The nuclear accentual fall in the intonation of Standard German

Ralf Benzmüller and Martine Grice
Institute of Phonetics, FR 8.7, University of the Saarland

1. Introduction

In this paper we investigate the F0 contours of nuclear falls in standard German and explore the consequences for autosegmental-metrical accounts of German intonation. Where the nuclear syllable is a considerable distance from the end of the phrase, one can observe two major stages in the F0 descent: a sharp fall followed by a plateau which extends up to the phrase boundary. The beginning of the plateau, which we refer to as the elbow, has been found in informal observations to be at variable distances from the peak.

Féry (1993) and Uhmann (1991), in their analyses of German intonation, assume that the nuclear pitch accent is bitonal, involving a High and a Low tone (H*+L), followed by a L% boundary tone at the end of the intonation phrase. This analysis correctly predicts that the contour consists of a fall and a plateau, the plateau being a consequence of interpolation between the trailing accentual L tone and the L% boundary tone. However, following inter alia Arvaniti et al. (forthcoming), who argue that the F0 correlate of an unstarred tone in a pitch accent is expected to occur at a fixed distance from the peak corresponding to the starred tone, these accounts would not be able to capture the degree of variability informally observed in the alignment of the elbow.

Grice et al. (1996) on the other hand analyse German nuclear falls as comprising a monotonal H* pitch accent followed by a L tone which is independent of the pitch accent. This is an edge tone for the intermediate phrase (L-). They acknowledge that there is not a simple linear interpolation between the H* peak and the end of the intermediate phrase, as would be predicted by their analysis if L- were merely aligned with the phrase edge. They assume, as argued by Pierrehumbert (1980), Beckman and Pierrehumbert (1986) for English, that the L- already controls the F0 at some distance from the end of the phrase. However, they do not give any further indication of the exact domain controlled by the L- tone.

In the present study we investigate in an experimental setting whether the elbow in the F0 contour is aligned at a fixed distance from the H peak, either in absolute terms, measured in milliseconds, or in relative terms as the number of syllables after the nuclear syllable. We also investigate whether it occurs at the edge of a constituent such as the nuclear foot or nuclear word, the latter as argued for English by Pierrehumbert and Beckman (1988), based on evidence in Steele and Liberman (1987). Finally, we investigate whether the elbow is aligned with a lexically stressed syllable, in which case it would constitute a 'postnuclear' or 'postfocal' accent, in the sense of Ladd (1996) and Grice (1995) respectively.

Furthermore, since we are also not certain that the elbow is the point in the fall taken to be the correlate of a L tone, we examine the contours for other turning points which might have been taken to represent the L in the accounts where L is a trailing tone of the pitch accent.

Since we cannot be sure that the differences in the accounts of falls are not due to actual differences in the contours analysed rather than to the theoretical analyses employed, our study does not restrict itself to one utterance type only. We introduce differences in information structure (broad vs. contrastive focus, see Uhmann 1991) and illocutionary force (assertive vs. directive), both of which potentially affect the type of falling contour used.

2. Experiment 1

A preliminary inspection of a corpus consisting of 40 paragraphs read by two speakers showed considerable variation in the elbow of the falling nuclear contours produced. However, since interpreting the f0 trace was often made difficult by the presence of obstruents in the postnuclear stretch, and since some postnuclear stretches were merely one or two syllables long, it was difficult to make generalisations about the exact location of the elbow in relation to any one structural landmark in the texts. In the experiment described in this section a controlled corpus was designed in which the postnuclear stretch was long and contained as many sonorant segments as possible.

2.1 Method & Stimuli

In the test corpus, the following factors were controlled for:

- number of syllables following the nuclear syllable (2, 3, 4, 5)
- interstress interval, i.e. number of postnuclear syllables before the next lexical stress (0, 1, 2, 3)
- number of postnuclear syllables before a word boundary (0, 1, 2)

Table 1 shows the set of 8 sentences upon which the experiment was based. The predicted position of the nucleus is the syllable orthographically represented as WOH, WOHN or MOHN.

1. Die Schüler	sollen	die	WOH	nung	en	#	be	MA	len
2.			WOH	nung	en	#	-	MA	len
3.			WOH	nung	-	#	be	MA	len
4.			WOH	nung	-	#	-	MA	len
5.		den	WOHN	WA	gen	#	-	MA	len
6.		das	WOHN	mo	BIL	#	-	MA	len
7.		den	MOHN	-	-	#	-	MAH	len
8.	haben	den	MOHN	-	-	#	ge	MAH	len

Table 1: Basic sentence types. CAPITAL letters mark lexical stress; # indicates a word boundary. For translation see footnote [1]

[1] 1. The pupils ought to decorate the flats
2. The pupils ought to paint the flats
3. The pupils ought to decorate the flat
4. The pupils ought to paint the flat
5. The pupils ought to paint the caravan
6. The pupils ought to paint the camper
7. The pupils ought to mill the poppyseed

The eight basic sentences were used to elicit read versions of the following types of utterance:

(a) statements with broad focus

(b) statements with contrastive focus

(c) commands

The text for the broad and contrastive focus stimuli was identical to the basic sentence set. Slight modifications to the beginning of the sentences were necessary in the commands. They started with "Du wirst sofort ..." (You will ... immediately).

The resulting three sets of sentences were randomized and read twice by two female speakers of Standard German (GS and JM). This resulted in 48 tokens for each subject[2].

After recording, the sentences were digitised and analysed with ESPS xwaves. The f0 traces were manually parameterised, marking not only the peak and the elbow, but also a third point between the two, referred to as L1 below, which was observable in the majority of the contours.

H# the f0 maximum at the nuclear accent

L1 the point where the slope of the fall turns from convex to concave, usually well above the baseline

L2 the elbow, or point where the baseline level is reached, defined as first major deviation from a "regression line" working backwards from the lowest f0 at or near the end of the contour, ignoring microsegmental perturbations.

The above labels were aligned with the f0 contour without examination of the speech pressure waveform. Nuclear and postnuclear syllable boundaries were aligned to the speech waveform with the help of spectrograms. After labelling, values were extracted for the fundamental frequency at H#, L1, and L2 and the duration was calculated between the start of the nuclear syllable and the f0 peak (H#), between H# and L1, and between H# and L2. Additionally the alignment of H#, L1, and L2 with a particular nuclear or postnuclear syllable (S1 - S6) was calculated.

2.2 Results

The results for f0, timing and syllable counts are presented separately.

2.2.1 F0 Analysis

A MANOVA was performed with the utterance type and sentence type as independent variables and the frequency at H#, L1, and L2 as dependent variables. The values for each speaker were analysed separately.

The f0 values for H# differed significantly across the three utterance types ($F = 102.5$; $df = 2$; $p < 0.001$ for GS; $F = 127.1$; $df=2$; $p < 0.001$ for JM). For both speakers commands have the highest H peaks, followed by contrastive focus, followed by broad focus statements (see

8. The pupils have milled the poppyseed

2 Two stimuli were deleted by accident after recording. So the total number of stimuli is only 94.

figure 1 and table 2). Post hoc tests (Scheffé) showed that each utterance type differs significantly from all others ($p < 0.001$ for JM; $p = 0.002$ for GS). Thus commands, contrastive focus statements, and broad focus statements constitute distinct groups as far as f0 height is concerned.

There is less variation in the f0 values at L1 than in the H# values. However, despite the spread being small, significant differences were found ($F = 4.2$; df=2; $p < 0.03$ for GS; $F = 3.9$; df=2; $p < 0.04$ for JM). A Scheffé post hoc test showed almost significant differences between the commands and contrastive sentences for speaker GS ($p = 0.058$) and a significant difference between commands and broad focus statements ($p = 0.035$) for speaker JM. The differences were not found to be between the same utterance types across the two speakers.

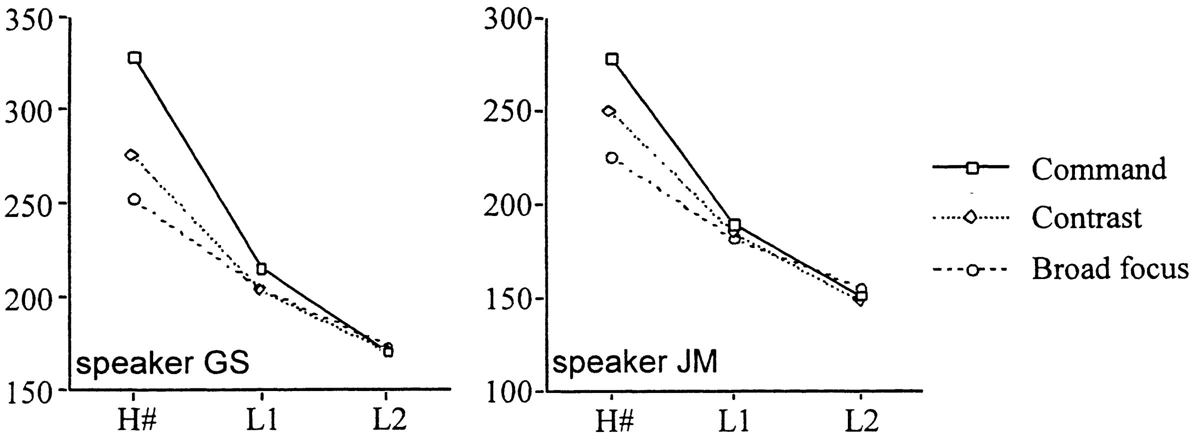

figure 1: f0 means of landmarks H#, L1, L2 in different utterance types for speakers GS and JM

	H# GS	L1 GS	L2 GS	H# JM	L1 JM	L2 JM
command	329.0 Hz (42.5)	215.7 Hz (17.8)	170.7 Hz (6.7)	278.9 Hz (15.4)	189.9 Hz (15.3)	151.5 Hz (6.1)
contrast	275.5 Hz (14.8)	203.7 Hz (9.0)	170.6 Hz (5.0)	249.5 Hz (11.7)	184.9 Hz (9.9)	148.5 Hz (3.0)
broad focus	252.1 Hz (14.8)	204.2 Hz (7.5)	173.7 Hz (3.6)	225.7 Hz (9.1)	181.2.Hz (9.7)	155.3 Hz (4.2)

table 2: means and standard deviation (in brackets) of f0 values in different types of utterance at position H#, L1, and L2 for speakers GS and JM.

The variation was even smaller for L2 values. A significant difference was found ($F = 9.7$; df=2; $p = 0.001$) only for speaker JM, the difference being between contrastive and broad focus sentences ($p = 0.001$).

Only H# values give a clear indication that the three utterance types differ consistently across the two speakers, where commands, contrastive focus statements, and broad focus statements constitute three distinct groups. For L1 and L2 there are differences which are less significant and less consistent across speakers. We now turn to timing and investigate whether there are any consistent differences in that area.

2.2.2 Timing

As for the frequency analyses, a MANOVA with certain durational measures as dependent variables, and utterance and sentence type as independent variables was carried out.

Across the 8 basic sentence types, the timing of L1 is relatively constant in relation to the f0 peak. The duration from S0 (beginning of the nuclear syllable) to H# is correlated with the duration from S0 to L1 ($r= 0.51$; $df = 2$; $p < 0.001$). No effect of sentence type and utterance type was found in duration from H# to L1 in either speaker. This points to L1 being located a fixed distance after the nuclear peak regardless of sentence or utterance type.

Duration of	Speaker	mean	st. dev.
H# to L1 (ms)	GS	121.0	32.4
H# to L1 (ms)	JM	176.9	40.6
H# to L2 (ms)	GS	352.7	108.3
H# to L2 (ms)	JM	438.4	107.8

table 3: Distance of H# to L1 and H# to L2 in ms, mean and standard deviation for speakers GS and JM.

Unlike L1, the position of L2 is highly variable as the means and standard deviations in table 3 show. Moreover it appears to be independent from H#, since the duration between S0 and H# is not significantly correlated with the duration from S0 to L2.

For H# to L2 only speaker JM showed significant differences for utterance type ($F = 3.9$; $df=2$; $p = 0.033$). The high variation in duration from H# to L2 is mainly caused by sentence type with highly significant differences ($F = 15.6$; $df=2$; $p < 0.001$ for JM; $F = 12.4$; $df=2$; $p < 0.001$ for GS). The extremes in the distance between H# and L2 were found for both speakers in the sentences with "Mohn malen" (JM mean = 304.5 ms; GS mean = 190.5 ms; interstress interval = 0) on the one hand and "Wohnungen bemalen" (JM mean = 530.7 ms; GS mean = 452.6 ms; interstress interval = 3) on the other. This indicates that L2 tends to occur later as the interstress interval is increased. However the absolute timing does not tell us which syllable L2 is aligned to. This is investigated in the next section.

2.2.3 Syllable alignment

Since absolute durational measures are speaker and speech rate dependent, and they do not give information about relations to the linguistic structure of the utterance, the location of the landmarks in the f0 contour in relation to syllables was examined.

The position of L1 is in 84.0 % of the cases one syllable after the syllable aligned with the nuclear peak. In 10.6 % of cases L1 occured two syllables after the H peak rather than one.

This was the case if the immediately postnuclear syllable was short[3], or it was part of a longer foot of 3 or more syllables[4]. In only 5.3 % it is in the same syllable.

The highly significant correlation between the distance of the peak and that of L1 from the beginning of the nuclear syllable referred to in section 2.2.2 would in fact predict that L1 is not bound to a particular syllable but rather to a set distance which happens to correspond mainly to the postnuclear syllable. A richer variety of syllable structures and weights would have to be introduced to test this hypothesis exhaustively. However the present data have not been designed to examine this question.

The syllables aligned with L2 are shown in table 4. L2 is most often (94.7 %) aligned with the syllable MA of "(be)malen". In 92 out of 94 (97.9 %) cases, L2 was aligned with a lexically stressed syllable after the nucleus[5].

Syllable	number	percent
MA (malen)	89/ 94	94.7 %
BIL (wohnmobil)	2 / 12	16.7 %
WA (wohnwagen)	1 / 12	8.3 %
len (malen)	2 / 94	2.1 %

table 4: Alignment of L2 with syllables - number and percentage

In the sentences containing "Wohnwagen" and "Wohnmobil" there is variation as to the lexically stressed syllable upon which the baseline is reached. In 16.7 % of the "Wohnmobil" sentences L2 is on BIL and in 8.3 % of the "Wohnwagen" cases it is on WA. This implies that there is a degree of choice on which stressed syllable to reach the baseline[6]. In most cases it was not the secondary stress in the compound, but the stress on the following verb which aligned with L2.

Now, if L2 is predominantly aligned with a stressed syllable, the question arises as to whether it could be interpreted as occuring at the boundary of a foot, since the stressed syllable always marks the beginning of a new foot. If the L tone is interpreted as an edge tone for the nuclear foot, then we would expect that L2 would have aligned more often with BIL of "Wohnmobil" and WA of "Wohnwagen" because these constitute the beginning of the

---

3 L1 occurs 2 syllables after the peak 3 times (out of 11) in "mohn gemahlen", twice (out of 12) in "Wohnmobil malen".

4 L1 occurs 2 syllables after the peak twice (out of 12) in "Wohnungen malen", and 3 times (out of 12) in "Wohnungen bemalen".

5 In fact, L2 is located in the stressed syllable more often than the starred tone of the pitch accent, here labelled as H# (77 out of 94 = 81.9 %)

6 In an additional recording session the two speakers read the basic sentence set twice as stylised contours. While speaker GS realised the step down always on the syllable MA of "(be)malen" (paint/ decorate), speaker JM chose WAG of "Wohnwagen" (caravan) and BIL of "Wohnmobil" (camper) as the stepdown location. Although this is a different contour, sometimes referred to as a 'stylised fall' (e.g. Ladd (1996)) the fact that the step down to mid occurred on the secondary stress of the compound words some of the time as well as on the lexical stress of the main verb supports the possibility of an alignment of postnuclear tones with lexical stress.

immediately postnuclear foot. This was in fact not the case (10/12 in Wohnmobil and 11/12 in Wohnwagen).

Additional evidence is given by the following findings: For each speaker we calculated the means of the location of L2 in the syllable MA- of "(be)malen", if it was located on that syllable. For speaker JM, L2 is located 56.2 % (154.7 ms) into the MA syllable and for speaker GS it is located 40.4 % (92.2 ms). Thus L2 is more oriented towards the center of the syllable than to its edges. There are no cases of L2 occuring immediately before the syllable MA. These findings lead us to reject the hypothesis that L2 occurs at the nuclear foot boundary or indeed at any foot boundary.

The cases where the foot and the word boundary are in different positions are particularily interesting for the question as to whether the fall is completed at the end of the nuclear word, which would have supported the analysis of the L tone as an edge tone of the word, as claimed for English. From the data shown in table 5 it is clear that word boundary does not play a role in the location of L2, because it does not occur in the vicinity of the word boundary at all.

	L2 2 or more syllables after word boundary	L2 one syllable either side of the word boundary
MOHN ge/be	11	0
WOHnung be	12	0
WOHnungen be	12	0
sum	35	0

table 5: Alignment of L2 in words where foot and word boundaries are distinct

We can conclude that L2 aligned neither with the right edge of the nuclear word nor with the right edge of the nuclear foot.

3. Experiment 2

In Experiment 1, L2 was, in the vast majority of the cases, located in the syllable MA of "(be)malen". This could be due to "(be)malen" being the last word in the utterance or to the fact that MA was the penultimate syllable. A follow-up experiment with one speaker and a reduced number of utterance types was carried out in order to test whether the final position of "(be)MAlen" led to the frequency of its alignment with L2.

3.1 Method and Stimuli

The 8 basic sentences were read by speaker GS in contrastive and broad focus statements. After "(be)malen" the word "wollen" was added and the beginning of the sentence was adjusted appropriately. The sentences had the following pattern:

"Die Schüler hatten die ... (be)malen wollen." (the pupils wanted to paint/ decorate the ...).
The empty slots were filled according to the basic sentence set in table 1. The recorded stimuli were analysed as in experiment 1.

3.2 Results

The results for the additional sentences show more variability as to the position of L2 than in experiment 1 (see table 6). L2 is in 68.8 % cases aligned with the syllable MA. In 75 % it is on a lexically stressed syllable. Despite the higher proportion of unstressed syllables aligned with L2 it is clear that neither the last word nor the penultimate syllable is the preferred location of L2. In fact it never occurred on the final word "wollen".

	broad focus		contrast	
	ma	else	ma	else
mohn	2		1	1 len
wohnwagen	1	1 WA	1	1 gen
mohn ge/be	2		2	
wohnung	1	1 len	1	1 ung
wohnmobil	2		1	1 BIL
wohnung be	1	1 be	2	
wohnungen	1	1 en	1	1 en
wohnungen be	2		1	1 en
Sum	12	4	10	6

table 6: Number of occurences of L2 aligned with syllables in two types of utterance: broad focus and contrastive sentences.

4. Discussion

We have been able to observe two landmarks after the peak, rather than one. The first, L1, is a flex point where the fall turns from convex to concave. The second, L2, is the beginning of the plateau at which the fall reaches the baseline. In section 4.1 we discuss whether differences in utterance type reflected in the height of the peak affect the timing of the fall. In section 4.2 and 4.3 we address the status of L1 and L2 as corresponding to phonological entities. Finally in section 4.4 we explore the implications of our findings for autosegmental-metrical theory.

4.1 F0 height and timing

The experiments have shown that there are consistent differences in the height of the H peak across utterance type, indicating three distinct F0 ranges for broad focus statements, contrastive statements, and commands. However, this height distinction does not appear to affect the timing of L1 and L2 in relation to the peak. Leaving open the questions as to whether the F0 height differences are local or global, and whether they are phonological or phonetic, we can assume that the phonological analysis of all three falling contours consists of the same number of tones. We have not found any differences in the contours analysed which could have resulted in the differences between Féry and Uhmann on the one hand, who incorporated a L tone into the accent (H*+L), and Grice et al. on the other who treated the L tone as independent of the accent.

4.2 The status of L1

It was found that the timing of L1 was correlated with that of the foregoing H peak. It might be argued that the L1 therefore corresponds to the L trailing tone of Féry's and Uhmann's H*+L pitch accent. However the truly phonological status of the L1 correlate is unclear. There might be an entirely phonetic explanation for the presence of a discontinuity in the F0 descent. We might for instance take it to be a reflection of the fact that relaxation and contraction of different sets of muscles are involved at different stages in the fall. Assuming that the first part of the fall is achieved by relaxing the crycothyroid muscle, which is known to have a raising effect when contracted, the flex point could mark the stage at which an equilibrium is reached, after which further lowering is achieved only by the contraction of lowering muscles such as the sternohyoid (cf. Strik & Boves 1992, Beckman et al. (1995), and Erickson et al. (1994)). If this is the case, then we would not necessarily expect L1 to occur at a fixed time from the peak, regardless of the distance in Hz between the two points, unless the increased tension needed to produce higher peaks leads to a faster return to the equilibrium level. This issue requires further investigation.

4.3 The status of L2

The elbow of the fall, referred to as L2, was shown, as predicted from our preliminary investigations, to occur at variable distances from the peak, depending on the text of the sentences. In their discussion of English intonation, Pierrehumbert and Beckman (1988) suggest that the intermediate phrase edge tone is simultaneously an edge tone for the nuclear word. In the present study of German, we found no evidence in favour of treating L2 as the correlate of a word-final edge tone. In fact, in cases where word boundaries did not coincide with other boundaries, L2 was, with only two exceptions which occurred in experiment 2, at least two syllables away from the word boundary. The only plausible candidate for a boundary aligning with L2 was the right edge of a foot. However, there were two indications which led us away from treating L2 as the correlate of a foot-final edge tone. First, in the two sentences containing nominal compounds 'Wohnwagen' and 'Wohnmobil', L2 occurred in the majority of cases near the edge not of the nuclear foot, but of a postnuclear foot. If we were positing a foot-final edge tone, the foot concerned would not be the nuclear one, neither would it be the final in the phrase, as experiment 2 has shown. Second, L2 usually occurred not at the boundary but well into the first syllable of the subsequent foot. It also rarely occurred before the boundary. This indicates that it is the stressed syllable itself which attracts L2 alignment.

Whereas in the sentences with nominal compounds there appears to be a choice as to where to align the L2 - supported by data on stylised stepdown contours[7] - it never aligned with the stressed syllable of the auxiliary verb 'wollen' in experiment 2. Thus semantic weight appears to be involved in the choice. All this - alignment with stressed syllable, speaker choice, and semantic weight - points to an accentual function of L2. However, the accentuation brought about by the L tone is clearly secondary, as it does not affect the primary focal structure, which means that the sentence stress remains unchanged. This secondary accentuation is akin to what Kohler (1996) includes in his German text-to-speech synthesis system as 'partial deaccenting', where postnuclear as well as prenuclear lexical stresses are not fully deaccented but rather receive a small degree of accentuation.

4.4 Implications for autosegmental-metrical theory

The accentual function of the L tone poses problems for the established positional definition of the nucleus as the last accent in the phrase (cf. Pierrehumbert 1980). However, German and English are not the only languages requiring a redefinition of the nuclear pitch accent. As Ladd (1996) points out, similar conclusions have been drawn for other languages which have been shown to have accents following the sentence stress: eg. Palermo Italian (Grice 1995), Maltese (Vella 1994), and Portuguese (Frota 1997). Solutions to this problem have been proposed by Grice (1995) and Ladd (1996). Grice associates accents to phrasal nodes in a prosodic tree in which the designated terminal element receives the nuclear accent, after it has percolated down along the strong branches. The nuclear accent is thus not the final accent in the phrase, but the accent which is associated with strong nodes only. Ladd accords the nuclear accent a special status and permits a following tone which may have accentual properties (given an accentable item in the postnuclear text), or may simply surface at the phrase boundary. He refers to this tone as a 'phrase accent'.

In the data reported on here, the L tone corresponding to L2 clearly resembles the phrase accent, in that it has a dual function: it is simultaneously a boundary tone for the intermediate phrase and a secondary accent. Although the issue of whether the accent should be represented as H* or H*+L is left open, the results presented here support the analysis of the fall as having a low target independent of the nuclear accent and suggest a complex function of this tone.

---

[7] cf. footnote 5

5. References

Arvaniti, Amalia, D. Robert Ladd and Ineke Mennen (Forthcoming): Stability of Tonal Alignment: the Case of Greek Prenuclear Accents, Journal of Phonetics.

Beckman, Mary E. and Janet B. Pierrehumbert (1986): Intonational Structure in English and Japanese. Phonology Yearbook 3: pp. 255-310.

Beckman, Mary, Donna Erickson, Kiyoshi Honda, Hiroyuki Hirai, and Seiji Niimi (1995): Physiological correlates of global and local pitch range variation in the production of high tones in English. in: Proc. ICPhS, Stockholm.

Erickson, Donna, Kiyoshi Honda, Hiroyuki Hirai, Mary E. Beckman, and Seiji Niimi (1994): Global pitch range and the production of low tones in English intonation. in: Proc. III. ICSLP, Yokohama, pp 651-654.

Féry, Caroline (1993): German Intonational Patterns. Tübingen.

Frota, Sonia (1997): Association, Alignment and Meaning: the Tonal Sequence HL and Focus in European Portuguese. In: Proceedings of the ESCA Workshop on Intonation, Athens, pp. 127-130.

Grice, Martine (1995): The Intonation of Interrogation in Palermo Italian: Implications for Intonation Theory. Tübingen.

Grice, Martine, Matthias Reyelt, Ralf Benzmüller, Jörg Mayer and Anton Batliner (1996): Consistency in Transcription and Labelling of German Intonation with GToBI, Proc. IV. ICSLP, Philadelphia, pp. 1716-1719.

Kohler, Klaus (1996): Parametric Control of Prosodic Variables by Symbolic Input in TTS synthesis. In: van Santen et al. (eds.) Progress in Speech Synthesis, Springer, New York, pp. 459-475.

Ladd, D. Robert (1996): Intonational Phonology. Cambridge University Press, Cambridge.

Pierrehumbert, Janet B. (1980): The Phonology and Phonetics of English Intonation. MIT Cambridge, Bloomington.

Pierrehumbert, Janet B. and Mary E. Beckman (1988): Japanese Tone Structure. Cambridge, London.

Steele, S. and Mark Liberman, (1987): The shape and alignment of rising intonation, JASA 81, Supl. 1, p. 79.

Strik, Helmer & Lou Boves (1992): A physiological model of intonation. in: Proc. of the Dept. of Language and Speech, University of Nijmegen, Vol. 16/17, pp. 96-105.

Uhmann, Susanne (1991): Fokusphonologie. Niemeyer, Tübingen.

Vella, Alexandra (1994): Prosodic Structure: Intonation in Maltese and its Influence on Maltese English, PhD dissertation, University of Edinburgh.