Hierarchical organization and tonal scaling[*]

Hubert Truckenbrodt and Caroline Féry

1 Introduction

1.1 Overview

In this paper we present and discuss the results of an experimental study on German intonation. A number of related goals are pursued.

For one thing, we are interested in replicating core results on English of Bob Ladd's (1988) classic paper. These results rest on a comparison of two different structural configurations among three clauses (domains of partial F0 resetting) A, B, and C, one employing the structure [A[BC]], and the other the structure [[AB]C]. The replication strengthens Ladd's core conclusions about effects of hierarchical structure on tonal scaling.

Another goal relates to a phenomenon of clause-final upstep that many German speakers show. Truckenbrodt (2002, 2007b) argues that upstep supports a particular implementation of Ladd's ideas suggested by van den Berg, Gussenhoven, and Rietveld (1992). We are here interested in studying the interaction of upstep with the clausal configurations [A[BC]] and [[AB]C]. Our observations will allow us to further strengthen Ladd's perspective and the implementation of it by van den Berg et al. (1992).

The results of the study bear on the formulation of the syntax-prosody mapping. The syntactic clause structures [A[BC]] and [[AB]C] must be mapped to isomorphic prosodic structures if they are to be able to affect the intonation. Reinforcing Ladd's conclusion about the different intonation of the two structures allows us to argue that suggestions of the *match theory* of Selkirk (2009, 2011) and a recursive model of the prosodic structure as proposed by Féry (2009, 2011, 2015) do better empirically than an account in terms of alignment and wrapping (Selkirk 2005, Truckenbrodt 2005).[1]

The remainder of this introduction reviews relevant background. After that, section 2 introduces the experimental method and section 3 presents the results. In section 4 we argue for a prosodic distinction between the two experimental conditions and discuss consequences for the

[1] The material presented here overlaps with Féry and Truckenbrodt (2005), a paper that builds on an earlier version of the current article (cited there as the manuscript Truckenbrodt and Féry 2004). The way in which final lowering among intonation phrases strengthens the conclusions about the crucial intermediate constituent (section 3.2) and the theoretical conclusions concerning the syntax-phonology mapping (section 4) seem to us important enough to publish in a revised version of this earlier manuscript, despite the time that has passed since then.

syntax-phonology mapping. Section 5 sums up our conclusions. Additional details of our method are given in an appendix.

1.2 Background on hierarchical organization and tonal scaling

Here we introduce some general background that leads up to a discussion of the results of Ladd (1988) and the model of van den Berg et al. (1992).

This paper adopts the autosegmental-metrical analysis of intonation developed in Pierrehumbert (1980), Beckman and Pierrehumbert (1986), Pierrehumbert and Beckman (1988) and later work (see Gussenhoven 2004, Jun 2005, 2014 and Ladd 2008 for recent cross-linguistic overviews and discussion). The F0-contour of an utterance is analyzed in terms of H and L tones. In intonation languages, these belong either to pitch accents (H*, L*+H etc.) or edge tones (here H%, L%). Turning points in the F0-contour are taken to be evidence for the presence of a H or L tone. Underlying the F0-contour is a language-specific tonal system, with specific phonological properties and pragmatic meanings (see e.g. Gussenhoven 2000, Pierrehumbert and Hirschberg 1990, Bartels 1999, Truckenbrodt 2012), as well as phonetic implementation (see e.g. Bruce 1977, Liberman and Pierrehumbert 1984, Pierrehumbert and Beckman 1988, Ladd and Schepman 2003). The present study is concerned with the phonetic implementation of the tones, in particular the assignment of phonetic values to H tones.

Background for our discussion is the phenomenon of *downstep* of successive H tones, i.e. the successive lowering of the phonetic values of the H tones under language-specific phonological conditions.[2] Initially in a new prosodic domain, such downstep can be undone in what is sometimes called a *reset*. Frequently found is the *partial reset* (Ladd 1988, van den Berg et al. 1992, Truckenbrodt 2002, 2007b, Laniran and Clements 2003), which represents an incomplete return to the utterance-initial F0 height. We are here interested in the suggestion of Ladd (1988, 1990) that partial reset involves lowering of one larger domain relative to another such larger domain. Implemented in terms of a horizontal reference-line that is lowered between the domains, as in van den Berg et al. (1992), this takes the form shown in (1). Downstep among accentual H tones proceeds away from the phrasal reference-line in each domain as shown by the dots in (1). The reset in the second domain is partial rather than complete because the phrasal reference-line to which the reset returns is itself lowered between the two domains. This lowering of the reference-line is also referred to as *downstep*, though it is downstep not among accents but among the reference-lines of two larger domains.

---

[2] Experimental studies of downstep include Liberman and Pierrehumbert (1984) on English, Pierrehumbert and Beckman (1988) on Japanese, Prieto et al. (1996) on Mexican Spanish, and Laniran and Clements (2003) on Yoruba. A discussion of different downstep triggers in different African languages can be found in Odden (1995).

(1)
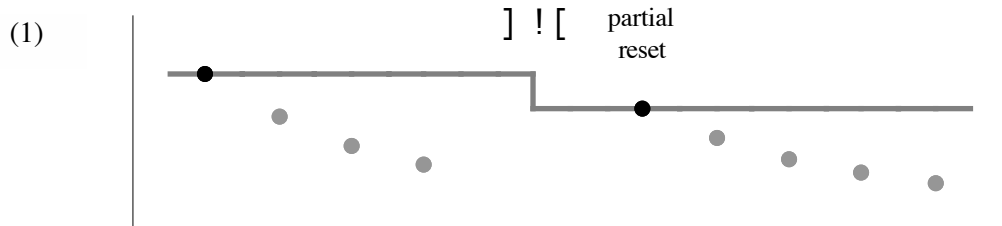
The specific combination of these earlier suggestions that we employ is that *prosodic sisterhood in the phonology corresponds to a lowering of the reference-line in the phonetics*. Evidence for the connection of sister-nodes to domain-lowering is provided by Ladd (1988). The stimuli of Ladd's experiment can be described as the embedding of one partial reset among sister-nodes within another partial reset among sister-nodes. The sentences in (2) illustrate his two experimental conditions.

(2)   but/and condition [A *but* [B *and* C]$_X$]: Ryan has a lot more money *but* [Warren is a stronger campaigner *and* Allen has more popular policies]$_X$
         and/but condition [[A *and* B]$_X$ *but* C]: [Allan is a stronger campaigner *and* Ryan has more popular policies]$_X$ *but* Warren has a lot more money

Both experimental conditions consist of three clauses A, B, and C. In the first condition, [B *and* C]$_X$ form an embedded constituent, while in the second condition, [A *and* B]$_X$ form such an embedded constituent. The embedded constituent thus formed, here labeled X, is then joined with the remaining clause on the highest level by the conjunction *but*, thus giving [A *but* X] in the but/and condition, and [X *but* C] in the and/but condition.

        Ladd's lowering among sister-nodes, modeled in terms of the phonetic reference-lines, is illustrated in (3) for these more complex cases. Intuitively it can be thought of as applying the schema seen in (1) twice in each condition. First, within X, there is lowering of a (here: grey) reference-line between the two clauses that are sisters inside of X (i.e. between B and C in the but/and condition and between A and B in the and/but condition). Second, the constituent X is also assigned a reference-line of constant height. The sisterhood-relation between X and its sister then leads to a second application of the schema in (1), so that there is lowering between X and its sister-node (i.e. between A and X in the but/and condition and between X and C in the and/but condition). This is shown by the black reference-line in (3). The three layers, indicated by the black reference-line, the grey reference-line and the dots, combine in the following way: the phonetic height of the leftmost element of one layer is equated with the reference-line of the next higher layer. Thus, in the but/and condition, the height of B, initial in the higher X, is equated with the height of X; in the and/but condition, the height of A, initial in the higher X, is equated with the height of X; and in each

3

clause, the leftmost accent H is equated in its height with the height of the reference-line of that clause.

(3)
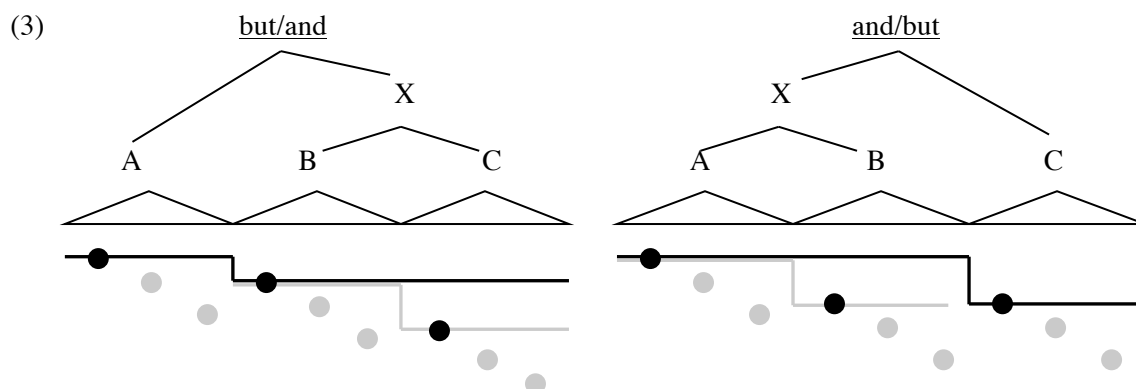
The predictions of this model can be assessed in the height of the clause-initial peaks. As shown, in both conditions, the first peak in B is predicted to be lowered relative to the first peak in A. This is due to the sister-relation [A X] in the but/and condition (left side of (3)), and due to the sister-relation [A B] in the and/but condition (right side of (3)). Further, in the but/and condition, B and C are sisters inside of X. Lowering among these sisters predicts that C is further lowered relative to B. The prediction for the but/and condition is therefore successive lowering among the clause-initial peaks of A, B, and C. This prediction could be derived by a number of alternative proposals that also account for the simple case in (1). For example, it could be maintained that with each larger domain-boundary, the reference-line is lowered by one step. Or it could be maintained that the lowering among the clause-initial peaks is due to global declination in the sense of Pierrehumbert (1980) (see also Ladd 1993, Shih 2000, 't Hart et al. 1990), an effect that generally makes values later in an utterance appear lower than values occurring earlier, all else being equal.

Crucially, then, the different structural configuration in the and/but condition predicts a different relative scaling between the clauses B and C. In this structural configuration, C is not sister to B. Rather, C is sister to X = [A & B]. Therefore C is predicted to be scaled one step lower than the reference-level of X = [A & B]. As shown in (3), this means C is expected to be scaled one level lower than the initial reference-level of A, but C is not predicted to be lowered relative to B.

This contrast between the two conditions was found in the English data studied by Ladd: initial-peak lowering was observed across speakers between A and B and between B and C in the but/and condition. The crucial and/but condition also showed initial-peak lowering between A and B across speakers and conditions; between B and C, however, there was either a smaller amount of lowering or even some raising, depending on speakers and the number of accented words. (We return to some other findings of Ladd 1988 in section 3.2.)

We find this to be a strong argument for a direct effect of higher structure on tonal scaling. An alternative account that postulates lowering with each larger domain-boundary will predict identical lowering among clause-initial peaks in both experimental conditions. Similarly, an account in terms of global declination without sensitivity to higher structure will amount to identical lowering in the two experimental conditions. Only the hypothesis that the lowering among clause-initial peaks (partial reset) is a phonetic reflex of higher sisterhood-relations correctly predicts that there is such lowering in three out of four cases of adjacent clauses, but crucially no such lowering in the fourth case, between B and C in the and/but condition.

1.3 German intonation and upstep

In this paper, we follow the suggestions for unified German ToBI transcriptions and tonal analysis developed in Grice and Baumann (2002) and Grice et al. (2005); the suggestions are further developed in Grice et al. (2009). Downtrends among high tones (downstep, final lowering, declination) in German are experimentally studied in some detail in Grabe (1998), Truckenbrodt (2004, 2007b) and in Féry and Kügler (2008).

Upstep in German, in the sense relevant here, is established in Truckenbrodt (2002, 2007b) and investigated in connection with focus in German in Féry and Kügler (2008).[3] For those speakers of German who use the contour L*+H (...) H% in intonation-phrase-final position when the utterance is continued after the intonation phrase, one or both of the two H tones are scaled to an upstepped level that is arguably comparable to the initial height of that clause. Truckenbrodt's phonetic analysis of upstep is illustrated in (4). Upstep, like reset, is claimed to be a return to the phrasal reference-line. Unlike reset, however, upstep occurs in domain-final position, and thus before the phrasal reference-line is downstepped. The upstepped peak is thus comparable in height to the earlier clause-initial peak, and separated from the following partial reset by the lowering of the phrasal reference-line between the two clauses.

---

[3] Pierrehumbert (1980) employs the term *upstep* in a different sense in her analysis of English, namely to denote a postulated process of raising of the lower and upper bound of the register by a H-phrase accent which affects a following L% or H% tone. In Grice and Baumann (2002), Grice et al. (2005) and Grice et al. (2009) the term *upstep* is used in a more restricted sense; there H-^H%, where '^' stands for upstep, transcribes a sequence in which H% is higher than H-. Further phenomena that are related to upstep are reported in Truckenbrodt (2004, 2005).

(4)
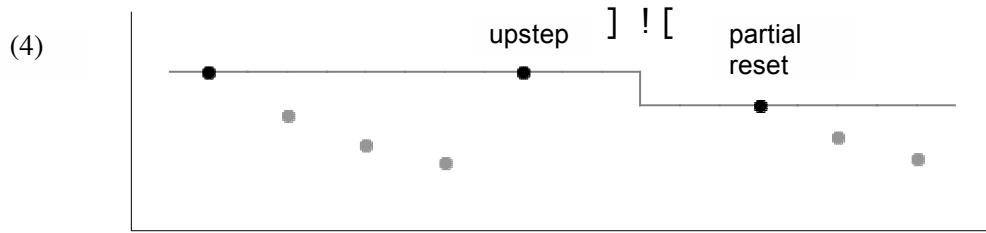
Upstep on nuclear H* accents in German is reported in Féry and Kügler (2008). There it occurs optionally on a nuclear accent in sentences with wide focus. It occurs obligatorily on a nuclear accent on an element with narrow focus.

The present study explores for German whether upstep provides independent support for Ladd's account of his two experimental conditions. (5) illustrates the expectations about the scaling of upstep in nested structures. Upstep is expected to target the respective clause-initial height in clauses A and B in the but/and condition, and in clause A in the and/but condition. Of particular interest is upstep in clause B of the crucial and/but condition. The model leads to the expectation that upstep at the end of B might here target the (black) phrasal reference-line of X = [A and B], just before it is lowered towards C (see the higher grey dot in (5) on the right). However, such a scaling of upstep in B is only one predicted possibility, the other one being upstep to the (grey) reference-line of B, as is also shown with the lower grey dot in (5) on the right. Since the grammar allows both possibilities, we expect variation that is reflected in an average value of this point.

(5)
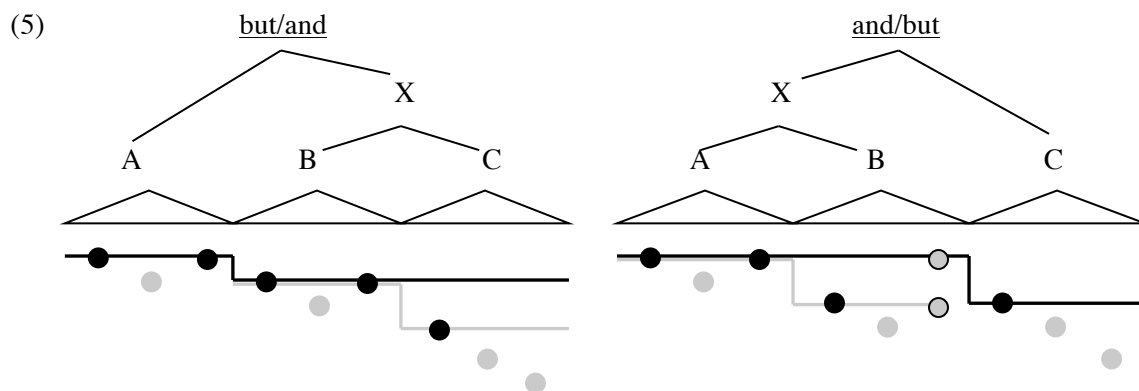
If the influence of the high reference-line of X could be documented in this condition, it would provide independent evidence for Ladd's account (in the implementation we are considering here). Recall that the explanation for why C is not lowered relative to B in the and/but condition is that C is lowered relative to its sister X = [A & B]. If, now, upstep in clause B of the and/but condition targets this reference-line of X = [A & B], we would have independent evidence for this reference-line in a position just before it is lowered for clause C.

2 The experiment

In the present section we illustrate the stimuli of our experiment with their prosodic and tonal analysis, and some aspects of the measurements. Additional details of the method are described in the appendix. The tonal analysis is the basis for the phonetic measurements. The phonetic values of the tones, in turn, are expected to reflect the different structures (we argue: prosodic structures) of the different experimental conditions.

2.1 Stimuli: AX- and XC-conditions

We conducted pilot studies, which led us to a design for our German stimuli that differs in some details from that of Ladd. The two contrastive conditions of our experiments contained only two expected accents per clause. The first peak in each clause would allow for the relevant comparisons among clause-initial peaks. The second peak in each clause would allow for an assessment of upstep. The three clauses (A, B, and C) were combined into the structures [A *während* [B *und* C]$_X$] and [[A *und* B]$_X$ *während* C] (translation: [A *while* [B *and* C]$_X$] and [[A *and* B]$_X$ *while* C), comparable to Ladd's but/and and and/but conditions. The conditions are here named *AX-condition* ('A *while* X') and *XC-condition* ('X *while* C'). Each stimulus was presented in the context of a question that draws attention to the important higher constituent X. This division semantically corresponds to the division induced by *während* 'while'. In the examples in (6), the clauses are separated by commas, and the crucial constituent X in both conditions is highlighted by square brackets. Underlining highlights the positions of word stress in which accents are expected, on which more will be said below.

(6)   a. AX-condition (corresponding to Ladd's but/and condition): A *während* [B *und* C]$_X$

Warum meint Anna, dass Handwerker teurere Autos haben als Musiker?
'Why does Anna think that craftsmen have more expensive cars than musicians?'

Weil	der	Maler	einen	Jaguar	hat,	*während*	[die	Sängerin	einen	Lada	besitzt,
because	the	painter	a	Jaguar	has	while	the singer		a	Lada	owns

und	der	Geiger	einen	Wartburg	fährt]
and	the	violinist	a	Wartburg	drives

'Because the painter has a Jaguar, while the singer owns a Lada, and the violinist drives a Wartburg.'

b. XC-condition (corresponding to Ladd's <u>and/but</u> condition): [A *und* B]_X *während* C

Warum meint Anna, dass Musiker teurere Autos haben als Sportler?
'Why does Anna think that musicians have more expensive cars than sportsmen?'

[Weil   die <u>Sä</u>ngerin einen <u>Ja</u>guar hat, *und* der <u>Gei</u>ger   einen <u>Daim</u>ler besitzt],
because the singer   a   Jaguar has and the violinist a   Daimler owns

während der <u>Ring</u>er   einen <u>La</u>da fährt.
while   the wrestler a   Lada drives

'Because the singer has a Jaguar, and the violinist owns a Daimler, while the
wrestler drives a Lada.'

16 stimuli of the AX-condition and 16 stimuli of the XC-condition were constructed in a comparable way, using the words shown in (7). Stressed syllables are underlined.

(7)   a. professions
               sportsmen: <u>Ru</u>derer, <u>Ring</u>er (rower, wrestler)
               craftsmen: <u>Ma</u>ler, <u>We</u>ber (painter, weaver)
               musicians: <u>Sä</u>ngerin, <u>Gei</u>ger (singer, violinist)
               hospital staff: Neuro<u>lo</u>ge, He<u>bam</u>me (neurologist, midwife)
       b. cars
               inexpensive cars: <u>La</u>da, <u>Wart</u>burg
               expensive cars: <u>Daim</u>ler, <u>Ja</u>guar
       c. verbs
               be<u>sitzt</u>, fährt, hat (owns, drives, has)

These words were permuted in the stimuli so as to eliminate effects of individual words on F0 height (microprosodic effects or effects of the position of stress within the word) in the comparison of clause-initial peaks with each other and in the comparison of clause-final peaks of the first and second clause with each other. In the 16 sentences of each condition, each of the eight professions occurred twice as the subject of the first clause, twice as the subject of the second clause, and twice as the subject of the third clause. Each of the four cars occurred four times as the object of the first clause, four times as the object of the second clause, and four times as the object of the third clause. Of the three verbs, *hat* occurred six times at the end of the first clause, six times at the end of the second clause and six times at the end of the third clause. The other two verbs each occurred five times at the end of the first clause, five times at the end of the second clause and five times at the end of the third clause.

With these stimuli, upstep on the second peak of the clauses A and B would be manifested as the absence of downstep in this position, i.e. the second peak would be comparable in height to the first. To test whether the subjects would otherwise show downstep among clause-internal peaks, a further set of stimuli was included, here called *no-X condition*. These stimuli also consisted of three clauses, though this time with the simpler structure [A, B, and C]. The clauses here each contained three expected accent locations. An example is shown in (8).

(8)   no-X condition:

Warum meint Anna, dass ihre Nachbarn teure Autos haben?

Weil M<u>öl</u>ler und <u>Hum</u>mel einen <u>Ja</u>guar haben, <u>Mey</u>er und <u>Ler</u>ner einen <u>Daim</u>ler besitzen, und <u>Woll</u>mann und <u>Leh</u>mann einen BM<u>W</u> fahren.

'Why does Anna think that her neighbors have expensive cars?

Because Möller and Hummel have Jaguars, Meyer and Lerner own Daimlers, and Wollmann and Lehmann drive BMWs.'

The 18 stimuli of the no-X condition all used the words shown in this illustration. Here the six family names, the three car names, and the three verbs were again systematically permuted in order to avoid influences of individual words on the results.

2.2 Prosodic analysis of the stimuli

The prosodic analysis employed is illustrated in (9) with an example from the AX-condition. Two phrasal prosodic levels are postulated. The lower level is that of the phonological phrase, which we refer to as *p-phrase* in the following. At this level one accent heads each p-phrase. These domains and the positions of the accents can be predicted by the SAAR of Gussenhoven (1983, 1992), by the modified analysis of Truckenbrodt (2007a), or in the analysis proposed by Féry (2011) using recursive prosodic phrasing. The stimuli are constructed to contain two p-phrases in each clause A, B, and C.

Further, each clause itself is expected to form a higher prosodic domain. We analyze these as intonation phrases (Beckman and Pierrehumbert 1986, Nespor and Vogel 1986, Selkirk 2005), which we refer to as *i-phrases* in the following. We return to this issue in section 4. The strongest (nuclear) stress in each i-phrase is taken to be assigned in the rightmost p-phrase (Uhmann 1991, Selkirk 1995a, 2005, Truckenbrodt 2007a, Féry 2011). In non-final i-phrases, the position of nuclear stress is the position of upstep in our experiment. An important issue is whether the conjunction of two clauses in the syntactic constituent X is mirrored in the prosodic structure by a further i-phrase. We indicate such an i-phrase in (9) and (10) and return to this question below. The prosodic analyses of all three conditions are schematically shown in (10).

9

(9)   [x   ]   i-phrases
      (x)(x   )   p-phrases
      Weil der <u>Ma</u>ler einen <u>Ja</u>guar hat,

      ⌈[x   ][x   ]⌉ i-phr.
      (x   )(x   )(x   )(x   ) p-phr.
      während die <u>Sä</u>ngerin einen <u>La</u>da besitzt, *und* der <u>Gei</u>ger einen <u>Wart</u>burg fährt

(10)   AX-condition   XC-condition
         A   B   C   A   B   C
      [x] ⌈ [x] [x]⌉   ⌈[x][x]⌉ [x]
      (x)(x) (x)(x) (x)(x)   (x)(x)(x)(x)   (x)(x)

      no-X condition
         A   B   C
      [x][x][x]   i-phrases
      (x)(x)(x)(x)(x)(x)(x)(x)   p-phrases

2.3 Phonetic evaluation of the stimuli

Five speakers were recorded. They are here referred to as S1 to S5. The appendix provides additional details about the speakers, the recordings and the measurements. The current section presents aspects of the phonetic evaluation that are particularly relevant to an understanding of our results.

A tonal analysis was fit to each token recording, based on the tonal inventory of German ToBI (Grice and Baumann 2002, Grice et al. 2005; see also Grice et al. 2009). The most frequent tonal pattern found across speakers is illustrated in (11) and (12). Example (11) is from the AX-condition and illustrates the tonal pattern found in both the AX- and the XC-condition. Example (12) illustrates the no-X condition.

(11)   Warum meint Anna, dass Handwerker teurere Autos haben als Musiker?
       'Why does Anna think that craftsmen have more expensive cars than musicians?'

              L*+H   L*+H   H%
      [Weil der <u>Ma</u>ler einen <u>Ja</u>guar hat]

               L*+H   L*+H   H%
      [während die <u>Sä</u>ngerin einen <u>La</u>da besitzt]

               L*+H   (H+)L*   L%
       [und der <u>Gei</u>ger einen <u>Wart</u>burg fährt]

       'Because the painter has a Jaguar, while the singer owns a Lada, and the violinist drives a Wartburg.'

(12)   Warum meint Anna, dass ihre Nachbarn teure Autos haben?
        'Why does Anna think that her neighbors have expensive cars?'

            L*+H   L*+H   L*+H   H%
        [weil Lehmann und Möller einen BMW haben]

            L*+H   L*+H   L*+H   H%
        [Hummel und Meyer einen Jaguar besitzen]

            L*+H   L*+H   (H+)L*   L%
        [und Lerner und Wollmann einen Daimler fahren]

        'Because Lehmann and Möller have BMWs, Hummel and Meyer own Jaguars,
        and Lerner and Wollman drive Daimlers.'

    A decision was made to base the phonetic results on measurements of the tones shown in
(11) and (12): non-final rising L*+H pitch accents and non-final high H% boundary tones, final
falling L* or H+L* pitch accents and a final low L% boundary tone.

    The phonetic analysis only included token recordings that showed clause-initial L*+H
accents in all three clauses A, B, and C, as well as H% boundary tones at the end of both the first and
second clauses. A second pitch accent of clauses A and B that differed from a L*+H rise was not a
reason for excluding the token from analysis. Instead, the measurement of this pitch accent was then
skipped.

    In most cases, at least 14 out of 16 recordings for the AX-condition and at least 14 out of 16
recordings for the XC-condition could be included in the measurements of a given speaker by our
criteria. The two exceptions to this are the XC-condition of speaker S2, where our criteria led to 6 out
of 16 tokens being measured, and the AX-condition of speaker S3, where the criteria allowed for
measuring 7 out of 16 token recordings. In the clauses A and B of the AX- and XC-conditions taken
together one measurement of the nuclear accent (L*+H) had to be skipped. See section 4 of the
appendix for an overview of the data retained for analysis.

    For the measurements of upstep in the sequence L*+H H%, only one phonetic value, the
highest in the area of +H H%, enters into the phonetic evaluations below. As shown in Truckenbrodt
(2007b), upstep in such tonal sequences may occur either on the clause-final H% or on the preceding
+H of the nuclear rise, or on both. Therefore, the best approximation to the upstepped height seems to
be the highest value in this area for each token.

3 Phonetic height that reflects hierarchical structure

Section 3.1 shows the results of the no-X condition. In section 3.2, our crucial replications of findings of Ladd (1988) are presented with results of the AX- and XC-conditions. Section 3.3 shows our findings for upstep in the AX- and XC-conditions.

The measured values were normalized. The normalization uses the following linear transformation.

(13)   *transformed_value = (measured_value - L%[speaker])/(H1[speaker] - L%[speaker])*

L%[speaker] and *H1[speaker]* are calculated on the pooled values of the AX- and XC-conditions. *L%[speaker]* is the average of the L% value (at the end of clause C) for a speaker, and *H1[speaker]* is the average of the H1 values (initial in clause A) for the speaker.

The crucial results rest on the comparison of the AX- and the XC-conditions. The no-X condition is longer than the other two conditions. Therefore the normalization, including the values of L%[speaker] and H1[speaker], was calculated separately for the no-X condition.

3.1 No-X condition

Figure 1 shows the averaged measurements of the recordings of the no-X condition. They are normalized and pooled across the five speakers.
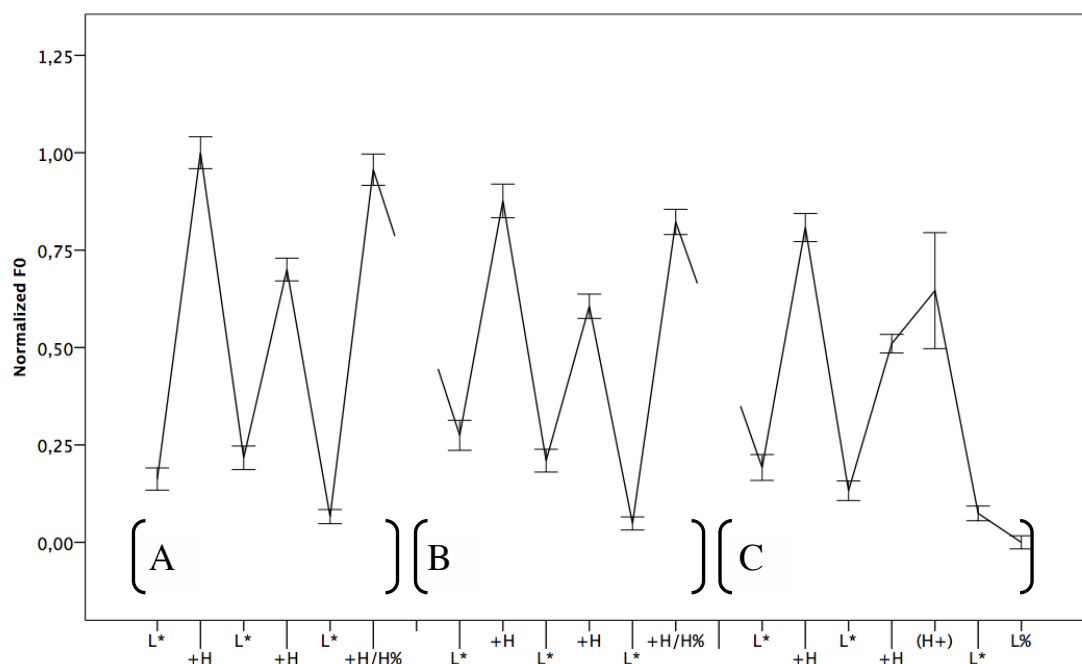
Figure 1. Pooled normalized measurements and 95% confidence intervals of the no-X condition. The brackets at the bottom of the plot separate the values within the three clauses A, B, and C. ("(H+)" is a measurement point related to a final H+L* in clause C only for S1, S3, and S4, who employ a final H+L* in clause C.)

The no-X condition allows us to make the following points. First, there is a clear pattern of downstep between the first and the second H peak in each clause. Second the high values at the end of the clauses A and B do not continue the downstepping pattern of the first two peaks of the clause. Rather the last high values return to approximately the height of the initial peak of their clauses, i.e. they display upstep. Together, these two observations show that there is i-phrase-internal downstep and i-phrase-final upstep for the speakers we recorded.

The AX- and the XC-conditions contain only an initial and a final peak in each clause. In those cases, a similar height of the second, clause-final peak to the first, clause-initial peak might simply seem like the absence of downstep or upstep. However, the no-X condition shows that the recorded speakers show a pattern of downstep and upstep relative to the i-phrase as has been described in the previous studies. If, therefore, the AX-condition and the XC-condition show a similar height of the second peak to the first in non-final clauses, we are justified in analyzing this as upstep preceding an i-boundary.

We do not analyze the relation between the three clauses in the no-X condition. We think that it is possible that individual speakers superimpose their own structure [AB]C or A[BC], since none of these options is enforced by the experiment (see Kentner and Féry 2013 for evidence that speakers do impose such a structure on a ternary sequence of names).

13

The preceding discussion has shown the points that we want to make with reference to the no-X condition. We will not return to the no-X condition below.

3.2 Comparison of clause-initial peaks in AX- and XC-conditions: Evidence for the effect of hierarchical structure on tonal scaling

Figure 2 shows the comparison of the three clause-initial peaks for the pooled normalized data of all five speakers. These clause-initial peaks allow for an assessment of the scaling of the three clauses relative to each other.
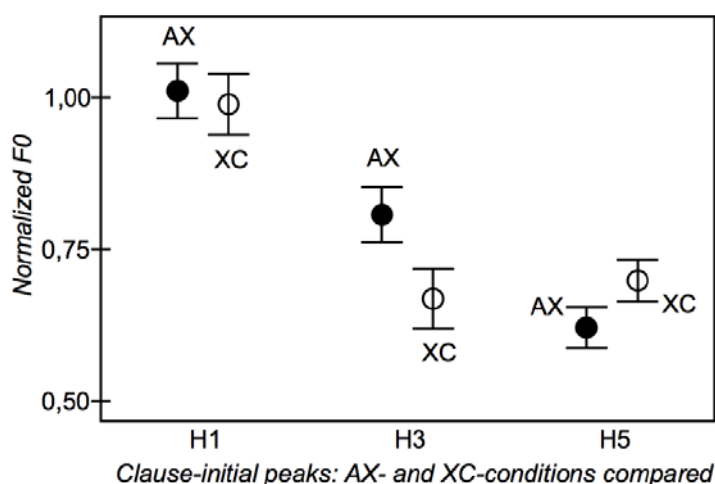

Figure 2. Averaged values and 95% confidence intervals for the clause-initial peaks in the normalized pooled values of speakers S1–S5. H1, H3, and H5 are the respective initial peaks of the clauses A, B, and C. Filled dots represent the AX-condition, circles represent the XC-condition.

Figure 2 shows a replication of the crucial result of Ladd (1988) that we reviewed above. In the AX-condition (Ladd's but/and condition), there is lowering not just between clauses A and B (here: between H1 and H3) but also between clauses B and C (here: between H3 and H5). By contrast, the XC-condition (Ladd's and/but condition) shows lowering between clauses A and B (here: between H1 and H3), but not between clauses B and C (here: between H3 and H5). This is confirmation of Ladd's claim that higher structure affects tonal scaling as well as of the particular effect of higher structure on tonal scaling that Ladd postulated: lowering among larger domains affects nodes that are structural sisters in the representation. As reviewed in (14), this correctly predicts the absence of lowering between clauses B and C in the XC-condition, where C is not downstepped relative to B but relative to its structural sister X = [A & B] and thus relative to the initial height of A. At the same time, it correctly predicts lowering among all other adjacent clauses, including lowering between B and C in the AX-condition, where C is a structural sister of B.

14

(14)   AX-condition   XC-condition

        A   [B   C   ]x   [A   B   ]x   C

        H1   H3   H5   H1   H3   H5
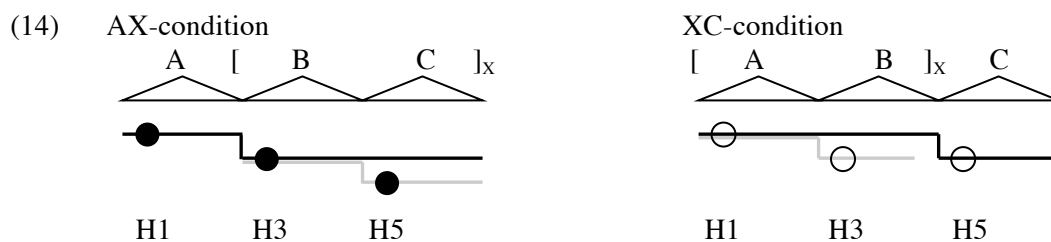
We also replicated other results of Ladd (1988). These are discussed with reference to Figure 2 as well. For one thing, as in Ladd's data, the utterance-initial peaks (H1) in both conditions are comparable in height. Ladd's model does not lead us to expect a difference in utterance-initial height, and it is thus reasonable to expect that a similar utterance-initial level is used by the speakers in both conditions.

Second, there is a clear difference in the height initially in clause B between the two experimental conditions. In Figure 2, this is visible at point H3, which is higher in the AX-condition than in the XC-condition. The difference is highly significant.[4] The difference is similarly found by Ladd (1988) but not predicted by the core of Ladd's model reviewed above, since clause B is lowered by one step relative to clause A in both conditions. Ladd (1988) describes this in a formulation that correlates boundary size with the strength of the reset: "Clause-initial accent peaks are higher following a stronger boundary" (p.541).[5] We here relate this to final lowering. Final lowering is known as a factor in sequences of downstepping accents. These have the shape of exponential decay towards an abstract reference-line in their non-final accents; however, the final accent deviates from this pattern and shows a step of downstep that is deeper than expected (Liberman and Pierrehumbert 1984). The relevant sense of 'final' could be utterance- or phrase-final or 'final in the downstep sequence' in the English data of Liberman and Pierrehumbert (1984) and in the Mexican Spanish data of Prieto et al. (1996). Truckenbrodt (2004) argues that at least some amount of final lowering is found in accent sequences in German in the position 'final in the downstep sequence' even where this is not final in an utterance or final in an i-phrase. In the cases at hand this does not make any additional predictions.

We suggest generalizing the concept of final lowering from the height of accents to the height of phrasal reference-lines. (15) shows how the different senses of finality apply to these. As shown, H3 (i.e. the reference-height of clause and i-phrase B) is final in X in the XC-condition but not in the AX-condition. If the reference-line of B undergoes final lowering in the XC-condition

---

[4] H3 in AX: 0.80, in XC: 0.69; $t_{66} = 4.14$, $p < 0.001$ in a two-tail t-test. Since the total number of normalized measurements in position H3 is 67 in AX and 66 in XC, the average of XC was added to the XC-values once in this calculation.

[5] In Féry & Truckenbrodt (2005), the difference in the height of H3 in the two conditions was accounted for by a principle that we called 'the deeper the steeper' to the effect that the more deeply embedded the constituents are, the steeper the downstep is.

because it is final within the larger X here, then final lowering can account for the difference between the two values of H3 in Figure 2.

(15)   Predictions of final lowering and downstep at the phrasal level

	AX-condition [ₓ　　　　　　　]			XC-condition [ₓ　　　　　]		
	H1	H3	H5	H1	H3	H5
a. final lowering			final in X final in utterance		final in X	final in utterance
b. downstep		one step down	two steps down		one step down	one step down

A similar account can be given for the distinction in height between H3 of the AX-condition and H5 of the XC-condition in Figure 2. These have each undergone only one step of downstep among phrases (in the case of H5 of the XC-condition, this is the crucial step of lowering relative to the constituent X = [A B]). Therefore downstep among phrases alone would lead us to expect that they would be of comparable height. The fact that H5 of the XC-condition is nevertheless lower than H3 of the AX-condition can be explained in terms of final lowering. Clause B of the AX-condition, as we have seen, is not final in any sense. Clause C of the XC-condition, on the other hand, is final in the utterance. Final lowering in its application to phrases can thus explain that H5 in the XC-condition is lower than H3 in the AX-condition.

Taken together, downstep among phrases and final lowering applied to phrases, both crucially affected by the constituent X, provide a good analysis of the values in Figure 2. The analysis thus strengthens the claim of the relevance of higher structure for tonal scaling both in terms of the effect of X on downstep among phrases, and in terms of the effect of X on final lowering applied to phrases.

In sum, we have replicated core findings of Ladd (1988) in regard to clause-initial peaks in the two experimental conditions A[BC] vs. [AB]C. The most important aspect is that C is lowered relative to B in A[BC], where C is sister to B, while C is not lowered relative to B in [AB]C, where C is not sister to B. In the latter case C is lowered relative to its sister [AB] instead. This provides confirmation of a core aspect of Ladd's proposal about the effect of hierarchical structure on tonal scaling: lowering among higher domains is systematic, and tied to a structural configuration of sisterhood. We have also replicated other aspects of his results, which we analyze in terms of final lowering applied to phrases. We have argued that this provides additional support for the role of the higher structure in tonal scaling.

16

3.3 Relative height of upstepped values

(16) shows our expectations and allows us to illustrate the test question concerning upstepped values. We expect the upstepped values to be of the same height as the immediately preceding clause-initial values in clauses A and B of the AX-condition (H1 ≈ H2, H3 ≈ H4), and in clause A of the XC-condition (H1 ≈ H2). To the extent that this is the case, we can test the question we are interested in, which concerns upstep in clause B of the XC-condition. The relevant tone is the upstepped H4 of the XC-condition, which we call H4/XC in the following. Does H4/XC target the reference-line at the initial height of B (H3 ≈ H4 in XC), or does upstep here find the abstract reference-line of the constituent X, just before this reference-line is lowered for clause C in the XC-condition (H1 ≈ H4 in XC)?

(16)   AX-condition   XC-condition

A   [B   C   ]ₓ   [A   B]ₓ   C

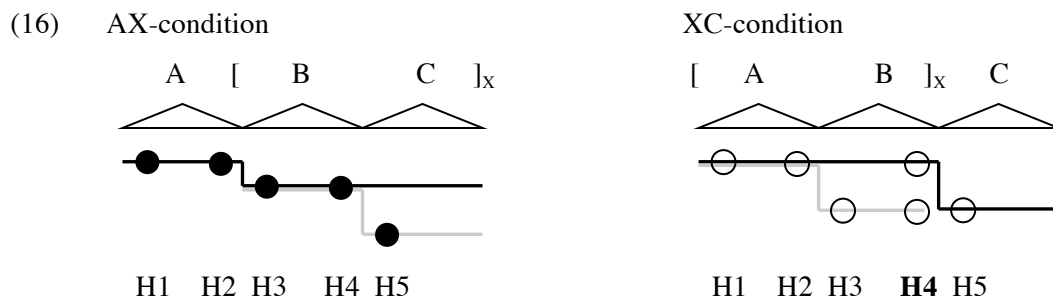H1   H2 H3   H4 H5   H1   H2 H3   **H4** H5

Table 1 shows a comparison of the clause-initial peaks H1 and H3 with the upstepped values H2 and H4, separated by speakers. The first three columns show the comparisons where we expect no difference in height between a clause-initial peak and the following upstepped peak. The last two columns bear on the test-case in the XC-condition and compare H4 with both H3 and H1 in the XC-condition. Distinctions that are significant after Bonferroni-adjustment for each speaker (p < 0.05/5) are boldfaced.

	AX		XC		
	H1, H2	H3, H4	H1, H2	H3, H4	H1, H4
S1	282 Hz, 281 Hz $t_{13} = 0.17$ $p > 0.05$	245 Hz, 244 Hz $t_{13} = 0.46$ $p > 0.05$	278 Hz, 279 Hz $t_{15} = -0.21$ $p > 0.05$	249 Hz, 258 Hz $t_{15} = -2.08$ $p > 0.05$	**278 Hz, 258 Hz** **$t_{15} = 5.47$** **$p < 0.001$**
S2	261 Hz, 262 Hz $t_{14} = -0.21$ $p > 0.05$	244 Hz, 233 Hz $t_{14} = 2.88$ $p = 0.012$	256 Hz, 260 Hz $t_5 = -0.45$ $p > 0.05$	236 Hz, 237 Hz $t_5 = -0.18$ $p > 0.05$	**256 Hz, 237 Hz** **$t_5 = 4.13$** **$p = 0.009$**
S3	273 Hz, 269 Hz $t_6 = 0.59$ $p > 0.05$	234 Hz, 262 Hz $t_6 = -2.90$ $p = 0.027$	262 Hz, 268 Hz $t_{12} = -0.68$ $p > 0.05$	**204 Hz, 260 Hz** **$t_{12} = -7.53$** **$p < 0.001$**	262 Hz, 260 Hz $t_{12} = 0.17$ $p > 0.05$
S4	140 Hz, 146 Hz $t_{14} = -1.68$ $p > 0.05$	134 Hz, 137 Hz $t_{14} = -1.14$ $p > 0.05$	141 Hz, 146 Hz $t_{14} = -1.26$ $p > 0.05$	**121 Hz, 139 Hz** **$t_{14} = -4.99$** **$p < 0.001$**	141 Hz, 139 Hz $t_{14} = 0.61$ $p > 0.05$
S5	**163 Hz, 183 Hz** **$t_{15} = -7.43$** **$p < 0.001$**	**156 Hz, 169 Hz** **$t_{15} = -6.28$** **$p < 0.001$**	**163 Hz, 186 Hz** **$t_{15} = -6.87$** **$p < 0.001$**	**151 Hz, 171 Hz** **$t_{15} = -8.60$** **$p < 0.001$**	163 Hz, 171 Hz $t_{15} = -2.76$ $p = 0.014$
	Upstep to expected levels? (Premise for interpr. of H4/XC)		H4/XC: upstep to level of H3 or of H1?		

Table 1. Comparisons of upstepped values H2 and H4 with earlier clause-initial values H1 and H3, using paired-sample t-tests. Boldfacing highlights the differences that are significant after Bonferroni-adjustment for each speaker ($p < 0.05/5$, i.e. $p < 0.01$).

Notice that speaker S5 shows significant distinctions in the first three columns. For this speaker, then, upstep does not normally target the clause-initial height in this data; rather, upstep is unexpectedly high. We therefore put this speaker aside for the moment, and return to him below. For the remaining four speakers, there are no significant distinctions in the first three columns of Table 1, i.e. upstep is broadly comparable in height to the immediately preceding clause-initial value in the three cases where we expect equal height. For these four speakers, S1–S4, Figure 3 shows pooled normalized values of clause-initial peaks and upstepped peaks.
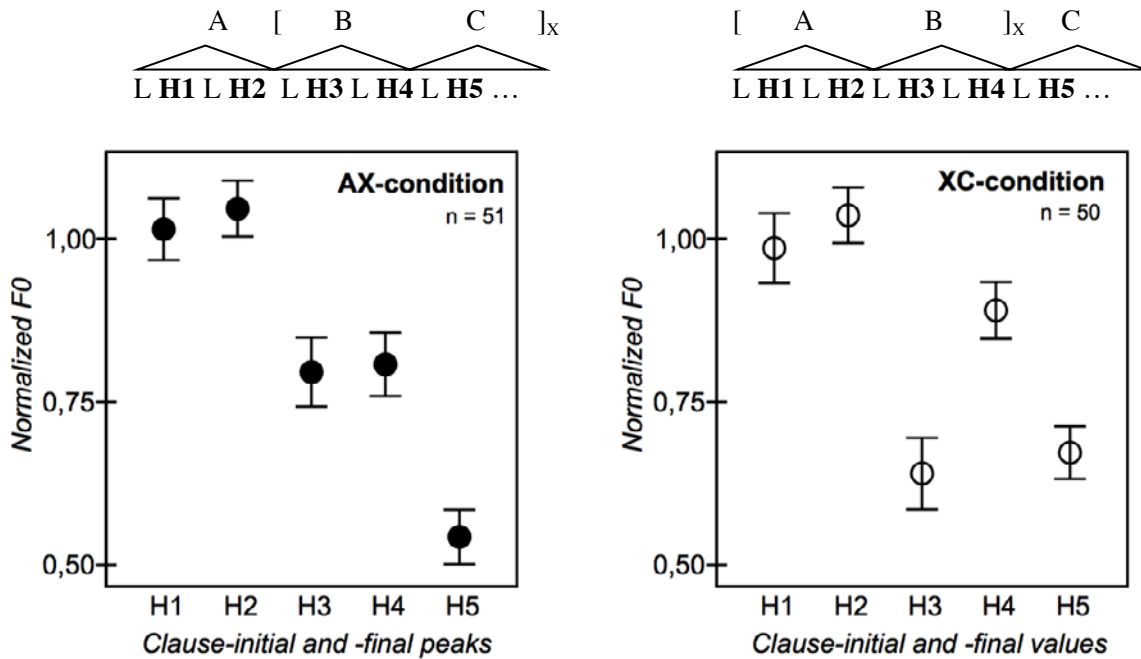
Figure 3. 95% confidence intervals for upstepped values H2 (clause A) and H4 (clause B) in relation to the three clause-initial values H1, H3, and H5 in the normalized pooled data of speakers S1–S4.

The plots of Figure 3 illustrate that H1 ≈ H2 and H3 ≈ H4 in the AX-condition, and H1 ≈ H2 in the XC-condition. For the crucial test-case, it can be seen that H4/XC is much higher than the preceding H3. This strongly suggests that the high reference-line of X plays a role as a target for upstep in H4/XC.

H4/XC is also lower than H1 in XC in Figure 3. The results in the two righthand columns of Table 1 suggest that speakers show individual preferences as to the scaling of H4/XC. These results are compatible with speakers S1 and S2 scaling H4 to the lower reference-line of B in (16), and speakers S3 and S4 scaling H4 to the higher reference-line of [AB] in (16). The latter group might thus primarily be responsible for the greater height of H4 relative to H3 in Figure 3, XC-condition. The former group might primarily be responsible for the lower height of H4 relative to H1 in the same plot. This is compatible with the model we are pursuing, which allows both options of scaling H4 in (16).

For completeness, we briefly return to speaker S5, whom we put aside earlier. Figure 4 shows the data of this speaker separately.
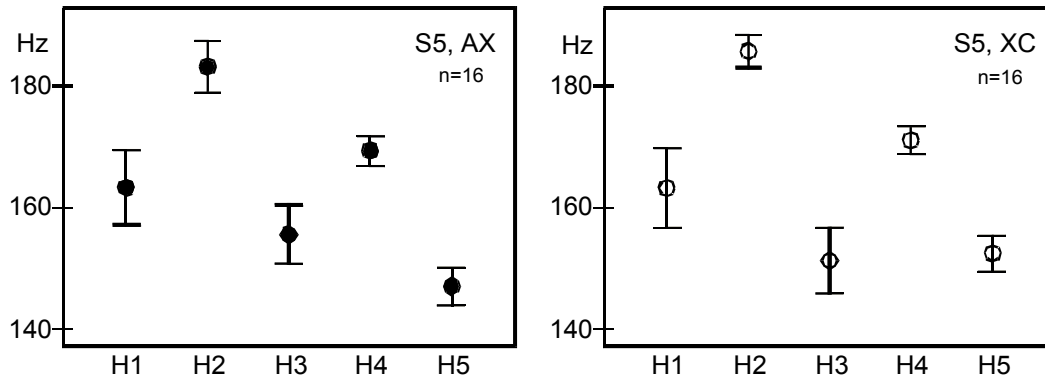
Figure 4. Averaged values and 95% confidence intervals for the clause-initial and upstepped peaks of speaker S5.

It is clear in both conditions that each upstepped value H2 and H4 is higher than the peak immediately preceding it. We interpret this in terms of the suggestion of Kentner and Féry (2013), who formulate a principle called *Proximity* to express the fact that speakers reduce the first boundary in a phrase grouping two constituents in the same p-phrase. Anti-proximity accounts for the fact that speakers increase a boundary before a following boundary to express the separation of constituents. Petrone et al. (2014) argue for an understanding of Proximity as a strategy that speakers may use over and above the default prosodic mapping: where Y and Z are otherwise mapped to separate p-phrases, a constituent [Y Z] can be made salient by being mapped to a single p-phrase. In the current experiment, speaker S5 seems to have made such an effort in the domains [H1 H2] and [H3 H4].

4 On the prosodic representation of the two experimental conditions

In this section we discuss the recursive i-phrasing of the clauses A, B, C, and X in the experiment.

Ladd (1986, 1992) proposes that the constituent that is here labeled X forms a compound intonation phrase, consisting of two embedded intonation phrases. This is a compound intonation phrase of B and C in the AX-condition, A[BC], and a compound intonation phrase of A and B in the XC-condition, [AB]C. We think that this is correct for principled reasons. Much of the work in prosodic phonology since Selkirk (1980) and Nespor and Vogel (1986) (see Truckenbrodt 2007a and Selkirk 2011 for review) has been guided by the *Indirect Reference Hypothesis* (Inkelas 1989:9), according to which phonological rules may not access syntactic structure directly. Rather, syntactic structure, in these accounts, plays a role in conditioning a prosodic constituent structure, and only this prosodic constituent structure may be referred to by the phonology. On standard assumptions, the interface between phonology and phonetics occurs after the mapping from syntax to phonology. Therefore, the phonetic implementation (such as the determination of tonal height) will not have access to the syntactic structure either. Syntactic structure can only indirectly affect the phonetic scaling, to the extent that it leads to prosodic structure that affects the phonetic implementation. (See Kabagema-Bilan et al. 2011 for a more detailed discussion of this issue.) It also seems that this

position converges with the standard assumption that it is the prosodic structure that affects the phonetic implementation (Pierrehumbert 1980, Pierrehumbert and Beckman 1988, Kent and Netsell 1971, Fougeron and Keating 1997 and much recent work following up on them).

On this view, Ladd's (1988) results and ours provide empirical support for recursive i-phrasing: $[A]_I[[B]_I[C]_I]_I$ in the AX-condition and $[[A]_I[B]_I]_I [C]_I$ in the XC-condition. At the level of the individual clauses A, B, and C, there is evidence that each of them constitute separate i-phrases in the observation that they are domains of partial F0-resetting in Ladd's experiment, and domains of upstep in our experiment. At the same time, the higher i-phrases X = [BC] in the AX-condition and X = [AB] in the XC-condition must also be mapped to i-phrases, so that the distinctions between the two experimental conditions can be accounted for.

How, then, does the syntax-phonology mapping derive that each of the syntactic constituents X, A, B, and C turn into coextensive i-phrases? Since this is a mapping to an isomorphic structure, one might think that it should be straightforward. However, it turns out that not all mapping accounts are equipped to derive the isomorphic recursive structure.

The three clauses A, B, and C of Ladd's experiment are root sentences in the sense of Downing (1970): they are not embedded in a higher clause that has a predicate of its own. Similarly, the combined clause [AB] in the XC-condition and the combination [A B C] or all three clauses are root sentences. We think that our stimuli are best viewed in the same way. Though our sentences have the shape of typical embedded German clauses (verb-final word order), they occur without overt embedding in our experiment. Downing (1970) argues that root sentences are obligatorily separated by intonation phrase boundaries from surrounding material at their left and right edges. If we think of this suggestion in terms of syntax-prosody alignment (Selkirk 1986, 1995b), however, the two experimental conditions are wrongly mapped to identical prosodic structures, as shown in (17). (17a) is a sketch of the syntactic structure. (17b) shows the i-phrasing derived by left- and right-alignment of clause boundaries with i-phrases.

(17)   Insufficient mapping by syntax-prosody alignment of clauses with i-boundaries

	AX-condition						XC-condition					
a.	A	[	B	C	]$_X$		[	A	B	]$_X$	C	
b.	[	] [	] [	]			[	] [	] [	]		

Each of the clauses A, B, and C would form an i-phrase, delimited at its left and right edges by i-boundaries. The resulting i-boundaries would also satisfy the alignment requirement for the higher constituents. In particular X = [BC] in the AX-condition now also has an i-boundary at its left edge, and an i-boundary at its right edge. Similarly, the constituent X = [AB] in the XC-condition has an i-boundary at its left edge and an i-boundary at its right edge. The account would not derive the

prosodic distinction between the two conditions that gives rise to the different clause-initial height relations in Ladd's experiment or in ours.

No improvement can be gained by the addition of a Wrap-constraint to the account. The constraint WRAP-XP is suggested to interact with edge-alignment of XPs in Truckenbrodt (1995, 1999). A Wrap-constraint for the relation between clauses and i-phrases is formulated in Selkirk (2005) and Truckenbrodt (2005), with differences in detail that are not relevant here. The requirement, applied to the structures at hand, is in both formulations that *each clause must be contained in an i-phrase*. The addition of this constraint would derive the prosodic structures in (18b) if wrapping suppresses alignment, and the prosodic structure in (18c) if wrapping does not suppress alignment.

(18)   Wrong mapping by alignment and wrapping

AX-condition   XC-condition

a.   A   [B   C   ]$_X$ [A   B   ]$_X$ C

b. [   ]   [   ]

c. [   ]   [   ]
    [   ] [   ] [   ]   [   ] [   ] [   ]

In both structures the single overarching i-phrase would satisfy the wrap-requirement for all lower clauses in both conditions. In particular, the constituent X = [BC] is correctly wrapped by the large i-phrase in the structures on the left, and the constituent X = [AB] is correctly wrapped by the large i-phrase in the structures on the right. Wrapping, like alignment, does not provide an incentive to also map these intermediate constituents to the prosodic structure. However, these are the constituents that crucially distinguish the two conditions.

An account that correctly distinguishes the two structures is the match theory of Selkirk (2011). This theory postulates that syntactic words are mapped to identical prosodic words, syntactic XPs are mapped to identical phonological phrases, and syntactic clauses are mapped to identical intonation phrases. In addition, Selkirk suggests replacing Downing's (1970) notion of root sentence with that of the illocutionary clause, following up on remarks in Potts (2005) about the connection to speech acts. Her theory includes an additional constraint that maps illocutionary clauses to identical intonation phrases. In the context of our paper, it is not important whether all clauses or only illocutionary clauses are matched to intonation phrases. For concreteness, we employ the constraint that matches illocutionary clauses to intonation phrases, given in (19).

(19)   *Match illocutionary clause* (Selkirk 2011)

The left and right edges of an illocutionary clause must correspond to the left and right edges of an intonation phrase.

In this account each illocutionary clause (or root sentence) is directly mapped to an intonation phrase. This has the desired effect. It achieves that the constituents X = [BC] in the AX-condition and X = [AB] in the XC-condition are not only syntactic constituents but also matching prosodic constituents, as shown in (20).

(20)   Correct mapping by matching clauses and i-boundaries

AX-condition   XC-condition

a.   A   [B   C $]_X$ [A   B $]_X$ C

b. [   ]   [   ]
         [   ]   [   ]
    [   ] [   ] [   ]   [   ] [   ] [   ]

We therefore think that Ladd's results and ours provide empirical support for the move from Downing's formulations in terms of edges alone to the stronger requirement of the match theory that the two edges that are aligned must be edges of the same prosodic constituent, the one that is matched to the relevant illocutionary clause (or root sentence).

Selkirk's match theory assumes a good deal of recursion in prosodic structure, following in this an early suggestion of Ladd (1986). In other earlier prosodic accounts, recursion was either assumed not to exist (Selkirk 1986, Nespor and Vogel 1986) or was derived as an exception (Selkirk 1995b).

An exception has been the model proposed by Féry (2009, 2011, 2015) and Kentner and Féry (2013) which rests on the fact that prosodic structure is recursive, not only at the level of the p-phrase, but also at the level of the i-phrase. In Féry (2015), it is shown with a match model that embedded relative clauses and complement clauses often avoid recursivity by extraposition, but that recursive i-phrases are nevertheless allowed in the prosody of German. Féry (2009) and Kentner and Féry (2013) show that syntactic and prosodic recursivity mirror each other.

We also mention that the relation between speech acts and intonation phrases that is postulated in Selkirk (2011) is explored in detail in German in Truckenbrodt (to appear), supporting Selkirk's match constraint.

We add that we think that our conclusion holds even in the presence of functional motivation for indicating a boundary. Our experimental set-up was designed to draw attention to the semantic relevance of the constituent X in both conditions. We think, however, that this can only lead to an i-

phrase corresponding to X if the grammar contains an incentive to map X to an i-phrase to begin with. Without such a mapping relation in the grammar, there is no sense in which the phonetic cues for the i-phrase for X = [AB] provide a cue for a corresponding syntactic structure to the listener.

In conclusion, each root sentence (or illocutionary clause, or speech act) is mapped to an intonation phrase. This is a strengthened requirement of the suggestion of Downing (1970) that root sentences have intonation phrase boundaries at their left and right edges. The way in which Downing's suggestion needs to be strengthened supports the match format of relating syntactic and prosodic constituents in Selkirk (2011).

5 Conclusion

In our experiment, we have replicated the core findings of Ladd (1988) in regard to the scaling of clause-initial peaks in nested structures. In addition, we have shown that structural distinctions of the kind investigated by Ladd also play out in the scaling of German clause-final upstep in a way that confirms Ladd's account. Our findings lend support to the following conclusions:

(a) Downstep applies (in the typical case) among hierarchical sister-nodes, allowing for the application of downstep within downstep, as postulated by Ladd (1988, 1990).

(b) These downstep relations are sensibly modeled using the phrasal reference-lines of van den Berg et al. (1992) (see also the extension of this model in Truckenbrodt 2007b) with reference-lines constant for a given higher constituent, and lowered among sister-nodes.

(c) Clause-final upstep in German, where it applies, involves scaling on this phrasal reference-line, as suggested in Truckenbrodt (2002, 2007b).

(d) The data investigated here require an isomorphic mapping from root sentences (or illocutionary clauses, or speech acts) to intonation phrases. The mapping needs to distinguish $A[BC]_X$ from $[AB]_X C$, where A, B, C, and X are root sentences. To achieve this, the account of Downing (1970) that root sentences are bounded by intonation phrase edges needs to be strengthened. The right kind of strengthening cannot be achieved in terms of alignment in interaction with wrapping. It is given by the match theory of Selkirk (2011) and a recursive account of the prosodic structure (Féry 2009 2011, 2015): (illocutionary) clauses are matched to intonation phrases.

References

Bartels, Christine (1999). *The intonation of English statements and questions. A compositional interpretation*. New York and London: Garland.

Beckman, Mary E., and Janet B. Pierrehumbert (1986). Intonational structure in Japanese and English. *Phonology Yearbook* **3**. 255-309.

Berg, Rob van den, Carlos Gussenhoven, and Toni Rietveld (1992). Downstep in Dutch: Implications for a model. In Gerard J. Docherty and D. Robert Ladd (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press. 335-367.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: CWK Gleerup.

Downing, Bruce Theodore (1970). *Syntactic structure and phonological phrasing in English*. PhD dissertation, The University of Texas.

Féry, Caroline (2009) Syntax, information structure, embedded prosodic phrasing and the relational scaling of pitch accents. In Nomi Erteschick-Shir and Lisa Rochman (eds.) *The sound of syntax*. Oxford and New York: Oxford University Press. 271-290.

Féry, Caroline (2011). German sentence accents and embedded prosodic phrases. *Lingua* **121**. 1906-1922.

Féry, Caroline (2015). Extraposition and prosodic monsters in German. In Lyn Frazier and Ted Gibson (eds.) *Explicit and implicit prosody in sentence processing*. Amsterdam and London: Benjamins.

Féry, Caroline, and Frank Kügler (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* **36**. 680-703.

Féry, Caroline, and Hubert Truckenbrodt (2005) Sisterhood and tonal scaling. *Studia Linguistica. Special Issue "Boundaries in intonational phonology"*. **59.2/3**. 223-243.

Fougeron, Cecile, and Patricia A. Keating (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* **101**. 3728-3740.

Grabe, Esther (1998). *Comparative intonational phonology: English and German*. PhD dissertation, Universiteit Nijmegen.

Grice, Martine, and Stefan Baumann (2002). Deutsche Intonation und GToBI. *Linguistische Berichte* **191**. 267-298.

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*. Oxford and New York: Oxford University Press. 55-83.

Grice, Martine, Stefan Baumann, and Nils Jagdfeld. (2009). Tonal association and derived nuclear accents. The case of downstepping contours in German. In Sabine Zerbian, Laura Downing, and Frank Kügler (eds.) *Tone and intonation in a typological perspective*. *Lingua* **119**. 881-905.

Gussenhoven, Carlos (1983). Focus, mode and the nucleus. *Journal of Linguistics* **19**. 377-417.

Gussenhoven, Carlos (1992). Sentence accents and argument structure. In Iggy M. Roca (ed.) *Thematic structure. Its role in grammar*. Berlin, New York: Foris. 79-106.

Gussenhoven, Carlos (2000). The lexical tone contrast of Roermond Dutch in Optimality Theory. In Merle Horne (ed.), *Prosody: Theory and experiment*. Dordrecht: Kluwer. 129-167.

Gussenhoven, Carlos (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Inkelas, Sharon (1989). Prosodic constituency in the lexicon. dissertation, Stanford University.

Jun, Sun-Ah (ed.) (2005). *Prosodic typology. The phonology of intonation and phrasing*. Oxford and New York: Oxford University Press.

Jun, Sun-Ah (ed.) (2014). *Prosodic typology II. The phonology of intonation and phrasing*. Oxford and New York: Oxford University Press.

Kabagema-Bilan, Elena, Beatriz López-Jiménez, and Hubert Truckenbrodt (2011). Multiple focus in Mandarin Chinese. *Lingua* **121**. 1890-1905.

Kent, R. D., and R. Netsell (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica* **24**. 23-44.

Kentner, Gerrit, and Caroline Féry (2013). A new approach to prosodic grouping. *The Linguistic Review* **30**. 277-311.

Ladd, D. Robert (1983b). Even, focus, and normal stress. *Journal of Semantics* **2**. 257-270.

Ladd, D. Robert (1986). Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook* **3**. 311-340.

Ladd, D. Robert (1988). Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of America* **84**. 530-544.

Ladd, D. Robert (1990). Metrical representation of pitch register. In John Kingston and Mary Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press. 35-57.

Ladd, D. Robert (1992). Compound prosodic domains. Occasional Papers of the Department of Linguistics, University of Edinburgh, Edinburgh, Scotland.

Ladd, D. Robert (1993). On the theoretical status of 'the baseline' in modeling intonation. *Language and Speech* **36**. 435-451.

Ladd, D. Robert (2008). *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press.

Ladd, D. Robert, and Astrid Schepman (2003). Sagging transitions between high pitch accents in English: Experimental evidence. *Journal of Phonetics* **31**. 81-112.

Laniran, Yetunde O., and G. N. Clements (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics* **31**. 203-250.

Liberman, Mark, and Janet Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In Mark Aronoff and Richard T. Oehrle (eds.) *Language, sound, structure:*

Studies in phonology presented to Morris Halle by his teacher and students. Cambridge: MIT Press. 157-233.

Nespor, Marina, and Irene Vogel (1986). *Prosodic phonology*. Dordrecht: Foris.

Odden, David (1995). Tone: African languages. In John Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, Mass.: Blackwell. 444-475.

Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgrefe, Isabell Wartenburger, and Barbara Höhle (2014). On the production and perception of prosodic cues in bracketed lists. Manuscript, Centre for General Linguistics (ZAS) Berlin, Laboratoire Parole et Langage, CNRS UMR 7309, and University of Potsdam.

Pierrehumbert, Janet B. (1980). *The phonology and phonetics of English intonation*. PhD dissertation, MIT.

Pierrehumbert, Janet B., and Mary E. Beckman (1988). *Japanese tone structure*. Cambridge, Mass.: MIT Press.

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. Pollack (eds.) *Intentions in communication*. Cambridge: MIT Press. 271-311.

Potts, Christopher (2005). *The logic of conventional implicatures*. Oxford and New York: Oxford University Press.

Prieto, Pilar, Chi-Lin Shih, and Holly Nibert (1996). Pitch down trend in Spanish. *Journal of Phonetics* **24**. 445-473.

Selkirk, Elisabeth (1980). Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff and Mary-Louise Kean (eds.) *Juncture: A Collection of Original Papers*. Saratoga, CA: Anma Libri. 107-129.

Selkirk, Elisabeth (1986). On derived domains in sentence phonology. *Phonology Yearbook* **3**. 371-405.

Selkirk, Elisabeth (1995a). Sentence prosody: Intonation, stress, and phrasing. In John Goldsmith (ed.), *The handbook of phonological theory*. Cambridge, Mass: Blackwell. 550-569.

Selkirk, Elisabeth (1995b). The prosodic structure of function words. In Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.) *Papers in Optimality Theory. University of Massachusetts Occasional Papers 18*. Amherst, Mass: GLSA. 439-469.

Selkirk, Elisabeth (2005). Comments on intonational phrasing in English. In Sónia Frota, Marina Vigário, and Maria João Freitas (eds.) *Prosodies. With special reference to Iberian languages*. Berlin, New York: Mouton de Gruyter. 11-58.

Selkirk, Elisabeth (2009). On Clause and Intonational Phrase in Japanese: The Syntactic Grounding of Prosodic Constituent Structure. *Gengo Kenkyu 136*.

Selkirk, Elisabeth (2011). The syntax–phonology interface. In John Goldsmith, Jason Riggle, and Alan Yu (eds.) *The handbook of phonological theory*, 2nd edition. Oxford: Blackwell.

Shih, Chi-Lin (2000). A declination model of Mandarin Chinese. In Antonis Botins (ed.), *Intonation. Analysis, modeling and technology*. Dordrecht: Kluwer.

't Hart, John, René Collier, and Antonie Cohen (1990) *A perceptual study of intonation. An experimental phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Truckenbrodt, Hubert. (1995). Phonological phrases: Their relation to syntax, focus, and prominence. PhD dissertation, Massachusetts Institute of Technology.

Truckenbrodt, Hubert (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* **30**. 219-255.

Truckenbrodt, Hubert (2002). Upstep and embedded register levels. *Phonology* **19**. 77-120.

Truckenbrodt, Hubert (2004). Final lowering in non-final position. *Journal of Phonetics* **32**. 313-348.

Truckenbrodt, Hubert (2005). A short report on intonation phrase boundaries in German. *Linguistische Berichte* **203**. 273-296.

Truckenbrodt, Hubert (2007a). The syntax-phonology interface. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press. 435-456.

Truckenbrodt, Hubert (2007b). Upstep of edge tones and of nuclear accents. In Carlos Gussenhoven and Tomas Riad (eds.) *Tones and tunes. Volume 2: Experimental studies in word and sentence prosody*. Berlin, New York: Mouton de Gruyter. 349-386.

Truckenbrodt, Hubert (2012). Semantics of intonation. In Claudia Maienborn, Klaus von Heusinger and Paul Portner (eds.) *Handbook of semantics. An international handbook of natural language meaning. Volume 3*. Berlin: de Gruyter. 2039-2069.

Truckenbrodt, Hubert (to appear). Intonation phrases and speech acts. In Marlies Kluck, Dennis Ott, and Mark de Vries (eds.) *Parenthesis and ellipsis: Cross-linguistic and theoretical perspectives*. Berlin.

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

Appendix: Details of the method

1. Participants

The five speakers were students at the University of Potsdam. They were monolingual native speakers of German in the sense that the first time they had learned a second language was in school. They were in their twenties and came from the northern half of Germany. S1, S2, and S3 were female, S4 and S5 were male. The speakers were reimbursed for their time.

2. Recordings

The recordings were made in a quiet room on a DAT tape recorder. The experimenter left the participant alone in the room, with the tape recorder running, after brief initial instructions as to beginning and ending the session. The participants went through the experiment in the form of a PowerPoint presentation in a self-paced manner. The instructions familiarized the participants with the procedure and made them practice the procedure with four examples. The procedure for each stimulus consisted of two steps. The question and the answer were presented together on the screen. The participant was asked to take the time to understand the connection in meaning between the question and the answer. Then, upon pressing the 'forward' key of the presentation, the visual display remained unchanged and a pre-recorded voice was played by the computer, reading out the question. The participant then read the stimulus as an answer to that question. The set-up and the instructions included the option of repeating the two steps for a particular stimulus in case of hesitation during rendition, or in case the rendition was not felt to be natural. The instructions further specified that we were interested in a normal, narrative tone of voice, and in a normal rate of speech ('zügige, nicht verweilende Sprechgeschwindigkeit'). Included was also the initial information that there are two parts of the experiment, the options of re-reading the instructions and of going over the practice recordings again, as well as suggestions about taking breaks.

First the stimuli of the AX- and XC-conditions were presented, pseudo-randomized as a group. No fillers were employed in this group. For one thing, the task of concentrating on the connection between question and answer before each rendition required close attention and pausing; it seemed to us that this would itself prevent repetitive renditions from arising. For another, this concentration is demanding; if fillers of the same kind had been added, we would have worried that the subjects would not have been able to keep up their concentration for this task through all the stimuli.

After a suggested break, the stimuli of the no-X condition followed. These stimuli all employed the same prosodic pattern and there was no special cognitive task preceding each rendition. Filler sentences were therefore interspersed to minimize the occurrence of repetitive routines.

3. Measurements

The recordings were analyzed using the acoustic speech analysis software Praat. The recordings were manually divided into labeled sub-strings with the help of spectrograms. This was performed by student research assistants at the University of Potsdam, and checked by the authors. The divisions assigned included accent domain-boundaries (Gussenhoven 1983, Ladd 1983b) as well as beginnings and ends of the accented syllables. Relevant possible accents were H*, L*, L*+H, L+H* H+L* and

H+!H*. These were assigned jointly by the two authors based on a combination of auditory impression and the acoustic course of the F0. It turned out that there were no cases of disagreement.

Acoustically, a postulated L*+H accent in clause-initial position showed an F0 minimum around the onset of the stressed syllable (this F0-minimum was then analyzed as the L* part), and a following rise terminating before the end of the p-phrase (the F0-maximum of the remainder of the p-phrase was then analyzed as the +H part). A postulated L*+H accent in clause-final utterance-medial position showed a similar position of the F0-minimum, a following rise, and a high turning point at which the rise turns into a horizontal (or more shallowly rising) plateau (this further turning point was then analyzed as the +H part of the pitch accent). A final H% edge tone was assigned at the end of such a final plateau on the clause-final verb. Accents perceived as clearly high or falling on the stressed syllable were not included in the category L*+H. This perceptual assessment was done by listening to the clauses. Where the perception was not clear between low or high on the stressed syllable, a L*+H was assigned, since this assignment is not in contradiction to the auditory impression. This occurred only in a few cases. In cases of distortions of the pitch-track due to an obstruent sound, a measurement was made in an area that was judged to be just outside of the distortion caused by the obstruent.

4. Exclusion of values

As explained in the text, the criterion for including a recording in the evaluation of phonetic scaling was the assignment of clause-initial L*+H pitch accents with the L* part just before or early in the stressed syllable in all three clauses, as well as H% at the end of both the first and second clauses. An example pitch-track in which this criterion is satisfied is shown in Figure A1. Here the non-final pitch accents are labeled L+H, and the final one L.
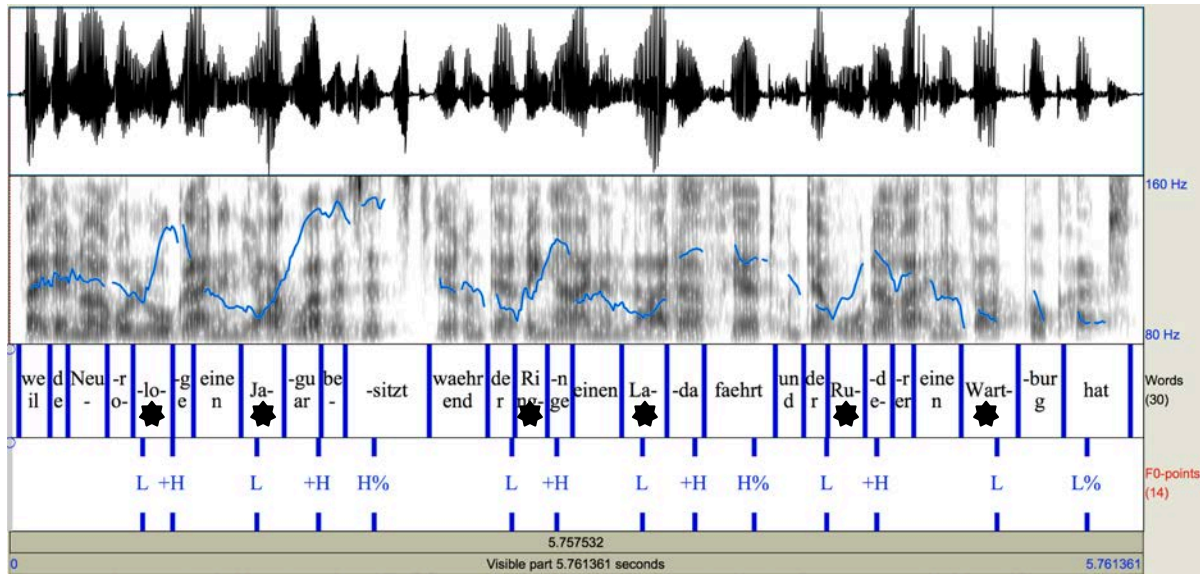
Figure A1. Example pitch-track from the AX-condition. *Weil der Neurologe einen Jaguar besitzt, während der Ringer einen Lada fährt, und der Ruderer einen Wartburg hat.* 'Because the neurologist owns a Jaguar while the wrestler drives a Lada and the rower has a Wartburg.' Speaker S4. The accented syllables are marked by stars.

The criteria for inclusion were applied jointly by the two authors. Column 2 in Table A1 shows that most recordings of the no-X condition met these fairly strict criteria. The third column shows that L*+H was also assigned with great regularity in the nuclear position of clauses A and B (otherwise individual measurements were skipped).

Speaker	# of recordings with three clause-initial L*+H and two clause-final H%	L*+H on nuclear accents of clauses A,B	utterance-final pitch accents
S1	16/18 (twice diff. pitch accent)	36/36	16 H+L* (2 H+!H*)
S2	18/18	36/36	11 L* (7 L*+H)
S3	16/18 (once diff. p.a., once no H%)	35/36	9 H+L* (7 H+!H*, 2 H*)
S4	18/18	34/36	10 H+L* (6 L*, 2 H+!H*)
S5	18/18	36/36	17 L* (1 H*)

Table A1. Frequency of the core characteristics of the no-X condition. The second column shows the number of recordings that met the criteria for inclusion in the phonetic analysis. Reasons for exclusion are given in brackets in this column. The third column shows the assignment of L*+H in the second and third positions of clauses A and B, and in the second position of clause C.

The same criteria for inclusion in the phonetic analysis were used in the AX- and XC-conditions. The frequencies of the tones are shown in Table A2.

Speaker & cond.	# of recordings with three clause-initial L*+H and two clause-final H%	L*+H in position of upstep in A and B	utterance-final pitch accents
S1, AX	14/16 (twice no H%)	32/32	15 H+L* (1 H+!H*)
XC	16/16	32/32	14 H+L* (2 H+!H*)
S2, AX	15/16 (once no H%)	32/32	11 L* (5 L*+H)
XC	6/16 (ten times no H% in B)	32/32	13 L* (3 L*+H)
S3, AX	7/16 (nine times diff. pitch accent)	32/32	13 H+L* (3 H+!H*)
XC	13/16 (once no H%, twice diff p.a.)	32/32	12 H+L* (4 H+!H*)
S4, AX	15/16 (once no H%)	31/32	13 H+L* (3 L*)
XC	15/16 (once no H%)	32/32	11 H+L* (5 L*)
S5, AX	16/16	32/32	16 L*
XC	16/16	32/32	16 L*

Table A2. Frequency of the core characteristics of the AX- and XC-conditions. The numbers of recordings that met the criteria for inclusion in the phonetic analysis are shown in the second column. The third column shows the assignment of L*+H in the second positions of clauses A and B. The fourth column shows the utterance-final pitch accents that were assigned.

As mentioned in the text, there were two cases in which fewer than half of the recordings could be included in the phonetic analysis. Speaker S2 often marked the larger internal boundary of the XC-condition with a L% preceded by an accentual fall instead of a rise followed by H%. Further, speaker S3 sometimes used a different pitch accent in clause-initial positions (in most of these cases H* had been assigned).