Phonetics and Phonology of Thematic Contrast in German

Bettina Braun

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Key words	Abstract
annotation	

Abstract

It is acknowledged that contrast plays an important role in understanding discourse and information structure. While it is commonly assumed that contrast can be marked by intonation only, our understanding of the intonational realization of contrast is limited. For German there is mainly introspective evidence that the rising theme accent (or topic accent) is realized differently when signaling contrast than when not. In this article, the acoustic basis for the reported impressionistic differences is investigated in terms of the scaling (height) and alignment (positioning) of tonal targets.

Subjects read target sentences in a contrastive and a noncontrastive context (Experiment 1). Prosodic annotation revealed that thematic accents were not realized with different accent types in the two contexts but acoustic comparison showed that themes in contrastive context exhibited a higher and later peak. The alignment and scaling of accents can hence be controlled in a linguistically meaningful way, which has implications for intonational phonology. In Experiment 2, nonlinguists' perception of a subset of the production data was assessed. They had to choose whether, in a contrastive context, the presumed contrastive or noncontrastive realization of a sentence was more appropriate. For some sentence pairs only, subjects had a clear preference. For Experiment 3, a group of linguists annotated the thematic accents of the contrastive and noncontrastive versions of the same data as used in Experiment 2. There was considerable disagreement in labels, but different accent types were consistently used when the two versions differed strongly in F0 excursion. Although themes in contrastive contexts were clearly produced differently than themes in noncontrastive contexts, this difference is not easily perceived or annotated.

Key words (continued):
annotation
contrast
information structure
intonation
ToBI

1 Introduction

In this paper the prosodic marking of thematic contrast in German is analyzed. Thematic contrast is used to establish an explicit or implicit contrast to an element

Language and Speech

that is recoverable from the previous discourse or the situation (cf. Büring, 1997b; Jacobs, 1997), as exemplified in the following literary quotation.

> *Ben stand auf. Ich muss noch was tun. Er sagte das so, als sei ich der, der nichts tun muss* (Timm, 2003, p. 156).

> 'Ben stood up (and said): "I still have some work to do." He said this in such a way as to imply that I was the one who had nothing to do.'

In reading this literary quotation aloud, a German speaker would first produce the utterance *Ich muss noch was tun* ('I still have some work to do') with a neutral intonation where the pronoun is unaccented and 'tun' receives a falling accent. However, in light of the third utterance ("He said this in such a way as to imply that I was the one who had nothing to do"), he would pronounce this sentence in a different way because the last utterance indicates that there is a contrast between the two referents (Ben and the literary first person) and between the predications made about them (one has to work, the other one doesn't have to). In that case, there would probably be a rising accent on the pronoun, a sustained high F0, and a falling accent on 'tun' (this contour was first described under the term "hat pattern" by Cohen and 't Hart, 1967). Following Firbas (1964) and Steedman (2000), among others, the respective persons — which are given in the situation — are termed *thematic*, while the predications that are made about them — which constitute new information — are denoted *rhematic*.

Although in German, there are also other means to encode contrast, such as word order variation or particles,[1] this example shows that contrast can be retrieved from the prosodic realization only. The focus of the present article lies in investigating the prosodic realization of thematic material produced in contrastive and noncontrastive contexts.

In this introduction, first GToBI, the most recent description of German intonation is summarized. Besides, the phonetics and phonology of the hat pattern, of thematic (prenuclear), and rhematic (nuclear) accents in German are reviewed. This is followed by a discussion of previously reported meaningful differences in pitch accent realization and their formalization in intonational phonology. Then the literature on thematic contrast in German is summarized.

---

[1]   Since German has a relatively free constituent order, thematic contrast can be expressed by scrambling constituents. In main clauses, the finite verb occupies the second position, while the first position (German: Vorfeld) can be filled with a variety of constituents. In neutral, noncontrastive statements, the Vorfeld is occupied by the subject or by a preposition phrase (e.g., *Marlene studiert in Hamburg*, 'Marlene is studying in Hamburg' or *Im Mai fahre ich nach England* 'In May, I will go to England'). Other constituents in the Vorfeld (such as direct or indirect objects, and finite verbs), however, render the sentence contrastive (e.g., *Meiner Schwester habe ich ein Buch gegeben* 'To my sister I gave a book' or *Geschlafen hat keiner von uns* lit: 'Slept has none of us; None of us slept'). These marked and rather uncommon sentence structures (Weber & Müller, 2004) are often realized with certain intonation contours to ease the processing load (Weber, Grice, & Crocker, 2006).

1.1
German intonation

There are many different descriptions on the intonation of German, which cannot be reviewed in full detail here (Fox, 1984; Isačenko & Schädlich, 1966; Kohler, 1991a; Moulton, 1962; Pheby, 1975; Uhmann, 1991; von Essen, 1964; Wunderlich, 1991, among others). We therefore concentrate on GToBI, the most recent description on German intonation. When discussing the hat pattern, we will also refer to other autosegmental-metrical (AM) descriptions, such as Féry (1993), Grabe (1998), and Wunderlich (1991).

GToBI (German Tone & Break Indices, see Grice, Baumann, & Benzmüller, 2005) was developed along the lines of the annotation system for American English, MAE-ToBI (Mainstream American English Tone and Break Indices, see Beckman, Hirschberg, & Shattuck-Hufnagel, 2005); for earlier descriptions see also Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert, & Hirschberg (1992) and Beckman & Ayers (1997). In GToBI, there are six accent types, H*, L+H*, L*, L*+H, H+L*, and H+!H*. Grice et al. (2005) only discuss the meaning of a small set of complete contours (consisting of one pitch accent, a phrase accent, and a boundary tone), but not the meaning of individual accents in isolation. All the nuclear accents can theoretically also appear in prenuclear position. Generally, however, prenuclear accents in German are mainly rising (H*, L+H*, L*+H, and more rarely L*), while nuclear accents are more varied.

Labeling accentual distinctions in autosegmental phonology is not straight-forward. Consistency studies conducted with a group of labelers often give rise to disagreement between particular accent distinctions, see Pitrelli Beckman, and Hirschberg (1994) and Herman and McGory (2002) for American English, as well as Grice, Reyelt, Benzmüller, Mayer, and Batliner (1996) for German. In the consistency study on GToBI by Grice et al. (1996), most confusions were found between the accent pairs L+H* and H* (28% — expressed as percentage of confusions involving one of the 2 accents), between L*+H and L* (17%), between L+H* and L*+H (16%), and between H* and H+!H* (15%).

1.1.1
Phonetics and phonology of hat patterns

Sentences with contrastive themes are often claimed to be realized with a hat pattern (Büring, 1997b; Wunderlich, 1991, among others). The most striking characteristics of hat patterns is that they have two accents involving a high tone but there is no low target between these two accents. That is, F0 stays high (or declines slightly) between the two accents. In the phonological literature, hat patterns are described in a variety of ways, some of which will be outlined here: Wunderlich (1991) called this pattern *Brückenakzent* ('bridge accent') and described it phonologically as H* H L*. The thematic accent (H*) is what he calls a default accent, the rhematic accent is realized with a fall. In between there is a floating H tone which is not associated to any of the two accents but determines the high pitch between them. Féry (1993) distinguished between two different hat pattern contours, H* H*L (in her model a complete linking of two H*L pitch accents) and L*H H*L (two fully realized accents). She discusses that

the former pattern is widely used in different contexts (i.e., information structures), while the second one is restricted to topic-comment structures (L*H marking the topic or theme and H*L the comment or rheme). According to her, however, the difference between these two contours "is not always phonologically clear-cut" (p. 151). Grabe (1998) describes the hat pattern as H* > H*+L, that is, with a deleted trailing tone of the first accent. Grice et al. (2005) do not discuss the hat pattern in their article, nor in the training materials.

Summarizing, both the initial and the final accent in a hat pattern have been described with different accent types in German: theme accents are said to be H* and L*+H, rheme accents (H+)L* and H*L. Therefore the only reliable characteristic of the hat pattern seems to be the lack of a low tonal target between the two accents.

1.1.2
Phonetics and phonology of thematic accents

Although often claimed as such, a hat pattern is not always indicative of contrastive themes. Kohler (1991a), for instance, identified this contour as "ideally suited for matter-of-fact reading in German" (p. 328). More and more impressionistic prosodic descriptions on contrast in German have emphasized that it is especially the initial thematic F0 rise that differs in contrastive and noncontrastive contexts (Büring, 1997b; Féry, 1993; Jacobs, 1997). Büring (1997b), for instance, impressionistically described the difference between contrastive and noncontrastive themes as a difference in pitch excursion of the rising accent (p. 52). Jacobs (1997), on the other hand, proposed that it is not the pitch rise per se that signals a contrastive theme but rather the fact that the rise is preceded by a noticeable trough. He does not make a comparison to noncontrastive themes, however.

In the recent literature on alignment, sentence-initial rising accents have been extensively studied. It has been shown that under certain conditions, both L and H tones in a rising accent show stable alignment. Stable segmental anchors for low tonal targets preceding an accentual rise have been reported by Arvaniti, Ladd, and Mennen (1998); Caspers and van Heuven (1993); Prieto, van Santen, and Hirschberg (1995), among others. Recent work on various languages has suggested that accentual peaks are also consistently aligned with the segmental structure. For instance, Modern Greek prenuclear rising accents were shown to have high targets that were aligned at a fixed distance from the beginning of the first poststressed vowel (Arvaniti et al., 1998). The presence of two fairly stable segmental anchor points was replicated for English rising accents by Ladd, Faulkner, Faulkner, and Schepman (1999).[2] The alignment of the peak (H tone) in an accentual rise can be influenced by various factors, such as the proximity of an upcoming prosodic boundary or accent, phonological vowel length, and speech rate (Caspers & van Heuven, 1993; Prieto et al., 1995; Silverman & Pierrehumbert, 1990; Steele, 1986). The realization of thematic accents further

---

[2]   The findings for Dutch prenuclear rises are somewhat less clear-cut. Ladd, Mennen, and Schepman (2000) report different alignment patterns for the end of the rise depending on the syllable structure of the accented syllable. If there is a phonologically long vowel, the peak is found at the end of the vowel, but when there is a short vowel, the peak is found within the following onset consonant.

depends on dialect. For German, Atterer and Ladd (2004) investigated the alignment properties for northern and southern German speakers and reported that the trough before the prenuclear rise was significantly later for southern German speakers than for Northerners. A similar trend was observed for the high target at the end of the rise but that difference was not statistically significant.

1.1.3
Phonetics and phonology of rhematic accents

Irrespective of whether sentences in contrastive contexts are realized with a hat pattern or not, the rhematic accent is part of the overall contour. Rhematic (or nuclear) accents have been extensively studied in the past and are therefore not the main topic of this article. Generally, nuclear accents are more varied than prenuclear ones. GToBI, for instance, recognizes the following nuclear accents in declarative contexts: H* L−% is used for neutral statements, L+H* L−% for contrastive assertions, L*+H L−% for self-evident or sarcastic assertions, H+!H* for established facts, and H+L* L−% for soothing or polite requests. For German, Kohler (1991b) has shown a categorical distinction between two accent types: early peaks (comparable to GToBI (!)H+L* or H+!H*) contrast with medial peaks (GToBI H* L−, !H* L−, L+H* L−, L+!H* L−).[3]

With respect to alignment, Dilley, Ladd, and Schepman (2005) tested two conflicting hypotheses regarding nuclear L+H* pitch accents in Dutch. Their data lends support to the view that tonal targets are aligned with particular segmental landmarks (segmental anchoring hypothesis) rather than at a fixed temporal distance from each other (constant interval hypothesis). Schepman, Lickley, and Ladd (2006) investigated the alignment in Dutch nuclear bitonal pitch accents and argued that nuclear accents are less affected by stress clash than prenuclear ones because the peak in nuclear accents is generally aligned earlier than in prenuclear ones. In their view, the alignment in nuclear accents is mainly influenced by the phrase accent which overrides any effects of stress clash.

1.1.4
Meaningful gradual differences in accent types

Gradual variations in accent types have been reported for both thematic and rhematic accents. Liberman and Pierrehumbert (1984), for instance, showed for English that the height of nuclear accents is highly predictable from the context they appear in. Increasing the F0 height of an accent has been reported to increase its perceived prominence (Kohler & Gartenberg, 1991; Rietveld & Gussenhoven, 1985; Terken, 1991) and emphasis (Ladd & Morton, 1997; Nolan, 1995). Also, Bartels and Kingston (1994) found that the slope of an accentual rise was an important perceptual cue to rhematic contrast. Wichmann, House, and Rietveld (2000) investigated the effect of the position of an utterance in a text (initial, medial, final) on the scaling and alignment of sentence-initial accents. They reported that the peak in rising accents was

---

[3]   GToBI is missing the accent type H*+L, which is used in many other AM descriptions of German to denote a fall onto the accented syllable (for a discussion on this issue see Grice et al., 2005, p. 79). To describe an accentual fall, GToBI uses a L− phrase accent, which is aligned with the poststressed syllable.

realized higher and later when the utterance occurred more initially in the paragraph. Gussenhoven (2002) has recently claimed that "higher peaks will tend to be later than lower peaks" (p. 52), arguing that it takes longer to reach a higher peak. However, this is no firm prediction as he also regards peak delay as a substitute for peak height (Gussenhoven, 2002, 2004). He links a raised peak to an increase in (local) effort and argues that speakers and listeners tacitly know that it takes longer to reach a higher peak than a lower one (assuming the same speed of pitch change). So, listeners are inclined to interpret both later peaks and higher peaks as showing increased effort. These observations raise an interesting issue as to how different prosodic variables interact. Arvaniti and Garding (to appear) investigated how different degrees of emphasis affect the realization of nuclear H*, L*+H, and L+H* accents in two American English dialects. Speakers from Southern California switched from an H* accent to a L+H* accent with increased emphasis while speakers from Minnesota implemented the increased emphasis phonetically by delaying and raising the peak.

Various proposals have been made to use a more fine-grained approach to pitch accent marking. These are based on the observation that there are consistent differences in the alignment and scaling of accents that cannot be captured with standard AM notation. Based on data from several languages, Ladd (1983) proposed to use the features [delayed peak], [downstep], and [raised peak] to mark differences in intonational phonology. A similar line of analysis was taken by Prieto (2004). She showed that there are consistent differences in peak height and alignment that signal different sentence types in Peninsular Spanish. In questions, the prenuclear peak is realized higher and earlier than in statements. She also analyzed this distinction in terms of the extrinsic features [+ delayed peak] and [+ raised peak]. Prieto, D'Imperio, and Gili-Fivela (2005) proposed a different solution for Central Catalan. In this language there are four rising accents, three of which have the low tone aligned with the start of the stressed syllable: a rise with a peak aligned within the stressed syllable, a rise with a peak delay, and a rise with a peak delay where the peak is close to the end of the word; furthermore, there is a post-tonic rise. In Standard AM-theory, the first three rising accent types cannot be distinguished. Since the end of the rise is associated with different landmarks, Prieto et al. (2005) argue to use secondary associations to distinguish between the different variants of rising L+H* accents.

1.2
Previous work on thematic contrast

Thematic contrast has received some attention in linguistics. Contrastive themes (also called *I-Topics*, see Jacobs, 1997) establish a contrast to another given element. This may be explicitly mentioned (as in semantic parallelism or double contrast pairs, such as "Frida went to the party, but Mary washed her hair," see Zeevat, 2004) or left implicit (e.g., *Meine Frau hat keine fremden Männer geküsst*, 'My wife didn't kiss other men', see Büring, 1997b). Sentences with implicit contrastive elements can point to alternatives to the current theme the speaker wants to talk about (Büring, 1997b; Krifka, 1999; Umbach, 2001). Also, sentences with a contrastive theme can evoke a scope inversion if two quantifiers or a quantifier and a negation particle are

involved (Büring, 1997a; Krifka, 1998; Wunderlich, 1991) and give rise to contrastive implicatures.[4]

Semantic formalisms associate information structure constituents with intonational categories of some sort (e.g., A and B accents in Jackendoff (1972) and Büring (1997b), or L+H* and H* in Steedman (2000)). Therefore, the adequacy of the formalisms partly depends on the kind of intonational categories referred to; especially it is important that they are sufficiently distinct prosodically and that they can be reliably annotated. Since prosodic realizations were shown to vary with regional background (Atterer & Ladd, 2004; Bruce & Gårding, 1978; Gilles & Peters, 2004; Peters, 1999; Peters, 2004) and from speaker to speaker, semantically meaningful accent categories should be broad enough to include intra- and interspeaker variability, but at the same time narrow enough so that meaningful phonetic variation is captured. This demand may be difficult to fulfill, especially in the light of the high proportion of intertranscriber disagreement on the identity of accent types in intonational annotation.

It appears that the semantic distinction between contrastive and noncontrastive themes has not been sufficiently analyzed in prosodic terms yet. Except for a related production experiment conducted by Mehlhorn (2001) there is only introspective evidence for a prosodic difference in contrastive and noncontrastive contexts. Mehlhorn's experiment mainly aimed at eliciting hat patterns; she hence recorded sentences with contrastive word orders that are discussed as being realized with hat patterns in the literature (such as sentences with particular syntactic structures, e.g., containing split-NPs, sentences with quantifiers, and complex sentences with more than one topic); additionally, she recorded a corpus of syntactically neutral sentences in contrastive and noncontrastive contexts. The data were analyzed by averaging the F0 contours over all the speakers in the two conditions. Unfortunately, the sampling procedure is left unclear and statistical analyses are not described. She found that contrastive themes had a steeper rise, a higher F0 range, and a longer syllable duration. Further, it was reported that sentences with contrastive topics were generally produced with a hat pattern. Mehlhorn (2001) also conducted two perception experiments where participants indicated the prominent words in the sentences, their relative prominence, as well as the direction of the F0 movement. The findings of these experiments indicate that themes in contrastive contexts are perceived as prominent while themes in noncontrastive contexts are not. Furthermore, themes in contrastive contexts are perceived as being more prominent than the corresponding rheme accents. However, these experiments do not inform us about the communicative function of different productions, that is, whether they are interpreted differently.

This paper is divided into three parts. Experiment 1 forms the main part; there, the acoustic differences between identical sentences in contrastive and noncontrastive

---

4   A formalization for the noncontrastive sentence 'I still have some work to do' is shown in (1), while the contrastive version of the same sentence is formalized as shown in (2):

(1) $\exists x[I'(x) \land have\text{-}to\text{-}work'(x)]$

(2) $\exists x[I'(x) \land have\text{-}to\text{-}work'(x)] \land \exists y[y \neq x] \land \neg\, have\text{-}to\text{-}work'(y)]$

contexts are described. The target sentences are syntactically unmarked and contain only full lexical forms for thematic and rhematic constituents to make the productions in the two contexts comparable. Thematic accents are intonationally annotated and the acoustic differences between themes in contrastive and noncontrastive contexts are compared. This also includes an analysis of how the acoustic variables interact with each other and the problems this poses for annotating these accents. Since many previous studies emphasized the importance of the hat pattern for signaling thematic contrast, the number of hat patterns in contrastive and noncontrastive contexts are compared. In addition, the rhematic accent type is analyzed. In Experiment 2, the perception and interpretation of sentences produced in contrastive and noncontrastive contexts is investigated. To this end, a subset of productions was selected and linguistically naïve subjects chose which of two productions was more appropriate in a contrastive context. Subjects' choice was analyzed in terms of the acoustic differences between the contrastive and noncontrastive version of a sentence pair to provide a first indication about which variables are important for listeners. In Experiment 3, results from a labeling experiment are described. Productions in contrastive and noncontrastive contexts are annotated in isolation by a group of linguists to examine whether the difference between contrastive and noncontrastive themes can be expressed in terms of different accent types. Similar to Experiment 2, the label differences between the contrastive and noncontrastive version are analyzed in terms of the respective acoustic differences. The results are further compared to the disagreements reported in the GToBI consistency experiment (Grice et al., 1996).

2 Experiment 1

It is assumed that pragmatically distinct contexts (i.e., contrastive and noncontrastive ones) are reflected in the speaker's discourse model, which influences speech planning and hence the prosodic realization of utterances. Unlike the introductory example above, utterances with full lexical forms in the thematic constituent (instead of pronouns) are analyzed. Theme marking cannot be compared in sentences involving pronominalized thematic constituents since accented pronouns are already very marked, regardless of the actual intonational realization. The following questions are addressed:

1. Is there a phonological difference in theme accent type in sentences produced in contrastive and noncontrastive context?

2. Are there more hat patterns in contrastive than in noncontrastive contexts?

3. Is there a difference in the distribution of rheme accent types in contrastive and noncontrastive contexts?

4. Is there an acoustic difference between identical sentences produced in contrastive and noncontrastive contexts? If yes, what are the differences and how do the variables interact? Especially, the correlation between peak scaling and alignment is investigated since delayed peaks may be a substitute for higher peaks (Gussenhoven, 2002).

2.1
Method

Provided that contrastive themes are indeed realized differently than noncontrastive themes, we should be able to observe these differences across different speakers and across differences in phonological and metrical structure. Therefore, a fair amount of variation was included in the experimental items. The stimuli varied in the number of syllables in the target word, phonological vowel length of the stressed syllable, the position of the word stress in the thematic constituent, and whether the thematic constituent was a preposition phrase or a noun phrase. Furthermore, subjects were not chosen from a particular variety of German, but came from different regions.

Short paragraphs were designed and the target sentences were embedded in these larger contexts that controlled both the information structure and the contrastiveness. Identical sentences were hence produced in a contrastive and noncontrastive context. Analyzing alignment in continuously read text differs from most previous studies on alignment; these often had lists of isolated, segmentally controlled, sentences to be read aloud (cf. Arvaniti et al., 1998; Atterer & Ladd, 2004; Ladd et al., 1999, among others).

2.1.1
Participants

Twelve native German speakers, seven female and five male, voluntarily participated in the recording. They were between 23 and 36 years old. At the time of recording they were graduate and postgraduate students, or staff members of Edinburgh University; they had been in Edinburgh for periods ranging from a few months to four years. None of them had any known speech or hearing problems and they were naïve with respect to the purpose of the experiment. The participants all spoke Standard German but originated from different parts of Germany (8 northern German speakers and 4 southern German speakers, for more detail see Table 7 in the Appendix).

2.1.2
Materials

Target sentences started either with a subject noun-phrase (for example, *Die Malayen leben von der Landwirtschaft* [5] 'The Malaysians live from agriculture') or a prepositional phrase (e.g., *Im Januar ist es frostig*, 'In January, it is frosty'). These word orders are very frequent in German (Weber & Müller, 2004) and are not inherently contrastive. Target sentences consisted of a simple theme-rheme structure, with the theme identifying a topic (an animated subject, a location or a time) and the rheme making a proposition about it. Example 3 shows the information structure of a target sentence (location of theme and rheme accents is indicated by small capitals).

(3) Die MaLAYen leben von der LANDwirtschaft

       theme   *rheme*

---

[5]   The German word 'Malayen' actually refers to the ethnic group whereas 'Malaysier' would be the correct word for the nationality. The incorrect term was chosen here for the sake of sonority and none of the readers objected against it.

Sentences were constructed so that the rheme accent (nuclear accent) would fall on the same constituent in contrastive and noncontrastive contexts; this way the influence of the rhematic accent on the phonetic implementation of the thematic accent is reduced. Target constituents were chosen to be maximally sonorant and to be groupable with other items in a "set of alternatives," for example, the noun-phrase 'the Romans' (which can be contrasted with e.g. 'the Teutons'). In total, 12 target sentences were constructed; they are listed in Table 8 in the Appendix.

Two conditions were investigated, contrastive and noncontrastive context. In noncontrastive contexts (Example 4), the thematic element is entirely background information (given), while the theme in contrastive contexts (Example 5) is further split into a focused element (Malayen) and the background.

(4) non-contrastive: Die MaLAYen leben von der LANDwirtschaft

  background   *background*   *focus*

      theme   *rheme*

(5) contrastive: Die MaLAYen leben von der LANDwirtschaft

  background focus   *background*   *focus*

      theme   *rheme*

Twelve paragraphs containing five to six sentences each were constructed for each condition (on average 5.5 sentences in noncontrastive contexts and 5.6 in contrastive contexts). To reduce the influence of paragraph intonation (Sluijter & Terken, 1993), the position of the target sentence in the paragraph was matched across contexts. Target sentences were not paragraph-final to avoid lowering of F0 and creaky voice.

Two sample paragraphs, translated into English, are shown in Table 1; the German originals are printed in Table 6 in the Appendix. For the noncontrastive context, the target constituent or a hypernym was introduced as a topic early in the paragraph and resumed in the target sentence (e.g., Malaysia, the Malaysians). In contrastive paragraphs, a proposition about an alternative element occurred in the context (e.g., the Indonesians, the Malaysians). To link the two elements to one another, the context also contained a sentence stating the commonality between the two alternatives (e.g., *Malaysien und Indonsien sind Nachbarländer im Südchinesischen Meer.* 'Malaysia and Indonesia are neighboring countries in the South China Sea'). The contrast is always established by a semantic parallelism (Prevost, 1995; Theune, 1999).

In addition, 12 distractor paragraphs on the same topic as the target paragraphs were constructed to prevent subjects from expecting identical sentences in paragraphs about certain topics. Sixteen other filler paragraphs, which were unrelated to the experimental items, were constructed. This resulted in 52 paragraphs.

Table 1

English translation of a sample noncontrastive and contrastive paragraph. The target sentences are highlighted in bold face

Noncontrastive context	*Contrastive context*
Many Europeans don't know much about Malaysia. The country consists of two islands. To ease the communications between the two parts, almost every household has a computer with Internet access. However, Malaysia is not a highly technological country. **The Malaysians live from agriculture.** They are neither especially poor nor rich.	Malaysia and Indonesia are neighboring countries in the South China Sea. Despite their geographical adjacency, their living and working conditions differ tremendously. In Indonesia, tourism is very important and many people work in this sector. **The Malaysians live from agriculture.** They have mainly focused on the cultivation of rice.

2.1.3
Procedure

Participants received written instructions to read the paragraphs at normal speed as fluently as possible. They were given a pile of 52 A5 cards that contained the paragraphs. They were told to silently scan the texts before reading them aloud. If paragraphs contained too many hesitations or false starts they were asked to read them again. Participants were recorded in a sound-proof room in the Department of Theoretical and Applied Linguistics at Edinburgh University. Data was stored simultaneously on DAT-tapes and on a PC with a sampling rate of 44.1kHz. The presentation of the paragraphs was block-wise randomized, separating the contrastive and noncontrastive versions of a given target sentence by a minimum of five other paragraphs. Recording sessions lasted between 20 and 30mins. After the recording, participants were asked whether they noticed that identical sentences had appeared in different paragraphs.

2.2
Analysis

In total, 86 utterance pairs were analyzed. The recordings of three speakers were discarded before analysis; two of them read too fast and monotonously, which made the identification of F0 minima and maxima almost impossible and therefore unreliable.[6] One speaker had a very emphatic reading style and inserted many intermediate phrase breaks at various places in the sentences. One sentence (*In Amerika besitzen viele eine Waffe*, 'In America, many people own a weapon') had to be discarded from analysis because the rheme accent was realized on different constituents in the two conditions. For one speaker half of the sentence pairs were lost due to recording failure. A further six sentences had to be excluded because of hesitations and mispronunciations in the target sentence for one of the two conditions (also the repetitions of these paragraphs could not

---

[6]   Problems with monotonous speakers have been reported in various production experiments, see for example, Atterer and Ladd (2004), p. 189; Ladd et al. (1999), p. 1548; and Ladd and Schepman (2003), p. 86.

be used because they showed a considerable amount of hyperarticulation as speakers wanted to correct their previous mistake). The unequal number of items per speaker is not crucial since the experiment was designed mainly for within-subject comparisons.

Data were analyzed using Xwaves; F0 tracking was conducted with the built-in F0 tracking algorithm (get_F0), which is based on normalized cross correlation function and dynamic programing (cf. Talkin, 1995); default values of 10 ms frame shift and 7.5 ms autocorrelation window were used. Artifacts introduced by the F0 tracking algorithm (F0 doubling or halving) were manually corrected. Up to a sequence of five missing F0 values were linearly interpolated. Then, the F0 contour was smoothed using a seven-frame window (7.5 ms each) with mean smoothing (for a comparison of different smoothing techniques on the data, see Braun, 2005).

2.2.1
Intonational labeling

All thematic constituents were produced with a rising F0 movement and there was both a low and a high tonal target. Following the guidelines of GToBI (Grice et al., 2005), accents whose stressed syllable was perceived as high were labeled as L+H* and accents whose stressed syllable was perceived as low were annotated as L*+H (as will be discussed later, L+H* accents strictly speaking have the low tonal target in the pretonic syllable and reach the peak late in the accented syllable; this however, was rarely the case in the present data—L+H* was therefore marked on purely perceptual grounds).

Further, the frequency distribution of the nuclear accents was analyzed in more detail. Since nuclear accents have been shown to be categorically distinct in German (Kohler, 1991b), rhematic accents are grouped into falling accents (collapsing the AM accent types H+L* and !H+L*) and high ones (collapsing H* L−, !H* L−, L+H* L−, L+!H* L−).

With respect to the overall contour, the number of hat patterns were counted in contrastive and noncontrastive contexts. Operationally, hat patterns were defined as cases in which F0 did not drop down to a low tonal target between the prenuclear and nuclear accent. It is expected that there are more hat patterns in contrastive contexts (Büring, 1997b; Jacobs, 1996; Mehlhorn, 2001).

2.2.2
Acoustic labeling

Data annotation was done on the segmental and suprasegmental level, concentrating around the area of the F0 rise. Label points are depicted in Figure 1. Standard segmentation criteria were followed for segmental annotation, using both information from the acoustic signal and a broadband spectrogram. Following Ladd et al. (2000) and Atterer and Ladd (2004), four segmental landmarks in the area around the stressed and poststressed syllable were marked:

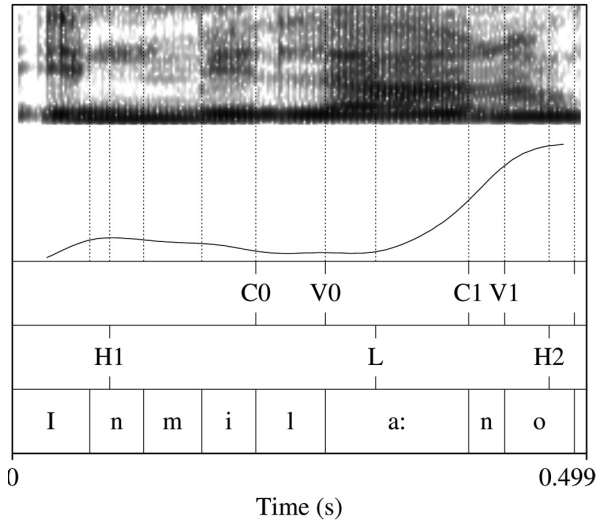C0: Consonantal onset of the stressed syllable

V0: Onset of the stressed vowel

C1: Consonantal onset of the poststressed syllable

V1: Onset of the poststressed vowel

Figure 1

Segmental, suprasegmental, and phonemic labels of a sample constituent (In Milano). The stressed syllable lies between C0 and C1

Suprasegmentally, the following events in or before the test words were marked:

H1:   Highest point before the fall. In most cases this value was found in the middle of the vowel of the first unstressed syllable of the prosodic word.

L: Local minimum preceding the rise. If there were more consecutive points with an equally low F0 value (differing no more than 1 Hz), the last low point was marked.

H2:   First local maximum after the end of the F0 rise. If there were more consecutive points with an equally high F0 value (differing no more than 1 Hz), the first high point was marked.

2.2.3
Dependent variables

Four groups of acoustic variables were investigated: F0 variables, temporal variables, alignment variables, and the slope of the rise.

Five F0 variables were analyzed, the absolute F0 value in Hz for the three tonal events (F0(H1), F0(L), and F0(H2)), the magnitude of the F0 fall and the magnitude of the subsequent rise, both in Hz and in semitones (ΔF0(fall), ΔF0(rise)).

The temporal variables consisted of the duration of the whole utterance (utt-dur), of the thematic constituent (theme-dur), the stressed syllable (syll-dur), and the stressed vowel (vowel-dur). Further, the duration of the F0 fall (t(L)−t(H1)) and the F0 rise (t(H2)−t(L)) were measured. They are referred to as Δt(fall) and Δt(rise).

Alignment variables represent a link between the segmental and suprasegmental tier. It is still an open methodological issue whether alignment should be measured in

absolute terms (referring to some given anchor point) or in proportion to the duration of the segment or syllable the event appears in (see discussions in Atterer & Ladd, 2004; Prieto et al., 1995; Schepman et al., (2006); Silverman & Pierrehumbert, 1990). Therefore, alignment was expressed both absolutely and proportionally to the duration of the stressed syllable. In German, the peak in prenuclear accents is reached in the poststressed syllable only (Atterer & Ladd, 2004); therefore, peak alignment was calculated with respect to the start of the poststressed vowel: al(H2,V1). The alignment of the trough was calculated to the start of the stressed syllable, al(L,C0), and to the start of the stressed vowel, al(L,V0). Positive values indicate that the tonal target is located after the segmental landmark. Finally, the slope of the F0 rise was calculated by dividing the F0 excursion by its duration, slope(rise).

2.3
Results

None of the subjects reported to have noticed that identical sentences appeared in different paragraphs.

2.3.1
Qualitative analysis—Theme accents

L*+H and L+H* accents were equally distributed in contrastive and noncontrastive contexts (examples of 2 different theme accents are presented in Fig. 2). In 72.1% of the cases the contrastive and noncontrastive version of a sentence pair was annotated with the same accent type (32 times with L+H* and 30 times with L*+H).
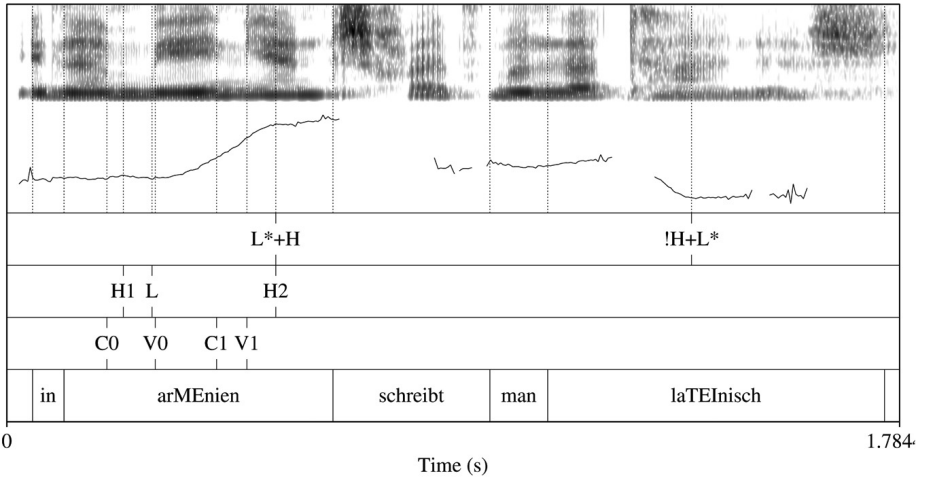
Figure 2

Noncontrastive (panel (a)) and contrastive (panel (b)) realization of the sentence *In Armenien schreibt man lateinisch* ('In Armenia, the Latin alphabet is used')

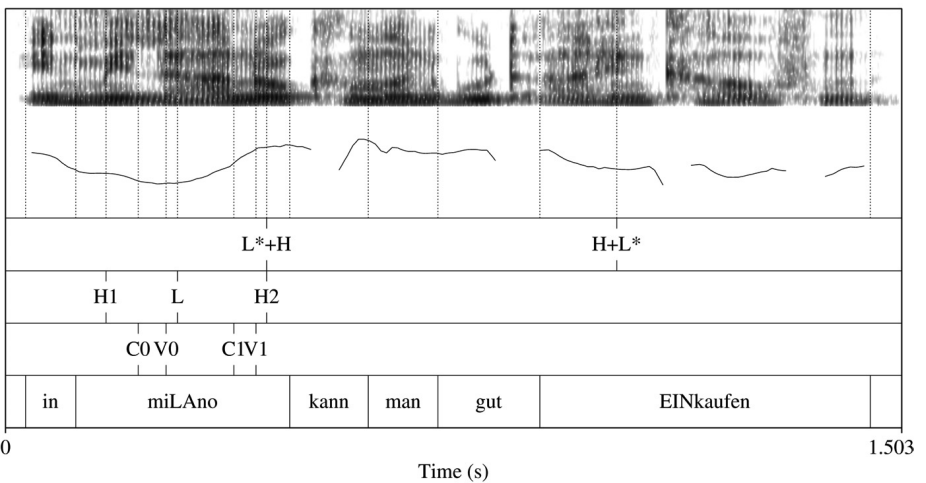Panel (a)
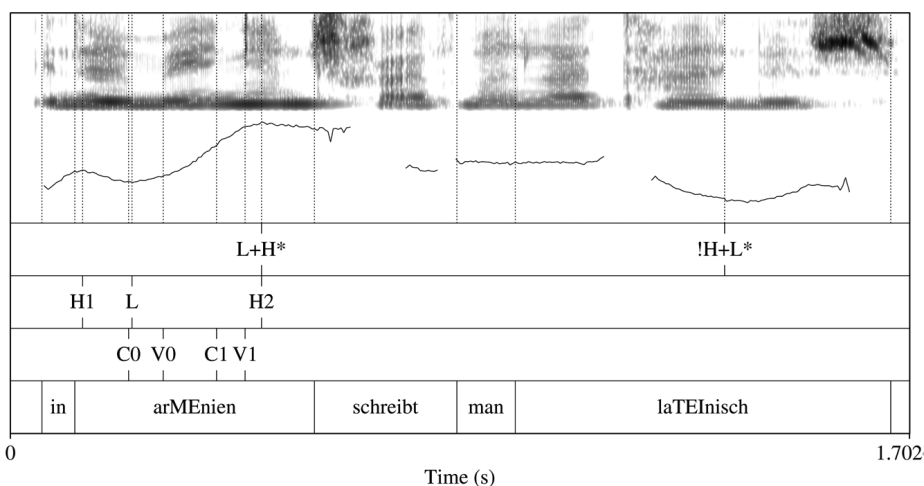
Panel (b)

2.3.2
Qualitative analysis — hat patterns

Overall, 18.6% of the sentences were realized with a hat pattern; hat patterns were not more frequent in contrastive than in noncontrastive contexts ($\chi^2 < 2$). Figure 3 shows a hat pattern in both noncontrastive (upper panel) and contrastive context (lower panel).
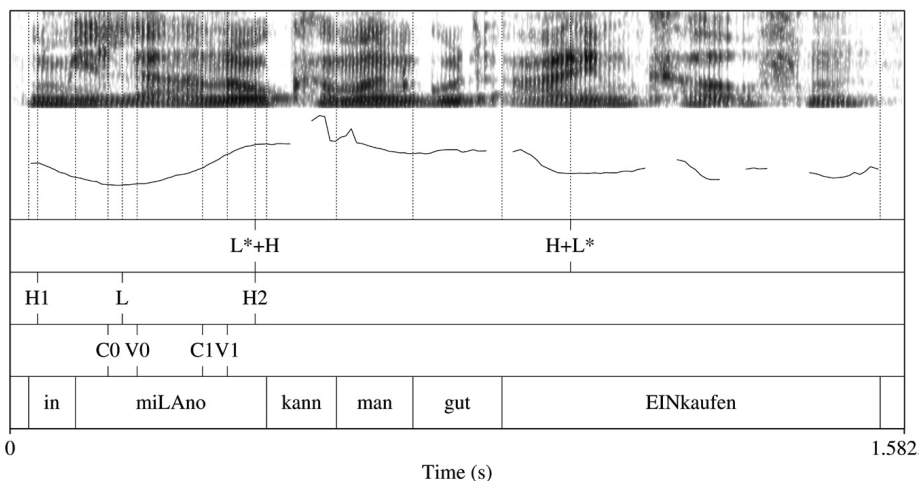
Figure 3

Noncontrastive (panel (a)) and contrastive (panel (b)) realization of the sentence *In Milano kann man gut einkaufen* ('In Milano, shopping is great')

Panel (a)

Panel (b)

2.3.3
Qualitative analysis—rheme accents

The frequency distribution of the rhematic (nuclear) accent is shown in Table 2. There were significantly more falling rheme accents in the data than high ones ($\chi^2 = 7.53$, $df = 1$, $p < .01$). There was no difference in accent distribution in noncontrastive contexts, but there were significantly more falling rheme accents than high ones in contrastive contexts ($\chi^2 = 13.44$, $df = 1$, $p < .01$).

Table 2

Frequency distribution of rheme accents in contrastive and noncontrastive contexts

	Noncontrastive context	*Contrastive context*	*Total*
High accent	42	26	68
Falling accent	44	60	104
Total	86	86	172

2.3.4
Theme accent realization in contrastive and noncontrastive context

Contrast is a within-subjects factor, so paired *t*-tests were calculated initially for the acoustic variables. Using this test, the inherent variability in the experimental items is reduced; only the variability caused by different subjects remains. Because of multiple *t*-tests, the standard significance level of $p = .05$ was adjusted to $p = .004$ (Bonferroni correction[7]). Except for the alignment of the trough, all variables were

---

[7]   The Bonferroni correction only adjusts the significance level based on tests for independent

normally distributed. Therefore, *t*-tests were performed on 15 variables, nine of which differed significantly in contrastive and noncontrastive context. The results of the paired *t*-tests for the different groups of variables are presented in Tables 3 and 4.

Table 3

Means and *SD*s of F0 variables in contrastive and noncontrastive contexts, *t* value, and significance value of a 2-tailed paired *t*-test ($N = 86$). Significance level $p = .004$. Significant differences are highlighted in bold face

	Noncontrastive		Contrastive			
Variable	*mean*	*SD*	*mean*	*SD*	*t*	*p*
F0(H1) in Hz	168.1	51.4	167.1	49.7	.68	.5
F0(L) in Hz	154.9	47.7	151.9	47.8	2.52	.01
F0(H2) in Hz	218.0	78.9	227.7	79.3	3.65	<.0001
ΔF0(fall) in Hz	13.2	9.8	15.2	11.7	1.32	.19
ΔF0(fall) in st	1.4	1.0	1.8	1.4	1.89	.06
ΔF0(rise) in Hz	63.1	40.4	75.8	38.9	5.44	<.0001
ΔF0(rise) in st	5.4	2.2	6.5	1.9	6.87	<.0001

On average, prenuclear peaks in contrastive contexts were 9.7 ± 2.7 Hz higher than peaks in noncontrastive contexts.[8] This was also reflected in the magnitude of the F0 rise, which was 12.7 ± 2.4 Hz larger for contrastive contexts (equivalent to 1.1 ± 0.2 st).

Overall utterance duration did not differ significantly in the two contrast conditions. The duration of the thematic constituent was significantly longer in contrastive contexts than in noncontrastive ones, on average 19.6 ± 4.6 ms. Since the overall utterance duration was the same in the two conditions, the thematic lengthening was at the expense of the rhematic part of the utterance. The duration of the stressed syllable was also significantly longer in contrastive contexts than in noncontrastive ones (8.3 ± 2.7 ms) and so was the duration of the stressed vowel (5.8 ± 1.9 ms). There was no significant difference in the duration of the F0 fall prior to the accentual rise but the duration of the F0 rise took significantly longer in contrastive contexts than in noncontrastive contexts (20.5 ± 5.4 ms). Except for the differences in vowel and syllable duration, the reported duration differences between the two contexts lie in the upper region of reported just noticeable differences (JND) from $10 - 25$ ms (Klatt, 1976).

---

measures, such as the three F0 values for the different landmarks, all temporal variables (6), and the absolute alignment variables (3). The magnitude of the F0 rise and the F0 fall are calculated directly from the static F0 measures. Similarly, the proportional alignment-measures and the slope are only derived variables. The significance level has to be adjusted for 12 variables, resulting in $p = .05/12 = 0.0042$.

[8]   Error values represent the *SE* of the mean.

Table 4

Means and *SD*s of temporal and alignment variables in contrastive and noncontrastive contexts, *t* value, and significance value of a 2-tailed paired *t*-test ($N = 86$). Significance level $p = .004$. Significant differences are highlighted in bold face

Variable	Noncontrastive		Contrastive			
	mean	SD	mean	SD	t	p
utt-dur in ms	1611.6	351.4	1622.4	233.2	.99	.32
theme-dur in ms	436.6	88.2	457.1	88.4	4.30	<.0001
syll-dur in ms	152.7	36.1	161.0	39.2	3.02	.003
vowel-dur in ms	96.4	29.3	102.2	29.9	3.00	.004
Δt(fall) in ms	115.2	59.0	125.5	64.2	1.60	.11
Δt(rise) in ms	178.8	43.8	199.3	42.7	3.77	<.0001
al(H2,V1) in ms	20.9	30.2	42.4	35.0	5.44	<.0001
al(H2,V1)_prop	0.14	0.24	0.30	0.28	5.29	<.0001
slope(rise)	0.35	0.19	0.38	0.18	2.44	.02

The peak, calculated with respect to the start of the poststressed syllable, al(H2,V1), was 20.0 ± 3.7 ms later in contrastive contexts. The proportional alignment variables showed the same effect as the absolute alignment variables. The alignment of the trough was bimodally distributed (see Fig. 9 in the Appendix). One peak in the frequency distribution was located around 18 ms after the onset of the stressed syllable and one around 80 ms after the onset of the stressed syllable. The same distributions hold for contrastive and noncontrastive contexts alike, except that the bimodality is stronger in contrastive contexts. No *t*-tests were performed for the alignment of L.

2.3.5
### Correlation between dependent variables

In contrastive contexts, peaks were realized later than in noncontrastive ones. Whether or not an increase in F0 height was correlated with peak delay and the slope of the rise was analyzed. A correlation between peak height and peak delay would be in line with a strong version of Gussenhoven's view that peak height and peak alignment are adjusted together (Gussenhoven, 2002).

Since the raw data are subject to much variation, such as different F0 levels or speech rates, the correlation analyses are based on comparisons between the contrastive and noncontrastive realization of each utterance pair for every speaker. In other words, the noncontrastive version was taken as a baseline against which the deviation found in contrastive context was calculated. Peak height ratio was computed by dividing the absolute peak height in each contrastive version by the absolute peak height of each corresponding noncontrastive version. Similarly, slope ratio and peak alignment difference were calculated. Peak height ratio did not correlate with peak alignment difference but it correlated slightly with slope ratio (Pearson $r = 0.52$,

p =.047, *N* = 86). Figure 4 shows a scatterplot of peak height ratio plotted against peak alignment difference. It shows that there is a high degree of variation in conveying a contrastive theme (compared to a noncontrastive theme). In the majority of cases (46.5% of the utterance pairs), the peak was both later and higher in contrastive contexts compared to noncontrastive ones. In nearly three quarter of the cases (72.1%) the peak was higher in contrastive contexts than in noncontrastive ones (first and second quadrant), and in nearly two third of the cases (63.9%) the peak was later in contrastive contexts than in noncontrastive ones (second and forth quadrant).

Figure 4

Scatterplot of peak height ratio and peak alignment difference (of the alignment variable al(H2,V1) in ms). Each dot represents an utterance pair produced by a single speaker. To the right to the vertical line, peaks are later in the contrastive versions than in the respective noncontrastive ones; above the horizontal line, peaks are higher for contrastive contexts. Percentages indicate what proportion of the data falls into the respective quadrants
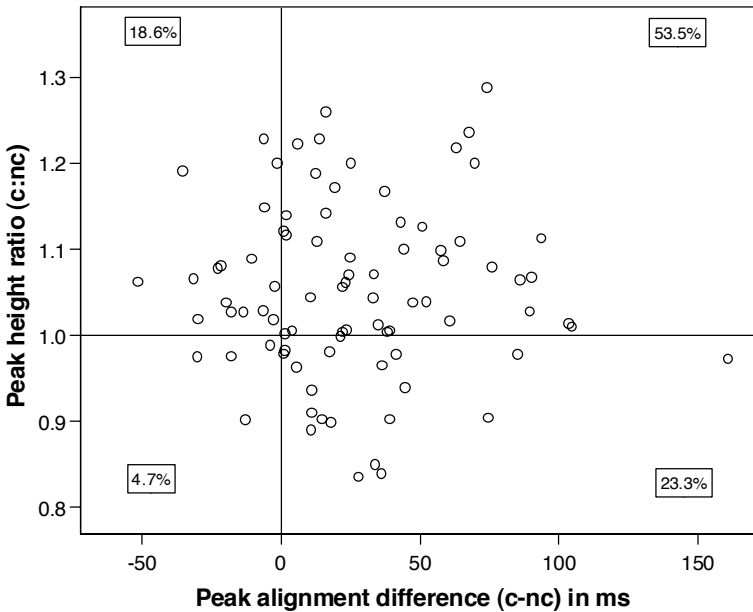
2.3.6
Further analyses

The present analyses are not suited to test whether the scatterplot in Figure 4 shows random variation or whether the use of peak height or peak alignment is conditioned by other factors, such as the dialectal origin of the speakers or the metrical pattern of the sentences. Therefore, the data were analyzed using multilevel modeling (Baayen, 2004; Goldstein, 1995; Quené & van den Bergh, 2004). In this design, fixed and random effects are analyzed on different levels and a factorial design is not necessary. Subjects and lexicalizations were treated as random factors; contrast, gender, region of origin, grammatical relation, metrical structure, phonological vowel length, number

of poststressed syllables, theme accent type, and rheme accent type as fixed factors. It turned out that the only relevant factors were gender, contrast, and theme accent type. Gender naturally affected the height of H1, L, and H2 only. Contrast had a significant effect on the absolute and proportional alignment of the peak, the height of the peak, the F0 excursion of the rise, and the duration of the rise; the magnitude of the difference is comparable to the results of the paired *t*-tests. Peaks were 21.1 ms later in contrastive contexts and on average 9.0 Hz higher, F0 excursion was 12.3 Hz larger in contrastive contexts which corresponds to a difference of 1.1 st. By analyzing more factors in this model, it turned out that segmental duration was not affected by contrast (contrary to the results of the paired *t*-tests). Theme accent type had a significant effect on the alignment of the trough (which was 81.9 ms later for L*+H accents than for L+H* accents), on the duration of the F0 fall and rise, as well as on syllable and vowel duration of the stressed syllable. More specifically, L*+H accents have longer stressed vowels (11.1 ms), longer stressed syllables (20.2 ms), longer F0 falls (78.2 ms) but shorter F0 rises (44.4 ms). Word and utterance duration as well as the slope of the rise were not affected by any of the factors.

2.4
Interim discussion

The production experiment successfully concealed the purpose of the experiment. None of the subjects reported to have noticed that they read identical sentences in different contexts. The use of distractor paragraphs on the same topic as the experimental paragraphs seemed sufficient to disguise the purpose of the experiment. Despite the variability in both subjects and items, significant differences between the acoustic realization of thematic material in contrastive and noncontrastive contexts were found. Themes in contrastive contexts were realized with a higher and later peak and the duration of the F0 rise was longer compared to themes in noncontrastive contexts. Given the results of Atterer and Ladd (2004) it is likely that regional differences existed in our data; however, the statistical power was probably too low to uncover them. The same holds for the other factors, such as metrical stress and phonological vowel length.

The prosodic annotation of thematic accents proved problematic. In GToBI, L*+H is described to have a low target within the accented syllable with the peak located in the next syllable or later. Furthermore, the accented syllable should be perceived as low. For L+H* where the accented syllable is perceived as high, the low target should preceed the accented syllable and the peak should be reached late in the accented syllable (Grice et al., 2005, pp. 65). In the present data—and supposedly in most kinds of prenuclear accents in German—the peak is reached almostly exclusively in the post-tonic syllable (cf. also Atterer & Ladd, 2004). Strictly speaking therefore, bitonal rising accents should be only of the sort L*+H. Perceptually however, the stressed syllable of some of these rising accents sounds high, while in others it sounds low (although this classification is sometimes hard). This was the major criterion for labeling an accent distinction in Experiment 1. The reliability of the intonational labels was not tested directly. However, theme accent type affected various acoustic variables, which indirectly corroborates the labeling. It was shown that L*+H accents had a later trough compared to L+H* accents but also that the vowel and syllable

duration of the stressed syllable were longer in L*+H accents. The lengthening of the stressed syllable, together with a low tonal target in the middle of the stressed syllable certainly contribute to the perception of a low pitch for L*+H accents. An idealized F0 track of the two accents is presented in Figure 5.

Figure 5

Idealized F0 track of a prenuclear L+H* and L*+H accent. The box represents the stressed syllable

L+H*   L*+H
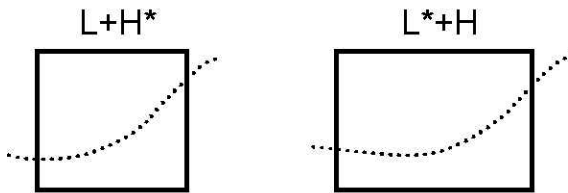
There is hence auditory and acoustic evidence for the presence of two different accent types. The different accent types are not related to contrastive or noncontrastive contexts, though. In two thirds of the cases, the contrastive and noncontrastive versions of a sentence pair were annotated with the same accent type. Acoustically however, prenuclear accent types did differ in contrastive and noncontrastive contexts: themes in contrastive contexts were realized with a later and higher peak. That is, irrespective of theme accent type, peaks in contrastive contexts were later than in noncontrastive contexts. Also, the duration of the F0 rise was longer in contrastive contexts and there was a larger rise excursion. This result runs counter to the finding that the high tonal target in bitonal rising accents is consistently aligned with the segmental structure (Arvaniti et al., 1998; Ladd et al., 1999). The test sentences were structurally and prosodically identical and speakers nevertheless produced a later peak in contrastive contexts. That is, peak alignment may not only be influenced by tonal crowding or other prosodic factors but also by the preceding semantic context (see also Wichmann et al., 2000). The correlation analyses revealed that speakers used either a higher or a later peak or both to mark thematic contrast. This is consistent with the view that peak delay can be a substitute for peak height (see especially Gussenhoven, 2004, pp. 90–92). The scaling and alignment of the low tonal target preceding the F0 rise was not affected by the context.

Note that the alignment of the peak and its height were not affected by a different theme accent type but solely by the context. This raises the question how prenuclear rising accents in German should be described phonologically and more specifically, how the difference in peak realization should be encoded. GToBI, for instance, provides a diacritic for peak delay (e.g., L+H* > or L*+H >) but not (yet) one for peak raising in intonation phrase initial accents.

Contrary to claims in the literature, hat patterns were not more frequent in contrastive than in noncontrastive contexts. This is apparently different in sentences with particular syntactic constructions (Mehlhorn, 2001; Steube, 2001) or when pronouns or adjectives (instead of nouns) are marked with the rising theme accent (see examples in Büring, 1997b). There is, however, an interaction between context

and rheme accent type. Contrastive contexts (i.e., a contrastive theme and rheme) provoked significantly more falling nuclear accents than high ones but high and falling accents were equally distributed in noncontrastive context. It is too early, however, to establish a connection between falling rheme accents and contrastive contexts (in the sense employed here).

In the remainder of this article, it is first tested whether the different productions led to interpretation differences (Experiment 2) and second whether the differences in prosodic realization are annotated with different accent types by a group of linguists.

3 Experiment 2

In Experiment 2 the linguistic interpretation of utterances produced in contrastive and noncontrastive contexts is investigated. This experiment served to test the perceptual significance of the differences between productions in contrastive and noncontrastive contexts.

3.1
Method

Subjects were visually presented with the start of a semantic parallelism and had to decide between two auditorily presented continuations. Semantic parallelisms were also used in the production data but there the two parts were separated by intervening sentences which is not the case here. The auditory stimuli were sentence pairs selected from the production data (see below). With this method we aim to investigate whether listeners can reliably relate an utterance produced in a contrastive context to a visually presented contrastive context.

3.1.1
Participants

Fourteen graduate and postgraduate students at the University of Nijmegen voluntarily took part in the experiment. They were native speakers of German with no known hearing problems. They were between 20 and 30 years old and all originated from Northern Germany (north of the Benrather line). Half of them were male, half of them female. All participants were naïve with respect to the purpose of the experiment. They had not taken part in Experiment 1.

3.1.2
Materials

Ten sentence pairs were pseudorandomly selected from the production data. The contrastive and noncontrastive versions of a sentence pair were from the same speaker each, but overall, sentence pairs from seven speakers were included (for further detail on the sentence pairs see Table 9 in the Appendix). One criterion was that the two versions of each sentence pair differed with respect to either peak alignment, peak height, or both. The respective differences between the contrastive and noncontrastive version are assumed to be above the just noticeable difference (JND) for these

dimensions: peak height ratio was larger than 1.1 st and the alignment difference larger than 50 ms, see Figure 6). Second, the rheme accent type was identical in the two versions of a sentence pair to minimize its effect on subjects' decisions. Seven sentences contained falling rheme accents, three sentences high accents; the unequal number of rheme accent types is due to the frequency distribution of the rheme accents in the production data which is in favor of falling accents.

Figure 6

Peak height ratio and peak alignment difference (al(H2,V1)) for the materials used in Experiment 2 and 3. The scale in this figure is identical to the one used in Figure 4. Sentence numbers refer to Table 9
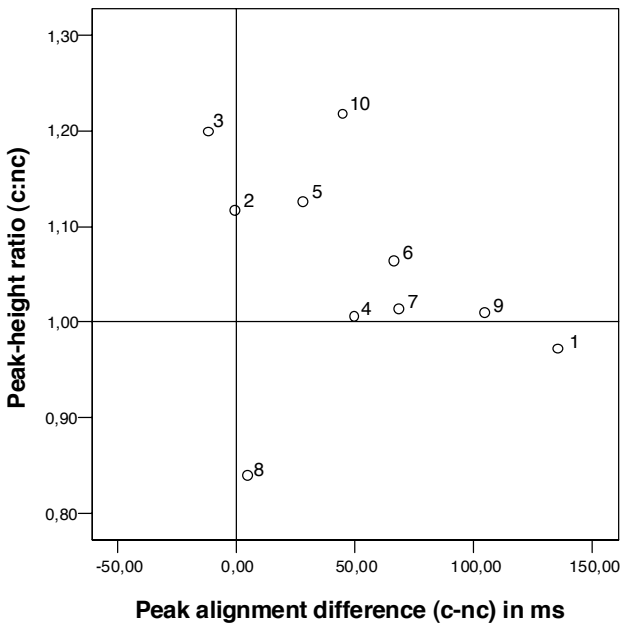
Three further sentence pairs were included for familiarization at the start of the experiment and one at the end. The utterances were presented as the second clause of a semantic parallelism. Semantic parallelisms consist of two clauses where both the thematic part and the rhematic part of the two clauses are contrasted (e.g., In Indonesia, tourism is very important and the Malaysians live from agriculture). For every utterance pair, an appropriate semantic parallelism was constructed (see Table 10 in the Appendix).

3.1.3
Procedure

The experiment was carried out on a PC; the stimuli were presented via headphones (Sony MDR-V150). The first clause of the parallel construction was shown visually on the screen, followed by three dots to indicate the continuation. There were

two loudspeaker symbols on the screen, one for the contrastive version, one for the noncontrastive version. After clicking on a loudspeaker symbol, subjects heard the contrastive or noncontrastive version of the same sentence read by the same speaker.[9] Every sentence pair was presented once.

Subjects were told in writing to read the clause on the screen and then to listen to the two possible continuations to the sentence by clicking on the two loudspeaker symbols on the screen. Their task was to select the version with the more appropriate speech melody. The loudspeaker symbols were labeled 'a)' and 'b)' respectively; participants were asked to tick the more appropriate version on paper or tick 'c)' if they perceived no difference at all between the two versions. Participants could listen to the two possible continuations as often as they wished. The experiment was self-paced. Stimulus presentation was randomized and the order of the contrastive and noncontrastive versions was reversed for half of the subjects.

3.2
Results

In the majority of cases (95.7%), subjects made a choice between the contrastive and noncontrastive version. Only five listeners used the category 'no difference', one subject twice (for Pairs 1 and 8), and four subjects once (for Pairs 2, 4, 5, and 6, respectively).

Contrary to expectation, the presumed contrastive version was not chosen more often than the noncontrastive one (54% overall, see Fig. 7). There were a few sentence pairs (Pairs 4, 7, and 9), however, for which there was a high agreement among listeners as to the more appropriate version. For sentence Pairs 4 and 9, subjects significantly preferred the contrastive version ($\chi^2 = 7.2$, $df = 1$, $p < .05$ and $\chi^2 = 5.78$, $df = 1$, $p < .05$, respectively) and for sentence Pair 7 they preferred the noncontrastive one ($\chi^2 = 8.64$, $df = 1$, $p < .05$).[10]

The pairs with consistent preference patterns (4, 7, and 9) all show a considerable difference in peak alignment between the contrastive and noncontrastive version (as, e.g., shown in Fig. 6). But so do the versions of Pair 10 for which there was no significant preference for any of the versions. Although the alignment differences for the two versions of Pairs 4, 7, and 9 all point to the same direction, listeners preferred the noncontrastive version for Pair 7, but the contrastive ones for Pairs 4 and 9.

To find out to which acoustic differences listeners were most sensitive, we tested for correlation between the number of contrastive responses to an utterance pair

---

[9] We are well aware of the criticism of unnaturalness and the uncontrollable effects of visual presentation, but we believe that parallel constructions are syntactically and semantically sufficiently marked to constrain the degrees of interpretational and intonational freedom. In not presenting the first part of the parallel construction auditorily, we avoid the as yet rather unexplored territory of interclause peak height relations.

[10] All χ^2-analyses for Experiment 2 compensate for the small sample by applying Yates' correction for discontinuity. They are based on an expected frequency of seven (half of the subjects). Because of sporadic choices for "no difference," this is a conservative statistic.

Figure 7

Number of contrastive, noncontrastive and "no-difference" choices for the 10 sentence pairs in Experiment 2. Asterisks mark a significant difference ($p < .05$) between the number of contrastive and noncontrastive responses ($\chi^2 > = 5.78$)
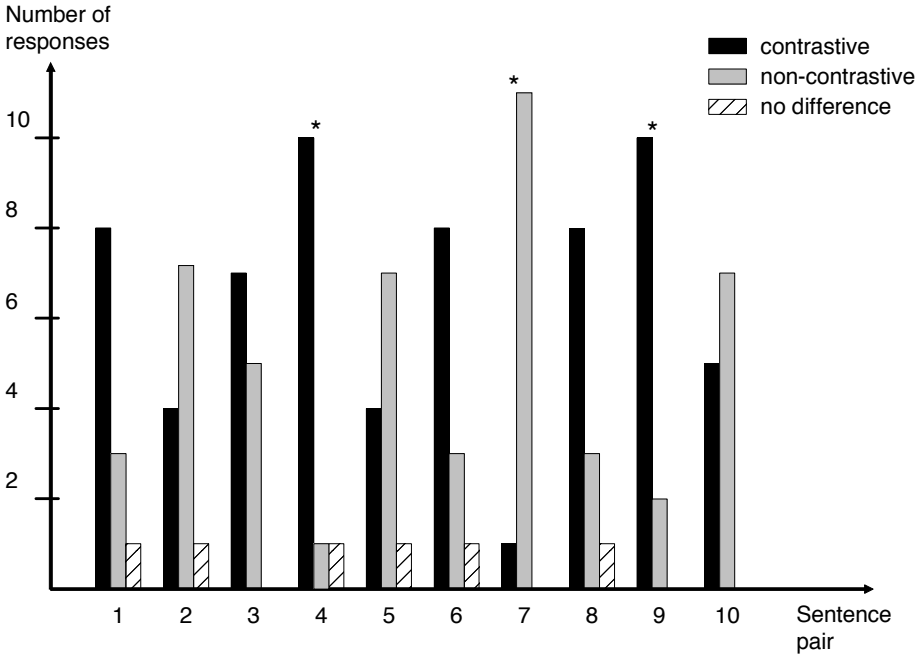
and several acoustic variables that describe the realizational difference between the two versions. Naturally, this does not replace controlled experiments with resynthesized stimuli in which the different acoustic variables can be independently manipulated;[11] these correlation analyzes only provide a first indication about what is important for listeners. The acoustic variables tested were (a) the difference in peak alignment between contrastive and noncontrastive realization, (b) the peak height ratio between contrastive and noncontrastive realization, (c) the F0 excursion ratio between contrastive and noncontrastive realization, and (d) the slope ratio of the rise between contrastive and noncontrastive realization. Using a Spearman's Rho, there was no correlation between the number of contrastive responses and any of these acoustic differences.

---

[11] None of the acoustic variables can be manipulated fully independently of the other ones. For instance, a raised peak needs to be accompanied by either an increase in slope, a raise of the trough, an earlier trough or a later alignment. Therefore, testing the perceptual relevance of these individual cues is a larger project.

3.3
Interim discussion

Subjects' responses indicated that there was no single sentence pair for which the two versions were particularly hard to distinguish. However, subjects' choice for the contrastive or noncontrastive version was at chance level. Although speakers robustly produced a difference, listeners could not map the contrastive versions to the contrastive contexts. Such an asymmetry in production and perception has been discussed under the term 'near-merger' (Labov, Karen, & Miller, 1991). The inconsistent performance might be explained in different ways. First, both the number of subjects and test items was small and most of the read speech samples might not have been distinct enough to signal a contrastive or noncontrastive theme accent unambiguously. More natural data elicitation, however, often results in the production of contrastive or additive particles (e.g., Dimroth, 2002); the use of lexical contrast markers then makes a prosodic comparison between contrastive and noncontrastive themes impossible. Second, there might be speaker specific ways of scaling and aligning sentence-initial pitch accents; since listeners did not have much exposure to the different speakers it is hence hard to interpret the relative acoustic differences. Third, delaying and raising the peak might be part of a very general mechanism of increasing the strength of the first accent. This need not be interpreted in terms of thematic contrast but could point to affective functions, such as friendliness or enthusiasm. Because of the multifunctionality of prosodic information (linguistic, extralinguistic, and paralinguistic) listeners' interpretation is not restricted to the linguistic functions of intonation (see also Andreeva & Barry, 1999; Braun, 2004, 2005). It might also be the case that listeners were sensitive to information other than the realization of the thematic accent, such as slight differences in voice quality, speech rate, and so forth.

Experiment 2 failed to show that naïve listeners can map the contrastive version of a sentence pair to a visually presented contrastive context (semantic parallelism). In Experiment 3 it is tested whether GToBI labelers discriminate prenuclear accents produced in contrastive and noncontrastive contexts.

4 Experiment 3

Annotation of thematic accents by the author revealed that there were two different kinds of thematic accents: L+H* and L*+H. However, the choice of accent type was not affected by context; they were equally distributed in the two conditions. Context rather affected the scaling and alignment of the accentual peak. Acoustic analyses revealed that prenuclear accents in contrastive contexts had later and higher peaks than those in noncontrastive contexts. While delayed peaks can be labeled in GToBI (using the diacritic '<'), there is not yet an established diacritic for intonation phrase initial raised peaks in GToBI (nor is the need for such a diacritic generally acknowledged). The aim of Experiment 3 was to investigate whether linguists trained in prosodic annotation discriminate the prenuclear accents originally produced in contrastive and noncontrastive contexts. To this end, the sentence pairs used in Experiment 2 were annotated by a group of linguists.

It is conceivable that especially the two versions of the sentence Pairs 4, 7, and 9 are annotated more reliably with different accent types than other sentence pairs, since naïve listeners showed a consistent preference for certain versions of these sentences (see Experiment 2); this rests on the assumption, however, that nonlinguists based their judgment on the acoustic differences in the realization of the theme accent only.

4.1
Method

Using a web-interface, linguists trained in using GToBI for intonation annotation labeled the prenuclear (thematic) accents of the sentence pairs used in Experiment 2.

4.1.1
Participants

Eight linguists from different German universities voluntarily participated in the labeling experiment. They had native or near-native command of German and were naïve with respect to the purpose of the experiment. All of them were trained in using GToBI for intonational annotation. No further information about the labelers is available as the experiment was anonymous.

4.1.2
Materials

The stimuli were the same 10 sentence pairs as used in Experiment 2.

4.1.3
Procedure

The experiment was web-based. The 20 stimuli were presented in pseudorandom order, separating the two versions of a sentence pair by at least two other sentences. Labelers annotated the prenuclear accent only. They could listen to the whole utterance or to the preverbal constituent in isolation, with the option to play the parts as often as they wanted. The F0 contour could be inspected by downloading the files. For annotation they could tick a box for the three most probable prenuclear accent types (H*, L+H*, and L*+H) as well as one for "other accent type" that they could specify in a special text field. They were further asked to briefly explain their choice and to rate the certainty of their judgment on a scale from 1 (very uncertain) to 5 (very certain).

4.2
Results

The average certainty was 3.51, ranging from 3.00 to 4.04 (mean values per subject). There were no specific stimuli that made labelers particularly uncertain.

The results of the labeling experiment are summarized in Table 5. To simplify the presentation of the results, the data are first discussed in terms of percentage of same or different labels for the two versions of a sentence pair. Only then the identity of accent types is analyzed.

4.2.1

Frequency of "same" and "different" labels

In half of the cases (52.5%), the two versions of an utterance pair were annotated with the same accent type, which is chance level ("same accent," however, does not mean that every labeler used exactly the same accent types to describe the realizations; as can be seen in Table 5, different annotators used rather different intonational categories to express their percept). As shown in Figure 8, there is some variation in the sentence pairs, however. There were three sentence pairs (Pairs 5, 7, and 9) for which the majority of annotators used different accent types for the contrastive and noncontrastive version (high black bars). Note that these three sentence pairs do not differ considerably from the other sentence pairs in terms of peak height and peak alignment between the contrastive and noncontrastive realization (see Fig. 6).

Figure 8

Number of same and different labels for the 10 sentence pairs in Experiment 3
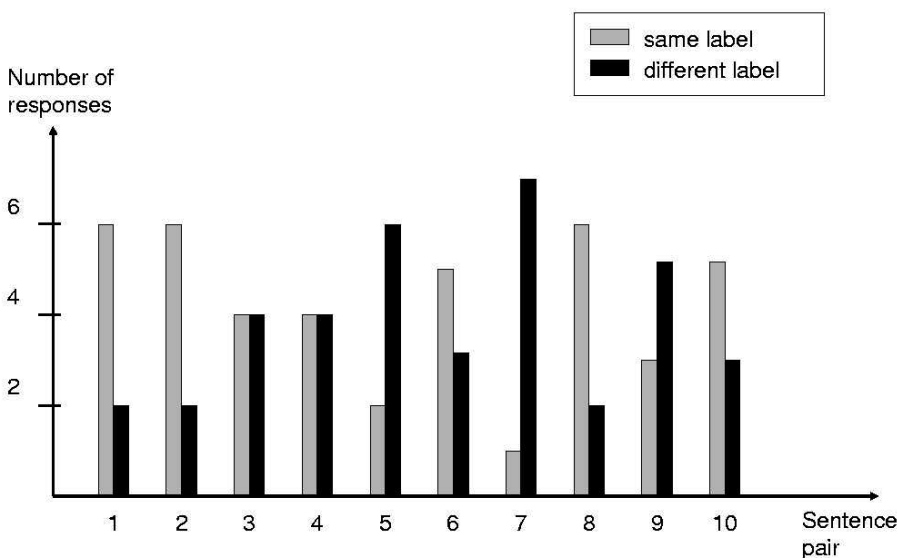
The sentence pairs that obtained most different labels are not exactly the same as the ones that naïve listeners most successfully classified. Listeners' choice was most consistent for sentence Pairs 4, 7, and 9, while most different labels are found for Pairs 5, 7, and 9.

Similar to Experiment 2, correlation analyses were conducted to find out to which acoustic differences the annotators were most sensitive. Correlations were calculated between the number of labelers who annotated the two versions of an utterance pair with a different accent type and (a) the difference in peak alignment between contrastive and noncontrastive realization, (b) the peak height ratio between contrastive and noncontrastive realization, (c) the F0 excursion ratio between contrastive and noncontrastive realization, and (d) the slope ratio of the rise between contrastive and noncontrastive realization. There was only a significant correlation between number

of "different" labels and F0 excursion ratio ($r = 0.85$, $p = .002$, $N = 10$, using the nonparametric Spearman's Rho). Labelers obviously were most sensitive to large or small F0 excursions in the accentual rise; the higher the F0 excursion ratio between contrastive and noncontrastive realization of the prenuclear accent, the more labelers annotated the two rises of a sentence pair with a different accent type. No such correlation existed with peak alignment differences, the genuine basis for intonational contrasts in autosegmental-metrical approaches since Bruce (1977). However, as already discussed in Experiment 2, the stimuli were not optimal to test this hypothesis.

Table 5

Labels of eight linguists for the 10 sentence pairs. Accent differentiations for the contrastive (c) and noncontrastive (nc) version of a pair are highlighted with bold face. Sentence pairs in italics (4, 7, and 9) resulted in significant preference patterns for nonlinguists

sentence pair	Ling1	Ling2	Ling3	Ling4	Ling5	Ling6	Ling7	Ling8
1 (c)	L+H*	L+H*	L+H*	L*+H	L*+H	H*	**L+H***	**L*+H**
1 (nc)							**L*+H**	**L+H***
2 (c)	L+H*	L*+H	L+H*	L*+H	L*+H	**L*+H<**	L*+H	**L*+H**
2 (nc)						**L+H***		**L+H***
3 (c)	**L+H***	L+H*	L+H*	**L*+H**	**L+H***	**H***	L*+H	L*+H
3 (nc)	**L*+H**			**L+H***	**L*+H**	**L+H***		
4 (c)	L+H*	L+H*	L+H*	L*+H	**L+H***	**L+H***	**L*+H**	**L*+H**
4 (nc)					**L*+H**	**H*<**	**L+H***	**L+H***
5 (c)	L+H*	**L*+H**	L+H*	**L*+H**	**L*+H**	**L+H***	**L*+H**	**L*+H**
5 (nc)		**L***		**H***	**H***	**H***	**L+H***	**L+H***
6 (c)	L*+H	L*+H	L+H*	L*+H	L*+H	H*+^H	L*+H	L*+H
6 (nc)	L+H*				L+H*		L+H*	
7 (c)	**L*+H**	**L+H***	L+H*	**L*+H**	**L+H***	**H*+^H**	**L*+H**	**L*+H**
7 (nc)	**L+H***	**L***		**H***	**H***	**H***	**H***	**H***
8 (c)	L+H*	**L*+H**	L+H*	**L*+H**	H*	L+H*	H*	L+H*
8 (nc)		**L+H***		**L+H***				
9 (c)	**L*+H**	L*+H	**L*+H**	L*+H	**H***	**L+H***	L*+H	**L*+H**
9 (nc)	**L+H***		**L+H***		**L+H***	**L+H*<**		**L+H***
10 (c)	L+H*	L+H*	**L+H***	H*	L+H*	L+H*	**L*+H**	**L*+H**
10 (nc)			**H***				**H***	**H***

4.2.2
Accent types

There was a high degree of variation in accent types. When annotators used the same accent type for both the contrastive and noncontrastive version of a sentence pair, there was a pronounced preference for L+H* (22 times), compared to L*+H (15 times), H* (4 times), and H*+^H (once).

Also, when the contrastive and noncontrastive versions were labeled with different accent types, many different accentual distinctions were used. Most frequent was a distinction between L+H* and L*+H (19 times), between L*+H and H* (7 times), and between L+H* and H* (5 times). One could argue that labelers have their own internal "representation" of GToBI-categories, which could still be distinct from other accent types in their particular system but need not fully coincide with the representation of other labelers. Analyzing the results of the labeling experiments for the individual annotators separately, also results in a high degree of variation, though. The first four labelers used the same accent type for annotating the contrastive and noncontrastive version for most of the sentence pairs, while the remaining four linguists used different accent types more often. Linguist 1 and 3 appear to have a strong preference for L+H*, whereas labeler four uses L*+H very frequently. Linguists 7 and 8 seem to be very consistent in annotating the contrastive versions with L*+H and the noncontrastive version with an 'H*-accent' (L+H* or H*).

Often, the explanations labelers provided are not part of the GToBI labeling criteria. Annotators commented, for instance, on the slope of the rise, F0 excursion, and made unusual presuppositions about prenuclear accents.

4.2.3
Consistency in transcription

Most disagreement was found in the use of L*+H, L+H*, and H*. One labeler (Linguist 2) used L* to describe accents that were transcribed as H* or L+H* by other labelers. Unfortunately, he/she didn't comment his/her choice; it was only remarked that no visual information of the F0 contour was used. In the consistency study on GToBI by Grice et al. (1996), most confusions were found between the accent pairs L+H* and H* (28% if expressed as percentage of confusions involving one of the 2 accents), between L*+H and L* (17%), and between L+H* and L*+H (16%). These are the confusions that are most frequent in our data as well. One labeler (Linguist 7) was especially careful and used many diacritics (such as a delayed peak). According to his/her comments, L+H* was used only when the rise was sufficiently steep and the F0 excursion large. For peak delay the diacritic '<' was employed. As becomes clear from Table 5 and Figure 6, this diacritic was not always employed when the peak was delayed. A further particularity of this annotator is that she/he used H*+^H as he/she found a sentence-initial low pitch accent (such as L*+H) unusual in German. GToBI, however, does well allow L*+H pitch accents to occur sentence intially.

4.3
Interim discussion

Labelers were not very consistent in annotating the thematic accents of the 20 sentences. Despite the use of identical materials, it is difficult to directly compare

the results from Experiments 2 and 3 because the tasks were rather different. While listeners in Experiment 2 heard two realizations of the same utterance by the same speaker pair-wise, labelers were presented with one realization at a time. Furthermore, listeners in Experiment 2 had the task to choose the more appropriate version in a given context, while labelers made a choice with respect to a particular accent type. Finally, listeners in Experiment 2 paid attention to the overall contour, while labelers concentrated on the prenuclear accent only.

Since there was a large variation in accent types used, it is conceivable that the accent types are either not sufficiently well defined or are interpreted differently by different annotators. In either case, the criteria for annotating particular accent types need to be clarified. There were three sentence pairs (Pairs 5, 7, and 9) for which the majority of annotators labeled the two versions with different accent types. It turned out that labelers were most sensitive to a small versus a large F0 excursion in the thematic rise. Notably, autosegmental-metrical contrasts are mainly based on alignment differences. MAE-ToBI and GToBI only offer a limited way to annotate differences in pitch range: These are diacritics to indicate that the pitch range is expanded (upstep ^H) or decreased (downstep !H); however, in the current model, these diacritics only describe changes in pitch range relative to an earlier high target and are hence not used for sentence-initial accents (as the ones that were labeled here). The high degree of disagreement in accent types might be partly attributed to the fact that intonation phrase initial accents are not generally marked with a pitch range diacritic. That F0 excursion is important becomes also clear by analyzing the annotators' comments who often mentioned extreme F0 excursions and steep rises. It has to be emphasized that the annotator's task was very difficult because they only heard isolated sentences and could not adjust to the peculiarities of the speakers. Also, speakers came from different regional backgrounds and there was no explicit training phase involved.

Intonational contrasts used in semantic formalisms, such as L*+H versus L+H*, are indeed often employed by annotators to discriminate the realizations in contrastive and noncontrastive contexts. However, the annotation was not very consistent. Further, given the high proportion of identical labels for the contrastive and noncontrastive versions and the large number of different accent distinctions otherwise, a simple 1:1 mapping from present-day GToBI accent types to thematic contrast seems premature.

5 General discussion

In Experiment 1 the realization of thematic accents in contrastive and noncontrastive contexts was compared. This production experiment revealed that sentence-initial themes in contrastive contexts were realized with later and higher peaks as well as larger F0 excursions and longer rise durations than themes in noncontrastive contexts. This result can be treated as very robust given that the statistical power was comparatively low (due to variability in both items and subjects). Although speakers produced different thematic accent types (L+H* and L*+H), these were equally distributed in the two contexts. The hat pattern was not produced very frequently and did not occur more often in contrastive than in noncontrastive contexts. As for rhematic accents,

there were significantly more falling rheme accents than high ones in contrastive contexts while both accent types were equally distributed in noncontrastive contexts. In Experiment 2, naïve listeners judged which of two versions (originally recorded in a contrastive and noncontrastive context) sounded more appropriate in a visually presented contrastive context. This task proved rather difficult and listeners performed mostly at chance level. Only for three (out of 10) sentence pairs, their choices were consistent: for two sentence pairs, they preferred the contrastive recording and for one sentence pair, they preferred the noncontrastive recording. In Experiment 3, a group of linguists annotated the prenuclear accents of the 20 sentences used in Experiment 2 (i.e., the contrastive and noncontrastive versions of 10 sentence pairs). The individual sentences were presented in isolation, not pair-wise. Overall, in half of the sentence pairs the two versions were annotated with the same accent type. For three sentence pairs, however, the majority of annotators labeled different accent types for the contrastive and noncontrastive version. Post hoc analyses revealed that in these cases, the two versions are characterized by a strong difference in F0 excursion.

Although the main focus of this paper lies in the realization of thematic accents, we start with a discussion of the hat pattern as this contour has been frequently associated with thematic contrast. Then we turn to the phonetics and phonology of thematic accents, which also incorporates a discussion of the labeling experiment. Overall, no strong conclusions can be drawn from Experiment 2 as the sample size was small and the results rather inconclusive.

5.1
Hat pattern

The results of the production experiment (Experiment 1) showed that — contrary to general belief — hat patterns were not produced very frequently and, more importantly, they were not realized more often in contrastive than in noncontrastive contexts. Hence their presence constitutes neither a sufficient nor a necessary condition for thematic contrast. So, why are hat patterns so often ascribed to contrastive themes in the literature? We can think of two explanations, one is phonetic in nature, the other concerns the materials used. As there are significantly more falling rheme accents (H+L*, !H+L*) in contrastive contexts than high ones (H* L−, !H* L−, L+H* L−, L+!H* L−), it is conceivable that for some linguists, a falling rhematic accent is equivalent to a hat pattern (in which F0 should stay high between the two accents, however). We are not aware of any studies in German that test for the perceptual difference between sentences in which F0 drops or does not drop between two accents. The other explanation for the discrepancy between our data and previous claims about hat patterns lies in the types of sentences used. In the examples found in the literature (that report hat patterns), the thematic pitch rise is often realized on parts-of-speech that only rarely receive an accent, such as quantors (*Alle Politiker sind NICHT korrupt*, 'Not all politicians are corrupt', example with scope inversion), pronouns (*Ich war ZUHAUSE*, 'I was at home'), verbal infinitives (*Regnen tut es NICHT*, 'It doesn't rain'), or adjectives (*Die weiblichen Popstars trugen KAFTANE*, 'The female pop stars wore caftans') — in these examples bold face marks the thematic rise, capitals the rhematic fall. Clearly, the sentences cited in the literature strongly lend themselves to a contrastive interpretation; this interpretation, however, is

probably not caused by the hat pattern itself but by a particular accent distribution (i.e., accenting preverbal quantors, pronouns, verbal infinitives, or adjectives). In these examples, the same semantic function is probably achieved by a double peak pattern instead of a hat pattern. In the sentences studied in this paper, on the other hand, the thematic accent was either associated with the noun of a preverbal noun phrase or with the temporal or locative nominal adverbial of a prepositional phrase. Accenting these constituents does not give rise to a contrastive interpretation in the same way as accenting pronouns, quantifiers, or adjectives does.

5.2
Phonetics and phonology of thematic accents

The production experiment showed that sentence-initial themes in contrastive contexts are distinguished from those in noncontrastive contexts by subtle phonetic differences, manifested in peak height and alignment, F0 excursion and duration of the prenuclear rise. These acoustic differences are especially significant, given (a) that there was a lot of variability in the items, so that statistical power was reduced and (b) that readers were not aware of reading identical sentences in pragmatically different contexts. The results support some but not all of the impressionistically ascribed differences between contrastive and noncontrastive themes. It was confirmed that contrastive themes are realized with a significantly larger pitch excursion, but it could not be established that the rise starts from a lower trough in contrastive contexts.

Speakers produced two kinds of prenuclear accents, L+H* and L*+H (all prenuclear rises unambiguously had both a low and a high tonal target). L+H* was labeled when the stressed syllable was perceived as high, while L*+H was labeled when the stressed syllable sounded low. This is not totally in line with standard GToBI criteria which take the position of the tonal targets into account as well (Grice et al., 2005). According to this description, the peak in L+H* accents should lie within the accented syllable which is only very rarely the case for prenuclear accents in German. That is, the perceptual and alignment criteria for L+H* do not match for prenuclear accents (while they do for nuclear accents). It certainly would be useful to establish criteria that are of direct relevance for labeling intonation phrase initial accents and such that are secondary only. Possibly, the difference between nuclear and prenuclear accents needs to be re-established in AM models; the different factors that influence peak alignment in nuclear and prenuclear accents (Schepman et al., 2006) could be seen as a further argument for this distinction. Generally, speakers' choice of thematic accent was not influenced by the pragmatic context; L+H* and L*+H were equally distributed in both contexts. Moreover, in almost two thirds of the cases, the contrastive and noncontrastive version of a sentence pair was annotated with the same accent type. Themes in contrastive and noncontrastive contexts do not seem to be distinguished by different accent types but by acoustic differences.

The low tonal target at the start of the F0 rise was not affected by the context but by the type of thematic accent. In L+H* accents the low tonal target was significantly earlier than in L*+H accents. On the other hand, the high tonal target at the end of the rise was not affected by prenuclear accent type but was used only for expressing contrast: the peak was moved further to the right and / or realized higher in contrastive contexts compared to noncontrastive contexts. This effect was observed

across speakers from different backgrounds, different sentence types, and different segmental structures. These factors could not be shown to affect the result significantly which might be due to the low statistical power of the production experiment. As it stands, the results suggest that peak height and peak alignment can be used (and are used) to express functional differences. The presence of different strategies to achieve the same communicative function is reminiscent of trading relations found in segmental phonetics, such as the perceptual equivalence of different cues to stop consonant perception (Best, Morrongiello, & Robson, 1981; Fitch, Halwes, Erickson, & Liberman, 1980), for a review see Repp (1982). Recently, Gussenhoven suggested that "peak delay can […] be used as an enhancement of, or even a substitute for, pitch rising" (Gussenhoven, 2002, p. 52). Gussenhoven's effort code provides a new framework to interpret intonational meaning.

5.3
How to integrate the difference between contrastive and noncontrastive themes in AM phonology?

Thematic contrast was encoded by different prosodic means (peak height and peak alignment), which is hard to reconcile with a simple distinction between different accent types, one for contrastive and one for noncontrastive themes (as assumed in semantic formalisms or by some phonologists). It is conceivable, however, that contrastive and noncontrastive contexts are not distinguished by a different accent type but merely by a shift in prominence (or local effort) of the thematic accent with respect to the rhematic one. This interpretation is consistent with the finding of Mehlhorn (2001) who reported that thematic constituents in contrastive contexts are perceived as more prominent than the corresponding rhematic constituents. Increased prominence of the thematic accent can be achieved by a variety of means, such as higher or later peaks (Arvaniti & Garding, to appear; Gussenhoven, 2002), an increase in the rate of pitch change (Bartels & Kingston, 1994), by lowering the trough before the rise (Jacobs, 1997), and/or by durational differences (Arvaniti & Garding, to appear, among others). This is not to say that later and higher peaks could not both be used to signal a phonological contrast; but this phonological contrast cannot be described (annotated) easily with the accent type distinctions currently available. One could imagine a feature [+ emph] that is attached to the rising L*+H or L+H* accent (in analogy to, for example, the feature "delayed peak" suggested by Ladd, 1983). Its phonetic implementation would involve delaying and/or raising the peak while leaving the low tonal target unaltered.

That accent type distinctions in AM phonology cannot be based on both alignment and scaling differences may be one reason why the results of the labeling experiment were so inconclusive (It has to be acknowledged, however, that the labelers' task was very challenging as they had to annotate prenuclear accents in isolated sentences produced by a different speaker each. This is more difficult than labeling the prenuclear accents of an identical sentence produced by the same speaker pair-wise (as done in Experiment 1) since in this setting the annotator can zoom in better on the differences). It seemed that annotators were most sensitive to small versus large F0 excursions in the prenuclear rise. Increased peak height has been associated with perceived prominence (Kohler & Gartenberg, 1991; Rietveld & Gussenhoven, 1995;

Terken, 1991, among others), emphasis, surprise/exceptionality (Ladd & Morton, 1997), and politeness (Chen & Gussenhoven, 2002); all functions that are gradual in nature. Abstracting away from the dichotomy contrast versus noncontrast, sentences with contrastive and noncontrastive themes are all part of hierarchical discourse structuring, which is marked by prosodic means (Ayers, 1994; Brown & Yule, 1983; Grosz & Hirschberg, 1992; Noordman, Dassen, Swerts, & Terken, 1999; Shriberg, Stolcke, Hakkani-Tur, & Tur, 2000; Sluijter & Terken, 1993; Wichmann, House, & Rietveld, 2000). Contrary to the binary function of thematic contrast marking, discourse structuring is less categorical (Chafe, 1994), so prosodic marking need not be necessarily binary: increased speaker effort is likely to be expended to support the listener in retrieving the discourse structure (Lindblom, 1990). For the listener, it is no doubt more difficult to activate a new concept or a concept different from the previous one than to simply add new information to the current theme (Chafe, 1994). From that perspective, the dichotomy contrast versus noncontrast has to be seen as a special case, one that is operationally definable and of semantic interest, but one that is embedded in the overall discourse structure.

Although the perception and labeling experiment failed to show that naïve listeners and phonologists can reliably tell the difference between thematic constituents read in contrastive and noncontrastive contexts, speakers encode this semantic distinction very robustly by fine prosodic detail. It was shown that this semantic distinction is not marked by different accent types but by a variety of acoustic cues (such as peak height and alignment). These differences are not easily captured by current AM descriptions of German. As they serve to increase the prominence of the thematic accent, a feature [+ emph] is proposed that is associated with peak raising and delay.

References

ANDREEVA, B., & BARRY, W. J. (1999). Intonation von checks in der Sofia Varietät des Bulgarischen. *Research Report Phonetics*, *University of the Saarland (PHONUS)*, **4**, 1–14.

ARVANITI, A., & GARDING, G. (to appear). Dialectal variation in the rising accents of American English. In J. Hualde & J. Cole, (Eds.), *Papers in laboratory phonology 9*. Berlin, New York: Mouton de Gruyter.

ARVANITI, A., LADD, D. R., & MENNEN, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, **26**(1), 3–25.

ATTERER, M., & LADD, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics*, **32**(3), 177–197.

AYERS, G. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics*, **44**, 1–49.

BAAYEN, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers in Psycholinguistics, Edmonton*, 1–45.

BARTELS, C., & KINGSTON, J. (1994). Salient pitch cues in the perception of contrastive focus. In P. Bosch & R. van der Sandt, (Eds.), *Focus and natural language processing*, vol. **1** (Intonation and syntax), 1–10, IBM Deutschland Informationssystem GmbH, Working Paper 6.

BECKMAN, M., & AYERS, G. M. (1997). *Guidelines for ToBI Labelling*. Tech. rep., Ohio State University. <http://www.ling.ohio-state.edu/research/phonetics/E_ToBI>.

BECKMAN, M. E., HIRSCHBERG, J., & SHATTUCK-HUFNAGEL, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun, (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). Oxford: Oxford University Press.

BEST, C. T., MORRONGIELLO, B., & ROBSON, R. (1981). Perceptual equivalence of acoustic cues in speech and non-speech perception. *Perception and Psychophysics*, **29**, 191–211.

BRAUN, B. (2004). Answers to the perception of thematic contrast and questions concerning the perception of thematic "non-contrast." In *Proceedings of the 2nd International Conference on Speech Prosody*, 685–688, Nara, Japan.

BRAUN, B. (2005). *Production and perception of thematic contrast in German*. Contemporary Series in Descriptive Linguistics, vol. **9**, Peter Lang Publishing.

BROWN, G., & YULE, G. (1983). *Discourse analysis*. Cambridge, U.K.: Cambridge University Press.

BRUCE, G. (1977). Swedish word accents in sentence perspective. *Travaux de l'Institut de linguistique de Lund*, **12**, Gleerup, Lund.

BRUCE, G., & GÅRDING, E. (1978). A prosodic typology for Swedish dialects. In E. Gårding, G. Bruce, & R. Bannert, (Eds.), *Nordic prosody*. Papers from a Symposium.

BÜRING, D. (1997a). The great scope inversion conspiracy. *Linguistics and Philosophy*, **20**, 175–194.

BÜRING, D. (1997b). *The meaning of Topic and Focus—the 59th Street Bridge Accent*. London: Routhledge.

CASPERS, J., & van HEUVEN, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica*, **50**, 161–171.

CHAFE, W. L. (1994). *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.

CHEN, A., & GUSSENHOVEN, C. (2002). Language-specific uses of the effort code. In *Proceedings of the 1st International Conference on Speech Prosody* (pp. 215–218), Aix-en-Provence, France.

COHEN, A., & t'HART, J. (1967). On the anatomy of intonation. *Lingua*, **19**, 177–192.

DILLEY, L. C., LADD, D. R., & SCHEPMAN, A. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics*, **33**, 115–119.

DIMROTH, C. (2002). Topics, assertions, and additive words: How L2 learners get from information structure to target-language syntax. *Linguistics*, **40**(4), 891–923.

FÉRY, C. (1993). *German intonational patterns*, vol. **285** of Linguistische Arbeiten. Niemeyer, Tübingen.

FIRBAS, J. (1964). On defining the theme in functional sentence analysis. *Travaux Linguistiques de Prague*, **1**, 267–280.

FITCH, H. L., HALWES, T., ERICKSON, D. M., & LIBERMAN, A. M. (1980). Perceptual equivalence of acoustic cues for stop-consonant manner. *Perception and Psychophysics*, **27**, 343–350.

FOX, A. (1984). *German intonation: An outline*. Oxford: Clarendon Press.

GILLES, P., & PETERS, J. (Eds.) (2004). *Regional variation in intonation*. Linguistische Arbeiten, Niemeyer, Tübingen.

GOLDSTEIN, H. (1995). *Multilevel statistical models*. New York: Halstead Press.

GRABE, E. (1998). *Comparative intonational phonology: English and German*. Ph.D. thesis, Katholieke Universiteit Nijmegen, The Netherlands.

GRICE, M., BAUMANN, S., & BENZMÜLLER, R. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun, (Ed.), *Prosodic typology*, (pp. 55–83), Oxford: Oxford University Press.

GRICE, M., REYELT, M., BENZMÜLLER, R., MAYER, J., & BATLINER, A. (1996). Consistency in transcription and labelling of German intonation with GToBI. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)*, 1716–1719, Philadelphia, U.S.A.

GROSZ, B., & HIRSCHBERG, J. (1992). Some intonational characteristics of discourse structure. *In Proceedings of the 2nd International Conference on Spoken Language Processing*, 429–432, Banff.

GUSSENHOVEN, C. (2002). Intonation and interpretation: Phonetics and phonology. In *Proceedings of the 1st International Conference on Speech Prosody*, 47–58, Aix-en-Provence, France.

GUSSENHOVEN, C. (2004). *The phonology of tone and intonation*. Research Surveys in Linguistics, Cambridge: Cambridge University Press.

HERMAN, R., & McGORY, J. T. (2002). The conceptual similiarity of intonational tones and its effect on intertranscriber reliability. *Language and Speech*, **45**(1), 1–36.

ISAČENKO, A. V., & SCHÄDLICH, H. J. (1966). Untersuchungen über die deutsche Satzintonation (Investigations on German sentence intonation). *Studia Grammatika*, **VII**, 7–64.

JACKENDOFF, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

JACOBS, J. (1996). Bemerkungen zur I-Topikalisierung (Remarks on l-Tropicalization). *Sprache und Pragmatik*, **41**, 1–48.

JACOBS, J. (1997). I-Topikalisierung (l-Tropicalization). *Linguistische Berichte*, **168**, 91–133.

KLATT, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**(5), 1208–1221.

KOHLER, K. J. (1991a). A model of German intonation. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, **25**, 295–360.

KOHLER, K. J. (1991b). Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, **25**, 115–185.

KOHLER, K. J., & GARTENBERG, R. (1991). The perception of accents: F0 peak height versus F0 peak position. A*rbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, **25**, 219–242.

KÖNIG, W. (1994). *dtv-Atlas Deutsche Sprache* (dtv-Atlas German) (revised ed.). Deutscher Taschenbuch Verlag.

KRIFKA, M. (1998). Scope inversion under the rise-fall contour in German. *Linguistic Inquiry*, **29**(1), 75–112.

KRIFKA, M. (1999). Additive particles under stress. In *Proceedings from Semantic and Linguistic Theory SALT 8*. Cornell: CLC Publications.

LABOV, W., KAREN, M., & MILLER, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, **3**, 33–74.

LADD, D. R. (1983). Phonological features of intonational meaning. *Language*, **59**, 721–759.

LADD, D. R., FAULKNER, D., FAULKNER, H., & SCHEPMAN, A. (1999). Constant "segmental anchoring" of f0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, **106**(3), 1543–1554.

LADD, D. R., MENNEN, I., & SCHEPMAN, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, **107**(5), 2685–2696.

LADD, D. R., & MORTON, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, **25**, 313–342.

LADD, D. R., & SCHEPMAN, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, **31**, 81–112.

LIBERMAN, M., & PIERREHUMBERT, J. B. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle, (Eds.), *Language sound structure* (pp. 157–233), MIT Press.

LINDBLOM, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. J. Hardcastle and A. Marchal, (Eds.), *Speech production and speech modeling* (pp. 403–439), Dordrecht, Kluwer Academic Publisher.

MEHLHORN, G. (2001). Produktion und Perzeption von Hutkonturen im Deutschen (Production and perception of hat patterns in German). *Linguistische Arbeitsberichte*, **77**, 31–57.

MOULTON, W. G. (1962). *The sounds of English and German*. Chicago: University of Chicago Press.

NOLAN, F. (1995). The effect of emphasis on declination in English intonation. In J. W. Lewis, (Ed.), *Studies in general and English phonetics* (pp. 241–254), London: Routledge.

NOORDMAN, L., DASSEN, I., SWERTS, M., & TERKEN, J. (1999). Prosodic markers of text structure. In K. van Hoek, A. Kibrik, & L. Noordman, (Eds.), *Discourse studies in cognitive linguistics*, Selected papers of the 5th International Cognitive Linguistics Conference, (pp. 133–148). Amsterdam: John Benjamins.

PETERS, J. (1999). The timing of nuclear high accents in German dialects. In *Proceedings of the 14th International Congress of the Phonetic Sciences* (pp. 1877–1880). San Francisco, U.S.A.

PETERS, J. (2004). *Regionale Variation der Intonation des Deutschen. Studien zu ausgewählten Regionalsprachen* (Regional variation in German intonation. Studies on selected regional varieties). Habilitationsschrift, Universität Potsdam/University of Nijmegen.

PHEBY, J. (1975). *Intonation und Grammatik im Deutschen* (Intonation and Grammar in German). Berlin: Akademie-Verlag.

PITRELLI, J., BECKMAN, M., & HIRSCHBERG, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP)*, vol. **2**, 123–126, Japan: Yokohama.

PREVOST, S. (1995). *A semantics of contrast and information structure for specifying intonation in spoken language generation*. Ph.D. thesis, University of Pennsylvania.

PRIETO, P. (2004). The search for phonological targets in the tonal space: H1 scaling and alignment in five sentence-types in Peninsular Spanish. In T. Face, (Ed.), *Laboratory approaches to Spanish phonology* (pp. 29–59). The Hague: Mouton de Gruyter.

PRIETO, P., D'IMPERIO, M., & GILI-FIVELA, B. (2005). Pitch accent alignment in Romance: Primary and secondary associations with metrical structure. *Language and Speech* (Special issue on variation in intonation, Ed. by Paul Warren), **48**(4), 359–396.

PRIETO, P., van SANTEN, J., & HIRSCHBERG, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, **23**, 429–451.

QUENÉ, H., & van den BERGH, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, **43**, 103–121.

REPP, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, **92**, 81–110.

RIETVELD, A. C. M., & GUSSENHOVEN, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, **13**, 299–308.

RIETVELD, T., & GUSSENHOVEN, C. (1995). Aligning pitch targets in speech synthesis: Effects of syllable structure. *Journal of Phonetics*, **23**, 375–385.

SCHEPMAN, A., LICKLEY, R., & LADD, D. R. (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics*, **34**(1), 1–28.

SHRIBERG, E., STOLCKE, A., HAKKANI-TUR, D., & TUR, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, **32**, 127–154.

SILVERMAN, K. E. A., BECKMAN, M. E., PITRELLI, J., OSTENDORF, M., WIGHTMAN, C., PRICE, P., PIERREHUMBERT, J. B., & HIRSCHBERG, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP)* (vol. **2**, pp. 867–870), Banff, Canada.

SILVERMAN, K. E. A., & PIERREHUMBERT, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman, (Eds.), *Papers in laboratory phonology I* (pp. 72–106), Cambridge, U.K.: Cambridge University Press.

SLUIJTER, A. M. C., & TERKEN, J. M. B. (1993). Beyond sentence prosody: Paragraph intonaton in Dutch. *Phonetica*, **50**, 180–188.

STEEDMAN, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, **31**(4), 649–689.

STEELE, S. A. (1986). Nuclear accent F0 peak location: Effects of rate, vowel, and number of following syllables. *Journal of the Acoustical Society of America*, **80**.

STEUBE, A. (2001). Grammatik und Pragmatik von Hutkonturen (Grammar and pragmatics of hat patterns). *Linguistische Arbeitsberichte*, **77**, 7–29.

TALKIN, D. (1995). A robust algorithm for pitch tracking (RAPT). In W. B. Klein & K. K. Palival, (Eds.), *Speech coding and synthesis*, Elsevier.

TERKEN, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, **89**(4), 1768–1776.

THEUNE, M. (1999). Parallelism, coherence, and contrastive accent. In *Proceedings of the 6th European Conference on Speech Communication and Technology* (vol. **1**, pp. 555–558), Budapest, Hungary.

TIMM, U. (2003). *Rot* (Red). Deutscher Taschenbuch Verlag.

UHMANN, S. (1991). *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie* (Focus phonology. An analysis of German intonation patterns within non-linear phonology). Tübingen: Niemeyer.

UMBACH, C. (2001). Relating contrast and contrastive topic. In I. Kruijff-Korbayova (ed.), *Proc. ESSLLI Workshop on Information Structure, Discourse Structure, and Discouse Semantics*, Helsinki, Finland.

Von ESSEN, O. (1964). *Grundzüge der hochdeutschen Satzintonation* (Basic lines of high German sentence intonation) (2nd ed.). Ratingen: Henn.

WEBER, A., GRICE, M., & CROCKER, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, **99**(2), B63–B72.

WEBER, A., & MÜLLER, K. (2004). Word order variation in German main clauses: A corpus analysis. In *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland.

WICHMANN, A., HOUSE, J., & RIETVELD, T. (2000). Discourse constraints on F0 peak timing in English. In A. Botinis, (Ed.), *Intonation: Analysis, modelling and technology, text, speech, and language technology* (vol. **15**, pp. 163–182), Dordrecht/Boston/London: Kluwer.

WUNDERLICH, D. (1991). Intonation and contrast. *Journal of Semantics*, **8**, 239–251.

ZEEVAT, H. (2004). Contrastors. *Journal of Semantics*, **21**(2), 95–112.

Appendix

Table 6

Original sample noncontrastive and contrastive paragraph. The target sentences are highlighted in bold face

Noncontrastive context	Contrastive context
Malaysien ist vielen Europäern weit-gehend unbekannt. Das Land besteht aus zwei Inseln. Zur Vereinfachung der Kommunikation zwischen den beiden Inselteilen besitzt fast jeder Haushalt einen Computer mit Internetverbindung. Trotzdem ist Malaysien kein hochtech-nologisches Land. **Die Malayen leben von der Landwirtschaft.** Sie sind zwar nicht reich, aber auch nicht besonders arm.	Malaysia und Indonesien sind Nachbarländer im Südchinesischen Meer. Trotz dieser geographischen Nähe unter-scheiden sich die Lebensgewohnheiten der Malayen und Indonesier stark. In Indonesien spielt der Tourismus eine grosse Rolle. Dieser Sektor bietet viele Arbeitsplätze. **Die Malayen leben von der Landwirtschaft.** Vor allem malaysischer Reis ist in ganz Asien bekannt.

Table 7

Information about the speakers. Speakers in brackets were excluded from analysis. The classifica-tion into northern and southern German is based on König (1994) and is the same classification as used by Atterer and Ladd (2004). Northern German speakers originate all from north of the Benrather line dividing broadly Northern dialects from Central and Southern ones

Speaker	Origin	Region	Sex
BK	Lower Saxony	North	female
CS	Lower Saxony	North	female
UB	Lower Saxony	North	female
(SZ)	North Rhine-Westphalia	North	female
IB	Baden-Wuerttemberg	South	female
(CH)	Bavaria	South	female
PK	Bavaria	South	female
JL	Palatinate	South	male
CZ	Hesse	North	male
MB	Berlin	North	male
DS	North Rhine-Westphalia	North	male
(SV)	Schleswig-Holstein	North	male

Table 8

Target sentences with information about the phonological vowel length of the thematic stress, the grammatical relation of the theme (subject-NP or PP), and the position of the stressed syllable in the theme. Stressed syllables are marked in bold face

| *Target sentence*	*long vowel*	*gramm. relation*	*pos. of stressed syllable*
Ital**ie**ner sind sehr **gast**freundlich. *'Italians are very hospitable'*	yes	NP	3 (penultimate)
Die Ka**na**ren sind ein **Wan**derparadies. *'The Canaries are a paradise for hiking'*	yes	NP	3 (penultimate)
Die Ma**lay**en leben von der **Land**wirtschaft. *'The Malaysians live from agriculture'*	yes	NP	3 (penultimate)
Die **Rö**mer waren sehr organi**siert**. *'The Romans were very organized'*	yes	NP	2 (penultimate)
Die **Ma**ler arbeiten viel im **Frei**en. *'Painters often work outside'*	yes	NP	2 (penultimate)
Mar**le**ne spielt Kla**vier** und kann **sin**gen. *'Marlene plays the piano and can sing'*	yes	NP	2 (penultimate)
Die **Lä**mmer haben **Angst** vor Menschen. *'Lambs fear humans'*	no	NP	2 (penultimate)
In Mi**la**no kann man gut **ein**kaufen. *'In Milano, shopping is great'*	yes	PP	3 (penultimate)
In Ar**me**nien schreibt man la**tei**nisch. *'In Armenia, the Latin alphabet is used'*	yes	PP	3 (penultimate)
In A**me**rika besitzen **vie**le eine **Wa**ffe. *'In America, many people own a weapon'*	yes	PP	3 (penultimate)
In **Bay**ern beginnen die **Fe**rien Anfang **Ju**li. *'In Bavaria, holidays start early in July'*	yes	PP	2 (penultimate)
Im **Ja**nuar ist es **fros**tig. *'In January, it is frosty'*	no	PP	2 (penultimate)

Table 9

Stimulus sentences used in Experiments 2 and 3. Translations are provided in Table 8

Sentence number	Speaker	Sentence	Rheme accent
1	BK	In Armenien schreibt man lateinisch.	falling
2	BK	In Milano kann man gut einkaufen.	falling
3	PK	Die Lämmer haben Angst vor Menschen.	falling
4	CS	In Bayern beginnen die Ferien Ende Juli.	falling
5	CS	In Milano kann man gut einkaufen.	high
6	DS	Im Januar ist es frostig.	falling
7	DS	Die Römer waren sehr organisiert.	high
8	IB	Italiener sind sehr gastfreundlich.	falling
9	JL	Die Lämmer haben Angst vor Menschen.	falling
10	UB	Italiener sind sehr gastfreundlich.	high

Figure 9

Histogram of the alignment of the trough relative to the start of the stressed syllable (in ms)
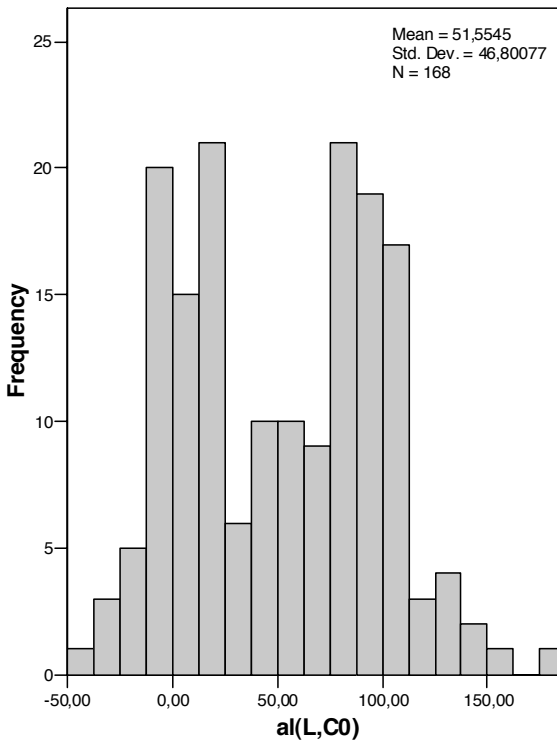
Table 10

List of semantic parallelism used in Experiment 2. Left column shows the visually presented sentence starts (and an English translation), right column the respective target sentence (for translations see Table 8). Numbers in brackets indicate the sentence number (see Table 9)

Start of sentence	*Target sentence*
Die Georgier haben eine eigene Schrift, und *'Georgians have their own writing system'*	In Armenien schreibt man lateinisch. (1)
Rom ist kulturell interessant, und *'Rome is culturally interesting'*	In Milano kann man gut einkaufen. (2 and 5)
Fohlen haben sich gut an Menschen gewöhnt, aber *'Foals have become used to people'*	Die Lämmer haben Angst vor Menschen. (Sentences 3 and 9)
Die Saarländer starten früh in die Sommerferien, aber *'People from the Saarland have early summer holidays'*	In Bayern beginnen die Ferien Ende Juli. (4)
Der December ist oft vergleichsweise mild, aber *'In December it is often relatively mild'*	Im Januar ist es frostig. (6)
Die Germanen waren ein wilder Haufen, aber *'The Germanic tribes were wild and uncivilized'*	Die Römer waren sehr organisiert. (7)
Dänen sind lieber für sich allein, aber *'Danes prefer to be on their own'*	Italiener sind sehr gastfreundlich. (8 and 10)