, modifications of the F0 contour varied to some degree from speaker to speaker, but even far stronger across the different word groups. The data suggested that adjustment strategies were not consistent across different word groups, but varied greatly. Due to word-dependent choices of adjustment methods of F0 contours on the one hand, and some inter-speaker variation on the other hand, it is difficult to identify clear word- and speaker-consistent strategies. Consequently, a study with more test subjects and more variable materials would be required to provide clarification on the influence of the words themselves. Perhaps even comparisons of different dialect areas would be needed, since relevance of that parameter was already indicated by Bannert & Bredvad-Jensen 1975 [1].

6. References

[1] Bannert, R.; Bredvad-Jensen, A.-C. 1975. Temporal organization of Swedish tonal accents: The effect of vowel duration. Working Papers 10, pp. 1–36, Phonetics Laboratory, Department of General Linguistics, Lund University, Lund.

[2] Bierwisch, M. 1966. Untersuchungen über Akzent und Intonation im Deutschen. In Bierwisch, M., ed., *Studia Grammatica VII*, pp. 99–201. Akademie-Verlag, Berlin.

[3] Erikson, Y.; Alstermark, M. 1972. Fundamental frequency correlates of the grave word accent in Swedish: the effect of vowel duration. Speech Transmission Lab., Quarterly Progress and Status Report 2–3, pp. 53–60, KTH, Stockholm.

[4] Gartenberg, R.; Panzlaff-Reuter, C. 1991. Production and perception of F0 peak patterns in German. Arbeitsberichte (AIPUK), pp. 29–113, IPDS, Christian-Albrechts-University, Kiel.

[5] Grabe, E. 1998. *Comparative intonational phonology: English and German*. Ph.D. thesis, Nijmegen University, Ponsen & Looijen, Wageningen.

[6] Grabe, E. 1998. Pitch accent realization in English and German. *J. of Phonetics*, 26: 129–143.

[7] Hanssen, J.; Peters, J.; Gussenhoven, C. 2007. Phrase-final pitch accommodation effects in Dutch. In *Proc. ICPhS*, vol. 2, pp. 1077–1080, Saarbrücken.

[8] Isačenko, A. V.; Schädlich, H.-J. 1966. Untersuchungen über die deutsche Satzintonation. In Bierwisch, M., ed., *Studia Grammatica VII*, pp. 7–67. Akademie-Verlag, Berlin.

[9] Isačenko, A. V.; Schädlich, H.-J. 1970. *A model of standard German intonation*. Mouton, The Hague.

[10] Ladd, D. R. 1996. *Intonational phonology*. Cambridge University Press, Cambridge.

[11] Niebuhr, O.; Ambrazaitis, G. 2006. Alignment of medial and late peaks in German spontaneous speech. In *Proc. of the 3rd Int. Conf. on Speech Prosody*, pp. 161–164, Dresden.

[12] Peters, B.; Pfitzinger, H. R. 2008. Duration and F0 interval of utterance-final intonation contours in the perception of German sentence modality. In *Interspeech*, pp. 65–68, Brisbane.

[13] Talkin, D. 1995. A robust algorithm for pitch tracking (RAPT). In Kleijn, W. B.; Paliwal, K. K., eds., *Speech coding and synthesis*, chap. 14, pp. 495–518. Elsevier, New York.

[14] von Essen, O. 1964. *Grundzüge der hochdeutschen Satzintonation*. Henn, Ratlingen, Düsseldorf.