Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Detecting Intonation Phrase boundaries in German laboratory speech by means of H tone upstep

Fabian Schubö

University of Stuttgart, Germany

fabian.schuboe@ifla.uni-stuttgart.de

Abstract

This paper proposes a linguistically informed method for automated detection of Intonation Phrase (IP) boundaries in German lab speech. The method makes use of H tone upstep, a phenomenon that applies to the nuclear pitch accent or boundary tone of a non-final IP. The H tone of these events is considerably higher in f0 scaling than the H tone of the immediately preceding pitch accent. This is made use of in order to test for the presence of IP boundaries at certain positions in an utterance: The scaling of the f0 peaks of two adjacent intonational events are extracted and compared in regard to their relative height. If the second value is not lower than the first one, H tone upstep can be assumed, which points to the presence of an upcoming IP boundary. The method is tested on a data set of 216 lab speech utterances and performs with an accuracy of 94%. It is meant to provide a tool for linguists from any background who are working on IP formation in German with data gained in elicited production studies.

Index Terms: Intonation Phrase, prosodic phrasing, H tone upstep, automated boundary detection, lab speech, German

1. Introduction

During speech production speakers organize discourse into chunks of speech, also called prosodic phrases. These phrases usually mirror the syntactic structure of a sentence in that their edges coincide with the edges of syntactic constituents, e.g. [1, p. 60]. Thus, they may help in the resolution of structural ambiguities. In German, it is commonly distinguished between the smaller intermediate phrase (ip) and the larger Intonation Phrase (IP), e.g. [2]. Phonetically, these phrases are instantiated predominantly by insertion of pauses, lengthening of pre-boundary segments, and specific tonal movement towards the end of the phrase. While it remains to be investigated how consistently the former two cues are used, it can be observed in elicited production data that tonal movement serves as a reliable means for the identification of IP boundaries, e.g. [3].

The present paper proposes a linguistically informed method for the detection of IP boundaries in German lab speech utterances, making use of the tonal characteristics preceding the boundary. In particular, the method deterministically decides about the presence/absence of IP boundaries at the edges of syntactic constituents based on the phenomenon of H tone upstep. In German, H tone upstep applies to H tones of nuclear pitch accents or boundary tones at the right edge of an IP [3]. The method assumes the presence of an upcoming IP boundary if a pitch accent comprising an H tone is followed by an element comprising a relatively higher f0 peak. The boundary is assumed to coincide with the right edge of the syntactic constituent that is completed next after the upstepped H tone. The method is tested on a data set of 216 lab speech utterances and performs with an accuracy of 94%.

2. Intonation in German lab speech

The method suggested here makes use of the intonational behavior of German speakers producing isolated sentences in lab speech. This section outlines the findings of prior studies in regard to the distribution of pitch accents and the realization of intonational events and discusses the patterns of relative f0 scaling on elements preceding an IP boundary.

2.1. The distribution of pitch accents

According to the Sentence Accent Assignment Rule (SAAR) [4], every predicate, argument, and modifier that is contained in the focused part of the sentence must be accented, with the exception of a predicate that is adjacent to an argument. Originally proposed for Dutch and English, the SAAR was also found to hold for German [5], [6]. From this it follows that, in German, nouns and verbs show a clear difference in regard to the distribution of pitch accents: Nouns, either as part of an argument or modifier, obligatorily bear a pitch accent whereas verbs need to be in a specific position in order to be accented obligatorily. This difference in pitch accent distribution for nouns and verbs has been found in empirical studies analyzing German lab speech: [5] analyzed 448 broad focus sentences with different structures in regard to their pitch accent distribution and found that sentences with an argument noun and a following verb were realized with a pitch accent on the noun in all cases whereas an additional pitch accent on the verb was rarely present (13%). Furthermore, they found that sentences with a modifier (comprising a noun) between an argument and the final verb were regularly realized with pitch accents on the argument noun (99%) and the modifier noun (91%), but pitch accents on the verb occurred only marginally (15%). Similarly, [7] analyzed 348 broad focus sentences and found that argument nouns were always realized with a pitch accent, whereas verbs were accented in 71% of cases (the authors attribute the high amount of accented verbs to their experimental design). The data of other studies analyzing sentences produced in isolation suggests that nouns (as parts of arguments) are regularly accented whereas verbs adjacent to an argument are not accented, e.g. [3], [8]. In sum, the results of prior studies suggests that nouns in German lab speech typically bear a

10.21437/SpeechProsody.2016-54

pitch accent whereas verbs show a fair amount of variation in this regard, probably depending on the design of the stimuli.

2.2. The realization of intonational events

2.2.1. Pitch accent realization

In German lab speech, broad focus sentences usually comprise rising or high pitch accents in non-final position and a falling or low pitch accent in final position, as can be observed in the data by [3], [8], [9]. [3] analyzed four speakers and found that they produced exclusively L*+H pitch accents in non-final position and H+L* in utterance-final position. This was also the predominant pattern produced by the five speakers analyzed in [8]. [9] presents perception and production data that shows that new information is primarily associated with high pitch accent types (H*) whereas given or accessible information is primarily associated with low or falling pitch accent types (e.g. L*, H+L*).

An IP comprises one or more pitch accents, of which the last one is termed *nuclear*, as it is perceived as stronger than the preceding ones. If there are more than two pitch accents in an IP, successive downstep applies among the H tones of the (pre-nuclear) pitch accents, i.e., the f0 peak of an H tone is lowered relative to the one of the preceding H tone [1], [3], [7] (see Figure 2 below). The nuclear pitch accent of a non-final IP is typically a bitonal rising pitch accent (e.g., L*+H), which may undergo upstep, i.e., the f0 peak of its H tone may reach about the height of the IP-initial H tone and may thus be scaled significantly higher than the H tone of the immediately preceding pitch accent [3] (see Figure 1a below). If a non-final IP contains only two pitch accents, the f0 peak of the H tone comprised by the nuclear pitch accent usually reaches a higher level than the one of the H tone comprised by the preceding pitch accent [3], [10]. In utterance-final IPs, this upstep phenomenon is absent [10].

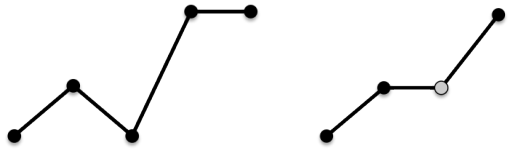
2.2.2. Boundary tone realization

Two types of boundary tones are commonly assumed for marking IPs, a low one (L%) and a high one (H%), e.g. [1], [2]. While the former is typically used to mark the end of a (declarative) utterance, the latter is typically used to mark the right edge of a non-final IP. The high IP boundary tone H% may be realized as part of a boundary tone combination, in which it is preceded by a lower level boundary tone L- or H- [2], [11]. As can be observed in the data by [3] and [11], the tonal scaling properties of an H% marking a non-final IP are as follows: First, the f0 peak of the H% usually targets the topline of the pitch register and may thus be upstepped in relation to the preceding H tone (see Figure 1b below); second, if the preceding pitch accent undergoes nuclear upstep, the f0 peak of the H% has about the same height as the one of the nuclear H tone (see Figure 1a below); and, third, if the H% is part of an L-H% boundary tone combination, the f0 peak of the H% may be lower than the one of the preceding H tone. The data in [3] and [11] suggests that an L-H% combination is not common for non-final IPs in neutral statements (only one of the speakers used for data elicitation realized this pattern). Furthermore, the data analyzed in these works suggests intra-speaker consistency in regard to the realization of boundary tone patterns.

2.3. Relative f0 scaling on the intonational events preceding an IP boundary

Based on the observations on the distribution and phonetic realization of pitch accents and boundary tones presented above, we can identify the two IP-final tonal patterns in Figure 1 as most common in German lab speech. The patterns are similar in that they involve a relative upstep in maximum f0 among the H tones of the IP-final events.

a. Upstep of nuclear H tone b. Upstep of boundary H tone

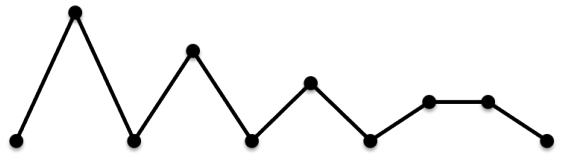


L* + H L* + ^H H% L* + H (H-) ^H%

Figure 1: *Nuclear contour patterns involving H tone upstep on the nuclear pitch accent (a.) and on the boundary tone (b.) (upstepped H tones are marked with a circumflex; nuclear pitch accents are underlined; the optional H target point is colored in grey).*

In the sequence in Figure 1a, this increase applies from the H tone of the immediately pre-nuclear pitch accent to the one of the upstepped nuclear pitch accent (cf. nuclear pattern 1 in [11]). Since upstep applies only to nuclear pitch accents (which are in final position of an IP), the occurrence of an IP boundary would be expected after the upstepped H tone (more specifically, at the right edge of the syntactic constituent on which the H tone is realized). An H% boundary tone may be realized between the nuclear pitch accent and the IP edge. If there is only little material in this location, the two H tones may conflate so that no boundary tone contour is detectable. In the sequence in Figure 1b, the increase applies from the H tone of the nuclear pitch accent to the one of the upstepped boundary tone. The upstepped H tone may be part of a boundary tone combination H-^H%, which comprises a preceding high target between the H tone of the nuclear pitch accent and the one of the H% boundary tone, as illustrated in Figure 1b. This intermediate target is not obligatory though (cf. nuclear patterns 2 and 3 in [11]).

In utterances where no medial IP boundary is present, the H tones of the pitch accents do not involve a relative upstep in f0, but are downstepped in relation to the preceding H tone, as illustrated in Figure 2 (the nuclear H+L* is exempt from downstep, as maximal lowering of H tones applies to the pre-nuclear pitch accent in German [10]). Thus, the common patterns of tonal events preceding an utterance-internal IP boundary and the tonal events within an utterance without an internal IP boundary differ in that the former involve a relative f0 upstep of H tones after a pitch accent at some point whereas the latter involves successive f0 downstep of H tones. This difference can be made use of in order to detect IP boundaries at certain positions in German lab speech utterances.



L* + H L* + !H L* + !H L* + !H H + L*

Figure 2: *Downstep pattern of pre-nuclear H tones (downstepped H tones are marked with an exclamation point; the nuclear pitch accent is underlined)*

3. Automatic detection of IP boundaries by means of H tone upstep

In order to detect IP boundaries on the basis of H tone upstep in German lab speech utterances, two points of maximum f0 are compared: the first point is the f0 value of a non-upstepped H tone and the second point is a later f0 value that is potentially the one of a following upstepped H tone. The first (non-upstepped) H tone is part of a pitch accent, since the upstep of a later H tone peak is relative to the H tone of a preceding pitch accent (as shown in Figure 1). The second point is the f0 maximum of a following element that is potentially followed by an IP boundary (since upstepped H tones of nuclear pitch accents or boundary tones are realized between the prenuclear pitch accent and the IP edge). If the value of the second point is not lower than the value of the first point, successive downstep is interrupted, which points to the presence of an upstepped H tone and thus to the presence of an upcoming IP boundary. Since IP boundaries usually coincide with the edges of syntactic constituents, e.g. [1, p. 60], the location of the boundary can be assumed to coincide with the right edge of the syntactic phrase that is completed next after the occurrence of the upstepped H tone.

In the following, a method for automatic extraction of the f0 values at the two points for comparison is proposed. The method involves the segmentation of recorded utterances into intervals of words or word groups whose right edge coincides with the right edge of a (lexical) syntactic constituent, as this is the relevant position for potential IP boundary insertion. In simple sentences, this usually corresponds to the right edges of the content words (however, if one or more auxiliaries follow the main verb, the relevant position is after the last auxiliary and not after the lexical verb). By segmenting the utterance into word groups containing exactly one content word plus the preceding function words, we arrive at a sequence of intervals of which each interval contains a maximum of one pitch accent and may additionally host one IP-level boundary tone. This is illustrated with a simple sentence in (1) and with a complex sentence in (2) (vertical lines indicate the boundaries of word group intervals). Since in German lab speech nouns generally carry a rising or high pitch accent (see section 2), it can be expected that all intervals containing a noun also contain an H tone. Intervals containing another content word, such as a verb, may however lack a pitch accent and thus not contain an H tone.

(1) | *Der Lehrer* | *hat dem Schüler* | *die Bücher* |*gezeigt*|
the teacher has the student the books shown
'The teacher showed the books to the student.'

(2) | *Cornelius* | *will* *dem Lehrer* | *melden* |
Cornelius wants the teacher report
| *dass Manuel* | *eine Brille* | *gestohlen hat* |
that Manuel a glasses stolen has
'Cornelius wants to report to the teacher that Manuel stole a pair of glasses.'

By comparing the maximum f0 values of two adjacent intervals, an IP boundary after every interval that follows an interval containing a noun can be detected. In (1), this method can detect a boundary after the second and the third interval, as the first two intervals contain nouns and can thus be expected to bear an H tone. In (2), it can detect a boundary after the second, third, and fifth interval, but not after the fourth interval (as the third interval does not contain a noun).

Note that microprosodic distortions in the f0 contour need to be dealt with, as they might augment the maximum f0 in an interval not comprising an upstepped H tone. Thus, the test stimuli for elicitation should be designed in a way that avoids potential microprosodic effects, e.g. by avoiding words containing obstruents. If this is not possible, it might be necessary to smooth the f0 contour or remove distortions manually.

In case an utterance contains more than one IP boundary, the second boundary can only be detected reliably if it occurs at least two intervals later. Otherwise the left hand interval would contain an upstepped H tone and could not be used as a reference. However, since the intervals contain only little material, a second IP boundary is unlikely to occur directly after a first one.

4. Production study

4.1. Background

This section presents a subset of data from a production study that applied the method described above as well as manual annotation of IP boundaries. The data was elicited in order to test for the presence of an IP boundary preceding a subordinate clause in a complex clause configuration in German. The original study [12] aims at testing for an impact of focus and givenness on phrasing decisions in complex clauses, but the data set discussed here is restricted to broad focus sentences comprising only discourse-new material. The hypothesis for this data set was that an IP boundary is regularly inserted preceding the edge of a subordinate clause in German.

4.2. Methods

4.2.1. Stimuli and recording procedure

The stimuli consisted of complex sentences comprising a complement clause embedded in final position, as illustrated in (2) above. The structure of the main clause was SUBJ AUX OBJ V and the structure of the complement clause was SUBJ OBJ V AUX. 18 items of this sort were recorded with twelve female subjects (n=216) between 20 and 29 years old. The target sentences were preceded by a context question eliciting broad focus. The stimuli were presented on a screen, one by one in a pseudo-randomized order and interspersed with filler

sentences. The subjects could familiarize with each question-answer pair before reading the target sentences out loud. Recordings were made in a soundproof booth at the University of Stuttgart (mono, 16 bit, 44.1 kHz).

4.2.2. Detection of IP boundaries

The recorded material was segmented into intervals according to the method described in section 3. The segmentation was performed by means of an automatic alignment tool [13], and further analyzed with the acoustic analysis software *Praat* [14]. The f0 contours of all productions were checked in regard to microprosody and distortions caused by obstruents were manually removed. The maximum f0 values of the last two intervals preceding the internal clause boundary (i.e. the second and the third interval) were extracted automatically and converted into semitones with average speaker minimum f0 as the reference. The converted f0 values were then compared for each target sentence. If the value of the third interval (containing the verb) was not lower than the value of the second interval (containing the object), the utterance was taken as comprising an upstepped H tone within the third interval and thus an IP boundary preceding the embedded clause.

Furthermore, all target utterances were manually annotated in regard to IP boundaries by the author. The decision about the presence of an IP boundary was made on the basis of auditory impression and visual inspection of the f0 contour using *Praat*.

4.3. Results

Manual annotation revealed that 169 out of the 216 utterances comprised an IP boundary preceding the embedded complement clause (78.2%). The results for the automated detection method are presented in Table 1. The first line provides the results for the 169 utterances that comprised an IP boundary: The automated method correctly detected 158 of these IP boundaries (93.5%) and falsely classified eleven utterances as not involving an IP boundary (6.5%). The second line provides the results for the 47 utterances that did not comprise an IP boundary according to manual annotation: The automated method correctly classified 45 of these utterances as not comprising an IP boundary (95.7%) and falsely classified two utterances as comprising an IP boundary (4.3%). In total, 203 out of the 216 utterances were correctly classified in regard to the presence/absence of an IP boundary, which yields an accuracy of 94%.

Table 1. *Results for the automated detection method (first line: results for the 169 utterances comprising an IP boundary; second line: results for the 47 utterances not comprising an IP boundary; third line: results for all 216 utterances)*

	Correct assignments	False assignments
IP present (169)	158 (93.5%)	11 (6.5%)
IP absent (47)	45 (95.7%)	2 (4.3%)
Total (216)	203 (94%)	13 (6%)

5. Summary and discussion

This paper suggested an automated detection method for IP boundaries in German that makes use of the tonal realization of intonational events in lab speech. The method requires the segmentation of utterances into intervals containing one content word and the preceding function words. A boundary can be detected after every interval that has a preceding interval containing a noun by comparing the maximum f0 values of the two adjacent intervals. If the f0 maximum of the second interval is not lower than the one of the first interval, H tone upstep can be assumed, which is followed by an IP boundary at the edge of the syntactic constituent. Using H tone upstep for automated IP boundary detection showed an efficiency of 94% on a data set of 216 lab speech utterances.

The method is somewhat restricted in regard to the locations it can check for the presence of IP boundaries: First, since the left hand interval must contain a noun, there might be constituent edges within a sentence that cannot be dealt with, e.g. if there is a medial verb. The test sentences should thus be designed in a way that controls for part of speech in regard to the relevant locations for IP boundary insertion. This problem can be prevented altogether if pitch accent annotations are available. In this case, an interval can be checked for an IP boundary at its right edge if the preceding interval comprises a pitch accent with an H tone. Another restriction concerns the positions where the IP boundaries occur: The present method can only be applied in order to check the edges of syntactic constituents, but not positions within these constituents. Further research is needed to explore in how far the method can detect IP boundaries in other syntactic positions and other types of sentences.

A major advantage of this method is that it is easy to apply. It is meant to be a tool for linguists of any background who are working on IP formation in German with data gained in the framework of elicited production studies. Further research is needed in order to test the method's capability to detect IP boundaries in naturally occurring speech.

6. Acknowledgements

Many thanks for helpful comments and discussion to Sabine Zerbian and two anonymous reviewers. Many thanks also to the participants of the production experiment.

7. References

[1] C. Féry, *German Intonational Patterns*, Tübingen: Niemeyer, 1993.

[2] M. Grice and S. Baumann, "Deutsche Intonation und GToBI," *Linguistische Berichte,* vol. 191, pp. 267-298, 2002.

[3] H. Truckenbrodt, "Upstep and embedded register levels," *Phonology,* vol. 19, pp. 77-120, 2002.

[4] C. Gussenhoven, "Sentence accents and argument structure," in *Thematic Structure: Its Role in Grammar,* Ignacio M. Roca (ed.), 79-106. Berlin/New York: Foris, 1992.

[5] C. Féry and L. Herbst, "German Sentence Accent Revisited," in *Interdisciplinary Studies in Information Structure,* vol. 1, S. Ishihara, M. Schmitz and A. Schwarz (eds.), pp. 43-75, 2004.

[6] H. Truckenbrodt, "Phrasal Stress," in The Encyclopedia of Languages and Linguistics, vol. 9, K Brown (ed.), pp. 572-579, Amsterdam: Elsevier, 2006.

[7] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *Journal of Phonetics,* vol. 36, 680-703, 2008.

[8] H. Truckenbrodt and C. Féry, "More on hierarchical organization and tonal scaling," Ms. University of Potsdam and University of Tübingen, 2003.

[9] S. Baumann, “Degrees of Givenness and their Prosodic Marking,” in *ZSM Studien*, vol. 2, C. M. Riehl and A. Rothe (eds.), pp. 35-55, Aachen: Shaker, 2008.
[10] H. Truckenbrodt, “Final lowering in non-final position,” *Journal of Phonetics*, vol. 32, pp. 313-348, 2004.
[11] H Truckenbrodt, “Upstep on edge tones and on nuclear accents,” in *Tones and tunes: experimental studies in word and sentence prosody*, C. Gussenhoven and T. Riad (eds.), pp. 349-386, Berlin: Mouton, 2007.
[12] F. Schubö, “Intonation Phrase formation in German sentences with clausal embedding,” University of Stuttgart, in prep.
[13] K. Gorman, J. Howell and M. Wagner, “Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech,” *Canadian Acoustics*, vol. 39, no. 3, pp. 192-193, 2011.
[14] P. Boersma and D. Weenink, “Praat: Doing Phonetics by Computer” [Computer program], version 5.3.32. Retrieved from http://www.praat.org/ [17 October 2012].