Perception of intonational contours on given and new referents: a completion study and an eye-movement experiment[1]

Caroline Féry, Elsi Kaiser, Robin Hörnig, Thomas Weskott and Reinhold Kliegl

1. Introduction

Two factors have facilitated the study of the interplay between speech perception and phonology: first, we now have the benefit of various technological advances that allow for the collection and analysis of data on perception that were unavailable 20 years ago, and second, theoretical progress. In the framework of Optimality Theory (Prince and Smolensky 2004; McCarthy and Prince 1993), research has led to the formulation of perceptually grounded constraints that interact with constraints motivated by other modules of the linguistic system. See for instance Smolensky (1996), Steriade (1995, 1999) and Flemming (1995), who have modeled in Optimality Theory the influence of perception on the usual production grammar. A further progress is proposed by Boersma (1998 et seq.), who models perception itself in Optimality Theory.

In this paper, by presenting a completion experiment and an eye-tracking study and by implementing the results in an Optimality-Theoretic model, we rely on both the technological and the theoretical advances. Traditionally, phonological theory draws heavily on production, a fact which is explained by the ease with which articulatory facts, like the position of the tongue, can be analyzed as opposed to the difficulties encountered in the study of perception. It has been observed repeatedly, however, that the way we perceive contrasts in sound structure has an influence on phonological systems and sound changes, and vice-versa, and that we articulate sounds or sound sequences depending on how we perceive them (see for instance Ohala 1981; Lindblom 1990; Steriade 1999). To cite well-known examples, palatalization arises as a consequence of coarticulation between a coronal and a high vowel, and final devoicing as a consequence of a reduction of all final obstruents to plain voiceless ones. If neutralization arises as a result of palatalization or final devoicing, contrasts are lost. But clearly, the maintenance of a contrast may be in some cases more important than the reduction of articulatory effort, even when perception is not maximally guaranteed, such as in non-prominent syllables (see Hume 1994 and Ćavar 2004, for instance). On a more abstract level, perception and production may not be separable into distinct cognitive modules, since these two components have necessarily evolved in parallel.

The phonetic articulation of sounds and the intonation of melodies are achieved by manipulating the physical acoustic waves, and their perception is translated into auditory impressions by the auditory system. Phonological systems, on the other hand, are symbolic in nature, and abstract away from concrete phonetic events. Once we have learned a language, we do not need to physically produce or perceive speech sounds in order to access its phonology. The problem that phonologists face is providing the best model of the relationship between phonetics and phonology. To this purpose, the

[1] This paper is part of the projects A1 and C1 of the SFB 632 'Information structure', funded by the DFG. We thank the organizers, Paul Boersma and Silke Hamann, and the participants of the workshop 'Speech perception in or outside phonology' of the 27th DGfS conference, where we presented this paper. Many thanks to Sabine Kern, Petra Grüttner and Anja Kuschmann for technical help.

classification of sounds in categories has proved very useful for the study of phonemic systems and allophonic alternations, as well as for the study of tonal patterns (see Pierrehumbert 1980 and below). The theoretical question which we try to answer in this paper is how tonal structures relate to the expectation of listeners, and to provide a model of our results.

Section 2 gives an overview of the main theoretical issues relating to intonation and its perception. Studies of the perception of tonal structures have observed categorical classes in the alignment of peaks (horizontal dimension) as well as the height (vertical dimension) of peaks (Pierrehumbert and Steele 1987; Kohler 1990; Ladd and Morton 1997). Other studies find that misplaced accents induce more difficulty than exaggerated or insufficient ones (Gussenhoven 1983; Birch and Clifton 1995; Hruska, Alter, Steinhauer and Steube 2001), with this difference being attributed to the difficulty of processing incorrect prosodic information as compared to merely inappropriate prosodic information. Only few processing studies have focused on the perception of qualitatively different accents, like rising and falling ones, and the kind of anticipatory expectation they elicit (but see Féry and Stoel 2006 for such a study in German).

In sections 3 and 4, we report two experiments which we conducted in order to examine these issues in German. Experiment 1 is a completion experiment with a forced-choice task between two possible object completions: one that refers to a character that has already been mentioned in the preceding discourse ('discourse-given') and one that refers to a referent that is being mentioned for the first time ('discourse-new'). In Experiment 2, the eye movements of listeners were tracked as they heard spoken descriptions of scenes, so as to test whether specific on-line referential processes are involved in the processing of accents. Upon hearing an early falling accent sequence, German listeners should expect the upcoming noun to be discourse-given, whereas an early rising sequence should trigger anticipation of a discourse-new referent. This experiment adapted the material and experimental design of Kaiser and Trueswell (2004), an eye-gaze experiment on word order in Finnish which showed that comprehenders extract pragmatic implications of word order incrementally as a sentence unfolds in real time, and even anticipate upcoming referents based on information encoded in word order.

The results of the sentence completion study and the eye-tracking experiment confirmed our predictions. Sentences with falling accents triggered more completions with the discourse-given character and showed anticipatory eye movements to the discourse-given referent at the onset of the second noun, even before participants had enough acoustic information to recognize this word. In contrast, sentences with rising accents triggered more answers with a discourse-new character – as compared to a (hypothetical) baseline – and showed anticipatory eye movements to a discourse-new referent.

Section 5 introduces a model of tone perception. An important finding of our experiments is that listeners have expectations regarding the discourse-status of the verbal arguments well ahead of their realizations. If listeners know that the German sentence they are processing is a declarative sentence, they can deduce the information status as discourse-new or discourse-given just by the way the tonal structure of the sentence is unfolding. As is shown in section 5, such a result speaks strongly in favor of phonological processing of spoken material, and a model of perception has to take these observations into account.

2. Background

2.1. Tonal structure of German

German is an intonation language, which means that it uses pitch accents to express grammatical relationships and information structure (see von Stechow and Uhmann 1986 for instance). In intonation languages, the tonal structure of utterances is relatively flexible. Most researchers who study intonation agree that two kinds of intonational markings are necessary: (i) boundary tones, which delimit prosodic phrases and (ii) pitch accents, associated with lexically stressed syllables which are focused or topicalized. Syllables lacking a lexical stress can carry a pitch accent if necessary for the information structure. We use the term 'focus' in the sense of 'prominent' or 'new' (Rooth 1985; von Stechow and Uhmann 1986 among others), and topic in the sense of 'aboutness' (see e.g. Reinhart 1981; Jacobs 2001; inter alia). If a constituent is a topic, in German it gets a special rising tone, and the sentence is felt to be 'about' this constituent. In our material, both focus and topic are signaled by pitch accents. If accenting is normal or unmarked (on the object in an SVO sentence), wide focus is assumed. The notion of wide focus is used in contrast with narrow focus, a situation in which only part of the sentence is prominent. In our sentences (see below), both wide and narrow focus are used. A pitch accent on the verb induces narrow focus on this constituent. An early rising accent signals a topic. We refer the reader to Rooth (1985, 1992) for a semantic definition of 'focus.' Here, it suffices to say that focus denotates the part of the sentence which the speaker wants to signal as new, or as particularly prominent. Many researchers also agree that tones are morphemic, and that they arise both in the lexicon (for tone languages) and in the postlexical grammar (for all languages), when words are put together to deliver meanings. Pitch accents may appear on any syllable in a sentence, if necessary for a felicitous conversation. The term 'nuclear accent' refers to an accent that is the last one in the sentence and which is perceived as the strongest one, whereas prenuclear accents are those realized early in the sentence, before the nuclear one. Nuclear accents and prenuclear accents are realized by a small inventory of pitch accents, the distribution of which is relatively well-studied. In German, nuclear accents are always bitonal (falling or rising), but prenuclear ones may be less complex and consist of just a high tone, or just a low tone. The nuclear accent is the last pitch accent in the Intonation Phrase (IP), an entity which corresponds roughly to a sentence. In a neutral realization of (1), for instance as an answer to the question 'What happens next?', the object *Krankenschwester* 'nurse' carries the nuclear accent and the subject *Arzt* 'doctor' has a prenuclear accent. The location of the neutral nuclear accent is conditioned by grammatical principles. In the kind of sentences which we focus on in this paper (transitive matrix clauses with SVO order), the default location of the nuclear accent is on the last argument of the verb, namely on the direct object (see von Stechow and Uhmann 1986; Cinque 1993; Féry and Samek-Lodovici 2006, among others). (In example (1), as in the other examples in this paper, the square bracket and the subscripted F highlight the focus structure of the sentence, namely the prominent or new information.)

(1)
L*H H* L L_I
[Der ARZT befragt gleich die KRANKENSCHWESTER]$_F$
'The doctor asks/questions soon the nurse'

In a declarative sentence, the nuclear accent has a falling contour, and the prenuclear accent is rising. This is indicated in (1) by means of the so-called autosegmental-

metrical notation system, which has been originally developed by Pierrehumbert (1980) for English, following a proposal by Bruce (1977) for Swedish, but which has in the meantime been adapted for a number of languages, among others for German (see Féry 1993; Grabe 1998; and Grice, Baumann and Benzmüller 2005; among others). L*H is a rising accent, the starred tone L* is associated with the accented syllable, and the following tone is a 'trailing' tone, which shows that the entire excursion is rising. Rising intonation (L*H) has been interpreted as expressing topicality, openness and non-finality. In contrast, a falling accent such as H*L expresses focus and finality. The pitch contour of this sentence, one of our stimuli, is reproduced in Fig. 1.

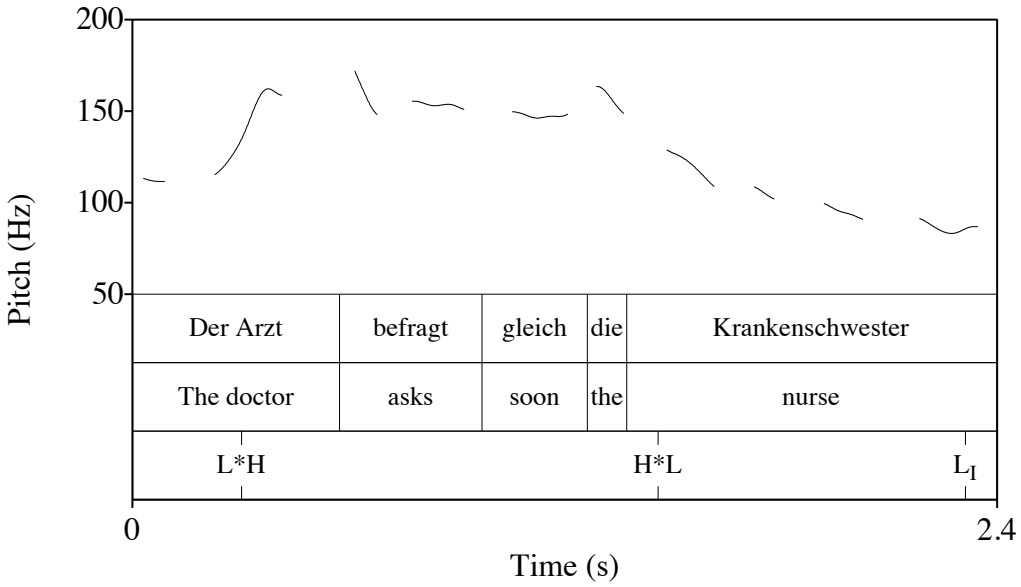

Fig. 1. Contour with final fall.

In most cases, nuclearity means finality, since the last accent is generally perceived as the most prominent of the sentence, even though its acoustic correlates may be less prominent than those of the prenuclear accents. This is due to the regular downstep of the melodic peaks which is part of the tonal characteristics of German, as well as of other languages. The H of the final H*L is lower than the H of the initial L*H. At the end of the phrase, the tone L_I indicates that the global contour of the sentence is falling. This is a boundary tone which is associated with the last syllable of its domain rather than with a prominent syllable. A final L_I contrasts with a final H_I, which is typical for a question intonation or a so-called continuation intonation. In a sentence like (1), the falling pattern is induced by the last pitch accent, located on the first and second syllables of *Krankenschwester*, and the remainder of the sentence, in this case the two last syllables of this word, are low and flat. Between the two accents, the voice may remain relatively high, since there is no further tonal specification between the last high tone of *Arzt* and the first high tone of *Krankenschwester*. However, there may in general be a dip between the two high tones of a sentence like (1), especially when more syllables intervene between the two H tones. This dip is found in other languages, as well (see Pierrehumbert 1980 for English).

2.2. Given and new in intonation

The distribution of pitch accents depends on the information structure of a particular sentence. Elements that are part of the background by virtue of having already been introduced into the discourse are deaccented, especially if they occur after the nuclear accent, i.e., in a 'postnuclear' position. This is illustrated by the question-answer pair in (2). In (2B), *Krankenschwester* is introduced by the preceding question (2A), and for this reason, it is no longer able to carry the nuclear accent of the sentence. It is now the noun phrase that answers the question, *der Arzt*, that carries the nuclear accent, and the rest of the sentence has a low and deaccented intonation. Notice that, in naturally-occurring dialogue of this type, the answer is often elliptical, i.e., the deaccented part of the sentence is elided. In fact, the repetition of the given material, though grammatical, is not completely natural.

(2) A: Wer befragt gleich die Krankenschwester?
who questions soon the nurse
'Who asks soon the nurse?'

H*L L_I
B: [Der ARZT]$_F$ befragt gleich die Krankenschwester
the doctor questions soon the nurse
'The doctor asks/questions soon the nurse.'

In contrast, in the question-answer pair in (3), *befragt* answers the preceding question and thus receives the nuclear accent. In this case, even though *der Arzt* is backgrounded, it still carries a rhythmical prenuclear accent. This prenuclear accent can be less prominent than in an all-new, entirely-focused utterance as in (1) above, but does not need to be (see section 2.3). As mentioned above, a marked rising tone in this position is readily interpreted as topical. This sentence is also part of our stimulus set and is illustrated in Fig. 2.

(3) A: Was macht der Arzt gleich mit der Krankenschwester?
what does the doctor soon with the nurse
'What does the doctor do with the nurse?'

L*H H*L L_I
B: Der Arzt [BEFRAGT]$_F$ gleich die Krankenschwester
the doctor questions soon the nurse
'The doctor asks/questions soon the nurse.'

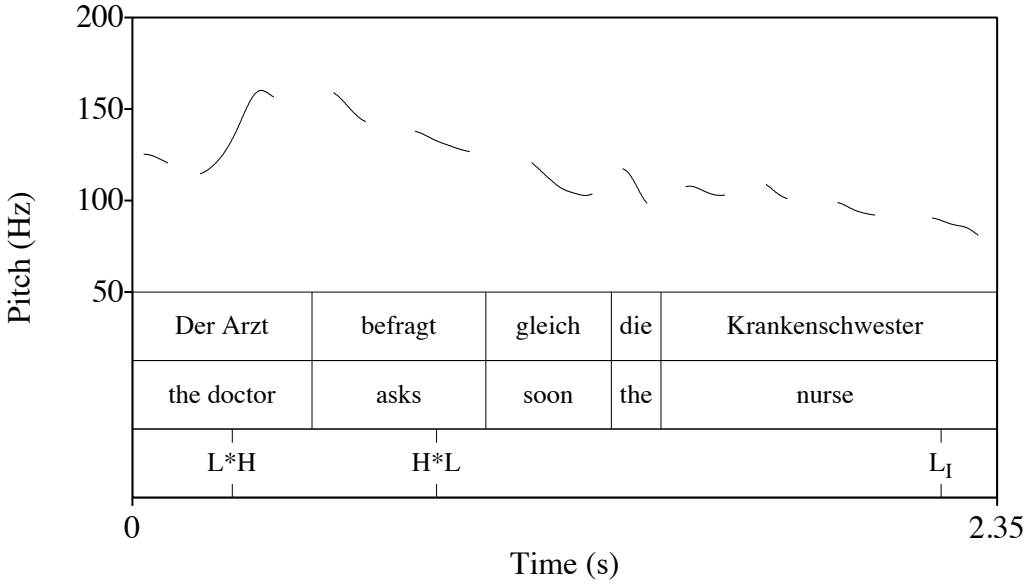

Fig. 2. Contour with an early fall and deaccented final material.

In German, a rising tone (L*H) must be followed by a falling tone (H*L) in a declarative sentence, since the last accent of a declarative is always falling. A falling tone is the last tone of an IP and there is no following accent in this case.

2.3. Perception of tones

Earlier research on the perception of tones has investigated the relationship between accent location and information structure, or between accent quantity and information structure, and the findings indicate that listeners are reasonably sensitive to the relation between accents and tones. Gussenhoven (1983) and Birch and Clifton (1995) examine the role of prenuclear accents on the verb in a VP consisting of a verb plus an argument or an adjunct in English, and find that a prenuclear accent on a verb is tolerated in a sentence where only the nuclear accent is required. In Gussenhoven's study, this result is much stronger when the VP consists of a verb plus argument (as in the VP *share a flat*, where an accent structure with a single accent on *flat* is preferred) than when it consists of a verb plus adjunct (as in *skiing in Scotland*, where a structure with two accents, one on *skiing* and one on *Scotland* is readily accepted). In an experiment in which listeners had to decide how well pairs of question-answer sequences make sense, Birch and Clifton also find that listeners accept both absence and presence of prenuclear accent on the verb in focused VP (with the nuclear accent on the postverbal element). The distinction between prenuclear, nuclear and postnuclear accents has been investigated in Dutch, (Nooteboom and Kruyt 1987; Krahmer and Swerts 2001), and German (Hruska, Alter, Steinhauer and Friederici 2001). The findings show that prenuclear accents on given material are readily accepted whereas unlicensed (post)nuclear accents are not. For example, the sentence in (1) with the nuclear accent on the object is perceived as deviant in the context of question (2A) which is asking about the subject of the sentence.

Studies bearing on the perception of tones as indicators of givenness and newness have been conducted for English by Jannedy (2002) and by Welby (2003), and

for German by Féry and Stoel (2005). These authors find a sensitivity not only to presence and absence or quantity of accents, but also to the quality of accents. Simply put, topics require topical accents and foci require focal accents. In these studies, a whole sentence was presented to listeners at once and judgments were elicited on the basis of the complete sentence plus a context preceding the sentence. These studies corroborate the finding that prenuclear accents on given material are much more readily accepted than unlicensed postnuclear accents. Unlicensed nuclear accents located later than the licensed ones are felt to sound inappropriate. The difference of acceptability between prenuclear and postnuclear accents is explained by the observation that a supplementary prenuclear accent can get an interpretation in which the prenuclear accent is information structurally prominent, for instance by virtue of being a topic. We will see below that the readiness to accept a prenuclear accent, but to reject a inadequate nuclear accent has influenced our results, as well.

Some studies have also explicitly addressed the question of whether the perception of intonation is categorical or gradient. This issue has proved to be particularly hard to tackle since tonal patterns are realized by real voices, with their own fundamental frequency, quality, intensity, range and so on, all of them being aspects which are gradient in nature. Nevertheless, some interesting clear categories have emerged. First, using a series of equal-sized temporal shifts of F0 peaks across the constant accented syllable /lo/ in the utterance *Sie hat ja gelogen* 'she's been lying', Kohler (1990) finds three intonation categories: 'early', 'medial' and late peaks, being associated with parallel changes along the semantic dimension from 'established' (early peak) to 'new' (middle) and 'surprising' (late).[2] As a second example, Pierrehumbert and Steele (1987) asked American English speakers to imitate 15 different synthesized realizations of the phrase *only a millionaire* which differed from each other in the timing of the intonational peak: the tokens were synthesized to form a continuum such that the peak of one token was 20ms later than the peak of the preceding token. Pierrehumbert and Steele hypothesized that if speakers were able to reproduce these fine differences, then peak alignment must be gradient. But speakers were not. Their realizations clustered around two values which are shown with the help of tone sequences in (4) and (5).

L_IH^*L H^*L H_I
(4) ONly a MILlionaire

L_IH^*L L^*HL H_I
(5) ONly a MILlionaire

A third, slightly more controversial example comes from Ladd and Morton (1997) who find categoriality in the interpretation of height of pitch accents in British English. There is an abrupt change in the interpretation of pitch accents, which skip form normal to emphatic when the height of the pitch accent is gradually increased.

The present study differs from the earlier studies in that it investigates the perception of tonal structures incrementally, before the sentence has ended. It assumes categoriality in the type of accents used: a falling accent is a final focus accent, and a rising accent is a non-final topical or pre-nuclear accent. Like production, speech perception is a process that occurs over time. As a sentence unfolds, listeners – on the basis of different kinds of information – develop expectations of what the speaker will

[2] Unfortunately, Kohler does not elaborate the concepts 'established', 'new' and 'surprising.'

produce. We assume that different kinds of phonological components exert an influence on the listener's expectations. Our study is not the first one to investigate the kind of hypotheses that listeners make on the basis of what they have heard at a particular point in time. The influence of the initial sequence of segments is a domain which has been extensively studied in the psycholinguistic literature. In the domain of syntax, word order also plays an important role, especially in languages with flexible word order like German and Finnish. But up to now, only very few studies have investigated the anticipatory effect that the not-yet-completed tonal structure has on listeners for the perception of entire sentences.

The next two sections present two studies that we conducted in order to investigate whether listeners have specific expectations about the information structure of sentences; in particular, whether listeners expect an upcoming constituent will be discourse-new (mentioned for the first time) or discourse-given (already mentioned in the discourse) on the basis of the intonational information they have heard. However, since listeners usually have no explicit knowledge of the tonal contours of utterances, it is not possible to investigate their expectations and intuitions concerning the fundamental frequency directly. Our investigation had thus to be indirect and relied on the following well-established facts about German (see Büring 1997; Grabe 1998, Grice, Baumann and Benzmüller 2005; Féry 1993 for more detailed discussion of these points):

- A declarative sentence has a global falling contour
- A bitonal falling accent is the last one in a declarative Intonation Phrase,
- A bitonal rising tone is not last in a declarative sentence
- A focus is realized with the last falling tone of the utterance, and
- A prenuclear accent (topical or not) is realized with a rising tone

By studying the kind of expectations that participants develop with respect to the discourse status (discourse-given versus discourse-new) of not yet pronounced referents, we were able to gather indirect evidence that listeners are indeed sensitive to the implications that certain tonal realizations bring with them.

3. Completion experiment (Experiment 1)

3.1. Material and procedure

The same set of sixteen experimental items was used in both the completion experiment discussed in this section and in the Visual World experiment discussed in section 4 (see Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy 1995 for details concerning the Visual World paradigm). The decision to use both types of procedure was motivated by a desire to compare results from an off-line experiment with those of an on-line experiment, in order to investigate the processing role of intonational information. It is important to note that although the off-line sentence completion experiment lets us investigate whether comprehenders use intonation to predict the discourse status of upcoming referents, an off-line study cannot tell us at what stage in the course of language processing this information is used. Both visual and auditory materials were adapted from Kaiser and Trueswell (2004) to meet the need of our design, and translated into German. The visual stimuli consisted of full-screen color pictures depicting three characters and other objects that made up coherent scenes. The three characters were of approximately the same size and positioned such that one was on the

left side of the picture, one was in the middle, and one was on the right. A sample picture with a patient, a doctor, and a nurse is shown in Fig. 3.

Fig. 3. A sample visual experimental stimulus.

As for the auditory stimuli, brief verbal passages were prepared that told a simple story involving the characters shown in the picture. The sample picture in Fig. 3 was accompanied by the following passage:

(6) a. An der Empfangstheke des Krankenhauses lehnen ein Arzt und eine Krankenschwester.
'On the hospital reception desk are leaning a doctor and a nurse,'

b. Die Uhr zeigt fast zwei.
'It is almost two o'clock.'

c. Der Arzt befragt gleich die Krankenschwester / die Patientin.'
'The doctor asks soon the nurse / the patient.'

d. Die Krankenschwester schaut nervös.
'The nurse looks nervous.'

Prior to the target sentence (6c), two of the characters were mentioned in the first sentence of the passage (the doctor and the nurse in (6a)) and hence were discourse-given. The third character (the patient) was not mentioned in the first two sentences and hence was still discourse-new. The target sentence was preceded by a distractor sentence which referred to an element in the picture different from the three characters (the clock in (6b)). The complete set of speech materials was recorded in one session with a male speaker of German. No acoustic manipulations were performed.

The subject of the target sentence always refered to an already-mentioned entity, whereas the object refered to either an already-mentioned referent (e.g. the nurse) or a discourse-new referent (e.g. the patient). The adverb *gleich* 'soon' was inserted between the verb and the postverbal object in all experimental items in order to separate the two words from each other. The pitch on *gleich* was low or high, depending on the accent on the postverbal noun (compare Figs. 1 and 2). As shown in (7), four versions of each

target sentence were recorded by crossing the discourse status of the referent of the object noun phrase (discourse-new vs. discourse-given). The pitch on *gleich* is also explicitely noted, since it is this word which serves as reference point in both our experiments. In the subsequent discussion, we will often use the terms 'high/low tone on *gleich*' simply as a convenient shorthand label for the two different intonational patterns we are investigating (late fall and early fall, respectively).

(7) a. *Late fall*, the object is new (congruent)

L*H H H* L L_I
[Der ARZT befragt gleich die PATIENTIN]$_F$
the doctor questions soon the patient

b. *Early fall*, the object is given (congruent)

L*H H*L L L_I
Der Arzt [BEFRAGT]$_F$ gleich die Krankenschwester

c. *Late fall*, the object is given (incongruent)

L*H H H* L L_I
[Der ARZT befragt gleich die KRANKENSCHWESTER]$_F$
the doctor questions soon the nurse

d. *Early fall*, the object is new (incongruent)

L*H H*L L L_I
Der Arzt [BEFRAGT]$_F$ gleich die Patientin

For the completion experiment, two variants of target sentence fragments were used. In the first one, called *late fall*, the object is realized with a falling nuclear accent (and the subject has a rising prenuclear accent). This is the variant illustrated in (1) and Fig.1, as well as in (7a and c). In the second variant, called *early fall*, there is a falling nuclear accent on the verb (and the subject has a rising prenuclear accent). This is the variant illustrated in (3) and Fig. 2, as well as in (7b and d). Fragments were obtained by truncating the target sentences so that the second noun phrase is removed, e.g., *Der Arzt befragt gleich* … 'The doctor asks soon …'. Two presentation lists were constructed by randomly combining the 16 target stories with 20 filler stories (also adopted from Kaiser and Trueswell 2004). Each pair of consecutive target items was separated by at least one filler item. Fillers were designed to vary in the number of characters in the picture and in whether the characters were mentioned in the story. Within a presentation list, eight of the target trials appeared with the high tone pattern and eight appeared with the low tone pattern. Reverse order lists were also generated to control for trial order.

Thirty-two native speakers of German were tested individually on a PC. Each trial began with the presentation of the picture stimulus. After a 1000 ms delay, participants heard the corresponding passage ending with the target sentence fragment via earphones. Participants were asked to complete the passage by ticking one of two options which were presented to them on a sheet of paper (e.g. the nurse or the patient).

3.2. Hypotheses

As noted in section 2.2, we hypothesized that in the late fall variant (with a high tone on *gleich*), the rising accent on the subject would be interpreted as the topical non-final accent which must be followed by another accent. We therefore expected participants to complete the sentence fragments more often with the discourse-new referent (e.g. the patient) in this condition. On the other hand, in the variant with the early fall (and a low tone on *gleich*), the falling contour on the verb should be interpreted as the last accent of the sentence. In this case, we expected participants to complete the sentence fragments more often with the discourse-given referent (e.g. the nurse).

Hypotheses:

H1: High tone on *gleich* (late fall) prompts completion with discourse-new referent.
H2: Low tone on *gleich* (early fall) prompts completion with discourse-given referent.

3.3. Results

The results of the completion task are summed up in Fig. 4. When hearing a low tone on *gleich*, the participants completed the target sentence fragments more often with the discourse-given referent than with the discourse-new referent (76% vs. 24%). When confronted with a high tone on *gleich*, completions with the discourse-new referent increased remarkably and were even somewhat more frequent than completions with the discourse-given referent (56% vs. 44%). Tests revealed that the difference between the answers to the two intonation patterns was statistically significant ($p < 0.05$) (see Weskott *et al.* submitted for detail).

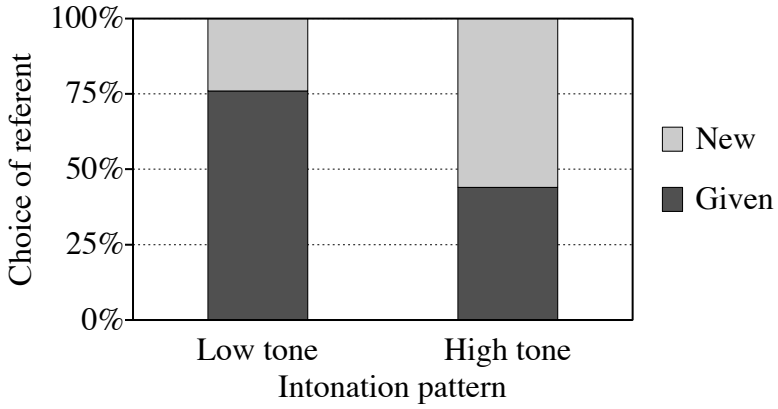

Fig. 4. Results of the sentence completion task.

3.4. Discussion

Our results show that the choice of the referent in the sentence completion was dependent on the intonational variant. This shows that participants were sensitive to prosody. Discourse-status depended on intonation as predicted: With a low tone on *gleich*, completions with discourse-given referents were chosen more frequently than with discourse-new referents. With a high tone on *gleich*, completions with discourse-given referents were reliably reduced. It is tempting to infer that only the low tone pattern influenced completions whereas the high tone pattern had no influence, since completions with referents of either discourse status were nearly equally frequent in this last condition. We have, however, independent evidence from a comparable study on

word order (not discussed in this paper, but see Weskott *et al.* submitted) that participants in general exhibit a preference for completions with discourse-given referents in this task (see also Kaiser and Trueswell 2004 for related discussion regarding a preference for already-mentioned referents). In light of these data, we take the results of the high tone condition as an indication that the rising accent counteracts the discourse-given bias and pushes participants to choose discourse-new referents more often than they otherwise would have.

When considering these results, one needs to keep in mind another factor that may have contributed towards the apparent discourse-old bias that we observed in the results. First, it is important to note that even the discourse-given character can, under certain circumstances, receive an accent – namely, when it is interpreted contrastively. In the semantic interpretation of focus, a contrastive accent is interpreted as generating a set of alternatives (in the sense of Rooth 1985, 1992). Thus, a rising accent on the subject can occur when the object is discourse-new information or when it is discourse-given and contrastive. As a result, a rising accent on the subject may also trigger completions with the already-mentioned referent as the object. However, the sentences used in this experiment did not provide any explicit cues that would push participants towards a contrastive interpretation of the object, and thus we regard the risk of contrast and discourse-oldness being confounded as very low.

Furthermore, since the discourse-new referent is displayed in the picture from the beginning, we cannot exclude the possibility that participants will interpret this character as 'visually given'. In other words, we might also expect some responses in which participants opted for the discourse-new (but visually given) referent even if the accent on the verb is falling. It is not unheard of for languages to distinguish between different kinds of givenness, see. e.g. Baumann (2005) and Birner and Ward (1998).

In sum, the results of the first experiment show that different intonational patterns create different expectations regarding the discourse status of upcoming referents. However, on the basis of the sentence completion experiment, we cannot tell at what stage of processing the intonational information is used. Are listeners extracting information from the intonational patterns at the same time as they parse the sense? Or is intonational information used at a later stage of processing, after the sentence fragment has already been comprehended? On the basis of the results of the off-line sentence completion task, we can conclude that listeners are sensitive to information that intonation carries regarding the discourse status of not-yet-mentioned entities, but we do not know whether intonational information is being processed at the same time as the propositional meaning of the sentence, or at a later stage of processing. In order to investigate this question, we used eye-tracking in a visual world paradigm (Experiment 2).

4. Visual World experiment (Experiment 2)

4.1. Material and procedure

The auditory stimulus material in this experiment consisted of the same passages used in Experiment 1, but now the sentences were not truncated, i.e., we used entire target sentences instead of target sentence fragments (sample items are given in (7). Four presentation lists were constructed by randomly combining the 16 target stories with 32 filler stories. Each pair of consecutive target items was separated by at least one filler

item. Within a presentation list, eight of the experimental trials appeared with the low tone pattern and eight trials appeared with the high tone pattern. For both of these prosodic patterns, the target sentence of half of the trials mentioned the discourse-given referent and the target sentence of the other half of the trials mentioned the discourse-new referent. Each target item was then rotated through these four conditions, generating four different presentation lists. The fillers had the same overall setting as in the preceding experiment, except that the final sentence of eight filler stories described the picture incorrectly. The total sample of 48 trials were individually randomized for each participant.

A new group of 40 native speakers of German (none of whom participated in the completion experiment), were tested individually. Each trial began with the presentation of the picture stimulus. After a 1000 ms delay, participants heard the corresponding passage via earphones. Participants were instructed to press a key whenever they noticed an erroneous description. While participants listened to the passages, their eye-movements were recorded by a SMI-iViewX eye-tracking system. The basic idea of the Visual World paradigm is to track which of the referents in the picture participants attend to by determining which referent is fixated at any point in time. We implemented the Visual World paradigm in order to check whether intonational cues can trigger anticipatory eye-movements towards the referent with the appropriate discourse status. Incorrect descriptions were included to ensure that participants would pay attention to both the visual and the auditory stimulus. The advantage of this method over the one used in Experiment 1 is that speakers do not perform a given task consciously, and that the method is less invasive. Eye fixations on referents indicate the likelihood of the referent as the intended object in the target sentence. It has being shown independently that people tend to look at what is being talked about. We can thus infer from the proportion of fixations how likely a potential referent was considered to be mentioned in each case (Allopenna, Magnuson and Tanenhaus 1998; and Tanenhaus, Magnuson, Dahan and Chambers 2000).

4.2. Hypotheses

According to our hypotheses, summed up below, an early fall melody (accompanied with a low tone on *gleich*) is interpreted as standing for an unaccented object, and participants will show a preference for fixating the discourse-given referent (e.g., the nurse) more often than the discourse-new referent. On the other hand, a late fall melody (accompanied by a high tone on *gleich*) is interpreted as signaling an upcoming accented (and hence discourse-new) object, and thus we predict that participants will tend to fixate the discourse-new referent (e.g., the patient) more often than the discourse-given one. This hypothesis claims that accent patterns trigger anticipatory eye-movements towards the discourse-given referent (when a falling accent has already been heard) or towards the discourse-new referent (when only a rising accent has occurred, and a falling accent is expected to follow) before the segmental information of the actual second noun phrase becomes available. In order to test this, we identified the offset (end point) of the adverb *gleich* 'soon' as the critical event at time $t_0 = 0$. The critical noun, of course, starts right at the offset of *gleich.* Thus, we predict a main effect of prosody (whether the voice is low or high) shortly after the critical event.

Hypotheses
H1: Low tone on *gleich* (early fall) prompts anticipatory fixations of discourse-given referent.
H2: High tone on *gleich* (late fall) prompts anticipatory fixations of discourse-new referent.

It follows from the hypotheses that two of the four experimental conditions are congruent, whereas two are incongruent. If the early fall pattern triggers anticipatory fixations of the discourse-given referent and the upcoming noun phrase turns out to refer to the discourse-given referent, the anticipatory fixations are correct. Likewise, if the late fall pattern triggers anticipatory fixations of the discourse-new referent and the upcoming noun phrase mentions the discourse-new referent, the anticipatory fixations prove to be correct. However, when the object noun phrase does not refer to the anticipated referent (i.e. when the object is discourse-given after a late fall pattern, or discourse-new after early fall pattern), we assume that attention has to be reallocated from the anticipated referent to the actually mentioned referent. We predict that this reallocation of attention is reflected in a delayed increase of preferred fixations of the mentioned referent.

4.3. Results

We report the results within the time window beginning 1000 ms before and ending 1000 ms after the critical event (offset of *gleich*). This time window covers nearly the whole target sentence. Fixations were determined by using a velocity-based saccade detection algorithm (Engbert and Kliegl 2003). Fixation probabilities for the discourse-given referent and the discourse-new referent were averaged within ten segments à 200 ms. In Fig. 5, we plot difference scores, which were calculated by subtracting the probability of fixating on the discourse-new referent from the probability of fixating on the discourse-given referent. Accordingly, positive scores indicate preferred fixations of the discourse-given referent, whereas negative scores indicate preferred fixations of the discourse-new referent.

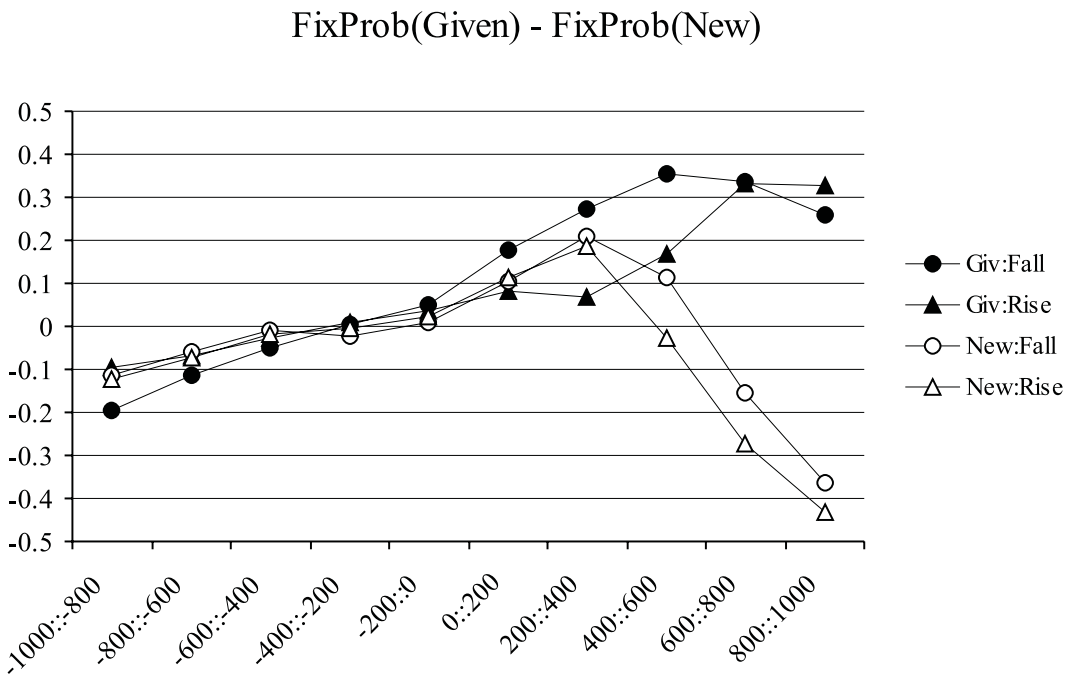

Fig. 5. Fixation probability for the discourse-given referent minus fixation probability for the discourse-new referent in Experiment 2 for target sentences mentioning the discourse-given (Given) and the discourse-new (New) referent, paired with an early fall accompanied with a low tone pattern on *gleich* (Low) or with a late fall accompanied with a high tone pattern on *gleich* (High)

Analyses of variance revealed an increasing tendency to fixate the discourse-given referent until about 300 ms after the critical event, indicating a general bias towards the discourse-given referent. In the time slices from 200 ms until 600 ms after the critical event, fixations of the discourse-given referent were reliably more frequent with a low tone on *gleich* as compared to a high tone on *gleich*. In other words, participants' fixations on discourse-old vs. discourse-new referents are significantly influenced by the intonational pattern of the sentence. It is well-known that it takes about 200 ms to program and launch an eye-movement (Matin, Shao and Boff 1993), and thus the eye-movements occurring 200 ms after the offset of *gleich* were actually programmed during the offset of *gleich*, i.e. before the onset of the subsequent noun (the object). Note also that the mean onset of the postverbal noun phrase was within that time window (113 ms after the onset of *gleich*). The finding that there is a significant difference in the 'given referent-new referent' difference scores already in the 200-400 ms time slice indicates that participants are indeed launching anticipatory eye-movements. In other words, they are making use of the intonation information very quickly and efficiently to make predictions about the next referent.

Finally, in the time slice from 400 ms until 600 ms after the critical event, the lexical content of the noun has an effect. Thus, in the incongruous conditions (a late fall/high tone on *gleich* followed by a given object, and an early fall/low tone on *gleich* followed by a discourse-new object), the anticipatory looks turn out to be incorrect, and thus we see a re-allocation effect, as participants make use of the lexical content of the noun to fixate the mentioned referent. In contrast, in the congruous conditions (a late fall/high tone on *gleich* followed by a new object, and an early fall/low tone on *gleich* followed by a given object), the lexical content of the noun confirms the expectations

participants built up based on the intonational pattern. Not surprisingly, this effect persists until the end of the of the time window. Note that by the end of the trials, participants' fixations mirror their overall target choices.

4.4. Discussion

Experiment 2 confirms the finding of Experiment 1, that participants are sensitive to intonation. Crucially, Experiment 2 also tells us about the time-course of this sensitivity. In addition to an early general preference to attend to the discourse-given referent, we saw in Fig. 6 that prosody affected fixations of referents very shortly after the critical event. The data suggest that the participants were launching anticipatory looks to one of the two referents even before they had enough phonological information to recognize the word – in other words, it seems that participants are very efficiently and quickly making use of the information carried by the intonational contour of the sentence. Moreover, the reallocation effect discussed above shows that whenever the anticipation prompted by the accent pattern turned out to be misleading, it took participants some time to shift their attention to the actually mentioned referent, as indicated by a slower increase of preferred fixations of the mentioned referent. As a whole, Experiment 2 shows that participants make use of intonational information very quickly during on-line processing, and that the discouse-level information carried by intonation, like that encoded in word order (see Weskott *et al.* submitted and Kaiser and Trueswell 2004 on Finnish) can be used to predict upcoming referents.

5. Modeling tonal perception[3]

On the basis of this observation and of existing models of phonological perception in Optimality Theory, we develop in this section a model of the perception of intonation, to our knowledge the first of this kind. Our results are important for a theory of perception of tonal contours since they provide evidence for a preplanning of upcoming tonal events. Hearers develop expectations about the yet-to-come tonal patterns while a sentence enfolds.

The material presented in the preceding sections leads us to assume three levels in perception, one of which being the phonological level, and the other two are those which Boersma (this volume) and Boersma and Hamann (this volume) call phonetic and underlying respectively, as shown in (8). The lowest level is the acoustic signal, in our case a melody or tune. The intermediate level is the surface form (the phonological level in our terminology), which corresponds to pitch accents, boundary tones etc. (see section 2) and the highest level is the semantico-conceptual one (the underlying form), which corresponds to the meaning of phonological tones. We thus follow McQueen and Cutler (1997) among others, who assume that perception involves two separate processes: perception *per se* (equivalent to the prelexical level of psycholinguistics) and recognition (corresponding roughly to the lexical level). The perceptual operation transforms the raw fundamental frequency into phonological objects, which can be understood as tonal morphemes (Liberman 1975; Pierrehumbert 1980; Ladd 1996), and the recognition operation interprets these morphemes as abstract concepts with their own meanings (topic in our example (8)).

[3] Many thanks to Paul Boersma and Silke Hamann who discussed the formal aspects of tonal perception with us.

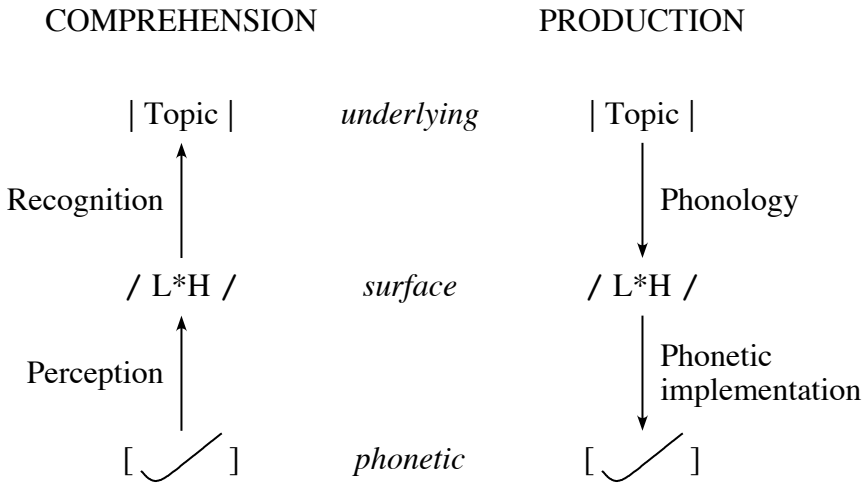

Fig. 6. Model of tone processing

According to the conventions of this book, the single arrows in Fig. 6 indicate that the interpretation of a rising melody as a pitch accent is language-dependent. In a language without pitch accent, this kind of pitch excursion may signal something else, like a lexical tone or the boundary of a phonological domain. It must be noted that the same kind of pitch excursion can also be interpreted as a boundary tone in German if it is located on an unstressed syllable at the end of an IP, for instance. In other words, the task of the hearer while perceiving a certain tonal pattern is to perceive it as a specific type of phonological object which is then identified as the bearer of a specific meaning, like topic, focus, new referent, etc. The two processes can happen in parallel since it must be the case that the process of identification (recognition) influences the process of perceiving (translation of F0 contours into tonal morphemes).

We do not dwell on the phonetic level here, and refer the readers to Boersma (this volume) who discusses several possible interpretations for it. Instead, we concentrate on the phonological level. There are two major arguments emerging from our study which speak in favor of the phonological level. First, as shown in section 3, pitch excursions are uninterpretable as such. It is the phonological system of a specific language which allows the listeners to interpret falling and rising tones as pitch accents, boundary tones or lexical tones. Listeners translate a rising pitch excursion into something abstract like a topic because they understand it as a certain kind of pitch accent, and pitch accents have themselves meanings (see for instance Pierrehumbert and Hirschberg 1990; Steedman 2000; and Gunlogson 2001 for the meaning of tones). Second, the results of Experiments 1 and 2 suggest an intermediate level that allows listeners to plan ahead which tones are still to come, and to infer the discourse-status of the not yet introduced referents. As we demonstrated in this paper, when listeners start to process an utterance which they know to be declarative, they anticipate a final falling tone. As long as this falling tone has not been perceived, listeners keep expecting it, and process the tonal contours associated on referents, verbs and so on, according to this expectation. More specifically, every rising tone is interpreted as a non-final one.

We assume a model of grammar in which the language forms which underlie production and perception are the same, similarly to the parity model (Liberman and Whalen 2000) or to the Optimality-Theoretic model, among others which make this assumption. Since each of these two aspects of language needs the other one to function, and since they evolved simultaneously, this is a reasonable assumption to make. In this view, models of auditory perception are identical to models of production.

A caveat in necessary however: modeling perception and production must be carefully distinguished from the 'actions' or processing involved in language. It is evident that language perception involves different functions from those needed in production. The articulatory gestures are only truly present in production, and even if theories like the motor theory are right in postulating an articulatory component in perception, it is clear that perception involves an auditory and a visual component which are subordinate in production.

In order to make our model more specific, we propose an Optimality-Theoretic view on intonation, adapted from Gussenhoven (2004), who shows how prespecified tones are aligned and associated with texts. In the following we consider both how perception as well as production can be expressed in Optimality Theory. As Smolensky (1996) first showed, Optimality Theory provides a framework in which the constraints entering the evaluation of candidates work in two directions, making this model of grammar ideal for evaluating both production and perception with the same tools.

We assume that sequences of tones associate and are aligned with segmental material or with edges of prosodic constituents in ways determined by specific constraint rankings. Our experimental material contained three kinds of tones: a bitonal falling pitch accent H*L, used to express a 'focus', which was the final tone in our sentences, a bitonal rising pitch accent L*H standing for 'topic' or simply prenuclearity, and a final boundary tone L_I, which is responsible for declarativity. Of course, a complete grammar of the intonation of German includes more than this small inventory of tones, but for the sake of this paper, it is sufficient. The starred tones H* and L* associate with metrical heads of the Intonation Phrase, H and L concatenate to the right of the starred tone and L_I aligns with the right edge of the Intonation Phrase. For the sake of simplicity, we assume that H*L and L*H are associate as a single entity with the segmental material. In a more precise grammar, each of the tone entering them would have to be dealt with independent constraints.

Some of the relevant constraints are listed in (9). The first constraints are Associate constraints which require that segmental material independently specified for focus or topic associate with certain tones. The metrically most prominent syllable of such information structural categories has to associate with the starred tone. The last constraint in (9) is an Align constraint which lets the right edge of an Intonation Phrase coincide with the boundary tone L_I.

(9) a. ASSOCIATE (Focus, H*L): Associate the prominent syllable of a focus with H*L.
b. ASSOCIATE (Topic, L*H): Associate the prominent syllable of a topic with L*H.
c. ALIGN (L_I,R): Align L_I with the rightmost syllable of the IP

Optimality-Theoretic tableaux (10) and (11) illustrate this simple grammar for our sentences. In both tableaux, candidate (a) is the winner, since it fulfills all constraints. The other candidates violate one constraint each. The constraints are not ranked, as they do not interfere which each other here. The input consists of the morpho-syntactic material with its information structure. The right boundary of the Intonation Phrase is also indicated in the input, since L_I is aligned with this boundary. In (10), the whole VP is focused, whereas in (11), only the verb is focused. The stress pattern of the words entering the input is calculated independently, and can be considered as part of the input. The lexical stress of the input words is underlined in the input. The small caps in the candidates indicate which syllable is supposed to be accented. Of course, a bitonal

tone such as H*L, which stands for focus, is realized on certain words in a focus domain. In (10), the argument of the verb carries the accent, rather than the verb itself (see Féry and Samek-Lodovici 2006 for constraints to this effect). And in (11), the verb itself carries the bitonal tone. In (10) and (11), there is no topic, and the second constraint is vacuously fulfilled.

(10) [befragt die PATIENTIN]$_F$

…[befragt die Patientin]$_F$]$_I$	ASSOC (Foc, H*L)	ASSOC (Top, L*H)	AL (IP, L_I ,R)
H*L L_I ☞ a. [befragt die PaTIENtin]$_F$]$_I$			
H* L L_I b. [befragt die PaTIENtin]$_F$]$_I$	*!		
L_I H*L c. [befragt die PaTIENtin]$_F$]$_I$			*!

(11) [befragt]$_F$ die Patientin

…[befragt]$_F$ die Patientin]$_I$	ASSOC (Foc, H*L)	ASSOC (Top, L*H)	AL (IP, L_I ,R)
H*L L_I ☞ a. [beFRAGT]$_F$ die Patientin]$_I$			
H*L L_I b. [beFRAGT]$_F$ die Patientin]$_I$	*!		
H* L L_I c. [beFRAGT]$_F$ die Patientin]$_I$	*!		
L_I H*L d. [beFRAGT]$_F$ die Patientin]$_I$			*!

The operations of association and alignment are further restricted by general constraints on the intonation structure which regulate the usual association of tones with tone bearing units (TBUs, in German, syllables). The four most important ones according to Gussenhoven (2004) are: (i) the fact that TBUs are usually associated with no more than one tone (NOCROWDING, NOCONTOUR), (ii) tones prefer to be associated with only one TBU (NOSPREADING), (iii) tones and TBUs do not delete and are not freely epenthesized (MAX, DEP), and (iv) they do not change their value, even if a markedness constraint like OCP militates in this direction (IDENT): in German, for instance, a hat pattern, implying two adjacent high tones is well tolerated.

Interpretative constraints attribute meanings to tones: a bitonal H*L for instance is interpreted as a focus, and a L tone at the end of an Intonation Phrase is interpreted as the boundary tone of a declarative sentence. In the same way, a bitonal rising tone gets an interpretation as a non-final or as a topical accent. In hearing such a tone on the stressed syllable of a constituent, hearers know that more material is to come. The interpretative constraints (12) are the mirror images of the constraints in (9).

(12) a. INTERPRET (H*L, Focus): Interpret H*L as a focus.
b. INTERPRET (L*H, Topic): Interpret L*H as a topic.
c. INTERPRET (L_I ,R): Interpret L_I as the boundary tone of a declarative IP.

Tableau (13) illustrates how a bitonal tone on *Patientin*, the object of a VP, allows us to formulate hypotheses about the focus structure of this VP. This tone structure is compatible with a narrow focus on *Patientin*, but also on the whole VP. It is, however, not compatible with a narrow focus on the verb. In (14), a bitonal accent on *befragt* is compatible only with narrow focus on this verb (at least when the whole VP is new in the discourse).

(13) H*L stands for a wide focus domain

H*L L_I, befragt die Patientin	INTERP (H*L,Foc)	INTERP (L*H, Top)	INTERP (L_I, IP, R)
H*L L_I ☞ a. [befragt die PaTIENtin]$_F$			
H*L L_I ☞ b. befragt die [PaTIENtin]$_F$			
H*L L_I c. [beFRAGT]$_F$ die Patientin	*!		

(14) H*L stands for a narrow focus domain

H*L L_I, befragt die Patientin	INTERP (H*L,Foc)	INTERP (L*H, Top)	INTERP (L_I, IP, R)
H*L L_I ☞ a. [beFRAGT]$_F$ die Patientin			
H*L L_I b. beFRAGT [die Patientin]$_F$	*!		

In the same way, hearers interpret a rising tone as a topic. Tableau (15) is the Optimality-Theoretic tableau to this effect. Candidate (15b) is eliminated because a rising tone is interpreted as a focus.

(15) L*H stands for a topic

L*H, der Arzt	INTERP (H*L,Foc)	INTERP (L*H, Top)	INTERP (L_I, IP, R)
L*H ☞ a. [der ARZT]$_T$			
L*H b. [der ARZT]$_F$		*!	

Evidently, the model we have offered in this section needs to be elaborated. But, even if the analysis is rather short for reasons of space, we think that we have demonstrated the existence and importance of a phonological level between the phonetic and the underlying level in the perception of intonation**.** In an Optimality-Theoretic approach like the one developed here, the phonological tones are interpreted as pragmatic discourse concepts, and pragmatic discourse concepts are realized as phonological tones. The speaker produces tones in order to highlight some constituents and the hearer interprets the tones correctly. This allows speakers and hearers to communicate and comprehend not only lexical and semantic contents, but also the discourse structure like focus, topic, new and given information.

6. Conclusion

In phonology, there has been a bias towards the study of production rather than perception. The reason for this difference is to be found in the poor access to the mechanisms underlying perception as compared to the possibilities offered by vocal tract modeling or experimental work on speech production. In this paper, we have presented two experiments bearing on the issue of perception. Our domain of investigation has been tonal contours in German and the expectations that listeners have, on the basis of the global intonation contour of a sentence, concerning the information structure of not-yet-perceived constituents. The first experiment was a forced-choice completion task using truncated sentences, where subjects had to choose between two possible object completions: one of them referring to a discourse-given and the other one referring to a discourse-new character. The results show an overall tendency in favor of the given constituent, which is confirmed by similar experiments we performed bearing on word order, not presented in this paper. However, this preference is modulated by accent type: participants were more likely to choose the given referent when they heard a sentence with an early falling accent than when they heard a tonal pattern giving them reasons to assume that the last accent was still to come, in which case participants were more likely to opt for the discourse-new referent. The second experiment was an eye-tracking experiment, using the complete versions of the sentences that were truncated in the first experiment. Thus, the participants heard the same sentences, but this time in their entirety. We tracked the eye movements of the listeners while they listened to a tonal pattern varying between early accent on the verb and late accent on the object, and assumed that the amount of fixation of the referents reflect the probabilities that the referents are considered as possible objects. Again, we saw a clear effect of the prosody on the preference for the upcoming, not yet heard object. As a whole, the two experiments indicate that not only do listeners use intonation to predict the discourse-status of upcoming referents, they do so very quickly and efficiently during on-line processing.

Altogether, our findings illustrate that in an intonation language like German, specific tonal contours, reflecting specific accent patterns and pragmatic meanings, can result in expectations regarding the discourse status of a yet-to-be-heard constituent. Furthermore, our experimental results show that a certain amount of anticipation helps the listeners to interpret tonal contours. The time frame inside of which such contours are processed is much longer than the one needed for the perception (and processing) of single segments or single words. Pitch excursions need language-specific phonologies in order to be interpretable. A model of perception must take this fact into account, and integrate a phonological level able to model the grammar of intonation.

References

Allopenna, Paul D., James S. Magnuson and Michael K. Tanenhaus 1998 Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language* 38: 419–439.

Baumann, Stefan 2005 *The Intonation of Givenness. Evidence from German*. Ph.D. dissertation, Saarland University, Saarbrücken.

Birch, Stacy, and Charles Clifton, Jr.1995 Focus, accent, and argument structure: effects on language comprehension. *Language and Speech* 38: 365–391.

Birner, Betty J., and Gregory Ward 1998 *Information Status and Noncanonical Word Order in English*. Amsterdam/Philadelphia: John Benjamins.

Boersma, Paul1998 *Functional Phonology: Formalizing the Interactions Between Articulatory and Perceptual Drives*. Ph.D. dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, Paulthis volume Cue constraints and their interactions in phonological perception and production.

Boersma, Paul, and Silke Hamann this volume Introduction: models of phonology in perception.

Bruce, Gösta 1977 Swedish word accent in sentence perspective. *Travaux de l'Institut de Linguistique de Lund* 12.

Büring, Daniel 1997 *The 49th Bridge Accent*. Berlin: Mouton de Gruyter.

Ćavar, Malgorzata 2004 *Perception and Production of Polish Palatals*. Ph.D. Dissertation, University of Potsdam.

Cinque, Guglielmo 1993 A null theory of phrase and compound stress. *Linguistic Inquiry* 24: 239–297.

Engbert, Ralf, and Reinhold Kliegl 2003 Microsaccades uncover the orientation of covert attention. *Vision Research* 43: 1035–1043.

Féry, Caroline 1993 *German Intonational Patterns*. Niemeyer, Tübingen.

Féry, Caroline, and Ruben Stoel 2006 Gradient perception of intonation. In: Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky and Ralf Vogel (eds.), *Gradience in Grammar: Generative Perspectives*, 145–166. Oxford: Oxford University Press.

Féry, Caroline, and Vieri Samek-Lodovici 2006 Focus projection and prosodic prominence in nested foci. *Language* 82: 131–150.

Flemming, Edward 1995 *Auditory Representations in Phonology*. Ph.D. Dissertation, Stanford University.

Grabe, Esther 1998 *Comparative Intonational Phonology: English and German*. Ph.D. Dissertation, Universiteit of Nijmegen.

Grice, Martine, Stefan Baumann and Ralf Benzmüller 2005 German intonation in autosegmental-metrical phonology. In: Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 55–83. Oxford: Oxford University Press.

Gunlogson, Christine 2001 *True to Form: Rising and Falling Declaratives as Questions in English*. Ph.D. dissertation, University of California at Santa Cruz.

Gussenhoven, Carlos 1983 Testing the reality of focus domains. *Language and Speech* 26: 61–80.

Gussenhoven, Carlos 2004 *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hruska, Claudia, Kai Alter, Karsten Steinhauer, and Anita Steube 2001 Misleading dialogues: Human's brain reaction to prosodic information. In: Ch. Cavé, I. Guaïtella and S. Santi (eds.), *Oralité et gesturalité: Interactions et comportements multimodaux dans la communication*. 425–430. Paris: L'Harmattan.

Hume, Elizabeth 1994 *Front Vowels, Coronal Consonants and their Interaction in Nonlinear Phonology*. New York: Garland.

Jacobs, Joachim 2001 The dimensions of topic-comment. *Linguistics* 39: 641 – 681.

Jannedy, Stefanie 2002 *Hat Patterns and Double Peaks: The Phonetics and Psycholinguistics of Broad Versus Late Narrow Versus Double Focus Intonations.* Ph.D. dissertation, Ohio State University.

Kaiser, Elsi, and John C. Trueswell 2004 The role of discourse context in the processing of a flexible word-order language. *Cognition* 94: 113–147.

Kohler, Klaus J. 1990 Macro and micro F0 in the synthesis of intonation. In: Kingston, John, and Mary Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, 115–138. Cambridge: Cambridge University Press.

Krahmer, Emiel, and Marc Swerts 2001 On the alleged existence of contrastive accents. *Speech Communication* 34: 391–405.

Ladd, D. Robert 1996 *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, D. Robert, and Rachel Morton1997 The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics* 25: 313–342.

Liberman, Mark Y. 1975 *The Intonational System of English*. Ph.D. dissertation, MIT.

Liberman, Alvin M., and Douglas H. Whalen 2000 On the relation of speech to language. *Trends in Cognitive Sciences* 4: 187–196.

Lindblom, Björn 1990 Explaining phonetic variation: A sketch of the H&H theory. In: William J. Hardcastle and Alain Marchal (eds.), *Speech Production and Speech Modeling*, 403–40. Dordrecht: Kluwer Academic Publishers.

Matin, Ethel, K.C. Shao, and Kenneth R. Boff 1993 Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics*, 53: 372–380.

McCarthy, John J., and Alan S. Prince 1993 Generalized alignment. In: Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1993*, 79–153. Dordrecht: Kluwer.

McQueen, James, and Anne Cutler 1997 Cognitive processes in speech perception. In: William J. Hardcastle and John Laver (eds.), *The Handbook of Phonetic Sciences*, 566–586, Oxford: Blackwell.

Nooteboom, Sieb G., and Johanna G. Kruyt 1987 Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America* 82: 1512–1524.

Ohala, John J. 1981 The listener as a source of sound change. In: Carrie Masek, Roberta A. Hendrick, Mary Frances Miller (eds.), *Proceedings of the Chicago Linguistics Society*, 178–203. Chicago.

Pierrehumbert, Janet 1980 *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.

Pierrehumbert, Janet, and Julia Hirschberg 1990 The meaning of intonational contours in the interpretation of discourse. In: Philip R. Cohen, Jerry Morgan and Martha E. Pollock (eds.), *Intentions in Communications*, 271–311. Cambridge: MIT Press.

Pierrehumbert, Janet, and Susan Steele 1987 How many rise-fall-rise contours? In: *Proceedings of the XI Meeting of the International Congress of Phonetic Sciences*, Volume 3, 145–147. Tallinn: Estonian Academy of Sciences.

Prince, Alan, and Paul Smolensky 2004 *Optimality Theory: Constraint Interaction in Generative Grammar*. Cambridge: MIT Press.

Reinhart, Tanya. 1981 Pragmatics and linguistics: an analysis of sentence topics. *Philosopica* 27: 53–94.

Rooth, Mats, 1985 *A Theory of Focus Interpretation*. Ph.D. dissertation, University of Massachusetts, Amherst.

Rooth, Mats, 1992 A theory of focus interpretation. *Natural Language Semantics* 1, 75–116.

Smolensky, Paul 1996 On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27: 720–731.

Stechow, Arnim von, and Susanne Uhmann 1986 Some remarks on focus projection. In: Werner Abraham and Sjaak de Meij (eds.), *Topic, Focus and Configurationality*, 295–320. Amsterdam: John Benjamins.

Steedman, Mark 2000 Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31: 649–689.

Steriade, Donca 1995 Positional neutralization. Unfinished manuscript, University of California, Los Angeles.

Steriade, Donca 1999 Phonetics in phonology: The case of laryngeal neutralization. In: Matthew Gordon (ed.), *UCLA Working Papers in Linguistics* 2: 25–46.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard and Julie C. Sedivy 1995 Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–1634.

Tanenhaus, Michael K., James S. Magnuson, Delphine Dahan and Craig Chambers 2000 Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research* 29: 557–580.

Welby, Pauline 2003 Effects of pitch accent position, type and status on focus projection. *Language and Speech* 46: 53–58.

Weskott, Thomas, Robin Hörnig, Elsi Kaiser, Caroline Féry, Sabine Kern and Reinhold Kliegl submitted Information structure and the anticipation of discourse referents: Effects of word order and intonation.