Comparative Intonational Phonology: English and German

Cover design: Esther Grabe

Printed and bound by: Ponsen and Looijen bv, Wageningen

Comparative Intonational Phonology: English and German

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op maandag 16 maart 1998
des namiddags om 3.30 uur precies

door

Esther Grabe

geboren op 18 maart 1965 te Braunschweig

Promotores:	Prof. dr. A. Cutler
	Prof. dr. C. Gussenhoven

Manuscriptcommissie:	Prof. dr. D. R. Ladd (University of Edinburgh)
	Prof. dr. W. Klein (Max-Planck-Institute, Nijmegen)
	Dr. T. Rietveld

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

To my parents and my brother

and in memory of my grandparents

Acknowledgements

First and foremost, I would like to thank my promotors, Anne Cutler from the Max Planck Institute for Psycholinguistics, and Carlos Gussenhoven from the University of Nijmegen for their generous and enthusiastic guidance. Without Anne Cutler's unfailing support, and without the many inspiring and entertaining discussions I have had with Carlos Gussenhoven, this study would not be what it is. I am deeply grateful to them both. Additionally, during my time at the MPI, I have received much kind support from Prof. Dr. William Barry of the University of Saarbrücken in Germany, both of an intellectual and practical nature.

I also thank the members of Carlos Gussenhoven's prosody working group at the University of Nijmegen. Carlos Gussenhoven, Judith Haan, Erwin Marsi, Brechtje Post and Toni Rietveld have contributed much to making my time in Nijmegen very enjoyable. At the University of Saarbrücken, Bistra Andreeva and Jacques Koreman provided stimulating discussion and a warm welcome. More generally, I am grateful to several people for many helpful comments and discussion. Alphabetically, they are Lyn Frazier, Nina Grønnum, Bob Ladd, Francis Nolan, and Ian Watson.

Als nächstes möchte ich dem Rektor der Realschule Maschstraße in Braunschweig herzlich für seine großzügige Unterstützung danken. Alle Aufnahmen in deutscher Sprache, die in dieser Arbeit vorgestellt werden, wurden an der Realschule Maschstraße durchgeführt und die freundliche Hilfe, die ich dabei von Herrn Wolf, meiner Mutter, und einer grossen Anzahl von Lehrerinnen, Lehrern, Schülerinnen und Schülern erhalten habe, ist von großer Bedeutung für diese Arbeit. Vielen herzlichen Dank!

With respect to the German recordings, special thanks go to my mother, who teaches at the Realschule Maschstraße. Her assistance with organising the recordings, and her advice on experimental subjects was invaluable. The English recordings were carried out at the Phonetics Laboratories at the Universities of Cambridge and Oxford. I have received much support from Francis Nolan in Cambridge and John Coleman and Ian Watson in Oxford, for which I am very grateful. For technical advice and help with the recordings, I thank Geoff Potter and Andrew Slater.

For kind and prompt technical assistance at the MPI, I am indebted to Daan Broeder, Inge Doehring and Christa Hausman-Jamin. Dirk Janssen was always willing to offer cheerful support and herbal tea whenever I got stuck. Also, I would like to thank Dirk Janssen, Toni Rietveld from the University of Nijmegen and Ian Watson for advice on statistics. Moreover, I am grateful to the 'Verwaltung' at the Max Planck, specifically Rian Zondervan, and Norbert Nielen, and to Karin Kastens who manages the Library. I have much appreciated Karin's helpfulness and expertise.

During my time the MPI, I have much enjoyed sharing offices with Ralf Nüse, Dirk Janssen and Arie van der Lugt. Brechtje Post from the University of Nijmegen has been an excellent colleague, house-mate, and friend. Brechtje also translated the summary of this thesis into Dutch.

Finally, I am deeply grateful to Ian Watson for looking after me in the final stages of preparing this thesis, and for providing me with invaluable intellectual and moral support.

Introduction and Background

Chapter 1

1 Introduction

Intonation is defined by Beckman (1995) as 'all aspects of the perceived pitch pattern that the speaker intends for the hearer to use in understanding the utterance, or that the hearer does use whether intentionally controlled by the speaker or not'. These pitch patterns of speech have been described by O'Connor and Arnold (1973) as *significant, systematic*, and *language-specific*. Taken together, the terms *significant* and *systematic* indicate why intonation is assumed to have phonological structure. In traditional analyses of segmental structure, phonology has been seen as concerned with those differences which a given language exploits to convey lexical identity, and thus to convey different meanings. Similarly, two utterances which differ solely in intonational structure can differ in meaning. Additionally, just as the segmental inventories of languages consist of a limited number of phonemes, the number of distinctive pitch patterns is limited.

The third characteristic of intonation, its language-specificity, forms the topic of the present study. Phonemic inventories vary across languages, and so do the inventories of possible pitch patterns. Two intonation systems are contrasted here; those of Southern Standard British English and Northern Standard German. In the literature, views on the presence or absence of cross-linguistic differences between the intonation systems of these languages have, at times, been extreme. Some authors have considered the two systems to be identical while others have asserted them to be fundamentally different. Thus, there is currently no consensus as to whether or not the two languages make use of the same basic set of intonation patterns. This investigation, therefore, focuses on basic structural aspects of intonation, that is, the inventory of pitch patterns available in the two languages. Other aspects of cross-linguistic variability such as the combination of patterns, their frequency of occurrence or their meanings in discourse are not addressed; these can only properly be studied once the taxonomy of distinctive patterns has been established.

The linguistic framework in which the comparison will be made is the *Autosegmental-Metrical framework* (for an overview see Ladd, 1996). This framework was chosen principally for its flexibility. Earlier traditions, such as that of the British school (e.g. Crystal, 1969, O'Connor and Arnold, 1973) describe intonation in terms of a single, unilinear representation, either as a set of holistic tunes or as linear successions of auditory categories. However, within the Autosegmental-Metrical (AM) framework, tunes may be represented at several linguistic levels, both phonetic and phonological. This may be vital for cross-linguistic studies, as it permits analyses in which two languages may be shown to differ at one level of representation and be similar at another. The hypothesis is that previous disagreements as to whether English and German intonation are very similar or very different may be resolved when a sufficiently rich system is used for analysis.

Ladd (1996: 119) suggests that cross-linguistic differences among intonation languages may be classified using a taxonomy of parameters derived from the description of segmental phonology and phonetics within British linguistics. According to this taxonomy, distinctions in intonational structure may be *systemic*, *phonotactic*, *realisational* or *semantic*. *Systemic* refers to differences in the inventory of intonational categories; *realisational* to distinctions in the way these categories are realised. *Phonotactic* refers to differences in the permitted structure of tunes, and *semantic* involves differences in intonational meaning; for instance, the same tune may signal continuation in one language and finality in another. The cross-linguistic study presented here will concentrate on systemic and realisational differences, on the assumption that these have to be established before differences in intonational meaning or function can be investigated[1].

1.1 Outline

The first two chapters of this study are introductory. The following two are corpus-based; they present the findings of auditory and acoustic analyses of directly comparable German and English speech data. Chapters 5 and 6 are experimental; they take up hypotheses arising from the corpus analyses presented in Chapters 3 and 4. The final chapter presents a summary and conclusions.

Chapter 1 summarises previous studies of German and British English intonation. All comparative studies predate the advent of autosegmental-metrical representations, and therefore, influential monolingual studies of each language within the autosegmental-metrical tradition will also be reviewed.

Chapter 2 discusses methodological issues which arise within the autosegmental-metrical framework and develops the basic structure of the descriptive system to be used

---

[1] *Note that it is commonly accepted that the relationship between accent placement and focus is highly similar in English and German. See Max Planck Institute Annual report 1996: 13 (Ruiter and Wilkins, 1996, eds.) for a study on accent placement and anaphoric reference carried out by the present author which shows that in German, accenting and deaccenting work in the same way as has been shown for English and, incidentally, Dutch (see Gussenhoven, 1992).*

in this study. Additionally, the cross-linguistic corpus of speech data collected for the purposes of the present study, and the presentation of evidence are described.

Chapter 3 presents a corpus analysis of Northern Standard German, which has been less extensively studied than Southern British English. The phonological and phonetic properties of the data are presented and illustrated with fundamental frequency traces. This involves the elaboration of the basic descriptive system developed in Chapter 2.

Chapter 4 is comparative, making use of data from a parallel English corpus. Hypotheses about cross-linguistic differences and similarities are developed. It is proposed that the languages may be represented as having the same underlying phonological structure but differing in phonetic implementation.

Chapters 5 and 6 present experimental investigations of two hypotheses emerging from Chapters 3 and 4. Chapter 5 provides systematic cross-linguistic evidence suggesting that English and German differ in pitch accent accommodation effects, and Chapter 6 shows a difference in the acoustic implementation of downstep.

Chapter 7 summarises and discusses the evidence. It concludes that English and German share a common inventory of phonological representations but differ in the way these representations are realised phonetically.

2 Past Research on German and English Intonation

The following review of contrastive studies concentrates, as far as possible, on standard varieties of German and British English. The contrastive studies discussed were carried out in a variety of descriptive frameworks, all of which may be described as 'holistic', that is, none of them explicitly described intonation with more than one level of linguistic representation.

2.1 Contrastive studies

2.1.1 General remarks

In the literature, views on the contrast between English and German intonation have been extreme; some authors have claimed the intonation systems of the languages to be fundamentally different, but others have asserted them to be identical (see Scuffil, 1982 for an overview). Barker (1925), for instance, suggests that English and German intonation are fundamentally different. In her handbook of German intonation for English students, she contrasts an English passage transcribed with English intonation with the same passage transcribed with German intonation. 'The ludicrous effect produced by the wrong intonation of his mother tongue', Barker points out, 'will convince the student, if nothing else can, that English and German intonation are fundamentally different'. German intonation is claimed to differ most clearly from that of English in that it is 'jerky' and should not be spoken with an 'indolent drawl'. The pitch variations in German speech are greater, and the quiet, even tones of English intonation need to be replaced by louder and more energetic tones. Fifty years later, Pürschel (1975) would appear to support Barker's view. He argues that the apparent inability of German learners of English to use intonation patterns of English appropriately even after several years of teaching suggests fundamental differences. Fundamental cross-linguistic differences are claimed also in Féry's (1993) study of German intonation (see section 2.2.3.4 below). However, the hypothesis that English and German intonation are quite different is not supported by the findings of contrastive studies by Kuhlmann (1952: 206), Schubiger (1965), Esser (1978: 51) and Scuffil (1982: 72). These authors assert that differences between English and German intonation are not fundamental. However, none of the authors is explicit on the nature of the claimed non-fundamental differences. Yet other authors have claimed that English and German intonation are virtually identical. Kingdon (1958: 267), for instance, points out that English and German have very similar intonation systems, and Moulton (1966) suggests that the systems are not only identical but also used in much the same way.

In the literature, the discrepant views on contrastive German and English intonation are accompanied by two further conceptions. Firstly, authors commonly state that too little contrastive information is available German and English intonation (Moulton, 1966, Pürschel, 1975, Bald, 1976)[2], and secondly, authors point out that we know more about the intonation of English than about that of German. As early as 1965, Schubiger stated that the investigation of English intonation had reached a point where its form had been explored almost to perfection. The intonational structure of German, on the other hand, had been explored less thoroughly. Almost twenty years later, Scuffil (1982) and Fox (1984) still point to widespread disagreement about the basic facts of German intonation. Since Scuffil's (1982) contrastive study of English and German intonation, the most recent to my knowledge, a great deal of further research has been carried out on the intonation of English, and models such as the autosegmental-metrical system proposed in Pierrehumbert (1980) have been widely adopted. Within the study of German intonation, however, no comparable consensus is apparent (Jin, 1990:3, Möbius, 1993: 31).

---

2 *Anderson (1979) would appear to be an exception. He points out that from the literature, a relatively clear picture of structural similarities and differences emerges. Scuffil (1982: 74), however, contradicts Anderson's assertion and points to wide disagreement in the literature.*

2.1.2 Specific remarks

The specific remarks summarised in this section will be restricted to 'realisational' and 'systemic' cross-linguistic differences. Unfortunately, explicit comparative remarks are rare, and again, more detailed systemic and realisational information appears to be available for English (Bald, 1976:45). Isolated remarks on realisational differences come from a set of comparative studies involving American English and German as well as other languages carried out by Delattre and colleagues in the sixties (1965a, 1965b). Delattre (1965) suggests that in German, the rising part of a falling accent takes the shape of an 'S', followed by a sharp fall to a flat low level which give the impression of being separated from what precedes. He then compares the effect to English speakers hearing the sequence *street-car level* as *street | car-level* (other things being equal). In English, the 'S' is reversed, and constitutes the falling, rather than the rising part of the accent. This observation led Delattre, Poenack and Olsen (1965) to suggest that in English the general form of intonation is wave-like, whereas in German it can be compared to the blade of a saw.

Figure 1 *Phonetic differences between English and German intonation (adapted from Delattre, 1965); note that English and German intonations were produced on an English text).*

Anderson (1979) describes the difference between English and German intonation illustrated in Figure 1 as one involving 'falling' vs. 'rising' emphasis in the pitch accent. Additionally, he suggests a number of rather general tonal and rhythmic differences between German and English. Firstly, the 'neutral' pattern in English and German is said to be realised differently. In German, the pattern is the equivalent of a 'flat hat', that is, rising pitch on the first accented syllable followed by falling pitch on the second (see 't Hart, Collier and Cohen, 1990 for the term 'flat hat'). In English, the first and second accents are rising-falling. In short sentences, however, 'declarative falls' and 'interrogative rises' are claimed to be near-identical, and generally, the suggestion is that the basic tonal inventory is very similar. A minor difference characterises the stretch between the last accent in the phrase and the following boundary, which is said to be lower in pitch and more monotonic in German. Anderson also suggests rhythmic differences. German speech is said to give 'more equal weight' to each syllable and contain longer series of unstressed syllables whereas American English is claimed to be characterised by a stronger 'stress beat'.

Gibbon (1984:35) informally lists a number of cross-linguistic differences which have been claimed to distinguish English and German. German pitch contours are claimed to have a higher proportion of level tones, level stretches and jumps between levels rather than glides; and compound tones tend to be less frequent than in English (except in dialects). However, Gibbon also points out that observations of this kind are doomed to remain atomistic unless the broader framework of language use is taken into account. A successful comparison needs to consider factors such as speaking style and register; otherwise, 'the systems' of two languages one appears to contrast may, in fact, be 'phantoms'.

Finally, Trim (1988: 240) states that German speakers tend to treat all non-nuclear prominent syllables alike (mid to high level or low rising, according to region) and this may produce an effect of monotony and lack of rapport in an English listener. Moreover, German speakers are said to use intensity rather than pitch range for emphasis (some support for this claim has been provided recently in Gut, 1995)[3] and this may sound aggressive to English listeners. English speakers, on the other hand, are said to make use of the first stressed syllable in an intonation phrase as an index of cheerfulness, arranging subsequent syllables on a descending scale. In German, this feature is said to be altogether absent. Trim suggests that this may mean that German speakers fail to establish the emotional atmosphere in a way expected by an English interlocutor. Additionally, in German discourse, the end of a turn is said to be commonly signalled by a falling nuclear tone. This is likely to be perceived by English listeners as an attempt to impose a belief. In English, on the other hand, turns are frequently concluded with rising or falling-rising tones which are said to invite comment from a listener. More generally, Trim points out that German intonation may make speakers sound bleak, dogmatic or pedantic, and as a result, English listeners may consider them uncompromising and self-opinionated (often to the German speakers' surprise). Germans, on the other hand, feel that the pitch of an English speaker's voice wanders meaninglessly if agreeably up and down. Additionally, 'they [the English speakers] often turn out to have meant something quite different from what they actually said, showing them to be devious and hypocritical behind that infamous snobbish reserve and meretricious facade of gentleness, such that butter would not melt in the mouth!' (Trim, 1988: 244).

---

3 Similar observations may have led Klinghardt (1920: 23) to point out that 'Wir Deutsche sind bei allen unseren germanischen Brüdern weithin als Schreier bekannt' ('among our Germanic brothers, we Germans are well known to be yellers').

2.1.3 Summary and discussion

A survey of the relatively small number of previous studies comparing English and German intonation suggests that authors have generally agreed that we know very little about this comparison, but disagreed on almost all aspects that have been investigated. Why might this disagreement have arisen? Firstly, because researchers have compared information collected in different descriptive traditions, focusing on different aspects of intonational structure, and may have described different speaking styles characterised by different realisations of the languages' intonational systems. The problem does not only apply to cross-linguistic comparisons; it is compounded by similar difficulties applying specifically to German. Scuffil (1982: 51) points out that studies of German intonation are not only marked by a variety of theoretical approaches, but there is also less agreement on the facts than is the case for English. Similarly, Möbius (1993:1) states that even the attempt to survey studies investigating German intonation is considerably hindered by individual contributions being based on completely different theoretical assumptions. Less agreement on the intonational phenomena to be described emerges than from studies investigating English intonation.

A second reason may be that researchers have addressed more than one of the linguistic functions of intonation at a time. Intonation has multiple functions in speech, intonation patterns play a role in discourse, they may signal paralinguistic information such as tenderness or anger, they may convey semantic information such as 'non-routineness' and they may signal syntactic structure. Accounts of English and German which combine several aspects of intonation in its description without explicitly motivating the combination may have complicated cross-linguistic comparisons.

A third reason may be that researchers have assumed that intonation could be modelled with only one level of linguistic representation, the exact status of this representation being unclear. Were it the case that English and German differed at one level of representation but not at another, then a unilinear system would not be able to deal with this. Studies within a unilinear system might then come up with either the 'highly similar' or the 'very different' view, depending on which aspect of intonation they investigated, or whether their investigative technique was auditory or instrumental. However, before the 1970s, no widely accepted non-holistic framework for the description of intonation was available.

Since then, however, considerable theoretical advances have been made. The 'autosegmental-metrical framework', which has become widely accepted, may be said to combine O'Connor and Arnold's three premises of intonational significance, systematicity and language-specificity with a departure from the unilinear representation of intonation. Instead, perceived intonation contours are broken down into a number of linguistic representations, which allow, for instance, a clear separation between cross-linguistic differences involving the phonological system of a language and those reflecting phonetic surface distinctions arising despite a shared phonological inventory. This theoretical advance, combined with technological progress which allows extensive speech corpora to be stored, labelled and widely disseminated, has opened up new avenues for cross-varietal research .

2.2 Monolingual Autosegmental studies

The following sections will provide a brief summary of the autosegmental-metrical framework (for a comprehensive overview of the autosegmental-metrical framework see Ladd, 1996). This will be followed by a more detailed presentation of those autosegmental-metrical studies on which the system used here was based.

2.2.1 The autosegmental-metrical framework

Researchers working within the autosegmental-metrical framework postulate that English and German tunes may be represented as having more than one level of linguistic representation. Basic to all systems is the assumption that intonation patterns may be decomposed into a number of primitives (primitive only at the intonational level, as each represents a synchronisation of two prosodic events, one tonal and one rhythmic). In English and German these primitives are pitch accents, that is, pitch movements anchored to stressed syllables, and boundary tones, which are pitch movements accompanying rhythmic discontinuities at the phrase edge[4]. The tonal properties of primitives are transcribed by using the letters H and L, which stand for high and low events in fundamental frequency and pitch, and the rhythmic properties by assigning a '*' following the letter transcribing the tone associated with a stressed syllable in the case of pitch accents and a '%' in the case of boundary tones. In representations of English and German intonation, pitch accents are commonly assumed to be either monotonal or bitonal, and boundary tones to be monotonal (see Pierrehumbert, 1980, Ladd, 1983a, Gussenhoven, 1984, Beckman and Pierrehumbert, 1986, Pierrehumbert and Beckman 1988 and Lindsey, 1985 for English, and Wunderlich, 1988, Uhmann, 1991 and Féry, 1993 for German).

The primitives are used to transcribe intonation at the phonological level, and languages may differ in their inventory of primitives. The phonological representation is mapped onto a phonetic realisation via a set of phonetic realisation rules, which are again language-specific (note that in this study, and with respect to intonation, 'phonetic' will refer to the combined auditory impressions of pitch, length and loudness).

---

4 *For pitch accent theory see Bolinger, e.g. 1958, 1986.*

Some authors assume one level of intonational phrasing (e.g. Pierrehumbert, 1980, Gussenhoven, 1984, Lindsey, 1985, Uhmann, 1991, Féry, 1993); others, following Beckman and Pierrehumbert (1986) assume two; the *intonation phrase* (IP) and a level of phrasing below the IP, the *intermediate phrase*, a prosodic constituent smaller than the intonation phrase. In systems following the 'Beckman-Pierrehumbert' approach, the intonation phrase is delimited by a boundary tone (transcribed with a percent sign following the high or low tone), and the intermediate phrase is delimited by a *phrase accent* (transcribed with a dash following the tone). However, it not always clear which criteria distinguish an intermediate phrase from an intonation phrase proper, and the concept of the phrase accent itself is somewhat controversial (see Ladd, 1983a: 746 and 1996: 89 for a critique and section 2.2.2.1 below).

2.2.2 Studies on English

2.2.2.1 Pierrehumbert (1980)

The first comprehensive autosegmental-metrical account of English was offered in Pierrehumbert's (1980) doctoral thesis, and this account has been influential ever since. Combining insights from previous work by Liberman (1975), Liberman and Prince (1977), and Bruce (1977), Pierrehumbert proposed a comprehensive account of American English intonation using two pitch levels associated with metrically strong syllables and intonation phrase boundaries. Previous authors had assumed four pitch levels, which led to considerable ambiguity in the resulting system (Pike, 1945, and Trager and Smith, 1951, suggested accounts of this type; for a critique, see Bolinger, 1951). Moreover, Pierrehumbert set new standards of experimental verification in intonation analysis by (a) making an explicit distinction between phonological and phonetic levels of representation, and (b) providing a set of mapping rules from one level to another (Ladd, 1996: 3). Note, however, that unlike in studies of intonation carried out within the British school of intonation analysis (e.g. Crystal, 1969), in Pierrehumbert's study 'phonetic' refers to the acoustic representation of fundamental frequency only (see Nolan, 1990 for a discussion of different views on levels of representations in phonetics). Within the British school, 'phonetic' may refer to the acoustic realisation of intonation, but more commonly, the term refers to the auditory impression of a specific contour when analysed by a trained phonetician.

In Pierrehumbert's system, each intonation phrase must consist minimally of a pitch accent, an initial and a final boundary tone. Additionally, each intonation phrase must have a *phrase accent*. The phrase accent was borrowed from Bruce's (1977) description of Swedish and posited by Pierrehumbert to account for F0 movement on and following the last pitch accent in the intonation phrase. Taken together, the phrase tone and the boundary tone account for the difference in complexity which frequently distinguishes intonation phrase final and non-final pitch accents. The phonological inventory Pierrehumbert posits for English is shown in (1); it claims that pitch accents may be monotonal or bitonal and may be 'right-headed' or 'left-headed', that is, either the first or the second element in a bitonal accent may be associated with a metrically strong syllable. Additionally, accents may be downstepped, that is, the high element of a pitch accent may be lowered in the pitch range relative to a preceding high tone. Downstep allows Pierrehumbert to account for the pitch patterns of English with only two pitch levels, despite cases in which a high tone is absolutely lower in the register than another high tone.

(1)	**Pitch accents**		**Phrase accents**	**Boundary tones**
	H*+L	H+L*	H-	H%
	L*+H	L+H*	L-	L%
	H*+H			
	H*			
	L*			

The phrase accent (or phrase *tone*, as phrase accents do not represent complete pitch accents but part of a pitch accent) allows Pierrehumbert to distinguish between two types of nuclear fall, a terminal fall which goes all the way down to the hypothesised baseline of a speaker's register, and a vocative fall which stops well above the baseline (1980: 74). This difference, Pierrehumbert argues, cannot be captured in models accounting for intonation patterns as sequences of F0 changes rather than sequences of F0 targets. In such models, she states, the declarative and the vocative fall involve more or less falling pitch[5]. In Pierrehumbert's system, the terminal fall is decomposed into H* L- L% and the vocative fall into H*+L H-L%. H*+L in the vocative contour is said to differ from H* in the declarative in that H*+L triggers downstep of a following high H- phrase accent. As a result, H- is lowered beyond the location expected in the normal course of an utterance (basically, to a mid-level). The lowered H- tone, in turn, is said to trigger upstep of the final L%, and thus, a fall in F0 ending mid is generated. Figure 2 below schematises the difference in F0 between vocative and terminal declarative falls as well as the transcriptions Pierrehumbert suggests.

5 Note, however, that Crystal (1969) describes this distinction as a categorical difference between a nuclear fall proper and a suspended nuclear fall. His system accounts for intonation patterns as sequences of F0 changes.

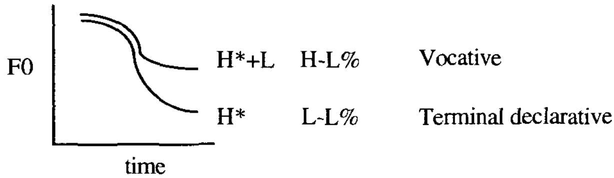


Figure 2 *Terminal declarative and vocative falls in Pierrehumbert (1980).*

Without the phrase accent, the distinction between vocative and terminal falls cannot be made in Pierrehumbert's system (see Ladd, 1983a and 1986 for a similar point)[6]. However, as the transcription of the vocative contour in Figure 2 shows, transcriptions involving phrase accents are rather complex. Additionally, the transcriptions do not reflect the structural similarity between declarative and vocative contours very well. This latter problem emerges also in the following example (adapted from Pierrehumbert 1980: 227):

L*+H L*+H L*+H L*+H L* H-H%
(2) Do you **real**ly be**lieve** Ebe**ne**zer was a **deal**er in Mag**ne**sium?[7]

The question in (2) is accounted for by Pierrehumbert as a series of rising accents, the last of which is transcribed as L* H-, that is, as phonologically different from the preceding accents. Additionally, the final L* H- is followed by H%. This transcription captures the difference in realisation between the phrase-final rising accent and the preceding rises: in (2), the pitch range of the final rise is larger. A more straightforward account of this difference appears to be L*+H H% with the H% capturing the final rise. However, in Pierrehumbert's system, this transcription is not possible. The trailing H in L*+H does not trigger upstep of the H%, and the final rise is not accounted for. Only the phrase accent H- triggers upstep, and this means that the phrase-final pitch accent must be transcribed as L* H- rather than as L*+H. An apparent phonological difference between the final accent and the preceding accents is the result. If one were to assume, alternatively, that boundary tones are implemented relationally rather than absolutely, i.e. an H% boundary tone is always higher than an immediately preceding H, then a transcription without a phrase accent would capture the pattern equally well, and reflect

---

6 *The vocative fall is captured more straightforwardly in Gussenhoven (1984) as H*L with HALF-COMPLETION, a modification which prevents F0 crossing the middle of a speaker's register and without HALF-COMPLETION. This account captures the falling character of both accents as well as their differences.*

7 *Pierrehumbert (1980) illustrates this example with a fundamental frequency trace. In this trace, the IP-final H- is marked as if it was located higher in the register than preceding H- tones. It is not clear why this was assumed to be the case; at the end of the phrase, the trace rises smoothly and no one event in the trace rather than another appears to be associated with the location of H-.*

the structural and semantic similarity between the rising accents in the phrase. Intuitively, the final pitch accent does not seem to differ in meaning from the preceding pitch accents. Further comments on the phrase accent can be found in section 2.4 of the following chapter.

Pierrehumbert posits one level of intonational phrasing; the intonation phrase, which is obligatorily delimited by a high or a low boundary tone. High boundary tones are motivated by sharp upwards movements in F0 at the phrase edge in the absence of a stressed syllable, but low boundary tones are not realised by equivalent downward F0 movement. Nevertheless, Pierrehumbert suggests that the description of intonation is considerably simplified if we assume that there is a low counterpart to H%, and accounts for the difference in phonetic implementation between high and low boundary tones by a special phonetic implementation rule. This rule states that phrase accents are spread before tones which are phonetically equal or higher, but not before those which are lower. Thus, L- does not spread before L% which is claimed to be lower, but rather interpolates with L% and this accounts for the gradual drop in F0 which is often observed in intonation phrases ending low (Pierrehumbert, 1980: 47)[8]. Note, however, the hypothetical status of the claim that L% is lower than L-. Evidence comparable to that for H% being higher than H- is not available.

2.2.2.2 Intonational phrasing: Beckman and Pierrehumbert (1986) and Ladd (1986)

In 1986, two proposals were published within the autosegmental-metrical tradition suggesting a level of intonational phrasing below that of the intonation phrase (Beckman & Pierrehumbert, 1986 and Ladd, 1986). Although in effect, they addressed the same issue, the authors proposed their models of phrasing for different reasons. Ladd's proposal aimed to account for apparent mismatches between tonal and rhythmic cues to intonational phrasing. Traditionally, an intonation phrase had been defined by (a) the presence of a 'nuclear' accent (see Cruttenden 1986: 48 for the notion of nucleus), and (b) rhythmic breaks or pausing. However, as Ladd pointed out, some phrases appear to contain two nuclear accents, not separated by an audible rhythmic break, and intonational tags can be delimited by pauses but nevertheless not bear an accent. To account for such apparently mismatched cues, Ladd posited a recursive two-level intonational phrase structure. The lower level, the Tone Group, was defined on the basis of tonal information whereas the higher level, the major phrase, was set off by audible rhythmic breaks.

---

8 *It is not immediately obvious why tones should spread only when followed by a tone which is at the same level or higher, but not when followed by a lower tone. Possibly, however, Pierrehumbert's rule reflects a more general asymmetry between high and low tones in American English (e.g. high tones are downstepped, but low tones are not).*

Beckman and Pierrehumbert's (1986) proposal on the other hand, explicitly built on Pierrehumbert's (1980) approach. The authors proposed the 'intermediate phrase', a level of intonational structure below that of the intonation phrase. Intermediate phrases are delimited by Pierrehumbert's (1980) phrase accent, and account for a wide range of intonational phenomena such as intonation phrases with multiple nuclei, similar tonal patterns in lists and intonational tags (for tags see e.g. Gussenhoven, 1990)[9]. Moreover, intermediate phrases were claimed to be the domain of downstep (for downstep see also Pierrehumbert and Beckman, 1988). Beckman and Pierrehumbert's model of intonational phrase structure contrasts with Ladd's in that both levels are defined on the basis of tonal information only, whereas in Ladd's proposal, one is defined on the basis of rhythmic and the other on the basis of tonal information.

However, both proposals leave a number of questions open. Firstly, clear acoustic cues distinguishing the two levels of phrasing proposed appear to be elusive, and to my knowledge, no comprehensive study contrasting them is available. In spontaneous speech, Ladd's major intonational phrases will not necessarily be delimited by audible prosodic breaks. Beckman & Pierrehumbert do not suggest any clear acoustic cues beyond the similarity in tonal structure between successive intonation phrases in lists. In tags, however, which are also accounted for as intermediate phrases, no such similarity needs to emerge; a tag may be unaccented, and is then dissimilar in patterning from a preceding host phrase. This point will be taken up again in section 1.2.7 below, where the view taken on phrasing in this study will be discussed.

2.2.2.3 Downstep: Pierrehumbert (1980), Liberman and Pierrehumbert (1984) and Pierrehumbert and Beckman (1988)

Pitch tends to decline over the course of phrases and utterances, and this effect has been one of the most widely studied properties of speech (Ladd, 1984, 1996). Pierrehumbert (1980) was the first to model downtrends in English fundamental frequency as 'downstep', that is, a local, step-wise lowering of pitch at specific accents rather than as a property of the complete intonation phrase. The notion of downstep is important to her account of (American) English and the AM framework in general, because it permits a modelling of tunes as linear sequences with only two pitch levels H and L, despite the fact that within one tune, some high targets may be lower than others. Pierrehumbert proposed, specifically, that downstep is triggered by an alternating sequence of H and L tones. In some of her descriptions, an L tone is simply there to lower the F0 value of a following H tone and has no direct manifestation in the F0 contour (see section 2.2.1

9 *In Pierrehumbert (1980), phrase accents immediately precede a boundary tone and cannot occur between pitch accents.*

above for a brief discussion of the phrase accent, which was introduced as a direct result of this proposal).

Pierrehumbert's work was further developed in Liberman and Pierrehumbert (1984) and Pierrehumbert and Beckman (1988), and the model of downstep first presented in Liberman and Pierrehumbert (1984) is probably the most explicit one currently available for (American) English. An experimental investigation led the authors to propose four characteristic aspects of downstepped sequences; (1) the value of each accent peak in the sequence may be expressed as a constant proportion of the one immediately preceding; (2) therefore, the steps between successive pairs of accents decrease; (3) English has 'final lowering', that is, the final accent in a sequence appears lower in F0 than predicted by the location of the immediately preceding accent, and (4) the final low in each IP is constant for each speaker. Their findings led them to suggest that downstep may be modelled with an exponential decaying curve. 'Final lowering' explains why the last accent in their sequences does not fit this curve.

Beckman and Pierrehumbert (1988) further elaborated details of the model of downstep first proposed in Liberman and Pierrehumbert (1984). They explicitly distinguished between three sources of downtrend in F0, 'catathesis', 'declination' and 'final lowering' (in later work, the authors replaced the term 'catathesis' with 'downstep', the earlier term). Catathesis has been covered by the summary of downstep given above. Declination is defined by the authors as a lowering of the pitch range which operates in time from the beginning of the utterance, without regard to the tonal description. Unlike downstep, declination is not a process whose domain is the intermediate phrase; rather, it appears to operate at some larger level of structure. Its existence in English is controversial, but the authors did find evidence for it in Japanese. Final lowering happens at the ends of declarative sentences; and is defined as a gradual compression and shift of the pitch range which occurs in anticipation of the end of a declarative utterance. It affects the scaling of accents as well as postnuclear tones. Finally, Beckman and Pierrehumbert point out that the many studies investigating downtrends do not adequately separate the effects of catathesis and final lowering from those of declination. Downtrends in English and German will be dealt with in more detail in Chapter 6.

2.2.2.4 An autosegmental-metrical feature model: Ladd (1983)

Ladd (1983b) presents an autosegmental-metrical feature model of intonational phonology based on work by Bruce and Gårding (1978) and Pierrehumbert (1980). His aim was to remedy shortcomings in previous work along the same lines, specifically, a difficulty in expressing a number of phonological and function generalisations based on overall contour shape. The central problem with Pierrehumbert's and similar work, Ladd points out, is an excessive concern with the perceptual and acoustic details of F0. However, such details play a secondary role in understanding the linguistic structure of intonation. More important are the functional distinctions of intonation, which have been extensively investigated in non-instrumental models of intonation (e.g. the British tradition). Ladd bridges the gap between instrumental and functional approaches by positing 'a systematic taxonomy of intonational phonetics' which is to serve as a basis for analysing the phonology of intonation. This taxonomy would then allow researchers to state generalisations about function.

The system Ladd uses to illustrate his point has four pitch accents (H, L, HL and LH) and two boundary tones (L% and H%). The leftmost tone is automatically associated with a metrically strong syllable, and therefore the '*' is omitted. Ladd then exemplifies his proposal with three phonetic features: [delayed peak], [raised peak], and [downstep]. For instance, the difference between a HL accent with a delayed peak and one without a delayed peak corresponds to that between rise-fall and a fall in the British tradition; the advantage in Ladd's system is that the structural and semantic similarity of the fall and the rise-fall is captured more explicitly. Similarly, [raised peak] captures the similarity between an accent with an extra high peak and one without.

Ladd's account of downstep differs from Pierrehumbert's in that it does not posit a sequence of H and L as a downstep trigger. Instead, downstep is claimed to be an independent speaker choice and accounted for by a downstep feature. Ladd (1983) proposes downstep to be a feature of intonational peaks involving an independent speaker choice. A criticism levelled against Ladd's downstep feature, however, is that it overgenerates, in that it allows tones to be downstepped in isolation although downstep is generally assumed to be a relational phenomenon (Grice, 1995a). It allows for the first H accent in a phrase to be downstepped, although this does not appear to happen[10]. Therefore, Ladd (1983) proposed that the downstep trigger be marked on the accent preceding the one that is downstepped. Then, an initial accent could not be downstepped. However, as Grice (1992) points out, this has the disadvantage that, theoretically, a low tone could be downstepped, and again, this does not appear to be the case. In a modified proposal (Ladd, 1990b, 1993a) accounts for downstep as a metrical relationship between intonational constituents (see Liberman, 1975 for a similar proposal for English).

---

10 *Note, however, that this problem depends on the way one defines downstep. If one assumes that an accent can 'step down' not only from another high accent, but also from an initial high boundary tone, then one could account for a falling accent preceded by high preaccentual pitch as, e.g. %H !H*+L.*

2.2.2.5 Two levels of phonological representation: Gussenhoven (1984)

Gussenhoven's (1984) autosegmental account of British English is based on the nuclear tones recognised in the British tradition. Three of these tones, the rise (L*H), the fall (H*L) and the fall-rise (H*LH) he takes as basic; all other patterns observed are derived from the basic tones. Thus, while Pierrehumbert asserts that English lacks rules which alter tonal values or delete tones (1980: 3), and that therefore underlying and derived phonological representations are identical, Gussenhoven posits just such tonal alteration rules, and suggests that English intonation is best accounted for with an underlying and a surface level of phonological representation[11]. Similarly to analyses of connected speech processes in segmental phonology which, for example, posit a process of assimilation to change the realisation of /s/ towards /ʃ/ in *she packs shorts,* Gussenhoven's tonal rules alter the realisations of tones in certain contexts. In this respect, his system is the only one in the AM tradition to meet an objection raised by Crystal to intonational analyses in general (1969:40); "there is the hidden assumption that, having done an analytic survey of the basic functional 'blocks' of intonation, the synthesis of these blocks into connected utterance is simple. All the evidence goes to suggest that this is not the case, and that connected speech makes important modifications to the units into which it can theoretically be broken down." Other AM systems postulate (implicitly) that there should be no difference between the intonational structure of citation forms and that of continuous speech. This would appear to imply that intonation is different from the segmentals of speech, i.e. that only segmental structure undergoes connected speech processes and suprasegmental structure does not.

In Gussenhoven's account, two kinds of operations may change tonal values, and their domain of application differs. 'Modifications' apply to nuclear tones and 'linking rules' to prenuclear tones. Four modifications are proposed to apply in British English: DELAY, which can turn a fall into a rise-fall by delaying the peak of the fall relative to the accented syllable, STYLISATION which creates a spreading mid-tone[12], for instance in calling contours, HALF-COMPLETION which accounts for tones failing to run their full course[13], and RANGE, which runs orthogonal to the other three modifications in that it affects nuclear tones as a whole and expands or compresses their realisations. RANGE appears to be somewhat problematic; this modification does not match the other categories well in that it is claimed to be gradient rather than categorical. This makes it difficult to see how one decides whether RANGE has applied or not or whether it applies

---

11 *Note, however, that despite Pierrehumbert's (1980) claim that General American English lacks rules which alter tonal values, her account includes downstep, a process which lowers high tones.*

12 *This feature was borrowed from Ladd (1978).*

13 *Gussenhoven defines HALF-COMPLETION as 'the failure of the tone to cross the mid-line' (1984:222).*

by default at all times, just in differing degrees. Also, RANGE may be confused with HALF-COMPLETION.

Gussenhoven indicates that in combination, nuclear tones and modifications specify twelve nuclear tones, although the gradient status of RANGE makes this claim problematic. Furthermore, Gussenhoven points out that the twelve tones cannot capture all nuclear contours found. The remaining contours, which occur less frequently, are accounted for by combining modifications (for details, cf. Gussenhoven, 1984:232 and Gussenhoven, 1988).

'Linking rules' optionally reduce the realisations of prenuclear accents and thereby account for the differences between nuclear and prenuclear accent patterns which are an integral part of descriptive systems put forward within the British tradition (e.g. Crystal, 1969, O'Connor and Arnold, 1973). Theoretically, any tone can be linked to any following tone, but in practice, not all combinations occur equally frequently. Two types of linking may apply; partial linking and complete linking (see Figure 3 below). Partial linking results in the slope of the fall or rise following an accented syllable being more gradual than that characterising unlinked nuclear tones. Complete linking is rather difficult to describe in purely auditory terms (this partially explains why completely linked contours have been described as categorically different from unlinked contours in the British tradition).

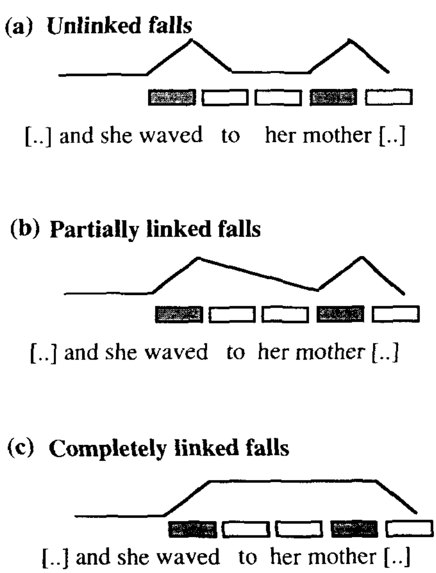


Figure 3 *Prenuclear accents in Gussenhoven (1984).*

Gussenhoven points out that this analysis of prenuclear accents enables one to group together semantically similar contours which are treated as categorically different in the British tradition. The stylised examples in Figure 3 illustrate his point (the contours represent auditory impressions and the shaded boxes the accented syllables). In the British tradition, both (a) and (b) could reasonably be analysed as falling head plus falling nucleus contours. (c), on the other hand, needs to be described as a level head plus falling nucleus, that is, a categorically different contour, despite its intuitive similarity to the other two. Gussenhoven's analysis captures the intuitive similarity between the three contours but it can also express the differences between them. The argument is that all three contours are underlyingly HL HL. In (a), no linking rules have applied, in (b), partial linking has applied and in (c) complete linking. Thus, the strength of Gussenhoven's approach lies in its ability to capture structural similarities and differences at more than one level of representation.

A further advantage of Gussenhoven's approach involves its ability to capture parsimoniously differences and similarities between different speaking styles. It is reasonable to assume that speaking styles differ from each other not only with respect to the choice and distribution of pitch accents, but also with respect to the way specific accents are realised. Spontaneous speech, for instance, is likely to be characterised by more instances of 'accent linking' that read speech (see Figure 3). In the Pierrehumbert system, each instance of 'linking' is accounted for as a separate accent choice, and needs to be stated separately. In the Gussenhoven system, the difference does not involve accent choice; linked and unlinked realisations of H*+L are derived from a single underlying level of representation. One may state that spontaneous speech is characterised by more frequent applications of linking than read speech.

Finally, Gussenhoven's approach differs from other AM approaches discussed in this study in that it proposes meanings for the modifications postulated to apply at the surface level of phonological representation[14]. However, intonational meaning is notoriously hard to pin down, largely because of its context dependency, and the meanings Gussenhoven suggests for his modifications reflect this difficulty to some extent[15]. DELAY is said to signal to listeners that the accented word relates to something 'non-routine' and 'very significant'. STYLISATION, on the other hand, is said to signal 'routineness' (as may be claimed to be signalled in calling contours, for instance). The meaning of HALF-COMPLETION appears to be even harder to define than that of the other modifications; 'unconvincingness' is mentioned tentatively. However, if one may account for the difference between a terminal declarative fall and a vocative fall in the

[14] *See Pierrehumbert and Hirschberg (1990) for a compositional theory of intonational meaning which suggests (partial) meaning to tones in the tonal inventory.*

[15] *For an experimental investigation of intonational meaning in Dutch, see Grabe et. al, 1997.*

way suggested, then 'unconvincingness' would not appear to be appropriate. Finally, differences in RANGE are related to different degrees of insistence. Linking rules are not said to affect the meaning of intonation phrases directly, rather, their application may be related to differences in focus structure.

2.2.2.6 Gussenhoven (1984) vs. Pierrehumbert (1980)

Gussenhoven (1984) presents a model of English intonation which, in many ways, parallels traditional analyses of segmental phonetic structure. As in phonemic analysis, a set of primitive, phonologically contrastive categories of intonational structure is posited, and the realisation of these primitives is governed by a set of phonetic implementation rules which are (a) sensitive to segmental structure and (b) language specific. Additionally, the primitives may be realised either directly, or they may undergo phonological adjustments when several categories are combined into an intonational phrase structure. These adjustments systematically modify the underlying structure when basic categories are combined in continuous speech[16]. It is these adjustments which most obviously distinguish Gussenhoven's model from that presented in Pierrehumbert (1980).

The difference between the models is apparent especially when we compare the authors' solutions to modelling the distinction between IP final (i.e. nuclear) and non-final accents. Final pitch accents are characterised by a phonetically richer realisation than non-final ones, that is, they tend to exhibit a larger inventory of pitch accent shapes. In principle, there are two ways of accounting for this distinction. Either one reduces the realisations of non-final accents in some way, or one enriches the realisation of final accents. Gussenhoven favours the first solution; he accounts for reduced prenuclear realisations with a linking rule, that is, a phonological adjustment. Pierrehumbert prefers the second; she does not make use of phonological adjustments, and in her account of American English, final accents are followed by a phrase tone and boundary tone, and differences in the realisation of prenuclear accents are handled by a richer set of phonetic realisation rules. Figure 4 illustrates the basic differences between the models. The figure shows that Gussenhoven's system assumes two levels of phonological representation, the underlying level, at which the primitives are specified, and a surface level. The surface level is derived from the underlying level via a set of phonological adjustment rules, that is, the modifications and linking rules. The phonological surface structure is then translated by phonetic realisation rules into the phonetic realisation. In Pierrehumbert's

---

16 *Gussenhoven suggests that primitives and modifications are morphemes. For instance, the effect of adding DELAY to H*+L can be compared to the addition of the past tense morpheme <-ed> to the stem of the verb to walk.. The meaning of walk is changed, but not altered beyond recognition. Similarly, DELAY adds something to the meaning of H*+L (e.g. 'surprise'), but does not change its character completely.*

model, the phonological representation has only one level, but a richer set of phonetic realisation rules accounts for differences in surface structure.

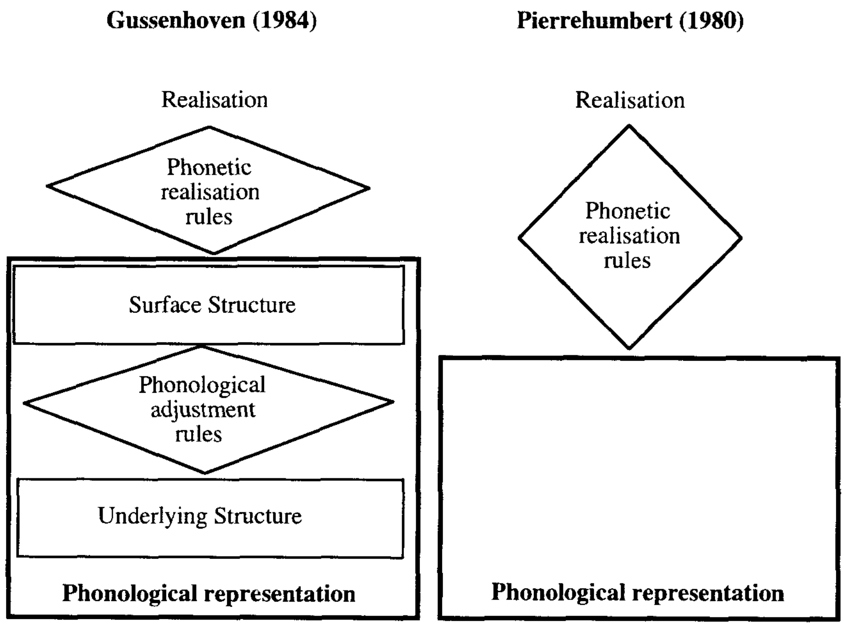


Figure 4 *Summary of differences between models of English intonation proposed in Pierrehumbert (1980) and Gussenhoven (1984).*

The advantage of positing two levels of phonological structure rather than just one can be illustrated by a comparison of Pierrehumbert's and Gussenhoven's accounts of the difference between terminal falls, vocative falls and terminal rise-falls. Pierrehumbert accounts for the difference as one involving different choices from the phonological inventory. The fall is transcribed as H*L-L%, the vocative fall as H*+L H-L%, and the rise-fall as L*+H L-L%. In Gussenhoven's system, on the other hand, the three contours are derived from the same phonological category H*+L . The terminal fall is basic, and is not modified; underlying and surface representations are identical. The vocative fall is represented as H*+L with HALF-COMPLETION, and the rise-fall as H*+L with DELAY. Thus, in Gussenhoven's system, the structural and semantic similarity between the three types of fall is captured explicitly, whereas in Pierrehumbert's system, the similarity between the contours is much less obvious (see also Ladd's 1983 critique of the contour classification generated by the Pierrehumbert system). Table 1 below contrasts the two analyses of terminal fall, vocative fall and terminal rise-fall.

	Terminal fall	*Vocative fall*	*Terminal rise-fall*
Pierrehumbert 1980	H*L-L%	H*+L H-L%	L*+H L-L%
	Be li n d a	Be li n da	Bel i n d a
Gussenhoven 1984	H*+L	H*+L, HALF-COMPLETION	H*+L, DELAY

Table 1 *Pierrehumbert's and Gussenhoven's analyses of three nuclear falling contours.*

A further advantage of Gussenhoven's system is that tonal changes which characterise particular speaking styles (e.g. careful vs. casual speech) do not need to be specified separately every time they occur, but can be stated as applying to a set of utterances as a whole. For instance, in casual speech, we often find that prenuclear accents, trailing tones or boundaries are deleted. In Pierrehumbert's system, every instance of deletion has to be specified separately, because every observed difference between contours is taken to reflect a different choice from the phonological inventory. The Gussenhoven system is more parsimonious; we can state that casual speech differs from careful speech in that it has more DELETION.

In summary, Gussenhoven's system appears to be (a) more parsimonious and (b) more flexible. Not only can it account straightforwardly for the structural similarities and differences which characterise pitch accents, but it can also capture structural similarities between larger stretches of utterance. The Pierrehumbert system offers less transparent transcriptions and cannot account for intonational differences distinguishing different speaking styles in any obvious way. We need to concede, however, that experimental evidence in favour of Gussenhoven's system is not easy to come by. Generally, controlled data supporting linguists' intuitions of semantic and structural similarities between contours are scarce, and in the absence of such data, one may argue that it is more consistent to represent what appears to be a surface categorical difference between two intonational surface structures as just that, a categorical difference, and no more (see 't Hart, Collier and Cohen, 1990 for such an approach).

One experimental study, however, has compared the nuclear tone taxonomies proposed in Pierrehumbert (1980) and Gussenhoven (1984). and this study supports Gussenhoven's system. Gussenhoven and Rietveld (1991) asked American English subjects to estimate the semantic contrast in paired nuclear tones, and their judgements were correlated with the sets of theoretical differences predicted in the two systems. The results showed that Gussenhoven's system was a better predictor of the experimental scores.

Finally, note that some of the differences in Pierrehumbert's and Gussenhoven's systems may result from actual difference between British English and American English intonation. Clearly, the systems do differ in some respects. For instance, in American English, nuclear high rises or rise-level contours are far more commonly used for statements than in British English. At times, such difference may have led the authors to regard different types of distinctions as more relevant than others. However, the experimental subjects tested in Gussenhoven and Rietveld (1991) were American, rather than British English speakers. The finding that Gussenhoven's system was nevertheless a better predictor of the experimental scores than Pierrehumbert's system can be interpreted to suggest that Gussenhoven's account has some validity for American English also.

2.2.3 Studies on German

Within German intonation research, there appears to be less agreement on basic facts than in research on English (Möbius, 1993), and far fewer studies have been carried out within the autosegmental framework. These will be reviewed in the following sections, as well as one earlier non-autosegmental-metrical study (Isačenko and Schädlich, 1966), which is included because the authors were the first to model German intonation with two pitch levels.

2.2.3.1 Isačenko and Schädlich (1966)

Partly in response to shortcomings in the Trager & Smith style levels analysis, Isačenko & Schädlich modelled German intonation with two pitch levels and tested their model with perception experiments using synthesised speech. On the basis of their findings, they suggested that the basic elements of German intonation involve one rising and one falling pitch change ('Tonbrüche'). Falls and rises are associated with the 'ictus' of a stressed syllable, that is, its voiced section. Changes either precede the 'ictus' or follow it. Two types of rises and two types of fall result, and are illustrated in Table 2 below. Intuitively, Table 2 summarises the options available in phonetic surface structure of German falls and rises very well, and in a later AM analysis summarised below, Féry (1993) lists patterns which would appear to correspond to those proposed by Isačenko and Schädlich (1966). Féry's (1993) transcriptions are given in Table 3. A comparison between Table 2 and Table 3 shows that Isačenko and Schädlich's system differs from Féry's in that it is perfectly symmetrical; the authors simply list logical options: pitch may step up before a stressed syllable, or after a stressed syllable. Féry's later account shows that the way such options are modelled in the AM framework depends on more than the logical possibilities available for a particular accent's surface realisation.

Generally, the way intonational categories are modelled in the AM framework reflects (a) their distribution and (b) the degree to which they are similar or different, both structurally and semantically. Féry's modelling, which reflects these concerns, is foreshadowed by comments Isačenko and Schädlich make about the characteristics of their basic categories. They say that pre- and post-ictic rises differ in their distribution. The post-ictic rise (L*+H) is the 'rise proper'; it can appear in prenuclear position as well as nuclear position. The pre-ictic rise (H*), on the other hand, cannot appear in nuclear position; in fact, we would be left with a sentence fragment, were a pre-ictic rise to appear intonation phrase-finally. In prenuclear position, on the other hand, a pre-ictic rise is said to 'foreshadow' a following fall, either pre- or post-ictic. In Féry's system, which is based on Gussenhoven's (1984) approach, H* appears in prenuclear position only and is derived via a linking rule from H*+L, but L*+H can be nuclear or prenuclear. The difference between pre- and post-ictic fall is said by Isačenko and Schädlich to be distinctive, and, again, this observation is reflected in Féry's system.

	A **rising pitch change**	**B** **falling pitch change**
I pre-ictic	die Kinder	die Kinder
II post-ictic	die Kinder	die Kinder

Table 2 *Isačenko & Schädlich's inventory of falls and rising pitch changes in German. Adapted from Isačenko & Schädlich (1966: 60).*

	A **Rises**	**B** **Falls**
I pre-ictic	H*	!H*+L
II post-ictic	L*+H	H*+L

Table 3 *AM categories of German intonation proposed in Féry (1993) which appear to correspond to those proposed by Isačenko and Schädlich (1966).*

The difference between Isačenko and Schädlich's falling accents, on the other hand, is not claimed to be distinctive, and both may appear in prenuclear and nuclear position.

Again, this observation appears to be reflected in Féry's system, where these pitch changes are modelled as the same pitch accent H*+L, either downstepped relative to a preceding accent or not.

2.2.3.2 Wunderlich (1988)

Following Pierrehumbert (1980) and Ladd (1983), Wunderlich (1988) presents an autosegmental-metrical account of German which distinguishes between a phonological and a phonetic component (in practice, however, the article concentrates on the phonological component and devotes only a few general comments to phonetic implementation). Wunderlich's system describes a limited set of intonation patterns established on the basis of an examination of F0 traces (no information regarding speakers, dialects or speaking style is given). He illustrates these patterns as in Figure 5 below (my translation of the impressionistic names of the different accent patterns). The small brackets indicate the location of the accented syllables, and the brackets in bold the right or left edge of an intonation phrase.

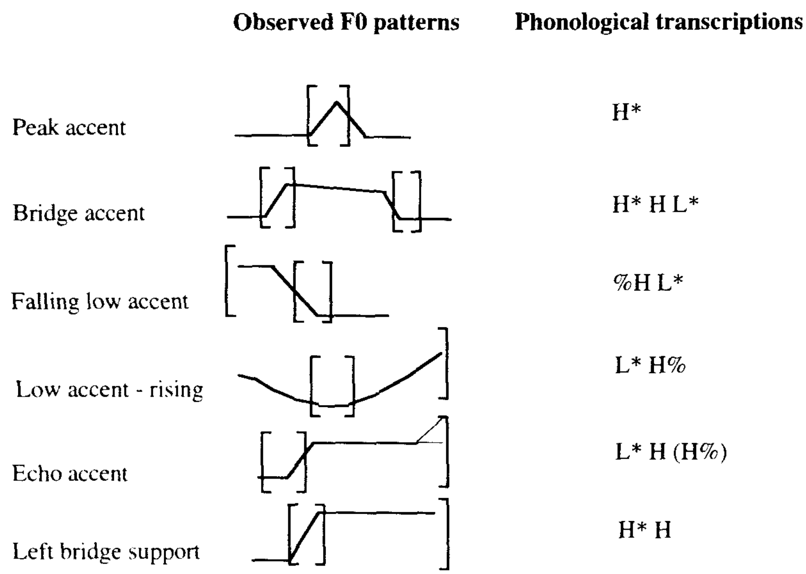


Figure 5 *Wunderlich's accents patterns of German. Adapted from Wunderlich (1988: 11).*

As can be seen in Figure 5, Wunderlich's bridge accent represents a combination of two accents, made up from the 'left bridge support' (left) and the falling low accent (right). In the 'echo accent', Wunderlich comments, the F0 peak is reached in the post-accentual syllable, but within the accented syllable in the 'left bridge support'. All patterns were attested in perception experiments, but no details or references are given. However, the status of the brackets surrounding the H% boundary tone in the Echo accent remains unexplained; do they indicate that the presence of H% is optional? And why do we not find a similar opposition for the 'left bridge support'? Moreover, the term Echo accent represents a category mismatch; here, function rather than form is implied, whereas all other accent terms refer to form only.

Wunderlich posits three types of phonological entities; pitch accents, boundary tones, and 'non-boundary tones' (presumably similar to trailing tones in Pierrehumbert's system). He then bases his autosegmental-metrical account of the F0 patterns illustrated in Figure 5 on two phonological oppositions, (a) presence or absence of a boundary tone, and (b) high (H) vs. low (L) tone. Moreover, he assumes a distinction between 'marked' and 'unmarked' patterns, and suggests that unmarked patterns do not need to be specified. The distinctions he makes are summarised in Table 4 below (again, the translations into English are mine).

	Unmarked	**Marked**
Boundary tone	L	H
non-boundary tone	L	H
pitch accents	H	L

Table 4 *Phonological distinctions and markedness. Adapted from Wunderlich (1988: 19). 'Non-boundary tone' is Wunderlich's term for a trailing tone.*

These oppositions are the basis for his transcription of the patterns observed as shown in Figure 5 above on the right.

Wunderlich's transcriptions raise a number of questions. Firstly, what is the basis for the marked-unmarked classification of tones in Table 3? Secondly, if only marked patterns need to be specified, why are both low and high accents specified, although high is the default? With respect to boundary tones, it appears that in Figure 5, the unmarked case, that is, the low boundary tone, has not been specified. Thirdly, what is the status of the non-boundary tone? These unresolved questions combined with a lack of clearly laid out evidence and transparent motivations for the modelling make Wunderlich's claims too weak to generate clear predictions.

2.2.3.3 Uhmann (1991)

Uhmann's autosegmental-metrical analysis of German intonation concentrates on the relationship between intonation and focus. Her observations are based on the analysis of two read speech corpora, in which focus structure was systematically manipulated (her test sentences were embedded in dialogues). Corpus I was read by eight speakers (six female and two male) and Corpus II by four speakers (two female and two male; no information about linguistic background or age is given). However, the data she presents are from two selected speakers from each set.

On the basis of her corpus analyses, Uhmann proposes that German has four pitch accents: H*+L, H*, L*+H and L*. These accents are said to have been 'extracted' from the corpus. The differences between H*+L and H* and between L*+H and L* are motivated by differences in the realisation of postaccentual syllables. Monotonal pitch accents, she states, do not influence the F0 of following syllables but bitonal pitch accents do.

Intonation phrase boundaries were modelled on the basis of F0 measurements at IP on- and offset. A bimodal distribution of F0 values at IP offset leads Uhmann to posit two boundary tones H% and L%. No similarly clear distribution is observed at IP onset, but nevertheless, Uhmann also posits two boundary tones in this position, although the boundary may also remain unspecified. However, this proposal fails to capture the clear discrepancy between IP onsets and offsets which her data reveals.

In terms of the number of representational levels, Uhmann's system resembles that of Pierrehumbert (1980) rather than that of Gussenhoven (1984). One level of phonological representation and one level of phonetic implementation are assumed, and no explicit distinctions are made between nuclear and prenuclear accents. Phonological adjustment rules are not discussed. On the other hand, Uhmann posits only one level of intonational phrase structure (unlike Beckman and Pierrehumbert 1986), and no phrase tone, and in this respect, her system resembles that of Gussenhoven.

2.2.3.4 Féry (1993)

The most comprehensive AM study of German intonation to date was presented in Féry's (1993) *German intonational patterns*. Her findings were based on the analysis of a corpus of 100 sentences read by three native speakers of Standard German (375 tokens, because some sentences were read with more than one realisation).

Féry's study had two aims; firstly, to give an autosegmental-metrical account of the phonological properties of German intonation, and secondly, to investigate the influence of a number of linguistic factors on the tonal pattern of utterances. Among these were focus-structure, topic-comment structure and scope. The following summary of her findings will concentrate on the phonological system she posits.

Féry's system differs from Uhmann's in that it posits two levels of intonational phrasing (the intonational phrase and the intermediate phrase), but, again, no phrase accent. Féry argues that phrase accents are not needed in German, either to describe the pitch movement between the last pitch accent and the boundary tone, or to delimit the intermediate phrase. In Féry's system, the pitch movement between the last pitch accent and the boundary is accounted for by the trailing tone of the last bitonal pitch accent which spreads to the end of the intonation phrase (all the nuclear pitch accents she posits are at least bitonal). Féry's intermediate phrase is not delimited by a phrase accent because of the 'phrasing which exists independently of the tone structure anyway' (1993: 72). One optional, final boundary tone is posited, but no initial boundary tone(s). No low boundary tone is said to be needed, because in overall falling contours, there is no tonal movement marking the end of an IP.

Within intonation phrases, Féry posits three nuclear accents: H*L, L*H and L*HL. The tritonal accent replaces Ladd's (1983) 'delayed peak'. She argues that German does not have a feature 'delayed peak', citing evidence from perception studies by Kohler (1987, 1991). Kohler synthesised continua between falling accent contours with 'early', 'middle' and late peaks. Féry suggests that if German had the 'delayed peak' feature, listeners should have been sensitive to the difference between the late and middle peaks in Kohler's studies; in fact, listeners were found to make a categorical distinction between early and middle peaks, but not between middle and late peaks. Given this, however, it is not clear why Féry postulates the additional tritonal nuclear pitch accent, considering that the distinction between middle and later peak appears to be smaller than that between middle and early peak (and this distinction is the one she accounts for with an 'early peak' feature; see below). In Gussenhoven's (1984) system, which Féry to some extent adopts, the phonological distance between two different nuclear accents is actually greater than that between variants of the same accent with or without a modification.

Following Gussenhoven (1984), Féry posits two modifications, STYLISATION and EARLY PEAK. The term STYLISATION is borrowed from Ladd (1978) and accounts for calling contours, whereas EARLY PEAK accounts for high preaccentual pitch. This second modification is based on the categorical distinction between 'early' and 'middle peak' observed by Kohler (1987, 1991). The nuclear-prenuclear distinction, finally, is accounted for by Gussenhoven's tone linking rules.

A special problem in German which Féry raises involves the 'hat pattern' (t'Hart, Collier and Cohen, 1990). Féry states that German has two different types of hat patterns, whose derivation is not straightforward. Her account of the derivation postulates that the patterns contain different accents but have, by coincidence, the same form. Hat contour 1 is analysed as a completely linked sequence of two H*L pitch accents. After linking has applied, the structure of an H*+L H*+L sequence is H* H*+L. Hat contour 2 consists of two fully realised accents L*+H H*+L. The difference between the contours is said to be not always phonologically clear-cut, and their lack of distinctiveness in some contexts is compared to neutralisation in segmental phonology. Féry's account of the two types of hat contour may be compared to Wunderlich's (1988) distinction between a 'Bridge accent' and an 'Echo accent' (see Figure 5 earlier). Wunderlich based his distinction on the alignment of the F0 contour with the stressed syllable. In the first element of the Bridge accent, F0 rises throughout the accented syllable and then levels out into a plateau. In the Echo accent, the F0 patterns is the same, but aligned later; now the rise continues beyond the accented syllable. The Bridge accent appears to be comparable to Féry's hat contour 1, where the first accent is H*, and the Echo accent resembles hat contour 2 which begins with L*+H.

Unlike Uhmann (1991), Féry (1993) contains a section on downtrends, in which she provides fundamental frequency traces illustrating examples of downstep in German. These contours are discussed briefly in the light of accounts of downstep in English posited by Pierrehumbert (1980), Liberman and Pierrehumbert (1984) and Ladd (1983), and Féry concludes that more research into downstep in German is needed.

Féry's account of German intonation leaves open a number of questions[17]. Firstly, the phonetic realisations of the intonational categories posited are not discussed. Further data on accent and realisation are required, for instance, on the distinction between the two types of hat pattern suggested. Also, a discussion of the theoretical implications of the boundary tone asymmetry posited would be desirable. Secondly, the status of Féry's intermediate phrase is unclear; specifically, it is not obvious how an intermediate phrase can be distinguished from an intonation phrase. The intermediate phrase is said to be delimited by the trailing tone of the last pitch accent in it. But how can a trailing tone have a delimiting function? In intonation phrases, this same trailing tone is claimed to spread up to the IP boundary. Thirdly, as Féry points out, more evidence is needed on downstep in German.

Of special interest to the present study are hypotheses offered by Féry about the difference between English and German intonation. Her hypotheses are summarised and briefly discussed below.

(1) The set of possible postnuclear realisations is more restricted in German than it is in English (1993: 61).

---

17 *See also Gussenhoven's (1994) and Ladd's (1994) reviews of Féry's study.*

(2) In English, the phrase accent is needed to control the melody between the nuclear accent and the boundary tone. In German, the nuclear accent is generally followed by an abrupt fall or rise immediately after the nuclear accent and not by a boundary tone at the end of the intonation phrase (1993: 74).

(3) There is no tonal movement marking the end of a falling IP in German (1993: 72).

The claim under (1) motivates the one optional boundary tone postulated by Féry as opposed to the two phrase accents and two boundary tones delimiting the intonation phrase in Pierrehumbert's account of English. However, no comparative data is offered to support this claim. Moreover, at least with respect to the intonation phrase, Féry's optional boundary tone would appear to generate more rather than fewer boundary options (i.e. 'low' in H*L without a boundary tone, 'high' in L*H without a boundary tone, and 'extra high' in L*H with a high boundary tone H%). This would suggest a larger rather than a smaller range of postnuclear realisations than the two available in English.

Claim (2) can be interpreted to imply that in English the nuclear accent is not generally followed by an abrupt fall or rise immediately after the nuclear accent. However, Pierrehumbert's (1980) results suggest otherwise. In English H*L-L%, for instance, the fall takes place on the postaccentual syllable after which the contour levels out gradually (see Pierrehumbert 1980: 187, figure 2.32 B).

The suggestion under (3) also implies a cross-linguistic difference. English is assumed to exhibit tonal movement marking the end of a falling IP, but German is not. However, again, there is counter-evidence. Pierrehumbert's (1980) examples of intonation phrases with L% do not show downward movement in F0 at the end of the IP, and therefore Pierrehumbert suggests a special phonetic implementation rule which accounts for the apparently asymmetrical realisation of high and low IP boundaries. Moreover, an AM account of English has been proposed by Lindsey (1985) which assumes that English has high but no low boundary tones.

Clearly, some of Féry's proposals require further investigation. Accent and boundary realisation in English and German, the question of unspecified intonation phrase boundaries, and downstep in German are among the issues addressed in the following chapters.

2.3 Prosodic labelling: ToBI

The contrastive autosegmental-metrical analysis of English and German presented in the present study was based on evidence from a directly comparable corpus of English and

German speech data. Contrasting these data required (a) that they should be prosodically labelled, and (b) that the labels should be comparable. This precluded the use of pre-existing systems such as Pierrehumbert's for English and Féry's for German which are not directly comparable. The following sections will briefly discuss two relatively widely used, very similar AM prosodic labelling systems which have recently been proposed for English and German.

2.3.1 English ToBI

In 1992, the ToBI system for prosodic labelling[18] was proposed as a standard for the transcription of General American English, General Australian and Southern Standard British English (Silverman et al. 1992, see also Beckman and Ayers 1994). The labelling system was the joint initiative of a group of researchers and based on Pierrehumbert (1980) and subsequent revisions of her work by Beckman and Pierrehumbert (1986) and Pierrehumbert and Beckman (1988). The development of ToBI was motivated by a need to establish a commonly used and understood system to indicate prosodic features in labelled computer corpora of speech (Ladd 1996: 94).

Briefly, ToBI transcribes intonation as a linear sequence of prosodic events on parallel tiers, principally the tone tier and the break index tier. On the tone tier, five pitch accents (H*, H+!H*, L*, L+H*, L*+H) and a two level intonational phrase structure are transcribed. Downstep is indicated by a '!' symbol. Both the intonation phrase and the smaller intermediate phrase may end high or low (H% vs. L% and H- vs. L-). On the break index tier, the degree of coherence between adjacent words is labelled on a scale from '0' (highest level of coherence) to '4' (least coherent). '0' is defined in terms of connected speech processes such as cliticisation; '1' describes most medial word boundaries in connected speech; '3' delimits an intermediate phrase and '4' marks a full intonational phrase boundary.

The exact linguistic status of ToBI has remained somewhat vague. Specifically, it is unclear whether ToBI is intended to provide phonetic transcriptions of intonation, phonological transcriptions, or possibly neither. When the transcription system was first introduced, it appeared to be phonetic rather than phonological in nature. The authors pointed to a need for a single standard for prosodic transcription analogous to the IPA for phonetic segments, and they appeared to suggest that ToBI was developed to meet this need (Silverman et al. 1992: 867). However, this parallelism is questionable; an IPA transcription of speech may be made without applying linguistic decisions (i.e. nonsense words may be transcribed), but ToBI labelling requires linguistic decisions. For instance,

---

18 *ToBI stands for 'Tones and Break Indices'; 'tones' refers to intonation, and 'break indices' to rhythmic structure.*

transcribers need to be able to identify the stressed syllables with which pitch accents are associated. Thus, it appears that ToBI is not a phonetic transcription system[19]. Whether ToBI is strictly phonemic, however, is also questionable. For instance, we find that a distinction is made between falling nuclear accents with a smaller or a larger onglide in pitch and fundamental frequency on the stressed syllable (labelled as H* L-L% and L+H* L-L%). This distinction is not made in the British school of intonation analysis, which also claims to transcribe phonological differences. Moreover, in an evaluation of ToBI, L+H* is described as a minor variant of H*, and the categories L+H* and H* are collapsed (Pitrelli et al., 1994)[20]. Apparently, ToBI is not strictly phonemic either. The impression that ToBI represents a compromise between a phonetic and a phonemic transcription system is reinforced by discrepancies in the system between labels which are minimally abstract, such as, for instance those referring to the distinction between H* L-L% and L+H* L-L%, and those transcribing intonation phrase boundaries which are relatively indirect. L* H-L%, for instance, transcribes a rise-to-mid, requiring a phonetic implementation rule 'upstep' which raises the final L% to the level of the preceding H-.

To summarise, it appears that in its current state, ToBI represents an uneasy compromise. Ladd (1996: 95) points out that ToBI is first of all a set of conventions for labelling prosodic features, aimed at making large corpora of speech more useful for research, Clearly, when developing such conventions, compromise is required. Whether a compromise between a phonetic and a phonological transcription is the best solution, however, may be questioned. A labelling system which explicitly distinguishes between a narrow level of transcription, which is minimally abstract and a broader, more obviously phonological level combined with more detailed explorations of the status of both levels may be preferable to the type of compromise offered by ToBI.

2.3.2 German ToBI

On the basis of the ToBI system developed for American English (henceforth 'EToBI'), a unified single ToBI system for German (ToBIG or GToBI) emerged in 1995 (Grice et al. 1996). Contributions came from ToBI-style systems developed in parallel at the universities of Braunschweig, Saarbrücken, and Stuttgart (Batliner and Reyelt, 1994, Grice and Benzmüller, 1995, Mayer, 1995). The GToBI inventory contains the five pitch

---

19 *Ladd (1996: 95), for instance, has pointed out that ToBI is not some kind of high-tech IPA for prosody. Moreover, Pierrehumbert (1980), on whose work ToBI draws, has argued that there are no phonetic representations other than those derived directly from the physical signal (see Nolan, 1990 for a critique). A phonetic ToBI would appear not to fit in with this view.*

20 *Grice et al. (1996) present an evaluation of German ToBI, in which L+H* and H* are the categories which are most frequently confused. However, L+H* was also confused with L*+H, and authors conclude that L+H* can therefore not be treated as a subcategory of either H* or L*+H.*

accents of EToBI plus one further accent H+L*, and again, downstep is indicated by the '!' symbol. Additionally, GToBI differs from EToBI in that intermediate phrases do not have to contain an accent (this is obligatory in EToBI). Inter-transcriber consistency among labellers using GToBI has been evaluated (Grice et al. 1996), and the results appear to be comparable to those obtained in a similar study using EToBI (Pitrelli et. al, 1994). In the GToBI labelling test, 71% inter-transcriber consistency was achieved for pitch accents whereas 68.3% was achieved in the EToBI test. Part of this agreement was on whether or not an accent was present (87% in GToBI vs. 80.6% in EToBI) and which accent was present (51% in GToBI vs. 64.1% in EToBI). 33% of the disagreement on which pitch accent was present in GToBI involved the accent pair L+H* and H*, but L+H* was also confused with L*+H, that is, in some cases, a falling accent L+H*L- must have been confused with a rising accent L*+H; a rather worrying finding. In the EToBI test, the results for L+H* and H* were merged, which suggests that transcribers may not have distinguished between them reliably. Thus, in both EToBI and GToBI, transcribers tend to agree relatively reliably on the presence or absence of an accent, but the agreement for the type of accent present appears to be rather low. Nevertheless, the evaluators of GToBI conclude that GToBI is already adequate for the transcription of databases in German. The evaluators of EToBI conclude that the EToBI convention and its training materials have been refined to the point that they can be used fruitfully for the labelling of prosodic phenomena in speech databases. Considering the levels of agreement found, however, both conclusions seem somewhat hasty. Presumably, the users of prosodically transcribed databases require more than just a reliable indication of whether an accent is present or not. Developers of speech synthesis systems, for instance, are likely to be interested in accurate information about the type of accent used in a specific utterance as well as in information about accent distribution.

Some criticism has been levelled at GToBI by Kohler (1995). Firstly, Kohler questions the status of the phonological model underlying the GToBI. He points out that the underlying model for EToBI is the one developed for American English by Pierrehumbert and colleagues. With respect to GToBI, however, it is not entirely clear whether the transcription system is based on an independent analysis of German intonation or whether the notational device of American ToBI has simply been transferred to German[21]. Considering that English and German intonation are nowhere near as different from each other as for instance, German and French intonation, Kohler concedes that this might not necessarily be a big problem, but if this is an appropriate approach, then it needs clear phonetic and phonological justification. GToBI may in some part be based on Féry's (1993) study of German but this analysis constitutes only a

---

21 *Similarly to EToBI, GToBI appears to be a compromise, driven to some extent by the need for an agreed labelling system for prosody that could be used in the VERBMOBIL project (see Batliner and Reyelt, 1994).*

partial model of German intonation and is functionally orientated towards focus and grammatical phrasing rather than constituting a formal phonological model.

Possibly in response, Grice et al. (1996: 1717) point out briefly that English and German are closely related languages which share a similar rhythm and intonation structure. However, there are differences in the inventories of pitch accents (GToBI has a pitch accent H+L* which EToBI does not have), and in the phonetic realisation of the pitch accent categories the languages share.

Secondly, Kohler questions the re-introduction of Pierrehumbert's (1980) H+L* and points to the lack of justification for this decision. It is unclear, he states, whether the decision was made on language-independent grounds or because the labelling of German made it mandatory. This criticism may not be entirely fair, however. One may assume that the decision to introduce H+L* was made on language-independent grounds, since Saarbrücken ToBI, one of the contributors to GToBI, was developed on the basis of Map Task data and has H+L* (see Anderson et al., 1991 for the Map Task). Additionally, Kohler's own work appears points towards a categorical distinction between early and medial F0 peaks in German nuclear falling accents (Kohler, 1987a).

Thirdly, Kohler points out that the theoretical objections to ToBI are compounded by practical problems; we do not know how the phonological categories given in ToBI are realised in the phonetics. Transcribers are, in fact, given some examples of the phonetic realisation of EToBI and GToBI labels in training materials available by anonymous ftp from the Linguistics Department at Ohio State University, and the Phonetics Department at Saarbrücken University. However, these examples are unlikely to suffice in their present form. For instance, neither training set provides systematic comparisons of the realisation of a specific pitch accent on different segmental material, and in English and German, pitch accents may be realised quite differently in different contexts. For instance, in English, the peak of an accent may shift to the right or left, depending on the amount of sonorant segmental material contained in the stressed syllable or the number of syllables following before an intonation phrase boundary intervenes (see van Santen and Hirschberg, 1994, Silverman and Pierrehumbert, 1990 for segmental effects on pitch accent realisation in English). Such changes in peak location may compound the confusions between the categories L+H* and H*. In German, on the other hand, an H* L- pitch accent does not involve a fall in F0 when realised on an IP-final syllable with a small proportion of sonorants (see Chapter 5 of the present study). This may lead a transcriber to label the pitch accent as H*H- rather than as H*L-. Thus, detailed information about segmental influences on acoustic patterns is required for successful prosodic labelling, but such information is not given in the EToBI and GToBI training materials. Finally, the acoustic phonetic realisation of a specific label is not likely to be identical in different varieties of American English and German; transcribers need to be aware of this, and they need to know what to expect. When combined, the difficulties which inexperienced labellers face are likely to render a successful application of GToBI or EToBI doubtful.

3 Summary

This chapter has summarised previous contrastive accounts of English and German intonation. The survey has shown that authors have agreed that we know little about this particular contrast but have disagreed on most of the aspects which have been investigated. Consequently, some authors have claimed that the intonational structures of English and German are quite similar, but others have claimed them to be fundamentally different. In the present chapter, it was argued that the disagreement is likely to have arisen because (a) generally, research on German intonation is characterised by less agreement about basic facts than English intonation, (b) researchers may have compared data which are not directly comparable (e.g. utterances analysed in different descriptive traditions), and (c) researchers have assumed that intonation can be modelled with only one level of linguistic representation. English and German may, however, differ at one level of representation and be similar at another. Additionally, the linguistic status of the representations which have been used often remains unclear. A relatively recently developed linguistic framework which allows for a description of intonation contours on several linguistic levels is the autosegmental-metrical framework. In this framework, a distinction may be made between cross-linguistic differences involving, for instance, the phonological systems of two languages and those reflecting phonetic surface distinctions arising despite a shared phonological inventory. Accordingly, this is the framework used for cross-linguistic comparison in this study.

As English and German have not been compared previously within the autosegmental-metrical framework, a number of relevant monolingual autosegmental-metrical accounts of English and German intonation were summarised. The summary illustrated the range of approaches which have been taken within this framework. The differences between two influential systems, the one proposed by Pierrehumbert (1980) and the one proposed by Gussenhoven (1984) were discussed in detail, and it was suggested that Gussenhoven's approach is better suited to cross-linguistic research. The principal strength of Gussenhoven's system lies in its ability to capture structural similarities and differences at two levels of phonological representation. English and German may well be felicitously described as not differing at the underlying level of phonological representation but differing at the surface level, and Gussenhoven's system would allow for such an account. Pierrehumbert's system, on the other hand, which posits only one level of phonological representation, does not allow for an account of this type.

To conclude, the relatively small number of previous contrastive studies on English and German intonation have generated some hypotheses about cross-linguistic similarities and differences, but the lack of agreement among researchers suggests that there is scope for further research. Tightly constrained studies are needed which address the realisation of one or more clearly specified aspect of intonation in a restricted number of conditions. For instance, discoursal aspects of intonation may be compared across languages in one specific speaking style, or the speaker attitudes conveyed by certain patterns may be compared across different social groups. Moreover, the linguistic background of experimental subjects needs to be controlled for. The research presented in the following chapters is restricted to structural aspects of intonation patterns produced in one speaking style, and the speakers were closely matched for language background and age. The assumption was that cross-linguistic data about basic structural characteristics need to be available first, before other issues such as discoursal or attitudinal differences may be fruitfully addressed.

An Approach to Cross-language Comparison of Intonation

Chapter 2

1 Introduction

This chapter discusses theoretical and practical considerations constraining the contrastive analysis of English and German intonation presented in the following chapters. The practical considerations involve questions of analytic technique. The theoretical considerations lead to the proposal of an autosegmental-metrical system for direct comparison of German and English which differs in one or more aspects from all of the previously suggested language-specific autosegmental-metrical systems. Such a system was required because no AM studies are available which have analysed English and German in directly comparable variants of the framework. A cross linguistic study, however, requires languages to be compared in, as far as possible, the same system.

2 Theoretical considerations

Ideally, an intonational system for cross-linguistic comparison would combine previous insights about basic similarities between the languages with the smallest number of assumptions about language specific characteristics. Also, it would be flexible enough to capture similarities and differences between contours within and across languages.

To obtain such a tool, researchers have two options. Either they choose a previously developed language-specific account that matches best the ideal system described above, or they develop a relatively simple compromise system which combines insights from a number of studies. In the present study, the second option was preferred. Cross-linguistic studies are based on the assumption that linguistic systems may differ across languages. This suggests that a transfer of linguistic categories from one language to another is likely to hinder rather than help the discovery of language-specific characteristics.

For English, the simplest and most flexible system was judged to be that proposed by Gussenhoven (1984). Gussenhoven posits three basic pitch accents (rather than Pierrehumbert's seven), a limited set of modifications, and one level of intonational phrasing. Féry's system for German has borrowed some features such as tone linking from Gussenhoven and will therefore be the starting point for German.

The following subsection on theoretical considerations will begin by defining the use of terms such as stress, accent and intonation phrase. Then, the question of the 'accentual cut' will be discussed; in principle, an accent may be defined relative to the pitch movement that immediately precedes the accented syllable, or with respect to what follows it (and this is how 'accentual cut' is defined here). Previous studies of English and German have not always agreed on where the accentual cut should be made. This will be followed by a discussion of intonational phrasing. As outlined in Chapter 1, some studies of English and German intonation posit one level of intonational phrasing, but others posit two. Then the question of intonational phrase boundary specifications will be discussed. Finally, an outline of the basic AM system proposed for cross-linguistic analysis will be given. A discussion of practical considerations involving questions of analytic technique will conclude the chapter.

2.1 Stress, accent and intonation phrases

In the area of stress and accent, terminological confusion abounds. Especially stress is notoriously difficult to define, and the definition researchers subscribe to depends to some extent on which aspect of stress they investigate. The following comments will be brief, and are intended to define the terminology used in the present study. For more detail, see, for instance, Cutler and Ladd (1983).

Researchers investigating the metrical properties of speech may define stress as a linguistic system which allocates different degrees of prominence to different syllables. The English word *elocution*, for instance, may be described as having three different degrees of stress. The strongest beat falls onto the third syllable *-cu-*, the second strongest on the first syllable *el-*, and the second and last syllable are not stressed. The constraints governing the degrees of stress, the distribution of stress and its exact realisation differ from language to language. We may find that in British English, elocution has three degrees of stress, but in Singapore English, two levels at most appear to be discernible (Low, forthcoming). Moreover, in British English, stress is relatively variable, but in Czech, for instance, stress is fixed; words are nearly always stressed on the final syllable. Variations in stress assignment result in different languages being characterised by different speech rhythms. The rhythm of British English is determined to a large extent by strong beats falling on the stressed syllables of words, and continuous speech can be segmented into rhythmic feet which begin with a stressed syllable and continue up to the next stressed syllable (see Abercrombie, 1967 for rhythmic feet, and Couper-Kuhlen, 1983 for a study of English speech rhythm). In French, on the other hand, stress beats regularly occur on the last syllable of a prosodic constituent which is often larger than a single word. Cross-linguistic differences of this type have led researchers to suggest a difference between 'stress-timed' languages such as British or American English and 'syllable-timed' languages such as French. Experimental evidence supporting this distinction, however, is scarce. Also, there is evidence showing that a classification of languages into stress-timed and syllable-timed overgeneralises. For instance, Low and Grabe (1995) showed that the rhythm of British English differs substantially from that of Singapore English. In Singapore English, successive vowel duration are more nearly equal than in British English, giving the impression of syllable-timing.

Researchers investigating the intonational properties of speech also use the concept of stress, but in their work, the term is used somewhat differently. Following Bolinger's (1958) theory of pitch accent in English, they distinguish between three phenomena; (word) stress, (pitch) accent and intonation (Cutler and Ladd, 1983: 141). Word stress is defined as an abstract property of a word in the lexicon (e.g. we know that the second syllable of the word *around* is potentially the more prominent one); accent refers to pitch movement at stressed syllables in actual utterances (in *I said aROUND* vs. *around the CORner*), and intonation refers to the combination of pitch accent and other sentence level pitch features such as pitch direction at boundaries and the relative height of accent peaks.

Auditorily, a syllable may be defined as accented when it is (a) stressed and (b) pitch prominent (Nolan, 1984). Pitch prominence is achieved if one or more of the following holds:

(a) the syllable is spoken on a perceptibly moving pitch
(b) the syllable manifests a pitch jump
(c) the syllable marks a change in the direction of pitch movement (e.g. from level to rising).

Acoustically, word stress involves a number of parameters. A stressed syllable will have more extreme formant values, greater duration, a steeper closing phrase of the glottal waveform with results in greater amplitude and more high-frequency energy in the spectrum (see e.g. Laver, 1994). Accent, on the other hand, is cued primarily by fundamental frequency movement. Early experiments by Fry (1958) showed that fundamental frequency is the strongest cue to accent in English, followed by duration and amplitude. However, later work by Beckman (1986) suggests that a measure of 'total amplitude' (reflecting a combination of amplitude and duration measures) is a good correlate of the accented syllable. Finally, the overall rhythmic and accentual pattern of an utterance may also cue accent on a particular word (Grabe and Warren, 1995).

The potential prominence distinctions to which the acoustic manifestations of stress, accent and, additionally, syllable weight may lead to in speech are summarised in Figure 1 below, which is similar to one found in Bolinger's (1964) (see also Liberman and Prince, 1977, Bolinger, 1986, and Beckman and Edwards, 1994). At the lowest level of contrast (full vs. reduced syllable), a prominence distinction is made primarily by vowel quality[1], at the second level by stress, and at the highest level by accent. Also, the schema shows that prominence distinctions made by stress or accent are syntagmatic phenomena; a syllable is accented only in comparison to a syllable that is not, and a stressed syllable is stressed only because there are other syllables that are unstressed.

In the present study, accent will be defined auditorily as suggested by Nolan (1984). Stress is taken to be an abstract property of particular syllables which specifies, amongst other things, how intonation can be aligned with a text, namely, in English and German, pitch accents are aligned with stressed syllables. Auditory and acoustic contrasts between stressed and unstressed syllables are of interest only in as far as they relate to analysis of tonal structure.

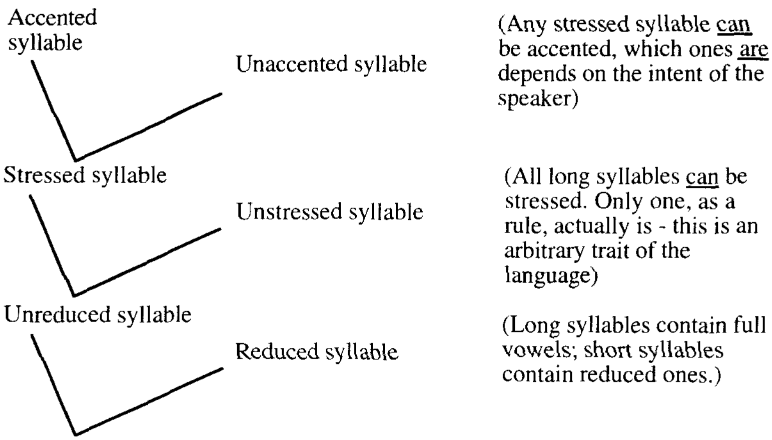


Figure 1 *Prosodic prominence hierarchy. Adapted from Bolinger (1964)*[2].

---

1 *See Fear, Cutler and Butterfield, 1995 for an experimental investigation of the strong-weak syllable distinction in English. The authors show that in production, unstressed unreduced vowels differ significantly both from stressed, full vowels and from reduced vowels. Nevertheless, listeners make a binary categorical distinction between strong and weak syllables on the basis of vowel quality, i.e. a syllable with a full vowel is classed as strong and one with a reduced vowel as weak.*

2 *Note that Bolinger (1964) refers to the unreduced / reduced syllable distinction as a long / short syllable distinction. This may be confusing, as 'long' and 'short' may be taken to refer to a phonological distinction in vowel length as in bat vs. bard rather than to a distinction in relative syllable prominence as in baton vs. butter.*

In one guise or another, the intonation phrase (IP) is a construct common to most studies of intonation (e.g. Trager and Smith's (1951) 'phonemic clause', O'Connor and Arnold's (1973) 'tone group', Crystals' (1969) 'tone unit', Pierrehumbert's (1980) 'intonation phrase', and Ladd's (1986) 'major phrase'). Ladd (1986: 311) points out that while there are differences of detail among these constructs, they share a number of properties. Firstly, they assume that IPs are the largest phonological chunk into which utterances are divided, and that the boundaries of this chunk may be phonetically specified. Secondly, an IP is assumed to have a specifiable intonational structure, including at least one accent. Finally, IPs are taken to match up, in some poorly understood way, with elements of syntactic or discourse-level structure (for problems with this 'standard' definition of the intonation phrase, see Ladd, 1986).

Cruttenden (1986: 36) points out that most analysts assume that the phonetic correlates of boundaries between intonation phrases can be determined much more straightforwardly than is really possible. No single auditory or acoustic correlate is available, and characteristics tend to involve different combinations of features from a bundle of acoustic and perceptual boundary signals. Boundary features include discontinuities in pitch between sections of utterance (frequently between major syntactic constituents, and in read speech often observable when there is punctuation), pauses, phrase-final lengthening and a slowing-down of speaking rate. Also, discontinuities in pitch in the absence of stressed syllables can be interpreted as evidence of boundary tones, and pattern repetition can provide evidence of phrasing; often, one finds that the patterns of larger chunks of utterances are repeated, for instance in lists or coordination structures, and such repetitions may be taken to indicate the presence of intonation phrase boundaries. With inexperienced readers and in spontaneous speech, however, one cannot expect to be able to identify all intonation phrase boundaries with a similar degree of certainty. In practice, Cruttenden points out, several phonetic cues or none at all may be available. The assignment of intonation phrase boundaries is therefore bound to be somewhat circular. We establish those cases in which boundary location is relatively clear, and note the internal intonational structure occurring in such cases. These internal criteria then help us to make decisions in cases where the external criteria are less clear-cut. In difficult cases, we may even resort to grammatical or semantic criteria. Thus, Cruttenden argues that IP boundaries cannot always be determined with any degree of certainty, especially in spontaneous speech. Accordingly, this first autosegmental-metrical comparison of English and German is based on read, rather than spontaneous speech (see section 2.1 in Chapter 3 for a description of the materials). In read speech, the identification of intonation phrase boundaries tends to be easier to determine than in spontaneous speech, because readers will be guided by punctuation provided in the written text.

2.2 The question of the 'accentual cut'

Drawing up a basic autosegmental-metrical system for cross-linguistic comparison requires some theoretically motivated choices about the internal structure one assumes pitch accents to have. One needs to decide on the 'accentual cut', that is, the section of speech accompanying the stressed syllable that one takes to reflect the realisation of an intonational category. Here, in principle, all models of intonation have three choices, and in previous studies of German and English two of the available options are employed[3]. The first group of authors assumes that accents are left-headed, and in that case, the relevant section of contour begins at an accented syllable and continues up to the following accented syllable (e.g. Gussenhoven, 1984 and Ladd, 1986 for English and Uhmann, 1991 and Féry, 1993 for German). In models which assume that pitch accents are left headed, the first element of a bitonal pitch accent is marked with a star and followed by an unstarred 'trailing' tone. House (1995) points out that left-headed accents are traditional in the British school of intonation analysis (e.g. O'Connor and Arnold, 1973, Crystal, 1969, Cruttenden, 1986). The choice of left-headed accents in English and German is not unrelated to the rhythmic structure of these languages; in both languages, rhythmic feet are left-headed (e.g. Selkirk, 1982)

A second group of authors has opted for a mixed-headed approach, which allows both right- and left-headed accents (e.g. Pierrehumbert, 1980, EToBI, GToBI). Here, accents have trailing or leading tones, and this proposal contrasts sharply with the view taken on the accentual cut in the British school. In the British school, a pitch accent may be associated with the head of a stress foot (Abercrombie, 1964) but in a mixed headed system, an accent with a leading tone crosses a foot boundary. Grice (1995a, b) offers an account which offers a possible reconciliation of these positions. Grice suggests a more complex internal structure for the pitch accent than other mixed-headed approaches do. The structure she proposes for the pitch accents resembles that of the prosodic word in Nespor and Vogel (1986), and is illustrated in Figure 2. In Grice's pitch accent, leading tones, which may cross a foot boundary, appear under the weak supertone node. The strong supertone node dominates tones corresponding to the nuclear tone in the British Tradition, and Gussenhoven's (1984) and Ladd's (1986) pitch accents.

---

[3] *The third option, which is not discussed in the text, is to propose that all accents are right-headed. In that case, the relevant section of contour is assumed to precede and include the stressed syllable, but as far as I know, no exclusively right-headed approach has been suggested within an autosegmental analysis of intonation for any language so far.*

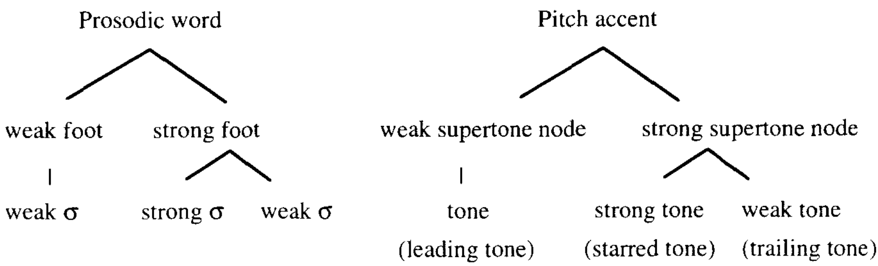


Figure 2 *The structure of the pitch accent in Grice (1995a, b).*

Note, however, that despite the apparently potentially tritonal structure pitch accents have in Figure 2, the accents which this structure generates must be either right-or left-headed; tritonal accents are not permitted. Therefore, to avoid tritonal accents, a constraint is required, stipulating that for English, either the pitch accent node or the strong supertone node branches.

A similar account is suggested in House (1995). House suggests a pitch accent structure essentially identical to Grice's, but unlike Grice, who posits only monotonal and bitonal accent, House also allows for tritonal accents. However, House does not state how the generation of right-headed accents is prevented in her pitch accent structure, and again, constraints are needed. The issue may be resolved by assuming that the minimal structure of an accent is not monotonal, as House assumes, but left-headed and bitonal[4], as shown in Figure 3 below. Taken together, the minimal pitch accent structure in Figure 3 and the maximal structure in Figure 2 ensure that the notion of left-headedness is preserved, that leading tones differ from trailing tones, and only left-headed accents are generated. As House states, a potentially tritonal pitch accent structure of the type she suggests allows us to capture useful generalisations and natural class-characteristics amongst related contours. This is more difficult in a mixed-headed approach where accents must be left- or right-headed.

---

4 *All nuclear accents are assumed to be underlyingly bitonal in Gussenhoven's (1984) analysis of English and Féry's (1993) analysis of German.*

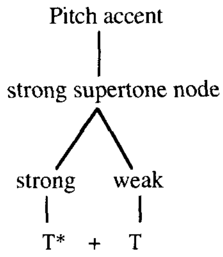


Figure 3 *Minimal structure of the pitch accent assumed in this study.*

As pointed out above, the autosegmental models of German drawn up by Uhmann (1991) and Féry (1993) present a left-headed account of pitch accents, as do Gussenhoven's analyses of English (1984) and Dutch (1988, 1992). The similarities between English, German and Dutch rhythmic and tonal structure (in all three languages, stress feet are left-headed) suggest that German pitch accents are indeed likely to be best portrayed as left-headed, with a pitch accent structure similar to Grice's (1995a,b) and House's (1995) accounting for leading tones. This was the view adopted here.

2.3 Intonational phrase structure

In Chapter 1, it was pointed out that the models of intonational phrasing proposed in Ladd and Beckman and Pierrehumbert (1986) involve reasonably similar two-level intonational phrase structures for English but differ in why we should need more than one level of phrasing. Ladd's account is motivated by the distribution of prosodic cues to phrasing - in his view, a sentence with two nuclear accents without an audible prosodic break in between is best represented as two minor intonational phrases embedded in one major phrase. The problem with this view is that in spontaneous speech, major intonational phrases are not necessarily delimited by audible prosodic breaks either. Beckman and Pierrehumbert point out that IPs should be able to have more than one phrase accent (in effect: more than one nuclear accent), and that there appears to be greater cohesion between intermediate phrases than between intonational phrases. However, Beckman and Pierrehumbert do not address the question of why there is a sense of greater cohesion between intermediate phrases, and present as two separate issues the matter of greater cohesion and the fact that intermediate phrases appear to capture similarities in tonal structure.

The discrepancies in motivation between Ladd's and Beckman and Pierrehumbert's accounts may suggest that the authors are describing different two-level phrase structures, but this seems unlikely. Both models offer intuitively convincing reasons for proposing an additional level of phrasing, their reasoning is not incompatible and their differences are not fundamental. Therefore, if their models describe the same phonological construct, then why the discrepancies in motivation and defining characteristics? And why are there no compelling reasons for choosing one model over the other?

In the present study, it is suggested that this is because the two models address different subsections of the same question, and this is why neither model accounts comprehensively for the distinctions which apparently characterise intonation phrases in English. Earlier work on intonation within the British school, specifically that of Trim (1959, 1988) and Crystal (1969) appears to suggest a potentially more successful way of dealing with the evidence. Some of Trim's and Crystals comments suggest that the reasons Ladd and Beckman & Pierrehumbert put forward for proposing minor tone units are in fact part of the same phenomenon: 'Tone Unit Dependency' .

In 1969, Crystal pointed out that researchers rarely acknowledge that tone units do not exist in isolation, but happen in sequence in connected speech. Because researchers tend to ignore this, there is a wide gap between what we know about the intonation of isolated phrases and what we know about the prosody of connected speech. The source of this problem, Crystal says, is a fundamentally false assumption about the nature of connected speech, namely that intonation is purely additive, that one can join up independently acquired tone units and in this way create normal utterances. Crystal's point is illustrated by some of the attempts that have been made to incorporate prosody into speech synthesis - one source of unnaturalness stems from the fact that connected speech is frequently made up from individual tone units with default intonation contours (Prevost and Steedman 1994). In fact, it has long been clear that accent patterns in successive tone units relate to one another (e.g. the given/new distinction, Nooteboom and Kruyt, 1987). Work of scholars such as Palmer (1922) who distinguished between co-ordinating and sub-ordinating sequences of tone units and Schubiger (1953) who noted that in complex sentences, the choice of accent patterns in successive tone groups is not free, motivated Trim (1959) and later Crystal (1969) to suggest structural dependency relations between successive tone units. These dependencies solve a number of problems in intonational analysis. Crystal noted tonal collocation between tone units, i.e. the repetition of the same nuclear pitch accent. This led him to suggest the theory of tonal subordination, a structural relationship between successive tones which accounts for stronger or weaker cohesion between them (first mentioned in Crystal and Quirk, 1964). The theory of tonal subordination relates to Beckman and Pierrehumbert's comments about subjectively felt greater cohesion between minor phrases. Trim's system, on the other hand, explains the behaviour of intonational tags (e.g. reported speech tags, or vocative tags) by allowing for anuclear tone units, defined as strongly dependent ('cliticised') on the immediately preceding tone unit.

From Trim's article and Crystal's work we can derive three kinds of dependency which structure tone units into two levels of intonational phrasing. We find the strongest level of dependency between anuclear tags and the preceding tone unit, where the pitch movement of the tag depends on that of the preceding nuclear accent; one might call this an asymmetric dependency. At a lower level of dependency, we find tonal collocation, where a pitch accent pattern is repeated. This relationship is symmetric, as it involves two tone units of the same type, i.e. with the same (nuclear) accent. The third structural relationship characterises independent tone units; there is no dependency.

Tone Unit Dependency Hierarchy

increasing dependency ↑

Type of dependency	Tone units involved
Asymmetric	Nuclear and anuclear
Symmetric	Nuclear, tonally collocated
None	Independent

Figure 4 *Tone unit dependency hierarchy in English.*

In the present study, it is suggested that the tone unit dependency hierarchy in Figure 4 explains Ladd's and Beckman and Pierrehumbert's intuitions about intonational phrasing in English. Symmetric dependency accounts for apparent mismatches between rhythmic and tonal structure. It explains why we feel that a traditional intonation phrase has two sub-components if it has two nuclear accents - this is because it does, in fact, consist of two units of phrasing, but the dependency between the units has integrated them into one larger unit. This is why we feel that there is some sort of cohesion between intermediate phrases within an intonation phrase. Asymmetric dependency explains why intonational tags are licensed to have a rhythmic break on either side. This is because the strong tonal dependency keeps the prosodic phonological structure intact, despite the rhythmic break.

Assuming a tone unit dependency hierarchy means that there is no need to propose that English has more than one kind of intonational phrase. In principle, the intermediate phrase falls out from Crystal's theory of tonal subordination; intermediate Phrases are successive tone units characterised by symmetric structural dependency. Different degrees of structural dependency result in perceived distinctions between intonational tags, intermediate phrases / tone groups, and independent phrases.

Evidence for the intermediate phrase in German is scarce. Uhmann (1991) assumes only one level of intonational phrasing, and Féry's (1993) proposal is not worked out in detail. GToBI assumes two levels of phrasing, but again, detailed auditory and acoustic evidence for this proposal is not yet available. The tone unit dependency hierarchy which appears to explain a number of facts about intonational phrasing in English combined with the lack of evidence for the intermediate phrase in German suggest that an AM system assuming one level of phrasing is more likely to be suitable for a first AM comparison of the two languages than one assuming two levels. This is the approach taken in the following chapters.

2.4 Intonation phrase boundary specifications

The approach the present study takes towards intonation phrase boundary specifications will be discussed next. Generally, AM systems following the Beckman-Pierrehumbert approach assume that each intonation phrase must consist minimally of a pitch accent, a phrase accent and a final boundary tone (whether initial boundary tones are obligatory, is not always equally clearly stated). A number of other authors, however, have suggested, more or less explicitly, that low boundaries may not need to be tonally specified. Bing (1979: 126) and Ladd (1983a: 745), for instance, analyse vocative chants and other stylised contours as not having a final boundary tone, and Ladd explicitly doubts that every audible prosodic boundary must be associated with a tone (1983a: 729). Lindsey (1985: 53) discards the low boundary tone for English altogether. Whenever there is no evidence of a high boundary tone, he takes low pitch to be the default case in standard British and American and argues that low pitch is inserted phonetically rather than by phonological rule. Cabrera-Abreu, 1994 does not specify low boundaries in her analysis of English either (note, however, that Cabrera-Abreu argues that we need not specify low in general). In her analysis of German, Féry (1993) motivates the lack of a low boundary tone delimiting her intonation phrase with the absence of downward tonal movement, and points towards an issue relevant to the discussion of whether all intonation phrase boundaries must have a tone: tonal structure is by no means the only acoustic correlate of phrasing. Grønnum (1992) has commented on the lack of convincing evidence for the existence of a phonological category L% in standard British English, and a phonological analysis without L% appears to be supported by a number of studies which have shown that phrase-final low boundary tones can take on some speaker-specific default value (e.g. Liberman and Pierrehumbert, 1984). This may be taken to suggest that L% may not be an independently chosen phonological category. If low boundaries reflect a default rather than an independently chosen phonological category, then the specification L% would have a somewhat different status from all other tones in the phonological inventory. All other tones are commonly assumed to represent 'active' choices on behalf of the speaker.

Gussenhoven's (1984) phonological analysis of Southern British English, on which the system proposed here is based, does not make use of a low boundary tone. In later work, however, Gussenhoven and colleagues (Gussenhoven, 1991, van den Berg *et al.* 1992), add to Gussenhoven's system a further intonational domain above the level of the IP, the 'scaling domain' (SD), which is equivalent to the utterance, and this domain may be delimited by a low boundary tone. In a system operating with the IP and the SD, then, an IP which is SD final may be delimited by a high or a low boundary tone, but an IP which is SD internal can only be specified with a high boundary tone[5].

The view proposed in the present study is that boundary tones may be language and dialect-specific. Consider, for instance, the realisation of IP boundaries in different varieties of English. Pierrehumbert (1980) has shown that low IP boundaries (H*L-L%) do not exhibit clear downward movement of F0 at the phrase boundary. The fundamental frequency trace from an utterance produced by a Northern Irish English speaker in Figure 5, however, does exhibit downward F0 movement at the phrase boundary (Nolan and Grabe, 1997)[6].

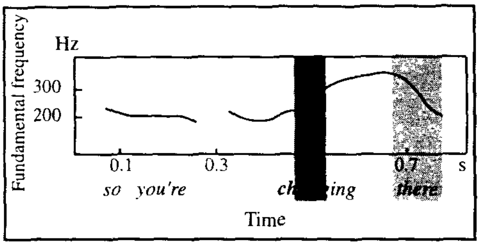


Figure 5 *Adapted from Nolan and Grabe (1997).*

The dark grey section in Figure 5 indicates the location of the accented syllable, and the light grey section the pitch movement at the phrase boundary which takes place in the absence of a stressed syllable. Accounting for this type of pitch pattern in a system such as Pierrehumbert's, which posits obligatory high and low boundary tones is not straightforward. The obvious transcription L*+H H- L% is not available, because Pierrehumbert's upstep rule raises the final L to the level of the preceding H. One might, of course, posit the absence of an upstep rule for Northern Irish English, but then the transcription would (a) no longer model the cross-linguistic difference and (b) no longer be able to capture the pattern L*+H H% with upstep, should such a pattern exist in

---

[5] *The system proposed in this study follows Gussenhoven as far as the IP; the investigation of acoustic and auditory cues to intonational phrasing above IP level lies outside the scope of this study.*

[6] *Figure 5 is based on data from an corpus analysis of Northern Irish English carried out by Lowry (1997).*

Northern Irish English (see also Ladd, 1996: 145 for a similar point concerning Glasgow English).

If we assume, however, that IP boundaries are not obligatorily associated with a boundary tone, the apparent dilemma can be solved relatively easily. One may posit that Northern Irish English has a boundary tone L% but the variety of American English which Pierrehumbert analysed does not.

2.5 Basic AM system proposed

This section summarises the AM system used for cross-linguistic comparison in the following chapters. Its basic characteristics are the following:

(1) All accents are represented as left-headed.
(2) Only one level of intonational phrasing is indicated (the intonation phrase).
(3) Phrase accents are not assumed to be needed.
(4) Intonation phrase boundaries can be left tonally unspecified.
(5) The system has two levels of phonological representation, in addition to one level of phonetic implementation.

The basic pitch accent inventory contains two bitonal pitch accents, which correspond to falling and rising nuclear tones in the British Tradition. These are the tones which all previous studies of English and German intonation have posited for the two languages, and they will be represented as H*+L and L*+H. The inventory of boundary specifications and phonological adjustment rules which mediate between underlying and surface levels of phonological representation will emerge from the corpus analyses presented in Chapters 3 and 4.

This section will conclude with some brief comments on the intonational terminology used in the following sections in this study. As Grice (1995a) points out, within the British school, some inconsistency may be observed regarding the use of the term 'nucleus'. The term has been applied to either the last salient pitch movement in an IP (i.e. starting on a stressed syllable and continuing up to the end of the IP) or to the syllable rendered accented by that particular pitch movement. This ambiguity will be avoided here by referring to the accented syllable as the 'nuclear syllable' and to the complete pitch movement starting on it and continuing up to the IP boundary as the 'nuclear tone'. What exactly the term 'nucleus' refers to in the AM approach appears to be somewhat unclear also. It may refer (a) to the last starred element in a phrase, (b) the last pitch accent in the phrase, whether bitonal or monotonal or (c) to the last pitch accent plus following boundary tone. Here, the terms will be used as follows. The last starred element in the intonation phrase is associated with the 'nuclear syllable'. 'Nuclear tone' refers to the last pitch accent in the phrase plus following boundary specifications. The term 'nuclear accent', however, will be used also, and this will refer to the last pitch accent in the phrase without boundary specifications. Thus, for instance, L*+H H% transcribes the nuclear tone, L*+H the nuclear pitch accent and L* is associated with the nuclear syllable. The British system does not recognise a division into pitch accents and boundary tones, and thus, here, only the terms 'nuclear syllable' and 'nuclear tone' correspond to AM tonal constituents. However, for the purposes of this study, some terminological parallelism appears desirable. Therefore, the AM use of the term 'nuclear accent' defined here will be taken to correspond to the last 'simplex' accent in the IP in the sense of the British Tradition, that is, for instance, the fall in a fall-rise (for simplex vs. complex nuclear, cf. e.g. Cruttenden, 1986: 58). However, this is not the way this term is used in the British school.

However, despite the obvious differences between the British model and the AM approach, there are also points of convergence. Roach (1994), for instance, discusses to what extent the intonational categories of the British school may be expressed in 'ToBI', an AM prosodic labelling system (Silverman *et al.* 1992, Beckman and Ayers, 1994). Specifically, it appears that the auditory phonetic percepts which the British school describes as a 'fall' and a 'rise' and the AM system as 'a high pitch level on a stressed syllable followed by a low pitch level' and 'a low pitch level on a stressed syllable followed by a high pitch level' refer to the same intonational category, that is, falling or rising pitch either on or immediately following a stressed syllable. Considering the range of possible transcriptions AM systems seem to offer for a what may be referred to simply as a 'fall' or a 'rise', and considering that one may, at times, wish to refer to the auditory percept of an intonational category without committing oneself to a specific AM representation, it seems reasonable to assume that auditory labels such as 'fall' or 'rise' may be used alongside AM transcriptions. This is the approach followed in this study. However, when the terms 'fall' and 'rise' are used, the aim is to refer theory-neutrally to the auditory percepts of the pitch events discussed rather than to invoke the theoretical framework proposed in the British model.

3 Practical considerations

3.1 Analytic techniques

Crystal (1969: 7) discusses the different senses in which the term 'analysis' has been used in linguistic research. For instance, 'analysis' may refer to auditory analysis, to articulatory analysis, instrumental analysis, statistical analysis, structural description or phonological analysis., and at times, this can be confusing. Crystal defines his use of 'analysis' as 'the explication of the non-segmental contrasts perceived [in his data] as meaningful by postulating a set of prosodic systems within which they may be defined and interrelated'. The specific method used to arrive at the end product of such an analysis (e.g. auditory, instrumental etc.) is referred to as an 'analytic technique'. Although the present study is carried out within a phonological framework different from that used by Crystal (1969), the essence of his view of analysis is adopted here The purpose of the present analysis was to establish a set of intonational categories which may be classified as capable of conveying differences in meaning. The analytic techniques adopted will be discussed in the following sections.

3.2 F0 as a narrow phonetic transcription?

Beckman (1995) suggests that one may analyse an intonational system by using the F0 contour as a 'narrow phonetic transcription', combined with careful listening and drawing of stylised contours (which, presumably, combine acoustic information and auditory impressions). She advocates the use of a transcription system such as, for instance, ToBI only when the analyst knows what the phonologically different categories in the language in question are. If one is not completely sure, then one should not begin by using a symbolic 'narrow phonetic transcription' but rather do the 'real work' first by carefully observing the F0 trace and establishing the categories. As no previous intonational investigation of the specific variety of Northern Standard German analysed was available, and as it was unclear whether the intonational categories established for other varieties of German were directly transferable, Beckman's comments were taken as a pedagogical guideline and careful listening was supplemented with an examination of F0. The view of F0 as a narrow phonetic transcription of intonation, however, was not adopted. The reasons for this were the following. Firstly, it is difficult to accept that F0 may function as a 'narrow phonetic transcription' because F0 represents more than an acoustic correlate of intonational categories. It also contains evidence of other aspects of phonetic structure, for instance of microprosodic variations caused by voiceless obstruents. This means that researchers using F0 as a guideline cannot use all of the information available, but rather need to use it selectively. F0 is subject to microprosodic variations which reflect segmental rather than prosodic structure. Thus, researchers need to know about the interaction of F0 and segmental structure and in some way 'filter' out the latter. Although F0 represents an acoustic correlate of pitch, it does not represent pitch exactly. A narrow phonetic transcription, on the other hand, claims to be rather more exact. Moreover, it implies discrete phonetic categories, but F0 as such is continuously variable. Secondly, F0 represents less than a 'narrow phonetic transcription' of intonation would. As is well-known, the acoustic correlates of accent involve more than pitch, which has F0 as its acoustic correlate; length (duration) and intensity (amplitude) are relevant also, even if pitch is often the most salient correlate of accent. This means that the F0 track reflects only part of the acoustic information that an auditory analysis uses.

In summary, an approach to intonation analysis which concentrates on F0 appears to be too inclusive of irrelevant detail and too exclusive of acoustic correlates other than F0 which contribute to the auditory impression of intonation. As Crystal (1969: 14) points out, the analyst needs to find a middle way; a compromise between a purely acoustic and a purely auditory method. Accordingly, the corpus analysis presented in the next two chapters was based on auditory analysis combined with supplementary reference to F0. Differences in length and intensity which form an intrinsic part of the overall auditory impression of an accent pattern, and their acoustic correlates duration and amplitude, however, will not be addressed. This restriction is motivated by the nature of the speech data analysed; corpus data are less well suited to establishing relative differences in duration and amplitude and better suited to establishing interactions between F0 and segmental structure. Also, arguably, F0 is a fruitful acoustic phenomenon to concentrate on, as it has been shown to be the most salient correlate of accent (Fry, 1958)[7]. As will be described in Chapter 3, the auditory analysis was carried out by systematic comparisons of intonation patterns produced by different speakers in identical contexts and by the same speakers in different contexts, and the categories established in the auditory analysis claim to have phonological status. F0, on the other hand, was assumed to be no more than a continuously variably acoustic record of the main perceptual aspect of intonation; that is pitch was not assumed to have phonetic status as such.

[7] *Note that this is an interpretation of Fry's results. Fry investigated cues to the location of lexical stress, and found F0 movement to be the most salient cue.*

3.3 Auditory technique

In the previous section, the use of F0 as a narrow phonetic transcription of intonation was rejected, and the use of a combined auditory / acoustic technique was advocated. The term 'auditory', however, requires some further discussion and definition. Crystal (1969: 14) points out that the term 'auditory' is not particularly clear; it may mean either 'auditory sensation' or 'auditory interpretation'. In what follows, this issue will be discussed with reference to two concepts discussed in 't Hart, Collier and Cohen (1990); these are 'perceptual equality' and 'perceptual equivalence'. Both are involved in an auditory analysis of intonation. Perceptual equality, which relates to 'sensation' refers to arguably involuntary listening processes. Perceptual equivalence relates to 'interpretation', that is, to linguistic decisions made by the analyst on the basis of pitch changes assumed to be the product of voluntary actions on the part of a speaker. 'Perceptual equality' will be discussed first.

In perception experiments carried out by 't Hart, Collier and Cohen (1990), naive listeners judged a resynthesised utterance with a close-copy stylisation of F0 to be perceptually equal to the same resynthesised utterances where F0 remained unchanged. The authors argue that this is so because close-copy stylisation removes microprosodic fluctuations from F0 which are not produced voluntarily by the speakers and therefore not part of the message communicated. The changes in intonational structure which the speaker produces intentionally, on the other hand, are kept intact. Although I do not want to argue that close-copy stylisation is what happens in a researcher's mind when he or she analyses an intonation contour (for instance, as the authors point out, at times, differences in intrinsic pitch CAN be heard), the fact that close-copy stylisations were shown to be perceptually equal to those with original F0 contours allows us to relate the concept of perceptual equality to auditory analyses of intonation. Listening to an intonation contour involves in some way an involuntary filtering out of microprosodic detail in F0[8]. In Crystal's (1969) terms, the sense of 'auditory' relevant to perceptual equality involves auditory sensation rather than interpretation.

The second concept introduced in 't Hart, Collier and Cohen (1990) is 'perceptual equivalence'. This concept is relevant to the perception of voluntary changes in intonational structure made by a speaker, and the interpretation of these changes. A listener carrying out an auditory analysis needs to decide whether two contours are of the same type or not. 't Hart, Collier and Cohen define perceptual equivalence as follows: 'if for a speech utterance two different courses of F0 are similar to such an extent that one is

---

8 *The perception of duration and amplitude involve other mental processes, which are also relevant to the auditory impression of intonation, but as F0 is the acoustic correlate of intonation this study concentrates on, these processes are not considered any further here.*

judged as a successful imitation of the other, we say that there is perceptual equivalence between the two.' Relevant to auditory analysis is the notion of 'successful imitation' (and a successful imitation of an intonation contour is something that not only phoneticians but most naive native speakers can produce and judge). The assumption is that if a contour represents a successful imitation of another contour, but is produced on different lexical material, then it is reasonable to assume that the two contours are of the same type. In the present study, 'being of the same type' means that the contours are assumed to have the same phonological structure. However, to avoid misunderstanding and to show that in this study the angle from which the concept of perceptual equivalence is looked at is somewhat different from that in 't Hart, Collier and Cohen (1990), 'perceptual equivalence' is replaced by 'auditory phonetic equivalence'.

A third concept which may be added at this point is that of 'auditory relatedness'. This concept relates to the question of phonological distance between contours which are modelled as categorically different, and is harder to define than auditory equality and equivalence. Analysts feel that there are differing degrees of phonological distance between contours, grouping together contours which are (a) structurally similar and (b) do not obviously differ in meaning. These are the minimum requirements of 'auditory relatedness'. 'Auditory relatedness' is to do with the idea that there are natural classes of intonation contours. 't Hart, Collier and Cohen's (1990: 50), for instance, refer to such natural classes of contours as 'melodic families' and House (1995) talks about 'families of contours'.

The notion of grouping intonation patterns has been a concept in the British school of intonation analysis for some considerable time. For instance, we find it in O'Connor and Arnold's (1973) 'tone groups'[9]. The authors state that, in principle, if one combined all the parts of tunes which they recognise in their analysis of colloquial English, one would find that the total number of possible pitch patterns in English is 105. However, this is not realistic because some meaning differences between patterns are so slight that they would be difficult to define in any very helpful way. Then the authors define as members of a tone group all those tunes that share one or more pitch features and convey the same attitude on the part of the speaker. This approach would appear to be similar to that of Gussenhoven.

9 *Within the British school, 'tone group' is more commonly used to refer to the intonation phrase.*

4 Speech data: A directly comparable corpus of German and English read speech

4.1 Introduction

In Chapter 1, the findings of previous cross-linguistic studies of English and German intonation were outlined. The discussion of the literature showed that, at times, researchers have disagreed strongly about how similar or different English and German intonation might be. Three reasons for this disagreement were suggested. Firstly, researchers compared the languages in analytic frameworks which were not directly comparable or had been drawn up on the basis of one language and had then been transferred to the other without prior analysis of that second language as a system in its own right. Secondly, some comparisons failed to distinguish clearly enough between phonetic and phonological levels of analysis and did not consider that the languages might be similar at one level but different at another. Finally, researchers did not work on directly comparable samples of speech, and some might have compared quite different speaking styles.

This study compares English and German in the autosegmental-metrical framework, which distinguishes explicitly between different levels of intonational representation. As the two languages have not yet been described in the same variant of the autosegmental-metrical framework, an a basic system for comparison was drawn up for comparison in the preceding sections of the present chapter. The remaining issue, that is, the question of what samples can be fruitfully compared, is discussed in the following sections.

The corpus of English and German speech data compared in this study contained read speech. For a first comparison of the intonational structures of two languages, read speech is useful because it allows a relatively constrained elicitation of intonation patterns; the speaker's prosodic options are limited by syntactic structure and guided by punctuation, and speaking rate is slower and usually less variable than in spontaneous speech. Moreover, intonation phrase boundaries may be determined with some degree of certainty.

The aim in setting up the corpus was to obtain directly comparable, orthographically transcribed and intonationally labelled German and English speech data with time-aligned fundamental frequency traces. The analysis was carried out using waves(tm), an Entropic Research Laboratory product, in conjunction with the 'transcriber' script which is part of English ToBI (Silverman et al., 1992; Beckman and Ayers, 1994). The script displays a speech wave and a time-aligned fundamental frequency trace plus a number of empty labelling templates where intonational transcriptions as well as other information may be entered. Time-aligned spectrograms which are needed to establish exact alignment of fundamental frequency trace and segmental structure can be generated using waves(tm). The original ToBI labels, however, were not used, and the tone labels in the transcriber script were replaced by labels reflecting the basic AM system developed as a starting point for cross-linguistic comparison.

4.2 Materials

When speech data for intonation analysis is elicited, constraints on subjects' interpretations of experimental materials are desirable. Cross-speaker and cross-language comparisons are facilitated when the number of different patterns produced by different speakers in identical contexts is limited (the underlying assumption being that speakers' choices of specific intonation patterns are context-dependent). The materials used to elicit the corpora collected for this study were based on Grimm's fairy tale 'Little Red Riding Hood', which is equally well known in Great Britain and Germany, and a more recent, English version of the same story (Langely, 1992). Using a well-known story ensured that subjects would interpret the materials similarly. Also, fairy tales tend to be produced in a fairly standardised speaking style, which is very suited to intonation analysis. Because they are read to children, they are produced at a moderate speed, and, just as in child-directed speech, pitch excursions are relatively large. This makes it easier to analyse the speech auditorily and to investigate the alignment of F0 with segmental material. Also, fairy tales cover a wide range of emotional states and are therefore likely to elicit a wider range of intonation patterns than materials consisting, for instance, of isolated sentences. Lastly, some of the traditional repetitions which occur in Grimm's fairy tales (e.g. here: *All the better to <u>hear</u> you with!* [..] *All the better to <u>see</u> you with! All the better to <u>eat</u> you with!*) are useful because one can examine the perceptual and acoustic aspects of equivalent intonation pattern aligned with different stretches of segmental material.

The English and German versions of the fairy tale were re-written to maximise their suitability for the purpose of this study (see Appendix A). Firstly, the content of the stories and the story line were kept as similar as possible. Secondly, some high frequency words with a low proportion of sonorants were replaced by words with a higher proportion of sonorants so that F0 traces would be less interrupted (for instance, the words 'Rotkäppchen' and 'Little Red Riding Hood' which contain a relatively large proportion of non-sonorant segments were replaced by 'Anna', which, in this particular version of the fairy tale, was supposed to be Little Red Riding Hood's real name). Thirdly, the syntactic structure of the stories were kept as similar as the languages would allow, and a wide variety of syntactic constructions and discourse features were included (e.g. syntactic tags, appositions, coordination structures, reported speech, direct speech, vocatives, appositions)[10]. The aim was to elicit as wide a variety of intonational structures as possible within a relatively short, coherent story. The stories are given in full in Appendix A.

4.3 Elicitation

Five German and five English subjects produced the materials. The German recordings were made in a quiet room at a secondary school in Braunschweig; the English recordings in a soundproof booth at Cambridge University. The data was recorded on DAT tape on a Sony TCD-D3 DAT recorder with a Sony Electret Condenser microphone 737.

4.4 Subjects

The German recordings were made at the Realschule Maschstraße in Braunschweig in northern Germany. Five female speakers aged between 16 and 18 were recorded. All had been born in Braunschweig, and so had their parents; they were attending the same school (a 'Realschule', a type of secondary school), and had lived in Braunschweig all their lives. Thus, one can reasonably assume that they spoke the same variety of Northern Standard German ('Hochdeutsch') and used the same intonational systems. Each recording session was started by asking the subjects to tell the experimenter some basic facts about themselves and their family background. The purpose of this was partly to put subjects at their ease and to familiarise them with being recorded (none of them had been recorded before), and partly to gather information about their language background and that of their parents.

For the British subjects, a similar degree of homogeneity was harder to achieve. Received Pronunciation ('RP', Wells, 1982), the variety of English comparable to 'Hochdeutsch; is largely found in southern England[11], but mobility in Britain appears to be higher than in Germany and class distinctions as well as multicultural influences are more clearly felt. Also, there is a stronger sense of social class than in Germany. The five female speakers taking part in the English recordings were undergraduates and postgraduates of Cambridge University, and aged between 19 and 24. They saw themselves as speaking RP, and this judgement was confirmed by an English phonetician; they were born in the south of England, and 'assuming there was such a

10 *Both versions were subsequently checked informally by native speakers of English and German who judged them to be 'native' English and German texts.*

11 *RP also functions as a prestige norm in the British Isles, and is widely spoken in other parts of the country. The relevance of Hochdeutsch as a prestige norm is less clearly felt in Germany (this is certainly true in the North).*

thing as class' rated themselves as middle or upper middle class. All of them had moved to different parts of southern England at some stage in their lives. Again, the recordings were initiated by collecting information about the speakers and their language background.

The data was digitised at 16 KHz on a HPA4032A in waves(tm) 5.0.2 under UNIX. The size of the corpora is as follows:

German		**English**	
Speaker	**Duration (min)**	**Speaker**	**Duration (min)**
JH	4.5	KP	3.8
JN	4.5	KS	4.4
MM	4.7	JS	5.9
NF	4.5	AT	4.8
SV	4.6	LC	5.0
Total	22.8	**Total**	23.9

Table 1 *Duration of German and English corpora.*

Table 1 suggests that all the German subjects read at approximately the same rate (no lengthy pauses occurred). For the English subjects, KP appears to have read somewhat faster than the others and JS appears to be slower. However, closer inspection of the data shows that these differences were not actually caused by differences in these speakers' articulation rate but rather by the durations of pauses; JS left long, dramatic pauses especially within dialogues whereas KP proceeded through the text more briskly.

4.5 Labelling

The data were labelled orthographically using the ToBI transcriber script. On the tone tier, the auditory impressions of intonational patterns were labelled using the following inventory:

(1)	**Pitch accents**	**Boundary specifications**		**Diacritics**
	H*+L	H%	%H	>
	L*+H	L%	%L	!
		0%	0%	

A pitch accent was transcribed as H*+L or L*+H when the trailing tone following the accented syllable appeared in the postaccentual syllable. If the trailing tone appeared to be realised later than the postaccentual syllable, a diacritic '>' was added and the accent was marked as H*+>L or L*+>H, with the '>' indicating displacement of the trailing tone to the right. Downstep was indicated by a '!' symbol preceding the downstepped tone. One level of intonational phrasing was indicated. Initial and final IP boundaries were labelled as H% when they exhibited upward pitch movement at the phrase boundary in the absence of a stressed syllable and as L% when there was downward pitch movement. Boundaries whose tonal specification did not differ from that of the immediately preceding trailing tone were marked as 0%. Note that '0%' is not assumed to reflect a phonological category but is a place holder indicating the end of an intonation phrase which does not appear to be associated with a tone. The label 0% was used rather than, for instance, the boundary inventory offered in German ToBI, the assumption here being that the labelling should reflect, as closely as possible, actual observations of pitch and F0. GToBI labels would not have reflected an absence of pitch movement at IP boundaries as straightforwardly as the labelling adopted here.

The break index labelling template was used to mark the vocalic sections within the stressed syllable of accented words. This was to allow within- and cross-language comparisons of fundamental frequency alignment on stressed syllables. The miscellaneous tier was used for notes and comments on intonational phrase structure.

4.6 Presentation of evidence

Pitch patterns may be illustrated visually in several ways. In the British tradition, for instance, some authors have illustrated their observations with so-called tadpole diagrams (e.g. O'Connor and Arnold, 1973). Tadpole diagrams depict different levels of prominence with smaller and larger dots and pitch movement by means of 'tails' following the dots. Figure 6 below shows an example of an intonation phrase with three rising prenuclear accents followed by a nuclear fall.

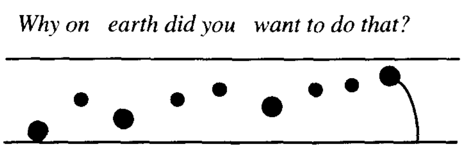


Figure 6 *Tadpole diagram. Adapted from O'Connor and Arnold (1973: 38).*

However, considering that some readers might find it difficult to assess to what extent a tadpole diagram can be taken as representative of any native speakers' perception of intonation rather than just that of the author's, and considering that relatively objective acoustic evidence in the form of F0 was available (even if F0 is clearly not equivalent to the perception of intonational structure), it was decided to illustrate the contrasts established in this study primarily with F0, and to arrange F0 traces to reflect the way in which the auditory analysis was carried out. Additionally, auditory evidence will be approximated via stylised contours which are similar to tadpole diagrams but provide some more information such as the association of an auditory pattern with syllable structure.

Many studies providing acoustic evidence of intonation illustrate the patterns they discuss with F0. However, it is not always possible to derive from such figures detailed information about the relationship between the trace and the associated text because no information is given about the alignment of the trace with the associated segmental material. In this study, an attempt was made to make the acoustic data more accessible by marking in each trace subsections of the accented syllable (in the first instance, this involved solely the vocalic portion, excluding onset and coda[12], but later, the complete syllable rhyme was marked[13]). Secondly, in the auditory analysis, each pattern produced in a specific context was contrasted with other patterns in two ways, and these comparisons are reflected in the F0 diagrams. On the one hand, a specific pattern was compared with patterns produced by other speakers in exactly the same context. This provided 'paradigmatic', cross-speaker information about the representative status of a contour, and the relevant F0 traces gave information about the alignment of this contour with segmental structure (as there were five speakers, and there were always five instances of a specific pattern). Then, the pattern was compared with apparently similar patterns produced by the same speaker in different contexts. This 'syntagmatic' comparison gave an impression of auditorily equivalent contours on different words. Figure 7 below illustrates the structure of the F0 displays which will be shown in the following section. The acoustic comparisons shown schematically in Figure 7 reflect the auditory comparisons which were carried out.

Figure 7 shows that in the displays illustrating the contrasts, F0 patterns are plotted on the same scale vertically (Hz). On the horizontal scale (time), the duration of utterances is normalised, that is, for all speakers, the same utterance is plotted as if it had the same duration (e.g. five renditions of the name *Anna* are aligned with each other by rescaling the F0 traces from speakers 2, 3, 4 and 5 to the duration of the trace from speaker 1). This means that the fundamental frequency patterns of utterances produced by different speakers are optimally comparable.

12 *In a small number of cases, where segmentation was hard to justify on acoustic grounds, preceding or following liquids or nasals were included; relevant cases are indicated in the text.*

13 *The syllable rhyme rather than the vocalic section was marked after the rhyme had been established as the relevant subsection of the syllable for the alignment of H*+L.*

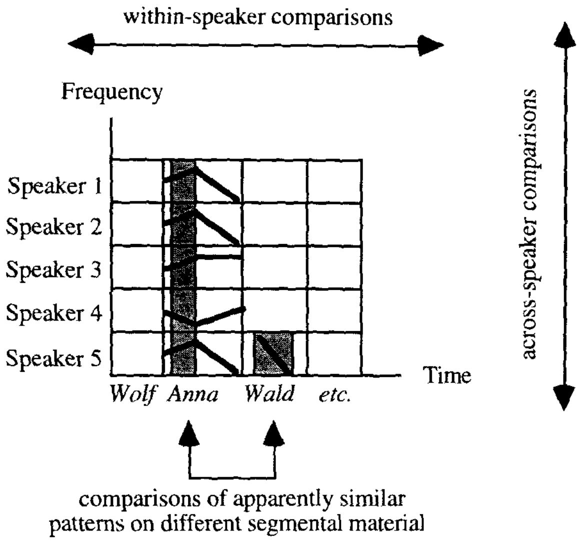


Figure 7 *F0 display of 'paradigmatic and 'syntagmatic' contrasts in the analysis.*

The displays were made as follows. First, speech wave and time-aligned fundamental frequency traces were displayed using waves(tm) in conjunction with the ToBI transcriber script (Beckman and Ayers, 1994). Then, sonorant portions of accented syllables were determined by inspection of the speech wave and time-aligned spectrograms and labelled. Subsequently, F0 traces for relevant sections of utterances were saved as segments and redisplayed in waves(tm), using the same window size for each section from each speaker to allow comparisons across speakers (note that these comparisons were not time-aligned). The markers delimiting the sonorant sections of accented syllables were displayed by attaching the relevant label file to the fundamental frequency window. The trace file was then saved as a '.tif' file using programs 'xwd' and 'xv' under UNIX and exported to a Macintosh Quadra 800. There, the file was redisplayed, F0 was retraced in Aldus Freehand 3.1 and the sonorant sections of the accented syllables were shaded in. Retracing the files permitted a more flexible data presentation, and saved disk space. Appendix C gives one comparison of original traces and retracings which shows that the match between originals and retracings is very close.

The approach to analysis presented in this chapter has the following advantages. At the auditory level, systematic comparisons of contours produced by different speakers in identical contexts help to establish those characteristics of a contour which are relevant to its identity. Also, information about potential speaker-specific preferences may be gathered. Comparing contours suspected to be equivalent produced by the same speaker on different lexical material helps to distinguish contours which are genuinely different from those whose differences result from systematic but purely mechanical effects of segmental structure.

Secondly, the approach allows a comparison of the choices different speakers make in identical contexts. In identical contexts, we may find evidence for natural classes of contours, which may then be contrasted with classes characterising other contexts. Evidence may be collected about auditory characteristics shared by families of contours, that is, related contours appearing in identical contexts which do not appear to differ substantially in meaning but which appear to be categorically distinct in their realisations (auditorily as well as in F0).

In the acoustic domain, the marked subsections of accented syllables allow comparisons of the alignment of F0 traces and segmental material within and across speakers and within and across languages. Marking, in the first instance, the vowel rather than the rhyme of the accented syllable or the complete syllable makes it possible to collect detailed information about segmental reference points of F0 alignment. At least theoretically, it is possible that F0 movements are sensitive, for instance, to the onset-rhyme distinction. Additionally, information is given about the extent to which F0 traces illustrating one and the same phonological category may vary within and between speakers, for instance, as a function of the structure and/or duration of the associated segmental material. This issue is relevant in a language such as German which appears to truncate accents on syllables containing a small proportion of sonorant segments (Grønnum, 1989).

5 Summary

The present chapter has discussed theoretical and practical considerations prior to the cross-linguistic comparison of English and German. First of all, the terminological confusion surrounding the terms *stress*, *accent* and *intonation phrase* was discussed and the use of these terms in the present study was defined. Next, the question of the accentual cut was discussed; some analysts have suggested that the accent inventory of English is best accounted for as exclusively left-headed (e.g. Gussenhoven, 1984), but others have posited a mixed-headed inventory (e.g. Pierrehumbert, 1980). In section 2.2 of the present chapter, it was argued that a left-headed inventory offers the most obvious starting point for the comparison of two languages in which rhythmic feet are left headed. Section 2.3 considered intonational phrase structure; an analyst needs to decide on how many levels of intonational phrasing he or she assumes English and German have. In the literature, one- and two-level structures have been suggested. In section 2.3 of the present chapter, an account of intonational phrasing was suggested which assumes only one type of phrase, the intonation phrase, but assumes a number of dependency relationships between intonation phrases. These dependencies are suggested to account more successfully for the phenomena which have led other authors to propose a distinction between the intonation phrase and the intermediate phrase. Intonation phrase boundary specifications were discussed next. Pierrehumbert (1980) assumes that every intermediate phrase boundary and every intonation phrase boundary must be specified with a tone. As a direct result, some of her boundary transcriptions are relatively indirect; they do not reflect the phonetic realisation of intonation phrase boundaries very straightforwardly. In this study, intonation phrase boundaries can, but do not have to be specified with a tone. In principle, an intonation phrase may be delimited by a rhythmic discontinuity such as a pause alone; a tone is specified only if there is tonal movement at the boundary (in the absence of a stressed syllable).

A discussion of analytic technique followed; specifically, in the analysis of intonation, should one rely primarily on acoustic analysis, or on auditory analysis, or should carry out a combination of both? The shortcomings of an approach relying largely on fundamental frequency were discussed, and a combination of auditory and acoustic analysis was advocated.

The final sections of the present chapter focused on the type of speech data suited to a cross-linguistic comparison of intonation within a framework not previously applied. A directly comparable corpus of read speech data was argued to be a felicitous starting point. The corpus materials designed for the purposes of the present study were discussed, and the elicitation method, the choice of subjects, the prosodic labelling of the data and the presentation of the evidence were described.

The following chapter will present evidence from Northern Standard German. In Chapter 4, the German data will be compared with data from Southern Standard British English.

Northern Standard German

Chapter 3

1 Introduction

The following sections present the results of the auditory and acoustic analyses of the German corpus. The presentation of evidence will begin with nuclear falling accents (a nuclear falling accent is defined as the last significant pitch movement in the intonation phrase falling from or on a stressed syllable). Next, prenuclear falls will be discussed. After that, rising accents will be considered, and again, nuclear and prenuclear rises will be described separately. Auditory phonetic impressions of falls and rises will be given, followed by acoustic evidence in the form of F0 traces. Specific attention will be given to the shape of accent patterns in F0 and their alignment with stressed syllables. Subsequently, the evidence in the corpus for phonological adjustment rules will be presented. After the presentation of accentual categories, phrase boundaries will be discussed. In the variety of German investigated, high boundaries exhibit pitch movement in the absence of a stressed syllable, but low boundaries do not appear to be accompanied by equivalent discontinuities in pitch.

The chapter concludes with the discussion of a specific problem in the analysis of German; that of 'rise-plateaux'. These are accent patterns involving rising pitch on the stressed syllable followed by level pitch. Two distinct types exist, which sound and look, at first sight, rather similar. This problem was raised previously by Féry (see section 2.2.3.4 in Chapter 1) as one involving two different types of 'hat patterns' and claimed to involve neutralisation. The present study shows (a) that the problem is not restricted to prenuclear position and (b) that despite their apparent similarity, the patterns can be distinguished on the basis of a number of auditory and acoustic characteristics.

2 Nuclear H*+L

In the following sections, evidence for nuclear falling pitch accents (H*+L) will be presented separately for accented words containing at least one postaccentual syllable and for IP-final monosyllabic words. The accent positions will be referred to as 'non-final' and 'final' respectively. The distinction is relevant to the present study because, at least in the variety of Northern Standard German analysed, fundamental frequency alignment on final and non-final accents differs substantially.

2.1 Non-final position

In the corpus of Braunschweig German, the majority of nuclear falls were produced as a relatively large, gliding rise in pitch on the accented syllable followed by a fall on the next syllable. This pattern is illustrated in Figure 1(a) below. A variant of the pattern shown in Figure 1(a) was observed, and is shown in 1(b). In 1(b), the accented syllable does not exhibit a gliding rise but more or less level pitch. Intuitively, however, 1(b) is closely related to (a). Figure 2 shows F0 traces illustrating a set of nuclear falls from the corpus.

Figure 1 *Alternative auditory phonetic impressions of a nuclear fall on* *<u>Zähne</u>* *'teeth'. The shaded boxes represent stressed syllables, the empty boxes unstressed syllables, and the contours model pitch movement*[1].

In Figure 2, the shaded sections in the F0 traces indicate the location of the vowel in the accented syllable. In the text underneath the traces, the stressed syllables are given in bold. The contexts in which each of the accented words were produced and the glosses are listed below the figure. All examples happen to have been produced as part of an address in direct speech.

Context

Ohren	Was für große Ohren du hast!
	'What big ears you have!'
Zähne	Was für große Zähne du hast!
	'What big teeth you have!'
Rotkäppchen	(girl addressed by mother) 'Little Red Riding Hood'
Morgen	Guten Morgen, meine Kleine.
	'Good morning, my dear.'

[1] *The phonetic transcriptions use the inventory proposed by the International Phonetic Alphabet. In Braunschweig German,* *<u>Zähne</u>* *is produced as* /tseːnə/ *and not as* /tsɛːnə/ *as tends to be the case for southern speakers.*

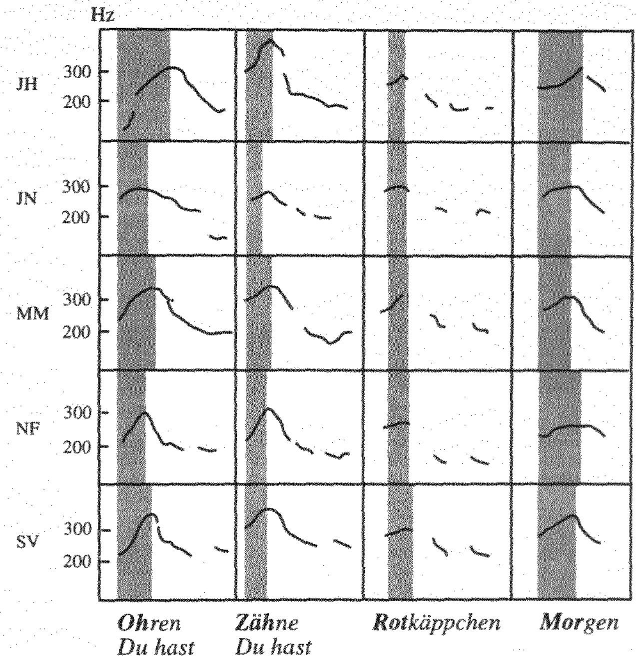


Figure 2 *F0 traces of nuclear falls in the German speech data.*

Some brief comments about the interpretation of F0 are called for at this point. When examining an F0 contour, one needs to bear in mind that the shape of F0 is influenced not only by tonal aspects of the utterance but also by segmental structure. First of all, F0 is interrupted by voiceless consonants (as the vocal folds do not vibrate in their production, there is no F0). Then, some consonants, especially obstruents, affect and change the course of F0. Voiceless obstruents may be accompanied by a dip into the consonant constriction and a subsequent fall starting from a much higher frequency than one would expect, and even voiced ones can cause perturbations (Beckman and Ayers, 1994). Also, vowels differ in intrinsic pitch, and this can make the realisation of the same pattern on different segmental material look different. Moreover, sometimes, pitch trackers make mistakes. They cannot usually deal with creaky voice which many speakers produce intermittently when producing low pitch, in which case F0 values appear scattered. Breathy voice is another problem; it may yield no trace at all. Finally, fluctuations in amplitude may cause halving or doubling errors where the trace appears at exactly half or twice the fundamental frequency it should have done, but such errors tend to be relatively easy to spot (see for instance, in Figure 2, the halving error on *hast* 'have' in JN's realisation of *Ohren Du hast* 'Ears you have'. For more details on the interpretation of fundamental frequency and examples see Beckman and Ayers, 1994.

Returning to Figure 2, the figure shows (a) that nuclear falls are realised as rise-falls in F0 (b) that the rise is variable in its extent, and (c) that the peak of the rise appears to be aligned with the right edge of the stressed vowel. The variability of the rise will be discussed first, followed by comments on peak alignment.

In the corpus, two apparently distinct variants of H*+L were observed, those with and those without a rising glide on the accented syllable. This observation may be taken to lend support to a categorical distinction between these variants. Indeed, in German ToBI, such a distinction is proposed and the variants are labelled as H*L-L% and L+H* L-L%. If the distinction between H*L-L% and L+H* L-L% is, as claimed, categorical, then it should be possible to classify the examples of H*+L shown in Figure 2 into those corresponding to auditory impressions 1(a) and 1(b) above. Systematic auditory comparisons were carried out. The accented word from each example was extracted and saved as separate speech file. Secondly, the separate files were compared across speakers per item (e.g. JH vs. JN's etc. realisations of *Morgen*) and across different items produced by the same speaker (i.e. JH's realisations of *Morgen* vs. her realisation of *Ohren* etc.). On the basis of these comparisons, the falls were grouped into those with a rising onglide (1a) and with a level onglide (1b).

The results of the auditory comparisons suggested that a classification of the nuclear falls in Figure 2 into H*L-L% and L+H* L-L% can be no more than tentative. Figure 3 below shows the results.

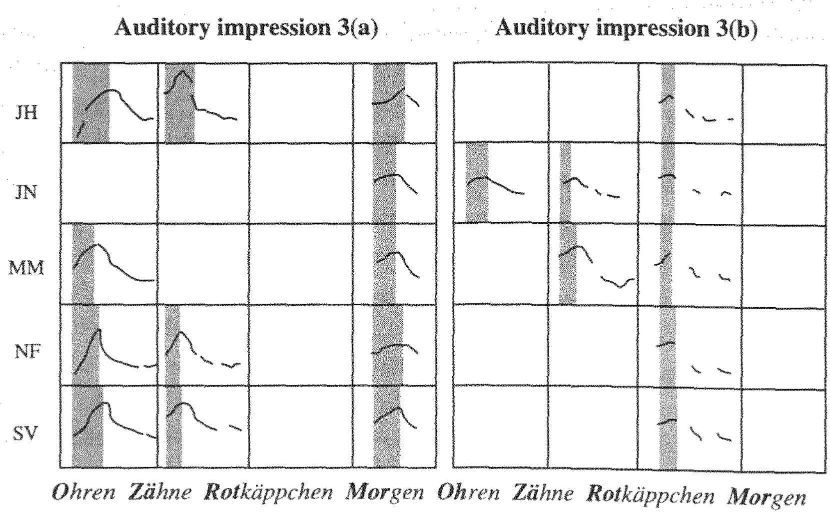


Figure 3 *Classification of F0 traces accompanying H*+L into realisations with a relatively large onglide (3a) and with a small onglide (3b).*

Figure 3 shows that some tokens which were realised differently, look quite similar in F0 (compare, for instance, the onglides on MM's *Morgen* and *Zähne*). Other tokens were difficult to classify on auditory grounds (i.e. it was not clear whether they were similar or different) and assigning them to one category rather than another appeared to be hard to justify. On some words, finally, the distinction did not appear to be made at all. All realisations of H*+L on *Rotkäppchen*, for instance, where the proportion of sonorant segments in the stressed syllable is rather small, were realised with a very small onglide. Realisations of H*+L on *Morgen*, on the other hand, where the proportion of sonorant material in the stressed syllable is relatively large, appeared to have rising onglides for all subjects. *Ohren* and *Zähne* exhibit both types of onglide, but these were the cases where a categorical decision could not be made with confidence. Thus, apparently, different types of onglide in H*+L exist, but they appear to vary gradiently from one another. A clear distinction may be made between realisations representing the extreme ends of the continuum, but this does not necessarily mean that the distinction between H*+L with a large and a small onglide is phonological in nature.

Secondly, Figure 2 above shows that the peak of the rise in F0 appears to be invariably aligned with the right edge of the stressed vowel (which happens to co-occur with the right edge of the stressed syllable for all examples apart from *Rotkäppchen*, where the coda of the stressed syllable is voiceless). This alignment appears to differ from that observed by Uhmann (1991: 159).

(a) F0 alignment in Uhmann (1991) (b) F0 alignment in Figure 2

Figure 4 *Uhmann's stylised F0 contour of H*+L in non-final position contrasted with the pattern observed in the present study*

Uhmann presents a stylised F0 contour for H*+L and she indicates the alignment of this contour with the stressed syllable. In her schematic impression, the accent peak is aligned in the middle of the accented syllable rather than at its right edge. The difference between the peak alignment observed in Uhmann's data and the data presented here may point towards a dialectal distinction, but unfortunately, Uhmann does not provide information about the linguistic background of her speakers[2].

[2] *An alternative explanation for the difference in peak alignment observed in Uhmann's data and the data presented here may be that Uhmann's syllable division of*

The F0 traces shown in Figure 2 above also suggests that peak location is not affected in any obvious way by the amount of preceding voiced segmental material within the syllable (i.e. compare *Morgen* and *Rotkäppchen*)[3]. This finding appears to contrast with evidence from Dutch, a related language, where there is evidence that an increase in sonorant material in the syllable onset pulls a peak leftwards (Rietveld and Gussenhoven, 1995) and also with data from English. In English, the duration of a syllable onset affects the timing of F0 on early parts of a syllable (van Santen and Hirschberg, 1994) and, more generally, longer syllables are reported to have later peaks than shorter syllables (Steele, 1986). Also, there is evidence that in English, the number of syllables following the accent within the stress group affects peak location (Steele, 1986), but Figure 2 does not show any comparable effect for German. For instance, *Morgen* was immediately followed by an intonation phrase boundary but *Ohren* was not, yet, there is no evidence of this affecting peak alignment (at this point, there is no evidence for Dutch available). Thus, so far, the evidence suggests that, in Braunschweig German, peaks are invariably aligned at the right edge of the stressed syllable. Note, however, that in three of the four examples shown in Figure 2 above, the right edge of the stressed syllable happens to be the right edge of the vocalic portion of the syllable. In the fourth (*Rotkäppchen*), the coda is voiceless (the <r> in *Morgen* is vocalised). Figure 5 below shows how nuclear falls are aligned on syllables with a voiced coda. Where it was possible to segment vowel and voiced coda, the coda is marked separately by the darker columns. Example *<gelben>* in Figure 5 shows that in stressed syllables with a voiced coda, the F0 peak is aligned with the offset of the lateral (vowel and lateral could not be separated in the spectrogram). Only one speaker produced *<wohnt>* (/voː ˈvoːn daɪnə oːmaː den/) with H*+L, but again, the peak is reached at the right edge of the coda.

Context

gelben	Es sind die gelben Pflaumen, die sie gern mag.
	'It's the yellow plums she likes.'
wohnt	Wo wohnt Deine Oma denn?
	'Where does your Grandma live?'

---

the name Xenja should be /kseː · nja/ rather than /kseːn · ja/. However, native German speakers are likely to judge the syllable division in Xenja to be /kseːn · ja/ rather than /kseː · nja/ because nja is not normally a permitted syllable in German (there is only a very small number of loan words which begin with the syllable nja).

3 *The differences in peak height which can be observed in Figure 4 across different tokens produced by the same speaker are contextually determined.*

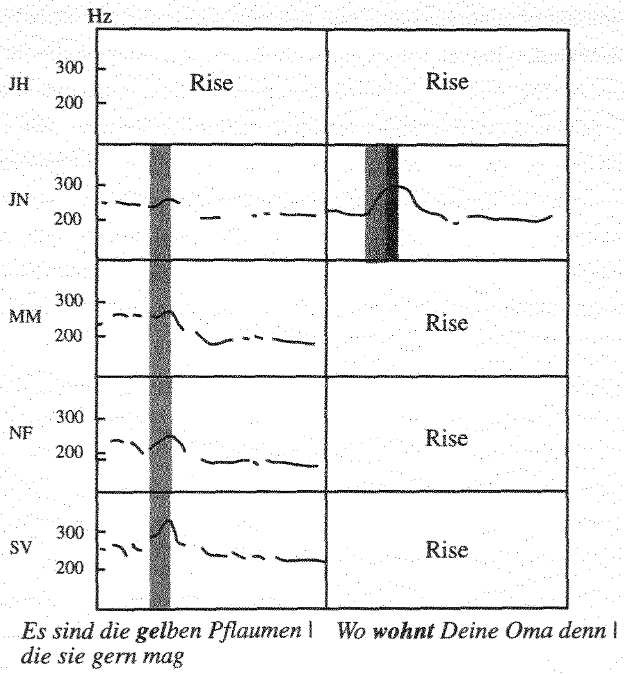


Figure 5 *Nuclear falls on accented syllables with voiced codas. In <gelben>, vowel and coda are included in the shaded section; in <wohnt>, vowel and coda could be separated.*

Combined, the F0 data in Figures 2 and 5 can be interpreted as suggesting that in nuclear H*+L, the F0 peak appears to be reached at the right edge of the stressed syllable. If there is no coda, the peak appears at the right edge of the stressed vowel, and if there is a voiced coda, the peak appears at the right edge of the coda. In all cases, the fall in F0 begins only after the stressed syllable.

To sum up, this section has shown that in Braunschweig German, in non-final position, nuclear H*+L is realised as a rising glide or level on the accented syllable followed by a fall on the following unaccented syllable. Auditorily, the rise is more variable than the fall. A categorical distinction between two types of rise can be made only tentatively. Rather, it is likely that some of the apparently categorically distinct auditory types shown in Figure 3(a) and 3(b) form the extreme endpoints of a continuously varying range of onglides.

In fundamental frequency, the peak of nuclear H*+L appears to be invariably aligned with the right edge of the stressed syllable. Peak location appears to be unaffected by the amount of sonorant segmental material within the stressed syllable or the number of unaccented syllables following within the remainder of the intonation phrase. The extent of the onglide towards the peak varies with the segmental composition of the accented syllable and a tentative distinction between two types of onglide can be made only on syllables with a relatively large proportion of sonorants.

The question arises of whether the variable onglide observed in H*+L should be included in the modelling and transcription of nuclear falls or not. In principle, the pitch accent structure proposed in Grice (1995a, b) and House (1995), adapted as discussed in Chapter 2, captures the evidence.

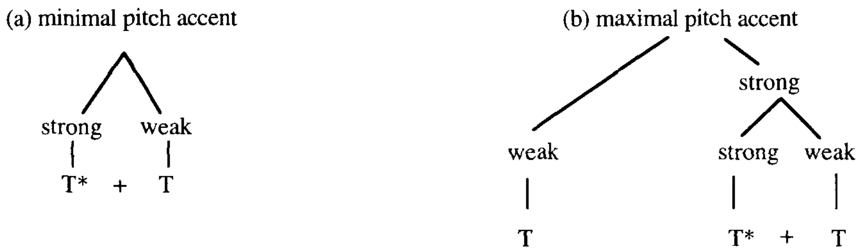


Figure 6 *Branching structure of the pitch accent.*

Figure 6 (a) models the realisation of H*+L with a level onglide and 68(b) models that with a rising onglide (LH*+L). The variable nature of the onglide in (b) is accounted for, because the leading L tone is located under the weak node of the branching pitch accent, and the auditorily falling impression of the whole accent is captured because H*+L is represented within the strong branch of the accent. However, the case for representing the onglide with a leading L in the phonology is not very strong. The following counter-arguments speak against such an analysis. Firstly, the observed realisational differences in the onglide appear to be restricted to accented syllables with a high proportion of sonorants. In other words, the distinction can only be made if the stressed syllable has a certain segmental structure (neutralisation is not impossible, of course, but a distinction which is only made when the segmental context is right is not likely to carry as much weight as one which is made regardless of segmental context). Secondly, the onglide is more variable than the actual fall in pitch and F0. Whether the distinction between rising and level onglides can be made with confidence if no comparable utterances from other speakers are available is doubtful. Thirdly, the onglide appears to be less clearly conditioned by context than the fall. In identical contexts, all speakers chose nuclear falls, but their choice of onglide varied. Finally, the extent of the onglide is likely to depend also on the speaker's register; an analyst may be more likely to decide that an onglide is present when the pitch range is large than when it is compressed. Thus, a phonological distinction between LH*+L and H*+L seems poorly motivated. Alternatively, the high target in H*+L may simply have a range of possible phonetic realisations.

The distinction between falls with different types of onglide has only recently been noted in the literature. As referred to above, in their presentation of 'Saarbrücken-ToBI' (i.e. a German ToBI) Grice and Benzmüller (1995) propose a distinction between falling accents which they transcribe as L+H* L- L% and H*L- L% which seems to be similar to one proposed for American English in EToBI (Silverman et al, 1992, Beckman and Ayers, 1994). The distinction proposed to apply to American English originated in Pierrehumbert's (1980) study. In Gussenhoven's (1984) analysis of British English, however, the distinction is not made, and neither Uhmann (1991) nor Féry's (1993) studies of German recognise the distinction. Earlier studies of German intonation refer only to the variant with the rising onglide. For instance, Delattre (1965) and Delattre et al. (1965) describe the falling accent characterising their 'terminal intonation' as consisting of a rising glide on the prominent syllable with following weak syllables low. This description fits in with the percept illustrated under (a), which happened to be the more commonly observed variant in the corpus. Scuffil (1982: 67) states that most terminal nuclei in German contain a rising tone and Wittig (1956: 80) claims that in German, overall falling tone sequences are frequently made up from consecutive rising accented syllables. Similarly, Trim (1964) comments on accented syllables in the body of German sentences being rising glides. An auditory impression of a nuclear fall with a level onglide, on the other hand is not commented on in any of these studies, and this may be because the variants are quite similar and both share the characteristic of being overall 'falling' pitch accents.

2.2 Final position

This section discusses the auditory impression and F0 alignment of nuclear H*+L in IP-final position. Obviously, the realisation of H*+L on final monosyllables must differ from the realisations of H*+L in non-final position (e.g. the accented syllable may no longer rise throughout, otherwise there would be no fall). In IP-final position, nuclear falls were produced as falls in pitch throughout the stressed syllable. This is illustrated in Figure 7. F0 examples are given in Figure 8 (JN and MM produced downstepped realisations, which are not relevant to the point discussed here).

Context

Wolf	Ich bin der Wolf.	'I'm the wolf.'
Teich	..., am Teich.	'..., next to the pond.'

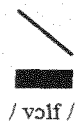


Figure 7 *Auditory impression of nuclear falls on monosyllables. Wolf has the same meaning as in English.*

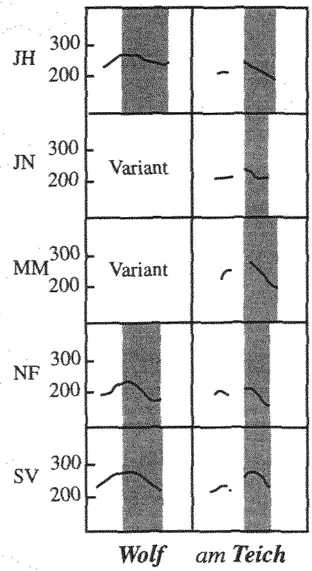


Figure 8 *Nuclear falls on monosyllabic words.*

As the previous section showed that F0 alignment in H*+L relates to the syllable rhyme rather than the stressed vowel, the shaded sections no longer indicate the stressed vowel but the sonorant portion of the rhyme. Note, however, that sonorant onset are excluded.

In Figure 8, *Wolf* /vɔlf/, which has a voiced onset, is contrasted with *Teich* /taɪç/ 'pond', which does not. Despite the auditory impression of a straight fall in F0, <*Wolf*> exhibits evidence of the onglide which spanned the complete duration of the stressed syllable in H*+L in non-final position. Now, this rise appears to be confined to the syllable onset (in this case a voiced labiodental fricative). The peak of the fall is aligned near the beginning of the vocalic segment. In *Teich* which does not have a voiced onset, for most of the speakers, the rise has disappeared altogether for all speakers except SV.

Uhmann (1991) stylises the F0 alignment of H*+L in IP-final position as in (a) in Figure 9; (b) and (c) are stylised examples from the present corpus.

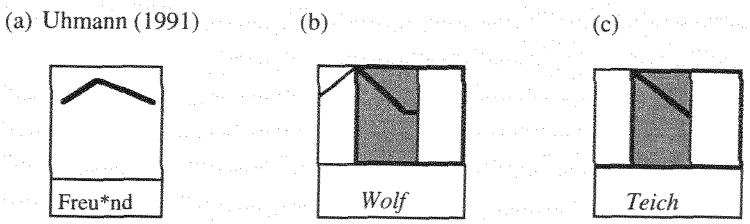


Figure 9 *Uhmann's stylised F0 contour for one-syllable words contrasted with stylisation of Wolf and Teich from the corpus. The shaded sections in (b) and (c) stand for the sonorant portion of the rhyme.*

At first sight, Uhmann's stylisation (a) looks similar to that for *Wolf* (b). However, it is doubtful that the <r> in her example *Freund* ([fχɔɪnt] in Braunschweig German, 'friend') is actually voiced, regardless of how it is produced (German <r> may be produced as a uvular fricative, uvular trill or apical trill). Therefore, her stylisation should have looked like 9(c) to be similar to that observed in the present study. Again, it appears that F0 alignment in Uhmann's data differs from F0 alignment in Braunschweig German.

Figure 10 below summarises the evidence for onglide realisation and peak location in nuclear H*+L. (a) illustrates alignment in non-final position, (b) in final position with a voiced onset and (c) in final position with a voiceless onset (NB. the schematic representations make no claims about segmental timing but aim to show how F0 trace is aligned on different words; secondly, no claims are made about the pitch range depicted).

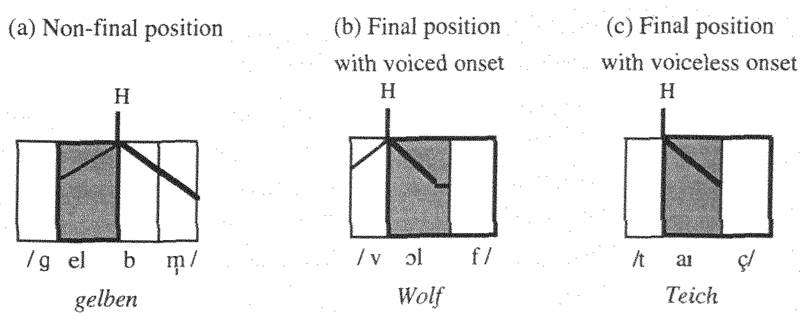


Figure 10 *Onglide and fall of nuclear H*+L in non-final and final position.*

In non-final position (a), F0 on the stressed syllable rises. The peak of the rise is aligned with the right edge of the stressed syllable. In final position (b), the onglide is realised when there is a voiced onset. The F0 peak is aligned with the beginning of the rhyme. If there is no voiced onset (c), there tends to be no onglide at all. Again, the peak is aligned with the beginning of the rhyme. These findings suggest that in Braunschweig German, the realisation of H*+L in F0 is governed by the metrical structure of the accented word, and the rhyme of the stressed syllable. In non-final position, the peak is aligned at the right edge of the rhyme, and in final position, the peak is aligned at the left edge of the rhyme.

2.3 Truncation

The previous section showed that in final position, H*+L is realised as a straight fall in F0 throughout the stressed syllable. In the acoustic realisation, the F0 peak appears to be aligned at the left edge of the rhyme and is followed by a fall throughout the stressed syllable. This section discusses the auditory and acoustic realisations of H*+L on syllables with a very small proportion of sonorant segments. Grønnum (1989) investigated fundamental frequency patterns on longer and shorter 'stress groups' (defined as consisting of a stressed syllable and succeeding unstressed syllables, if any) in a number of Danish dialects. She found that on very short stress groups, F0 is truncated - that is, when voiced segmental material is scarce, a fall in F0 does not run its full course (i.e. the fall does not extend as far downwards in the register as it does when there are more sonorant segments) but simply ends earlier. In the same paper, Grønnum also provided some evidence for truncation in Northern Standard German. However, her German data are not straightforwardly interpretable; they appear to offer evidence not only for truncation, but also for compression[4].

F0 traces for potential 'truncation candidates' are given in Figure 11 below (traces for MM are not given because the speaker produced downstepped falls on the relevant word).

---

4 *English, by contrast, has been described as a compressing language (Ladd, 1996). In compressing languages, the rate of F0 change increases on syllables with a small proportion of sonorants, so that the complete pattern can be accommodated in the relatively shorter time available (see also chapter 5 of the present study).*

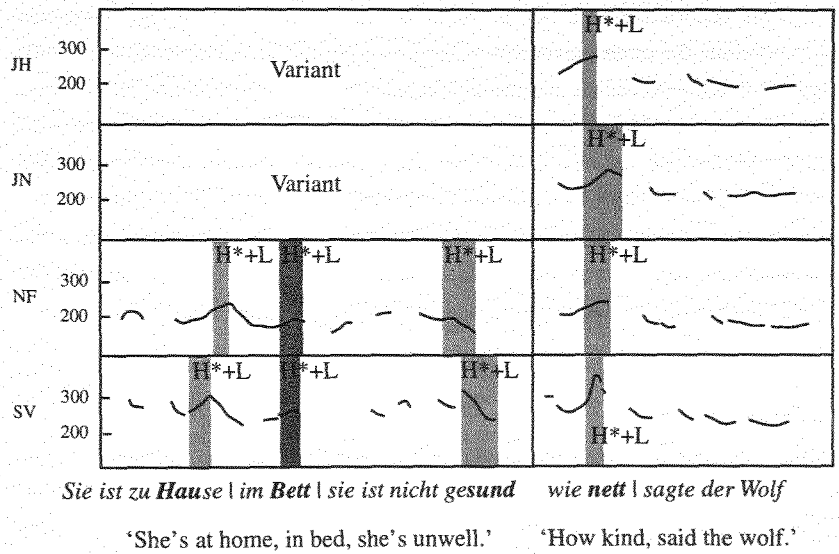


Figure 11 *Nuclear falls on monosyllables with a small proportion of sonorants.*

Example <*Hause*> in Figure 11 illustrates differences in the alignment of F0 in non-final position (*Hause*), and final position (*Bett* and *gesund*). On *Bett* (/bɛt/) which contains a small proportion of sonorants, the fall in F0 does not appear to run its full course but ends earlier[5]. This observation can be interpreted to suggest that German truncates H*+L on syllables with a small proportion of sonorants. Alternatively, one may argue that the pitch accent on *Bett* is different in nature from the preceding and the following accent. However, this is not likely to be the case. The utterances illustrated consist of three co-ordinated syntactic constituents produced as three coordinated intonation phrases, each with a single falling accent. As mentioned in Chapter 2 (section 2.3), coordination structures tend to be produced with equivalent intonation patterns[6], and accordingly, in the example <*Hause*>, all three accents sound like nuclear falls (i.e. the accent on *Bett* is heard as having falling pitch also).

On the word <*nett*> (/nɛt/), the onglide begins on the nasal but continues into the vowel, and for speakers JN and SV, we see some evidence of a fall on the following vowel though much less than on examples <*Wolf*> and <*Teich*> earlier (see Figure 8).

---

5 *In Bett, the initial voiced plosive could not be separated from the vowel. Note that Bett exhibits a peak in F0 similar to Wolf in Figure 8, but both the onglide and the very small fall take place on the vowel.*

6 *The boxes labelled 'variant' contained downstepped realisations of H*+L whose shape in F0 differs from non-downstepped realisations. Nevertheless, these were H*+L, so the point about coordination structures being produced with equivalent intonational choices is not invalidated.*

The following reported speech tag, however is low. As intonational tags have been argued to continue the pitch movement of their host phrases (e.g. Gussenhoven, 1990), this suggests that the preceding accents should specify a low target and be H*+L rather than, for instance, H*. Moreover, when a native speaker of German replaces the word *nett* with the phrase *lieb von Dir* 'nice of you', and repeats the complete phrase (i.e. *Wie lieb von Dir, sagte der Wolf*), then *lieb von Dir* is clearly realised as H*+L, that is, a rise on the accented syllable followed by a fall on the postaccentual syllable. Thus, the example <*nett*> further supports the hypothesis that Northern Standard German truncates H*+L on syllables with a small proportion of sonorants.

2.4 Summary

The corpus evidence presented has shown that in non-final position, nuclear H*+L is realised as a rise on the stressed syllable followed by a fall on the following syllable(s). In principle, two types of onglide could be distinguished, but whether the distinction is truly categorical seems doubtful. In final position, H*+L appeared to be realised as a straight fall in pitch, regardless of the segmental composition of the syllable.

In F0, H*+L is realised as follows. In non-final position, the pitch accent looks like a rise-fall, with the F0 peak invariably aligned with the right edge of the stressed syllable. In final position, a rise-fall appears if the syllable onset is voiced, and a straight fall if it is not voiced. When the onset is voiced, the F0 peak appears to be aligned at the left edge of the syllable rhyme (i.e. at the left edge of the vocalic portion of the syllable). Finally, on syllables with a small proportion of sonorants, the fall in F0 is truncated but apparently nevertheless realised as a fall in pitch[7]. The discrepancy between truncation in F0 and 'no truncation' in the auditory impression appears to support the view expressed in section 3.2 of Chapter 2, namely, that the view of F0 as a narrow phonetic transcription of intonation needs to be treated with caution. The discrepancy between acoustic truncation and apparent auditory stability of H*+L suggests that F0 and auditory impression cannot necessarily be assumed to reflect the same level of intonational representation.

The apparently invariable alignment of the F0 peak in German H*+L at either edge of the syllable rhyme would appear to differ from peak alignment in Dutch (Rietveld and Gussenhoven, 1995). In Dutch, the alignment of F0 peaks is timed with reference to the beginning of the stressed syllable[8], rather than the rhyme, and English,

---

7 *In other words, a relationship of auditory phonetic equivalence appears to hold between truncated and non-truncated H*+L.*

8 *Rietveld and Gussenhoven (1995) also give evidence of an effect of the length of the sonorant rhyme on the alignment of the high target. In rhymes consisting of a vowel plus a following nasal, the high target was aligned further to the right than in rhymes consisting of a vowel and a following plosive.*

where the peak is aligned at some fixed percentage of overall syllable duration (van Santen and Hirschberg, 1994). Figure 12 below tentatively illustrates the cross-linguistic evidence for the realisation of H*+L in non-final position. In German, the peak of the accented syllable remains aligned with the right edge of the rhyme, no matter what the segmental composition of the onset might be. In Dutch and English, on the other hand, the composition of the onset affects peak alignment. In Dutch, the peak moves further left when the onset gets longer, and in English, where peak location is calculated with reference to the duration of the accented syllable, the peak moves leftwards also as the syllable gets longer (here shown by arbitrarily setting the reference point at 50% of complete syllable duration).

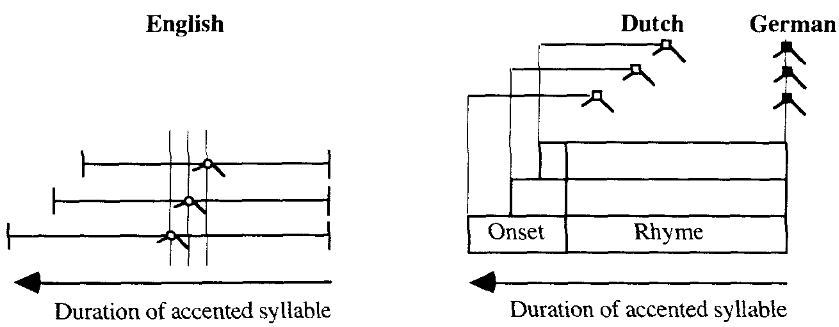


Figure 12 *Peak alignment as a function of syllable onset in German, English and Dutch based on evidence from van Santen and Hirschberg, 1994, Rietveld and Gussenhoven, 1995 and the corpus analysed in the present study.*

3 Prenuclear H*+L

In the auditory analysis, prenuclear realisations of H*+L fell into three structurally and auditorily distinct types. The first was auditorily equivalent to H*+L in nuclear position. The other two variants appeared to correspond to the application of Gussenhoven's (1984) partial and total tone linking rules (see Chapter 1, section 2.2.2.5).

In the variant corresponding to partial linking, the trailing L did not appear in the postaccentual syllable but was displaced to the right. This accent was transcribed as H*+>L. The difference between an intonation phrase with [H*+L H*+L]$_{IP}$ and [H*+>L H*+L]$_{IP}$ is best compared to that between an IP with two nuclear accents (split focus) and one in which a prenuclear accent is followed by a nuclear accent (one focal accent is subordinate to another).

In the other variant of H*+L, the one which corresponded to Gussenhoven's total linking, there was no auditory or acoustic evidence of a trailing L and the high target appeared to be spread up to the following pitch accent. This variant was transcribed as H*>. Figure 13 below shows F0 traces illustrating the three variants. The variants which were not auditorily equivalent to nuclear H*+L are marked with an arrow. Note that the three different realisations of H*+L were produced in identical contexts.

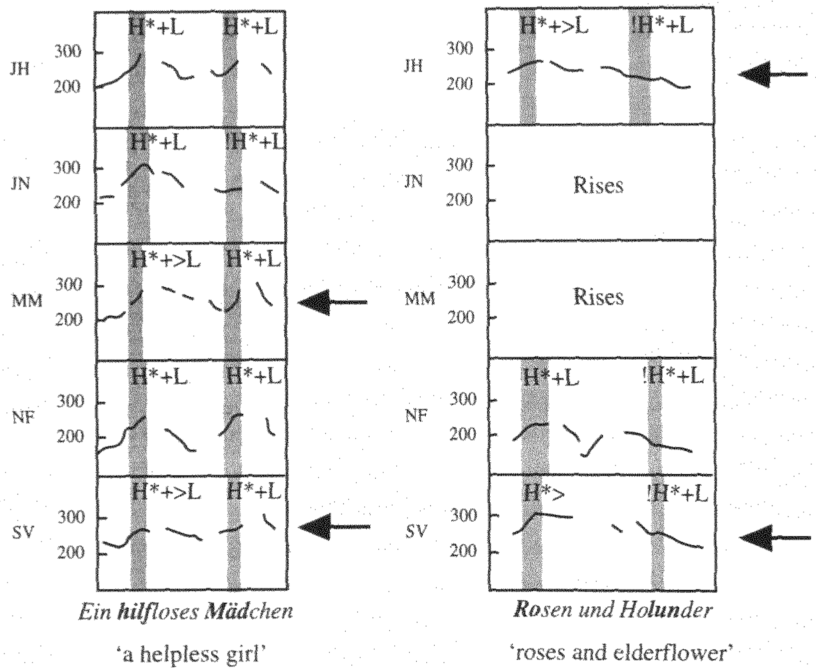


Figure 13 *Two variants of prenuclear falls.*

Figure 14 shows auditory impressions for the three patterns shown in Figure 13 (the issue of downstep is not taken into consideration).

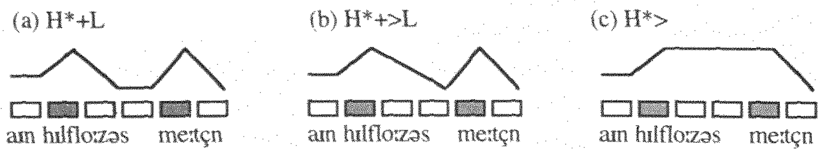


Figure 14 *Auditory impressions of three variants of prenuclear H*+L.*

Figure 15 shows that the prenuclear variants H*+>L and H*> share the auditory and acoustic characteristics of the initial part H*+L, that is, the rise throughout the accented syllable with the peak aligned on its right. The second part, the fall, can apparently be modified. Either the fall is more gradual (H*+>L), or it disappears altogether (H*>), and in that case, the F0 value of the peak of the rise appears to be spread rightwards.

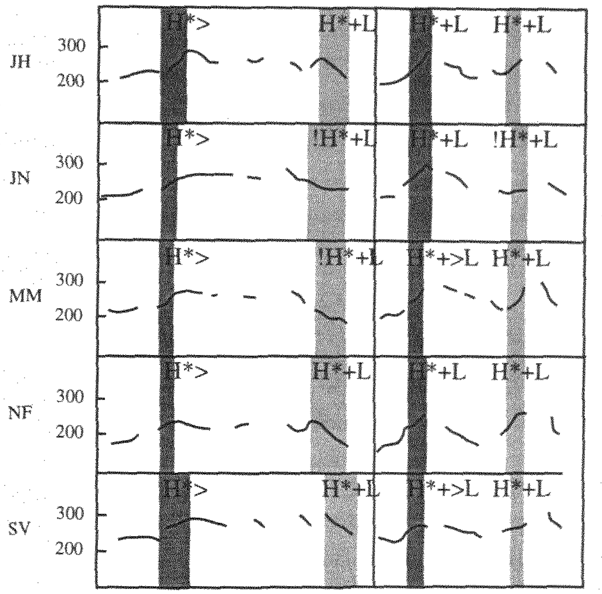


Figure 15 *H*> (H*+L with DELETION) contrasted with H*+L and H*+>L.*

Additionally, the example <*Rosen*> in Figure 13 showed that all three variants H*+L and H*> may be produced in identical contexts. However, in the example of H*>, the following H*+L was downstepped. Further examples of H*> followed by a non-downstepped H*+L are given in Figure 15 for <*Deine Oma ist nicht gesund*>and contrasted with the <*hilfloses*>. Here, all speakers made the same accent choice. In Figure 15, the F0 patterns transcribed as H*> H*+L appear to correspond to Féry's (1993: 149) 'hat contour 1' which she transcribes as a sequence of two completely linked H*L pitch accents, to Wunderlich's (1988: 11) 'bridge accent' and to 't Hart, Collier and Cohen's hat pattern (1990) in Dutch.

At this point, the reader may feel that the distinction between H*+L, H*+>L and H*> is no more convincing than, for instance, the one between H* and L+H* which is postulated in ToBI. And indeed, the distinction between the three variants of H*+L is not as substantial as the one between H*+L and L*+H, for instance. The difference between the ToBI distinction, and the one postulated here is that in ToBI, the distinction between H* and L+H* is not obviously different in rank from that between H* and L*+H. In the approach taken in the present study, the difference between the three variants of H*+L is taken to be different in nature from that between H*+L and L*+H. The three variants of H*+L are assumed to be derived from the same accent choice H*+L. Therefore, even if we find that at times, the distinction between the variants is less clear cut than other distinctions, then we do not have a major problem. In ToBI, H* and L+H* are assumed to be categorically different choices from the accent inventory, and this means that we need to expect this distinction to be as stable as any other distinction captured by the ToBI accent inventory.

The following section presents evidence for nuclear rises (L*+H). Again, evidence for nuclear rises in non-final, final, and prenuclear position will be given separately.

4 Nuclear L*+H

4.1 Non-final position

The auditory impressions of nuclear rises appeared to be more varied than those of nuclear falls. In principle, four options could be distinguished, but these did not appear to be contrastive. The auditory characteristic all nuclear rises shared was that the postaccentual syllable appeared to be higher in pitch than the accented syllable. Figure 16 illustrates the auditory impressions.

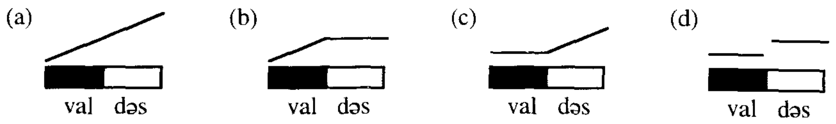


Figure 16 *Auditory impression of nuclear L*+H in non-final position.*

Figure 17 shows F0 patterns for L*+H in non-final position (all examples happen to be 'continuation rises' and appeared intonation phrase-final but not utterance final). One speaker (MM) produced a nuclear fall on *<ergriff sie>*.

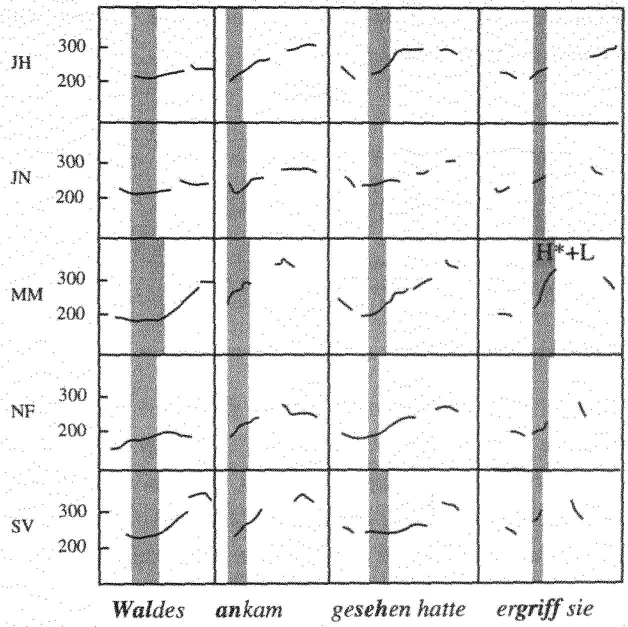


Figure 17 *Nuclear L*+H.*

Context

Waldes Es war einmal am Rande des grossen Waldes [..]
'Once upon a time at the edge of the big forest, ...'

ankam [..] als Rotkäppchen bei ihrer Oma ankam, [..]
'... when Little Red Riding Hood reached Grandma's house, ...'

gesehen hatte [..] weil Rotkäppchen noch nie zuvor einen Wolf gesehen hatte, [..]
'...because Little Red Riding Hood had never seen a wolf before, ...'

ergriff sie [..], und er ergriff sie, [..]
'..., and he grabbed her, ...'

In F0 also, we appear to find a greater range of possible F0 configurations than for nuclear H*+L. Note also the similarity of MM's realisation of H*+L and NF's and SV's realisations of L*+H. These were auditorily distinct; in MM's realisation of H*+L, the preaccentual syllable was lower in pitch than the accented syllable, but in NF's and SV's realisations, it was higher in pitch. Generally, the characteristic which all F0 patterns of L*+H appear to share is that unlike in H*+L, where the peak of the rise is reached within the stressed syllable, in L*+H, the peak is reached within the following syllable. As a reminder, Figure 18 below compares NF's realisation of L*+H on *ankam* 'arrived' with her realisation of H*+L on *Zähne* in Figure 2 earlier.

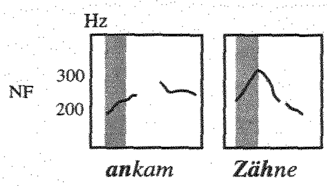


Figure 18 *Comparison of L*+H and H*+L.*

The alignment of nuclear rises with the stressed vowel appears more varied than that of nuclear H*+L also. In H*+L, the F0 peak appeared invariably at the right edge of the stressed syllable, but in L*+H, no similar invariable alignment was observed. In non-final position, a rise on the stressed syllable may begin at the left edge of the stressed syllable, within the stressed syllable, or at its right edge. At times, there may be a dip in F0 on the stressed syllables but the dip may also precede it.

Additionally, Figure 17 suggests that nuclear rises may end in one of two ways. They either reach their peak in the postaccentual syllable and then slump or level out (e.g. JH's <*gesehen hatte*>) or they continue rising beyond the postaccentual syllable until the end of the intonation phrase (<*gesehen hatte*>: JN, MM, NF, SV). On the example <*ergriff sie*>, this is especially clear. For all speakers apart from JH, the trace falls slightly, but for JH, it continues to rise towards the phrase edge. No equivalent distinction at the following phrase-edge was observed for nuclear falls. The distinction between the two types of rises in Figure 17 suggests that the continuously rising tokens are delimited by a high boundary tone, but apparently, this boundary tone is optional, because not all tokens continue to rise. Boundary specifications will be discussed in more detail in section 7 below.

4.2 IP-final position

In IP-final position, nuclear L*+H was produced as a rise in pitch. Additionally, at times, an small upstep in pitch could be discerned between the beginning of that rise and preceding unstressed syllables. This upstep was not interpreted as a clue to a categorical distinction, because versions of L*+H with and without upstep were produced in identical contexts by all five speakers. Figure 19 illustrates the auditory impression of L*+H, and Figure 20 shows F0 examples of nuclear rises in IP-final position.

Figure 19 *Auditory impression of nuclear L*+H in IP-final position. los, means 'off'.*

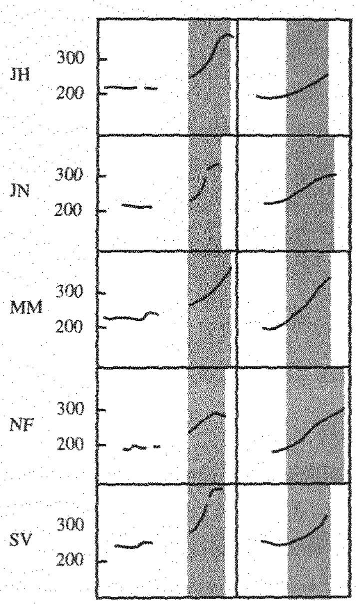


Figure 20 *L*+H in final position.*

Context

Du Und wer bist Du?
'And who are you?'

los Dann marschierte Anna mit ihrem Korb voller guter Dinge los, [..]
'So Anna went off with her bag full of goodies'

Figure 20 shows that in <*Du*>, the starting point of the rise in F0 is aligned at the left edge of the vowel for all speakers. The beginning of the rise, however, is not necessarily the lowest point in the IP; the preceding unaccented syllables are, in absolute terms, lower. In <*los*>, the lowest F0 appears to be aligned with the voiced onset /l/ (excluded from the shaded section).

4.3 Summary

The preceding sections have shown that in non-final position nuclear L*+H may have a number of auditory variants, which do not appear to be categorically distinct. In all variants, the postaccentual syllable has higher pitch than the accented syllable. On IP-final monosyllables, L*+H is realised as a straight rise in pitch. Generally, less information could be gained from the corpus on the F0 alignment of L*+H than emerged for H*+L. The shapes of L*+H in F0 and the beginning of the actual rises appeared to be considerably more variable than for H*+L. Erickson et al. (1995: 180) have commented on considerably greater difficulty in modelling low tone in English than modelling high tone. The difficulty appears to reflect more complicated interactions between the physiological control mechanism at work when pitch is lowered than when pitch is raised. The evidence from the Northern Standard German corpus presented in the preceding sections appears to suggest that in German, this observation might be applicable also.

5 Prenuclear L*+H

Auditorily, prenuclear falls did not appear to differ in any obvious way from nuclear falls (for auditory impressions, see Figure 16 above). Figure 21 shows F0 traces of prenuclear rises in complete intonation phrases. For *rote* in Figure 21, all speakers made the same intonational choices, that is, a prenuclear rise followed by a downstepped nuclear fall (downstep will be discussed in section 6 below). <*Mädchen*> and <*zeige*> were produced by all speakers apart from NF as L*+H L*+H. The traces show that the F0 patterning of prenuclear L*+H is not observably different from that of nuclear L*+H; patterns rise throughout the stressed syllable and reach a peak in the following syllable. No distinct variants equivalent to H*+>L and H*> were observed (this is not to say, however, that these variants do not exist; clearly, the evidence presented here is restricted to the corpus analysed).

Context

rote	Sind es rote Pflaumen oder gelbe Pflaumen, die Oma lieber mag?
	'Is it red plums or yellow plums Grandma prefers?'
Mädchen	Das Mädchen erinnerte sich an die Worte ihrer Mutter, die ...
	'The girl remembered her mother's words which...'
zeige	Ich zeige Dir einen Ort wo Du welche finden kannst.
	'I'll show you a place where you can find them.'

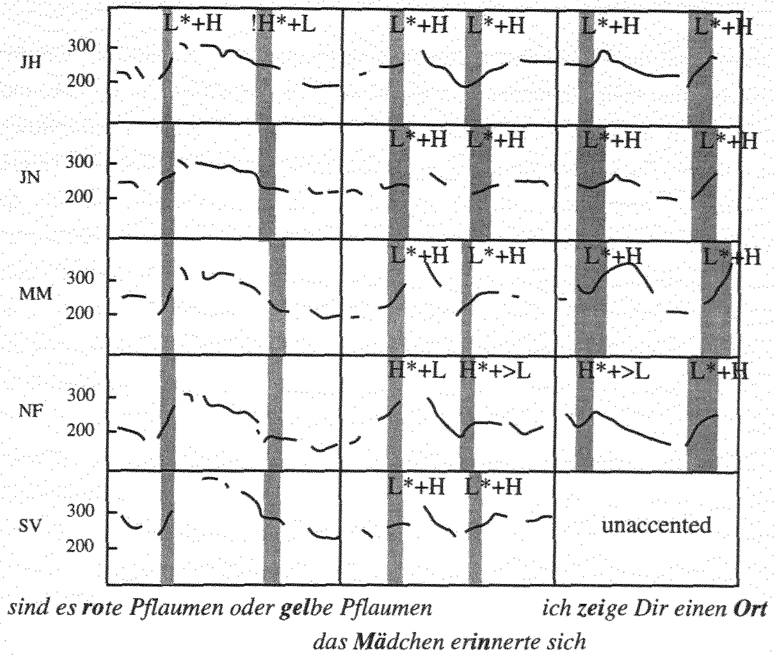


Figure 21 *Prenuclear L*+H.*

Note, finally, that L*+H followed by a low target may, at times look strikingly similar to H*+L and cannot be distinguished unless the alignment with the stressed syllable is carefully observed. A comparison between NF's realisation of H*+L on *Mädchen* in Figure 21 above and MM's realisation of L*+H on the same word illustrates this point. The apparent similarity results from the specific peak alignment of H*+L which appears to characterise Northern Standard German, namely, at the extreme right of the stressed syllable. As a result, the rise in F0 on the stressed syllable is visually rather salient. The difference between L*+H and H*+L in prenuclear position is illustrated in Figure 22 below with the German utterance *die blühenden Linden* ('the blossoming lime trees').

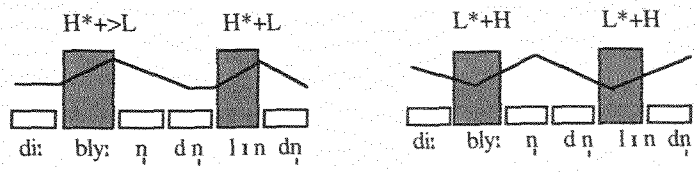


Figure 22 *Schematic illustration of F0 alignment in H*+>L H*+L and L*+H L*+H.*

Figure 22 shows that the F0 patterns accompanying prenuclear rises and falls can be rather similar, but that their alignment with segmental structure differs. In H*+>L, the rise is completed within the stressed syllable, but in L*+H, the rise begins on and follows the stressed syllable. The surface similarity in F0 patterning between H*+>L and L*+H shown in Figure 22 combined with the apparently gradiently variable onglide in realisations of H*+L may explain why in an evaluation of German ToBI transcribers had some difficulty in distinguishing accent patterns transcribed there as H*, L+H* and L*+H (Grice et al., 1995: 1719). The preceding sections have shown that a considerable amount of detail about the alignment of F0 with segmental material is needed to distinguish H*+>L and L*+H, and in the case of the onglide, a categorical distinction is questionable.

To sum up, the prenuclear rises observed in the corpus of Braunschweig German did not appear to differ in any obvious way from nuclear rises. No auditorily distinct variants similar to those of H*+L (i.e. H*+>L and H*>) could be discerned. In F0, no obvious differences between nuclear and prenuclear L*+H were established either.

6 Phonological adjustments

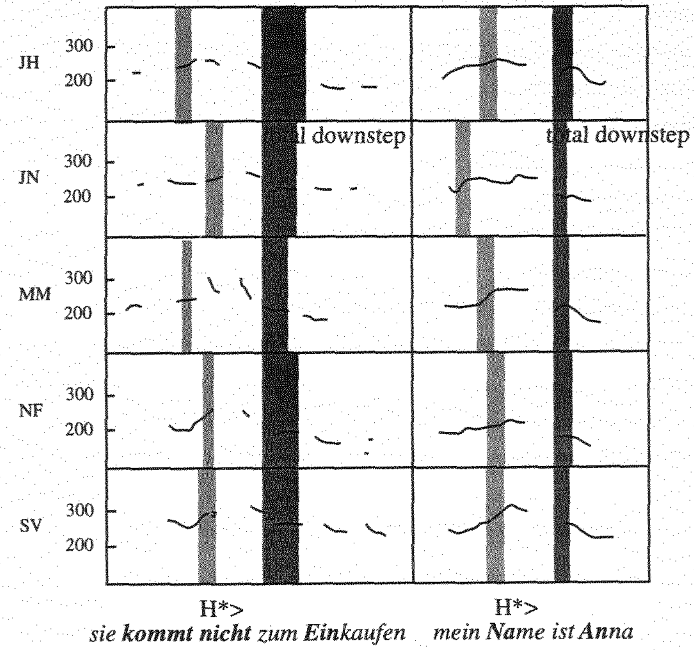
This section will discuss the evidence in the corpus for phonological adjustment rules which account for the modifications applying to basic accents H*+L and L*+H in continuous speech. With respect to the corpus investigated, the discussion aims to be exhaustive. Note, however, that the present study was restricted to one particular speaking style, with a similar rate of delivery across subjects. Other adjustments additional to those postulated in the present section are likely to apply in other speaking styles. In the analysis of segmental phonetic structure, the assumption is that we cannot expect to find evidence of all the connected speech processes which have been found to apply generally in a language in all utterances produced in that language. For instance, in fast speech, we are likely to find more evidence of place assimilation than in slow, careful speech. Similarly, in intonation, it is unlikely that evidence of all possible pitch accent adjustments will emerge from the analysis of one particular type of text.

In his analysis of British English, Gussenhoven (1984) distinguishes between two types of phonological adjustments; 'modifications', and 'linking rules'. Modifications DELAY, HALF-COMPLETION, STYLISATION and RANGE are shown to affect nuclear accents; partial and total linking affects prenuclear accents. Briefly, in the present study, no evidence was found for DELAY, HALF-COMPLETION, or STYLISATION, but intuitively, they seem applicable to German. RANGE was suggested by Gussenhoven to be gradient, and was not considered here (only categorical adjustments are discussed). Instead of RANGE, in the present study, DOWNSTEP is treated as a phonological adjustment rule. Including DOWNSTEP in the set of adjustment rules implies a departure from Gussenhoven's view of modifications. In his (1984) study, the application of a modification changed the meaning of an intonation phrase. DELAY, for instance, changes the meaning of a pattern from 'neutral' to 'highly significant'. DOWNSTEP does not result in comparable changes in meaning; instead DOWNSTEP affects the focus structure of an utterance (see, for instance Féry 1993: 157). From a purely structural point of view, however, there is no obvious reason why DOWNSTEP should not be treated as a phonological adjustment. A downstepped accent tends to sound equivalent to a non-downstepped accent when listened to in isolation, and looks similar in F0. However, it is defined by being lower in pitch and F0 than an immediately preceding accent than can be expected were the utterance characterised by declination, but not downstep (for declination, see, for instance, Cohen and 't Hart, 1967, and for different sources of downtrends in utterances see Pierrehumbert and Beckman, 1988).

Examples of downstep applied to H*+L are shown in Figure 23 below (the downstepped stressed syllable in !H*+L is shaded in dark grey). Auditorily, downstepped H*+L appeared to be equivalent to non-downstepped H*+L (positioned lower in the register) for all speakers but JN. In JN's utterances, the high target did not appear above the following low target in the register, but at the same level[9].

Figure 23 *Auditory impressions of partial and total downstep. Note that the figure illustrates two possible realisation of an IP-final !H*+L. The immediately preceding H*+L accent from which the !H*+L in the figure is downstepped is not shown.*

9 *JN's totally downstepped H*+L was not an isolated occurrence; examples of total downstep were found throughout the corpus and were produced by all speakers.*



'she can't go shopping' 'my name is Anna'

Figure 24 *Partial and total downstep in German. JH and MM accented* kommt *in <Einkaufen>, the other speakers accented* nicht.

The distinction between partial and total downstep shown in Figure 24 may have led the authors of the German version of ToBI to suggest an additional pitch accent H+L*. The distinction between !H*+L and H+L* would then reflect the difference between partial and total downstep. However, anticipating evidence presented in Chapter 6, the distinction between !H*+L and H+L* appears to be gradient rather than categorical. Therefore, in what follows no distinction between !H*+L and H+L* will be made. All downstepped falls will be transcribed as !H*+L.

A further modification which nuclear H*+L appeared to undergo was DELETION, similar to the deletion of a low trailing tone in totally linked H*> H*+L. Figure 25 below shows examples from speaker NF, who appeared to have a preference for this pattern and produced it frequently. Scattered examples from other speakers were found also.

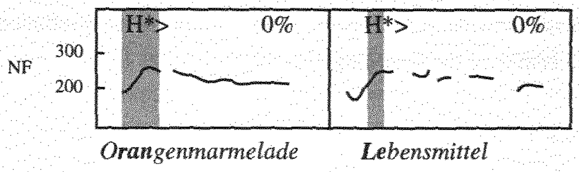


Figure 25 *Two examples of nuclear DELETION.*

Context

Orangenmarmelade	Sie füllte einen Korb mit ein paar reifen, gelben Bananen, einem Glas Orangenmarmelade, Margarine und [..] 'She filled a basket with some ripe yellow bananas, a glass of orange marmalade, margarine and ...'
Lebensmittel	Ich packe ein paar Lebensmittel für sie ein, [..] 'I will pack some food for her, and ...'

The patterns shown in Figure 25 look very similar to those shown for prenuclear H*> in section 3 of this chapter. Again, the peak is reached within the accented syllable, rather than after the accented syllable, as is the case for L*+H. An example of prenuclear H*> from one speaker is repeated below, for the purpose of comparison.

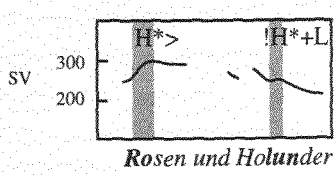


Figure 26 **Prenuclear L*+H.**

Further evidence for a nuclear pattern very similar to that of prenuclear H*> is given in Figure 27(a) below, which shows F0 traces from a list of downstepped compounds (this figure does not show evidence from the corpus but is taken from data recorded for an experimental investigation of downstep presented Chapter 6). The patterns illustrated were produced in the same experiment. Figure 27(b) shows why the pattern in 27(a) can be assumed to be underlyingly H*+L (note also the total downstep of the last item in the list in 27a and the partial downstep in 27b)[10]. Both patterns were produced by all speakers throughout the course of the recordings[11].

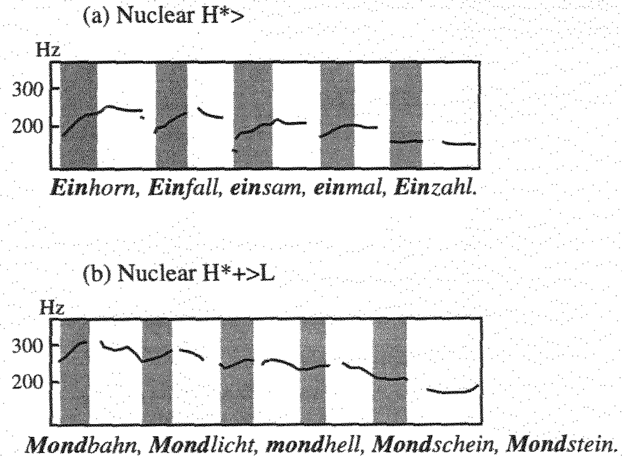


Figure 27 *More evidence of nuclear pattern similar to prenuclear H*>.*

In sum, in Braunschweig German, we find very similar prenuclear and nuclear patterns which can be accounted for as underlyingly H*+L with DELETION of the L. This finding may be accounted for by maintaining the distinction between prenuclear linking rules and nuclear modifications and stating that total linking and DELETION happen to have the same effect. Alternatively, one might dispense with the distinction between linking rules and modifications and postulate one unified set of phonological adjustment rules which may, at least in principle, apply to any accent, regardless of that accent's position in the intonation phrase. In the present study, the second option was preferred. Instead of positing two different mechanisms to account for patterns which do not appear to differ in auditory impression, F0 alignment or, intuitively, meaning, DELETION was assumed to apply potentially to prenuclear and nuclear accents. Considering that nuclear and prenuclear H*+L can be auditorily equivalent, this appeared to be a defensible solution.

---

[10] *The succession of fully realised H*+>L accents tended to be produced at the beginning of the recording session and the versions resembling prenuclear H*> towards the end. This may suggest that DELETION is a connected speech process which is more likely to happen in casual speech.*

[11] *The patterns shown in Figure 28(a) were realised as a gliding rise on the stressed syllable followed by level pitch on the following syllable. Those in Figure 28(b) were heard as a rising glide on the stressed syllable followed by a fall in pitch on the following syllable.*

Finally, note that nuclear H*> appeared IP-finally in the corpus but not utterance-finally, and that all intonation phrases ending in H*> appeared to be in some way dependent on preceding and following intonation phrases within the utterance (see chapter 2, section 2.3 for dependencies between intonation phrases). This may suggest that the application of DELETION is in some way constrained by discourse structure. Prenuclear deletion appears to tie together more closely the accents in an intonation phrase (this is why a linking rule has been proposed to account for this pattern), while nuclear DELETION might tie together intonation phrases (i.e. H*> might signal to a hearer that a section of utterance has ended but that the information contained in this section is in some way dependent on that given in the following section).

Dispensing with the distinction between linking rules and modifications requires the pattern transcribed as H*+>L to be reanalysed, as Gussenhoven's (1984) notion of 'partial linking' is no longer available. Here, the trailing L appears not on the postaccentual syllable, as in H*+L, but on a syllable further to the right. In Féry (1993) who follows Gussenhoven (1984) this pattern is accounted for as partially linked H*+L H*+L. In the present study, a phonological adjustment DISPLACEMENT is proposed to account for H*+>L. In English, this adjustment appears to apply to prenuclear accents only.

To conclude, instead of adopting Gussenhoven's distinction between linking rules and modifications, in the present study, a unified set of phonological adjustments is proposed which can potentially apply to all pitch accents in an utterance[12]. Constraints on the application of adjustments are assumed to be language specific. Such constraints can be stated felicitously if we assume that tunes have an internal structure such as one posited in the British School; that is, if we accept that tunes are divided into elements such as the prehead, the head and the nucleus (see Ladd, 1996:211 for a combination of the intonational phrase in the AM framework and a syntagmatic tune structure). For instance, the corpus evidence presented in this chapter suggests that in German, adjustments apply to prenuclear accents more frequently than to nuclear accents. Considering the semantic contribution nuclear accents are assumed to make to a phrase in and German and English (see Gussenhoven, 1984 for English), this is not surprising, and the observation may be taken to correspond to the commonly observed differences between nuclear and prenuclear accents in other languages.

---

12 *Generally, phonological adjustments are taken to reflect the numerous functions of intonation in speech. Some adjustments affect focus structure (e.g. DOWNSTEP), others affect meaning (DELAY), some reflect intonational connected speech processes (DISPLACEMENT, DELETION) and others relate to discourse structure (e.g. nucler DELETION in German). However, these suggestions are preliminary, and more work on the details of this proposal is needed.*

7 Intonation phrase boundary specifications.

This section will discuss the evidence in the German corpus for tonal specification of intonation phrase boundaries. The discussion will be restricted to IP-final boundary tones; the question of IP-initial preaccentual pitch exceeds the scope of this study[13].

Generally, in the corpus of Northern Standard German, upward pitch movement was observed on some IP-final unstressed syllables, but equivalent downward pitch movement was absent. This observation would appear to confirm findings by Wunderlich (1988), and Féry (1993). Secondly, the data showed that boundaries following L*+H, H*+L and H*> may remain level after the target for the trailing tone has been realised. Boundaries labelled 'level' were in the majority; upward pitch movement characterised only 78 out of 917 IP boundaries (5 speakers), no other boundaries appeared to require a tonal specification different from that of the immediately preceding tone. The boundaries exhibiting a rise in pitch were labelled as H% and boundaries which did not exhibit pitch movement were labelled as 0%. 0% was used rather than the boundary conventions used by systems following the Beckman-Pierrehumbert approach on the assumption that in a first analysis of a specific variety of a language, the labelling should reflect, as closely as possible, actual observations of pitch and F0. In German ToBI, for instance, a 'high level' boundary following L*+H is transcribed as H-L%, but this implies the application of an upstep rule which raises the L% to the level of the H-. However, clearly, such a rule cannot be postulated a priori, but only after the analysis of the data has shown the need for it.

F0 traces illustrating IP boundaries with and without F0 movement are shown in Figure 28 below. The left panel shows the final section of intonation phrases with a nuclear pitch accent H*+L; the right panel shows the final section of phrases with a nuclear pitch accent L*+H.

The boundaries following H*+L shown in Figure 28 (a) will be discussed first. 30 (a) shows a set of vocative tag constructions labelled as H*+L 0% 0%[14]. The first IP ends after *hören kann* ('to hear you with'), and the second after the tag *meine Kleine* ('my dear'). The last pitch accent in the phrase is associated with *hören*, and after the fall in F0 following the peak, the IP exhibits no further F0 movement. The following tag remains low and level, and again, no movement delimits its right edge. Auditorily, the

---

13 *More evidence is required for criteria which would allow us to decide whether IP-initialSuch constraints can be stated felicitously if we assume that tunes have an internal structure such as one posited in the British School; that is, if we accept that tunes are divided into elements such as the prehead, the head and the nucleus (see Ladd, 1996:211 for a combination of the intonational phrase in the AM framework and a syntagmatic tune structure). preaccentual pitch should be left unspecified, accounted for by a leading tone, or by a boundary tone, or by a mixture of all of these. For a study of IP-initial high preaccentual pitch in Dutch, see Grabe et al. (1997).*

14 *For intonational tags see, for instance, Ladd, 1986 and Gussenhoven 1990.*

pitch level at the end of each IP did not seem to differ from that of the lowest point of the preceding fall in F0. The traces shown in Figure 28(a) can be taken as representative for all 'low' boundaries in the German corpus. Figure 28(b) shows F0 traces of intonation phrases with a nuclear pitch accent L*+H. The boundaries of these phrases either rose above the F0 value of the trailing H or remained at the same level. Auditorily, JH, JN and SV's realisations contained a double step-up in pitch, but MM and NF's realisations did not.

One 'fall-rise' was found in the corpus, and is shown in Figure 29 (speaker MM).

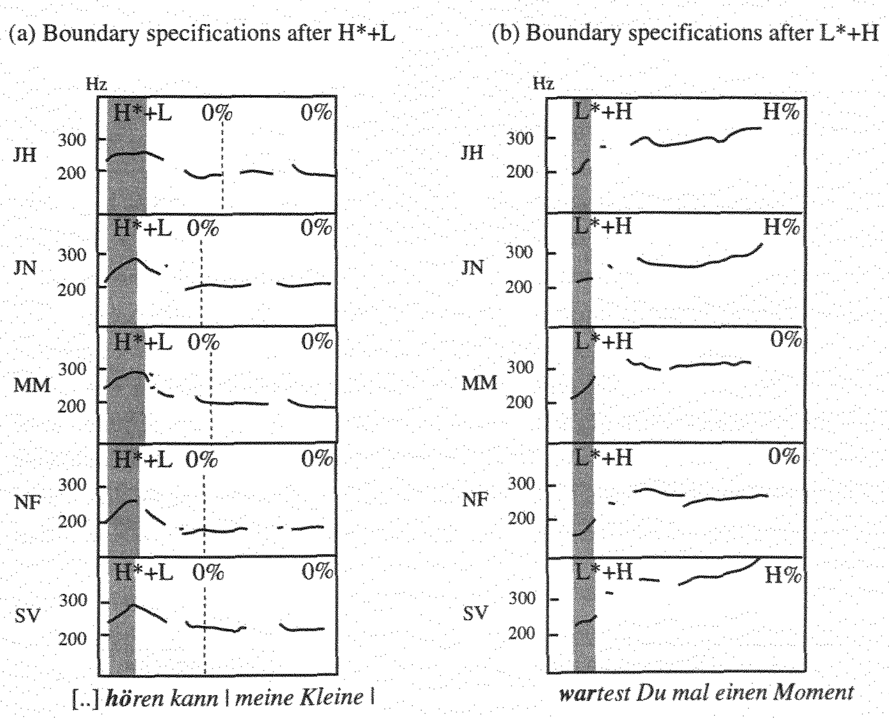


Figure 28 F0 traces of 'high rising' (JN, SV) and 'high level' IP boundaries (NF).

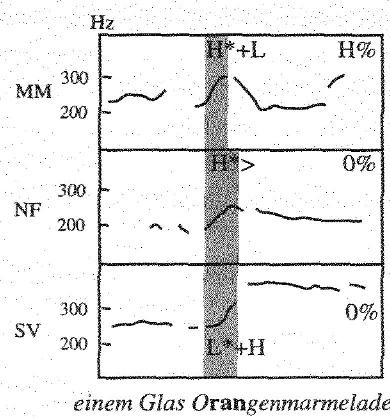


Figure 29 *H*+L H% contrasted with H*> 0% and L*+H 0%.*[15]

The fall-rise in Figure 29 is contrasted with an example of H*> followed by a level boundary and one of L*+H followed by a level boundary (the distinction between nuclear rise-plateaux represented as H*> 0% and L*+H 0% is discussed in the following section). Figures 30 and 31 illustrate all types of IP boundaries found in the corpus.

The findings will be discussed using the schematic auditory impressions given in Table 1. The table on the left summarises the observed boundaries and the one on the right the boundaries which might have been observed but of which there appeared to be no evidence. Boundaries following H*+L are shown in the first row, and those following L*+H in the second. Boundaries which exhibit the same pitch level as the lowest point of the preceding falls are referred to as 'level' and those associated with change upwards as 'high'.

	Observed boundaries	
	'level'	*'high'*
H+L*		
L+H*		

*** Boundaries**
'low'

Table 1 *Schematic auditory impression of nuclear pitch accent - IP boundary configurations observed in the corpus.*

15 *The slump in F0 following the accented syllable in NF's realisation of* Orangenmarmelade *may be attributed to declination.*

Table 1 shows that falls and rises may either continue level, or may rise again. Falls do not, however, appear to fall further; and rises do not continue level and fall again at the phrase boundary. Similar observations appear to hold for American English also and are stated somewhat less explicitly in Pierrehumbert (1980). Note, however, that Table 1 claims to apply to American English and Northern Standard German only. Northern Irish English, for instance, exhibits the pattern shown under L*+H - 'low' (see also Cruttenden, 1986: 139). Figure 30 below shows an example from Lowry (1997). The accent on *northern,* which may be represented as L*+H is followed by a sharp fall on *line* in the absence of a stressed syllable.

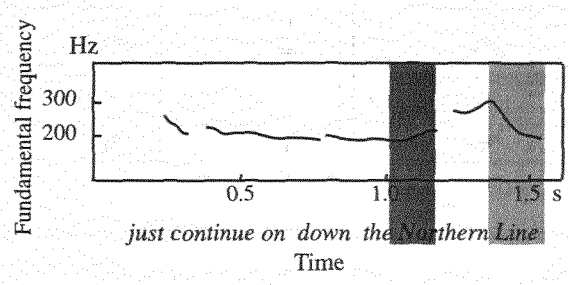


Figure 30 *A fall at an IP boundary following L*+H in Northern Irish English.*

The asymmetry in the high and low boundary specifications may be accounted for in more than one way. On the one hand, one might relegate the asymmetry into the area of phonetic realisation. This is the approach taken in systems following Pierrehumbert's (1980) approach. There, high rising boundaries following L*+H are accounted for as L*+H H-H%. The plateau which often appears between the trailing H and H% is accounted for by a spreading rule, which states that a tone spreads when the next tone is equal or higher. Thus, H- spreads before H%. High level boundaries following L*+H are accounted for as L*+H H-L%; here, an upstep rule raises the L% to the level of the preceding H-. Low level boundaries following H*+L are accounted for as H*L-L%[16]. L- does not spread because L% is taken to be lower, and accounts for the gradual drop in F0 which may follow the last accent in the phrase.

Alternatively, one may take the apparent boundary asymmetry to reflect a difference in phonological specification, and this is the approach taken, more or less explicitly, by Gussenhoven (1984), and Lindsey (1985) in their accounts of English and by Wunderlich (1988) and Féry (1993) in their accounts of German. In the present study, this second option was preferred. The relatively small proportion of boundaries in the corpus which appeared to require an independent specification suggests strongly that it

16 *F0 tends to drop gradually following H*L- in Pierrehumbert's examples, and gradually declining postnuclear stretches were found in the present corpus also.*

may be more economical for transcribers to consider labelling with tones only those IP boundaries which differ tonally from the preceding specification. However, the observation also raises the more fundamental question of whether every intonation phrase boundary requires a tonal specification.

Table 2 below contrasts transcriptions in systems following Pierrehumbert (1980) (left) with the alternative transcriptions proposed in this study (right)

	Pierrehumbert (1980)		**Alternative approach**	
	Level	High	Level	High
Fall	H*L- L%	H*L- H%	H*+L Ø%	H*+L H%
Rise	L*+H H-L%	L*+H H-H%	L*+H Ø%	L*+H H%

Table 2 *Transcriptions following Pierrehumbert (1980) are shown on the left and the alternative transcription proposed in the present study are shown on the right.*

Table 2 shows that in Pierrehumbert's approach, the H*+L-level configuration is transcribed as H* L-L%. The apparent mirror-image 'L*+H-level', however, is not accounted for by the 'mirror-image' in the transcription (L*H-H%) but as L*+H H-L%. What comes closest to a mirror image of H*L-L% in the transcription (L*+H H-H%) accounts for the rise-high configuration (the falling-rising pattern H*L-H% is relatively uncontroversial and will not be discussed further).

The Pierrehumbert transcription appears to have some disadvantages. Firstly, the transcription does not reflect the apparent 'mirror-image' relationship between the fall-level and the rise-level. Instead, fall-level and rise-level are grouped together by virtue of the final L%. This appears to imply that the high boundary following the rise in the rise-level is underlyingly low, an assumption which requires some justification which does not appear to be given. Secondly, the transcription of the rise-high pattern comes closest to mirroring that of the fall-level, but the shapes do not mirror each other. Finally, this particular way of accounting for the asymmetrical realisation of boundaries in General American and Southern Standard British English appears to be somewhat inflexible when it comes to transcribing other varieties of English which exhibit intonational patterns not included in the set discussed here. As mentioned above, in Northern Irish English, a rise-plateau-slump pattern has been observed which sounds similar to the pattern illustrated under 'rise-low' in Table 1. In this pattern, a rising pitch accent is followed by a level stretch which then exhibits a downwards movement at the phrase boundary. This pattern appears to correspond to Pierrehumbert's transcription L*+H H-L%, as long as one assumes that the upstep rule is not applicable in Northern Irish English. However, if this pattern were transcribed as such, then the transcription would no longer reflect the cross-varietal difference (see also Ladd, 1996: 146 for Glasgow English, which has IP boundary options which are similar to those observed in Northern Irish English).

The alternative account proposed in this study is illustrated in Table 2 on the right. Boundaries are assumed to be tonally specified only if they introduce a 'new' tonal target, that is, one which is different from the one specified by the tone which immediately precedes the boundary. Tonally unspecified boundaries are transcribed as '0%' and receive their realisation targets from the last specified tone in the IP. Rising boundaries, on the other hand, are proposed to be tonally specified, with H% to be implemented as a pitch level relatively higher than the one which immediately precedes, whether high or low.

The advantages of the system proposed here appear to be that it (a) reflects the 'mirror-image' relationship between fall-levels and rise-levels, (b) does not suggest that the rise-high is the mirror-image of the fall-level, and (c) does not suggest that rise-levels have underlying low boundary tones. Also, the Northern Irish rise-plateau-slump pattern may now be accounted for as L*+H L%, that is, as the mirror image of H*+L H%.

More generally, the evidence may be interpreted to suggest that boundary tones are language and dialect-specific. Northern Irish English may have only L% whereas Southern Standard British English has only H%. Other dialects or languages may not be characterised by any tonally specified boundaries at all, or they may exhibit some combination of the three types suggested here. In other words, just like the inventory of pitch accents which has been shown to vary across languages, the inventory of boundary specifications is proposed to be language and dialect-specific.

In summary, the evidence from the corpus suggests that in Northern Standard German, an intonation phrase may either be associated with a high boundary tone, or not be tonally marked. High boundary tones are transcribed as H% and tonally unspecified boundaries as 0%. '0%' transcribes the end of the intonation phrase and specifies an insertion point for a second phonetic target of the immediately preceding tone. Figure 31 illustrates the choices speakers are assumed to have following the nuclear accent.

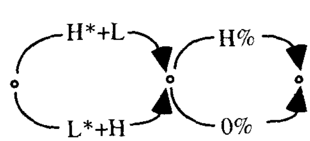


Figure 31 *Nuclear accent - boundary options at the underlying level of tonal structure.*

8 A problem - nuclear and prenuclear rise-plateaux

This chapter will conclude with the discussion of a specific problem in the analysis of German; that of 'rise-plateaux'. In Northern Standard German, two distinct types of rise-plateau appear to exist, one involving an underlyingly falling and the other involving an underlyingly rising accent, but the exact auditory and acoustic characteristics distinguishing the two types are not as easy to determine as the characteristics of some other contrasts. The problem was raised previously by Féry as one involving two different types of 'hat patterns' and claimed to involve neutralisation. The patterns were said to differ in phonological structure but to have, by coincidence, the same form (see section 2.2.3.4 in Chapter 1 for a summary of Féry's analysis). The corpus analysis carried out for the purposes of the present study suggested (a) that the problem is not restricted to prenuclear position and (b) that despite the similarity of the two patterns, they can be distinguished on auditory and acoustic grounds.

Figure 32 illustrates examples of nuclear rise-plateaux from the corpus. Excluding MM's realisation of *Orangenmarmelade*, the F0 patterns in Figure 32 look very similar. However, they did not sound the same (although careful listening was required to establish this). As the transcriptions show, NF's rise-plateaux appeared to be variants of H*+L with DELETION, whereas all other rise-plateaux were L*+H with a following unspecified IP boundary. In examples transcribed as L*+H 0%, an upstep in pitch could be discerned between the stressed syllable and the following unstressed syllable. In the examples transcribed as H*> 0%, no such upstep was observed.

Context

Orangenmarmelade	Sie füllte einen Korb mit ein paar reifen, gelben Bananen, einem Glas Orangenmarmelade, Margarine, und ein paar Vollkornsemmeln. 'She filled a bag with some ripe, yellow bananas, a jar of Orange marmalade, margarine and some wholemeal rolls'
Lebensmittel	Ich packe ein paar Lebensmittel für sie ein und ich möchte, daß Du sie ihr bringst. 'I'm packing up some food and I'd like you to take it to her'

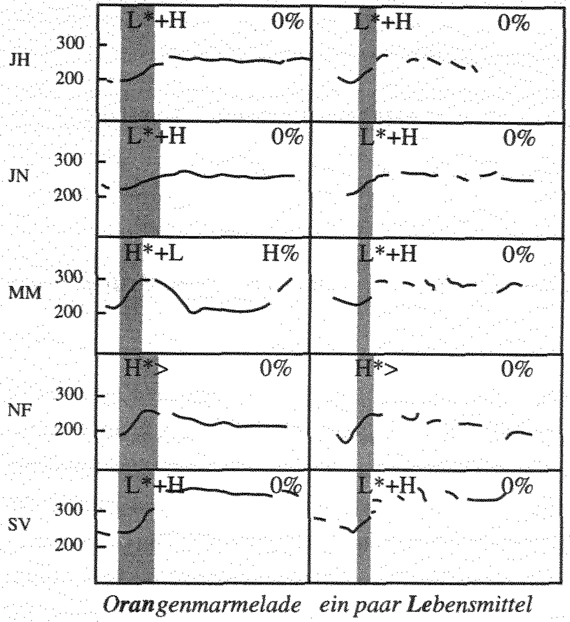


Figure 32 *Rise-plateaux in the corpus.*

In fundamental frequency, the characteristics distinguishing the two rise-plateaux are similar to the distinguishing features of rises and falls discussed in previous sections of this chapter. The H*> 0% rise-plateau is realised as a rise on the stressed syllable with the peak of the rise aligned at its right edge. The following F0 trace is level or slumps slightly. In L*+H 0%, F0 rises beyond the stressed syllable.

Further criteria distinguishing the two types emerge from the intonational context in which a rise-plateau is found. In isolation, the the distinction between the two types of plateau may not be as obvious as the distinction between H*+L and H*>, for instance, but when a list of either of the two types of rise-plateaux is produced, the distinction is very clear. First of all, the H*> 0% rise-plateau is downstepped (see Figure 27a in the present Chapter), but the L*+H 0% rise-plateau is not (generally, downstep of low tones has not been observed in German or English, and the assumption is that only high tones can step down). Accordingly, NF's realisation of <*Orangenmarmelade*> was followed by downstepped accents on the following two listed items (*Margarine und Vollkornsemmeln*), but JH, JN and SV's realisations were not. In MM's realisation, the following accent was also H*+L, and the final accent on *Vollkornsemmeln* was downstepped. Figure 33 below shows the difference schematically; a sequence of H*> 0% IPs is illustrated in (a) on the left, and a sequence of L*+H 0% IPs is in (b) on the right.

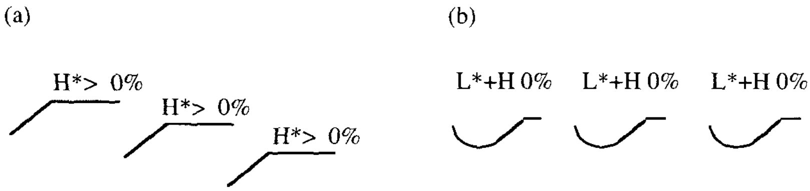


Figure 33 *Schematic illustration of the difference between two types of rise-plateaux in sequence.*

In summary, the two types of rise-plateaux can be distinguished by a set of auditory and acoustic IP internal and IP-external criteria. Internal criteria involve (a) the presence or absence of an upstep in pitch between the stressed syllable and the following unstressed syllable, and (b) the alignment of the rise in F0 with the stressed syllable. An external criterion is downstep; H*> steps down, but L*+H> does not. The two types of hat patterns (i.e. patterns involving prenuclear rise-plateaux) observed by Féry (1993) were observed in the corpus analysed in this study also, and the account given for nuclear rise-plateaux above applies to them also.

9 Summary

This chapter has discussed the auditory and acoustic intonation patterns observed in a corpus of Northern Standard German read speech. The patterns were accounted for with two basic pitch accents H*+L and L*+H, a set of categorical phonological adjustment rules (DOWNSTEP, DISPLACEMENT and DELETION) and an optional IP boundary tone H%. The adjustment rules proposed differ from those postulated in Gussenhoven (1984) in that Gussenhoven's distinction between prenuclear linking rules and nuclear modifications is no longer maintained. The proposed treatment of intonation phrase boundaries differs from that posited in systems following the Beckman-Pierrehumbert system in that IP boundary specifications are assumed to be language-and dialect specific. In particular, Northern Standard German is not assumed to have a low boundary tone.

In the acoustic realisation of H*+L, a range of apparently gradient variation effects were observed. These appeared to be the following:

(1) The rise in F0 on the stressed syllable
In H*+L, the rise in F0 on the stressed syllable appeared to range from 'rising' to 'level', but although an auditory distinction could be made at the extreme ends of the continuum, the overall auditory impression of H*+L involved a fall in pitch .

(2) The location of a downstepped F0 peak in the register
In IP-final downstepped H*+L, the location of the F0 peak ranged from a level above that of the final low to that of the final low. Again, at the extreme ends of the continuum, this distinction could be heard, but in many cases, a categorical distinction could not be made.

(3) The F0 excursion of the fall following the peak.
A less marked fall in F0 was observed in H*+L on syllables with a smaller proportion of sonorants than on syllables with a relatively large proportion of sonorants. Again, this did not affect the overall auditory impression of H*+L as a fall in pitch. Even in cases where the fundamental frequency on the stressed syllable was virtually level, the accent appeared to be realised as a fall in pitch[17].

Figure 34 below summarises the phonological modelling of HL.

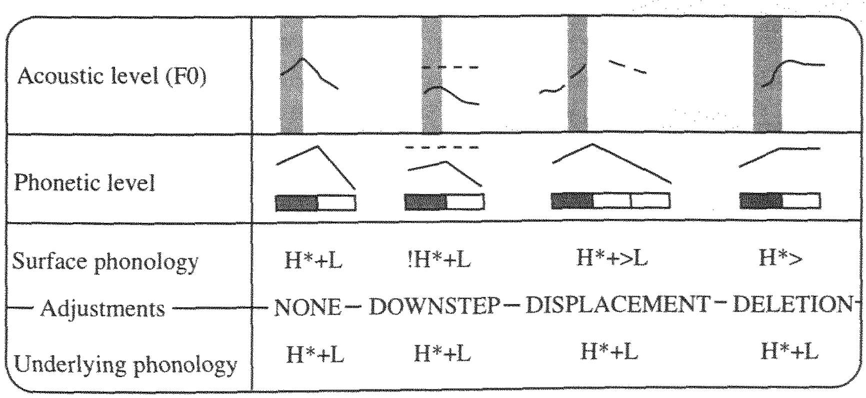


Figure 34 *Categorical phonological adjustments of H*+L.*

Three linguistic levels of representation are shown (phonetic, surface phonological and underlying phonological) and F0 traces are shown for illustrative purposes. The phonological level of representation models an underlying level of phonological structure and a phonological surface level derived via adjustment rules. In effect, the

17 *Note that a distinction between a 'falling' and a 'level' accent is made in German (H*+L vs. H*> in the present study). In H*>, no fall in pitch is heard.*

phonological surface level corresponds to the transcription system used in the corpus analysis. The phonetic level of representation models the auditory impression. Figure 34 postulates an underlying phonological category H*+L. H*+L may either be unmodified (as under NONE), and in this case, underlying and surface phonological structure are identical. Alternatively H*+L may be modified by DOWNSTEP, which lowers H*+L relative to a preceding accent, DISPLACEMENT, which moves the trailing L beyond the postaccentual syllable (how far this L may be moved requires further experimental investigation), and DELETION, which removes the trailing L altogether. At the phonetic level, unmodified and downstepped H*+L and H*+L with DISPLACEMENT involve falling pitch, but H*> does not. Here, the pitch following the accented syllable remains level. The rise in pitch on the stressed syllable, however, is equivalent to that observed in the other variants of H*+L.

Figures 35-37 below illustrate the acoustic variation effects observed. These effects are assumed to be gradient and do not form part of the phonological structure, although, at the extreme ends of the continuum they form, acoustic variation effects may be perceptible. The first phonetic variation effect which will be postulated is 'Expansion'. Expansion affects the excursion of the rise in F0 on the stressed syllable.

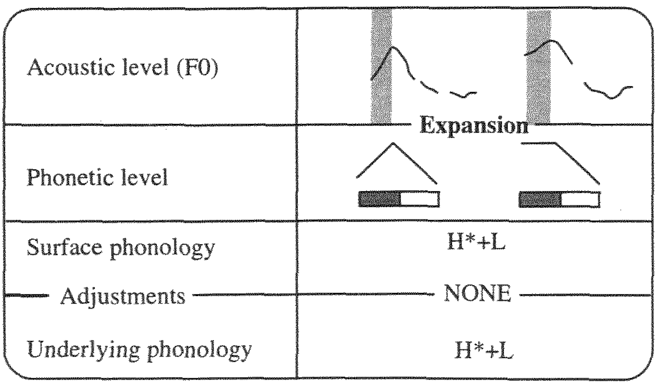


Figure 35 *Expansion: gradient acoustic variation. Note that the auditory impressions show the extreme endpoints of a continuum between falling accents with rising and level onglides.*

The acoustic variation effect summarised in Figure 36 below relates to the distinction between partial and total downstep (see section 6 of this chapter). This effect will be referred to as 'Final peak lowering'. In IP-final downstepped accents, the location of the F0 peak appeared to be located either above the level of the final low or at the same level. Again, the distinction appeared to be gradient.

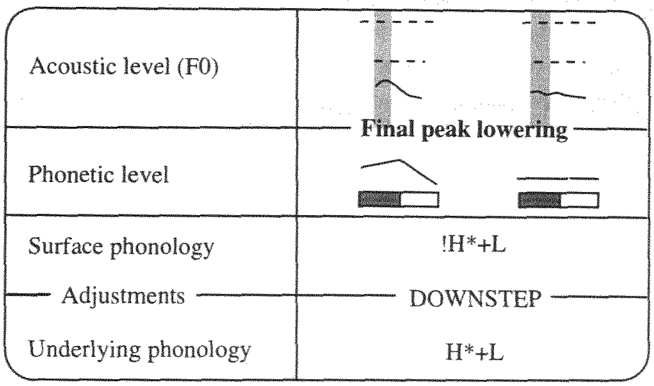


Figure 36 *Final accent lowering: gradient acoustic variation. Again, the auditory impressions show the extreme endpoints of a continuum between partial and total downstep. The dotted line indicates the location of the immediately preceding F0 peak.*

Figure 37, finally, shows the acoustic variation effect referred to as truncation in section 2.3 of this chapter. Note that unlike expansion, which did not appear to be linked reliably to the availability of sonorant segmental material, truncation is conditioned by segmental structure. If sonorant segmental material is scarce, truncation will apply. Conversely, if a considerable proportion of sonorant segments is available, a fall in F0 will surface.

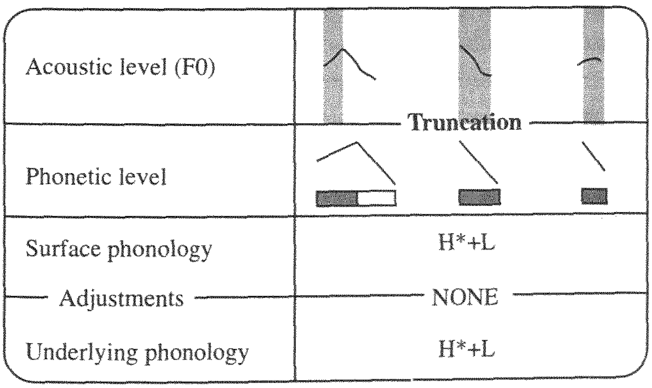


Figure 37 *Truncation: gradient acoustic variation. The contrast between the long and the short grey boxes on the right represents the difference between syllables containing relatively larger and smaller amounts of sonorants.*

Figure 37 shows (a) the difference in shape and alignment of F0 in non-final and final position and (b) that the fall in F0 following the peak does not run its full course when the syllable contains only a small proportion of sonorants. Note, however, that the auditory impression of a truncated fall in F0 is nevertheless that of a falling accent. This discrepancy between auditory impression and acoustic realisation may be taken to support the separation of acoustic and phonetic levels of representation discussed in Chapter 2 (see Nolan, 1990 for a discussion of the concept of a 'phonetic level of representation'). Acoustically, a truncated fall looks completely different from a non-truncated fall, but both sound like falls. If one assumed that the phonetic level of representation was identical to the acoustic level, then a difference in auditory impression should be observed when truncation has applied. However, native speakers hear both as falling. Non-native speakers, on the other hand, when informally asked whether they hear truncated falls as falls, tend to be less sure than native speakers. Apparently, the phonetic level of intonational representation is not necessarily accessible to non-native speakers.

The discussion will now move on to the modelling of L*+H. In the corpus data, L*+H did not appeared to be affected by phonological adjustments. Figure 38 below models the evidence.

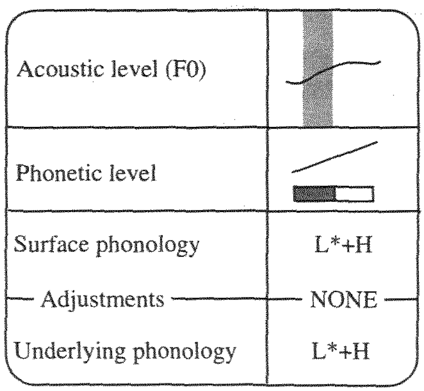


Figure 38 *Modelling of L*+H. The auditory impression illustrated at the phonetic level is one example of the four options illustrated in Figure 18, section 4.1.*

No obvious acoustic variation effects applying to L*+H were observed. However, auditorily and acoustically, the details of the realisation of L*+H appeared to be less stable than those of H*+L. Auditorily, the shared characteristic of all realisations of L*+H appeared to be that pitch on the postaccentual syllable appeared to be higher.

Acoustically, the peak of the rise in F0 was always reached in the postaccentual syllable rather than on the stressed syllable.

To summarise, the present chapter has given an account of the intonation patterns observed in a corpus of Northern Standard German. The evidence was discussed at four levels of linguistic representation; (1) the underlying phonological level, (2) the surface phonological level, (3) the acoustic representation (F0) and (4) the phonetic level (auditory impression). At the underlying level of phonological structure, two basic pitch accents H*+L and L*+H, and an optional IP boundary tone H% were postulated. At the surface level of phonological structure, where basic accents are combined into the intonation patterns characterising continuous speech, changes in the structure of the basic accents were accounted for by the categorical phonological adjustment rules DOWNSTEP, DISPLACEMENT and DELETION. In the acoustic implementation of the surface phonological patterns, three gradient effects were observed in F0. Two of these, referred to in the present chapter as expansion and final peak lowering, are accounted for as categorical changes in other autosegmental-metrical analyses of German. In the present study, they are not assumed to be represented in the phonology. Expansion applies to the realisation of H*+L; here, gradient variability was observed in the rise in F0 on the stressed syllable. German ToBI accounts for this effect by positing a categorical difference between a smaller rise transcribed as H* L-L% and a larger rise transcribed as L+H* L-L%. The findings of the present chapter, however, would appear to call this distinction into question. Gradient final peak lowering was observed in IP-final downstepped F0 peaks. Again, in German ToBI, two categorically different intonational categories, that is, !H*+L and H+L*, are postulated, but in the present study, this distinction is questioned[18] (see also Chapter 6). Thirdly, truncation was observed in the F0 excursion of the fall following the peak (see Chapter 5 for a discussion of whether truncation is best accounted for as phonetic or as phonological). Truncation appeared to be conditioned by the proportion of sonorant material the pitch accent is realised on, and was hypothesised to be gradient also. None of the previous autosegmental-metrical analyses of German have commented on this effect.

The following chapter will compare the evidence from Northern Standard German with parallel evidence from Southern British English. The comparison will involve the four levels of linguistic representation discussed in this chapter, that is, an underlying and a surface phonological level, the acoustic level and the auditory phonetic level.

---

18 *Note that neither Uhmann (1991) nor Féry (1993) postulate a categorical difference between L*+H L-L% and H*L-L% or !H*+L and H+L*.*

Northern Standard German and Southern Standard British English

Chapter 4

1 Introduction

This chapter compares the evidence from Northern Standard German presented in Chapter 3 with comparable evidence from Southern Standard British English. The comparison will follow the structure of the Chapter 3, that is, nuclear falls will be discussed first, followed by prenuclear falls, nuclear rises and prenuclear rises. Again, evidence will be presented separately for non-final and final H*+L. Then, the cross-linguistic evidence for phonological adjustments will be presented, followed by a comparison of intonation phrase boundary specifications. Generally, cross-linguistic differences will be discussed in detail, but similarities will be touched upon only briefly.

2 Comparison

2.1 Nuclear H*+L

2.1.1 Non-final position

The auditory realisations of English and German H*+L appear to be very similar. If at all different, German H*+L involves a more clearly defined step-down in pitch than English H*+L, which is more glide-like. Also, German H*+L seemed to sound more often like a rise-fall than English H*+L. However, these comments are tentative.

In German H*+L, a wide range of possible onglides were observed, that is, in some cases, the stressed syllable was realised with rising pitch, at other times, with level pitch or it was intermediate between rising and level. The same observation applied to the English data, and again, some of the onglides appeared to be auditorily distinct, but others were intermediate. In fundamental frequency, English, H*+L in non-final position tended to be realised as rise-fall, and the extent of the rise was variable. Again, no obvious cross-linguistic differences appeared to apply. Fundamental frequency examples of English H*+L are shown in Figure 1 below.

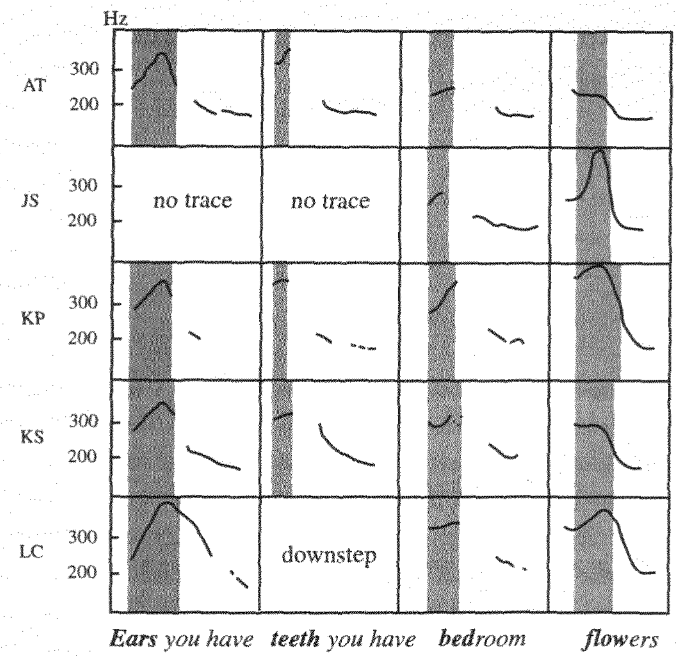


Figure 1 *Nuclear H*+L in English.*

Context

Ears	What big ears you have!
teeth	What big teeth you have!
bedroom	Over here, in the bedroom.
flowers	Maybe you'd like to take your Grandma some flowers.

An absence of cross-linguistic differences in nuclear falls is reflected also in the pitch accent inventories of English and German ToBI. However, in both ToBIs, a categorical distinction is made between a nuclear fall with a rising onglide (L+H* L-L%) and one with a level onglide (H*L-L%)[1]. Yet, the evidence presented for German in the previous chapter of this study did not support a categorical distinction between L+H*L-L% and

1 *Note, however, that in an evaluation of English ToBI (Pitrelli et al., 1994), L+H* was described as a 'minor variant' of H*, and the analysis of the results, the categories H* and L*+H were merged. Thus, no information is available on how reliably English transcribers made the distinction.*

H*L-L%. The evidence for a categorical distinction between L+H* L-L% and H*L-L% in English appeared to be questionable also.

Figure 2 below shows a tentative separation of onglides into tokens of H*+L realised with a level onglide and those realised with a rising onglide.

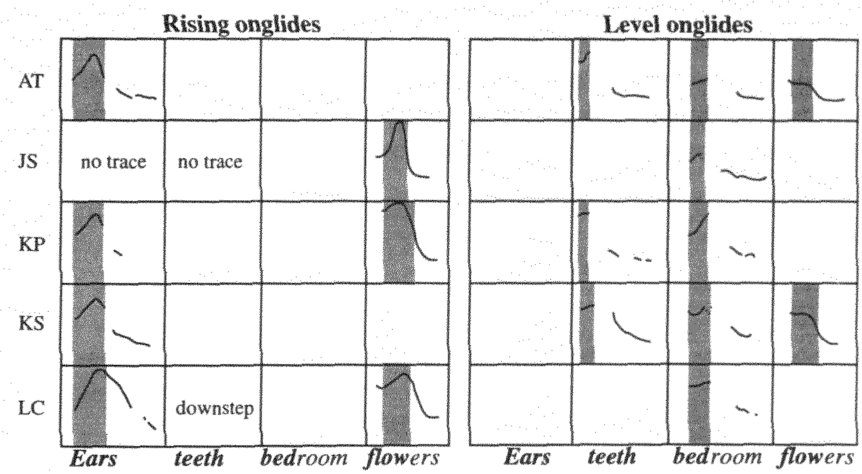
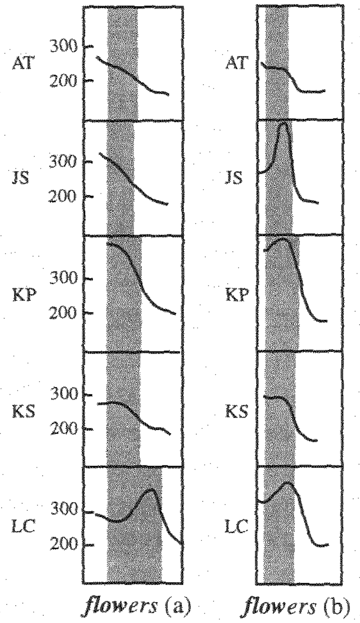


Figure 2 *Tentative classification of nuclear H*+L in English into realisations with rising and level onglides.*

Again, the separation was, at times, difficult to make and required repeated comparisons. A reliable correlation between auditory impressions of onglides and their realisations in F0 appears to be doubtful also, compare, for instance, the realisations in F0 of the auditorily level onglides on *bedroom* with the auditorily rising onglides on *flowers*.

Also, Figure 2 shows that the distinction was apparently not made on examples <*teeth*> and <*bedroom*>, where the stressed syllable contained a majority of non-sonorant segments. An absence of the distinction on syllables with little sonorant material was noted in German. In the German data, the distinction appeared to be made only when syllables contained a relatively large proportion of sonorant segments. In the English example <*ears*> in Figure 2, however, where the stressed syllables contained a majority of sonorants, the distinction was not made either. In <*ears*>, all accents were realised with a rising onglide. The only word where a distinction was made was <*flowers*>.

Figure 2 might be interpreted to suggest (a) that in English, the distinction is more closely linked to the segmental structure of the stressed syllable (i.e. when there is more scope for voicing, the onglide rises and when there is little, it is level) or (b) that the distinction is more closely linked to context. Some contexts condition a rising onset and others condition a level onset. However, neither interpretation is likely to be wholly appropriate. The first claim is contradicted by the example <*flowers*>. Here, the distinction is made, although the stressed syllable contains a relatively large proportion of sonorants. This suggests that the distinction between rising and level onglides is unlikely to be more strongly linked to segmental structure than in German. The claim that the distinction might be conditioned by context is also contradicted by the example <*flowers*> and further evidence illustrated in Figure 3 below. In the experimental materials, the word flowers happened to appear twice, in two different contexts (given below the figure). In both contexts, both versions of onglide were produced. Also, different speakers used both versions or either version.



Context (a): Maybe it's flowers Anna likes.

Context (b): Maybe you'd like to take your grandma some flowers.

Figure 3 *<Flowers> produced with a rising and a level onglide in two different contexts.*

Thus, the corpus data analysed in the present study suggest that the distinction between realisations of H*+L with rising and level onglides is gradient, both in Northern Standard German, and in Southern Standard British English. Accordingly, both will be represented as H*+L.

2.1.1.1 Peak alignment

English and German H*+L did not appear to differ with respect to their overall shape in F0 or the types of onglide observed. In one other respect, however, the languages did differ. In Northern Standard German, the peak of the rise in F0 appeared to be invariably aligned at the right edge of the stressed syllable, but in English, the peak was reached earlier, that is, within the stressed syllable rather than at its right edge (see Figure 1).

Figure 4 below provides comparative evidence. The *<flowers>/<Morgen>* comparison on the left shows that in the English example *<flowers>*, the F0 peak is aligned earlier than in the German example *<Morgen>*, and that the fall in the English example begins within the stressed syllable rather than following it. The comparison on the right shows that the distinction is not observed when the stressed syllable contains little voiced segmental material.

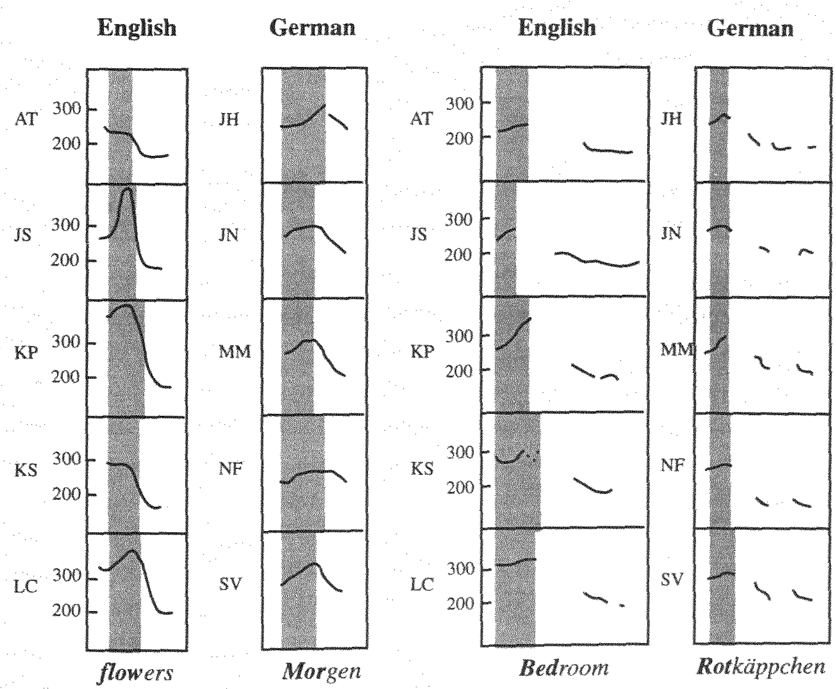
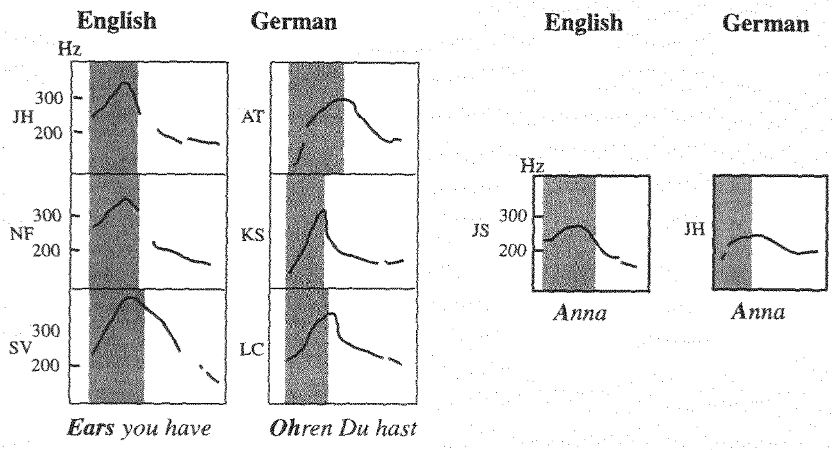


Figure 4 *Peak alignment in English and German.*

Note, however, that the examples shown in Figure 4 were not produced in identical contexts across languages. Thus, one might argue that the observed differences in peak alignment are in some way related to the different contexts in which the tokens were produced. Figure 5 below, however, shows that the context is not likely to be responsible. The examples shown were produced in identical contexts, and again, the difference in peak alignment can be observed.



What big ears you have.
Was für große Ohren Du hast.

My name is Anna.
Mein Name ist Anna.

Figure 5 *Peak comparison in identical contexts across languages.*

Figure 6 illustrates the alignment of F0 peaks on stressed syllables with a voiced coda. It shows that in English H*+L, the fall begins either within the coda or before the voiced coda. In German, on the other hand, the fall only ever begins after the end of the syllable rhyme.

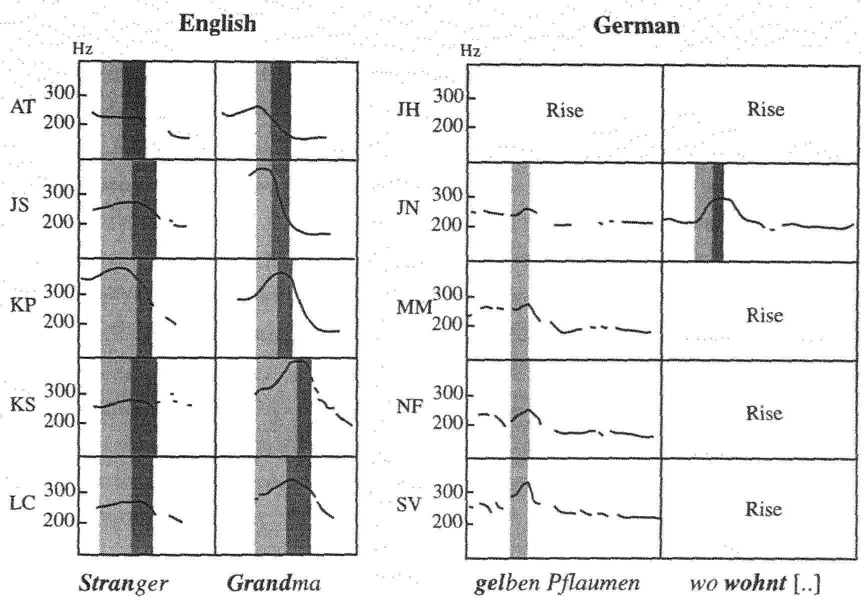


Figure 6 *Alignment of the peak in nuclear H*+L realised on syllables with a voiced coda in English on the left and German on the right.*

Context English	*Stranger*	A stranger was waiting for her
	Grandma	I'm going to see my Grandma.

Clearly, more systematic acoustic evidence of the claimed cross-linguistic difference in peak alignment is needed. However, the data from the cross-linguistic corpus was not ideally suited to acoustic measurements. The segmental structures of words with H*+L could not be directly compared across languages, nor was the number of tokens with a specific segmental structure comparable. Therefore, peak alignment in English and German was measured and compared in data recorded for an experimental study presented in Chapter 5. That study used materials whose segmental structure was directly comparable across languages. Among the fillers were the (fictitious) English surnames '*Shelfer*' (/ʃel · fə/) and '*Leaner*' (/liː · nə/) and the German surnames '*Schilfer*' (/ʃɪl · fɐ/) and '*Liener*' (/liː · nɐ/)[2]. These words were (a) comparable, (b) their syllabic structure was uncontroversial (this is not always the case in English) and (c) peak location could be measured because they contained at least two sonorant segments in their stressed syllables (Figure 4 showed that on stressed syllables with a very small proportion of sonorants, the difference in peak alignment cannot be observed). In the test

2 *The raised '·' indicates the syllable boundary.*

words, F0 peak location was defined as the distance between stressed syllable onset, F0 peak and rhyme offset. In the *Leaner/ Liener* contrast, the beginning of the stressed syllable was determined according to the onset of voicing on a wide-band spectrogram, and the rhyme offset which corresponded to the vowel offset was determined according to beginning and end of strong periodic energy in the second formant (see Fischer-Jørgensen and Hutters, 1982). In the *Shelfer / Schilfer* contrast, the on- and offsets of the stressed syllable were defined as fricative onset in the first and second syllable. 12 English and 12 German subjects took part in the experiment and each produced the relevant test words once. Table 1 shows the results of the measurements.

		English		**German**	
	Subject	*LE - Peak*	*Peak - RE*	*LE - Peak*	*Peak - RE*
Leaner /	1	99	90	203	0
Liener	2	214	65	182	0
	3	190	65	161	0
	4	133	81	94	0
	5	169	78	88	0
	6	112	100	131	0
	7	146	88	198	0
	8	177	50	156	0
	9	219	73	117	0
	10	170	62	228	20
	11	75	91	154	10
	12	120	94	216	0
Shelfer /	1	150	172	261	0
Schilfer	2	209	67	279	0
	3	193	110	235	0
	4	193	177	219	0
	5	138	125	229	30
	6	141	50	229	0
	7	217	117	308	0
	8	188	80	253	10
	9	235	110	269	0
	10	250	40	235	20
	11	209	100	240	0
	12	224	90	235	0
	Mean	174	91	205	4

Table 1 *F0 duration measurements. in ms. Location of F0 peak from left and right edges of the stressed syllable in ms. LE = left edge; RE = right edge.*

Table 1 shows that in all English test items, the F0 peak appeared within the stressed syllable; and in most cases, within the middle (i.e. the distances of the peak from right and left edge were similar). In almost all of the German test words, on the other hand, the location of the F0 peak corresponded to the right edge of the vowel (there were three exceptions, but even there, the peak was very close to the right edge of the syllable rhyme).

The data given in Table 1 is illustrated in Figure 7 below. Figure 7 expresses the location of the F0 peak as a percentage of the overall duration of the stressed syllable In English, the peak appeared at 66% of overall syllable duration, but the German peak appeared at 98%, that is, at the right edge of the stressed syllable. These data are consistent with the corpus evidence; that is, they suggest that in English H*+L, the F0 peak is aligned within the stressed syllable, whereas in German, it is aligned at its right edge (for further detail of peak alignment in English, see van Santen and Hirschberg, 1994). Typical examples of F0 traces aligned with wide-band spectrograms from each language are given in Appendix D.

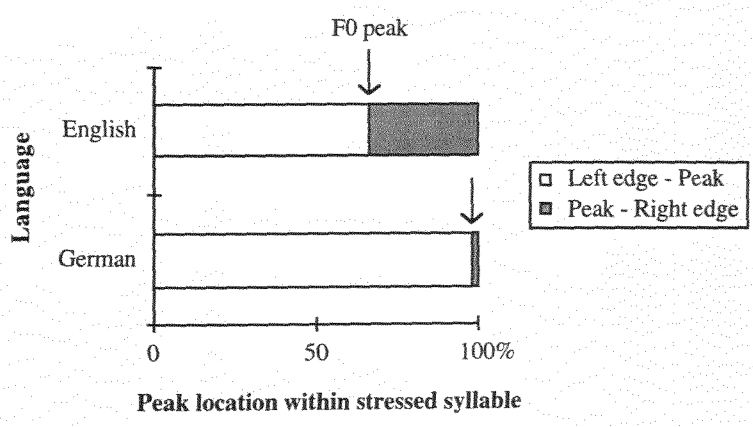


Figure 7 *F0 peak location in the test word expressed as a percentage of the overall duration of the stressed syllable.*

2.1.2 IP- final position

This section discusses the realisation of H*+L in final position. Generally, in IP-final position, English H*+L appeared to be associated with a wider range of auditory impressions than German H*+L. The auditory impressions ranged from rising-falling to falling. In F0, an onglide could be discerned more frequently than on equivalent German examples.

The preceding section showed that in fundamental frequency English and German differ in the way H*+L is realised in non-final position; the English peak is aligned earlier than the German peak. This observation suggests that we may find alignment differences in final position also. Fundamental frequency examples are given in Figure 8 and contrasted with examples from German. Figure 8 shows that in most of the English examples, a peak in F0 can be observed approximately in the middle of the syllable rhyme, even if the syllable onset is not obviously voiced (see <*girl*> which is classed as having a 'voiced onset' but is usually realised as a voiceless unaspirated stop in word initial position). In the German examples, on the other hand, the peak in F0 always appears near the onset of the rhyme.

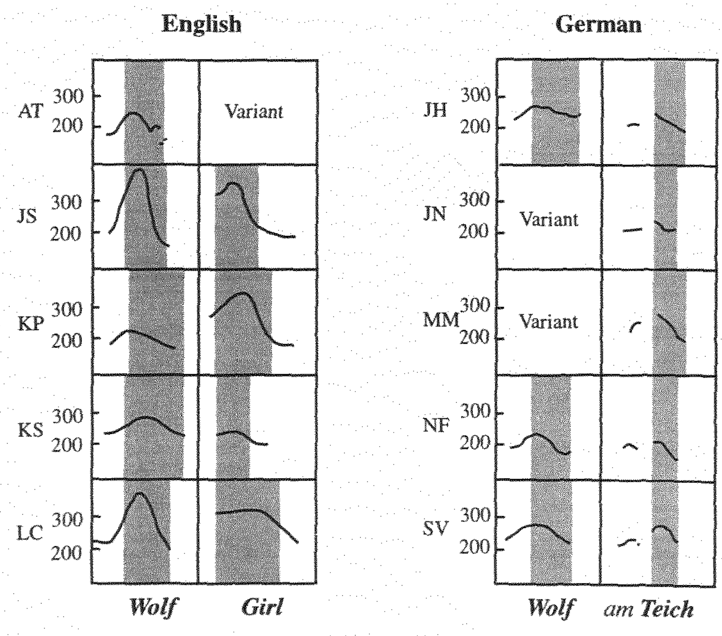


Figure 8 *Fundamental frequency traces of H*+L in IP-final position.*

Context English **German**

Wolf	I'm the Wolf.	*Wolf*	Ich bin der Wolf.
Girl	Little Girl!	*Teich*	Sie wohnt ...am Teich.

2.1.3 Truncation and compression

The data presented in Chapter 3 showed that in Northern Standard German, fundamental frequency is truncated when H*+L is realised on stressed syllables with a small proportion of sonorant segments. English, by contrast, is claimed to compress fundamental frequency in such cases. Figure 9 illustrates the evidence (note that the English and German examples were produced in identical contexts).

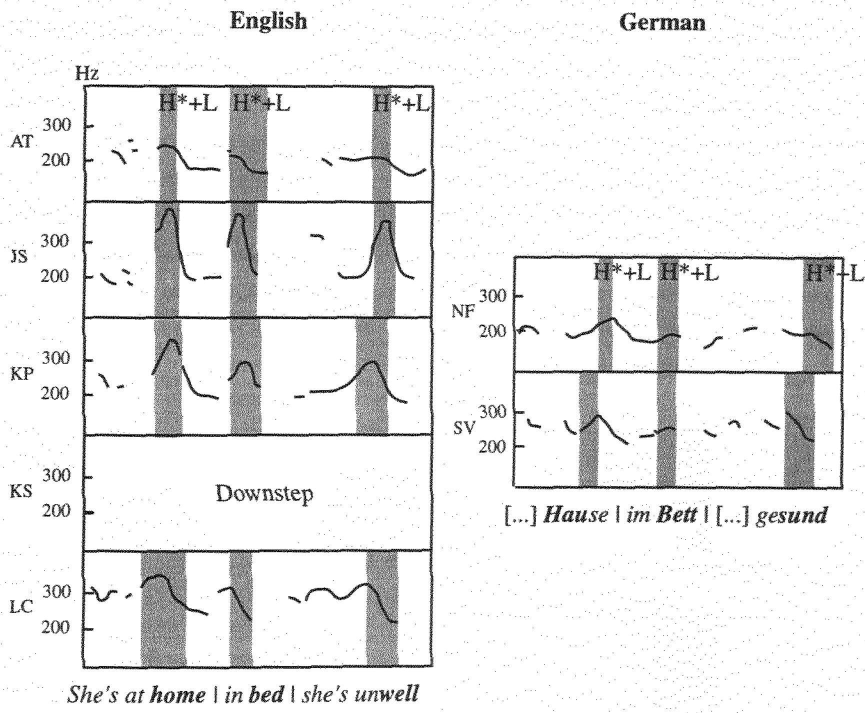


Figure 9 *H*+L associated with a syllable with small proportion of sonorant segments. The traces are lined up on the relevant comparison which is bed / Bett. The transcriptions at the top apply to all examples given.*

The figure shows F0 traces from utterances consisting of three coordinated syntactic constituents. As pointed out earlier, coordination structures in English tend to be produced with the same pitch accents (Schubiger, 1953, Trim, 1959, Crystal 1969), and this appears to apply to German also. The patterns presented in Figure 9 support this assertion, within each language, the three coordinated accent patterns are perceptually equivalent. In fundamental frequency, however, cross-linguistic differences emerge. In the English examples, the three accent patterns look quite similar; we see three falls or rise-falls in F0. The German patterns look rather different; on *Hause*, the accented syllable rises, on *gesund*, it falls and on *Bett* it is level. The accent patterns on the shortest words *bed / Bett* differ especially clearly across languages; in the English examples, the fundamental frequency falls, while on the German example produced in an identical context, it does not fall. This comparison suggests that English compresses H*+L in segmental contexts in which German truncates. Figure 10 below further supports the cross-linguistic difference in pitch accent accommodation suggested by the evidence in Figure 9. The trace of <*nett*> which was discussed in the previous chapter on German is contrasted with two similar English utterances (unfortunately, the comparison is not ideal, because the vowel in *nett* is short, whereas the vowels in *nice* and *kind* are diphthongs; on the other hand, the examples are very similar in meaning and were produced in comparable contexts). The example <*nice*> was produced in an identical context and F0 on nice falls. The German example <*nett*>, on the other hand, shows much less of a fall. Secondly, Figure 10 contrasts the English example <*nice*> which has a voiceless coda with the English example <*kind*>, which has a voiced coda. Despite the segmental differences, the traces on *nice* and *kind* look rather similar, suggesting compression of the F0 fall onto whatever voiced segmental material there is.

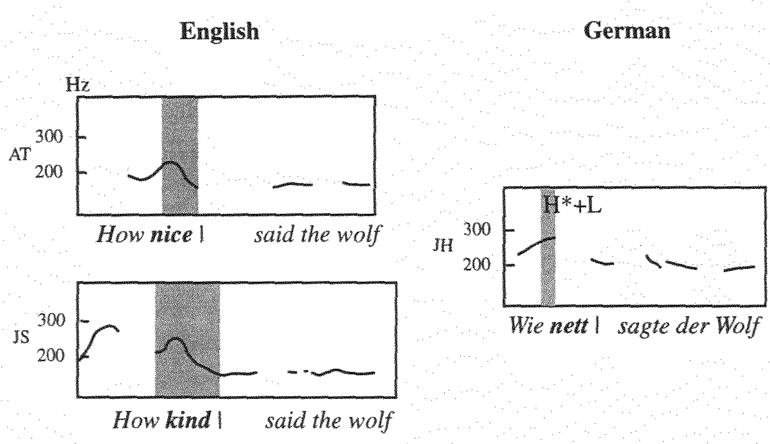


Figure 10 *Cross-linguistic comparison of <nice> and <nett> and monolingual comparison of <nice> and <kind>.*

2.1.4 Summary

The evidence presented in the preceding sections shows that both English and German are characterised by a pitch pattern transcribable as H*+L. Auditorily, the patterns are very similar, but in the acoustic implementation, cross-linguistic differences have emerged. Specifically, H*+L differs across languages in (a) peak alignment and (b) accent accommodation on stressed syllables which offer little scope for voicing. In non-final position, F0 peaks accompanying English H*+L appear to be aligned within the stressed syllable, but those accompanying German H*+L are aligned at the right edge of the stressed syllable. In final position, a rise-fall in F0 may be observed in English, but not, apparently, in German, where the peak appears at or very near the left edge of the rhyme. On syllables with a very small proportion of sonorants, F0 traces are compressed in English but truncated in German. Note, however, that truncation and compression appear to be 'acoustic' rather than 'phonetic' effects. Truncation may be observed in F0, but auditorily, truncated H*+L seems to be auditorily equivalent to non-truncated H*+L[3]. Similarly, compressed H*+L in English appears to be auditorily equivalent to non-compressed H*+L.

Figure 11 summarises the findings on F0 alignment presented so far with schematic F0 traces.

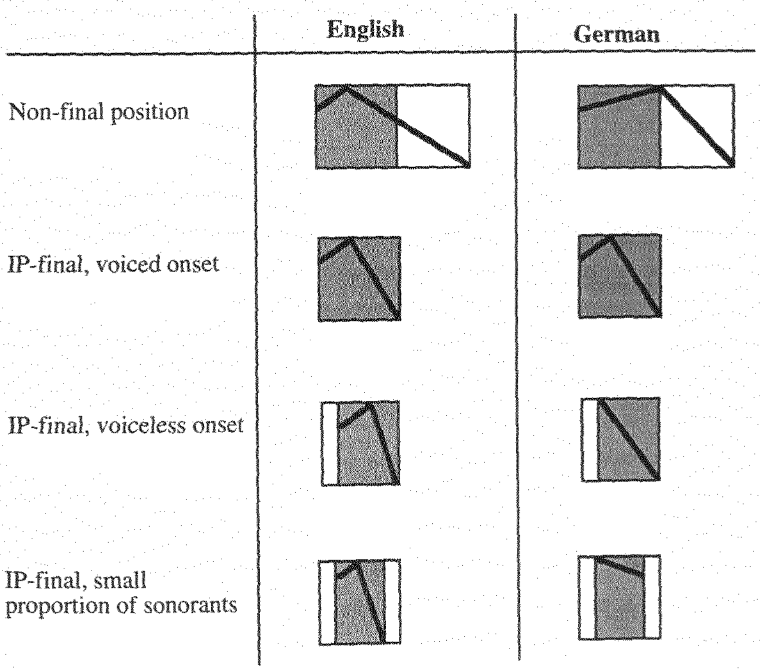


Figure 11 *Schematic comparison of F0 alignment in non-final position, 'long' monosyllables with and without a voiced onset and 'short' monosyllables.*

3 *This comment is based on my own intuition and informal discussions with other native speakers. Experimental verification of the claim, however, is currently lacking, and further research is needed. An investigation testing how native and non-native speakers hear these patterns would add to what we currently know about the relationship between F0 and pitch*

Figure 11 shows that in non-final position, the alignment of H*+L differs across languages. In English, the peak appears within the stressed syllable, but in German, it appears at its right edge. On IP-final monosyllables with a voiced onset, on the other hand, H*+L may look very similar in the two languages. Cross-linguistic differences re-emerge, however, when we compare IP-final accented monosyllables with a voiceless onset; in the English example, we still see a rise-fall in F0, but in the German example, the rising section does not appear to be implemented. Finally, the languages differ in the implementation of H*+L on stressed syllables with small proportion of sonorants (e.g. syllables consisting of a short vowel surrounded by voiceless consonants). F0 is truncated in German, but compressed in English.The data shown schematically in Figure 11 show clearly that the time alignment of accent patterns is language-specific. In the auditory phonetic domain, the patterns in Figure 11 may be described similarly; all involve a fall in pitch. Acoustically, however, the patterns differ substantially. Depending on the position of an H*+L accent in the intonation phrase and the segmental structure of the accented word, F0 patterns may look very similar or quite different across languages. Evidence for language-specific tonal alignment has been presented also in Arvaniti, Ladd and Mennen (1996). The authors investigated tonal alignment in Greek, and describe an LH pitch accent in which a low target is aligned just before the stressed syllable and a high target just after. This finding leads them to question the notion of the starred tone in the AM framework and to call for a clearer definition of the term 'tone' or 'tonal target'. The findings discussed in the preceding sections of the present study support Arvaniti et al.'s findings, and confirm the need for further research into the relationship between tonal association and alignment (see Grice, 1995 for a discussion of the terms association and alignment). Additionally, they show that criteria defining the notion of starredness in a particular language may be established when pitch and F0 patterns produced on comparable data are compared systematically across different positions in the intonation phrase. For instance, in German, the realisation of non-final H*+L with a relatively large onglide on the stressed syllable may, in principle, be hypothesised to reflect a tone sequence L* H L (ignoring the question of '+' signs for the time being), L H* L or H* L. If we compare the non-final realisation of this pattern with that of an equivalent pattern in final position, we see that the relevant targets appear to be HL rather than LHL (i.e. we no longer find evidence of the first low target). This means that we can argue that the star is assigned to the high rather than the low target, because it is the H and not the preceding L which is realised on the stressed syllable in final position. The assignment of the star to the high target receives further support from comparisons of the falling pitch accent in final position on syllables with more or fewer sonorant segments. When voiced segmental material is scarce, it is the position of the high target which is preserved rather than the position of the low target. Taken together, these observations suggest that the pattern in question involves a high target which is starred followed by a low target, i.e. H*+L.

In conclusion, the findings of this section show that in German, a starred tone can be realised in a number of ways in F0, and that criteria for starredness can be felicitously developed when accent realisations are compared (a) in different positions in the intonation phrase and (b) on words with different segmental compositions.

2.2 Prenuclear H*+L

The data presented in Chapter 3 showed that in Northern Standard German, prenuclear H*+L appears in three auditorily distinct variants. Firstly, prenuclear H*+L may be auditorily equivalent to nuclear H*+L and look similar in F0, secondly, the low target may be displaced to the right, beyond the postaccentual syllable (H*+>L) and thirdly, the low target may be deleted and the high target spread up to the following pitch accent (H*>). Equivalent variants of H*+L were observed in the English data and are illustrated with F0 traces in Figure 12. The versions of H*+L where either DISPLACEMENT or DELETION has applied are marked with an arrow. Note, however, that in the English data, examples of H*+L with DELETION occurred noticeably less frequently than in the German data.

Figure 13 illustrates examples of unmodified prenuclear H*+L contrasted with one example of prenuclear H*+L with DISPLACEMENT (marked with an arrow). It shows that prenuclear H*+L does not necessarily look different from nuclear H*+L (see F0 examples of nuclear H*+Ls in Figure 1 of this chapter. The auditory impressions of the three English prenuclear variants do not differ in any obvious way from the ones given for German, and the patterns shown appear to correspond to those accounted for in Gussenhoven's (1984) analysis of British English as an unmodified sequence of H*L, H*L H*L with partial linking (here: H*+L with DISPLACEMENT) and H*L H*L with total linking (here: H*+L with DELETION).

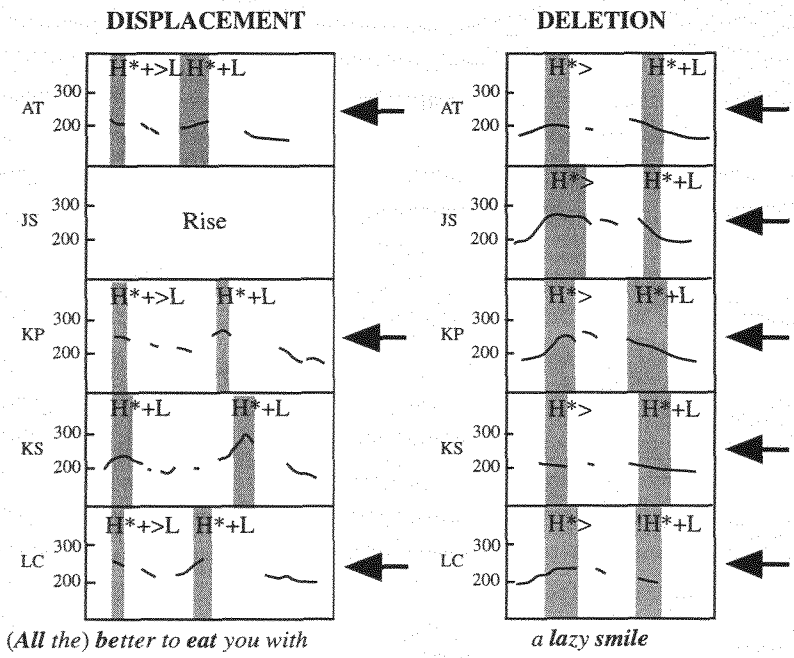


Figure 12 *Examples of prenuclear H*+L in English with and without DISPLACEMENT are shown on the left, and prenuclear H*+L with DELETION on the right.*

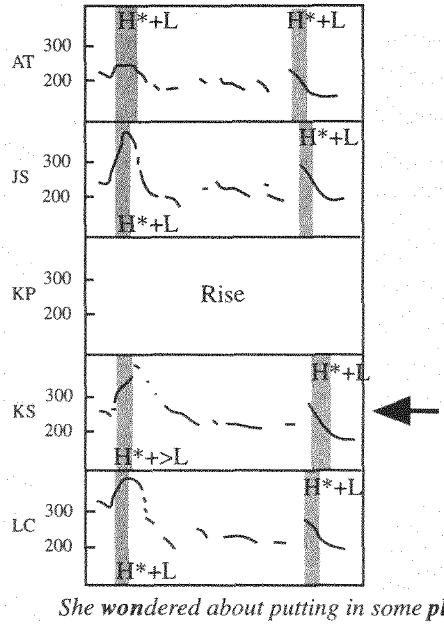


Figure 13
Unmodified prenuclear H+L contrasted with one example of H*+>L*

2.3 Nuclear L*+H

The English corpus contained a very small number of nuclear realisations of L*+H, and no examples of nuclear L*+H in non-final position[4]. Therefore, additional cross-linguistic data on nuclear L*+H were recorded. These data were needed not only to compare the auditory and acoustic realisations of L*+H across languages, but also to test whether realisations of L*+H in English and German would be characterised by a difference in dip alignment comparable to the cross-linguistic difference observed in peak alignment in H*+L.

In Southern Standard British English and Northern Standard German, nuclear rises are often produced in relatively long, carefully read lists of items. Examples of such successions of L*+H accents may be heard, for instance, when the winning National Lottery numbers are read out on television. Accordingly, three English and three German subjects were given lists of numbers and asked to imagine that these were the winning Lottery numbers and they had to read them out on television. This had to be done carefully, so that viewers could write them down while listening. The expectation was that this would lead the subject to read all pre-final numbers in a list with rising nuclear accents, signalling 'continuation', and the final item with a falling nuclear accent, signalling 'finality'. Lists of numbers appeared to be suitable for cross-linguistic comparison because English and German numbers are segmentally relatively similar and many of them have identical stress patterns[5]

In non-final position, nuclear L*+H was elicited with the numbers *ninety / neunzig* and *thirty / dreißig*. They are (a) comparable across languages and (b) contain different amounts of sonorant segmental material in the stressed syllable. Nuclear L*+H in final position was elicited with test items *one / eins* and *six / sechs*. Again, the pairs differ within languages in the amount of sonorant material in the stressed syllable and are comparable across languages. Specifically, *six / sechs* contains a very small proportion of sonorants, and the contrast between *six / sechs* and *one / eins* should show whether the cross-linguistic difference in accent accommodation shown to apply to H*+L applies to L*+H also (i.e. does L*+H in German truncate?). Other numbers were included as fillers.

---

4 *This observation does not imply a general absence of IP-final rises in the English corpus. 'Fall-rises' represented as H*+L H% occurred frequently.*

5 *Numbers under 10 with the exception of 7 are monosyllabic, and numbers up 20 can have the same stress pattern assuming items such as 14 in English are likely to be stressed contrastively on the first syllable when listed with other numbers. Fourteen is stressed on the second syllable in isolation whereas the parallel German word vierzehn is always stressed on the first syllable.*

Each subject read sixteen lists, in which each of the test items was embedded four times, either in a different position in the list or preceded and followed by different numbers. None of the test items were in list-final position.

The subjects were drawn from the same pool as those who took part in the corpus recordings. Again, the German participants were aged between 16 and 18, and had been born and brought up in Braunschweig. They attended the same school as the subjects who had read the corpus materials. The English subjects were undergraduates from Cambridge University and aged between 17 and 21. All had been born and brought up in the South of England.

2.3.1 Non-final position

Auditorily and acoustically, no obvious monolingual differences were observed between realisations of L*+H in non-final position with more or less sonorant material in the stressed syllable, and no auditory cross-linguistic differences emerged. Acoustically, however, a difference in F0 alignment could be observed in most cases (see the acoustic data given below). F0 alignment of L*+H in non-final position is shown in Figure 14.

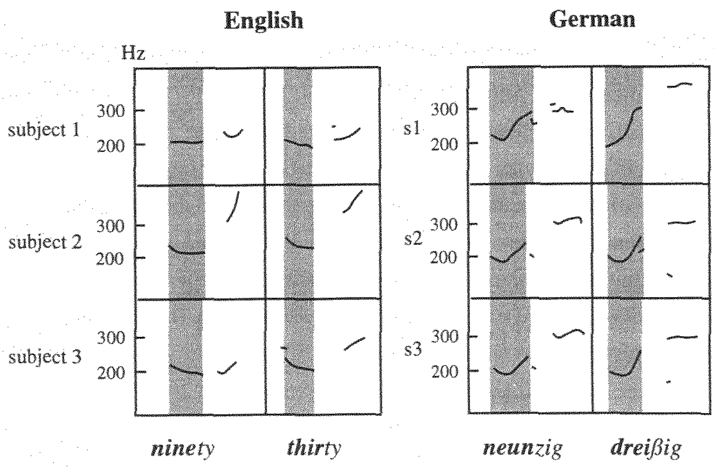


Figure 14 *Realisation of nuclear L*+H in non-final position for English on the right and German on the left.*

Figure 14 shows that in the English examples (left), the low target tends to be realised as a fall in F0, with the dip aligned at the right edge of the syllable rhyme. In contrast, in the German examples, we find the dip appears in the middle of the accented syllable rather than at its right edge (in fact, we find that the English rise looks like the mirror image of the German fall and the German rise like the mirror image of the English fall). Note, however, the difference between the quite uniform German realisations of nuclear L*+H produced in this more controlled study, and the wide range of possible realisations of nuclear L*+H observed in the German data (see section 3.3.1 in the previous chapter). The distinction illustrated in Figure 14 does not appear to be straightforwardly generalisable to different types of texts and different speaking styles. Moreover, the realisations shown in Figure 15 below were observed for English also. These findings suggest that in both English and German, the realisation of L*+H is less stable than that of H*+L.

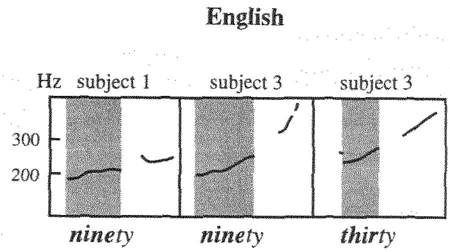


Figure 15 *Other realisations of F0 on the accented syllable in L*+H.*

Acoustic measurements of F0 dip alignment were taken on the data. As in the peak location measurements presented in section 2.1.1.1 of this chapter, the duration of the stretch between the dip in F0 and the right and left edges of the stressed syllable was measured. In one aspect, however, the measurements presented below differ from those taken on realisations of H*+L; there, in the example *Shelfer*, the fricative was included in the measurements, but in L*+H measurements, in the *thirty / dreißig* pair, the voiceless syllable onset was excluded from the duration measurements. The onset was excluded because an examination of the relevant F0 traces had already shown that in *ninety*, where the stressed syllable was fully voiced, the dip was aligned either at the right or at the left edge of the stressed syllable. In no case did the dip appear within the stressed syllable. Including the onset in the dip location measurements of *thirty* would have given the impression that the dip does, at times, appear in the middle of the syllable. The evidence from the test words with a fully voiced stressed syllable, however, suggested that this was not the case, at least not in the data collected in this particular experiment. In the *ninety /neunzig* pair, the beginning of the stressed syllable was determined according to the onset of voicing on a wide-band spectrogram, and the stop burst at the beginning of the second syllable. In the *thirty / dreißig* pair, duration was measured from the onset of strong periodic energy in the second formant of the vowel (see Fischer-Jørgensen and Hutters, 1982) and the onset of the stop at the beginning of the second syllable in the

English example and the onset of the fricative in the German example[6]. Table 2 shows the results of the measurements.

		English		German	
	Subject	LE-DIP	DIP-RE	LE-DIP	DIP-RE
ninety /	1	0	0	90	250
neunzig	1	0	0	80	200
	1	0	310	70	230
	1	0	303	110	200
	2	H*+L H%	H*+L H%	50	270
	2	0	240	40	210
	2	0	293	40	230
	2	350	0	0	260
	3	302	0	90	260
	3	270	0	120	240
	3	351	0	110	220
	3	H*+L	H*+L	130	170
thirty /	1	150	0	0	209
dreißig	1	191	0	92	210
	1	190	0	110	230
	1	189	0	120	160
	2	0	140	60	284
	2	210	0	creak	creak
	2	170	0	90	170
	2	160	0	60	190
	3	180	0	80	200
	3	178	0	60	230
	3	181	0	60	160
	3	184	0	100	200

Table 2 *F0 dip location measurements in test items produced four times by three speakers of each language. Note that '0 - 0' in English speaker 1's realisations of ninety indicates the absence of a dip, that is, the trace was level.*

Table 2 shows that in the English speakers' realisations of L*+H, the F0 dip was observed either at the right edge of the stressed syllable or at the left edge. In no case did the dip appear within the stressed syllable. In the German speakers' realisations of L*+H, on the other hand, the dip appeared within the stressed syllable (two exceptions). Figure 16 below illustrates the results. In the figure, dip location is expressed as a percentage of overall syllable duration. Figure 16 shows that in English, the dip appeared

6 *The syllable offset criteria for thirty / dreißig are not optimally comparable; theoretically, using vowel offset rather than stop and fricative onset would have been more consistent. With respect to taking measurements, however, the criteria chosen appeared to allow more consistent measurements.*

on either edge of the stressed syllable. In German, on the other hand, the dip was aligned within the first half of the stressed syllable. Again, the caveat is that the measurements presented in this section reflect dip alignment in systematically controlled elicitations, and that the findings cannot be generalised to the corpus (a) because the realisations of German L*+H appeared to be more variable in the corpus data than in more controlled realisations and (b) because no comparable examples of nuclear L*+H in non-final position were available in the English corpus. The acoustic measurements for H*+L presented earlier, on the other had, do reflect the F0 alignment observed in examples of H*+L in the corpus. Typical examples of F0 traces aligned with wide-band spectrograms are shown in the appendix.

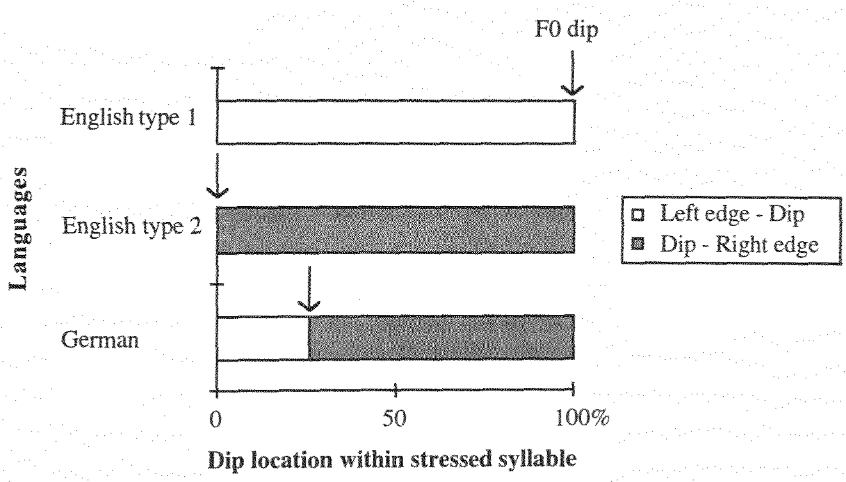


Figure 16 *Dip location expressed as a percentage of stressed syllable duration. The two bars at the top represent dip location at the right and left edge of the stressed syllable in English, and the bar at the bottom illustrates dip location within the stressed syllable in the German data.*

2.3.2 IP-final position

Figure 17 below shows F0 traces for nuclear L*+H in final position. Most obviously, Figure 17 shows that in IP-final position, L*+H appears to be implemented similarly in the two languages. Secondly, the contrast between realisations of L*+H on the German words *eins* and *sechs* show that unlike H*+L, L*+H does not appear to be truncated when voiced segmental material is scarce. This finding appears to suggest truncation in German applies to H*+L, but not L*+H. German L*+H appears to be compressed, just as in English. However, this finding requires further support from systematic experimental evidence. This evidence is presented in Chapter 5 of this study.

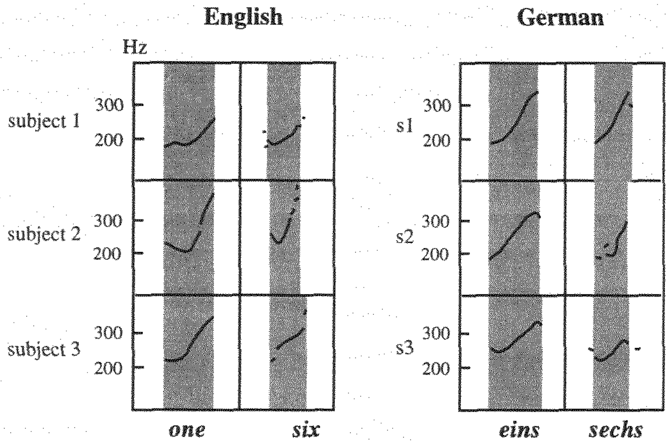


Figure 17 *F0 traces of nuclear L*+H in IP-final position in English (left) and German (right).*

2.3.3 Summary

The corpus materials did not provide a sufficient number of comparable examples of nuclear L*+H. Therefore, additional, systematically controlled data was collected which allowed comparisons of nuclear L*+H in non-final and final positions. The comparison did not show any obvious auditory differences between English and German realisation of L*+H, but appeared to suggest acoustic differences in non-final position. Specifically, in the English data, the lowest F0 was observed either at the right or at the left edge of the stressed syllable, whereas in German L*+H, an F0 dip appeared within the stressed syllable. However, the evidence presented for L*+H in the previous chapter on German suggest that this finding cannot be generalised to less systematically controlled realisations. In the German corpus, a range of possible realisations of L*+H were observed, and some of these were not characterised by a dip within the stressed syllable.

A comparison of L*+H in IP-final position showed (a) that the languages appear to implement L*+H in final position similarly, and (b) that the truncation effect observed for H*+L in German appears to be absent in L*+H.

2.4 Phonological adjustments

This section discusses the cross-linguistic evidence for phonological adjustments. Again, the discussion claims to be exhaustive with respect to the corpus investigated in the present study, but not with respect to the intonational system of Standard Southern British English and Northern Standard German as a whole.

The evidence presented in the previous chapter suggested that in the German data, phonological adjustments DOWNSTEP, DISPLACEMENT and DELETION applied. The analysis of the English data suggested the application of DOWNSTEP, DISPLACEMENT and DELETION also, but additionally, evidence emerged for two further adjustments suggested in Gussenhoven (1984): DELAY and HALF-COMPLETION. DISPLACEMENT will not be discussed any further because no cross-linguistic differences were observed. Differences were observed, on the other hand, in the application of all other phonological adjustments. DOWNSTEP differed across languages in its acoustic phonetic implementation[7]; DELETION applied to nuclear and prenuclear accents in German, but in prenuclear position only in English and HALF-COMPLETION and DELAY were absent altogether in the German data. DOWNSTEP will be discussed first.

Auditory and acoustic phonetic cross-linguistic differences were observed in the implementation of downstepped accents in F0. In the German data, the extent to which a downstepped accent was heard as falling appeared to vary on a continuum between partial and total downstep. The endpoints of this continuum appeared to be auditorily distinct, that is, an accent could be heard as falling or as low level. In the English data, no similar continuum of realisations was observed. All examples of downstepped H*+L were heard as partial downstep, that is, as involving a fall in pitch from the stressed syllable onto the following syllable. At times, pitch appeared to rise slightly on the stressed syllable before falling onto the following syllable, but this realisation did not appear to be categorically distinct from one in which pitch fell throughout the accented word. Fundamental frequency examples of DOWNSTEP in English are illustrated in Figure 18 below. Figure 18 shows that in F0, realisations of !H*+L involve either falling F0 throughout the accented word (e.g. on AT's *careful*) or a rise-fall (e.g. on LC's *minute*).

7 *'Acoustic phonetic' because the cross-linguistic difference was observed in F0 but could also be heard.*

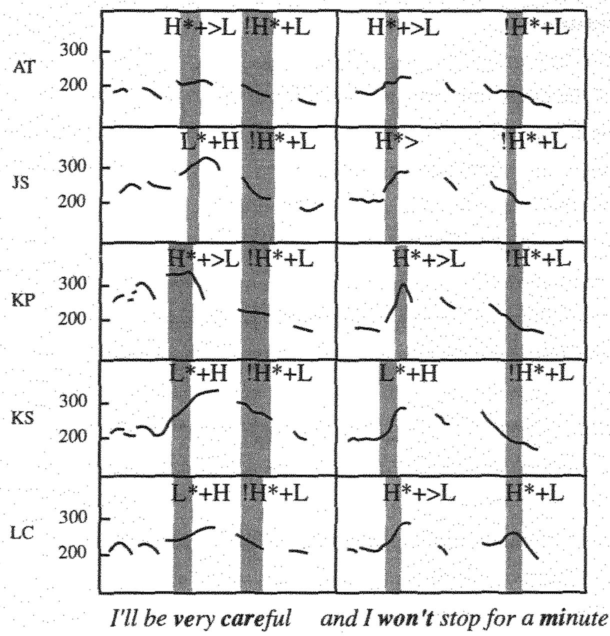


Figure 18 *Examples of downstep in English after L*+H, and variants of H*+L.*

Figure 19 contrasts an example of partial downstep in English with examples of partial and total downstep in German. Note that the examples shown in Figure 19 were produced in identical contexts.

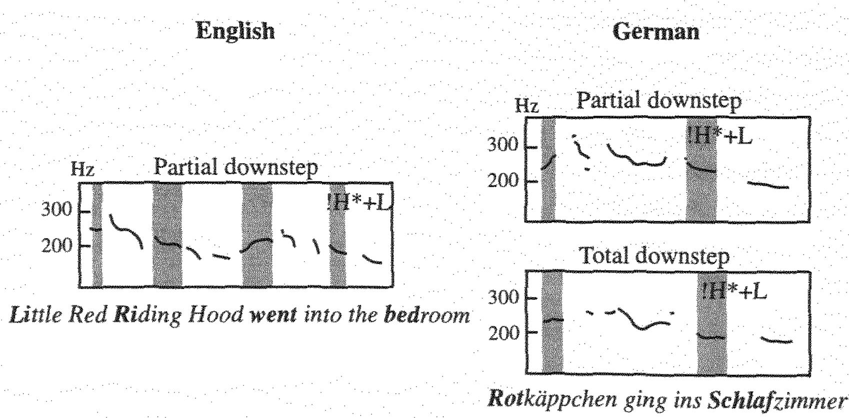


Figure 19 *A cross-linguistic difference in the implementation of downstep.*

Thus, it appears that English !H*+L is implemented as a fall in F0, but the implementation of German !H*+L is more flexible; on a gradiently varying scale, the high target may appear above the low target or at the same level. Experimental support for this claim is presented in Chapter 6.

With respect to the intonational contexts in which DOWNSTEP applies, no cross-linguistic differences were observed. In both languages, !H*+L is downstepped after a variety of accentual configurations, and both appear to have accentual as well as phrasal downstep (for 'phrasal downstep', see van den Berg et al, 1992; briefly, the idea is that not only pitch accents may be lowered relatively to each other within an intonation phrase, and intonation phrases can be lowered relatively to each other also). Figure 20 shows examples of !H*+L after H*+L, H*>, an intonation phrase H*+L 0%, and L*+H.

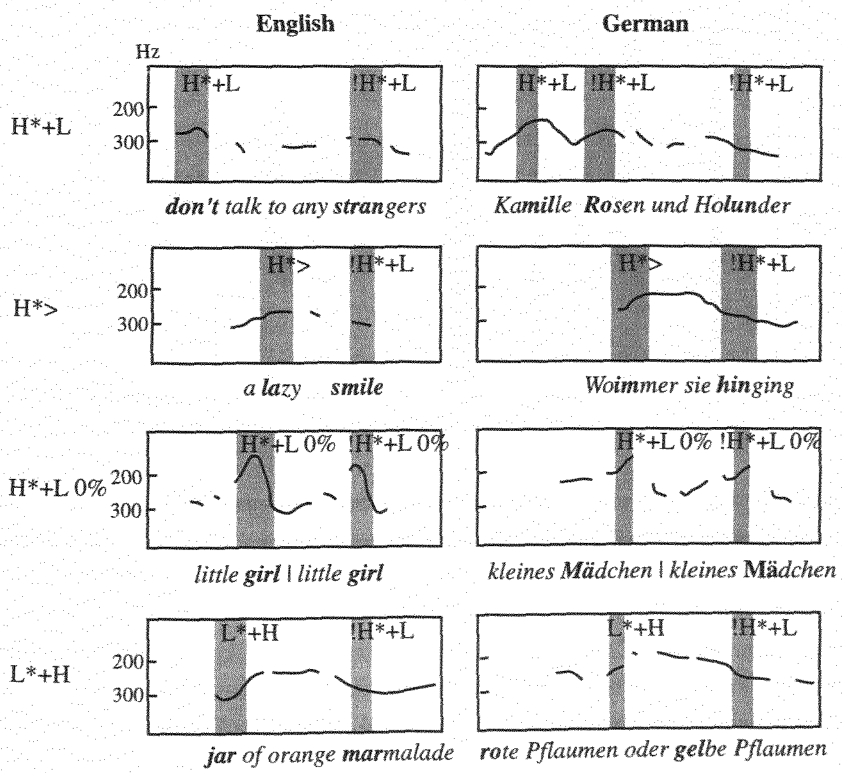


Figure 20 *DOWNSTEP following H*+L, H*>, H*+L 0% and L*+H.*

DOWNSTEP, DISPLACEMENT and DELETION were the only phonological adjustments observed in the German corpus and applied to H*+L only[8]. From the English data, evidence emerged for two further adjustments of H*+L, which appeared to apply exclusively in nuclear position; HALF-COMPLETION, discussed in Gussenhoven (1984), and DELAY, which was originally proposed by Ladd (1983a)

HALF-COMPLETION has been defined by Gussenhoven as 'the failure of the tone to cross the mid-line' (1984:222). Gussenhoven notes that HALF-COMPLETION differs from the other modifications because it does not obviously apply to all nuclear tones; evidence for half-completed rises is elusive. No evidence of half-completed rises emerged from the corpus investigated in the present study. F0 traces which may be interpreted to show an example of HALF-COMPLETION are given in Figure 21 (speaker KP).

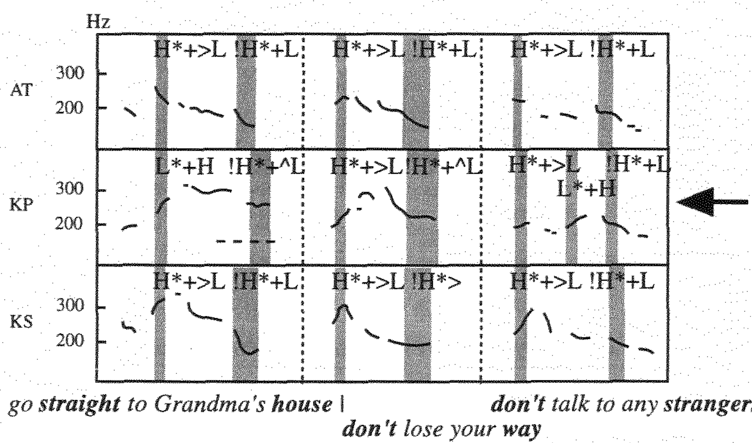


Figure 21 *An example of HALF-COMPLETION (speaker KP). Note that the beginning of the stressed syllable rhymes have been aligned across speakers at the beginning of each IP. H*+L with HALF-COMPLETION was labelled as H*+^L.*

The utterance shown in Figure 21 consists of three separate intonation phrases (the phrase boundaries are indicated by the dotted lines)[9]. For speakers AT and KS, the final lows at the end of each phrase seem to be comparable, but for Speaker KP, we see a

---

8 *Pierrehumbert (1980) posits a distinction between L*+H H-H% and L* H-H%. In the systems used in the present study, this distinction could be analysed as one between L*+H H% with or without DISPLACEMENT. However, no evidence for the distinction emerged in either the German nor the English data.*

9 *Note the 'mixed head' in KPs final IP (i.e. the two different prenuclear accents in <don't talk...>). Mixed heads were extremely rare in the corpus, and generally, 'heads' appeared to represent a single accent choice (see Ladd, 1986: 320).*

staircase effect, that is, the final low in the first IP is higher than the second, and the second is higher than the third. Auditorily, the accents on KP's *house* and *way* are mid-level, and appear to signal to a listener (a) the end of an IP and (b) continuation. H*+L with HALF-COMPLETION was transcribed as H*+^L, with the up-arrow indicating raising of the final L to the level of the preceding H[10]. Note, however, that this analysis is tentative, because the stepping pattern in KP's realisations of H*+L is reminiscent not only of HALF-COMPLETION, but also of phrasal downstep (van den Berg et al, 1992). As the corpus investigated in the present study contained only very few example of HALF-COMPLETION, this issue cannot be resolved at this point. Nevertheless, whether the appropriate account of the patterns in Figure 21 involves H*+L with HALF-COMPLETION or H*+L stepped down in successive phrases, the fact remains that a pattern equivalent to KP's realisations was not observed in German data. One example of a German utterance produced in the same context as the English examples in Figure 21 but without evidence of a pattern resembling HALF-COMPLETION is shown in Figure 22.

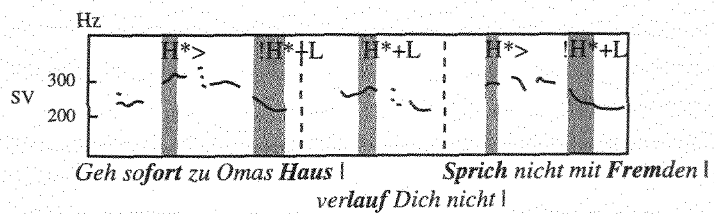


Figure 22 *Absence of HALF-COMPLETION in a German produced in the same context as the English utterances illustrated in Figure 22.*

Gussenhoven (1984) also suggests that English has a modification DELAY. DELAY turns a fall into a rise-fall by delaying the peak of the fall relative to the accented syllable. DELAY is said to signal to a listener that the accented word relates to something 'non-routine' and 'very significant'. In the corpus, DELAY was observed to apply to H*+L but not L*+H and semantic implications of DELAY appeared to be those suggested by Gussenhoven. An example of delayed H*+L on a trochee produced by speaker KP is given in Figure 23 below on the left and contrasted with examples of H*+L without DELAY.

10 *H*+^L in English appears to transcribe the opposite effect of total downstep in German which could have been given as ' vH*+L', but was not transcribed as such because the distinction between partial and total downstep appeared to be gradient. H*+L and H*+^L in English, on the other hand, appeared to be categorically different.*

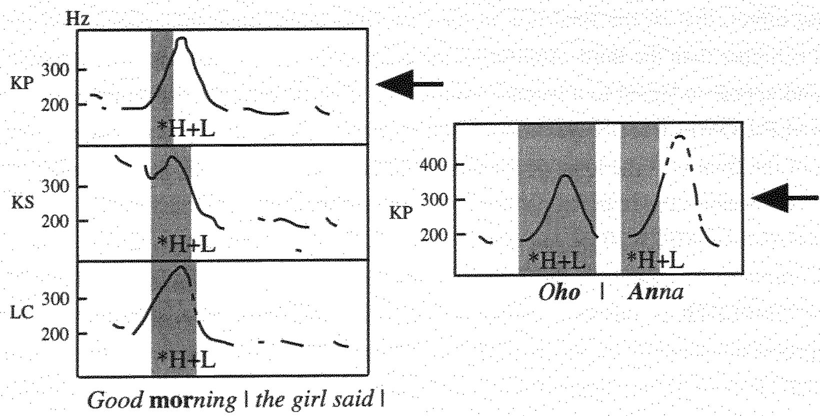


Figure 23 *Delay of H*+L (*H+L) in English.*

An example of DELAY on an IP-final monosyllable and an IP-final trochee is shown in Figure 23 on the right, again produced by speaker KP. The figures show that in KP's realisations, the F0 peak follows the stressed syllable rather than appearing within the stressed syllable. H*+L with DELAY will be transcribed as '*H+L'. *H+L indicates that the H follows the stressed syllable.

This section will be concluded with a brief discussion of the cross-linguistic evidence for DELETION. In the German data, DELETION was shown to apply to nuclear and prenuclear accents. The similarity between patterns represented as H*+L with DELETION (H*>) and L*+H with spreading of the H was discussed and a number of distinguishing criteria were suggested. In the English data, DELETION appeared to apply also, but was restricted to prenuclear position, and in English, the distinction between H*> and L*+H was relatively clear. Figure 24 compares fundamental frequency traces of prenuclear L*+H and prenuclear H*> in English (top) and German (bottom). Figure 24 shows that in the English examples, prenuclear H*> and L*+H are clearly distinct; in L*+H, the stressed syllable is low and followed by a rise in F0 on the following syllable. In H*>, we find the opposite type of alignment; now, a rising stretch of F0 precedes the stressed syllable and the F0 peak appears within the stressed syllable. In the German examples, by contrast, the F0 patterns accompanying L*+H and H*> are less clearly distinguished. Prenuclear L*+H rises throughout the stressed syllable and is followed by a relatively small step up on the postaccentual syllable. In H*>, the stressed syllable rises throughout also because the F0 peak in German is aligned at the extreme right of the stressed syllable, but the following step-up in F0 is absent. Thus, as a direct result of the peak alignment on German H*+L, hat patterns (and nuclear accent patterns) transcribed as L*+H and the H*> are less clearly distinct in German than in English.

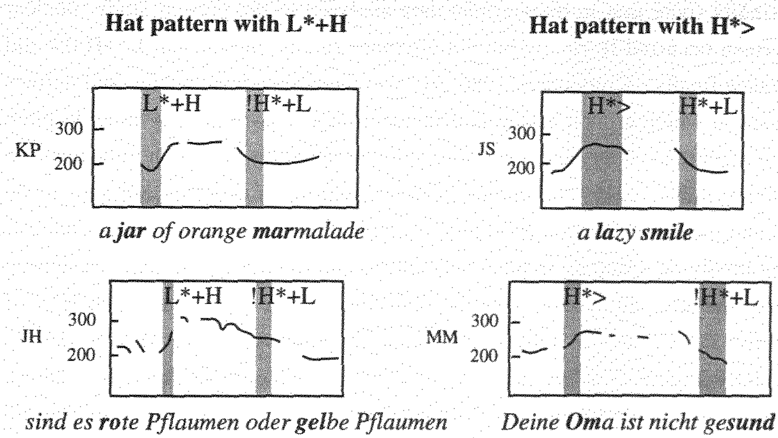


Figure 24 *Comparison of hat patterns with prenuclear H*> and L*+H in English (above) and German (below). Note that voiced syllable onsets are included in the shaded sections.*

As mentioned above, nuclear DELETION was observed in the German data, but not in the English data. Examples of the other type of German nuclear rise-plateau, that is, L*+H 0% were not found in the English data either. However, this finding is unlikely to suggest a general absence of nuclear rise-plateaux in British English. Crystal (1969: 215), for instance, recognises level nuclei. Most other authors, as Crystal points out, do not recognise this pattern (exceptions are Bolinger, 1945 and Kurath, 1964). Crystal does not group his level nucleus either with falling or rising tones, but lists levels separately (he claims that any phonological classification of levels with either falls or rises would be artificial). On the other hand, Crystal points towards two distinct functions of level nuclear tones, one which is similar to the function of a rise, and one which is similar to that of a fall. This functional distinction may suggest that Crystal's level nuclear tone transcribes two different patterns as the same phonological category. Crystal states that a level nucleus occurring at the end of a subordinate or correlative grammatical structure admits of substitution by (usually) a rising-type tone. This pattern could then be a rise-plateau transcribable as L*+H 0%. In other contexts, he points out, a level may function like a fall, and this pattern may be the same as the one transcribed as to H*> 0% or it may refer to H*+L with HALF-COMPLETION (H*+^L)[11]. On the other hand, H*> may also relate to level-nuclei functioning like a rise.

A mid-level nucleus in English has been recognised also by O'Connor and Arnold (1973: 8, 14, 43). The mid-level nucleus is claimed to occur IP-finally but not

11 *Crystal uses the term 'suspended fall' for what is referred to here as a half-completed fall.*

utterance-finally (note that in the German data analysed in the present study, nuclear H*> in German occurred IP-finally but not utterance finally, but nuclear L*+H 0% did occur utterance finally). Cruttenden (1986: 61) also recognises level nuclei, and defines them as cases in which a pitch accent is dependent on a step-up or step-down in pitch between the accented syllable and a previous syllable rather than on pitch movement on the accented syllable itself. The choice between step-up or down is not said to be significant and the most common level tone is claimed to be mid. Cruttenden's nuclear level accent involving a step-up in pitch may correspond to H*> which involves step-up in pitch onto the stressed syllable (see the realisation of the hat pattern with H*> of English subject JS). Cruttenden's level involving a step-down may be H*+L with HALF-COMPLETION (Figure 21 above shows a step down in pitch onto the stressed syllable). Note, however, that while Cruttenden assumes that the difference between his step-up and step-down level nuclei is not distinctive, in the present study the difference between H*> and H*^+L is claimed to be distinctive.

2.5 Intonation phrase boundary specifications

No cross-linguistic differences were observed with respect to intonation phrase boundary specifications. In the German data, a minority of IP-final syllables were shown to require independent tonal specification, and in all cases, the tone required was a high tone (78 out of 917 IP boundaries were labelled as H%). The same applied to the English data; no downward pitch movement was observed at IP boundaries in the absence of stressed syllables. However, 54 out of 885 IP boundaries (5 speakers) were labelled as H%.

Examples of H% following non-final H*+L (left) and final H*+L (right) are shown in Figure 25 below. Note that in the German data, a fall-rise H*+L H% did not occur on IP-final monosyllables.

On the left , Figure 25 shows an example of H*+L H% produced by speaker AT. On the IP-final unstressed syllable, F0 values clearly deviate from those on previous unstressed syllables. By contrast, in the realisations of speakers JS and LC, F0 on the IP-final syllable does not deviate very obviously from preceding F0 values on unstressed syllables nor from the F0 value of the preceding trailing low tone. No downward F0 movement is observed either. The examples on the right show examples of H*+L H% and H*+L 0% produced in identical contexts in final position.

Context

said So he said to Little Red Riding Hood: "As you're such a good girl ..."
girl The girl, remembering her mothers words [..]

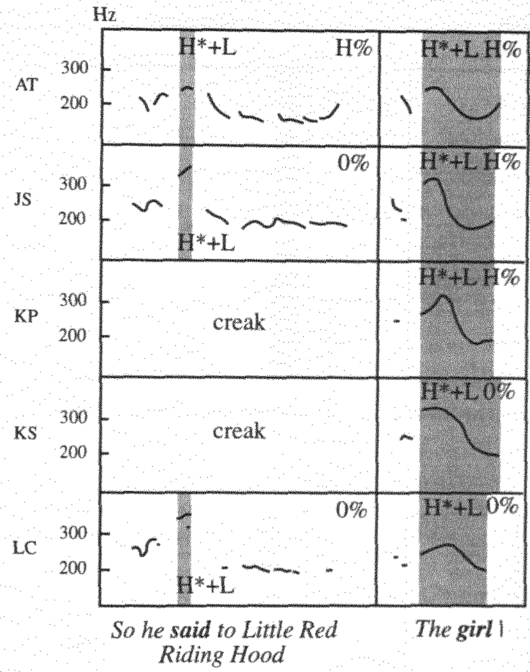


Figure 25 *Boundary specifications in British English following non-final accents (left) and IP-final accents (right).*

Figure 26 below summarises the tonal evidence on IP-final unstressed syllables in the two languages with examples produced in identical contexts. Representative examples are shown of all pitch accent - IP boundary configurations observed in the corpus (apart from the one fall-rise observed elsewhere in the German data; see section 3.6 in the previous chapter). The panels at the top show that in the two languages, after H*+L, pitch and F0 can continue virtually level up to the IP boundary. The panels at the bottom compare examples involving a rise in F0 (and pitch) at the phrase boundary. In the English examples on the left, H% is preceded by H*+L and in the German examples by L*+H. No examples of L*+H H% were found in the English data, although this pattern is common in English (see e.g. O'Connor and Arnold, 1973 or Cruttenden, 1986). As mentioned above, one case of H*+L H% was found in the German data.

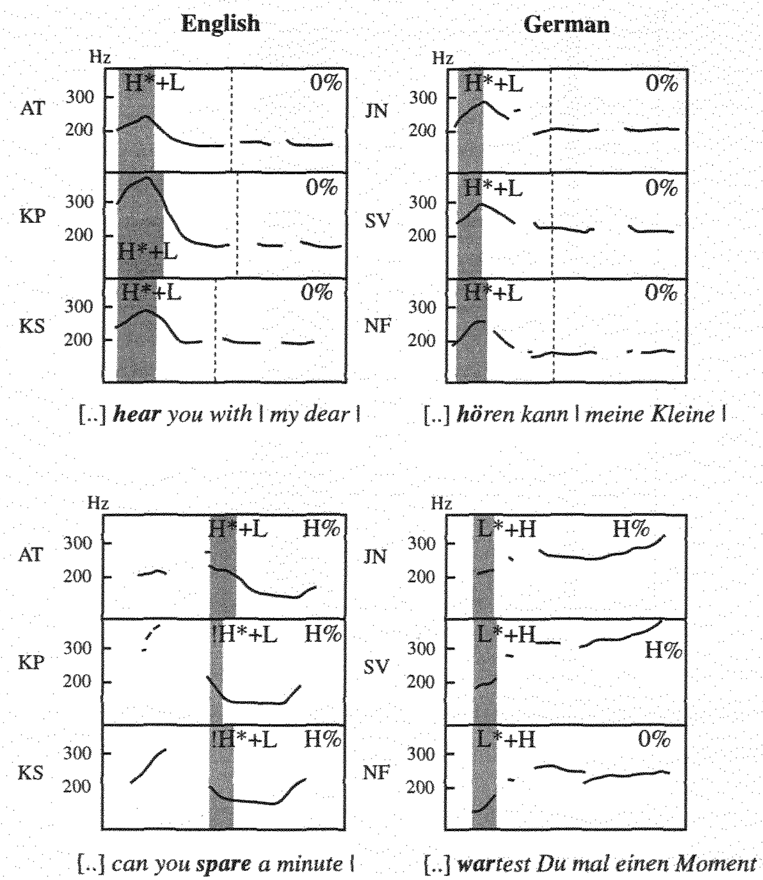


Figure 26 *Boundary comparison in identical contexts.*

3 Summary

This chapter has contrasted the findings from the German data presented in Chapter 3 with parallel evidence from English. The present section will (a) summarise the evidence from Southern Standard British English and (b) discuss the cross-linguistic similarities and differences which have emerged from the comparison of Southern Standard British English and Northern Standard German.

Just as the German data presented in the previous chapter, the English data was accounted for with two basic pitch accents H*+L and L*+H and a set of phonological adjustment rules. These model the changes the basic accents are assumed to undergo in continuous speech. The phonological modelling of H*+L in English is summarised in

Figure 28 below. Evidence of five phonological adjustments emerged: DOWNSTEP, HALF-COMPLETION, DISPLACEMENT, DELAY and DELETION. The auditory impressions for H*+L, H*> and H*> did not differ very obviously from the equivalent German realisations and are therefore schematised in the same way. The auditory impressions of H*+L with HALF-COMPLETION and DELAY in Figure 27 are transcribed as !H*^+L and *H+L and they sounded like a step-down to mid-level, and a rise-fall respectively[12].

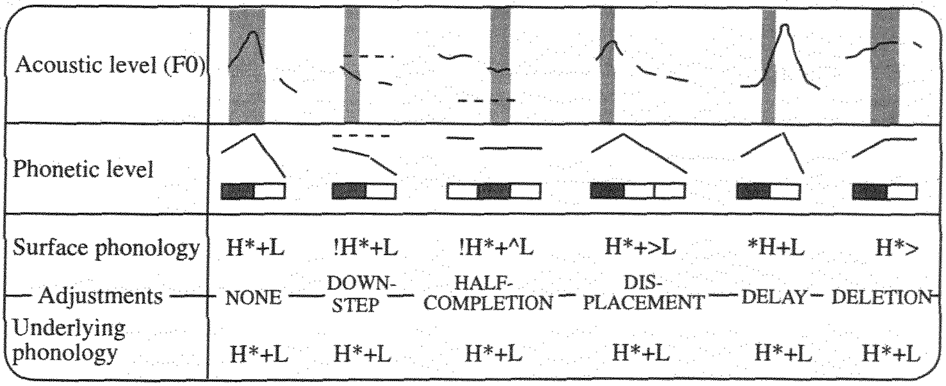


Figure 27 *Categorical phonological adjustments of H*+L. Note that the corpus did not contain examples of HALF-COMPLETION in non-final position or examples of HALF-COMPLETION without DOWNSTEP.*

Figures 28 and 29 below illustrate the acoustic variation effects observed in the English data. In the German data, three acoustic variation effects were observed; a gradient expansion of the onglide in unmodified realisations of H*+L, gradient final peak lowering in !H*+L and truncation when the stressed syllable offered little scope for voicing. The English data suggested that the onglide in H*+L was gradiently variable also. Gradient final peak lowering and truncation, on the other hand, were not observed; on stressed syllables with a small proportion of sonorants, F0 was compressed. Figure 28 illustrates the findings for expansion. Note that, in principle, expansion can apply to modified versions of H*+L also, but an auditory distinction between the endpoints of a continuum similar to that observed in unmodified H*+L was not observed.

---

12 *Note that !H*+^L transcribes a case of H*+L in which two phonological adjustments have applied: (1) DOWNSTEP and (2) HALF-COMPLETION. Examples of HALF-COMPLETION without DOWNSTEP may be assumed to exist ,but no examples were found in the English data. More generally, HALF-COMPLETION is presented with a note of caution. The effect of this modification is similar to the effect of phrasal downstep, and may be more adequately accounted for as such. This finding suggests (a) that we need more evidence of the acoustic phonetic effects of phonological adjustments and (b) that their theoretical status needs further consideration.*

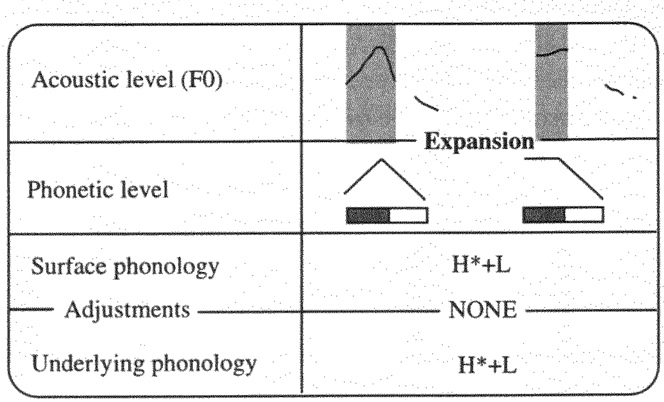


Figure 28 *Expansion: gradient acoustic variation in English.*

Figure 29 illustrates compression of H*+L on words offering successively less scope for voicing. It shows that H*+L is implemented as a rise-fall in F0 in non-final position. On IP-final accented monosyllables, whether 'long' or 'short' a similar rise-fall can be observed.

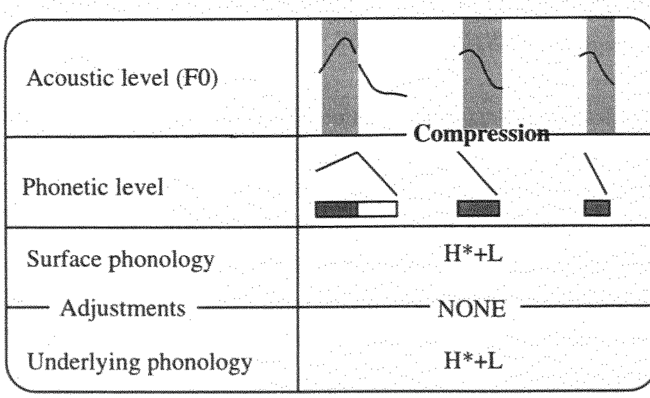


Figure 29 *Compression in H*+L: gradient acoustic variation in English.*

Figure 30 illustrates the phonological modelling of L*+H. In the corpus, examples of nuclear L*+H were scarce; and therefore, additional evidence was collected in a study designed specifically to elicit rises.

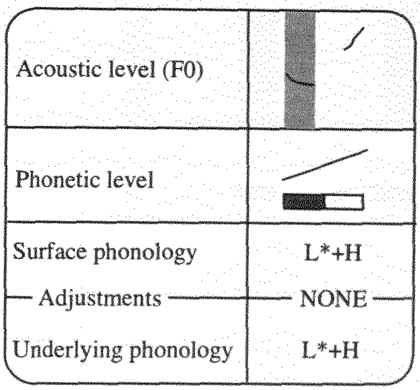


Figure 30 *Phonological modelling of L*+H in English.*

The additional experimental evidence of L*+H allowed a comparison between realisations of H*+L on words offering successively less scope for voicing. Figure 31 shows that English L*+H is compressed, just like English H*+L.

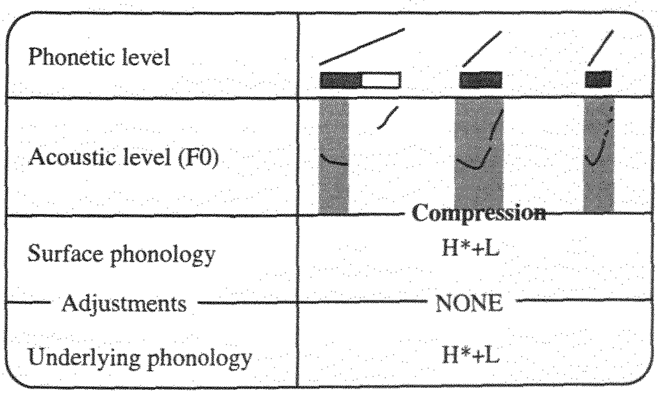


Figure 31 *Compression in English L*+H: gradient acoustic variation.*

Moreover, the additional data allowed a comparison of L*+H in German on words with fewer and fewer sonorant segments. The data showed that unlike German H*+L, German L*+H is compressed. Compression in German L*+H is illustrated in Figure 32.

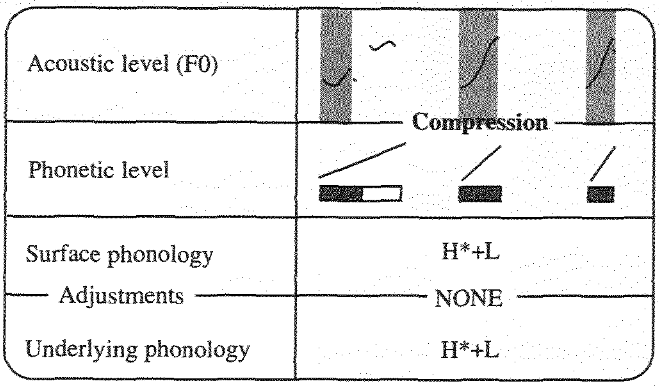


Figure 32 *Compression in German L*+H; gradient acoustic variation.*

In summary, the analysis of Northern Standard German presented in Chapter 3 and the comparison between Northern Standard German and Southern Standard British English presented in the present chapter have suggested the following cross-linguistic similarities and differences:

(1) Underlying phonological structure

It is suggested that the basic intonational inventory of English and German is shared; both languages have pitch patterns which can be accounted for as combinations of H*+L, L*+H and a boundary tone H%. Additionally, in neither language are tonal specifications of IP-boundaries obligatory.

(2) Phonological pitch accent adjustments in continuous speech

The speech data from both languages were interpreted to show evidence of phonological adjustments DOWNSTEP, DISPLACEMENT, and DELETION. Additionally, English H*+L appeared to be modified by HALF-COMPLETION and DELAY (note, however, that the evidence for HALF-COMPLETION was limited). In both languages, DOWNSTEP applied to prenuclear and nuclear accents but DISPLACEMENT applied to prenuclear accents only. DELETION applied to prenuclear and nuclear accents in the German data, but only to prenuclear accents in the English data. No evidence for HALF-COMPLETION and DELAY was found in German.

(3) Acoustic realisation

Both languages exhibit gradient 'expansion' of the onglide in H*+L (i.e. in both languages, H*+L can be realised with a rising, and intermediate or a level onglide).

Cross-linguistic differences were established in peak alignment; in English, the peak is aligned within the stressed syllable, but in German, it is aligned at its right edge. Differences in dip alignment in L*+H emerged from additional more controlled data, but no evidence for a difference in dip alignment was observed in the corpus (this is why 'dip alignment' is listed in brackets in Figure 33 below). Also, the languages differ in 'final peak lowering'. Gradient final peak lowering applied in the German, but not the English data. Finally, the languages appeared to differ in pitch accent accommodation effects; English compresses H*+L, but in German, H*+L is truncated. Figure 33 summarises the observed differences in acoustic realisation.

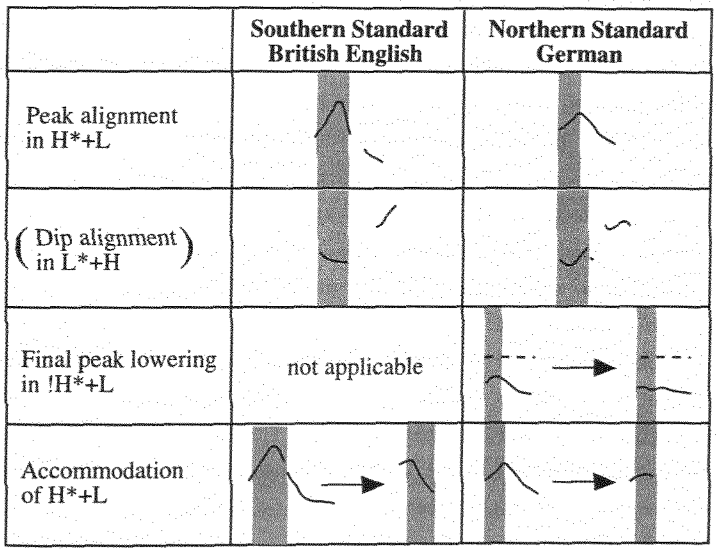


Figure 33 *Cross-linguistic differences in fundamental frequency implementation. The arrow indicates a gradient rather than a categorical change.*

(4) Phonetic realisation

No obvious auditory differences were established in the realisations of unmodified and modified basic categories assumed to be shared. However, the analysis also showed that (a) auditory differences cannot necessarily be assumed to directly reflect phonological distinctions and (b) the absence of an auditory difference across languages does not necessarily reflect the absence of acoustic differences.

In H*+L, the extreme endpoints of a gradiently variable continuum between rising and level onglides were heard as apparently categorically different in the two languages. However, a large number of intermediate onglides were heard also, which could not obviously be classified as either 'rising' or 'level'. This finding was interpreted to suggest that the auditory distinction, which could be made when the endpoints of the continuum were involved, was not, in fact, phonological[13]. Secondly, in German, an auditory distinction between partial and total DOWNSTEP could be made at times, but again intermediate realisations were observed also. Despite an apparently categorical difference characterising the endpoints of a gradiently variable scale, final peak lowering did not appear to be distinctive.

The observed difference between truncation and compression in the acoustic realisation is striking, and one might expect this difference to be reflected in the auditory realisation. However, truncated and compressed realisations of H*+L did not appear to be perceived as such by native speakers. In English, H*+L falls in F0 and is heard as having falling pitch, regardless of how little sonorant material the stressed syllable may contain. In German, on the other hand, H*+L is truncated in F0, but is also heard as having falling pitch by native speakers. This finding can be interpreted to suggest a separation between phonetic ('auditory') and acoustic levels of intonational representation.

Finally, in German H*+L, the F0 peak appears to be aligned later in the stressed syllable than in English H*+L. As pointed out in section 2.1.1.1 of this chapter, the auditory distinctions associated with this timing difference cannot be described straightforwardly. Indeed, there may be no salient differences when individual tokens are compared across languages. More generally, however, it is possible that the difference in peak alignment is reflected in German H*+L sounding more often like a 'rise-fall' in pitch than English H*+L, which is more 'fall' like. This observation may relate directly to the discrepancy observed in the literature between a broader consensus on the intonational categories of English than is the case for German (see Chapter 1, section 2.1.3). As a direct result of the German peak alignment, the rising part of what is auditorily described as a 'fall' is often very salient, both in pitch and in F0. Additionally, German H*+L is less clearly distinguished in its realisation from L*+H, which also involves rising pitch and F0. In prenuclear position, for instance, H*+L and L*+H may sound and look very similar, especially when L*+H is followed by a low target. In English H*+L and L*+H appear to be more clearly distinguished; the earlier peak alignment in H*+L makes the falling accent much less 'rise-like' than the German counterpart and distinguishes it clearly from L*+H. Thus, it appears that auditory categories such as 'falling' and 'rising' are very well suited to the analysis of English, because phonetic realisation and phonological patterns happen to match relatively well. In German, on the other hand, phonetic realisation and phonological patterns match less well, a 'fall' may be 'rising-falling' as well as 'falling', but this difference does not reflect a phonological difference. Moreover, 'rising' also describes a further,

---

13 *A phonological distinction is posited in the Beckman-Pierrehumbert system and transcribed as L+H* L-L% and H* L-L%.*

categorically different phonological category (L*+H). In the light of this overlap, it is no surprise that disagreement is found in the literature on German intonation concerning the basic phonological inventory. This observation may be interpreted as offering support for a description of both languages in terms of autosegmental-metrical categories, which account for pitch accents as targets implemented by language-specific realisation rules rather than in terms of auditory categories which are not directly comparable.

Pitch Accent Accommodation Effects

Chapter 5

1 Introduction and background

Ladd (1996) suggests that cross-linguistic differences among 'intonation' languages may be classified using a taxonomy of systematic phonological and phonetic parameters. Following a well-established tradition for the description of differences in segmental phonology and phonetics within British linguistics, he broadly suggests 'semantic', 'systemic', 'realisational' and 'phonotactic' distinctions in intonational structure. This chapter is concerned with 'realisational' distinctions between English and German, defined by Ladd as differences in phonetic detail involving no effect on the inventory of phonological contrasts. An understanding of realisational distinctions is relevant when one aims to establish an inventory of phonologically different intonational contrasts in the structure of a particular language; one needs to know which surface patterns represent systematically different realisations of one and the same underlying pattern and which indicate a difference in phonological representation. Similarly, realisational distinctions are relevant to cross-linguistic comparisons; an observed difference in accent realisation may reflect a difference in phonological structure or may be restricted to specific segmental contexts and are then best accounted for as a difference in the way in which essentially the same structure is phonetically realised.

Evidence for a realisational difference between English and German was presented in Chapter 4; on IP-final accented syllables with a small proportion of sonorants, H*+L appears to be truncated in German, but is compressed in English. However, the number of tokens in the corpus on which this difference could be observed was very small. Moreover, evidence of accent accommodation on L*+H was scarce. In the present Chapter, two experimental studies will be presented which investigate pitch accent accommodation effects in rises and falls in more detail[1]. The first experiment contrasts the realisation of H*+L and L*+H in English and German on words with successively less scope for voicing, and the second investigates the accommodation of L*+H followed either by 0% or by H% in German.

1 *An article based on the first study will appear in 'Journal of Phonetics' in 1998.*

2 Truncation and compression

One segmental context in which cross-linguistic realisational distinctions in intonation frequently surface is when the voiced segmental material available for pitch accent realisation is limited, for instance on the English word *shift*, where a short vowel is surrounded by voiceless consonants, or on the German word *Schiff*, where the same applies. Experimental evidence has illustrated two strategies which languages appear to adopt in such cases, and these have been referred to as 'compression' and 'truncation'. The term 'truncation' was suggested by Erikson and Alstermark (1972), who investigated the realisation of accent II in Swedish as a function of phonological vowel length. The authors discuss two ways in which the F0 contour of accent II may be modified with decreasing vowel duration; 'truncation', where a falling contour merely ends earlier, in the absence of rate adjustment, and 'rate adjustment', later referred to as 'compression', where, starting from the same level, the F0 contour of short vowels falls more rapidly than that of long vowels. Rate adjustment was taken to reflect a temporal reorganisation of the tonal contour, whereas truncation was not. Truncation and compression are illustrated in Figure 1, which is adapted from Erikson and Alstermark (1972).

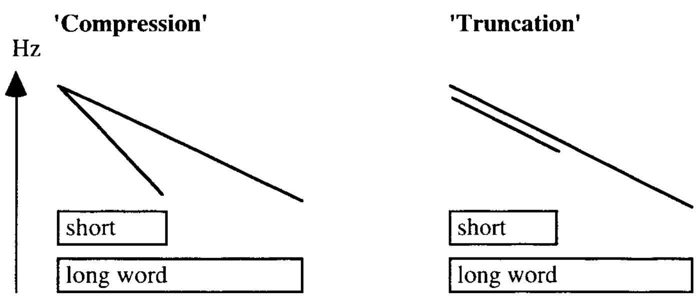


Figure 1 *Compression and truncation in words of different length.*

The authors found that a shortening of the vowel segment results in truncation of accent II, and in Stockholm Swedish, where accent II is often described as 'falling', truncation tends to obliterate most of this fall. Further work on accent realisation in Swedish was carried out by Bannert and Bredvad (1975) who replaced the term 'rate adjustment' by 'compression'. Their results show that the application of truncation and compression is dialect specific, that is, some dialects of Swedish truncate and others compress.

Grønnum (1989) investigated fundamental frequency patterns on longer and shorter 'stress groups' (defined as consisting of a stressed syllable and succeeding unstressed syllables, if any) in a number of Danish dialects, and found that all of them truncate short stress groups. In her work, Danish is described as a language characterised prosodically by one type of F0 pattern with a number of surface variations depending on the length of the stress group the pattern is associated with; in other words, in Danish, different surface realisations do not necessarily point towards different underlying phonological structures. In the same paper, Grønnum also provided some evidence for truncation in Northern Standard German. However, her German data are not straightforwardly interpretable; they appear to offer evidence not only for truncation, but also for compression, specifically for rising fundamental frequency patterns at phrase boundaries.

For English, on the other hand, there has not been a suggestion that speakers make use of truncation when segmental material is short; instead, it appears that pitch patterns are compressed. In fact, Ladd (1996) describes English as a 'compressing language par excellence'. Systematic experimental evidence, however, is not available.

A difference between English and German pitch accent realisation appears to be predicted not only by the comments in the literature but emerged also from the corpus analysis presented in Chapter 4. Figure 2 repeats one example.

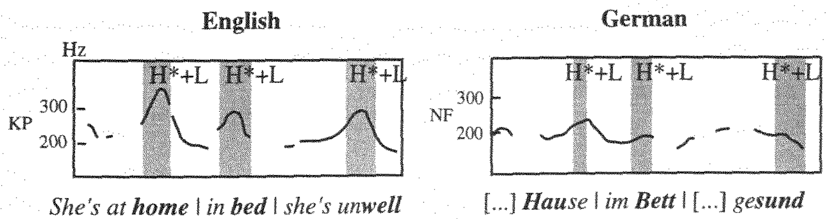


Figure 2 *Compression on bed is shown for English (left) and truncation on Bett for German (right).*

For each language, fundamental frequency traces are shown of three successive intonation phrases which happen to correspond to three syntactic phrases. In both languages, each IP contains one nuclear falling tone. For English, the F0 trace in Figure 2 shows three successive falls in F0. For German, however, the F0 trace shows clear evidence of a fall only in the first and the last IP, that is on *Hause* and *gesund*. In the second IP, that is, on *Bett*, we find little evidence of a fall. However, auditorily, to a native speaker, the accent on *Bett* appears to be no different in type from the preceding accent on *Hause* ('home') or the one following on *(nicht) gesund* ('unwell'). Nevertheless, in an AM analysis, which offers the possibility of distinguishing between an accent H*+L and an accent H*, one might take the F0 evidence to suggest that the speaker assigned H*+L accents to the first and last IP, and an H* accent to the one in the middle. After all, the word *Bett* is very short and the apparent fall in pitch may at least to some extent be due the auditory impression given by the surrounding falls. However, transcribing *Bett* as H* rather than H*+L may overlook a relevant generalisation: nuclear falls may be realised differently on words with different segmental structures. Support for the accent on *Bett* being of the same type as that on *Hause* and *gesund* comes from comments in the British school literature on the intonational patterns of coordinate structures. These have been noted frequently, for instance by Trim (1959), Schubiger (1958), Crystal (1969), Halliday (1967) and others. All authors point out that coordinate structures are characterised by some degree of pattern repetition. In accordance with this prediction, the English F0 traces in Figure 6 appear to exhibit the same pattern on each conjoint, and the German F0 traces are characterised by the same pattern on the initial and the final conjoint. This observation appears to suggest that the phrase between them may be of the same type. One may hypothesise, then, that underlyingly, the accent on *Bett* is falling, but as the word is very short, the fall has been truncated and surfaced as an apparently 'high' accent without evidence of a fall. Similar evidence of truncated rises was not observed.

The following sections describe a production study which was carried out to provide comparable evidence for pitch accent realisation on syllables with a small proportion of sonorants in German and English. A cross-linguistic difference in accent realisation was hypothesised for falling accents. Rises, on the other hand, were hypothesised to compress.

3 Experiment I

3.1 Method

3.1.1. Materials

Six 'surnames' with successively less scope for voicing were embedded in carrier phrases designed to elicit rising and falling accents on the test words. The English test items were *Sheafer* [ʃiːfə], *Sheaf* [ʃiːf], *Shift* [ʃɪft] and the German equivalents were *Schiefer* [ʃiːfɐ], *Schief* [ʃiːf], *Schiff* [ʃɪf]. The duration of voiced material was manipulated by reducing the number of syllables (two vs. one) and reducing phonological vowel length (/iː/ vs. /ɪ/). These particular test items and the particular way of reducing the proportion of sonorants in the test words were chosen for the following reasons. Firstly, the aim was to provide experimental data from naive speakers rather than from a trained phonetician. This imposed restrictions on the number of test items and fillers which could be included in the materials. Two lists of materials were intended to be produced, with one intonation contour each; in the first list, the test word was supposed to be produced with a nuclear falling tone, and in the second with a nuclear rising tone. All other intonational parameters should, ideally, be held stable (how this was achieved will be detailed below). The starting point for the choice of materials was the corpus analysis discussed above. There, clear examples of truncation were found on words with short vowels, surrounded by plosives (e.g. *Bett* 'bed'). Thus, a word of this type was included. Voiceless stops, however, tend to result in local disturbances in the F0 contour, making measurements difficult. Words containing a short vowel surrounded by voiceless fricatives such as *Shift / Schiff* appeared more suitable. Next, words with more scope for voicing were required for comparison. Because of vowel-intrinsic differences in F0, which might distort the measurements, the short vowel in *Shift / Schiff* was replaced by its phonologically long counterpart (*Sheaf / Schief*). The third and final length condition involved adding another syllable to the mid-length word. This was motivated as follows. In English and German, the acoustic realisation of nuclear falling and rising accents appears to involve maximally two syllables, giving a pitch peak within the first syllable and a fall onto the second. After that, we find no pitch changes of similar magnitude. Thus, it appeared possible to add further sonorant material to the word in the mid-length condition (*Sheaf / Schief*), while still adding this material to the prosodic domain which appears to be relevant to the realisation of a rising or falling nuclear accent.

A comparison of three rather than two length conditions in the experiment was motivated by the specific structure of the experimental materials. As was shown in Figure 6, in the German corpus, words with short vowels surrounded by plosives appeared to show no evidence of a fall in F0. This might indicate that truncation is a phonological process of L-deletion rather than a gradient acoustic effect. Including three length conditions in the materials, then, appeared to offer an opportunity to check whether truncation would apply gradiently between the mid-length and the longest condition.

The test items were embedded in carrier phrases and distributed between two lists. Rising accents were elicited via yes/no questions, and falls via simple statements. In both languages, yes/no questions frequently end in L*+H H% , whereas simple statements often end in H*+L Ø%. The test item was placed in phrase-final position, and followed by a phrase in apposition which was added as a control. Appositions tend to be produced with the same intonational pattern as the word or phrase they modify, and were therefore assumed to show evidence of the underlying phonological specification of a test word in case of truncation. Each list of sentences was preceded by a short introductory paragraph, given in (1) below. Carrier phrases designed to elicit falls are given in (2); those for rises in (3).

(1) *Anna and Peter are watching TV. A photograph of this week's National Lottery winner appears. Anna says: Look, Peter!*

(2) Carrier phrase for falls (test items are underlined):

English	*It's Mr. <u>Shift!</u> Our new neighbour!*
German	*Das ist doch Herr <u>Schiff</u>! Unser neuer Nachbar!*

(3) Carrier phrase for rises:

English	*Isn't that Mr. <u>Shift?</u> Our new neighbour?*
German	*Ist das nicht Herr <u>Schiff</u>? Unser neuer Nachbar?*

The materials were intended to elicit from naive subjects lists of sentences with identical intonational structures. Each subject was asked to begin by reading out the introductory paragraph which was supposed to set the scene, followed by the carrier phrases with the test items, one after the other. The list with the carrier phrases designed to elicit falls was read first, followed by the one designed to elicit rises (i.e. rises and falls were not mixed). On each list, the test items were semi-randomly interspersed with filler items (75% fillers on each list; the fillers were different names of one or two syllables). All items were read in the same order by all subjects; the longest word was always first, followed by the shortest and finally the mid-length word (with intervening fillers). In the written instructions, subjects were told that they were recording a 'pronunciation drill' for non-native speakers, and that it was therefore very important that all sentences were read 'the same way'. Non-native speakers would have to repeat these sentences and would find this difficult if they had not been read 'in the same way'. Additionally, subjects were asked to speak 'normally', i.e. not to pronounce words with exaggerated care as it was important for learners to hear 'normal', everyday German or English.

3.1.2. Subjects

12 German and 12 English female subjects read the materials. The English subjects spoke varieties of Southern British English, and the German ones a variety of Northern Standard German. The English subjects were undergraduates from Cambridge University aged between 18 and 23 and were recorded in a sound-treated room in the Phonetics Laboratory in the Department of Linguistics at the University. All speakers had been born in the south of England, and described themselves as 'middle class' or 'upper middle class'. The German recordings were made in a quiet room at a secondary school in Braunschweig. The speakers were 16-18 years of age, and would be rated as 'middle class'. All had lived in Braunschweig from birth.

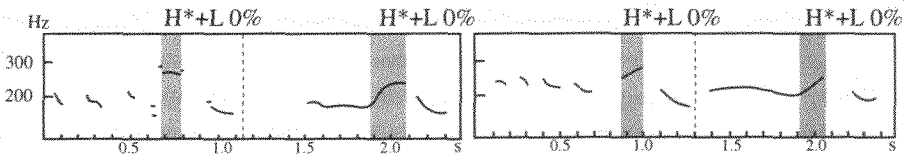
3.1.3. Auditory and acoustic analyses

The recordings were digitised at a sampling rate of 16 kHz and processed in the commercial software package waves(tm) on an HP workstation A4032A. The intonational structures of the utterances were analysed and transcribed by a combination of auditory analysis and inspection of the F0 trace. The analysis showed that the sentences in the 'statement' lists appeared to have been produced consistently as H*+L %, and the ones in the 'question' lists as L*+H H%. However, at this juncture, it should be pointed out that the evidence for the high boundary tone H% in the rises is not immediately obvious. On final bisyllabic words, rising accents are realised in both languages as a low on the stressed syllable and a rise on the following syllable. As the stressed syllable is only followed by one further syllable in the intonation phrase, an extra 'kick-up' in pitch at the IP boundary, which may be taken to reflect the presence of a boundary tone, cannot obviously be observed. However, it is intuitively clear that if one were to replace the *our neighbour / unser Nachbar* by a somewhat longer phrase such as *our neighbour over there / unser Nachbar da drüben*, a boundary rise can be observed. Therefore, it will be assumed that the rises are appropriately accounted for as L*+H H%; nothing in our treatment hinges crucially on whether the rises are seen as containing H%.

Figure 3 shows F0 traces from realisations of the test sequences by English speaker RF and German speaker BL (the longest test word is shown, and the patterns in the figure held across speakers). Also, the transcriptions are given. As can be seen, the cross-linguistic realisation of falling accents in English and German is quite similar, but a clear cross-linguistic difference can be seen for the rises. In English, the accented syllable tends to fall or be level whereas in German, it usually rises. Peak alignment, however, did not obviously affect the measure chosen to reflect truncation or compression and will therefore not be discussed further.

Falling accents

English **German**



*It's Mr **Sheafer** | our new **Neighbour** |* *[..] Herr **Schiefer** | unser neuer **Nachbar** |*

Rising accents

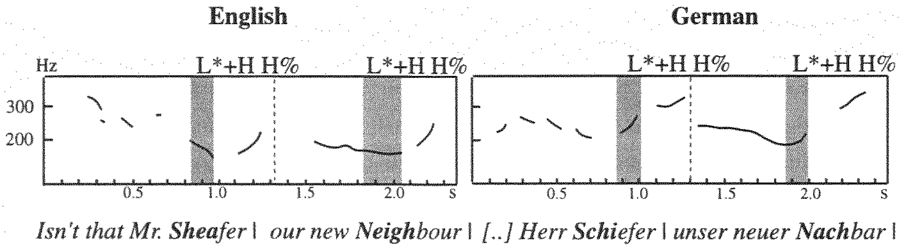


Figure 3 *F0 traces of carrier phrases surrounding the longest test word for English (left) and German (right). The patterns of the carrier phrases shown held across speakers and test items.*

As an acoustic correlate of truncation and compression, 'rate of fundamental frequency change' was chosen. The measure was calculated by dividing the maximum fundamental frequency excursion on a test word by the duration of fundamental frequency movement (details of how measurements were taken on monosyllabic and bisyllabic test words will be given below). F0 excursion and F0 duration were measured from the fundamental frequency trace in waves(tm). The measure was motivated as follows. The fundamental frequency contour represents the main acoustic correlate of the pitch contour of an utterance. This pitch contour is in a sense continuous, interrupted only by the 'accident' of voiceless segments and these interruptions are reflected in the fundamental frequency trace. It is not possible reliably to estimate fundamental frequency for these voiceless episodes, nor can we infer how they are perceived. If, for instance, fundamental frequency is level on the short vowel of *Schiff*, then we must presume that this is the length of pitch movement on which the listener must work, regardless of the fact that this level may be equivalent to a fall when there is voicing of greater duration. Therefore, the duration of fundamental frequency on the test words was defined as the duration of the observable acoustic equivalent of the pitch movement listeners hear on the test words.

Secondly, the fundamental frequency excursions on the test words were measured, again from the fundamental frequency trace in waves(tm). The highest and lowest points of the excursion were measured, always in one direction (i.e. from left to right). On monosyllabic words, this involved the peak at the beginning of the sonorant section of the syllable and the lowest point in the subsequent fall. This measured excursion was then divided by its duration. On bisyllabic words, however, measuring was not as straightforward. Firstly, the contour was interrupted by the voiceless fricative at the beginning of the second syllable, and secondly, as can be observed in Figure 7, the acoustic realisations of rises and falls did not involve constantly rising or falling F0; therefore, the appropriate measurement points were not immediately obvious.

F0 excursion on bisyllabic words was measured as follows. Firstly, on the assumption that the pitch contour is in a sense perceived as continuous and interrupted only in the acoustics by voiceless segments, the duration measurements on bisyllabic words included the section where the F0 contour is interrupted (e.g. in *Schiefer*, the duration of the labiodental fricative was included). Secondly, when measuring F0 excursion, the excursion over the whole word was measured rather than separate excursions on each syllable.

Clearly, this is not the only way in which measurements for 'rate of F0 change' could have been taken. F0 excursion might have been measured exclusively on the sections where F0 falls, and then divided by the duration of that fall, rather than by the duration of the complete F0 contour on the word. However, in that case, the duration measurements would have been restricted to the second syllable in German, and the second and part of the first syllable in English, because English and German differ in peak alignment (see Chapter 4 of the present study). In English, the peak is located within the stressed syllable, but in German, it is located at the right edge. That means that in English, the fall starts earlier and duration measurements would have had to involve part of the first syllable. Consistent measurements, however, would have been difficult to take; at times, no obvious peak was observed in the first syllable in the English data, and then the measurements would have involved the second syllable only, and at other times, F0 was level on the stressed syllable, and the exact location of the peak could not be determined.

However, the approach followed instead (i.e. the one described above) is not without disadvantages either. If, in two-syllable words, one measures the complete duration of F0, and not only the sections which are strictly speaking 'falling', then the measured duration is necessarily somewhat longer than it would have been had only the duration of the falling stretches been determined, and that means that the resulting slope is somewhat flatter than it really is.

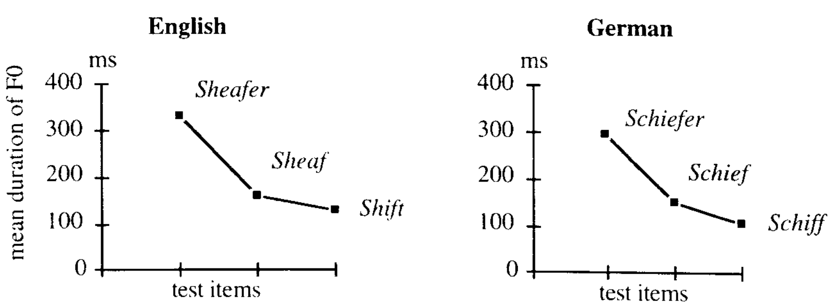
3.1.4. Predictions

For English, a significant increase in the rate of fundamental frequency change was predicted for falls and rises when sonorant material was shortened. This was taken to reflect compression. A difference between English and German was predicted for falls, reflected in the absence of a significant increase in the rate of fundamental frequency change in German. Rising accents in German, however, were predicted to compress, just as in English.

3.2. Results

The measurements were subjected to statistical analysis. Analyses of variance carried out separately for English and German established that the duration manipulation had the desired effect (univariate ANOVA, repeated measures; the pitch accent difference was treated as a repetition). The duration of fundamental frequency which had been measured for each test word decreased significantly from the longest to the shortest test word ($F_{[2,22]} = 149.27$, $p< 0.001$ for English and $F_{[2,22]} = 301.8$, $p< 0.001$ for German; significance levels unchanged by Greenhouse-Geisser correction) Planned comparisons showed that in English, the difference between the longest and the mid-length word and the longest and the shortest word were significant at 0.1 % level, and the difference between the mid-length and the shortest word was signficant at the 5% level. In German, differences between all three length conditions were significant at the 0.1% level (again, all significance levels unaffected by Greenhouse-Geisser correction). Figure 4 provides the mean duration values for the three test words. Figure 5 shows the mean values for fundamental frequency excursion. What is immediately obvious from Figure 5 is that, as predicted, the German falls pattern strikingly differently from the German rises and English rises and falls. Here, the range decreases dramatically from the longest to the shortest word, whereas no similar decrease can be observed for the other three patterns. Figure 6 shows the mean rate of fundamental frequency change across the test words. Again, we see that the German falls stand out - the rate of fundamental frequency change decreases, whereas in German rises and in English rises and falls, it increases.

Falling accents (H*+L Ø%)



Rising accents (L*+H H%)

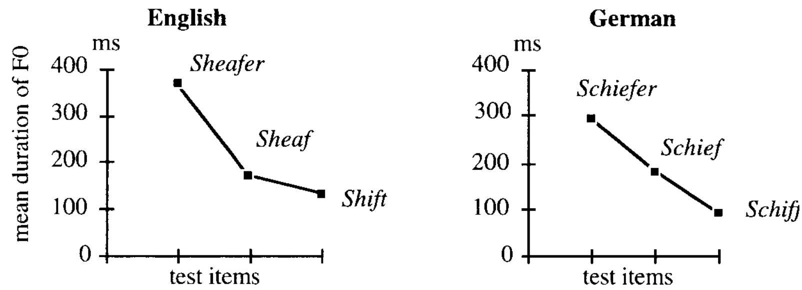
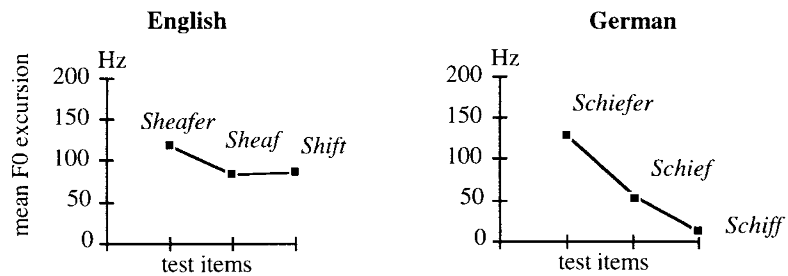


Figure 4 *Mean duration of F0 on test items.*

Falling accents (H*+L Ø%)



Rising accents (L*+H H%)

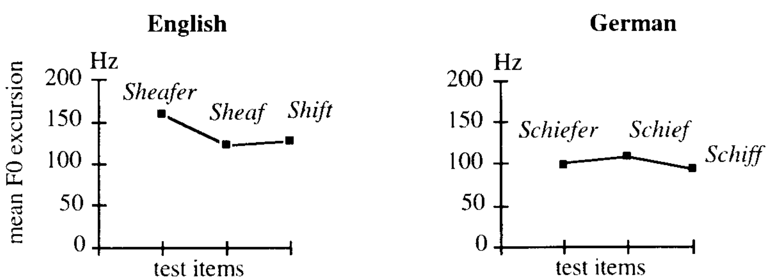


Figure 5 *Mean F0 excursion on test items.*

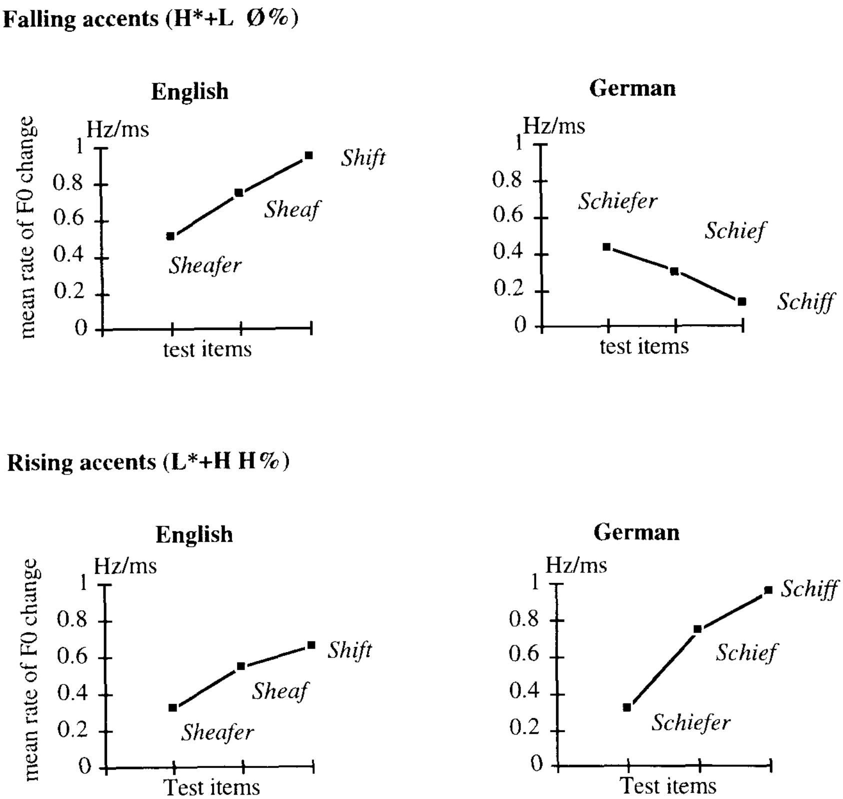


Figure 6 *Mean rate of F0 change on test items.*

An analysis of variance was carried out for the parameter 'rate of fundamental frequency change' with the factors Language, Word length and Pitch accent type (univariate ANOVA, repeated measures). Factors Language and Pitch accent were not significant, but Word length was ($F_{[2,22]} = 44.19$, $p<0.001$, $p<0.001$ after Greenhouse-Geisser correction). Moreover, the interaction between Language and Pitch accent was significant ($F_{[1,11]} = 19.16$, $p<0.01$, $p<0.01$ after Greenhouse-Geisser correction) as well as the interaction between Pitch accent and Word Length ($F_{[2,22]} = 11.35$, $p<0.01$, again unaffected by Greenhouse-Geisser correction). Then, the data sets for the two languages were processed separately.

An analysis of variance for English with the factors Word length and Pitch accent type showed no significant effect of Pitch accent, that is, falls and rises behaved in the same way, but a significant effect of Length ($F_{[2,22]} = 29.55$, $p< 0.001$ before and after Greenhouse-Geisser correction). No significant interaction between Pitch accent and LENGTH emerged. Thus, in English, the rate of F0 change increased significantly with decreasing segmental duration, confirming the prediction that English would compress pitch accents when segmental material was shortened. Planned comparisons showed significant differences between the longest and the mid-length word ($p < 0.001$), the longest and the shortest word ($p < 0.001$) and the mid-length and the shortest word ($p < 0.01$, all significance levels unaffected by Greenhouse-Geisser correction).

For German, again as predicted, the analysis revealed a significant main effect for Word length ($F_{[2,22]} = 6.31$, $p < 0.01$) and Pitch accent type ($F_{[1,11]} = 20.67$, $p < 0.001$). Planned comparisons showed that within rises, the difference between the longest and the mid-length word was significant ($p < 0.001$, Greenhouse-Geisser corrected to $p < 0.01$), the longest and the shortest word ($p < 0.001$, unaffected by Greenhouse-Geisser correction), and differences at the 5% level between the mid-length and the shortest word before Greenhouse-Geisser correction ($p < 0.07$ after correction). Thus, the rate of fundamental frequency change increased significantly from the longest to the shortest word, reflecting compression. Within falls, the difference between the longest and mid-length word was not significant, but the difference between the longest and the shortest word was ($p < 0.01$, unaffected by Greenhouse-Geisser correction). The difference between the mid-length and shortest word was significant at the 1% level before Greenhouse-Geisser correction and significant at the 5% level after correction ($p < 0.03$). Thus, in German, the rate of fundamental frequency change decreased significantly from the longest to the shortest word, reflecting truncation (significant differences at the 0.05 level between the three words).

A summary of the results is given in Figure 7. In the graphs, fundamental frequency excursion on the vertical axis is plotted against its duration on the horizontal axis. Figure 7 shows that in English, when words get shorter, the rate of fundamental frequency change increases, thereby preserving the rise or fall (NB: it was possible to normalise the fundamental frequency starting points in the range, because the differences were not significant). In German, on the other hand, accent realisation is asymmetrical. For rises, the F0 rate of change increases as the words get shorter, but for falls, it decreases. Figure 8 below supplies representative F0 traces of compressed rises and truncated falls from one speaker in each language. For English, the patterns shown held for all speakers, and for German, they held for all speakers in rises and in the longest and the mid-length condition for falls. The falls on the shortest word, however, exhibited a range of F0 patterns, albeit within a rather restricted F0 range. Three of the twelve German subjects produced a slight fall in F0 involving a drop of 4 Hz, 6Hz and 16 Hz, another three produced a rise-fall with excursions of 6, 7 and 12 Hz. The remaining six speakers produced virtually level F0. Thus, falling F0 is observed on *Schiefer* and *Schief*, but rarely, it appears, on *Schiff*.

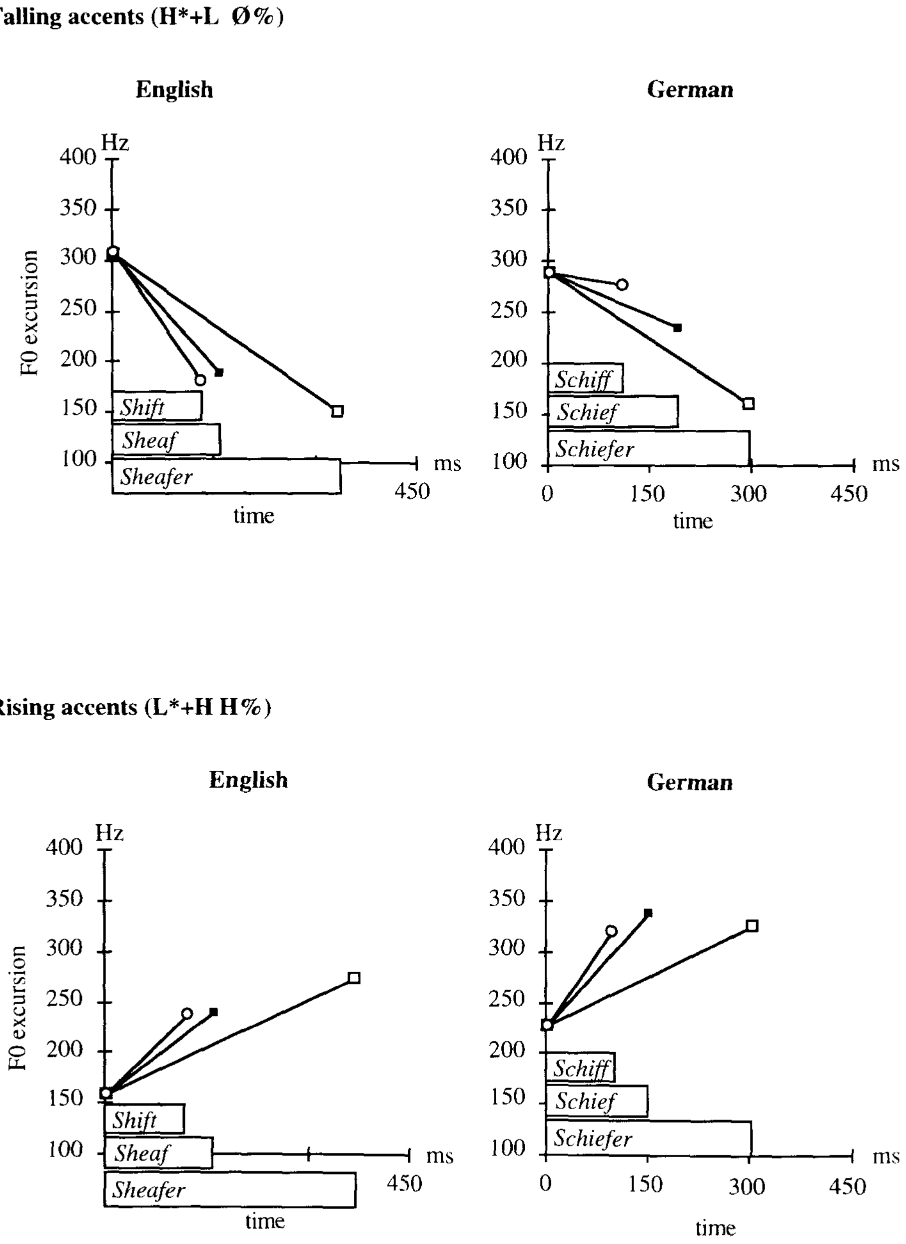


Figure 7 *Summary of results. F0 excursion on the vertical axis is plotted against time on the horizontal axis. Cross-linguistically, similar patterns may be observed for rises (below), but not for falls (above).*

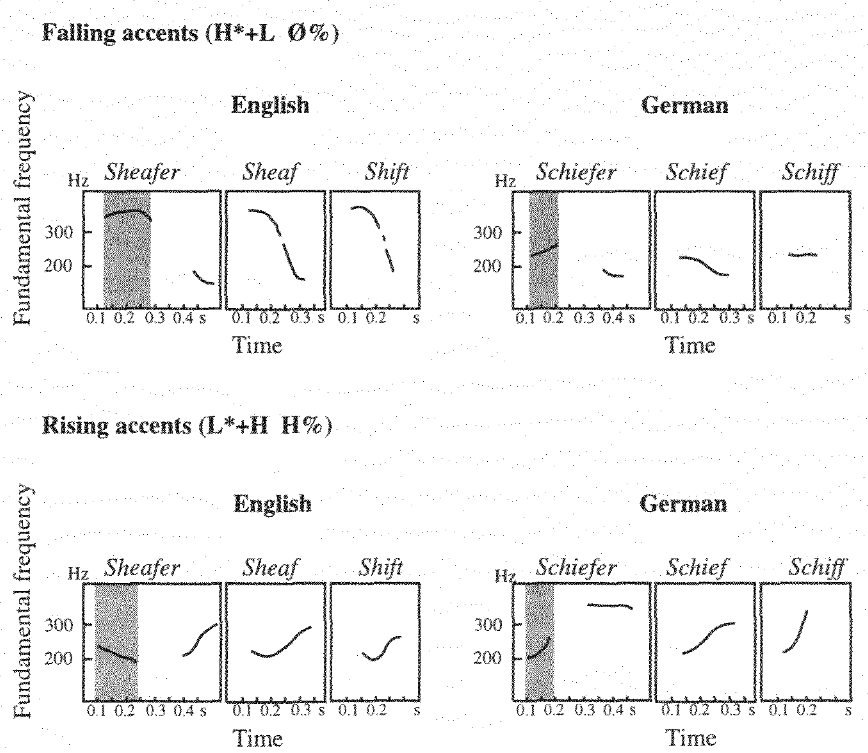


Figure 8 *Representative F0 traces of test items (English speaker JC; German speaker HW).*

To sum up, the results of the experimental investigation suggest a cross-linguistic difference in pitch accent realisation. When little sonorant material is available, English compresses, but German may truncate. Secondly, in German, accent realisation on segmental material with little scope for voicing is asymmetrical. Falls appear to be truncated, but rises are compressed, just as in English.

3.3 Discussion

A number of questions arise. Firstly, the results revealed a cross-linguistic as well as a monolinguistic difference in accent realisation. English falls compress but German falls truncate; German rises, on the other hand, compress. How might one account for these differences within the AM approach outlined earlier? Two options will be discussed. The first proposes that truncation results from a non-gradient change in the underlying phonological representation, and the second suggests that truncation may be a gradient acoustic accommodation effect. Both assume that compression is a matter of gradient acoustic implementation only.

Secondly, how generally may the results presented here be interpreted? Can we infer that truncation and compression will apply in the two languages whenever sonorant material is reduced?

3.3.1 Accounting for truncation and compression

The following sections will devote more space to the issue of truncation than to compression, as truncation appears to be the more interesting effect. Compression would seem to be quite straightforwardly interpretable as a matter of acoustic realisation even if further details such as, for instance, effects of peak alignment remain to be explored in detail. Crudely, on monosyllabic words with varying amounts of sonorant material, one might posit that targets are realised at the left and right edge of the sonorant material contained in the syllable nucleus (here defined as in Beckman, 1986, as the sonorant portion of the rhyme). In bisyllabic words, the targets appears to be realised within the syllable nuclei of the first and second syllables respectively.

Tentatively, two accounts of truncation will be proposed. The first assesses the evidence for the truncation effect on the shortest condition (*Schiff*) involving a phonological process of L-deletion. The second suggests truncation to be a gradient acoustic realisation effect.

3.3.2 Truncation: phonological or gradient acoustic?

Figures 8 shows F0 traces for German falls produced by speaker HW. From *Schiefer* to *Schief*, we observe an apparently gradient change; on *Schief* which has less sonorant material, the trace appears to be steeper than on *Schiefer*. The succession from *Schief* to *Schiff*, on the other hand, seems to be of a different nature; here, a non-gradient change appears to have taken place. *Schief* falls but *Schiff* exhibits evidence of a fall only for three of the twelve speakers. Although *Schiff* is clearly shorter than *Schief* (80 vs. 170 ms) one might nevertheless expect to observe at least the beginnings of a fall (arguably *Schief* begins to fall earlier than at 80 ms). This observation might be interpreted to suggest that the F0 traces on *Schiefer* and on *Schief* reflect the presence of a low target, but the one on *Schiff* does not. Here, the low target has been deleted, possibly because H*+L was associated with a single sonorant mora.

However, three observations appear to speak against this proposal. The first involves the apparently gradient difference between *Schiefer* and *Schief*. Here, we do not appear to find evidence of L-deletion; the F0 pattern on both words can be interpreted to reflect the presence of a low target. Nevertheless, the mean rate of F0 change on *Schief* is significantly slower than that on *Schiefer*, suggesting (a) there is no compression effect, and (b) that the acoustic realisation of the two words differs. Moreover, the mean F0 excursion on the shorter word was significantly smaller. These observations appear to point toward a gradient truncation effect.

The second observation involves the shape of German falls in F0 and further supports a view of truncation as a gradient acoustic effect. As can be observed in Figure 8, falls are realised in F0 not as a constant fall throughout the stressed syllable but rather as 'rise-falls' or 'level-falls'. On *Schiefer*, a falling accent is implemented as rising or level F0 on the stressed syllable followed by a fall onto the following syllable. On *Schief*, the fall starts off with a level, or sometimes a small rise in the early part of the vowel. On *Schiff*, we find a falling-rising or level F0 for nine of the twelve speakers. Shallow falls are rare. Thus, it appears reasonable to suggest that the level F0 we observe for speaker HW's realisation of *Schiff* in Figure 8 is in fact the equivalent of the level or falling-rising section observed in falls on longer words. Figure 9 illustrates this point. On the left, F0 traces from speaker HW illustrate the shape of F0 in German falls. On the right, a language-specific schematic representation of truncation is suggested for German.

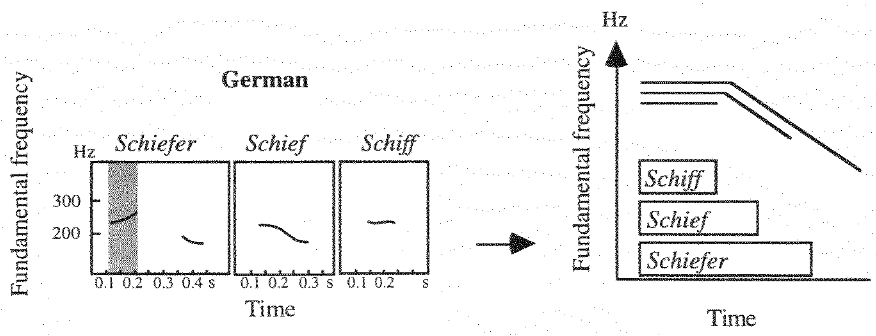


Figure 9 *A language specific schematic representation of truncation in German. On the left, F0 traces for German falls are repeated from Figure 8. The schematic representation is given on the right.*

Incidentally, Figure 9 also illustrates why the realisation of German falls on syllables with a small proportion of sonorants may be referred to as 'truncation'. For the three falling test words in German, the statistical analysis revealed a significant decrease in the rate of F0 change, and one might wonder why this decrease is referred to as 'truncation'. At first sight, this result appears to contradict the informal definition given in the introduction which stated that when truncation applies, F0 simply 'ends earlier'. That definition would appear to predict that when truncation applies, changes in F0 duration and range will be observed, but no changes should emerge in the rate of F0 change.

However, that definition was necessarily rather abstract, and could not include language-specific details concerning the shape of F0 in the realisation of particular accents. As mentioned earlier, German falls are realised as rise-falls or level-falls, that is, the fall is neither constant, nor does it necessarily extend over the complete duration of the syllable nucleus. The shape of F0 affects the measurement used to investigate truncation and compression in this study, but does not necessarily invalidate the definition of truncation as a case where F0 ends earlier when less sonorant material is available. Figure 9 shows that despite the apparent decrease in the rate of F0 change in the statistics, one may argue that F0 is truncated, that is, on words with less sonorant material, the trace simply ends earlier.

The third and final point which appears to speak against truncation being L-deletion comes from an observation about the auditory realisations of *Schiff*. Native speakers appear to hear this word as having 'falling pitch', no matter whether F0 is actually level, falling-rising or slightly falling[2]. Similarly, in the utterance illustrated in Figure 6, the truncated falling accent on *Bett* appeared to be an accent of the same type as the one in the preceding phrase on *Hause* and the one in the following phrase on *gesund*.

Taken together, the observations discussed in this section appear to weaken considerably an account of truncation as a phonological process of L-deletion. In the absence of convincing evidence suggesting that truncation is phonological rather than gradient acoustic, it seems reasonable to assume, at least tentatively, that the cross-linguistic results for falling accents can be interpreted as presenting a case of one and the same phonological representation being realised differently in two different languages. The following section will discuss an account of truncation as a gradient acoustic effect within the AM system suggested in this study.

3.3.3 Asymmetrical results for German

Assuming that truncation and compression reflect gradient acoustic implementation effects, this section considers the question of whether tonal elements or tonal 'constituents' within the nuclear tone may be called upon to account for the asymmetrical result in German. An approach following the British school cannot consider this question; nuclear falling and rising tones are seen as single phonological elements. In the AM approach followed here, however, falls are represented as consisting of two tonal elements (H*+L), and an unspecified boundary whereas rises consist of three (L*+H H%). As the experimental data suggested truncation for falls with an unspecified tonal

---

2 *Note, however, that the fall in pitch heard on Schiff is not as obviously a fall as is heard on the English word Shift. Comparing realisations of H*+L on Schiff and Shift, the impression is that Shift exhibits much more obviously a fall in pitch. Nevertheless, in German, the accent on Schiff appears to be heard as the same in kind as the accents on the longer test words Schief and Schiefer.*

boundary and compression for rises where the boundary was claimed to be specified, the boundary asymmetry may be responsible for the asymmetrical results. A possible account might posit that in German, when there is little scope for voicing, nuclear accents followed by low boundaries may be truncated, but those followed by high boundaries are compressed. In effect, this account suggests that truncation and compression effects in German are sensitive to the tonal specification of phrase boundaries.

This account generates a new prediction; it suggests that German rises which are followed by a tonally unspecified IP boundary should be truncated also. Unfortunately, no evidence either way is available at present, but native speaker intuition appears to suggest that in open lists which may be represented as successions of L*+H Ø%, rises are compressed also. Thus, native speaker intuition does not support an account of compression and truncation effects sensitive to tonal boundary specifications. Rather, the opposite appears to be suggested, namely, that the high boundary tone is not relevant at all. Considering that none of the test items were long enough to exhibit an effect of the high boundary tone on F0 in the form of an additional rise at the boundary, this account appears to reflect the data somewhat better than the one proposing the high boundary tone to be responsible for the asymmetry.

Thus, tentatively, it will be suggested that in German, truncation and compression involve the nuclear tone; all pitch movements from the accented syllable up the phrase boundary are affected. The application of either, however, appears to be conditioned by the nuclear pitch accent alone; following specified or unspecified boundary tones have no effect. Very tentatively, then, the result may be interpreted as reflecting a more general asymmetry between high and low tones, rather than the specific boundary asymmetry proposed in this study. Erickson *et al.* (1995) point out that, generally, the modelling of high targets appears to be considerably easier than the modelling of low targets. For instance, in downstepping sequences in English, the relationships between successive high targets appears to be proportionally constant across different pitch ranges (Liberman and Pierrehumbert, 1984), but there does not appear to be any corresponding relationship holding between successive low accents. Unstarred high trailing tones, on the other hand, do appear to step down. Thus, one may place the results presented in this study into the context of a more generally observed asymmetry between high and low tones, and accent realisation in German may be one instance of this.

3.4 Scope of results

At first sight, one might interpret the results of the experimental study presented here as suggesting that whenever sonorant material is reduced, truncation and compression will apply to some degree. However, this conclusion requires some qualification. As detailed above, in the materials, distinctions in the amount of sonorant material available were made in two ways; firstly, by reducing the number of syllables from two to one (*Sheafer* vs. *Sheaf*) and secondly by changing phonological vowel length (*Sheaf* vs. *Shift*). For the latter contrast the results may be interpreted relatively straightforwardly; truncation and compression appear to apply when less sonorant material is available. An increase in the rate of F0 change may be taken to reflect compression and the absence of such an increase suggests truncation. On comparing the mid-length and the shortest condition, then, it seems reasonable to conclude that pitch accents are compressed when less sonorant material is available.

Results from the first length distinction, i.e. the one between bisyllabic words and monosyllables, require some more detailed discussion. For the purpose of this experiment, *Sheaf / Schief* and *Shift / Schiff* were assumed to differ in one aspect only: phonological vowel length. The bisyllabic test words, however, differed from the monosyllabic 'long' condition (*Sheaf / Schief*) in two ways; firstly, in the proportion of sonorant material and secondly in number of syllables. Nevertheless, gradient differences were observed between all three test items; that is, a comparison between the bisyllabic condition and the monosyllabic conditions does not contradict the claim that less sonorant material will result in truncation and compression, regardless of whether the words involved are bisyllabic or monosyllabic. Stating the claim more strongly, however, would require further investigation. The materials tested here were designed such that the bisyllabic words would contain more sonorant material than the monosyllabic words. One may, however, think of bisyllabic words where this is not the case. For instance, the word *fire* may be argued to have two syllables but contain the same amount of sonorant segments as the monosyllabic word *file*. The experiment presented here does not show whether realisational differences apply in such cases or what shape they may take. Secondly, in a bisyllabic word, varying amounts of non-sonorant material may put more or less distance between sonorant syllable nuclei (e.g. *fire* vs. *fighter*), arguably separating potential realisation sites of tones, and differences of this kind may affect accent realisation. Thirdly, a similar number of non-sonorant segments may intervene between syllable nuclei, but the syllable nuclei themselves may differ in the amount of sonorant material they contain (e.g. *fighter* vs. *finder*). In the latter case, less sonorant material is available but considering the equivalent starting points of syllable nuclei, realisational differences are unlikely to apply.

Thus, clearly, there are several ways in which the amount of voiced material available for the realisation of pitch accents may be reduced, and this study addresses only one particular case. However, in a first study investigating realisational differences between German and English pitch accents, it appeared relevant to elicit materials as naturally as possible, and this restricted the number of 'sonorant conditions' which could be included. A more detailed follow-up study investigating other ways in which the amount of voiced material may be reduced would be of interest.

4 Experiment II

4.1 Method

In the discussion of Experiment I, it was proposed that truncation and compression may apply to the nuclear pitch accent rather than the nuclear tone. Specifically, native speaker intuition appeared to suggest that nuclear rises compress, regardless of whether the boundary is 0% or H%. However, no experimental evidence was available which might have supported this claim. The issue was addressed in a further production study, investigating German only. The follow-up study was restricted to German (a) because no cross-linguistic difference for compression had emerged from pitch accent accommodation Experiment I, and (b) because in German, nuclear rising tones without a boundary specification (L*+H 0%) appeared to be somewhat more frequent than in English. This meant that L*+H 0% was more likely to be elicited successfully from German naive speakers.

Accent accommodation was again investigated in a reading task. Materials were required which would elicit realisations of the nuclear tones L*+H 0% and L*+H H% on words with successively less scope for voicing. It was not immediately obvious, however, how such materials should be designed. An investigation of pitch accent accommodation effects requires speech data involving accents realised on very short syllables, but when an accented syllable is very short, how can one know what the boundary tone is? Thus, some diagnostic was required which would provide independent evidence of whether the nuclear tone on a very short word was L*+H 0% or L*+H H%. Such a diagnostic appeared to be offered by the specific accent patterns observed in the realisation of syntactic coordination structures. As mentioned in section 2.3 in Chapter 2, coordination structures are commonly produced with the same nuclear tone (Trim, 1959, Schubiger, 1958, Crystal, 1969, Halliday, 1967). This observation suggests that a longer word coordinated with an 'accommodation candidate' might provide independent evidence of the accentual structure of accommodation candidate. For instance, if items in a list are produced with successions of L*+H H%, and one of the words in the list is very short, the hypothesis is that this word is produced with the same nuclear tone as preceding and following words. Thus, in the second pitch accent accommodation experiment, accommodation candidates were embedded in lists designed to be produced with successions of either L*+H 0% or L*+H H%.

The lists were embedded into a short paragraph. The structure of this paragraph was modelled on sections of text from the cross-linguistic corpus presented in Chapters 3 and 4 which had consistently elicited either L*+H H% and L*+H 0%. A 'statement list', that is, a closed, declarative list of items was designed to elicit sequences of L*+H 0%, and a 'question list', consisting of two successive questions was predicted to elicit sequences of L*+H H%. Each list contained one or more polysyllabic words (the 'controls'), which were predicted to show clearly whether the accent pattern in question was L*+H 0% or L*+H H%, and words with successively less scope for voicing (the 'accommodation candidates') which were assumed to be characterised by the same pattern as the surrounding controls.

Three paragraphs were read by each subject. Each contained one statement and one question list, and the three paragraphs differed from each other in that the accommodation candidate became successively longer. In the statement lists, the accommodation candidates were *Fisch* (/fɪʃ/ 'fish'), *Fleisch* (/flaɪʃ/, 'meat') and **Fleisch**wurst (/flaɪʃvʊʁst/ 'sausage'), and controls were the words ***Brom**beermarmelade* (/bɹɔmbeamaːməlaːdə/) which means 'blackberry jam,' and ***Maul**beermarmelade* (/maulbeamaːməlaːdə/) which means 'mulberry jam'. Both controls were stressed on the first syllable. The questions lists used the same accommodation candidates and one of the controls. One version of the three paragraphs (the one with the shortest accommodation candidate) is given in (1) below. The paragraph and the test sequences (printed in bold) were kept as short and as natural as possible as this appeared to raise the chance of eliciting consistent productions.

(1) Eines Morgens sagt Rotkäppchens Mutter: "Rotkäppchen, deine Oma ist nicht gesund, und sie kommt nicht zum Einkaufen. Ich packe ein paar Lebensmittel für sie ein und ich möchte daß Du sie ihr bringst". Sie füllte einen Korb mit **Brombeermarmelade, Fisch, Maulbeermarmelade** und Birnenmarmelade. Rotkäppchen fand die Mischung merkwürdig und fragte:"Will sie wirklich **Brombeermarmelade? Und Fisch?"**

The predicted intonation structures were the following.

Statement list:	L*+H	0%	L*+H 0%	L*+H	0%
	***Brom**beermarmelade*		***Fisch***	***Maul**beermarmelade*	
			Fleisch		
			Fleischwurst		
Question list:	L*+H	H%	L*+H H%		
	***Brom**beermarmelade*		***Fisch***		
			Fleisch		
			Fleischwurst		

The materials addressed the question of whether compression involves the nuclear accent or the nuclear tone as follows. Assuming that coordination structures are produced with the same nuclear tone, and that this is L*+H 0% in the statement list and L*+H H% in the question list, the accommodation candidates *Fisch, Fleisch* and *Fleischwurst* would be associated with L*+H 0% in the statement list and with L*+H H% in the question list. If compression applies only when a boundary tone H% is present, i.e. if it involves the nuclear tone rather than the nuclear accent, then we should find compression in the question list, but not the statement list. On the other hand, if it is the nuclear accent which is compressed, regardless of the boundary specification, the we should find compression in both lists.

Fourteen female German subjects drawn from the same pool as described in Chapter 4 were asked to read the materials. The paragraph with the shortest word was read first, followed by the one with the mid-length word and the longest word. Again, subjects were told that the paragraphs were being recorded for foreign learners of German and that they should be read 'in the same way'.

4.1.1 Analysis

Auditory and acoustic analyses were carried out. The auditory analysis involved three steps. The first step addressed the question of whether the coordination structures had indeed been produced with the same nuclear tone. To this end, the auditory and acoustic realisations of the pre-final polysyllabic controls, *Brombeermarmelade,* and *Maulbeermarmelade* in the statement list were compared (the question list contained only two items, one of which was an accommodation candidate). The list-final item *Birnenmarmelade* was excluded from the comparison because the final accent in a list tends to be a fall; and the accommodation candidates were excluded also, because they were maximally bisyllabic and assumed to be too short to reflect reliably the distinction between L*+H 0% and L*+H H%. The prediction was that the two polysyllabic words *Brombeermarmelade* and *Maulbeermarmelade* should exhibit the same nuclear tone.

The second step of the auditory analysis compared productions in the statement list with those in the question list. The question was whether the lists had been consistently produced with L*+H 0% or L*+H H% respectively.

The third and final step addressed the question of compression. Were the accommodation candidates in the two lists compressed or not? Auditorily, compression was assumed to be signalled by a rise in pitch on all accommodation candidates.

Additionally, acoustic measurements were taken firstly to provide an acoustic correlate of the auditory distinction between phrases with and without H% and secondly to show whether the accommodation candidates would exhibit significant increases in the rate of F0 change when sonorant material was reduced.

Acoustic evidence of the presence or absence of H% was gathered by measuring on the polysyllabic controls (*Brombeermarmelade* and *Maulbeermarmelade*) (a) the F0 maximum in the postaccentual syllable and (b) F0 at the IP boundary.

On the accommodation candidates in both statement and question lists, F0 excursion and duration of F0 were measured, and the rate of F0 change was calculated. This was the acoustic measure which had been hypothesised to differentiate truncation and compression in the first experiment presented in this chapter. Again, significant increases in the rate of F0 change from the longest to the shortest accommodation candidate were assumed to signal compression. Thus, should the rate of F0 change increase in the statement as well as the question lists, then this would suggest that compression involves the nuclear accent, but if it increases only in the list with H%, then compression would appear to involve the nuclear tone. F0 excursion was measured from left-to-right and, as in the previous study, duration of F0 on the accommodation candidates was measured for the whole word.

4.2 Results

4.2.1 Auditory analysis

The auditory analysis of the data showed that all test items in statement and question lists had been produced with rising nuclear pitch accents, transcribed as L*+H. The realisation of German L*+H in F0 had been defined in Chapter 3 for trochees as a step-up in pitch from the accented syllable onto the postaccentual syllable and for monosyllables as a rise throughout the accented syllable. These patterns were observed in the data. Moreover, nuclear rise-plateaux transcribed as L*+H 0% were distinguished from those transcribed as H*>0% by the presence or absence of DOWNSTEP. In the data recorded for this experiment, DOWNSTEP did not apply to nuclear rise-plateaux.

Two slightly different auditory impressions appeared to reflect the presence of a boundary tone; a boundary was labelled with H% either if the last syllable of the test word exhibited a sharp rise in F0 or if F0 rose gradually from the H of L*+H towards the intonation phrase boundary. Sometimes an additional rise was heard on the last syllable. It appeared possible to make categorical decisions as to whether a boundary tone was present or not.

Next, the question was addressed of whether the controls *Brombeermarmelade* and *Maulbeermarmelade* in the statement lists had been produced with the same nuclear tones. The results are given in Table 1 below. 'Short list' stands for the list with the shortest accommodation candidate (i.e. *Fisch*), 'mid-length' list for the one including *Fleisch*, and the 'long list' contained the word *Fleischwurst*. Table 1 shows that of the 42 statement lists recorded, 26 were realised with H% delimiting both controls (*Brombeermarmelade* and *Maulbeermarmelade*), 15 with 0% delimiting both controls, and one pair being mixed (in that case, however, an relatively larger rhythmic discontinuity was observed between the control and the following accommodation candidate than in all other cases). Thus, the results of the auditory analysis clearly support the hypothesis that coordinated intonation phrases tend to exhibit the same nuclear tones: not only were all controls realised with a pitch accent L*+H but also the choice of boundary specification matched within individual realisations.

What is also apparent from Table 1 is that the second hypothesis, according to which statement lists would be produced with sequences of L*+H 0% and question lists with L*+H H%, is not borne out by the data. Subjects produced both L*+H 0% and L*+H H% in the statement lists.

Table 2 below shows a comparison between the accentual structures of the first control in the statement and question lists respectively (*Brombeermarmelade*). Realisations with 0% were predicted in the statement list and realisations with H% in the question list. Table 2 shows that, in fact, the statement list contains more realisations of H% than the question list. Clearly, it is not the case that statement lists are more likely to be produced with 0% than question lists[3].

	'Short list'		**'Mid-length list'**		**'Long list'**	
	Control 1	*Control 2*	*Control 1*	*Control 2*	*Control 1*	*Control 2*
s1	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**
s2	0%	0%	0%	0%	0%	0%
s3	0%	0%	0%	0%	0%	0%
s4	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**
s5	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**
s6	**H%**	0%	0%	0%	0%	0%
s7	0%	0%	0%	0%	0%	0%
s8	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**
s9	**H%**	**H%**	0%	0%	0%	0%
s10	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**
s11	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**
s12	0%	0%	0%	0%	**H%**	**H%**
s13	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**
s14	**H%**	**H%**	**H%**	**H%**	**H%**	**H%**

Table 1 *Distribution of boundary specifications in the statement list. All pitch accents were L*+H.*

3 *In the question list, only two intonation phrases were coordinated, the second of which contained the accommodation candidate (maximally bisyllabic). Thus, evidence suggesting whether the accommodation candidate ended in 0% or H% is not available.*

	Statement list		**Question list**	
	0%	H%	0%	H%
'short'	8	6	11	3
'mid'	6	8	7	7
'long'	6	8	11	3
	20	22	29	13

Table 2 *Boundary distribution on* <u>*Brombeermarmelade*</u> *in the three versions of the statement and question lists.*

The third stage of the auditory analysis involved the pitch patterns on the accommodation candidates. The first experiment presented in this chapter showed that rising accents compress. Thus, in the analysis of the present experiment, similar rises in pitch on the accommodation candidates on *Fisch, Fleisch* and *Fleischwurst* were assumed to reflect compression. Truncation of rising accents, on the other hand, had not been observed in experiment I. Since no single example of a truncated rise was available, the likely shape of such an accent can only be inferred by extrapolation of the pattern of truncated falls. Truncated falls were shown to be auditorily equivalent to non-truncated falls despite being virtually level in F0. However, when we compare the auditory impression of a truncated fall in German with that of a compressed fall in English, then the falls do sound quite different; the compressed fall in English involves a greater extent of pitch movement. Thus, we may conjecture that the difference between a compressed and a truncated rise in German is similar to that between a truncated fall in German and a compressed fall in English; when we compare them directly, one involves greater pitch movement than the other. Consequently, truncation of L*+H on the accommodation candidates *Fisch, Fleisch* and *Fleischwurst* was to be discernible auditorily by a smaller degree of upward pitch movement on the shortest word *Fisch* than on the longest word *Fleischwurst*.

The auditory analysis showed that (a) pitch rose on all accommodation candidates, and that (b) the perceived degree of pitch movement on all accommodation candidates was comparable. The shortest accommodation candidate *Fisch* did not appear to be associated with less pitch movement than the longest *Fleischwurst.*

In summary, the results of the auditory analysis suggest the following. Firstly, there appears to be a strong correlation between statement and question lists of the type investigated here and the choice of a nuclear accent L*+H. Secondly, the results confirm the hypothesis in the literature that lists tend to be produced with the same nuclear tone. Thirdly, it appears that neither type of list conditions the choice of a boundary tone H%

more strongly than the other (a Chi-square test showed no significant differences in the boundary distributions observed in statement and question lists (df = 1, χ^2=1.273)). Rather, the patterns in Table 1 appear to suggest that the choice between L*+H 0% and L*+H H% is a matter of personal preference.

4.2.2 F0 measurements

Acoustic measurements were intended to (a) provide acoustic evidence of the auditory distinction between L*+H 0% and L*+H H% and (b) to provide evidence of truncation or compression on the accommodation candidates.

As an acoustic correlate of the apparently categorical auditory distinction between L*+H 0% and L*+H H%, F0 maxima were measured on the postaccentual syllable (i.e. the one which was assumed to reflect the presence of the trailing H of the L*+H pitch accent) and on the IP-final syllable. These particular measuring points captured best the different types of F0 patterns which were taken to reflect H%. The different F0 patterns involved (1) a gradual rise from the F0 peak reflecting the H in L*+H up to the IP boundary, or (2) a sharp rise in F0 involving only the last syllable of the IP or the last two syllables or (3) a combination of both. The measurements were taken on the first control in the three statement lists (i.e. *Brombeermarmelade,* the first word in the list). After the measurements had been taken, the test items from the three lists were pooled (i.e. all realisations of *Brombeermarmelade*, regardless of the length of the accommodation candidate the word was coordinated with) and sorted by boundary type (i.e. 0% or H%). Then F0 excursions were sorted by size in ascending order. Figure 10 below shows the results.

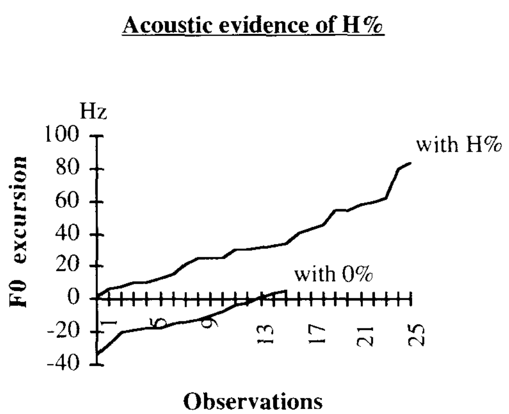


Figure 10 *F0 excursions between the postaccentual syllable and the final boundary for L*+H H% and L*+H 0% contours separately, in ascending order of magnitude.*

Figure 10 shows that in the majority of cases, the F0 excursion between the peak reflecting the trailing H and the IP boundary is larger for realisations transcribed as L*+H H% than for those transcribed as L*+H 0%. However, it is also clear that there is some degree of overlap. At first glance, this overlap may be taken to reflect the absence of a categorical distinction between L*+H 0% and L*+H H% in F0 patterns. However, the overlap may be explained with reference to the realisational options observed for L*+H H% in F0. H% may be reflected in a gradual rise in F0 towards the boundary and an additional rise at the boundary (Figure 11a), in a level pitch followed by a sharp rise (11b), or a gradual rise towards the boundary (11c).

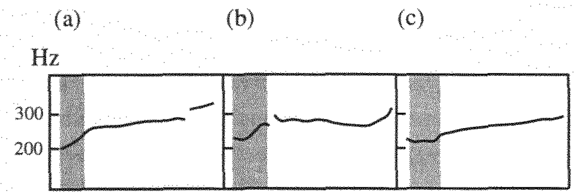


Figure 11 *Three possible realisations of L*+H H% on* ***Brombeermarmelade***.

Figure 11 shows that in principle H% may be reflected in a small rise in F0 on the IP-final syllable. However, F0 may drop slightly before this syllable is reached, and in that case, an F0 excursion measure involving the F0 peak on the postaccentual syllable and F0 at the boundary would not reflect the boundary tone, and the trace would appear to be level. However, the results of an alternative measure shown in Figure 12 below do not appear to differ greatly from the ones shown in Figure 10.

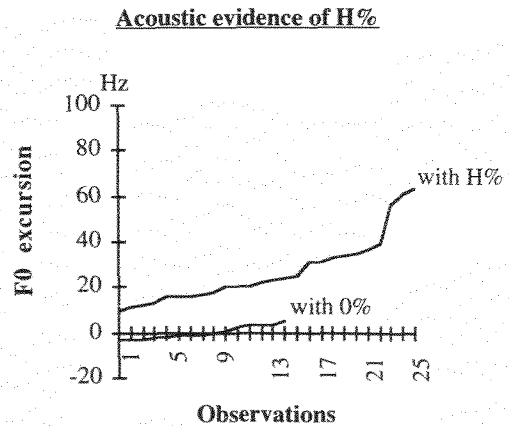


Figure 12 *F0 excursions between the penultimate syllable and the final boundary for L*+H H% and L*+H 0% contours separately, in ascending order of magnitude.*

Figure 12 shows the results of measuring the F0 minimum on the penultimate syllable and the F0 maximum on the IP-final syllable. This measure can capture the difference between IP boundaries realised as an upstep on the final syllable and those lacking the upstep, but it is unable to capture the auditory effect of, for instance, a gradual rise from the H up to the boundary.

In summary, it appears that there is auditory evidence distinguishing L*+H 0% and L*+H H%, but in fundamental frequency, the categorical auditory impression is harder to capture. L*+H H% is likely to exhibit either a greater F0 excursion between an F0 peak associated with the trailing H of L*+H and the IP boundary, or a greater F0 excursion between the last two syllables of the IP than L*+H 0%, but this difference is difficult to capture in a single acoustic measure.

The second purpose of the F0 measurements was to provide evidence of truncation or compression on the accommodation candidates in statement and question lists. Duration of F0 on the three test words and F0 excursion were measured, just as in the first pitch accent accommodation experiment, and the rate of F0 change was calculated. Figure 13 below illustrates the mean duration of F0 for the three test words and the mean F0 excursion.

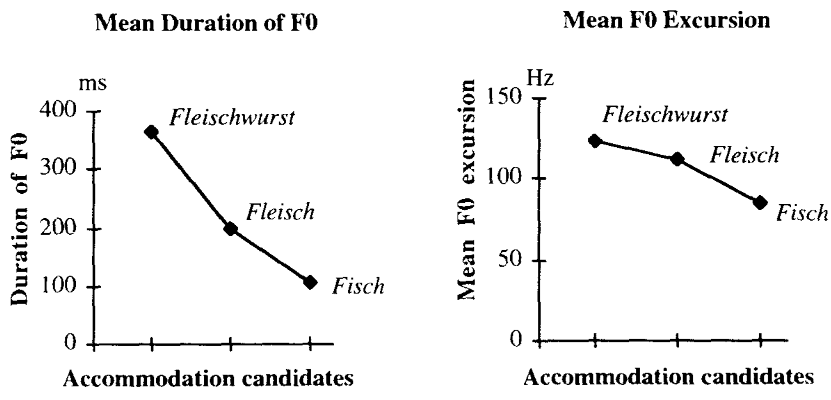


Figure 13 *Mean duration of F0 and mean F0 excursion on test words*[4].

The expectation had been that statement lists would be produced with L*+H 0% and question lists with H%. However, this turned out not to be the case; both lists contained

4 *The F0 excursion for shorter test words was smaller. However, at no time was the F0 excursion on Fisch as small as that measured on truncated H*+L. Compare the F0 excursion on Fisch in Figure 13 (mean = 79 Hz) with that of truncated H*+L on Schiff in Figure 5 above (mean= 7 Hz).*

examples of either nuclear tone. Therefore, the difference between lists with 0% and lists with H% could not be tested as originally planned. Instead of comparing the acoustic realisations of the three test items of different length in the statement with those in the question lists, the data from statement and question lists were pooled, and those items were selected in which subjects produced an equivalent boundary realisation on the preceding control (i.e. *Brombeermarmelade*) in all three conditions, i.e. they either produced realisations reflecting H% in all three controls or relaisations reflecting 0%. Items in which subjects produced, for instance, H% on *Brombeermarmelade* preceding the shortest accommodation candidate *Fisch*, but 0% on *Brombeermarmelade* preceding the mid-length candidate*Fleisch* were excluded from the analysis. This procedure yielded 11 sets of test sequences for 0% and 9 sets for H%. The two missing values in the H% sets were replaced with the relevant means of the 9 data sets available. Then the two datasets were subjected to a repeated measures ANOVA with the Factors Boundary (H% or 0%) and Length (*Fisch*, *Fleisch*, *Fleischwurst*). The results of the analysis showed that the difference between the sets with H% and those with 0% was not significant. The factor Length was significant ($F_{[2,20]}$=21.55, p< 0.001), but the factor Boundary was not. The interaction between Boundary and Length was not significant either (not also that the auditory analysis showed that all accommodation candidates were associated with rises in pitch, thus virtually excluding the possibility of any other accommodation effect but compression). Therefore, the acoustic measurements on items in statement and question lists were pooled and processed together[5]. The prediction was that in all items, the rate of F0 change would increase significantly from the longest to the shortest accommodation candidate.

An analysis of variance was carried out for the parameter ‘rate of fundamental frequency change’ with the factor Word Length. A significant main effect ($F_{[2,42]}$ = 24.31, p< 0.001) of Length emerged. The rate of F0 change increased significantly with decreasing segmental duration, confirming the prediction that compression would apply to all pitch accents when sonorant material was shortened. Planned comparisons revealed significant differences in rate of F0 change between the three words at the 1 % level. Figure 14 shows the mean rate of F0 change across the test words.

The results presented in Figure 14 suggest that nuclear rises compress, regardless of the boundary specification. A summary of the results is given in Figure 15. In the graph, fundamental frequency excursion on the vertical axis is plotted against its duration on the horizontal axis. The data supports the hypothesis that nuclear L*+H pitch accents compress, regardless of the following boundary specification.

---

5 *Note that at this stage, the items were not separated into those hypothesised to be L*+H H% and L*+H 0%.*

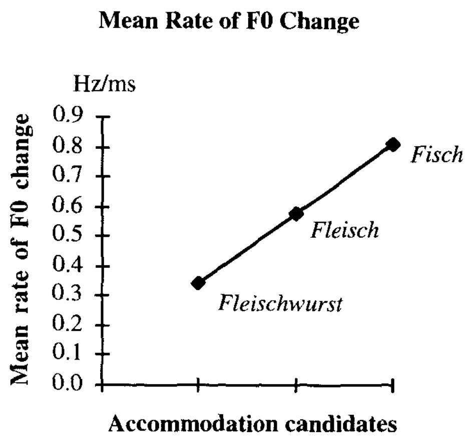


Figure 14 *Mean rate of F0 change on test words*

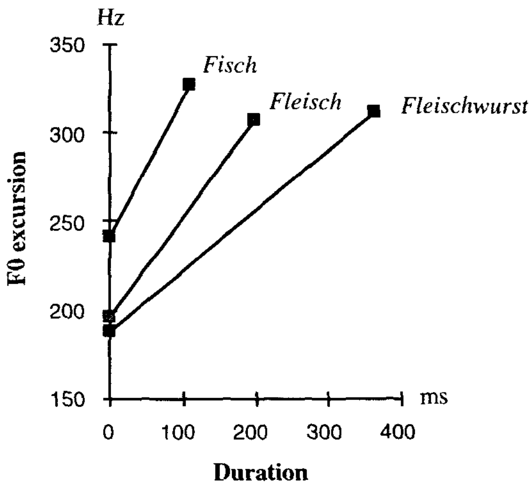


Figure 15 *Summary of results. Mean F0 excursion on accommodation candidates is plotted on the vertical axis against their duration on the horizontal axis.*

Figure 16 supplies representative F0 traces of the control *Brombeermarmelade* and compressed rises from one speaker. The patterns shown held for all speakers.

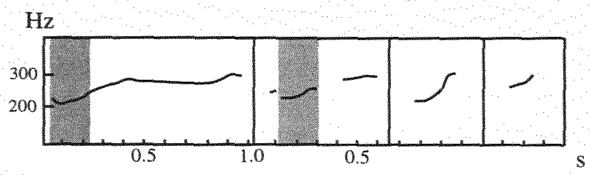
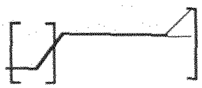


Figure 16 *Representative F0 traces of control and accommodation candidates.*

4.3 Discussion

The results of the pitch accent accommodation Experiment II appear to support the hypothesis that accent accommodation involves the nuclear pitch accent rather than the nuclear tone. The data set recorded contained coordination structures exhibiting the same nuclear tones within each coordination structure, and the nuclear tones were either L*+H 0% or L*+H H%. Nevertheless, all accommodation candidates which were coordinated with these nuclear tones appeared to compress, regardless of whether the preceding tone was L*+H 0% or L*+H H%. Secondly, the experiment showed that the materials consistently elicited realisations of L*+H on the test words, but the following boundary tone did not appear to be conditioned equally consistently. In individual lists, subjects chose either L*+H 0% for the complete list or L*+H H%. In only one case out of 42 were nuclear tones mixed. Additionally, many of the subjects produced consistently only one type of boundary specification or the other, and it is possible that the choice of 0% or H% is a matter of personal preference. The finding that boundary specifications appear to be a matter of personal preference may explain why Wunderlich's (1988:11) 'Echo accent' in German was drawn and transcribed as shown in Figure 17 rather than as two separate patterns (see Chapter 1, section 2.2.3.2). Wunderlich does not comment on the distinction shown in his representations, but the brackets in the transcription may suggest that the H% is in some way optional.

Echo accent



L* H (H%)

Figure 17 *Adapted from Wunderlich (1988:11).*

It is possible that in German nuclear rises, the boundary specification is in some way less relevant than the choice of pitch accent. Intuitively, the boundary specification adds little to the meaning of the phrase. However, there is little doubt that all speakers had both options, that is, both L*+H 0% and L*+H H% in their repertoire (e.g. the question list contained fewer instances of H% than the statement list).

A subsidiary finding involved the acoustic differences between L*+H 0% and L*+H H%. Apparently, a single acoustic correlate of an apparently categorical auditory distinction cannot be straightforwardly established. The auditory distinction between L*+H H% and L*+H 0% may correspond to (a) an F0 rise on the IP-final syllable, (b) a gradual rise from the postaccentual syllable or (c) a combination of (a) and (b),

Compression on German L*+H 0% and L*+H H% is modelled in Figure 18 below. Figure 18 shows that compression is preceded by a truncation of the F0 pattern on postaccentual syllables. Only when the pattern on postaccentual syllables has been truncated, is the fundamental frequency trace compressed on the two syllables on which the pitch accent L*+H is realised, that is, regardless of whether the underlying boundary is H% or unspecified.

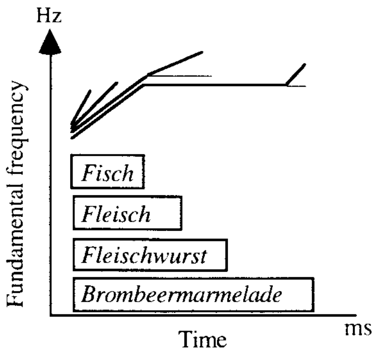


Figure 18 *Language-specific schematic representation of compression of L*+H 0% and L*+H H% in German.*

5 Summary

The first experiment presented in this chapter showed that English and German differ in the way they accommodate nuclear pitch accents when sonorant segmental material is scarce. In English pitch accents are compressed, that is, pitch rises and falls faster when a word is shorter. Acoustically, this observation appears to be reflected in an increase in the rate of F0 change. German, on the other hand, truncates falling accents. When sonorant material is scarce, the fundamental frequency trace simply ends earlier. However, German native speakers nevertheless appear to hear truncated H*+L as falling accents rather than level accents. Rising accents in German, on the other hand, are compressed, just as in English. Two accounts of truncation and compression were suggested. The first account involved the asymmetry which is claimed to characterise English and German IP boundaries. The suggestion was that German intonation is characterised by a pitch accent realisation effect 'truncation' which truncates the acoustic realisation of pitch accents. Truncation was said to be suspended when a pitch accent was followed by a high boundary tone. The second account suggested that truncation and compression effects relate to whether a nuclear pitch accent is rising or falling, and that the boundary specification did not play a role. German native speaker intuition appeared to support the second account.

A follow-up experiment on German provided some experimental support for the intuition that truncation and compression involve the nuclear accent rather than the nuclear tone. The second experiment tested whether accents in German compress regardless of whether the phrase boundary is tonally specified with H% or when it is not specified (0%). The results suggested that in German, rising pitch accents are compressed whether followed by 0% or H%. No evidence of truncation in L*+H 0% emerged. Additionally, the data showed that the experimental materials reliably elicited nuclear rises, but not the presence or absence of a boundary tone. Rather, H% boundary tones appeared to be speaker-specific choices. However, speakers did not mix nuclear tones within coordination structures; rather, coordination structures were produced with one nuclear tone or the other. This may suggest that H% in German is a speaker choice made at the level of the intonation phrase rather than at accent level.

Note, however, that unlike the results of experiment I which are based on auditory impressions and acoustic measurements taken directly from the accented words in question, the results of the experiment II are based on derived evidence. The test words were too short to allow a clear distinction between the nuclear tones compared (L*+H 0% and L*+H H%). Instead, the phonological representation of the test words was derived from structural information found in the intonational context. Consequently, the results of experiment II must be somewhat weaker than those of experiment I.

Language-specific schematic representations of truncation and compression were suggested for German. The representation of truncation in German showed why truncation does not appear to be a phonological process. The apparent absence of a fall in F0 on words with little scope for voicing appears to relate to the alignment of F0 on H*+L. The schematic representation for compression suggested that compression applies to the nuclear pitch accent rather than the nuclear tone. First, F0 is truncated on unaccented syllables following the nuclear syllable, and then F0 is compressed on the nuclear syllable.

Clearly, further research is needed. As pointed out in the discussion section of the first experiment, there are many ways in which the amount of sonorant material may be reduced in speech and other approaches may be taken towards acoustic measurements than the one taken here. Also, we know little about the pitch accent accommodation effects on intonational structure other than the ones investigated here; for instance, it is not clear how fall-rises are accommodated in English and German. English may compress fall-rises but German may truncate them.

The auditory effects of truncation require further investigation. Systematic perception experiments are needed (a) to confirm the impression that truncated falls are heard as falls by native German speakers, (b) to show how English native speakers perceive truncated falls in German, and (c) to show how English and German native speakers perceive the difference between truncated falls in German and compressed falls in English. Such a study may provide some evidence not only for language-specific realisation rules of pitch accents but also for language-specific rules for their perception.

Finally, it should be pointed out that the results presented in this chapter are restricted to Southern Standard British English and Northern Standard German. As has been shown for Swedish, different dialects do not necessarily follow the same realisations strategies (Bannert and Bredvad, 1975), and it is not impossible that other English dialects truncate and other German dialects compress.

Future research might investigate the physiological basis of truncation and compression, and this may shed more light on the strategies speakers employ when realising tones. Erickson *et al.* (1995) showed that in English, the infrahyoid strap muscles are active in the production of L tones, and this suggests the possibility that in languages which truncate falls, speakers do not make use of the strap muscles when segmental material is short. In languages where falls are compressed, on the other hand, they will attempt to do so, no matter how short the voiced portion of a word may be (Donna Erickson, personal communication). A replication of the experiment presented above combined with measurements of strap muscle activity could provide a test of this suggestion.

Downtrends

Chapter 6

1 Introduction

The auditory and acoustic analyses of read speech presented in Chapters 3 and 4 suggested (a) that both English and German have a pitch accent modification DOWNSTEP, and (b) that there may be cross-linguistic differences in the acoustic implementation of this modification. In English, an IP-final accent appears to be partially downstepped, that is, in F0, the target for the H* is always located above that of the following L, but in German, an IP-final accent can be partially or totally downstepped. This is reflected in the German version of ToBI (Grice et al., 1995), where a categorical distinction is made between !H*+L and H+L*. Figure 1 below shows examples of downstep from the corpus analysed for the purposes of the present study. Partial downstep in English is shown on the left, and partial and total downstep in German are shown in the right (the German examples were produced by two different speakers). All examples were produced in identical contexts. Note that the difference between partial and total downstep in German is auditorily salient.

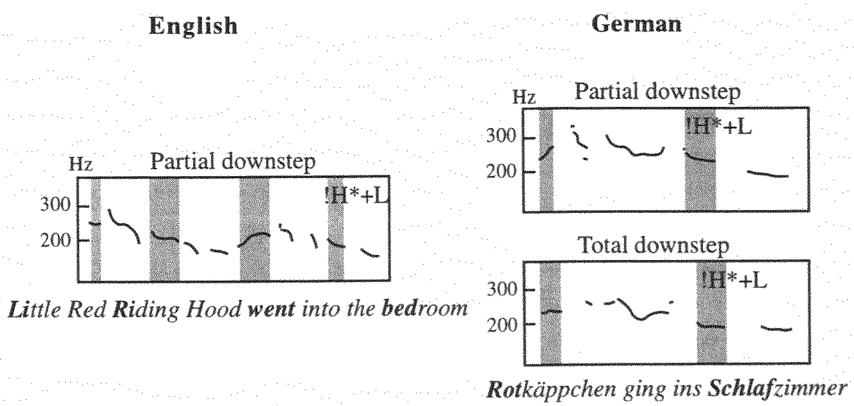


Figure 1 *Downstep options in read speech.*

However, in Chapter 4, the claim that English and German differ in the auditory and acoustic implementation of downstep was necessarily based on a small number of examples. Further data was needed to show how far the observed cross-linguistic difference could be generalised. Also, the corpus did not show whether the observed distinction in German was categorical in nature, and thereby potentially a matter of phonological structure, or whether it reflected a gradient, realisational option.

Two production experiments were carried out. The first investigated (a) the hypothesised cross-linguistic difference between English and German, and (b) the nature of the downstep distinction in German. The second addressed in more detail a discrepancy observed between the British English data investigated in the present study and the findings of a previous study on American English.

2 Background

At least since Pike (1945), it has been clear that F0 tends to decline over the course of phrases and utterances, and since then, this effect has become one of the most widely studied properties of fundamental frequency in speech (Ladd, 1984, 1996). The way in which declination should be modelled, however, has been a source of controversy, and a variety of views about the nature of declination have been put forward (see Ladd, 1984 for an overview). Nolan (1995: 242) offers a 'least controversial' definition; he suggests that declination may be seen as a statistical abstraction away from F0 contours; as long as one measures enough utterances and calculates means, a downward trend in F0 will emerge (note, however, that not all utterances must exhibit this downtrend; in questions, for instance, declination is often suspended, see e.g. Thorsen, 1980a).

Two main competing models of declination have emerged, the 'contour interaction model' and the 'tone sequence model'. Nolan (1995) uses the diagram shown in Figure 2 below to summarise essential differences between the models. In the contour interaction model (left), the scaling of successive accents is determined globally, that is, by an overall sloping contour associated with the complete intonation phrase. The assumption is that accent units find their place in the pitch range by latching onto the sloping utterance contour (Thorsen, 1980b, 1981, 1983). In contrast, in the tone sequence model, the notion of a sloping utterance contour is discarded. Instead, the model hypothesises that the pitch of successive accented syllables is determined locally, and within a 'two-accent window'. The location of each F0 peak in a sequence is calculated solely on the basis of the immediately preceding accent peak without reference to a global contour. Declination is then principally the result of a successive lowering of accented syllables (e.g. Pierrehumbert, 1980, Liberman and Pierrehumbert, 1984, Pierrehumbert and Beckman, 1988) and is referred to as 'downstep'.

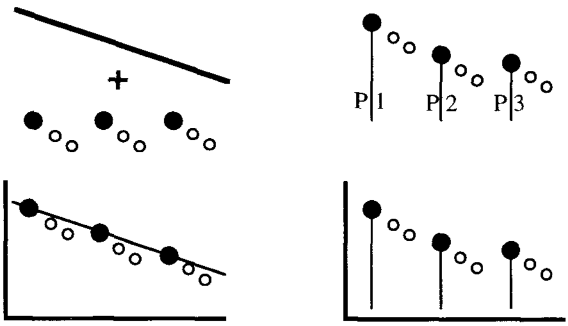


Figure 2 *Adapted from Nolan (1995). Filled circles represent accented and small open circles unaccented syllables. The sloping line on the left represents a sloping utterance contour.*

The experimental evidence given in the following sections of this chapter will be described within the tone sequence model. This is a practical rather than a theoretical decision; the tone sequence model is the one which has been widely adopted within the AM framework, and this is the framework adopted for the purposes of this study. Moreover, the aspects of fundamental frequency declination investigated in this chapter are restricted exclusively to those which appear to involve cross-linguistic differences, and the results are not claimed to be of sufficient generality to lend independent support to one model over the other.

The tone sequence approach to the modelling of fundamental frequency downtrends was first applied to English by Pierrehumbert (1980). The notion of downstep is important to her model of American English intonation and the AM framework in general, because it permits a modelling of tunes as linear sequences with only two pitch levels H and L, despite the fact that within one tune, some high targets may be lower than others. Pierrehumbert's (1980) work was further developed in Liberman and Pierrehumbert (1984) and Pierrehumbert and Beckman (1988), and the model of downstep first presented in Liberman and Pierrehumbert (1984) is probably the most explicit one currently available for (American) English.

Liberman and Pierrehumbert carried out several experiments which revealed three characteristic aspects of downstepped sequences. Firstly, the value of each accent peak in the sequence may be expressed as a constant proportion of the one immediately preceding given an appropriate mathematical transformation of the F0 space; secondly, English has 'final lowering', that is, the final accent in a sequence appears lower in the F0 range than predicted by the location of the immediately preceding accent; and thirdly, the final low in each IP is constant for each speaker. Liberman and Pierrehumbert's data led them to suggest that downstep may be modelled with an exponential decaying curve. Final lowering explained why the last accent in their sequences did not fit this curve. Figure 3 below summarises their findings.

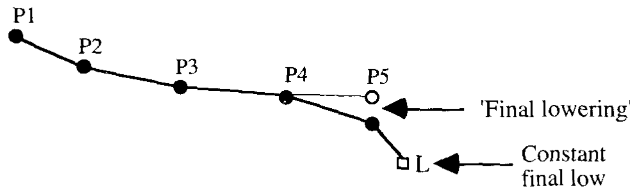


Figure 3 *The filled circles represent F0 peaks, and the empty one indicates where the last peak would be in an exponentially decaying curve in the absence of final lowering.*

The following sections will present several experimental investigations of fundamental frequency downtrends in English and German. Although the experiments presented were designed primarily to investigate potential cross-linguistic differences, rather than to confirm or challenge the details of Liberman and Pierrehumbert's model, the experiments carried out were modelled on Liberman and Pierrehumbert's experiment, and the expectation was that some of those authors' findings should be replicated. Specifically, successive accent peaks were predicted to form an exponentially decaying curve, with the steps between successive F0 peaks decreasing over the sequence, and evidence of final lowering was expected to emerge. The issue of the final low being constant for each speaker, however, was not addressed, as it was not directly relevant to the potential cross-linguistic differences investigated.

3 Downstep Experiment I

Downstep Experiment 1 was intended to establish (a) whether English and German differed in the acoustic implementation of downstep, and (b) whether the difference between partial and total downstep in German was categorical or gradient.

3.1 Method

Ten English and ten German subjects were asked to carry out two tasks, both based on Liberman and Pierrehumbert's (1984) 'berry name' experiment. In that experiment, three speakers (the authors and one other Bell lab employee) read 20 semantically bland lists of berry names such as *bayberries, raspberries and mulberries* etc. All lists had all different berries represented in all serial positions so that segmental effects on measured F0 could be assumed to be removed when peak F0 values in a given list position were averaged. The authors then measured peak F0 on each berry name in a list, as well as the final low, and used these data points to develop their model.

The experiment presented in the present study was intended to investigate downstep in productions from speakers naive to the purpose of the experiment rather than trained or semi-trained speakers. Therefore, the materials had to be modified. First of all, naive speakers cannot be expected to produce a very large number of downstepped sequences consistently without 'resetting' contours before the end of a list. Especially when sequences contain a relatively large number of items, naive speakers are likely to step down rather low on the first two or three accented words, reset the downstepping sequence, and start again higher in the register. Intuitively, on long lists, this seems easier than to produce consistently stepping patterns. Therefore, to reduce the number of lists to be read, each list was made up from initially stressed compounds which began with the same morpheme, for instance *Moonlight, moonlit, moonbeam, moonshine, moonstone* for English and *Mondbahn, Mondlicht, mondhell, Mondschein, Mondstein* for German (see Appendix B for the experimental materials). All compounds had two syllables; the first was fully voiced, and the second was kept short to keep the complete sequence as short as possible. Each of the subjects was then asked to read ten different sequences of this type, so that 100 downstep sequences were obtained for each language. During the German recordings, no attempt was made to condition the application of partial or total of downstep in any way; the application of either was hypothesised to be optional (i.e. in the corpus, partial and total downstep appeared in identical contexts).

As in previous experiments presented in this study, the subjects were told that the recordings constituted a pronunciation exercise for foreign learners of German or English. Additionally, they were asked to read 'casually', that is, not with exaggerated care, as the foreign learners needed to hear 'every day' German or English. An informal pilot experiment had shown that this last instruction was crucial. When subjects were asked to read the lists 'carefully', they were frequently reset, and produced with falling as well as rising accents within one sequence. In contrast, when subjects were asked to read casually, sequences were produced only with one type of pitch accent (H*+L) and quite consistently downstepped.

After having read the ten sequences (henceforth referred to as the 'production task'), subjects were asked to take part in a second task (the 'completion task'). The second task was designed to (a) collect more tightly controlled data on the realisation of the last accent in the phrase, (b) provide more data points (in the first experiment, some of the data was expected to be missing because of occasional resets), and (c) elicit data in which the final L could be measured. In the first task, where subjects produced five-accent sequences, some were expected to drop into creak when producing the final low in a list (this expectation was confirmed in a substantial number of cases). In addition, the completion task functioned as a backup. If subjects did not produce a sufficient number of consistently downstepped sequences in the first task, then the second task would still allow an investigation of a potential cross-linguistic difference in the realisation of the last accent.

In this task, subjects heard the initial fragments of 20 downstepped sequences over headphones and were asked to complete these sequences just as, in their view, the speaker would have done (the materials are given in Appendix B). They heard four accented words and were asked to fill in the fifth. While they listened, they read the relevant sequence on a sheet of paper, which also provided them with the required completion. Note that subjects were not asked to imitate the speaker's voice or the speaker's register, but simply to complete the sequences as if they were in the speaker's place[1].

The experimental stimuli were recorded by a female Southern British English speaker aged 20 drawn from the same pool as the experimental subjects, and myself, a native speaker of Northern Standard German from Braunschweig. Both speakers recorded complete five-accent sequences, which were then digitised in waves(tm). There, the last accent was removed, and an experimental tape was produced. On the tape, each downstep sequence was preceded by a warning tone and followed by a 5 second pause during which subjects were expected to complete the sequence. Subjects were given the opportunity to practise this procedure before the task was carried out.

Twenty female subjects took part in the two tasks. The ten English subjects were undergraduates from Oxford University and aged between 18 and 20. All came from the South of England and were judged by an English phonetician to speak a variety of Southern Standard British English. The data were recorded in a sound-treated booth in the Oxford University Phonetics Laboratory. The German subjects were aged 16 or 17 and drawn from the same pool as described in the previous chapter. The recordings were made in a quiet room at the Realschule Maschstraße in Braunschweig.

3.2 Analysis

The recordings from both tasks were digitised at a sampling rate of 16 kHz and processed in waves(tm) on an HP workstation A4032A under UNIX. An auditory analysis was carried out on the data from the production task to ascertain whether the sequences had been downstepped consistently. Table 1 below shows the number of items reset for each

1 *Naturally produced stimuli were used in this task rather than synthetic speech. In synthetic speech, the pitch ranges of the stimuli in the two languages could have been matched exactly. However, subjects might well have responded less favourably to being asked to complete the sequences as if they were in the speaker's place, if that speaker appeared to be a machine rather than a person. Adding some natural variation to the synthetic stimuli would have improved their suitability, but natural variation in downstep is difficult to mimic (e.g. what downstep factor one should use?)*

subject (each subject had read 10 items). Reset items were excluded from subsequent analyses.

Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Resets in English data	1	0	6	1	0	0	1	0	1	2
Resets in German data	1	6	1	0	0	1	0	1	2	0

Table 1 *Auditory analysis of the production task. Items reset for each English and German subject.*

Peak F0 was measured for each of the five accents produced. To avoid measuring F0 values resulting, for instance, from perturbations accompanying voiced stops, the highest F0 value appearing near the middle of the stressed vowel was taken rather than the highest F0 on the stressed syllable as such. Then, the F0 excursions between successive peaks were calculated for each subject, and the means were taken. These were then subjected to statistical analysis.

In the completion task, subjects had been required to produce the last accent only. Two measurement points were taken for each production; peak F0 and the lowest F0 in the following fall (i.e. measurements assumed to correspond to the target of the H* and the following L). Otherwise, measurements were taken as in the production task. Then, the excursion of the fall in F0 between the target of the H and that of the final L was calculated, and statistical analyses were carried out.

3.3 Results

3.3.1 Production task

Figure 4 below shows representative F0 traces of downstepped sequences from English and German. The German examples show an item produced with partial downstep and one with total downstep (produced by the same speaker); the English example shows partial downstep, the only version of downstep produced by the English subjects.

An analysis of variance (univariate, repeated measures) was carried out for the dependent variable 'step size in F0' with factors language (1,2) and Step (1,4). Step size rather than 'F0 peak location' was chosen as the dependent variable because overall, the German speakers had produced utterances with a somewhat higher pitch register than the English speakers (see Figure 6 below)[2]. Using step size meant that results of the statistical

[2] *On average, the German speakers were two years younger than the English speakers. The age difference may be responsible for the difference in pitch register.*

analysis would relate to the relationship between successive accents, the issue investigated here, and not to absolute differences between the two samples.

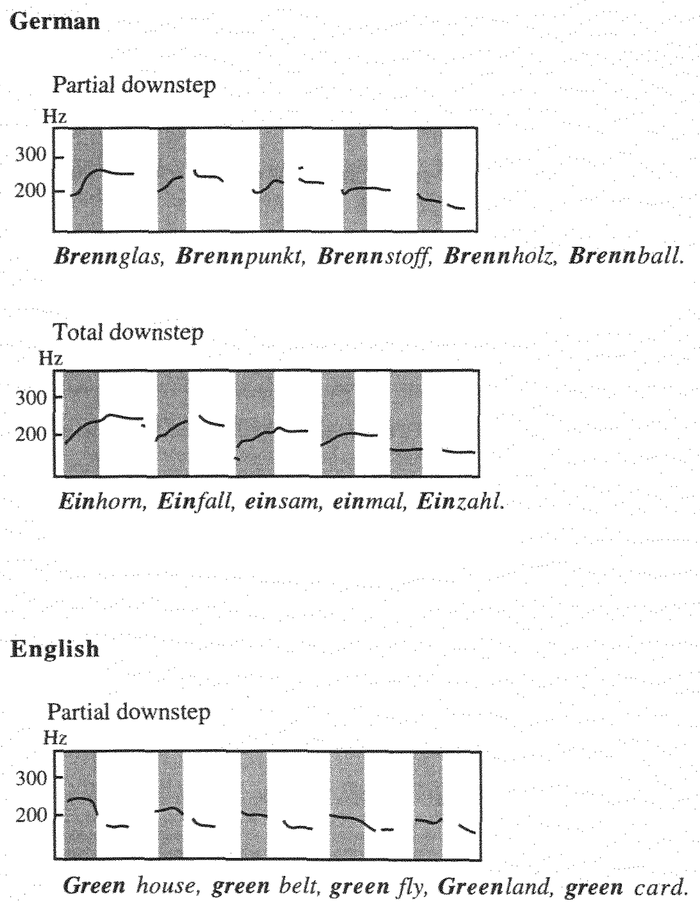


Figure 4 *Representative F0 traces of downstepped sequences for German and English. The German traces were produced by the same speaker.*

Significant effects of Language and of Step were predicted. The results confirmed these predictions; significant effects of Language ($F_{[1,9]} = 5.87$, $p<0.03$) and Step ($_{[3,27]} = 46.91$, $p<0.001$) emerged (significance levels unaffected by Greenhouse-Geisser correction, no significant interaction between Language and Step). Planned comparisons for Language within Step showed that in the two languages, the first three steps between F0 peaks did not differ significantly, but the last step did; in the German data, this step was significantly larger than in the English data ($p < 0.01$, significance level unaffected by Greenhouse-Geisser correction). Figure 5 below illustrates this finding. It shows the mean locations of F0 peaks in English and German. Overall, the curves for the two languages look quite similar, but in German, the step between the last two accents is relatively larger than in English.

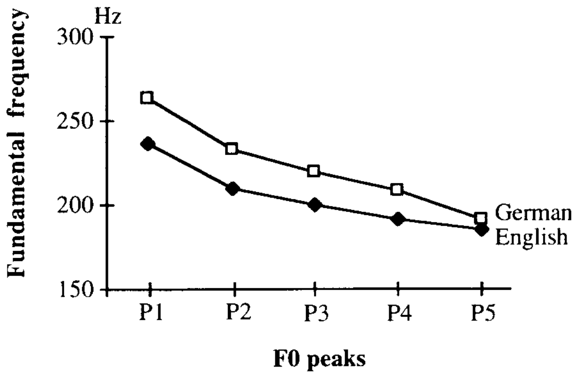


Figure 5 *Mean peak F0 in English and German downstepped five accent sequences. Means are shown of 100 English and 100 German contours.*

Additionally, the data were processed separately for English and German in order to establish whether the decrease in stepsize between accent peaks was significant or not. For both languages, a significant effect of Step emerged ($F_{[3,27]} = 26.53$, $p < 0.001$ for English and $F_{[3,27]} = 13.22$, $p < 0.001$ for German). Within English, significant differences at the 1% level distinguished the first step from the second, the third and the fourth step, but the second and the third step did not differ significantly from each other, nor did the third and the fourth (step sizes were 25.66 Hz, 9.78 Hz, 8.0 Hz and 6.3 Hz). Within German, the first step differed from the second and the fourth. The second step did not differ significantly from the third, and the third and the fourth step differed significantly at the 5% level before Greenhouse-Geisser correction ($p<0.07$ after correction, all other significance levels were unaffected by the correction). Note, however, that the last step in the German data was on average larger, not smaller than the preceding three steps (stepsizes were 30.47 Hz, 13.34 Hz, 9.92 Hz and 17.26 Hz respectively).

The results presented so far lend some support to the hypothesis that English and German differ in the acoustic implementation of downstep. The step between the last two accents is significantly larger in German than in English, and in German, the last accent in the sequence is often larger, not smaller than the preceding step. This finding may reflect that German has partial and total downstep and English only partial downstep. In a language with partial and total downstep, one would expect the mean value of the last F0 peak to be lower on average than in a language which has only partial downstep (assuming that a reasonable number of either appear in the data). However, alternatively, the results may reflect the fact that German has more 'final lowering' than English. Additional evidence is needed. If it were the case that in German, the standard deviation of the final step were larger than in English, then this would further support the hypothesis of German having partial as well as total downstep. This question was addressed with a further analysis of variance (univariate ANOVA, repeated measures) with the dependent variable 'standard deviation of F0' and, again, the factors Language (1,2) and Step (1,4).

The results showed that in German, the standard deviation of the final step was larger than in English. A marginally significant main effect of Language ($F_{[1,9]} = 6.63$, $p<0.05$) and a significant effect of Step ($_{[3,27]} = 8.57$, $p<0.001$) emerged (significance levels unchanged by Greenhouse-Geisser correction). The interaction between Language and Step was also significant ($F_{[3,279]} = 3.93$, $p<0.01$, $p<0.05$ after Greenhouse-Geisser correction). Planned comparisons investigating the effect of Language within Step showed no cross-linguistic differences for the first three steps. For last step, however, the standard deviation of F0 was significantly larger in German than in English ($p<0.001$). This finding further supports the hypothesis that German speakers have more options in the implementation of the last accent than English speakers do. Table 2 below gives the mean standard deviation of F0 excursion for English and German subjects.

	P1-P2	**P2-P3**	**P3-P4**	**P4-P5**
English	14.5	7.0	5.8	4.3
German	15.2	8.1	6.6	13.4

Table 2 *Mean standard deviation of F0 excursion in steps between successive accents (Hz).*

3.3.2 Summary

The results of the production task appear to support the hypothesis that English and German differ in the acoustic implementation of downstep. In German, the step between the last two accents is significantly larger than in English, though preceding steps do not differ significantly. Secondly, in German the last step exhibits a larger standard deviation than in English. Taken together, these findings can be interpreted to reflect partial and total downstep in the German data, and partial downstep in the English data. However, further evidence is needed, specifically about the relationship between the final peak in German and the following L. The results from the production task have not shown whether the final peak is ever stepped down as low as the final low and whether we may claim a categorical distinction between such cases and the ones where the final peak remains above the final low. Moreover, we do not know how the final peak in English relates to the following low. These issues were addressed in the analysis of the completion task data.

3.3.3 Completion task

The completion task was intended to provide detailed information on the realisation of the final accent in English and German downstepped sequences. Subjects appeared to have no difficulty in carrying out this task, and the completions they produced sounded, as expected, like the final item in a list. Figure 6 illustrates the mean peak F0 values in the experimental stimuli the German and English subjects heard (i.e. four accented words) and subjects' completions (i.e. the last word in the sequence). The figure shows that the value for mean peak F0 in the completions (P5) is higher than one might have expected considering the location of the last peak in the stimuli. Note, however, that the locations for P1-P4 shown for the stimuli represent means from one speaker in each language whereas P5 (the completions) represents mean peak F0 from 10 speakers with different registers. Also, the experimental task had not required the subjects to imitate the speaker's voice or speaker's register, so, presumably, speakers with a relatively higher register did not attempt to lower it to match that of the experimental stimuli[3].

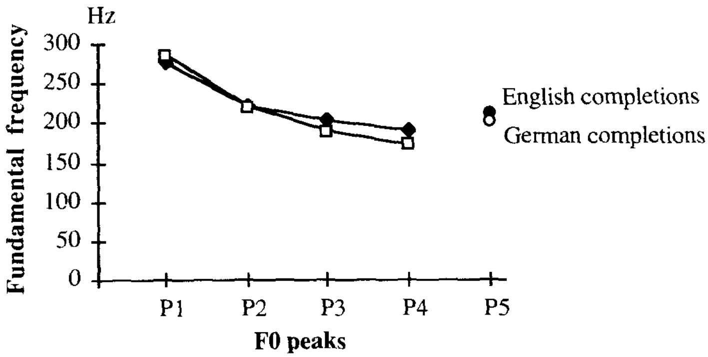


Figure 6 *F0 peaks measured on accents heard in the completion stimuli followed by F0 peaks in subjects' completions (the graph shows the mean of 200 English and 200 German completions).*

Figure 7 shows the mean peak F0 values measured in the stimuli recorded for the completion task with the last accent included. Here, the same cross-linguistic difference emerges as in the data from the production task. Again, for the German speaker, the step between the final pair of accents is larger than for the English speaker.

3 *Note that subjects were not matched for pitch range.*

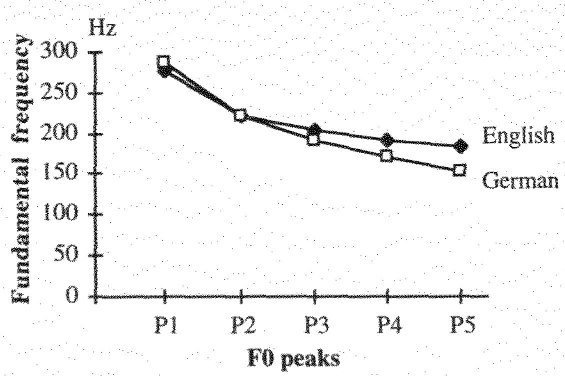


Figure 7 *Mean peak F0 in completion task stimuli.*

Figure 8 shows representative F0 traces for the accented items produced in the completion task. For English (left) a completion with partial downstep is shown, and for German (right), the figure shows one completion with partial downstep (top) and one with total downstep (bottom). Note at this point, however, that in general, in the German data, the distinction between partial and total downstep did not appear to be categorical, either auditorily or in F0. 'Extremes' at either end were easy to distinguish, but many in-between cases were observed also.

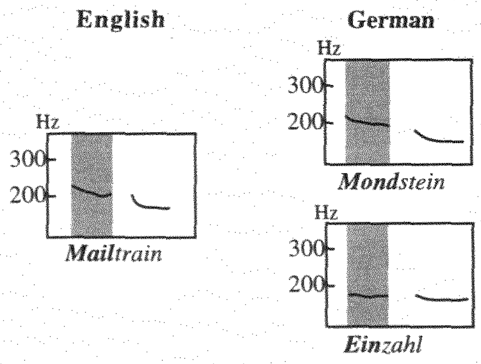


Figure 8 *F0 traces for accented items produced in the completion task. The German examples were produced by the same speaker.*

In the completions, the F0 excursion between P5 and the final low was calculated for each subject. Table 3 below gives the mean excursions for each language as well as their standard deviations. It shows that in English, the mean excursion of the final fall was larger than in German, but in German, the standard deviation of the fall was larger.

	English	German
Mean step size	40.44	25.03
Standard Deviation	11.52	15.24

Table 3 *Mean excursion and standard deviation (Hz) of the fall in F0 between the final peak (P5) and the final low.*

An independent samples t-test showed that the fall between P5 and the final low was significantly larger in English (df=398, t=-11.19, p< 0.001) whereas the mean standard deviation of this fall was significantly larger in German (df=38, t=-4.25, p< 0.001).

Figure 9 summarises the results. For maximal comparability, the data were normalised. The F0 of the final low was set to 181 Hz (the mean F0 of the final low in the German data) and the other data were recalculated accordingly[4].

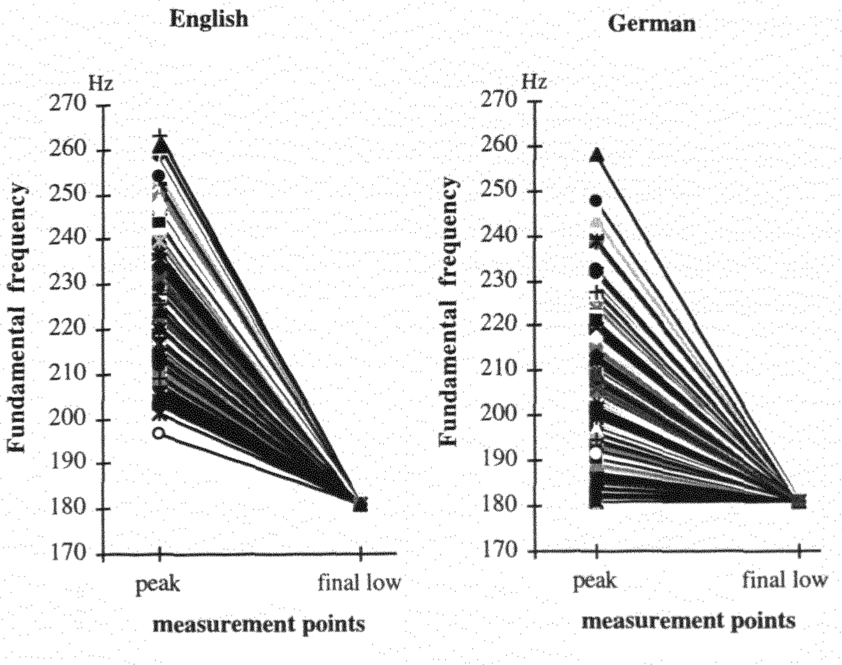


Figure 9 *F0 peaks and final lows measured in data from the completion task. All measured points are plotted.*

4 *The recalculation involved dividing each datapoint by the value in Hz of its final low and multiplying the result of this calculation by 181 (i.e. the mean F0 of the final low in the German data).*

Figure 9 shows that the English subjects produced a gradiently varying continuum of excursions for final falls. The differences between the languages arise because in German, the final fall may be virtually absent; the final peak may be stepped down to the level of the final low. In English, this does not appear to happen. Secondly, Figure 9 shows that there is no evidence of a categorical distinction between items with partial and items with total downstep. This would appear to confirm the auditory and visual impressions briefly referred to above. Thus, this result does not appear to lend support to the categorical distinction between !H*+L and H+L* which has been proposed in German ToBI.

The findings of the completion task further confirm the hypothesis that German and English differ in the realisation of the last accent. In German, the final fall is significantly smaller than in English, but this is because German speakers may step the last accent down as far as the final low. This interpretation is confirmed by the standard deviation of that fall being larger than in English. A categorical distinction between items stepped down partially or totally could not be established, either auditorily or in F0.

3.3.4 Modelling

The data from the production and completion tasks have shown a cross-linguistic difference in the implementation of the final accent in downstepped sequences. In German five-accent sequences, the step between the final pair of accents is significantly larger in than in English, but the final fall (i.e. the F0 excursion of the fall on the final pitch accent in the sequence) is significantly smaller. Additionally, the standard deviation of the peak F0 value of the last accent, and that of the final fall are larger in German. Figure 10 gives a schematic summary of the results.

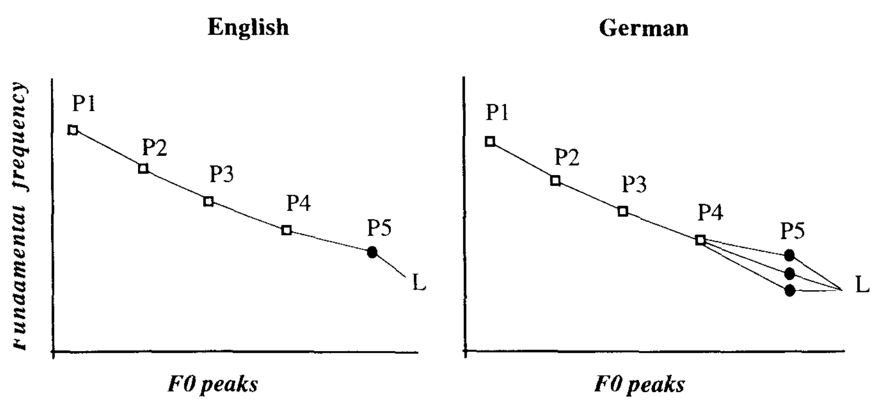


Figure 10 *Schematic summary of results. In the English data (left), the step between P4 and P5 is significantly smaller than in the German data (right). This is because German speakers place P5 on a continuum between the F0 level of P4 and the IP-final L.*

The results can be interpreted to suggest that English and German implement the final accent in a downstepped sequence differently. In both languages, the steps between successive pairs of accents get smaller as the sequences get longer, but when it comes to the final pair of accents, the languages differ. It would appear that German speakers have more options when it comes to placing the last accent, and this implies that the exact scaling of that accent is harder to predict than in English, where speakers do not appear to have similar options.

A remaining question asks to what extent the results of the experiment presented here replicate Liberman and Pierrehumbert's findings for American English. Among other points, the authors showed (a) a relationship between successive peaks which could be modelled as an exponentially decaying curve (and this implies that the steps between successive accents become smaller over the sequence), and (b) evidence of final lowering. Put simply, final lowering means that the last accent in a downstepped sequence does not fit into the exponentially decaying curve, but steps out of the curve; in a downward direction.

In the British English data presented here, we find that steps decrease between the first and second pair of accents, and the second and last pair. The step between the third pair does not differ significantly from either the preceding or following step. In German, all successive steps differed significantly from one another, and the last was larger than the one immediately preceding. It is not immediately clear how these findings should be related to Liberman and Pierrehumbert's; the authors did not carry out a statistical analysis of the type presented here, and therefore, the data are not immediately comparable. Nevertheless, at first sight, one might conclude that the German data fits Liberman and Pierrehumbert's model better than the English data; the German data *can* be modelled with an exponentially decaying curve, and there appears to be evidence of final lowering. From the English data, on the other hand, no evidence of final lowering emerged. A closer look at the German data, however, showed the apparently final lowering effect was not likely to reflect a constant lowering factor applied to the last accent in a downstepped sequence; rather, it appeared that German speakers have more freedom in the implementation of the final accent. In English, on the other hand, the implementation of the last accent appeared to be more restricted.

Thus, the results of the experiments presented in this chapter appear to suggest that unlike General American English, Southern Standard British English and Northern Standard German do not have final lowering. However, this conclusion may be too simplistic. In Liberman and Pierrehumbert's model, the notion of final lowering depends crucially on the assumption that downstepped sequences can be modelled with an exponentially decaying curve; it is not clear where exactly the cut off point is between an accent that is placed on the curve and one that is placed below it. This makes it difficult to assess when exactly final lowering can be claimed to have applied. How much lowering is final lowering? On the other hand, if one assumes that successive peaks do not form an exponentially decaying curve, but, for instance, a straight line, then the definition of final lowering is different. In that case, a significant difference between the pre-final and final step would have to be observed before the presence of 'final lowering' can be claimed. Thus, the evidence required for final lowering differs depending on the way one assumes downstep should be modelled. In this light, the data from the production experiment were re-examined.

Figure 11 shows mean peak F0 values measured on successive accented words separately for each of the ten English subjects (10 utterances per subject).

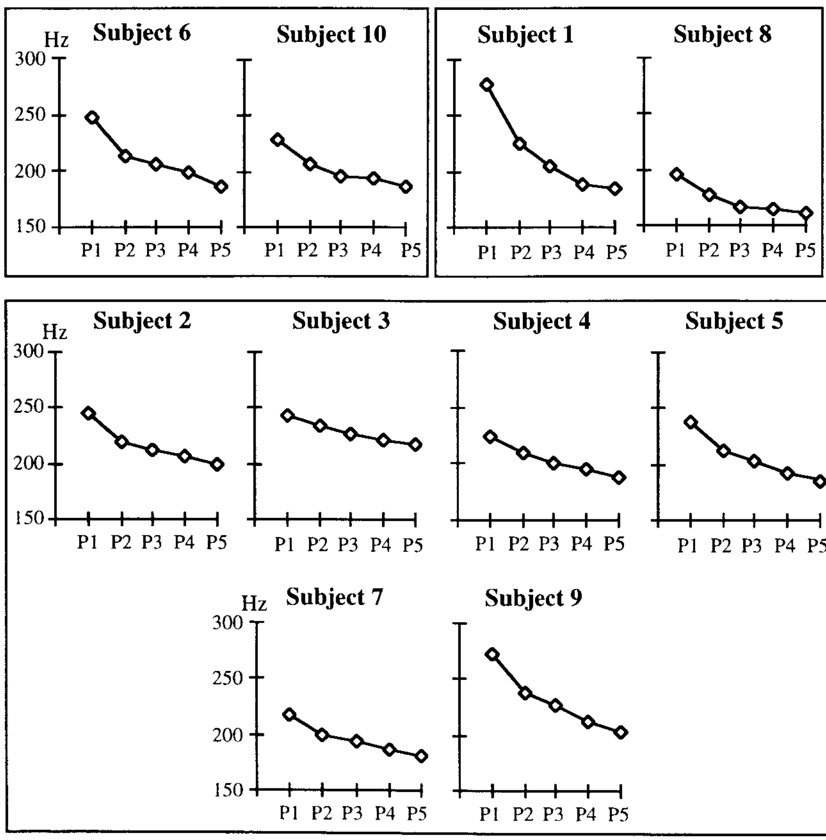


Figure 11 *Mean peak F0 values per English subject, grouped into those where the pattern resembles an exponential curve with final lowering (top left), an exponential curve without final lowering (top right) and those where the last four peaks form a straight line.*

Figure 11 shows that only two of the ten subjects produced a pattern which very obviously resembles an exponentially decaying curve with final lowering (subjects 6 and 10). In the realisations of the remaining subjects, the pattern resembles either a curve without final lowering (subjects 1 and 8) or a pattern in which the first step is relatively large, and the last four appear to form a straight line (subjects 2, 3, 4, 5, 7, 9).

Following Liberman and Pierrehumbert's approach to analysis, the structure of pitch accent progressions in the five accent sequences were examined statistically. The aim was to test the claim that the relationship of successive pitch accents is essentially exponential. Additionally, the possibility of cross-linguistic differences in the pitch accent progressions was assessed, and the variability between subjects was assessed (Liberman and Pierrehumbert tested only three subjects, who were trained or semi-trained; in the present study, ten English and ten German subjects were tested, and all were naive speakers).

The objective of fitting curves to complex data so that the effects of a number of different factors can be tested can be felicitously accomplished using General Linear Models. In effect, these allow both analysis of variance and multiple regression analyses of balanced or unbalanced data. The GLIM package (Healy, 1988) permits the experimenter to build a statistical model of the data interactively, testing the significance of individual factors and their interactions one by one. The significance of each factor is assessed by an ANOVA table in which (a) the sum of squares equals the reduction in variance brought about by adding that factor to the model (b) the residual sum of squares equals the remaining variance (c) the degrees of freedom are given (i.e. those in the factor, and those remaining in the data once the factor is added).

The variable to be modelled was the frequency of each pitch accent in the five accent sequences. The data were coded with the factors Language, Subject, Line, and Curve. The factor Language tested for cross-linguistic differences and the factor Subject for inter-subject variability. The factor Line tries to assess to what extent the decrease in pitch from one accent to another is due to a linear progression from one accent to the next; in other words, if by adding the Line factor, all the variability in the data is accounted for, then a linear curve is the appropriate way to model the progression. The factor Curve tries to assess whether the decrease in pitch from one accent to another can be modelled with an exponentially decaying curve. Curve was designed to approximate a very simple downstep model with a downstep constant of 0.5 (Liberman and Pierrehumbert give downstep constants 0.59, 0.64 and 0.62 for the three subjects they tested). Thus, the step size between successive pairs of accents is assumed to be halved at every step (Line = $step/2^{step-1}$). If Liberman and Pierrehumbert's model is broadly applicable to the data analysed in the present study, then the variable Curve should give a better fit with the data in the current experiment than the variable Line.

A simple model attempted was one which included Language, Line and Curve as factors[5]. All factors were highly significant (ANOVA tables are given in Appendix E). This finding suggests that the progression of pitch accent peaks in the downstep sequences results neither from an exclusively linear lowering, nor from an exclusively exponential one. Apparently, an equation involving both terms fits the data collected for the purposes of the present study best. Secondly, the two languages differ significantly. However, the curves fitted in this model are parallel, that is, at this stage, no factor has been included which tests whether the languages differ in the shape of their downtrend (in other words, the significant effect of Language may be simply due to a cross-linguistic difference in pitch range). The question of whether the languages differ in the shape of their downtrends is addressed by building into the model the interactions Language - Line and Language - Curve. These proved highly significant and suggest that, apparently, different terms must be added to the data from each language to make the equations fit (i.e. different proportions of linear and exponential terms). The addition of the factor Subjects showed that this factor was significant also ($p < 0.01$), not only across but also within languages. The addition of this factor did not affect any of the others already in the model. Clearly, a significant amount of variability can be expected in the implementation of downstepped sequences by naive subjects. Taken together, the results of the modelling show (a) that neither an exponential curve nor a linear progression of pitch accent to the next models the data perfectly, and (b) that native speakers vary significantly in their implementation of downstep.

How can we relate this finding to the question of whether English has final lowering or not? As discussed earlier, had the results of the GLIM analysis suggested that successive downstepped pitch accents form a straight line rather than an exponential curve, then the absence of a significant difference in step size between the pre-final and final step between downstepped accents in the English production experiment would have lent further support to the hypothesis that (British) English does not have final lowering. However, it appears that the pattern involves elements of a straight line as well as elements of an exponential curve. This finding suggests that final lowering is a concept which can neither be defined nor falsified straightforwardly. More research is called for.

3.3.5 Summary

In summary, the results of the first downstep experiment have shown the predicted cross-linguistic difference in the implementation of downstep in F0, but they have not unambiguously confirmed previous findings for English in the literature. Apparently, the

---

5 *Note that the factor Subjects is embedded within the factor Language, since all subjects must fall into just one of the two language categories. Subjects was therefore excluded from the first model tested to allow the significance of Language to be tested straightforwardly.*

progression of downstepped accents can be modelled as a pattern which contains elements of a straight line as well as an exponential curve, and clear evidence of final lowering is lacking.

However, the fact remains that Liberman and Pierrehumbert found evidence of final lowering across all subjects and conditions. Why the discrepancy? Should the results presented in this study be interpreted as suggesting that American English intonation exhibits an effect which British English intonation does not? This is not impossible, but in this particular case, the reason for the apparent discrepancy may be found elsewhere. A comparison between Liberman and Pierrehumbert's experimental stimuli and the ones used in the present study shows that the materials differ in one, apparently small aspect. In Liberman and Pierrehumbert's materials, the last two berry names were connected with <and>, but in the materials used in this study, the <and> was left out to ensure maximal regularity in the productions. Might the additional segmental material inserted between the last two berry names have been responsible for the effect? This question was investigated in a second downstep experiment.

4 Downstep Experiment II

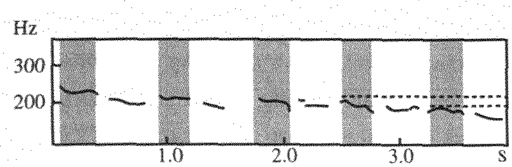
Pierrehumbert and Beckman (1988) discuss three sources of fundamental frequency downtrends: (1) declination, a time-dependent decrease in fundamental frequency over the course of an utterance, (2) downstep, a lowering of accents which goes beyond the effect of declination, and (3) final lowering, an effect which lowers the peak of an intonation phrase-final or utterance-final accent. Evidence from Liberman and Pierrehumbert's (1984) study suggested that of these three factors, downstep and final lowering account for fundamental frequency downtrends in English. The data presented in the previous sections of the present study, however, do not appear to lend support to all of Liberman and Pierrehumbert's findings. A downstep-type downtrend in fundamental frequency was observed, but evidence of final lowering was elusive. This finding may either suggest that Southern British English does not have final lowering or that the effect Liberman and Pierrehumbert report is not the result of final lowering, but the result of something else. The obvious candidate would be declination. This interpretation of Liberman and Pierrehumbert's data is supported by a difference in the structure of Liberman and Pierrehumbert's experimental materials and the materials used in the present study. In Liberman and Pierrehumbert's test sentences, the IP-final pair of berry names was connected with <and>, that is, between the final pair, additional segmental material was inserted between the stressed syllables. If downstep sequences in English are characterised by declination as well as downstep, then the effect of declination should become obvious when additional segmental material is inserted between otherwise regularly distributed stressed syllables. A second downstep experiment was carried out to test this hypothesis. If Liberman and Pierrehumbert's final lowering effect was, in fact, the result of declination, then this effect should be replicable in Southern British English.

A pilot study was carried out. One German and one English subject were asked to produce the materials from the production task (i.e. the ten sequences tested in that task) with and without *and* connecting the last two words in the list. The results appeared to suggest a 'final lowering' effect in both languages when the <and> was inserted. Figure 12 shows English and German examples. The step between the last pair of accents (indicated by the dotted line) appears to be larger in productions with <and> than in those without.

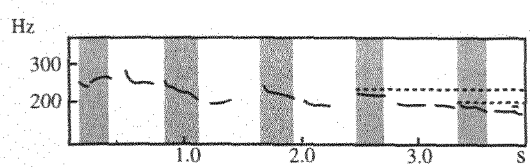
Additionally, to obtain some more controlled data, both subjects were asked to carry out the completion task with and without <and>. Figure 13 below shows the results. A t-test showed that in tokens with *and*, the last accent was placed significantly lower than in those without in English ($df=19$, $t=6.42$, $p<0.001$), and in German ($df=19$, $t=3.94$, $p<0.01$).

The findings of the pilot experiment appeared to suggest that the location of steps in successive downstepped accents is not independent of the amount of segmental material between them. Thus, a second production experiment was designed which was intended to investigate this observation in more detail. The hypothesis was that the steps between successive F0 peaks depend not only on their location in F0 but also on the amount of segmental material between them. The experiment was carried out with English subjects only, as no further cross-linguistic differences were expected, and Liberman and Pierrehumbert's experiments had investigated English data.

English

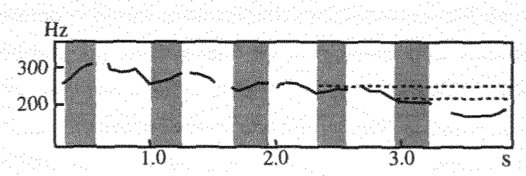


Ground *crew,* ***ground*** *fog,* ***ground*** *swell,* ***ground*** *floor,* ***grounds*** *man.*

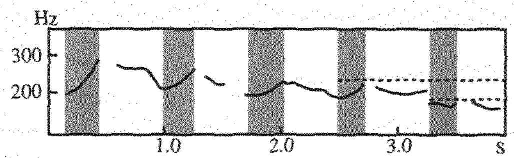


Ground *crew,* ***ground*** *fog,* ***ground*** *swell,* ***ground*** *floor, and* ***grounds*** *man.*

German



***Mond**bahn, **Mond**licht, **mond**hell, **Mond**schein, **Mond**stein.*



***Mond**bahn, **Mond**licht, **mond**hell, **Mond**schein und **Mond**stein.*

Figure 12 *Five-accent sequences produced with and without <and>. The dotted lines indicate the step between the last pair of accents.*

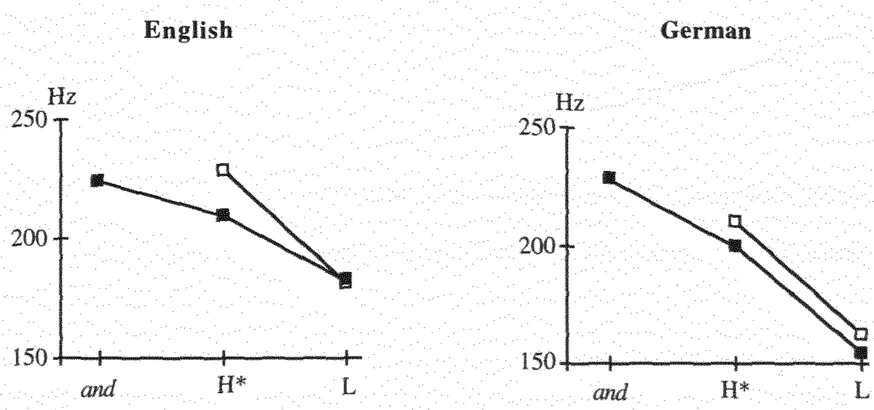
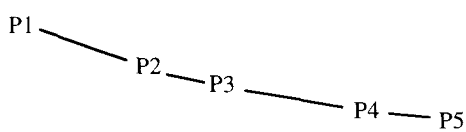


Figure 13 *Measurements of F0 in <u>and</u> (middle of vowel), peak F0 on the final accented word and final low in productions of the completion task with <and> / <und> (■) and without (□) .*

4.1 Method

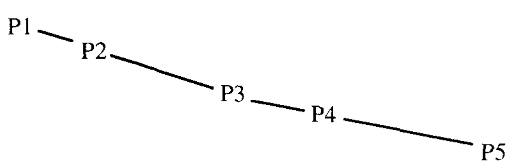
The stimuli were based on those used in the production task in the first downstep experiment, but now, the intervals between stressed syllables were varied systematically. Two sets of experimental materials were created; in the first set, the interval between stressed syllables was 'long-short' (two or three intervening syllables vs. one syllable, e.g. ***Moon***_lighting,_ ***moon***_lit,_ ***moon*** _landing,_ ***moon***_beam,_ ***moon***_flower_), and in the second, it was 'short-long' (1 syllable vs. 2/3 syllables, e.g. ***Moon***_lit,_ ***moon***_lighting,_ ***moon***_beam,_ ***moon*** _landing,_ ***moon***_struck._). Ten sequences of each type were designed. The predictions are illustrated in Figure 14 below.

Long-short



Moon_lighting,_ ***moon***_lit,_ ***moon*** _landing,_ ***moon***_beam,_ ***moon***_flower._

Short-long



Moon_lit,_ ***moon***_lighting,_ ***moon***_beam,_ ***moon*** _landing,_ ***moon***_struck._

Figure 14 *Predictions for the second downstep experiment.*

In the long-short list, a significantly larger step down in F0 was predicted between P1 and P2 than for the short-long list. Between P2 and P3, on the other hand, a relatively smaller step was predicted for the long-short list etc. Ten female undergraduates aged 18-20 from Oxford University took part in the experiment. Again, all were speakers of a variety of Southern Standard British English. The recordings were made in a quiet room.

4.2 Analysis

The data were digitised at a sampling rate of 16 kHz and processed in waves(tm) on an HP workstation A4032A. An auditory analysis was carried out to determine whether accent sequences had been produced without 'reset'. Those that had been reset were excluded from the analysis. Next, peak F0 was measured on the stressed morpheme of each compound in each sequence (see section 3.2 of this chapter for details). At this stage, a further four items were excluded. Creaky voice had caused a number of tracking errors which made it impossible to determine reliably peak height for all items. Table 4 below gives the number of items excluded from each list. Mean step sizes were calculated for each subject, which were then subjected to statistical analysis.

Long-short condition

Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
reset / creak	8	3	2	1	0	0	1	1	3	8

Short-long condition

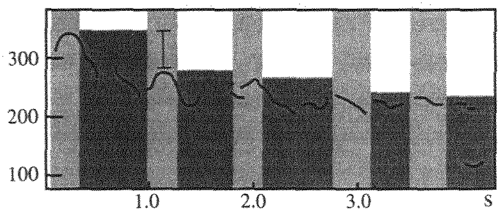
Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
reset / creak	6	0	4	1	0	2	4	0	1	8

Table 4 *Sequences excluded from the analysis for each subject.*

4.3 Results

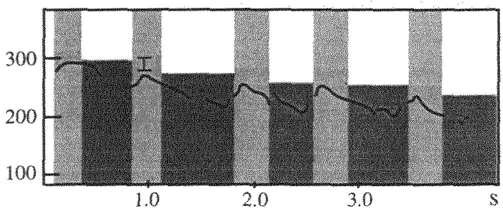
The data confirmed the predictions; when more segmental material was inserted between accented words, the steps down in F0 between them became larger; when there is less, they became smaller. Figure 15 shows representative F0 traces from one speaker from each language. The accented syllables are marked in light grey, and the steps down between successive accents in dark grey (note, however, that not all speakers realised the downstepped accent as H*+L; some produced non-final accents as H*>; see Chapter 3, section 3.2). A comparison between, for instance, the first pair of accents shows that the step is relatively larger in the long-short condition, where there is more segmental material intervening between accented words than in the short-long condition, where there is less.

Long-short condition



***Moon**lighting, **moon**lit, **moon** landing, **moon**beam, **moon**flower*

Short-long condition



***Moon**lit, **moon**lighting, **moon**beam, **moon** landing, **moon**struck*

Figure 15 *F0 traces from one speaker for one item in the long-short condition (above) and the short-long condition (below).*

An analysis of variance (repeated measures) was carried out for the dependent variable 'step size in F0' with the factors List (1,2) and Step (1,4). The statistics were calculated on the basis of the means for each subject, and the data from subjects 1 and 10 were excluded because most of those subjects' tokens had included a resetting of the downstep sequence (see Table 4 above).The results revealed a significant effect of Step ($F_{[3,18]} = 5.68$, $p<0.01$ before and after Greenhouse-Geisser correction), but not of List, and a significant interaction between Step and List ($F_{[3,18]} = 15.09$, $p<0.001$ before and after Greenhouse-Geisser correction). Planned comparisons for List within Step were carried out and showed that the first, third and fourth steps in the two lists differed significantly from each other at the 1% level before and after Greenhouse-Geisser correction. The second steps differed significantly from each other at the 5% level ($p<0.6$ after Greenhouse-Geisser correction and $p<0.4$ after Huynh-Feldt correction).

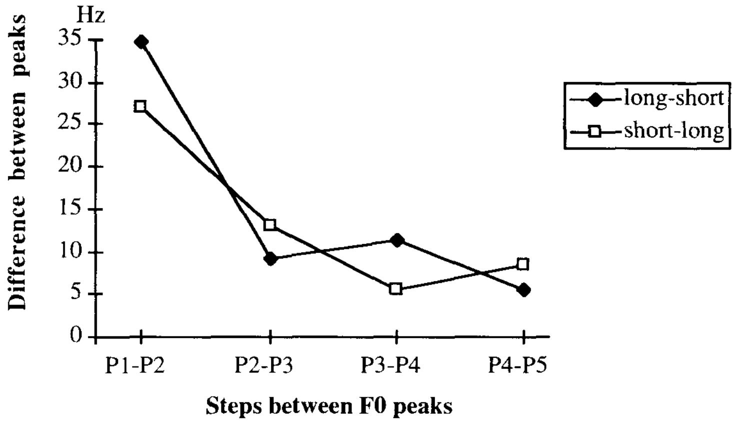


Figure 16 *Steps between successive F0 peaks in the two conditions averaged over all items. P1 stands for the first peak, P2 for the second etc..*

4.5 Discussion

The findings of the second downstep experiment suggest that the final lowering effect observed in Liberman and Pierrehumbert's data may indeed be at least partly explicable as the result of declination[6]. Apparently, the amount of segmental material between stressed syllables affects the location of successive accent peaks; steps between successive accent peaks increase when more segmental material is inserted, and decrease when there is less. More generally, the results can be interpreted to suggest that a model of downstep which seeks to predict the locations of successive accents in F0 locally, rather than globally, must take into account the duration of segmental material between the peaks as well as the locations of successive peaks in F0.

A number of questions remain. Firstly, what shape does the succession of peaks take when steps between them are not of equal length? Do they still form a more or less smoothly descending line? Figure 17 below illustrates the results of the F0 peak measurements. Two arbitrary units of distance have been inserted between peaks for 'long' distances between accents condition (e.g. ***Moon**lighting, **moon**lit)* and one unit for 'short' distances. conditions (e.g. ***Moon**lit, **moon**lighting*).

---

6 *Poser (1984) and Pierrehumbert and Beckman (1988) show that downstep and declination co-exit in Japanese.*

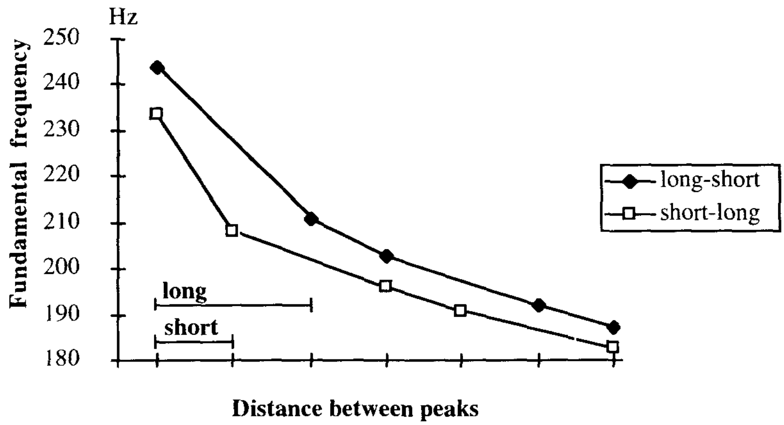


Figure 17 *Mean peak F0 values in long-short and short-long lists.*

Figure 17 shows that the pattern formed by successive peaks does not look particularly different from those observed in the first downstep experiment. As in the productions of most of the subjects tested in that experiment, in Figure 17, the last four accents form essentially a straight line. The pattern suggests that the amount of segmental material between stressed syllables may affect the realisation of the size of the steps between them, but not the overall pattern of the sequence.

5 Summary

The experimental evidence presented in this chapter has shown that English and German differ in the implementation of downstep in fundamental frequency. Downstep Experiment 1 showed that in five-accent sequences, the step between the last pair of accents was larger in German than in English, but all other steps did not differ across languages. The final fall between the last accent and the following low, however, was smaller in German. Moreover, the standard deviation in the step between the final pair of accents and in the final fall was larger in German. Combined, these results are interpreted as supporting the hypothesis drawn up on the basis of the cross-linguistic corpus study; Southern Standard British English has partial downstep, and Northern Standard German has partial as well as total downstep. The distinction between partial and total downstep in German, however, is gradient rather than categorical. This finding calls into question a categorical distinction between pitch accents !H*+L and H+L* which has been suggested for German in German ToBI (e.g. Grice and Benzmüller, 1995).

Secondly, the results from the first downstep experiment showed that 'final lowering', an effect claimed to characterise downstepped sequences in American English, appeared to be absent in British English. The apparent final lowering effect found in American English was hypothesised to be the result of declination rather than final lowering. A follow-up experiment was carried out to investigate this hypothesis. The results confirmed that the final lowering effect observed in Liberman and Pierrehumbert's study of American English could be explained by invoking the concept of declination.

Summary and Conclusion

Chapter 7

1 Introduction

Previous contrastive analyses of English and German intonation have disagreed on whether the intonation of the languages is quite similar or fundamentally different. The present study offers a resolution to this controversy. The combination of (a) an autosegmental-metrical approach to contrastive analysis and (b) directly comparable samples of speech has shown that the two languages can be described as having the same inventory of basic intonological categories. This explains why some authors have claimed that English and German intonation are very similar. The languages differ, however, in the acoustic phonetic realisation of the falling pitch accent H*+L. The peak in H*+L is aligned differently; H*+L is accommodated differently when sonorant material is scarce, and the implementation of IP-final downstepped H*+L is different. This explains why other authors have suggested that English and German intonation are fundamentally different.

The cross-linguistic differences in the realisation of H*+L which the present study has established shed light on a further discrepancy in the literature. They explain, at least to some extent, why analysts of English intonation have tended to agree on what the basic intonational categories of English are, but among analysts of German intonation, a comparable consensus has been lacking. In English, the phonetic realisation and the phonological structure of H*+L match relatively well. H*+L is realised as what one may call a 'prototypical' falling accent, that is, as a straightforward fall in pitch. In F0, the high target is realised as a peak within the stressed syllable and a fall onto the following syllable. This finding can be related to the success of the unilinear, auditory approach favoured by the British school. If phonetic realisation and phonological category match well, little ambiguity in the analysis is likely to arise. In German, on the other hand, phonetic realisation and phonological structure of H*+L do not match as straightforwardly as they do in English. H*+L is closer to a rise-fall in pitch and F0, and as a result, this category is harder to distinguish from phonological categories such as 'rising-falling' and 'rising' accents than is the case for English H*+L. This finding can be related directly to the lack of agreement on basic categories found among studies of German intonation.

The following sections summarise the preceding chapters and discuss the scope of the present study. Then, the methodological and theoretical implications of the findings will be discussed.

2 Summary

Chapter 1 of the present study surveyed the relatively small number of previous studies which have compared English and German intonation. The survey showed that these studies have produced a wide spectrum of opinions. In Chapter 1, it was argued that this spectrum of opinions may have arisen because the intonational structures of the languages may be similar at one level of linguistic representation and different at another. Unilinear approaches to intonation analysis, such as the ones which all previous contrastive studies on English and German intonation have taken, cannot account for cross-linguistic similarities and differences at different levels of representation. A multi-level approach to intonation analysis such as the autosegmental-metrical framework, on the other hand, can. Additionally, previous comparisons have not tended to generate hypotheses about cross-linguistic differences and similarities from utterances which were directly comparable across languages, and the generalisability of particular analyses to a group of speakers was not demonstrated.

In Chapter 2, an analysis within the AM framework was developed specifically for the comparison of English and German. Developing such a system was necessary because the languages had previously been accounted for in different versions of the framework. English and German versions of the ToBI system for prosodic labelling were argued to be a questionable starting point for comparative analysis. Firstly, the mixed-headed pitch accent inventory of English and German ToBI was drawn up largely on the basis of data from English. If one wishes to transcribe German intonation with a mixed-headed phonological inventory developed for the transcription of English, then one needs to have sufficient data on pitch accent realisation in German, otherwise, ambiguity between accent transcriptions may be the result. Such data, however, were lacking. Secondly, ToBI posits two levels of intonational phrasing, but the phonetic correlates distinguishing the two levels are not clearly specified. Finally, ToBI offers a relatively non-transparent account of intonation phrase boundary specifications. For the purposes of cross-linguistic comparisons, a more transparent boundary account is preferable. In response, in the basic comparative system developed in the present study, all pitch accents were represented as left-headed, and one level of intonational phrasing was argued to be sufficient. The treatment of intonation phrase boundaries differed from that offered in ToBI in that boundaries could be tonally specified, but did not have to be.

Additionally, following Gussenhoven (1984), the system used for comparison postulated two levels of phonological representation. At the *underlying* level, the basic accent and boundary tone inventory was specified. The *surface* phonological level accounted for changes in the tonal structure of the primitives when these were combined into phrases and utterances. Splitting the phonological level of representation into an underlying and a surface level is advantageous in cross-linguistic work because similarities and differences between contours can be captured more explicitly than in a system in which every difference between contours is reduced to a different pitch accent choice from the phonological inventory.

Chapter 2 was concluded with the presentation of a new method for cross-linguistic comparison of intonation. Directly comparable samples of English and German speech data were collected from five English and five German speakers matched as closely as possible within languages for age, 'class' and educational background. The speaker read a text which they were familiar with; the fairy tale 'Little Red Riding Hood'. A high degree of familiarity with the text ensured that the text was interpreted similarly by all speakers. Realisations of specific texts produced by different speakers in identical contexts were then compared within and across languages.

In Chapter 3, data from the Northern Standard German corpus were presented, and in Chapter 4, these data were compared with data from Southern Standard British English. The comparison showed that both languages can be accounted for with two basic pitch accents H*+L and L*+H and a boundary tone H%. The specification of a low boundary tone was argued to be redundant. At the surface phonological level, the categorical phonological adjustment rules DOWNSTEP, DISPLACEMENT and DELETION were proposed to apply. The same adjustments were found to apply in the English data, but additionally, evidence of Gussenhoven's (1984) modifications DELAY and HALF-COMPLETION emerged. However, the evidence for HALF-COMPLETION was limited, and it is possible that this modification is modelled more adequately as phrasal downstep. Also, the same data called into question Gussenhoven's distinction between two types of categorical phonological adjustments. In his (1984) account of English, modifications apply to nuclear accents, and linking rules to prenuclear accents. In the German data analysed in the present study, however, DELETION applied to both nuclear and prenuclear accents. Therefore, in the present study, modifications and linking rules were collapsed into a single group of phonological adjustments, and the application of particular adjustments to particular elements in the tune was suggested to be language-specific. Both findings, that HALF-COMPLETION cannot be clearly distinguished from phrasal downstep, and that the distinction between prenuclear and nuclear adjustments is not as clear cut as Gussenhoven (1984) suggests[1], call for further research into the

[1] *Note, however, that Gussenhoven (1984) investigated English intonation, and that the present study has replicated his findings for English. In English, the distinction between prenuclear linking rules and nuclear modification can be upheld. The evidence from German may suggest, however, that the application of particular modifications to*

acoustic and auditory effects of modifications and suggest that a revision of their theoretical status is needed.

Cross-linguistic differences emerged in the acoustic and the auditory phonetic realisation of H*+L. Firstly, the languages differed in peak alignment. In German, the F0 peak reflecting H*+L was invariably aligned with the right edge of the stressed syllable, but in English, the peak was aligned within the stressed syllable. Secondly, the languages differed in the accommodation of H*+L on syllables with a small proportion of sonorants. In German, H*+L is truncated, but in English, it is compressed. The third cross-linguistic difference involved the acoustic phonetic implementation of downstep. In German, the peak of an IP-final !H*+L can be stepped down to the level of the L, but in English, !H*+L always involves falling pitch and F0.

Chapters 5 and 6 presented experimental investigations of two hypotheses which had emerged from the corpus analysis. Chapter 5 further investigated the pitch accent accommodation effects *truncation* and *compression*. The data confirmed the hypothesised cross-linguistic difference. When sonorant segmental material is scarce, H*+L is compressed in English, but truncated in German. L*+H, however, is compressed in both languages. Two accounts of the asymmetry in the German results were discussed. The first suggested that in German, realisations of L*+H were compressed because they were followed by a high boundary tone H%. H*+L, on the other hand, was not followed by a tonal specification and could therefore be truncated. The second account suggested that truncation and compression apply to the nuclear pitch accent rather than to the nuclear tone, and that HL sequences truncate whereas LH sequences compress.

A follow-up experiment on German provided experimental support for the view that accent accommodation effects involve the nuclear accent rather than the nuclear tone (i.e. the nuclear accent plus the following boundary tone). Were it the case that L*+H accents compress, regardless of whether a following boundary specification was H% or 0%, then the nuclear accent hypothesis would be favoured. The experimental data supported this view. Both L*+H H% and L*+H 0% were found to compress. Additionally, the data showed that in the realisation of the experimental materials, the choice of the pitch accent L*+H was conditioned by the context, but that of the following boundary specification was not; the presence or absence of H% appeared to depend on the speaker. However, speakers did not mix nuclear tones L*+H H% and L*+H 0% within one list of coordinated intonation phrases. Once the first nuclear tone in a list had been chosen, all following members of the list were produced with that same nuclear tone. This finding suggests that speakers choose H% or 0% at the level of the complete coordination structure rather than at the level of individual intonation phrases within that structure.

---

particular elements in the tune is language-specific, and that in English, certain adjustments happen to be restricted to prenuclear accents.

Chapter 6 investigated downtrends in English and German. An experimental comparison of pitch and F0 patterns in downstepping lists confirmed the hypothesis that German has *final peak lowering* whereas English does not. Secondly, the data suggested that 'final lowering', an effect claimed to characterise downstepped sequences in American English (Liberman and Pierrehumbert, 1984), appeared to be absent in British English. The discrepancy between the results presented in Chapter 6 of the present study and earlier findings in the literature was hypothesised to be the result of declination. A follow-up experiment was carried out which lent support to this hypothesis. Apparently, in British English, but probably also in American English, downtrends in F0 need to be modelled with a combination of downstep and declination (this also the case in Japanese, Pierrehumbert and Beckman, 1988).

2.1 Scope

The findings of the present study suggest a number of topics for further research. Firstly, the present study has compared two standard varieties of English and German. Clearly, the findings cannot be generalised to every variety of English and German because the standard varieties are not representative, only exemplary. Similar comparative analyses of dialects within English and German are called for.

Secondly, the speech data analysed in the present study were restricted to one speaking style. The way in which the findings presented relate to the intonation of other speaking styles needs to be investigated. In spontaneous speech, for instance, it is likely that phonological adjustments such as DELETION have a higher frequency of occurrence than in the corpus investigated here. More generally, an investigation of the conditions under which phonological adjustments apply, as well as further detailed auditory and acoustic data on their nature is required.

Thirdly, the speech analysed was produced by female speakers aged between 16 and 22. Potential intonational variation in male and female speech or variation related to age lay outside the scope of this study (see Cruttenden, 1986: 134 for an overview of some studies on intonational variation).

Fourthly, the realisational differences between English and German may be investigated further. For instance, in Chapter 5, the question of whether truncation in German is phonetic or phonological was discussed. There, it was argued that truncation is likely to be phonetic, because the experimental evidence suggested that the process was gradient. However, the fact remained that on words whose only sonorant segment was a short vowel, frequently, truncated H*+L did not exhibit any evidence of a fall in F0. Rather, traces were level or rising-falling (within a very restricted F0 range). On words with a higher proportion of sonorants, on the other hand, a fall in F0 was invariably apparent (but this fall had a smaller F0 excursion than that observed, for instance, on bisyllabic words). This finding may suggest that on very short syllables, truncation is either phonologised or in the process of being phonologised (a small fall in F0 was observed at times). However, at present, this can only be a hypothesis and further research is needed. Additionally, perception experiments investigating the auditory impression of truncation by native speakers and non-native speakers would be expedient, and the physiological basis of truncation needs to be studied. Such investigations may shed more light on the strategies speakers employ when realising tones.

3 Implications

The findings presented in this study have methodological implications for cross-linguistic work on intonation and theoretical implications for current autosegmental-metrical models of intonational structure. The methodological implications will be discussed first.

3.1 Methodological implications

The present study differs from previous comparative studies on intonation in that it is based on samples of speech directly comparable within and across languages. Care was taken to choose subjects matched within language with respect to age, education and language background, and the possible interpretations of the materials were limited. Moreover, the corpus contained read, rather than spontaneous speech. Read speech was argued to provide a better starting point for a first autosegmental cross-linguistic comparison, because it allows a more constrained elicitation of intonation patterns, and intonation phrase boundaries may be determined with a higher degree of certainty. Analysing utterances produced by five comparable speakers means that the findings of the study can be generalised to a group of speakers and that individual findings have been replicated.

The results of the corpus analysis were illustrated with both auditory impressions and fundamental frequency traces. The acoustic data was made available in two ways. Firstly, in each trace, the location of the stressed syllable was indicated. This allowed detailed comparisons of the alignment of fundamental frequency with segmental material. Secondly, each pattern produced in a specific context was contrasted with other patterns produced in the same context and with similar patterns produced in different contexts. These comparisons provided 'paradigmatic', cross-speaker information about the representative status of a contour and its alignment with segmental structure, and 'syntagmatic' comparisons of auditorily equivalent F0 contours on different words (as there were five speakers, and there were always five instances of specific pattern).

This method brings a number of benefits. The comparison of intonation patterns produced by different speakers in identical contexts helped to establish the language-specific characteristics of contours and to reveal, for instance, a cross-linguistic difference in peak alignment. The comparison of fundamental frequency traces of auditorily equivalent contours on different lexical material produced evidence of pitch accent realisation effects such as truncation and of the cross-linguistic difference between truncation and compression. Additionally, the approach allows a comparison of the choices different speakers make in identical contexts. For instance, some speakers produced patterns analysed as H*+L with DELETION in a context in which other speakers produced H*+L without DELETION. Thus, in identical contexts, different speakers produced patterns which were closely related but systematically different. This finding lent support to an account of intonation which separates the phonological component into an underlying and a surface level. If different speakers produce structurally related but systematically different contours in identical contexts, then this suggests that the contours do not reflect unrelated choices from the phonological inventory, but are derived from the same primitives.

Finally, the combination of a corpus study generating hypotheses about cross-linguistic similarities and differences, and controlled experiments further exploring these hypotheses has been fruitful. The data presented in Chapters 5 and 6 have provided statistical support for cross-linguistic differences in accent accommodation and the implementation of downstep (*final peak lowering*). Additionally, these data showed that both truncation and final peak lowering are gradient. These findings could not have emerged from a corpus study alone, and they suggest that neither truncation nor final peak lowering should be accounted for as part of the phonological system. In the German ToBI inventory, however, which is not backed up by experimental evidence, !H*+L with final peak lowering is accounted for as phonologically different from !H*+L without final peak lowering.

The findings of the second downstep experiment confirmed the value of an approach to tonal analysis based on replicable experimental evidence. The results of a study using materials modelled closely on those used by previous investigators suggested an alternative explanation for a finding which appears to have been generally accepted as part of the tonal implementation of downstep in English, namely, that downstepped sequences have 'final lowering'. The data presented in Chapter 6 of the present study suggested that the apparent final lowering effect can be accounted for as the result of declination.

3.2 Implications for autosegmental-metrical theory

The results of the present study have implications for autosegmental-metrical theory. First of all, they provide evidence for Ladd's (1996) taxonomy of cross-linguistic differences in intonation; they confirm that languages may be similar at one level of representation and different at another. Not only does this finding show that an autosegmental-metrical approach to comparative intonation analysis is preferable to a unilinear approach, but it also explains at least some of the apparent confusion in previous unilinear comparisons of English and German intonation. If cross-linguistic similarities and differences at different levels of representation cannot be separated from each other, seemingly contradictory data are likely to emerge. One particular finding of the present study suggests that one may even need to go beyond Ladd's proposal, and assume a further level of representation at which the intonation of languages may differ. Apparently, German native speakers hear truncated falls as involving falling pitch, but English native speakers consulted informally were much less confident that they heard falling pitch. Thus, for German listeners , acoustic truncation in German does not appear to equal auditory phonetic truncation, but for English listeners, it does. Clearly, this finding requires experimental verification, but if it can be replicated, then we need to consider the possibility of cross-linguistic differences at an *acoustic realisational* as well as an *auditory realisational* level.

If we accept that cross-linguistic or monolinguistic accounts of intonation can be adequate only if they take into account different levels of representation, then it follows that analysts are required to motivate very clearly their decisions to assign a particular distinction to a particular level in their analysis. Often, a complete picture may appear only if a specific account has been verified by more than one kind of analytic technique or experimental procedure. For instance, some of the distinctions which are assigned to the phonological level in the ToBI inventory would benefit from further testing.

Not only must the decision to assign a particular distinction to a particular level in the analysis be motivated, but there must also be an explicit set of principles for mapping between the levels (see Ladd, 1996: Chapter 4). Cross-linguistic and cross-dialectal work on intonation requires that autosegmental-metrical accounts should be comparable, and accounts can only be comparable if constraints on mapping are stated more clearly than they frequently are in monolingual studies . In the ToBI system, for instance, which was developed on the basis of General American English, a low tone is realised as a local minimum in the F0 trace when following H*, but has no effect on the phonetic realisation after H-. Mapping a particular tone onto a number of very different surface realisations is not necessarily a problem in monolingual accounts of intonation. In cross-linguistic work, however, the limitations of such an approach are quickly reached. In General American English, we find two tonal options at the phrase boundary; in the absence of a stressed syllable, pitch may either rise or remain level. Falls in pitch at the boundary are absent. The absence of falls means that a boundary rise, for instance after

L*, can be transcribed H-H% and a boundary level as H-L%. An upstep rule explains why in H-H%, the H% is raised above the level of H-, and in H-L%, the L% is realised at the same level as the preceding H-. This approach does not work, however, when we use ToBI to compare General American English and Northern Irish English. Northern Irish English appears to have three boundary options; rising, level and falling. The upstep rule triggered by H- prevents us from transcribing the boundary fall as H-L%. If we attempt to solve the problem by assuming that upstep is a speaker choice, then the transcription H-L% arbitrarily covers two of the three boundary options available in Northern Irish English. Thus, as far as standard varieties of English are concerned, the phonetics-phonology mapping posited in ToBI may be no more than cumbersome; but when it comes to cross-linguistic and cross-varietal comparisons, the system runs into serious problems.

In a nutshell, the present study has provided evidence for cross-linguistic differences in intonation at different levels of representation. English and German intonation are very similar at the phonological level of representation but differ at the acoustic phonetic level. This finding shows that cross-linguistic comparisons of intonation cannot be restricted to a single level of representation.

References

Abercrombie, D. (1964). Syllable quantity and enclitics in English. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott and J. L. M. Trim (eds). *In honour of Daniel Jones*. London: Longman, 216-22.

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Anderson, A. H., Bader, M., Bard, E. G., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H., and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech* 34, 4, 351-66.

Anderson, K. O. (1979). On the contrastive phonetics of English and German intonation. Festschrift für Otto von Essen anlässlich seines 80. Geburtstages. In H-H. Wängler (ed.), *Hamburger Phonetische Beiträge* 25, 25-35.

Antonsen, E. H. (1966). Suprasegmentals in German. *Language* 42, 587-601.

Arvaniti, A., Ladd, D. R. and Mennen, I. (1996). What is a starred tone? Evidence from Greek. Paper presented at Laboratory Phonology V, Evanston, IL, Northwestern University.

Bald, W. (1976). Contrastive Studies in English and German intonation. *Papers and Studies in Contrastive Linguistics* 4, 37-47. Poznán.

Bannert, R. and Bredvad, A. (1975). Temporal organisation of Swedish tonal accent: the effect of vowel duration. *Working papers* 10, Phonetics laboratory, Department of General Linguistics, Lund University, Sweden.

Barker, M. L. (1925). *A Handbook of German Intonation*. Cambridge: Heffer.

Batliner, A., and Reyelt, M. (1994). Ein Inventar prosodischer Etiketten für VERBMOBIL. Verbmobil Memo 33, 1994.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: evidence for cue trading. *Journal of Memory and Language* 30, 644-63.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, M. and Ayers, G. (1994). Guidelines for ToBI labeling, version 2.0. Linguistics Department, Ohio State University.

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook* 3, 255-310.

Beckman, M. and Pierrehumbert, J. (1992). Comments on chapters 14 and 15. In G.J. Docherty and D.R. Ladd (ed.s.).*Papers in Laboratory Phonology II. Gesture, Segment, Prosody*, 387-97. Cambridge: CUP.

Beckman, M. E. (1995). Problems of intonation. In *Proceedings of theXIIIth International Congress of Phonetic Sciences*, 1, 450-7. Stockholm, Sweden.

Beckman, M. E. and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (ed.) *Phonological structure and phonetic form. Papers in Laboratory Phonology III*. Cambridge: CUP.

Berg, R. van den, C. Gussenhoven and T. Rietveld (1992). Downstep in Dutch: Implications for a model. In G. J. Docherty and D. R. Ladd (eds.) *Papers in Laboratory Phonology II. Gesture, Segment, Prosody*, 335-58. Cambridge: CUP.

Bierwisch (1971). Regeln für die Intonation deutscher Sätze. *Studia Grammatica* 7, 99-201.

Bing, J. (1979). Aspects of English prosody. Unpublished doctoral dissertation, University of Massachusetts.

Bolinger, D. (1945) The minimising downskip. *American Speech* 20, 40-4.

Bolinger, D. (1951). Intonation: levels versus configurations. *Word* 7, 199-210.

Bolinger, D. (1958). A theory of pitch accent in English. *Word* 14, 109-49.

Bolinger, D. (1964). Intonation: around the edge of language. Harvard Educational review 34, 282-96.

Bolinger, D. (1986). *Intonation and its parts.* Palo Alto, CA: Stanford University Press.

Bolinger, D. L. (1972a). *Intonation: Selected. Readings.* Harmondsworth: Penguin.

Bruce, G. (1977).*Swed.ish word accents in sentence perspective.* Lund: Gleerup.

Cabrera-Abreu, M. (1994). An alternative model of intonation applied. to English. Speech, Hearing and Language: Work in Progress, 8, University College London.

Cohen, A. and 't Hart, J. (1967). Declination: construct or intrinsic feature of speech pitch. *Phonetica* 39, 254-73.

Cohen, A., and 't Hart, J. (1967). On the anatomy of intonation. *Lingua* 19, 177-92.

Cooper, W.E. and Paccia-Cooper, J. (1980). *Syntax and Speech.* Cambridge, MA: Harvard University Press.

Couper-Kuhlen, E. (1993). *English speech rhythm: form andfunction in everyday verbal interaction.* Amsterdam: J. Benjamins,

Cruttenden, A. (1986). *Intonation.* Cambridge: CUP.

Crystal, D. (1969) *Prosodic systems and intonation in English.* Cambridge: CUP.

Crystal, D. (1997). *A Dictionary of Linguistics and Phonetics.* 4th edition, Oxford: Blackwell.

Crystal, D. and Quirk, R. (1964). *Systems of prosodic and paralinguistic features*feet. and

Cutler, A. and Ladd, D.R. (eds.) (1983). *Prosody: models and measurements.* Heidelberg: Springer.

Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French.* Heidelberg: Julius Groos.

Delattre, P., Poenack, E. and Olsen, C. (1965). Some characteristics of German intonation for the expression of continuity and finality. *Phonetica* 13, 134-61.

Erickson, D., Honda, K., Hirai, H. and Beckman, M. E. (1995). The production of low tones in English intonation. *Journal of Phonetics* 23, 179-88.

Erikson, Y. and Alstermark, M. (1972). Fundamental Frequency correlates of the grave word accent in Swedish: the effect of vowel duration. *Speech Transmission Laboratory. Quarterly Progress and Status Report* 2-3, KTH, Sweden.

Essen, O. v. (1964). *Grundzüge der hochdeutschen Satzintonation.* Ratingen: Düsseldorf.

Esser, J. (1978). Contrastive intonation of German and English. *Phonetica* 35, 41-55.

Fear, B., Cutler, A., and Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America* 97, 3, 1893-904.

Féry, C. (1993). *German intonational patterns.* Tübingen: Niemeyer.

Fischer-Jørgensen, E. and Hutters, B. (1981). Aspirated stop consonants before low vowels, a problem of delimitation, its causes and consequences. *ARIPUC* 15, 77-102.

Fox, A. (1982). Remarks on intonation and 'Ausrahmung' in German. *Journal of Linguistics* 18, 1: 89-106.

Fox, A. (1984). *A comparative study of English and German intonation.* Oxford: Clarendon Press.

Fry, D. B. (1958). Experiments in the perception of stress. Language and Speech 1, 126-52.

Gibbon, D. (1976). *Perspectives of intonation analysis.* Forum Linguisticum 9, Frankfurt am Main: Lang.

Gibbon, D. (1984). Metalocutions, structural types and functional variation in English and German. *Papers and Studies in Contrastive Linguistics* 13, 17-39.

Grabe, E. (1997). Comparative intonational phonology: English and German. In A. Botinis, G. Kouroupetroglou and G. Carayannis (eds.) *Proceedings of the ESCA Tutorial and Research workshop on Intonation: Theory, Models and Applications.* Athens, Greece.

Grabe, E. and Warren, P. (1995). Stress Shift: do speakers do it or do listeners use it? In B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic evidence. Papers in Laboratory Phonology IV.* Cambridge: CUP.

Grabe, E., Gussenhoven, C., Haan, J., Marsi, E., and Post, B. (1997). The meaning of intonation phrase onsets in Dutch. In *Proceedings of the ESCA Tutorial and Research workshop on Intonation: Theory, Models and Applications.* Athens, Greece.

Grice, M. (1995a). *The intonation of interrogation in Palermo Italian: implications for intonation theory.* Tübingen: Niemeyer.

Grice, M. (1995b). Leading tones and downstep in English. *Phonology* 12, 183-233.

Grice, M. and Benzmüller, R. (1995). Transcription of German intonation using ToBI-tones. The Saarbrücken System. Phonus 1, Research Report, Institute of Phonetics, University of the Saarland.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., and Batliner, A. (1996). Consistency in transcription and labelling of German intonation with ToBI. In *Proceedings of the Fourth International Conference on Spoken Language Processing*, 3, 1716 - 19. Philadelphia, USA.

Grønnum, N. (1989). Stress group patterns, sentence accents and sentence intonation in Southern Jutland (Sønderborg and Tønder) - with a view to German. *ARIPUC* 23, 1-85.

Grønnum, N. (1992). Comments on chapter 14. In G. J. Docherty and D. R. Ladd (eds) *Papers in Laboratory Phonology II. Gesture, Segment, Prosody*, 359-67. Cambridge: CUP.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents.* Dordrecht: Foris.

Gussenhoven, C. (1988). Adequacy in intonation analysis. In H. van der Hulst and N. Smith (eds.) *Autosegmental studies on pitch accent.* Dordrecht: Foris.

Gussenhoven, C. (1990). Tonal association domains and the prosodic hierarchy in English. In S. Ramsaran (ed.) *Studies in the pronunciation of English.* London, New York: Routledge.

Gussenhoven, C. (1991). Tone segments in the intonation of Dutch. In T. F. Shannon and J. P. Snapper (eds.) *The Berkeley Conference on Dutch Linguistics 1989*, 139-55. Lanham (MD): University Press of America.

Gussenhoven, C. (1992). Sentence accents and argument structure. In I. M. Roca (ed.) *Thematic structure: its role in Grammar.* Berlin: Foris.

Gussenhoven, C. and Rietveld, T. (1991). An experimental evaluation of two nuclear-tone taxonomies. *Linguistics*, 29: 423-49.

Gussenhoven, C. (1994). Review of Féry (1993). *Linguistics* 32. 555-8.

Gut, U. (1995). The intonation of German learners of British English. Unpublished M.Phil dissertation, University of Cambridge.

Halliday, M. A. K. (1967). *Intonation and grammar in British English.* The Hague: Mouton.

Hart, J. 't, Collier R., and Cohen, A. (1990). *A perceptual study of intonation.* Cambridge: CUP.

Healy, M. J. R (1988). *GLIM, an introduction*. Oxford: Clarendon Press.

House, J. (1995). Intonational stereotype: a reanalysis. In J. Windsor-Lewis (ed.) *Essays in Honour of Professor J. D. O'Connor*. London and New York: Routledge.

Isačenco, A. v. and Schädlich, H. (1966). Untersuchungen über die deutsche Satzintonation. *Studia Grammatica* 7. Berlin: Akademie-Verlag.

Jin, F. (1990). *Intonation in Gesprächen: ein Beitrag zur Methode der kontrastiven Intonationsanalyse am Beispiel des Deutschen und Französischen*. Tübingen: Niemeyer.

Kingdon, R. (1958). *The Groundworks of English Intonation*. London: Longman.

Klinghardt, H. (1920). *Übungen im Englischen Tonfall*. Cöthen, Leipzig: Quelle und Meyer.

Klinghardt, H. (1923). *Sprechmelodie und Sprechtakt*. Marburg: Elwert.

Klinghardt, H. and Klemm, G. (1920). *Übungen im englischen Tonfall für Lehrer und Studierende*. Cöthen: Otto Schulze Verlag.

Kohler, K. (1983). F0 in speech timing. In K. Kohler and C. Hoequist (eds.) *Studies in Speech timing*. Arbeitsberichte 20 (AIPUK), Institut für Phonetik, Universität Kiel.

Kohler, K. (1991). Studies in German intonation. Arbeitsberichte 25 (AIPUK), Institut für Phonetik, Universität Kiel.

Kohler, K. (1995). ToBIG and PROLAB - Two prosodic transcription systems for German compared. Workshop on Prosodic Labelling held at the 13th International Congress of Phonetic Sciences, Stockholm, Sweden. Unpublished manuscript, University of Kiel.

Kohler, K. J. (1987). Categorical pitch perception. In *Proceedings of the XIth International Congress of Phonetic Sciences*, 5, 331-3.

Kuhlmann, W. (1952). Vergleich deutscher und englischer Tonhöhenbewegung. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 6, 195-207.

Kurath, H. (1964). *A Phonology and Prosody of Modern English*. Heidelberg: Winter.

Ladd, D. R. (1978). Stylised intonation. *Language* 54, 3, 517- 40.

Ladd, D. R. (1983a). Phonological features of intonational peaks. *Language*, 59, 721-59.

Ladd, D. R. (1984). Declination: a review and some hypotheses. *Phonology Yearbook* 1, 53-74.

Ladd, D. R. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3: 311-40.

Ladd, D. R. (1990b). Metrical representation of pitch register. In J. Kingston and M. E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: CUP.

Ladd, D. R. (1992). Compound prosodic domains. Linguistics Department Occasional paper. Department of Linguistics, University of Edinburgh.

Ladd, D. R. (1993a). In defense of a metrical theory of intonational downstep. In H. v. d. Hulst and K. Snider (eds.) *The representation of tonal register*. Dordrecht: Foris.

Ladd, D. R. (1994). Review of Fery (1993). *Linguistische Berichte* 154.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: CUP.

Langley, J. (1992). *Little Red Riding Hood*. London: HarperCollins.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: CUP.

Leben, W. R. (1976). The tones in English Intonation. *Linguistic Analysis* 2, 69-107.

Liberman, M. (1975). The intonational system of English. Unpublished doctoral dissertation, Cambridge, MA: MIT.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry* 8, 249-336.

Liberman, M.Y. and Pierrehumbert, J. B. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff and R.T. Oehrle(eds.) *Language Sound Structure: Studies in Phonology Presented. to Morris Halle*, 157-233. Cambridge, MA: MIT Press.

Lindsey, G. (1985). Intonation and interrogation: tonal structure and the expression of a pragmatic function in English and other language. Unpublished doctoral dissertation, University of California.

Low, E. (forthcoming). Prosodic prominence in Singapore English. Doctoral dissertation, University of Cambridge.

Low, E. & Grabe, E. (1995). Prosodic patterns in Singapore English. *Proceedings of the Intonational Congress of Phonetic Sciences*, Stockholm 13-19 August: Vol. 3: 636-639.

Lowry, O. (1997). Intonation patterns in Northern Irish English. Unpublished MPhil dissertation, University of Cambridge.

Mayer, J. (1995). Transcription of German intonation. The Stuttgardt system. Unpublished manuscript, University of Stuttgardt.

Möbius, B. (1993). *Ein quantitatives Modell der deutschen Intonation.* Tübingen, Niemeyer.

Moulton, W. G. (1962). *The sounds of English and German.* Chicago, University of Chicago Press.

Nolan, F. (1984). Auditory and instrumental analysis of intonation. Cambridge Papers in Phonetics and Linguistics 4, 1-21.

Nolan, F. (1990). Who do Phoneticians represent? *Journal of Phonetics* 18, 453-64

Nolan, F. (1995). The effect of emphasis on declination in English intonation. In J. Windsor-Lewis (ed.) *Essays in Honour of Professor J.D. O'Connor.* London and New York: Routledge.

Nolan, F. and Grabe, E. (1997). Can ToBi transcribe intonational variation in the British Isles? In A. Botinis, G. Kouroupetroglou and G. Carayannis (eds.) *Proceedings of the ESCA Tutorial and Research workshop on Intonation: Theory, Models and Applications*. Athens, Greece.

Nooteboom, S.G. and Kruyt, J.G. (1987). Accent, focus distribution and the perceived distribution of given and new information: an experiment. *Journal of the Acoustical Society of America* 82, 1512-24.

O'Connor, J. D. and Arnold, G.F. (1973). *Intonation of colloquial English.* London: Longman.

Palmer, H.E. (1922). *English intonation, with systematic exercises*. Cambridge: Heffer.

Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan and M. E. Pollack (eds.) *Intentions in Communication*. Cambridge, MA: MIT Press.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, Cambridge, MA: MIT.

Pierrehumbert, J. B. And Beckman, M. E. (1988). *Japanese Tone Structure*. Linguistic Inquiry Monograph 15. Cambridge, MA: The MIT Press.

Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pitrelli, J., Beckman, M. E, and Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the Third International Conference on Spoken Language Processing*, Yokohama, Japan.

Poser, W. J. (1984). The phonetics and phonology of tone and intonation in Japanese. Unpublished doctoral dissertation, Cambridge, MA: MIT.

Prevost, S. and Steedman , M. (1994). Specifying intonation from context for speech synthesis. *Speech Communication* 15, 139-53.

Pürschel, H. (1975). *Pause und Kadenz. Interferenzerscheinungen bei der englischen Intonation deutscher Sprecher*. Tübingen: Niemeyer.

Rietveld, T. and Gussenhoven, C, (1995). Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics* 23, 375-85.

Roach, P. (1994). Conversion between prosodic transcription systems: 'Standard British and ToBI'. *Speech Communication* 15, 91-9.

Ruiter, J. de and Wilkins, D. (eds.) Annual Report 1996, Max Planck Institute for Psycholinguistics, 13.

Santen, J. P. H. van and Hirschberg, J. (1994). Segmental effects on timing and height of pitch contours. *In Proceedings of the International Conference on Spoken Language Processing* 94, 719-22.

Schubiger, M. (1935). *The role of intonation in Spoken English.* Cambridge, Heffer.

Schubiger, M. (1953). Notes on intonation of co-ordinate sentence and syntactic groups. *English Studies* 34: 268-73.

Schubiger, M. (1958). *English Intonation: its Form and Function.* Tübingen: Niemeyer.

Schubiger, M. (1965b). English intonation and German Modal Particles: a Comparative Study. *Phonetica* 12, 65-84.

Scott, D. (1982). Duration as a cues to the perception of a phrase boundary. *JASA* 71, 4: 996-1007.

Scuffil, M. (1982). *Eperiments in Comparative Intonation.* Tübingen: Niemeyer.

Selkirk, E. O. (1982). *The syntax of words.* Linguistic Inquiry Monograph 7. Cambridge, MA: The MIT Press.

Silverman, K. E. A. and Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston and M. E. Beckman, *Papers in Laboratory Phonology I* , Cambridge: CUP.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). ToBI: a standard for labeling English prosody. In *Proceedings of the Second International Conference on Spoken Language Processing*, 2: 867 - 70. Banff, Canada.

Steele, J. (1986) Nuclear accent F0 peak location: effect of rate, vowel, and number of syllables. *JASA* supplement 1, 80, s51.

Thorsen, N. (1980a). A study of the perception of sentence intonation: evidence from Danish. *JASA* 67, 1014-30.

Thorsen, N. (1980b). Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. *ARIPUC* 14, 1-29.

Thorsen, N. (1981). Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish: supplementary data. *ARIPUC* 15, 13-47.

Thorsen, N. (1983). Two issues in the prosody of standard Danish. In A. Cutler and D. R. Ladd (eds.) *Prosody: models and measurements*. Heidelberg: Springer.

Trager, G. L. and Smith, H. L. (1951). *An outline of English structure*. Norman, OK: Battenburg Press.

Trim, J. L. M (1959). Major and minor tone-groups in English. *Le Maître Phonetique,* 112, 26-9.

Trim, J. L. M. (1964). Tonetic stress marks for German. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott and J. L. M. Trim (eds). *In Honour of Daniel Jones*, London: Longman.

Trim, J. L. M. (1988). Some contrastive intonated features of British English and German. In J. Klegraf and D. Nehls (eds.) *Essays on the English language and applied linguistics on the occasion of Gerhard Nickel's 60th Birthday*. Heidelberg: Julius Groos, 235-49.

Uhmann, S. (1991). *Fokusphonologie: eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Tübingen: Niemeyer.

Wells, J. C. (1982). *Accents of English*. Cambridge: CUP.

Wittig, K. (1956). *Die Phonetik des Amerikanischen Englisch*. Heidelberg: Winter.

Wunderlich, D. (1988). Der Ton macht die Melodie - Zur Phonologie der Intonation des Deutschen. In H. Altman (ed.) *Intonationsforschungen*. Tübingen: Niemeyer, 1-40.

Appendix A - Corpus materials

1 English

Little Red Riding Hood

Once upon a time on the edge of the big forest lived a little girl whom everyone called "Little Red Riding Hood". This was because she always wore her favourite red bonnet, wherever she went. But her real name was Anna. One morning Anna's mum said, "Little Red Riding Hood, your grandma is unwell, and she can't go shopping. I'm packing up some food, and I'd like you to take it to her." She filled a bag with some ripe yellow bananas, a jar of orange marmalade, a tub of margarine and some wholemeal rolls. She wondered about putting in some plums but she had two kinds and wasn't sure which kind Grandma liked. So she asked Anna, "Is it red plums or yellow plums grandma prefers? "It's yellow plums she normally buys, replied the girl. Soon the bag was full.

Then her mother said "Listen to me, Anna, be careful in the forest!" "Go straight to Grandma's house. Don't lose your way. Don't talk to any strangers." "I'll be very careful," promised the girl, "and I won't stop for a minute". So off Anna went with her bag full of goodies and waved to her mother until she was well out of sight.

When Little Red Riding Hood got to the edge of the forest, a stranger was waiting for her, unexpectedly. The stranger called in a deep and silky voice,
"Little girl, little girl, can you spare a minute?" It was a wolf, oh dear! But Little Red Riding Hood had never seen a wolf before, so she wasn't scared. The wolf sauntered towards her and greeted her with a lazy smile. "Good morning, my dear," he said. "I'm the wolf and who are you?" "Good morning," the girl said politely. "My name is Anna". "Oho, Anna!", said the wolf, "and where are you going, Anna?" "I'm going to see my grandma," the girl replied. "She's at home, in bed, she's unwell." "How kind," said the wolf. And he thought to himself, "What a stroke of luck! A helpless little girl and a delicious grandma, and I haven't had any lunch yet! I will distract the girl so I can go and eat her grandma first. And I'll eat the girl later. But how can I distract Anna? Maybe it's flowers Anna likes." So he said to Little Red Riding Hood, "as you're such a good girl, maybe you'd like to take your grandma some flowers. I'll show you a place where you can find them. You will find Camomile, Roses and Elderflower." The girl, remembering her mother's words, knew that she had promised not to stop, but she liked the idea of bringing grandma flowers. So she followed the wolf, very cautiously. "Here we are, little girl" he said. "Oh, by the way, where does your grandma live?" "She lives at the far edge of the wood, next to the pond," Little Red Riding Hood replied, innocently. "How nice,"

the wolf said, "but now I must fly, I'm late for my lunch." And off he raced to grandma's house.

A little later, when Little Red Riding Hood reached the house, she was a little bit surprised to see that the door was open. "Grandma, where are you?" she called, loudly. "Over here, in the bedroom," answered a deep voice. Little Red Riding Hood went into the bedroom and up to her grandma's bed. When she got closer, she noticed something very strange. "Oh grandma!" she cried, "what big ears you have!" "All the better to hear you with, my dear," came the reply. Little Red Riding Hood went up a little closer. "Oh grandma, what big eyes you have!" she cried. "All the better to see you with, my dear," was the answer. Little Red Riding Hood took one more step. "Oh grandma, what big teeth you have!" At that moment the wolf leapt out of the bed and growled, "all the better to eat you with, my dear," and he grabbed her and tried to gobble her up. At the last minute, the door burst open and there stood mum with Grandma's frying pan in her hand. She had felt worried about Anna and had followed her to Grandma's house, and when she got there, she had realised that there was trouble. Mum raised up the frying pan and banged it on the wolf's head, as hard as she could. He did not move again.

Anna's Mum carefully cut open the wolf, and out jumped grandma. She was fine, happily. Then mum said, "I will teach that wolf a lesson." She filled his tummy with big heavy stones and sewed up his woolly belly. Then she opened the door and rolled him outside.

When the wolf woke up, he felt terrible! His head hurt, his tummy was swollen and very heavy. "Ooooooh," he mumbled, "I'll never eat another Grandma again." He never did and he never talked to strange girls again, either.

2 German

Rotkäppchen

Es war einmal am Rande des grossen Waldes ein kleines Mädchen, das alle "Rotkäppchen" nannten. Das war, weil sie ihre rote Lieblingsmütze trug, wo immer sie hinging. Aber in Wirklichkeit hieβ sie Anna. Eines Morgens sagte Annas Mutter: "Rotkäppchen, deine Oma ist nicht gesund, und sie kommt nicht zum Einkaufen. Ich packe ein paar Lebensmittel für sie ein, und ich möchte, daβ du sie ihr bringst". Sie füllte einen Korb mit ein paar reifen gelben Bananen, einem Glas Orangen-marmelade, Margarine und ein paar Vollkornsemmeln. Dann überlegte sie, ob sie ein paar Pflaumen hinein tun sollte, aber sie hatte zwei Sorten und wuβte nicht, welche Oma lieber mochte. "Sind es rote Pflaumen oder gelbe Pflaumen, die Oma lieber mag?" fragte sie Anna. "Es sind die gelben Pflaumen, die sie gern mag" antwortete das Mädchen. Bald war der Korb voll.

Dann sagte die Mutter: "hör mir gut zu, Anna, sei vorsichtig im Wald. Geh sofort zu Omas Haus. Verlauf Dich nicht. Sprich nicht mit Fremden." "Ich werde vorsichtig sein," versprach das Mädchen, "und ich werde nicht trödeln." Dann marschierte Anna mit ihrem Korb voller guter Dinge los, und winkte ihrer Mutter zu, bis sie nicht mehr zu sehen war.

Als Rotkäppchen an den Waldrand kam, wartete ein Fremder auf sie, unerwarteterweise. Der Fremde rief mit einer tiefen und weichen Stimme: "kleines Mädchen, kleines Mädchen, wartest du mal einen Moment?" Es war ein Wolf, oh weh! Aber da Rotkäppchen noch nie zuvor einen Wolf gesehen hatte, hatte sie keine Angst. Der Wolf schlenderte auf sie zu und begrüßte sie mit einem faulen Lächeln. "Guten Morgen, meine Kleine", sagte er, "ich bin der Wolf, und wer bist du?" "Guten Morgen", antwortete das Mädchen höflich. "Mein Name ist Anna". "Aha, Anna!" sagte der Wolf, "und wohin geht's, Anna?" "Ich besuche meine Oma," erwiderte das Mädchen. "Sie ist zu Hause, im Bett, sie ist nicht gesund." "Wie nett," sagte der Wolf. Und er dachte bei sich: "Was für ein Glück! Ein hilfloses Mädchen und eine wohlschmeckende Oma, und ich habe noch nicht zu Mittag gegessen! Ich werde das Mädchen ablenken und dann vorlaufen und ihre Oma zuerst verschlingen. Das Mädchen fresse ich dann später. Aber wie kann ich Anna ablenken?" Vielleicht sind es Blumen, die Anna ablenken." Also sagte er zu Rotkäppchen: "da du so ein nettes Mädchen bist, möchtest Du deiner Oma vielleicht ein paar Blumen mitbringen. Ich zeige Dir einen Ort, wo du welche finden kannst. Dort gibt es Kamille, Rosen und Holunder." Das Mädchen erinnerte sich an die Worte ihrer Mutter, die jedes Verweilen verboten hatten, aber ihr gefiel die Idee Oma Blumen zu bringen. Also folgte sie dem Wolf, sehr vorsichtig. "Da sind wir schon, kleines Mädchen," sagte er. "Oh, ganz nebenbei, wo wohnt deine Oma denn?" "Sie wohnt an der anderen Seite des Waldes, am Teich," antwortete Rotkäppchen unschuldig. "Wie nett", sagte der Wolf, "aber jetzt muß ich schnell weg, ich komme zu spät zum Mittagessen." Und er rannte so schnell er konnte zu Omas Haus.

Etwas später, als Rotkäppchen bei ihrer Oma ankam, sah sie zu ihrem Erstaunen, daß die Tür offen war. "Oma, wo bist Du?" rief sie laut. "Hier drüben, im Schlafzimmer", antwortete eine tiefe Stimme. Rotkäppchen ging ins Schlafzimmer und ans Bett ihrer Oma. Als sie näher kam, bemerkte sie etwas Seltsames. "Aber Oma," rief sie "was für große Ohren Du hast!" "Damit ich dich besser hören kann, meine Kleine," kam die Antwort. Rotkäppchen ging ein wenig näher. "Aber Oma, was für große Augen du hast!" rief sie. "Damit ich dich besser sehen kann, meine Kleine" war die Antwort. Rotkäppchen ging noch etwas näher. "Aber Oma, was für große Zähne du hast!" In diesem Moment sprang der Wolf aus dem Bett und knurrte: "Damit ich dich besser fressen kann, meine Kleine," und er ergriff sie und versuchte sie zu verschlingen. In letzter Sekunde ging die Tür auf und da stand Mutter mit Omas Bratpfanne in der Hand. Sie hatte sich Sorgen um Anna gemacht und war ihr zu Omas Haus gefolgt. Als sie dort ankam, wurde ihr klar, daß etwas nicht stimmte. Mutter hob die Bratpfanne und schlug dem Wolf damit auf den Kopf, so hart sie konnte. Danach rührte er sich nicht mehr.

Annas Mutter schnitt dem Wolf vorsichtig den Bauch auf und Oma sprang heraus. Es ging ihr gut, glücklicherweise. Dann meinte Mutter: "Ich werde diesem Wolf ein Lektion erteilen". Sie füllte seinen Magen mit großen schweren Steinen und nähte seinen Bauch wieder zu. Dann öffnete sie die Tür und rollte ihn nach draußen.

Als der Wolf aufwachte, war ihm sehr übel! Sein Kopf tat ihm weh und sein Bauch war geschwollen und sehr schwer. "Ooooooh" murmelte er, "Ich esse nie wieder eine Oma." Und er tat es auch nicht und sprach auch niemals wieder mit fremden Mädchen.

Appendix B - Experimental Materials

1 Chapter 5 - Accent Accommodation Experiment I

1.1 English

Note that only test sequences are given, not fillers.

To elicit sequences with H*+L H%:

Anna and Peter are watching TV. A photograph of this week's National Lottery winner appears. Anna says: "Look, Peter!

Isn't that Mr. Sheafer? Our new neighbour?"
Isn't that Mr. Shift? Our new neighbour?"
Isn't that Mr. Sheaf? Our new neighbour?"

To elicit sequences with H*+L 0%:

Anna and Peter are watching TV. A photograph of this week's National Lottery winner appears. Anna says: "Look, Peter!"

It's Mr. Sheafer! Our new neighbour!"
It's Mr. Shift! Our new neighbour!"
It's Mr. Sheaf! Our new neighbour!"

1.2 German

Anna und Peter sehen fern. Ein Lottogewinner wird vorgestellt. Anna sagt: "Na sowas!

Das ist doch Herr Schiefer! Unser neuer Nachbar!"
Das ist doch Herr Schiff! Unser neuer Nachbar!"
Das ist doch Herr Schief! Unser neuer Nachbar!"

Anna und Peter sehen fern. Ein Lottogewinner wird vorgestellt. Anna sagt: "Na sowas!

Ist das nicht Herr Schiefer? Unser neuer Nachbar?"
Ist das nicht Herr Schiff? Unser neuer Nachbar?"
Ist das nicht Herr Schief? Unser neuer Nachbar?"

2 Chapter 5 - Accent Accommodation Experiment II

Note that in the materials presented to the subjects, the test sequences did not contain a line break.

(1) *Eines Morgens sagt Rotkäppchens Mutter: "Rotkäppchen, deine Oma ist nicht gesund, und sie kommt nicht zum Einkaufen. Ich packe ein paar Lebensmittel für sie ein und ich möchte daB Du sie ihr bringst". Sie füllte einen Korb mit Brombeermarmelade, Fisch, Maulbeermarmelade und Birnenmarmelade. Rotkäppchen fand die Mischung merkwürdig und fragte: "Will sie wirklich Brombeermarmelade? Und Fisch?"*

(2) *Eines Morgens sagt Rotkäppchens Mutter: "Rotkäppchen, deine Oma ist nicht gesund, und sie kommt nicht zum Einkaufen. Ich packe ein paar Lebensmittel für sie ein und ich möchte daB Du sie ihr bringst". Sie füllte einen Korb mit Brombeermarmelade, Fleisch, Maulbeermarmelade und Birnenmarmelade. Rotkäppchen fand die Mischung merkwürdig und fragte: "Will sie wirklich Maulbeermarmelade ? Und Fleisch?"*

(3) *Eines Morgens sagt Rotkäppchens Mutter: "Rotkäppchen, deine Oma ist nicht gesund, und sie kommt nicht zum Einkaufen. Ich packe ein paar Lebensmittel für sie ein und ich möchte daB Du sie ihr bringst". Sie füllte einen Korb mit Brombeermarmelade, Fleischwurst, Maulbeermarmelade und Birnenmarmelade. Rotkäppchen fand die Mischung merkwürdig und fragte: "Will sie wirklich Birnenmarmelade? Und Fleischwurst?"*

3 Chapter 6 - Downstep Experiment I

3.1 English Production task

In English, you can find quite a lot of compounds starting with the same word. For example:

1. *Mainland, mainstreet, mainstay, mainstream,* ***mainline****.*
2. *Mailbag, mailvan, mail charge, mailman, mail train.*
3. *Earwig, earlobe, earring, eardrum, earphone.*
4. *Green house, green belt, green fly, Greenland, green card.*
5. *Barman, barwork, barmaid, barmeal, barstool.*
6. *Low born, low down, low bred, low brow, low-grade.*
7. *Air gun, air hole, air mail, airplane, airline.*
8. *Eyesore, eyelid, eyestrain, eyeball, eyebrow.*
9. *Moonlight, moonlit, moonbeam, moonshine, moonstone.*
10. *Ground crew, ground fog, ground floor, grounds man, groundswell.*

3.2 German Production task

Im Deutschen gibt es viele zusammengesetzte Worte, die mit dem selben Wort anfangen. Zum Beispiel:

1. *Mondbahn, Mondlicht, mondhell, Mondschein, Mondstein.*
2. *Einhorn, Einfall, einsam, einmal, Einzahl.*
3. *Lehrstuhl, Lehrsaal, Lehrfilm, Lehrplan, Lehrsam.*
4. *Neugier, Neubau, Neumond, Neuwahl, Neujahr.*
5. *Ölscheich, Ölbild, Ölstand, Ölbaum, Ölschlamm.*
6. *Wohngeld, Wohnblock, Wohnort, Wohnraum, Wohnheim.*
7. *Weinglas, Weinfass, Weinrot, Weinfest, Weinfan.*
8. *Blaulicht, Blaufuchs, Blaustrumpf, Blaugrau, Blauhelm*
9. *Weltmarkt, Weltbild, Weltall,Weltraum, Weltmeer.*
10. *Brennglas, Brennpunkt, Brennstoff, Brennholz, Brennball.*

3.3 English completion task

Subjects were asked to complete each sequence they heard with the word printed in bold.

1. *Godparent, godmother, goddaughter, god fearing,* ***godfather****.*
2. *Mainland, mainstreet, mainstay, mainstream,* ***mainline****.*
3. *Mailbag, mailvan, mail charge, mailman,* ***mail train****.*
4. *Gold medal, gold nugget, gold-rimmed, gold plating,* ***gold-digger****.*
5. *Wine drinker, wine cooler, wine cellar, wine lover,* ***wine grower****.*
6. *Barman, barwork, barmaid, barmeal,* ***barstool.***
7. *Eyesore, eyelid, eyestrain, eyeball,* ***eyebrow****.*
8. *Whale hunting, whale watching, whale fishing, whale loving,* ***whale breeding****.*
9. *Road mender, road user, road ranger, road runner,* ***road roller****.*
10. *Moonlight, moonlit, moonbeam, moonshine,* ***moonstone****.*
11. *Dog breeder, dog lover, dog owner, dog handler,* ***dog trainer****.*
12. *Air gun, air hole, air mail, airplane,* ***airline****.*
13. *Earwig, earlobe, earring, eardrum,* ***earphone****.*
14. *Ground crew, ground fog, ground floor, grounds man,* ***groundswell.***
15. *Low born, low down, low bred, low brow,* ***low-grade****.*
16. *Oil-tanker, oil-painting, oil level, oil filter,* ***oil heater****.*
17. *Green house, green belt, green fly, Greenland,* ***green card****.*
18. *Night-clubbing, night-marish, night-watchman, night flyer,* ***night rider****.*
19. *Newsagent, news item, news dealer, news monger,* ***newsreader****.*
20. *Gun barrel, gun powder, gun runner, gun slinger,* ***gun running****.*

3.4 German completion task

1. *Maulbeeren, Maulkörbe, Maultiere, Maultaschen,* ***Maulfäule****.*
2. *Mondbahn, Mondlicht, mondhell, Mondschein,* ***Mondstein****.*
3. *Einhorn, Einfall, Einsam, Einmal,* ***Einzahl.***
4. *Meerenge, Meerwasser, Meerbusen, Meerkatzen,* ***Meerschweinchen****.*
5. *Anlachen, ansagen, anstreichen, anseilen,* ***anmalen***
6. *Ölscheich, Ölbild, Ölstand, Ölbaum,* ***Ölschlamm****.*
7. *Blaulicht, Blaufuchs, Blaustrumpf, Blaugrau,* ***Blauhelm***
8. *Brieftasche, Brieföffner, Briefmarke, Briefwerbung,* ***Brieftaube****.*
9. *Blaustrümpfe, Blaulichter, Blaustifte, Blaumeise,* ***Blausäure****.*

10. *Weltmarkt, Weltbild, Weltall,Weltraum,* ***Weltmeer.***
11. *Mülltonnen, Mülltüten, Müllwagen, Müllmänner,* ***Mülleimer.***
12. *Weinglas, Weinfass, Weinrot, Weinfest,* ***Weinfan.***
13. *Lehrstuhl, Lehrsaal, Lehrfilm, Lehrplan,* ***Lehrsam.***
14. *Brennglas, Brennpunkt, Brennstoff, Brennholz,* ***Brennball.***
15. *Wohngeld, Wohnblock, Wohnort, Wohnraum,* ***Wohnheim.***
16. *Brandwunde, Brandwache, Brandsalbe, Brandmale,* ***Brandmauer.***
17. *Neugier, Neubau, Neumond, Neuwahl,* ***Neujahr.***
18. *Blutgruppe, Blutkörper, Blutsauger, Blutspender,* ***Blutschande.***
19. *Waldmeister, Waldwege, Waldbrände, Waldbeere,* ***Waldwiese.***
20. *Wahlsieger, wahlweise, Wahlrede, Wahlräume,* ***Wahlurne.***

4 Chapter 6 - Downstep experiment II

Long-short list

1. *News dealer, newsreel, news monger, news girl, newsreader.*
2. *Mailing list, mail box, Mail order, mailman, mail transporter.*
3. *Ear warmers, earwig, ear splitting, earmark, ear mufflers.*
4. *Wine drinker, wine glass, wine cooler, wine bar, wine lover.*
5. *Road roller, road show, road user, road hog, road runner.*
6. *Moonlighting, moonlit, moon landing, moonbeam, moonflower.*
7. *Eye surgeon, eyelid, eyewitness, eyeball, eyeliner.*
8. *Oil-painting, oil drum, oil level, oil well, oil filter.*
9. *Night blindness, night-mare, Nightingale, night owl, night rider.*
10. *Nail biting, nail file, nail varnish, nail brush, nail polish.*

Short-long list

1. *Newsreel, news dealer, news girl, news monger, newsboy.*
2. *Mail box, mailing list, mailman, mail transporter, mailvan.*
3. *Earwig, ear warmers, earmark, ear mufflers, earring.*
4. *Wine glass, wine drinker, wine bar, wine lover, wine press.*
5. *Road show, road roller, road hog, road user, road test.*
6. *Moonlit, moonlighting, moonbeam, moon landing, moonstruck.*

7. *Eyelid, eye surgeon, eyeball, eyeliner, eyestrain.*
8. *Oil drum, oil-painting, oil well, oil level, oilcan.*
9. *Night-mare, night blindness, night owl, Nightingale, nightcap.*
10. *Nail file, nail biting, nail brush, nail varnish, nailhead.*

Appendix C

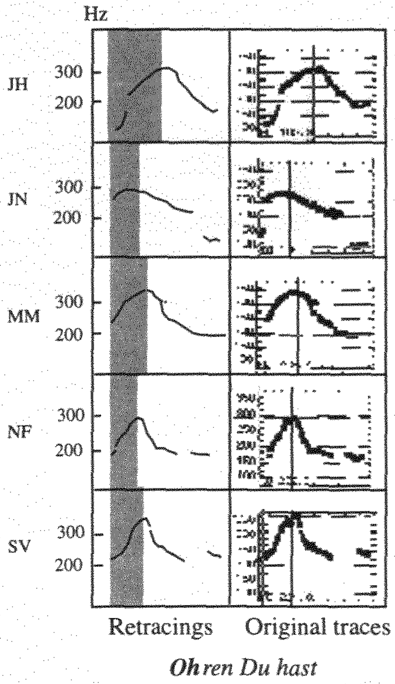


Figure 1 *Comparison of original F0 traces and retracings.*

Appendix D

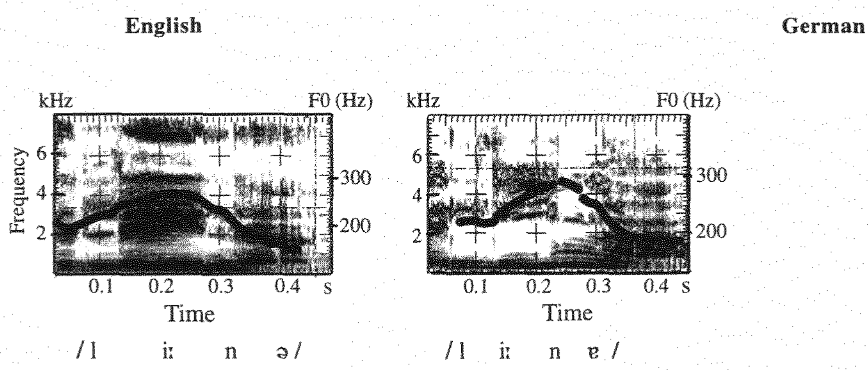


Figure 1 *Peak alignment in H*+L on test words <u>Leaner</u> / <u>Liener</u>.*

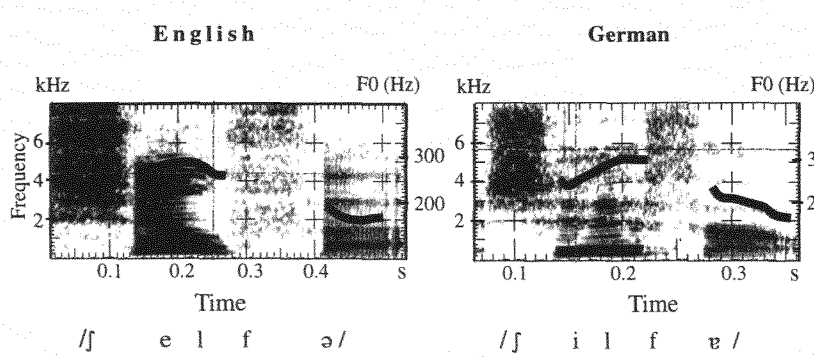


Figure 2 *Peak alignment in H*+L on test words <u>Sheafer</u> / <u>Schiefer</u>.*

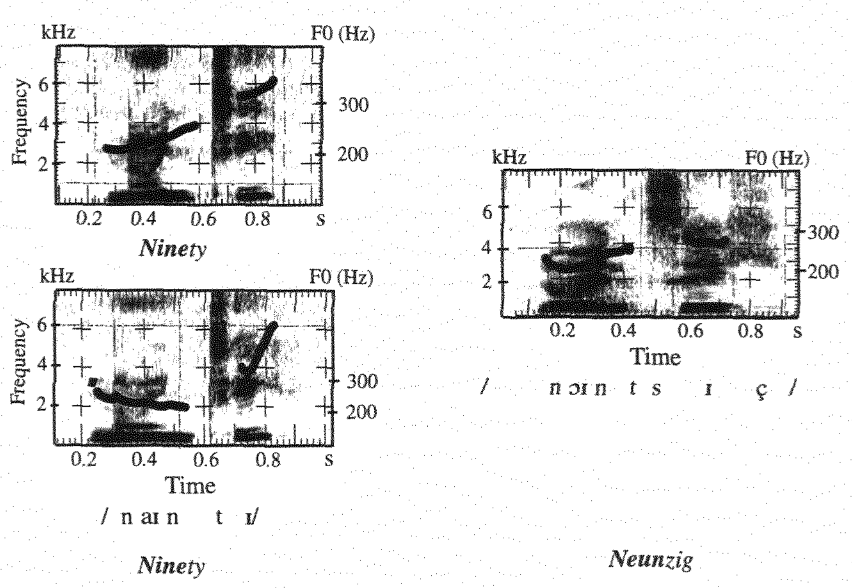


Figure 3 *Two examples of dip alignment in L*+H for English (left) and one example for German (right).*

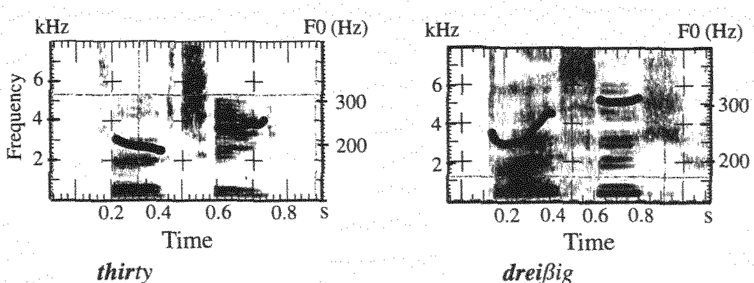


Figure 4 *Dip alignment in L*+H on test words <u>thirty</u> and <u>dreißig</u>.*

Appendix E

	df	ss	ms	F	p <
Line	1	9.944	9.944	11.44	.01
	901	783.41	0.869		
Curve	1	39.56	39.56	45.5	.01
	901	783.41	0.869		
Language	1	121.2	121.2	139.39	.01
	901	783.41	0.869		

Table 1 *ANOVA table for downstep model with factors Line, Curve and Language.*

	df	ss	ms	F	p <
Lang.line	2	12.98	6.49	7.61	.01
	899	766.62	.85		
Lang.Curve	2	39.64	19.82	23.25	.01
	899	766.62	.85		
Language	1	6.117	6.117	7.17	.01
	899	766.62	.85		

Table 2 *ANOVA table for downstep model with interactions Language - Line, Language - Curve and factor Language.*

Samenvatting

Het is een merkwaardige omstandigheid dat er in vergelijkend onderzoek naar de intonatie van het Engels en het Duits geen overeenstemming bestaat over de vraag of de intonatiesystemen van deze talen grote overeenkomsten vertonen of juist fundamenteel van elkaar verschillen. In dit proefschrift wordt een oplossing voor deze controverse voorgesteld. Door in de contrastieve analyse een autosegmenteel-metrische benadering toe te passen op direct vergelijkbare spraakcorpora wordt aangetoond dat beide talen met behulp van dezelfde inventaris van intonationele basiscategorieën (toonhoogteaccenten) beschreven kunnen worden. Dit verklaart waarom volgens een aantal onderzoekers de intonatie van het Engels en het Duits sterk met elkaar overeenkomen. De talen verschillen echter in de akoestisch-fonetische realisatie van toonhoogteaccenten, en dan met name wat betreft het dalende toonhoogteaccent H*+L. De oplijning van de toonhoogtepiek in situaties waarin weinig sonorant materiaal aanwezig is verloopt anders dan in situaties waarin sonorante consonanten aanwezig zijn, terwijl een gedownstepte H*+L (verder aangegeven als !H*+L) aan het einde van de Intonationele Frase op verschillende manieren wordt geïmplementeerd. Deze verschillen verklaren waarom andere onderzoekers menen dat Engelse en Duitse intonatiepatronen juist fundamenteel verschillend zijn.

De in dit proefschrift beschreven cross-linguïstische verschillen in de realisatie van H*+L werpen ook meer licht op een tweede tegenstrijdigheid in de literatuur. Zij verklaren - tot op zekere hoogte - waarom analyses van de intonatie van het Engels meestal dezelfde intonationele basiscategorieën onderscheiden, terwijl een dergelijke consensus niet bestaat voor het Duits. In het Engels geeft de fonetische realisatie van H*+L de fonologische structuur tamelijk nauwkeurig weer. H*+L wordt gerealiseerd met wat men een 'prototypisch' dalend accent kan noemen, dat wil zeggen, als een duidelijke daling van de toonhoogte: de hoge toon wordt gerealiseerd als een piek op de beklemtoonde lettergreep die direct gevolgd wordt door een daling naar de volgende lettergreep. Dit gegeven kan direct gerelateerd worden aan het succes van de relatief concrete auditieve benadering die door de Britse school wordt voorgesteld. Een grote overeenkomst tussen de fonologische categorie en de fonetische realisatie vergroot de kans op een eenduidige analyse. In het Duits daarentegen is deze overeenkomst minder duidelijk dan in het Engels. Wat de fundamentele frequentie betreft lijkt H*+L meer op een stijgend-dalend patroon en daarom is dit specifieke accent moeilijker te onderscheiden van stijgend-dalende en stijgende accenten dan in het Engels. Het gebrek aan overeenstemming over de basiscategorieën die in het Duits onderscheiden moeten worden hangt direct samen met deze bevinding.

Hoofdstuk 1 van het proefschrift geeft een overzicht van het relatief kleine aantal bestaande studies waarin de intonatie van het Engels en het Duits vergeleken worden. Het overzicht laat zien dat deze onderzoeken een breed scala aan opvattingen hebben opgeleverd. In Hoofdstuk 1 wordt betoogd dat deze uiteenlopende opvattingen het gevolg zijn het feit dat de intonationele structuur van de talen een bepaald niveau van de linguïstische representatie vergelijkbaar zijn, terwijl ze op een ander niveau juist van elkaar afwijken. De concrete, sterk op de fonetische vorm gerichte benaderingen die tot nu toe in al het contrastieve onderzoek naar de intonatie van het Engels en Duits zijn toegepast, bieden niet de mogelijkheid cross-linguïstische overeenkomsten en verschillen op verschillende representatieniveaus te beschrijven. Dit is uiteraard wel mogelijk in een benadering waarin meerdere niveaus van analyse worden aangenomen, zoals de autosegmenteel-metrische benadering. Bovendien hebben eerdere vergelijkende onderzoeken over het algemeen nagelaten hypotheses over cross-linguïstische overeenkomsten en verschillen te genereren op basis van uitingen die direct vergelijkbaar waren tussen de talen, en het is ook niet altijd duidelijk in welke mate de analyses generaliseerbaar waren over grotere groepen sprekers.

Hoofdstuk 2 beschrijft de AM methode van analyse die specifiek ontwikkeld werd voor de vergelijking van het Engels en het Duits. Het was noodzakelijk om een nieuwe methode te ontwikkelen, omdat de twee talen voorheen door middel van verschillende versies van de AM benadering beschreven zijn. De bestaande Engelse en Duitse versies van het ToBI systeem voor prosodische labeling, zo wordt betoogd, zouden een twijfelachtig uitgangspunt voor het comparatieve onderzoek gevormd hebben. Ten eerste is de ToBI inventaris van toonhoogteaccenten voor het Engels en het Duits, waarin bitonale accenten zowel links- als rechtshoofdig kunnen zijn, voornamelijk gebaseerd op Engelse data. Wanneer men Duitse intonatie wil transcriberen met een fonologische inventaris die is ontwikkeld voor de transcriptie van het Engels, moet men over voldoende Duitse data kunnen beschikken om ambiguïteiten in de transcriptie van accenten te kunnen vermijden. Deze data ontbraken echter. Ten tweede stelt ToBI dat er twee niveaus van intonationele frasering zijn, maar specificeert onvoldoende welke fonetische correlaten het mogelijk maken tussen die niveaus onderscheid te maken. Tenslotte biedt ToBI een tamelijk ondoorzichtige beschrijving van de specificaties van Intonationele-Frasegrenzen. Wanneer twee talen met elkaar vergeleken worden, is een eenvoudigere beschrijving van grenzen wenselijk. Als antwoord hierop zijn in het hier ontwikkelde comparatieve basissysteem alle toonhoogteaccenten linkshoofdig en wordt betoogd dat een enkele intonationele frasecategorie voldoet. Anders dan in ToBI is de tonale specificatie van Intonationele-Frasegrenzen optioneel. Dat wil zeggen dat Intonationele-Frasegrenzen net als in ToBI tonaal gespecificeerd kunnen worden, maar dat ook met een symbool dat alleen de grens markeert kan worden volstaan.

Daarnaast wordt in het systeem dat voor het vergelijkend onderzoek gebruikt is aangenomen dat in de fonologische representatie twee niveaus moeten worden onderscheiden, zoals is voorgesteld in Gussenhoven (1984). De basisinventaris van toonhoogteaccenten en grenstonen wordt op het onderliggende niveau gespecificeerd. Op het fonologische oppervlakteniveau worden veranderingen in de tonale structuur van deze basiselementen beschreven die optreden wanneer ze gecombineerd worden in frasen en uitingen. Voor cross-linguïstisch werk heeft deze opsplitsing van de fonologische representatie in een onderliggend niveau en een oppervlakteniveau het voordeel dat overeenkomsten en verschillen tussen intonatiepatronen explicieter beschreven kunnen worden. Immers, niet alle intonatieverschillen hoeven te worden herleid tot verschillen tussen toonhoogteaccenten.

Hoofdstuk 2 sluit af met een voorstel voor een nieuwe methode voor cross-linguïstisch onderzoek naar intonatie. Met behulp van vijf Engelse en vijf Duitse sprekers, die per taal wat betreft leeftijd, sociale achtergrond en opleidingsniveau met elkaar overeenkwamen, werden twee direct vergelijkbare spraakcorpora samengesteld. De sprekers lazen een tekst voor waarmee ze bekend waren: het sprookje Roodkapje. De hoge mate van bekendheid van deze tekst droeg ertoe bij dat hij door alle sprekers op soortgelijke wijze geïnterpreteerd werd. De realisaties van de betreffende tekst, die door de verschillende sprekers in een identieke situationele context was geproduceerd, werden vervolgens vergeleken binnen iedere taal en tussen de twee talen.

In Hoofdstuk 3 worden de data uit het Standaard-Noord-Duitse corpus gegeven, en deze worden in Hoofdstuk 4 met de data uit het Standaard-Zuid-Brits Engels vergeleken. Het vergelijkend onderzoek laat zien dat beide talen beschreven kunnen worden met behulp van de twee fonologisch onderliggende toonhoogteaccenten H*+L en L*+H en een grenstoon H%. Verder wordt betoogd dat de specificatie van een lage grenstoon overbodig is. Op het fonologische oppervlakteniveau dragen de categorische fonologische regels DOWNSTEP, DISPLACEMENT en DELETION zorg voor aanpassingen in de realisatie van de onderliggende tonale elementen. De aanpassingen die werden aangetroffen in de Duitse data bleken ook van toepassing te zijn op de Engelse data, maar in het laatste geval werd bovendien evidentie gevonden voor Gussenhovens (1984) modificaties DELAY en HALF-COMPLETION. De evidentie voor HALF-COMPLETION was echter beperkt, en het is mogelijk dat een model waarin downstep van Intonationele Frasen wordt aangenomen een adequatere beschrijving zou geven. Dezelfde data lieten ook ruimte voor twijfel over twee typen categorische fonologische aanpassingen die Gussenhoven onderscheidt. In zijn beschrijving van het Engels (1984) worden modificaties toegepast op nucleaire accenten, terwijl verbindingsregels op prenucleaire accenten van toepassing zijn. In de hier geanalyseerde Duitse data werd DELETION echter op zowel nucleaire als prenucleaire accenten toegepast. Daarom beschouwt deze studie modificaties en verbindingsregels als een enkele groep van fonologische aanpassingsregels, en wordt voorgesteld dat de toepassing van bepaalde aanpassingen op bepaalde elementen in de representatie taalspecifiek is. De bevindingen dat HALF-COMPLETION niet duidelijk onderscheiden kan worden van downstep van Intonationele Frasen en dat het verschil tussen prenucleaire en nucleaire aanpassingen niet zo scherp omlijnd is als Gussenhoven doet veronderstellen vragen allebei om nader onderzoek naar de akoestische en auditieve effecten van die aanpassingen, op grond waarvan hun theoretische status opnieuw bezien kan worden.

Tussen de twee onderzochte talen werden verschillen in de akoestische en auditieve fonetische realisatie van H*+L gevonden. Ten eerste vertoonden de talen verschillen in de oplijning van de toonhoogtepiek. In het Duits werd de F0 piek in H*+L steeds opgelijnd met de rechterkant van de beklemtoonde lettergreep, terwijl in het Engels de piek binnen de beklemtoonde lettergreep werd gerealiseerd. Ten tweede vertoonden de talen een verschil in de realisatie van H*+L op lettergrepen met weinig sonorant materiaal. H*+L wordt in het Duits getrunkeerd, d.w.z. niet volledig uitgevoerd, en in het Engels gecomprimeerd, d.w.z. in een kortere tijd uitgevoerd. Het derde cross-linguïstische verschil betrof de akoestisch-fonetische implementatie van downstep. In het Duits kan de piek in !H*+L aan het einde van de Intonationele Frase tot op het niveau van de L gedownstept worden, maar in het Engels wordt !H*+L altijd als een daling in de toonhoogte gerealiseerd.

Hoofdstuk 5 en 6 verslaan experimenteel onderzoek dat is uitgevoerd naar aanleiding van twee hypothesen die op grond van de corpusanalyse konden worden geformuleerd. In Hoofdstuk 5 worden de effecten van truncatie en compressie op de realisatie van toonhoogteaccenten nader onderzocht. De data bevestigden de hypothese dat H*+L in het Engels wordt gecomprimeerd en in het Duits wordt getrunkeerd wanneer er weinig sonorant materiaal aanwezig is. L*+H wordt echter in beide talen gecomprimeerd. Vervolgens worden twee analyses van de asymmetrie in de Duitse resultaten besproken. Volgens de eerste analyse worden de realisaties van L*+H in het Duits gecomprimeerd omdat ze door een hoge grenstoon H% gevolgd worden. H*+L wordt daarentegen niet door een tonale specificatie gevolgd en kan daarom getrunkeerd worden. Volgens de tweede analyse worden truncatie en compressie alleen op een nucleair accent en niet op een nucleaire toon toegepast, en een opeenvolging van H en L wordt getrunkeerd terwijl een opeenvolging van L en H gecomprimeerd wordt.

Een vervolgexperiment in het Duits leverde experimentele ondersteuning voor de opvatting dat aanpassingseffecten op accenten betrekking hebben op nucleaire accenten (d.w.z. het toonhoogteaccent) en niet op nucleaire tonen (d.w.z. het toonhoogteaccent plus de daarop volgende grenstoon). Als L*+H accenten hoe dan ook worden gecomprimeerd, ongeacht of de erop volgende grenstoonspecificatie nu H% of 0% (toonloos) is, dan ondersteunt dit de 'nucleair accent hypothese'. De experimentele data onderschreven deze opvatting. Zowel L*+H H% als L*+H 0% bleken gecomprimeerd te worden. Bovendien lieten de data zien dat de keuze voor het toonhoogteaccent L*+H in de realisatie van het experimentele materiaal bepaald werd door de context, terwijl de keuze van de volgende grensspecificatie dat niet was; de aan- of afwezigheid van H% bleek sprekerafhankelijk te zijn. Desondanks gebruikten sprekers de nucleaire tonen L*+H H% en L*+H 0% niet bij het voorlezen van een lijst van nevenschikkende Intonationele Frasen. Zodra de eerste nucleaire toon in een opsomming gekozen was werden alle volgende onderdelen van de opsomming met dezelfde nucleaire toon geproduceerd. Deze bevinding geeft aan dat sprekers op het niveau van de volledige structuur een keuze maken voor H% of 0%, en niet op het niveau van iedere individuele Intonationele Frase binnen de structuur.

Hoofdstuk 6 rondt het proefschrift af met een onderzoek naar *downtrends* in het Engels en het Duits. Een experimentele vergelijking van F0-patronen in opsommingen waarin downstep wordt toegepast bevestigt de hypothese dat in het Duits de laatste piek een extra verlaging ondergaat, hetgeen in de Engelse data niet het geval is. Deze bevinding is in tegenspraak met de resultaten die door Liberman and Pierrehumbert (1984) worden gerapporteerd voor het Amerikaans Engels, waar *final lowering* werd aangetroffen aan het eind van gedownstepte series accenten. In het proefschrift wordt verondersteld dat dit verschil in de resultaten tussen het Amerikaans Engels en het Brits Engels toegeschreven kan worden aan het effect van declinatie. De resultaten van een vervolgexperiment ondersteunden deze veronderstelling. Met andere woorden, het is mogelijk dat een combinatie van downstep en declinatie een verklaring biedt voor de data op grond waarvan eerder *final lowering* is vastgesteld.

Kort samengevat kan gesteld worden dat dit proefschrift evidentie aandraagt voor cross-linguïstische intonationele verschillen op verschillende niveaus van de representatie. De intonatiesystemen van het Engels en het Duits vertonen grote overeenkomsten op het niveau van de fonologische representatie, maar verschillen op het akoestisch-fonetische niveau. Verwacht mag worden dat cross-linguïstisch onderzoek er in het algemeen baat bij zal hebben van een dergelijke scheiding van representatieniveaus uit te gaan.

Curriculum Vitae

Esther Grabe studied Modern Languages (English, German) and General Linguistics at the Universities of Braunschweig and Münster in Germany. In 1988, she moved to Great Britain, where she spent one year as a teaching assistant for German at secondary schools in North Wales. She subsequently read General Linguistics at the University of Manchester and was awarded a Master of Linguistics in 1990. Between 1990 and 1994, she held research positions at the University of Leeds and Cambridge, investigating the production of fricatives in European Languages and the role of prosody in human language comprehension. In October 1994, she was awarded a stipend from the Max-Planck Gesellschaft which allowed her to carry out her doctoral research at the Max-Planck-Institute for Psycholinguistics in Nijmegen in the Netherlands. In October 1997, Esther Grabe returned to the University of Cambridge, where she holds a project funded by the Economic and Social Research Council with Dr. Francis Nolan. The project will provide an account of the phonetics and phonology of intonation in a number of British English dialects.

MPI Series in Psycholinguistics

1. The Electrophysiology Of Speaking. Investigations On The Time Course Of Semantic, Syntactic And Phonological Processing.
   Miranda van Turennout

2. The Role Of The Syllable In Speech Production. Evidence From Lexical Statistics, Metalinguistics, Masked Priming And Eletromagnetic Midsaggital Articulography.
   Niels. O. Schiller.

3. Lexical Access In The Production Of Ellipsis And Pronouns.
   Bernadette M. Schmitt

4. The Open-/Closed-Class Distinction In Spoken-Word Recognition.
   Alette Haveman

5. Gesture And Speech Production.
   Jan Peter de Ruiter

6. The Acquisition Of Phonetic Categories In Young Infants: A Self-Organising Artificial Neural Network Approach.
   Kay Behnke

7. Comparative Intonational Phonology: English And German.
   Esther Grabe