

TRANSCRIPTION OF GERMAN INTONATION THE STUTTGART SYSTEM

*Jörg Mayer, University of Stuttgart, Germany
joemayer@ims.uni-stuttgart.de*

CONTENT

Introduction	1
Tone tier	1
Pitch Accents	1
Inventory	1
Standard accents	1
Linking.....	6
Downstep.....	8
Uncertainty	9
Alignment	9
Boundary tones.....	9
Intermediate phrase boundary	9
Intonation phrase boundary	9
Initial IP boundary.....	11
Alignment	12
Break indices	12
ToBI conventions.....	12
Verbmobil conventions	13
Alignment.....	13
Orthographic transcription	13
Miscellaneous	13
Inventory	14
Alignment.....	14
References	14
Data References	14

1 INTRODUCTION

The Stuttgart System is an attempt to integrate the phonological analysis of German intonation done by C. Féry [1] and the ToBI labelling conventions [2], [3]. The system was developed as a tool within our overall aim which is to create a prosodic module for Discourse Representation Theory (DRT) [4]. DRT is a model-theoretic approach to discourse semantics which describes the interpretation of discourses as a dynamic two-level process. Hence, since our inventory of symbols is primarily motivated by phonological analysis and since the domain of processing these symbols is DRT, the criterion for describing fundamental frequency contours which is emphasized in our system is that only those intonational events should be labelled which are distinctive in the sense that one can assign them a function in the domain of discourse interpretation. A further consequence is, that the system reflects phonological pitch-accent linking rules which introduce a set of 'allotonic' accents. Beyond that a small set of default symbols enriches the standard ToBI notation. These default labels are filled in with phonetic content according to autosegmental spreading and alignment conventions.

2 TONE TIER

2.1 Pitch Accents

2.1.1 Inventory

Standard accents	Linking
H*L <i>fall</i>	H* <i>high target on accented syllable</i>
L*H <i>rise</i>	L* <i>low target on accented syllable</i>
HH*L <i>"early peak"</i>	..H <i>high trail tone</i>
L*HL <i>rise-fall/"late peak"</i>	..L <i>low trail tone</i>
H*M <i>stylized contour</i>	

Downstep	Uncertainty
!H*L <i>fall</i>	? <i>Uncertainty: accent type</i>
	*? <i>Uncertainty: accentuation</i>

2.1.2 Standard accents

There are two basic pitch accents H*L and L*H.

H*L a high target on the accented/tonic syllable followed by a falling pitch. If the accented syllable is the last syllable of an intonation unit (intonation phrase/intermediate phrase) the high target and the fall are realized on *one* syllable, namely the accented syllable. If there are syllables following the accented one within the same intonation unit, the high target is reached on the accented syllable followed by the first part of the fall which is continued on the next syllable. After H*L the F0 contour runs in the lower third

of the speaker's range parallel to the baseline until just before the next accented syllable or a phrase boundary.

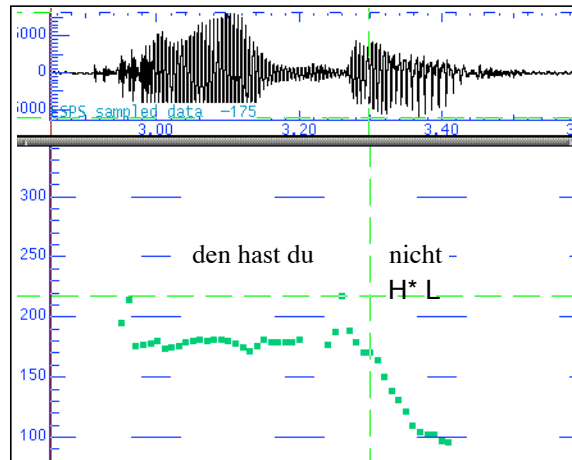


Fig. 1 Example for H*L; utterance *den hast du NICHT*; the vertical line indicates the onset of the accented syllable *nicht* which is the last syllable of the phrase.

L*H a low target on the accented syllable followed by a rising pitch. If the accented syllable is the last syllable of an intonation unit (intonation phrase/intermediate phrase) the low target and the rise are realized on *one* syllable, namely the accented syllable. If there are syllables following the accented one within the same intonation unit, the low target is reached on the accented syllable followed by the first part of the rise which is continued on the next syllable. After L*H the F0 contour runs in the upper third of the speaker's range (mostly parallel to the baseline) until just before the next accented syllable or a phrase boundary.

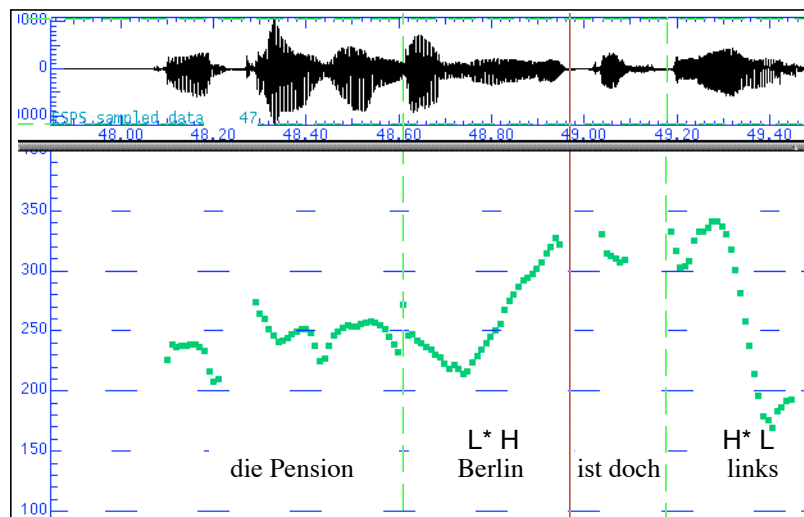


Fig. 2 Example for L*H followed by H*L; utterance *die Pension BERLIN ist doch LINKS*; vertical lines indicate word boundaries.

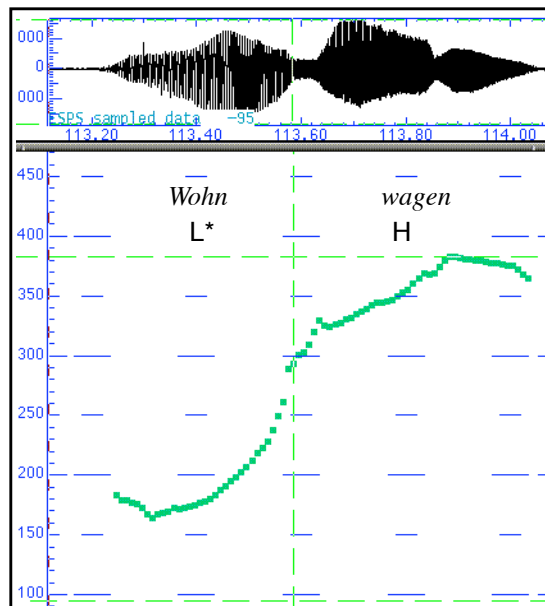


Fig. 3 Example for L*H; utterance *WOHNwagen*; the vertical line indicates the boundary between *Wohn* (the accented syllable) and *wagen*.

Besides the two basic accents there are three additional 'special' pitch accents (cf. [1], ch. 3).

L*HL (rise-fall) a low target on the accented syllable followed by a rise and a fall to a low level. If the accented syllable is followed by at least two additional syllables within the same intonational unit the low target (the starred tone) is realized on the accented syllable. The rise is expected to start on the accented syllable and to be continued on the next syllable. The fall starts on the first syllable after the accented one and ends somewhere in the second syllable. If the accented syllable is followed by only one syllable within the same intonation unit the low target should be on the accented syllable, followed by a rise, which is realized partially on the accented syllable and partially on the postaccentual syllable and a fall on the postaccentual syllable. In the third possible case – the accented syllable is the last syllable in its intonation unit – both the rise and the fall are realized on *one* syllable starting at a low target at the beginning of the accented syllable.

HH*L ("early peak") a high target on the preaccentual syllable followed by a fall on the accented syllable. This accent type must be realized on at least two syllables: an accented syllable and a preaccentual syllable, *which must be weak* (i.e. not stressable). If there is no syllable following the accented one the fall is realized on the accented syllable. If there is a postaccentual syllable the fall is divided into two parts, one part on the accented syllable and the other part on the postaccentual syllable. In any case a high target on the preaccentual syllable must precede the fall. There can be a downstepped target on the accented syllable but in most cases it is a plain fall starting at the height of the preaccentual pitch (or even lower but in the upper third of the speakers range).

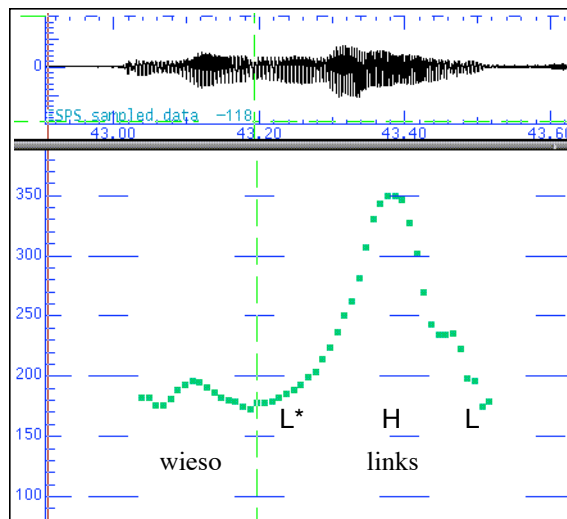


Fig. 4 Example for L*HL; utterance *wieso LINKS*; vertical line indicates word boundary.

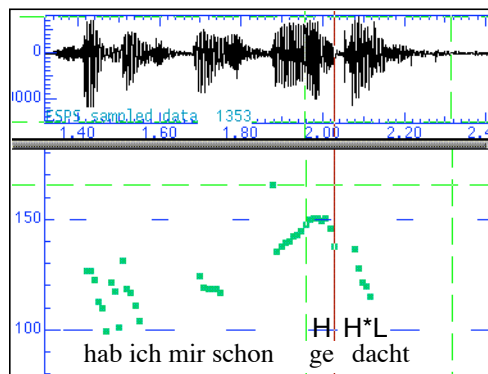


Fig. 5 Example for HH*L; utterance *hab ich mir schon geDACHT*; the first vertical line indicates the beginning of *ge*, the second line the beginning of *dacht*; a high target is reached on the syllable *ge* (prefix), the fall is realized on *dacht*.

The last accent type which is covered by our labelling system is very restricted. It is designated for describing the so called stylized contours which are often used in vocatives (cf. [1], ch. 3.2.1).

H*M ('M' encodes a tone in the middle of the speakers range) a high target on the accented syllable followed by a *high level tone* followed by a fall to the middle of the speakers range on the last syllable of the intonation unit. Words bearing this accent type are normally final in their intonation unit. H*M is the only accent type with a starred tone that can undergo spreading [1]. This means that if there are postaccidental syllables these syllables are associated with the H instead of the trail tone M and only the last postaccidental syllable in the intonation unit is associated with M. This results in the *high level tone* realized on the syllables between the accented syllable and the last syllable. If there is only one syllable to bear H*M the nucleus

of this syllable is often duplicated, so that a high target can be realized on its first part and on its second part a mid target is realized.

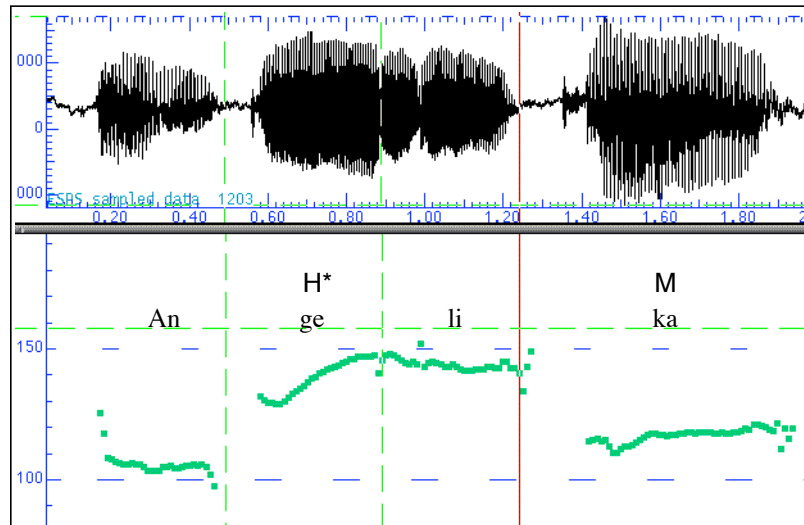


Fig. 6 Example for H*M; utterance *AnGElika!*; vertical lines indicate syllable boundaries; note the high level tone beginning on the accented syllable *ge* and continued on the postaccental syllable *li*; the trail tone M is realized on the last syllable of the intonation phrase *ka*.

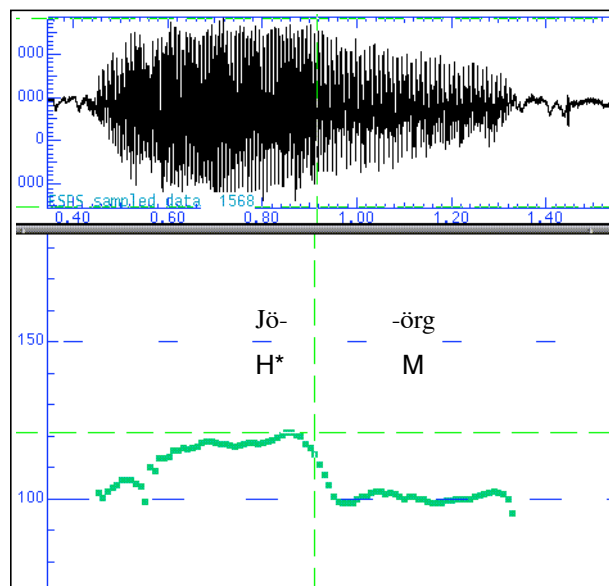


Fig. 7 Example for H*M; utterance *JÖRG!*; the vertical line indicates the boundary between the first half and the second half of the nuclear vowel.

Another (non-tonal) feature that may serve as an identifier of stylized contours is the considerable lengthening of the accented syllable and all postaccental syllables until the boundary of the intonation phrase is reached.

2.1.3 Linking

Linking rules (cf. [1]) permit that *prenuclear* pitch accents can split off their trail tone which is then either associated with the syllable before the next accented syllable (i.e. partial linking) or even completely omitted (i.e. complete linking). The application of these rules is dependent on speaking style, speaking rate, etc., and is claimed to be invariant to discourse interpretation. That is the linked/omitted trail tone will not change the meaning of the intonational contour. The linking processes are thus *allotonic*. The reason for introducing symbols that reflect these rules is to enable the labeler to annotate the application of such a rule so that there is not too much discrepancy between the labels and the actual contour.

H* a high target on the accented syllable that is not followed by an immediate fall. The course of the F0 contour after the high target depends on the type of linking. In case of partial linking the contour should roughly be interpolated between the stated tone (the high target) and the partial linked trail tone L, which is associated with the syllable just before the next accented syllable¹. This results in a smooth fall starting on the accented or the postaccentual syllable and ending on the next preaccentual syllable. In case of complete linking the F0 course depends on the next accent. The contour should look roughly like an interpolation between the stated tone of the linked accent (H) and the stated tone of the next accent (H or L). So if H* is followed by for example H*L F0 between the accented syllables runs in the upper third of the speakers range, if H* is followed by L*H the contour falls between the two accented syllables.

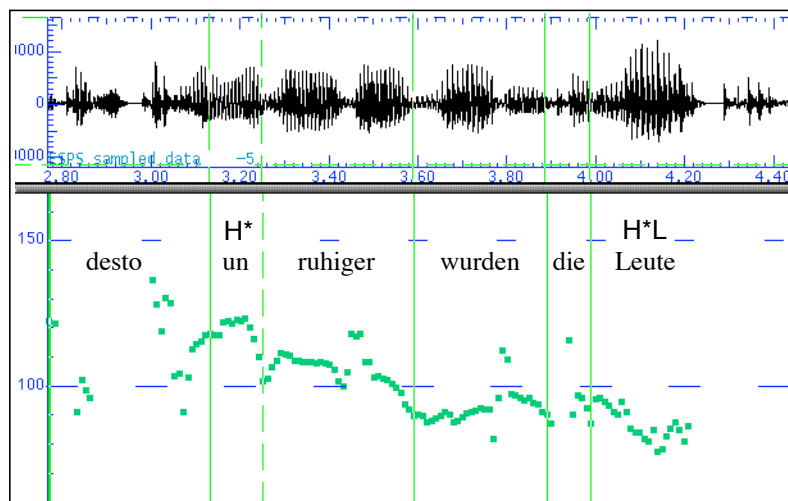


Fig. 8 Example for H* ... H*L (underlying H*L ... H*L); utterance ...*desto UNruhiger wurden die LEUTE*; the contour between the two accented syllables is high but slightly falling due to declination. (There are some instances of strong laryngalization: on *desto* the F0 algorithm fails, on the second syllable of *Leute* no pitch is detected at all.)

..L partial linking: the low trail tone of an underlying H*L accent; to be assigned to the next preaccentual syllable following the linked accent.

¹ Since only pre-nuclear accents can undergo linking there is always at least one pitch accent between the linked accent and the phrase boundary.

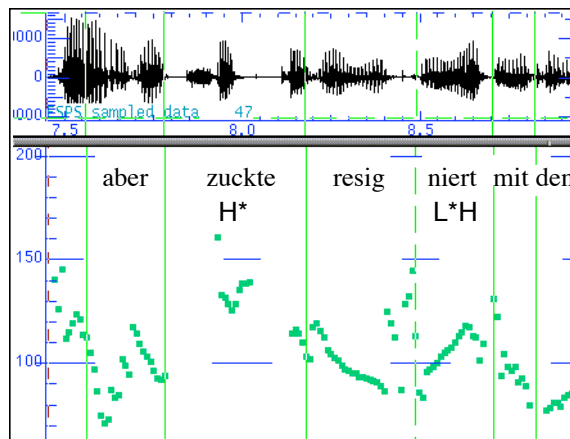


Fig. 9 Example for H* ... L*H (underlying H*L ... L*H); utterance ...*aber ZUCKte resigNIERT mit den...*

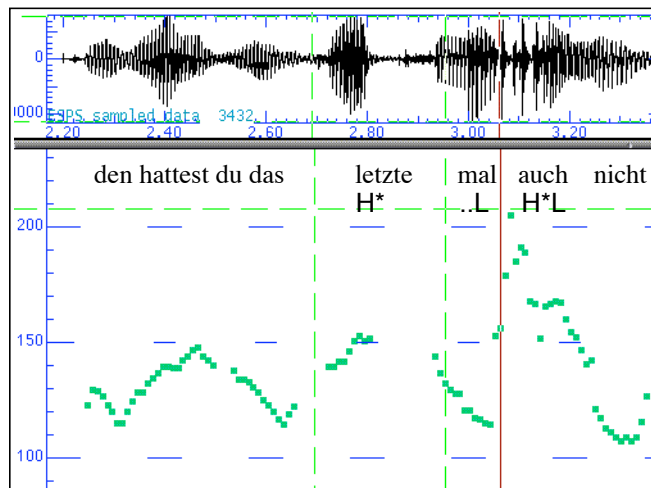


Fig. 10 Example for H* ..LH*L (underlying H*L ... H*L); utterance *den hattest du das LETzte mal AUCH nicht*; note the laryngalization at the beginning of *auch*.

- L* a low target on the accented syllable that is not followed by an immediate rise. The course of the F0 contour after the low target depends on the type of linking. In case of partial linking the contour should roughly be interpolated between the stated tone (the low target) and the partial linked trail tone H, which is associated with the syllable just before the next accented syllable. This results in a smooth rise starting on the accented or the postaccentual syllable and ending on the next preaccentual syllable. In case of complete linking the F0 course depends on the next accent. The contour should look roughly like an interpolation between the stated tone of the linked accent (L) and the stated tone of the next accent (H or L). So if L* is followed by for example L*H F0 between the accented syllables runs in the lower third of the speakers range, if L* is followed by H*L the contour rises between the two accented syllables.

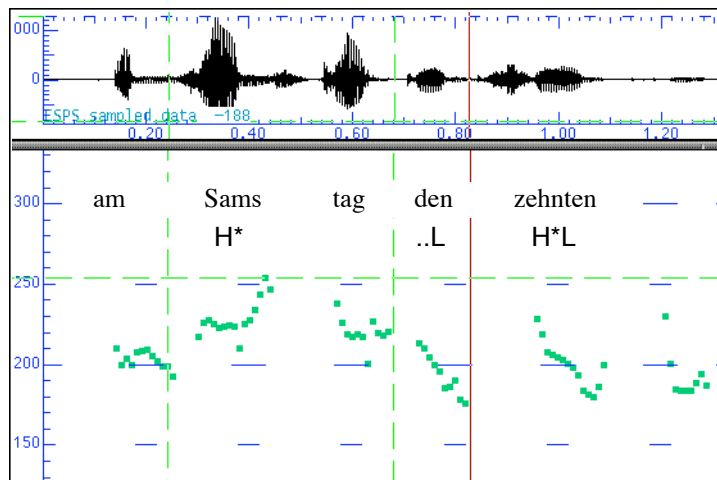


Fig. 11 Example of H* ..LH*L; utterance *am SAMstag den ZEHnten*.

..H partial linking: the high trail tone of an underlying L*H accent; to be assigned to the next preaccental syllable following the linked accent.

2.1.4 Downstep

For downstepped falling accents the system provides the symbol !H*L. (Downstep should be labelled though at present it is not sure whether it differs in meaning from H*L.)

!H*L see H*L; the !H*L accent must be preceded by at least one pitch accent with a starred H tone (H*L, HH*L, H*). The high target of !H*L is still high compared with the surrounding lows but it is lower compared with a preceding high target.

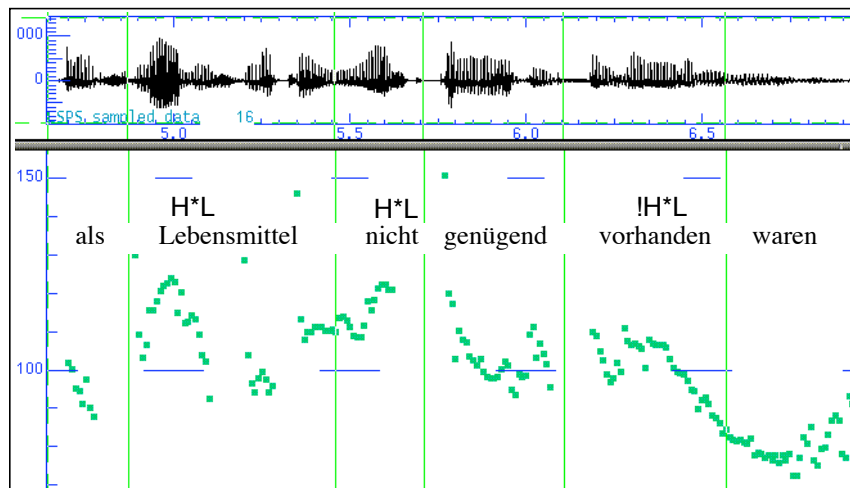


Fig. 12 Example of !H*L; utterance *als LEBensmittel NICHT genügend vor-HANden waren*; three falls in sequence, one not downstepped (on *nicht*) and one downstepped (on *vorhanden*).

2.1.5 Uncertainty

The system provides two symbols for different levels of uncertainty.

- ? is a diacritic that may be applied to each accent type described above (it may also be applied to boundary tones; see below). ? should be added to a standard symbol if the presence of an accent (boundary) is clear but the accent (boundary) type is questionable.²
- *? is used in cases, when the transcriber is uncertain even if there is a pitch accent at all.

2.1.6 Alignment

All pitch accents and the *? symbol are labelled approximately in the middle of the nucleus of the accented syllable (HH*L is also labelled in the nucleus of the accented syllable). Linked trail tones are labelled in the middle of the nucleus of the preaccentual syllable.

2.2 Boundary tones

Prosodic phrasing of utterances is described on two levels, intermediate phrases (ip) and intonation phrases (IP). These levels are ordered hierarchically: an intonation phrase contains at least one intermediate phrase, an ip contains at least one pitch accent.

2.2.1 Intermediate phrase boundary

There is only one default label to transcribe the presence of an intermediate phrase boundary³.

- (hyphen) indicates an intermediate phrase boundary (combined with ToBI break indices 2 or 3 / VM break index B2); ip boundaries at the end of an IP are not indicated, in this case only the IP boundary tone is labelled, the ip boundary is subsumed.

2.2.2 Intonation phrase boundary

For transcribing boundary tones at the end of intonation phrases three labels are provided.

- H% high boundary tone. To transcribe a *clear rise* on the last syllable of an intonation phrase. If the last syllable of an IP is accented both the pitch accent and the boundary tone are realized on the same syllable: if the last syllable is assigned H*L and the IP is assigned H% there should be a fall followed by a rise at the end of the IP. If the last syllable is assigned L*H

² ? substitutes the standard ToBI symbol X*?. In standard ToBI X*? is defined to mark uncertainty about the accent type while the presence of an accent is clear. In our system the difference is that the transcriber is forced to decide what accent type is most probable and then – if necessary – to add a question mark.

³ Specifying ip boundaries with H or L is claimed to be superfluous in the German intonational system [1]. The idea is that the default symbol is filled in with "tonal content" by spreading of the trail tone of the previous pitch accent. So the symbol is redundant since the presence of an ip boundary is also indicated by break indices (see below). However, for some applications (e.g. training of automatic recognition systems) it might be desirable to have all accent and phrasing informations in one label file, namely in the tones label file.

and the IP is assigned H% there is just a rise at the end of the IP. If the nuclear accent (i.e. the last accent) is not phrase final two possibilities appear: if the trail tone of the nuclear accent is L the contour runs in the lower third of the speakers range and rises on the last syllable. If the trail tone is H the contour runs in the upper third and there should be an additional rise on the last syllable.

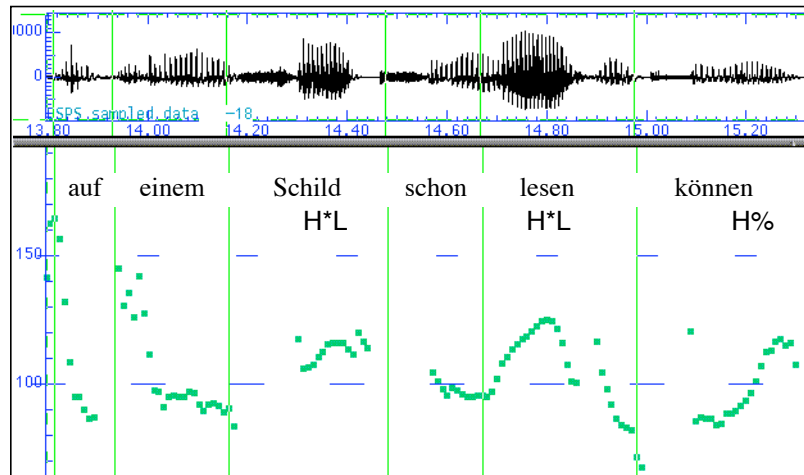


Fig. 13 Example for H*L H%; utterance ...auf einem SCHILD schon LESEN können; rise on the last syllable which is not an accent.

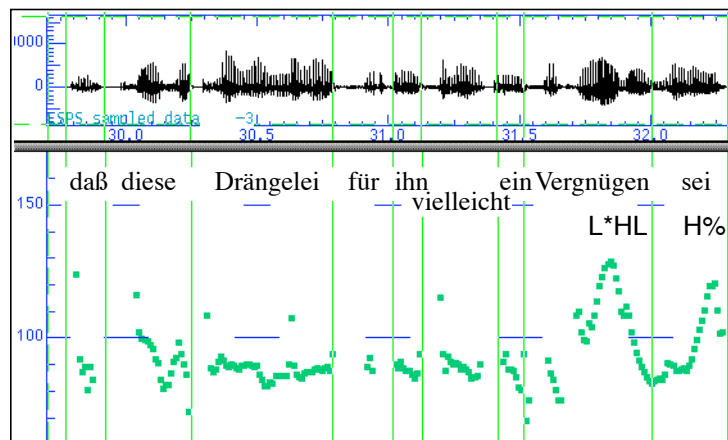


Fig. 14 Example for L*HL H%; utterance ...daß diese Drängelei für ihn vielleicht ein Vergnügen sei.

L% low boundary tone. To transcribe a fall on the last syllable of an intonation phrase. If the last syllable of an IP is accented both the pitch accent and the boundary tone are realized on the same syllable. But since the German intonation system allows only falling nuclear accents previous to a low boundary tone this makes not much difference. If the nuclear accent (only falling accents in front of L%) is not phrase final F0 is low until the end of the phrase with an additional fall on the last syllable.

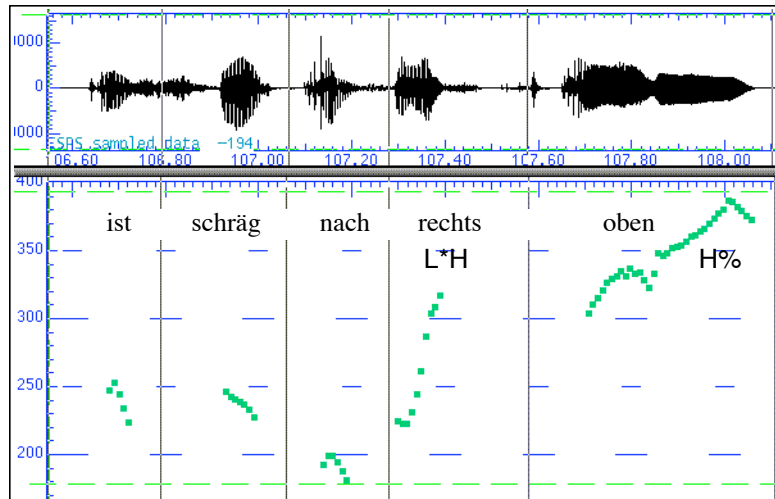


Fig. 15 Example for L*H H%; utterance *...ist schräg nach RECHTS oben*. Note the additional rise on the last (unaccented) syllable.

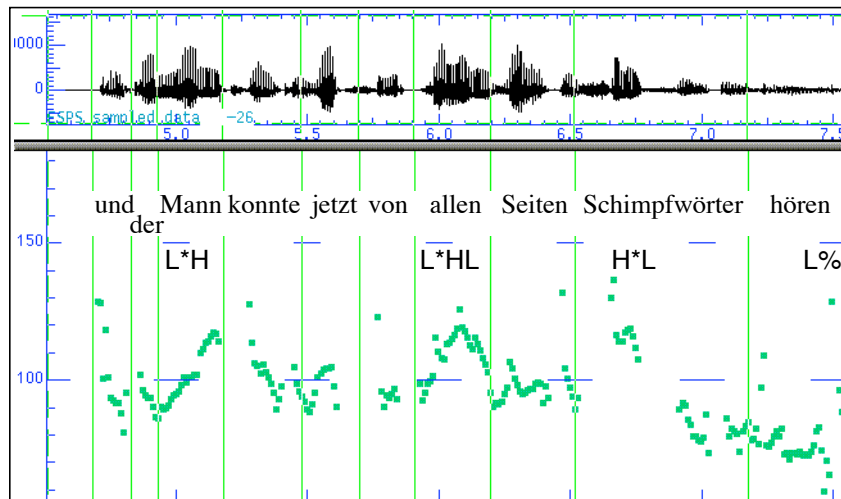


Fig. 16 Example for H*L L%; utterance *...und der MANN konnte jetzt von ALLEN SEITEN SCHIMPFWÖRTER hören*. (jitter at the end is due to laryngalization and breathiness.)

% default boundary tone for transcription of IP boundaries without or with only slight tonal movement. Interpretation by spreading of the nuclear accent's trail tone. IP ends with a high or low plateau, dependent on the previous trail tone, without considerable tonal movement on the last syllable.

2.2.3 Initial IP boundary

By default, pitch at the beginning of intonation phrases is expected to start in a low or middle part of the speakers range which is considered as being not a (initial) boundary tone and therefore is not transcribed. Only if pitch starts decidedly high and only when there is no other plausible explanation for an initial high pitch (e.g. a H accent on the first few syllables) the following label is used:

%H phrase initial high pitch.

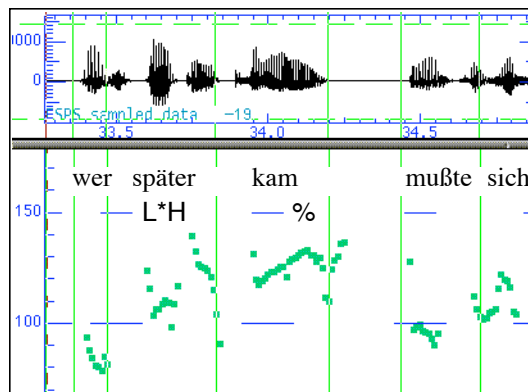


Fig. 17 Example for L*H %; utterance *wer später kam, mußte sich...*

For cases of disfluency there is an additional symbol (which should be used 'conservatively', i.e. not too frequently).

%r for restarted new intonation contours when the last contour was interrupted without being finished due to some disfluency.

2.2.4 Alignment

The phrase final boundary tones are labelled at the very end of the phrase (ip or IP), i.e. at the end of the last syllable of the phrase. (So phrase final the boundary tone label, the transliteration label, and the break index label should have the same alignment point.) The initial boundary tone %H and the %r symbol are labelled at the beginning of the phrase, in front of the first syllable of the IP.

3 BREAK INDICES

On the 'nature' of the break index tier:

"Break indices represent a rating for the degree of juncture perceived between each pair of words and between the final word and the silence at the end of the utterance. They are to be marked after all words that have been transcribed in the orthographic tier. All junctures – including those after fragments and filled pauses – must be assigned an explicit break index value; there is no default juncture type." ([2]: 31)

Two equal valued models can be used to transcribe pauses in the Stuttgart system: ToBI standard [2] and the *Verbmobil* conventions [5]⁴.

3.1 ToBI conventions (cf. [2]: 31-38)

The five basic break indices are the following:

- 0 for word boundaries in clitic groups
- 1 for 'normal' phrase-medial word boundaries

⁴ ToBI is the default model as it is international standard, the *Verbmobil* conventions can be used to improve data exchange with the German *Verbmobil* research group.

- 2 for mismatch between tonal marks and disjuncture marks:
"a strong disjuncture marked by a pause or virtual pause, but with no tonal marks; i.e. a well-formed tune continues across the juncture
OR
a disjuncture that is weaker than expected at what is tonally a clear intermediate of full intonation phrase boundary" ([2]: 35)
- 3 indicates typical boundary strength at intermediate phrase boundaries
- 4 indicates typical boundary strength at intonation phrase boundaries
- The indices 1, 2, and 3 may also be combined with the diacritic p:
- 1p *"an abrupt cutoff before an actual repair, or as if stopping to permit a repair or restart of some kind" ([2]: 36)*
- 2p *"a hesitation pause or prolongation of segmental material where there is no phrase accent [= ip boundary tone; JM] perceived in the intonation contour" ([2]: 36)*
- 3p *"a hesitation pause or a pause-like prolongation where there is a phrase accent in the tone tier" ([2]: 36)*

3.2 Verbmobil conventions (cf. [5]: 3-5)

- (B0 option for word boundaries in clitic groups; not used by now)
- B1 for 'normal' phrase-medial word boundaries
- B2 indicates typical boundary strength at intermediate phrase boundaries
- B3 indicates typical boundary strength at intonation phrase boundaries
- B9 for hesitations, prolongations, cutoffs; comparable with the p diacritic in ToBI.

3.3 Alignment

Break indices are labelled at the end of the word before the break, i.e. they are aligned with the orthographic labels and – if there are any – with the tonal boundary labels.

4 ORTHOGRAPHIC TRANSCRIPTION

In the orthographic tier the words of the utterance are simply transcribed orthographically⁵. The labels should be positioned exactly *at the end* of the actually transcribed word. For different reasons we have added the following symbols:

- <P> for pauses; to be labelled at the end of the pause.
- <slip/disfl> for slips of the tongue and other disfluencies.

5 MISCELLANEOUS

According to the ToBI standard [2] the miscellaneous tier is designated for transcribing different types of disfluencies. In addition, the Stuttgart System provides some labels to mark phonation type/voice quality.

⁵ The Stuttgart System will provide an option for an automatic phonetic transcription and alignment.

5.1 Inventory

silence	(sil) for silent pauses
breath	for audible breathing
laugh	for laughing
cough	for coughing
disfluent	(disfl) other disfluencies
laryngalization	(laryng) for instances of laryngalization
creaky_voice	(crk_voc) for creaky voiced sections
breathy_voice	(brth_voc) for breathy voiced sections
harsh_voice	(harsh_voc) for harsh voiced sections
whispery_voice	(whisp_voc) for whispered sections

5.2 Alignment

Unlike events in the tone tier, the break index tier, or the orthographic tier all events in the miscellaneous tier must be marked for both their ends *and* their beginnings because there is no strict succession of miscellaneous events (normally only few intervals of the actually labelled signal will be marked in the miscellaneous tier whereas in the other tiers the whole signal is successively transcribed). Two diacritics (both suffixes) serve to indicate beginning or end of the transcribed event:

< beginning
> end

For example `silence< ... silence>`

(The miscellaneous labels should be used rather as rough pointers to the transcribed event than as precise demarcations.)

6 REFERENCES

- [1] Caroline Féry (1993), *German intonational pattern*. Tübingen: Niemeyer.
- [2] Mary E. Beckman & Gayle M. Ayers (1994), *Guidelines for ToBI labelling*. Version 2.0, February 1994.
- [3] Julia Hirschberg & Mary E. Beckman (1994), *The ToBI annotation conventions*.
- [4] Hans Kamp & Uwe Reyle (1993), *From discourse to logic*. Dordrecht: Kluwer Academic Publishers.
- [5] Matthias Reyelt & Anton Batliner (1994), *Ein Inventar prosodischer Etiketten für VERBMOBIL*. Verbmobil-Memo-33-94, Juli 1994.
- [6] Martine Grice & Ralf Benz Müller (1994), *Transcription of German using ToBI-tones - the Saarbrücken system*. Ms., University of Saarbrücken.

7 DATA REFERENCES

Data shown in figures 1, 2, 3, 4, 5, 10, 15: Saarbruecken Map Task Corpus

Data shown in figures 8, 9, 12, 13, 14, 16, 17: The Kiel Corpus of Read Speech

Data shown in figure 11: German Verbmobil Corpus