

Prosodic Annotation

Contents: The procedures of prosodic labeling will be described, but not the procedures of segmentation, that may be done in parallel. These can be found in TechDok 8/5e.

Authors: Ulrich Reubold, Sonja Schmidt

Date: 05.01.2006

Version: 1.7

A) Work steps of prosodic labeling

1. Initial prosodic labeling (possibly with parallel initial rough segmentation)

- The prosodic labelers choose a User-Corpus-file from folders containing the raw files (“/data/data45/BITS_DATA/SEGMENTATION/vorbereitet/BLOCK10/SES100[1-4]”). These are TextGrid-files containing descriptive data; their names follow a certain pattern: “US100[1-4]????_0_PR.TextGrid”. The “PR” indicates the files that can be annotated prosodically.
- The command „erst-seg US100[1-4]????“ (without any extensions) opens PRAAT that loads a TextGrid- and the corresponding wav-file. TextGrid- and wav-file are edited together, so that the TextGrid can be processed. Before starting PRAAT, the user is asked: “Found only a PR TextGrid to work on – ok? (y/n)”. This question must be answered by the input “y”.
- The PRAAT editor displays oscillogram and spectrogram as well as a segmentation tier and a tier called „PRB“ (Prosodie Braunschweig), which already contains prosodic labels (GtoBI).
- PRAAT also displays a curve describing F0. A different range of possible F0-values is used for every speaker so that F0-extraction errors are more unlikely. The labelers can estimate whether the values are reliable when they zoom in. Especially phone-transitions and laryngalizations may be critical for the F0-extraction.
- If the segment tier is called „MAU“, its segmentation (and segmental labeling in extended German SAM-PA) is made fully automatically, so that the labeler has to do a manual segmentation first. If the tier is called „SAP“, segmentation is already manually made and, therefore, the labeler can start prosodic labeling.
- The given labels result from a manually done prosodic labeling with data of another speaker, but the same sentences, made at IMS, University of Stuttgart. Within our data, these labels has been automatically moved to the corresponding positions. By using given labels, processing time at IPSK shall be saved, as our labelers just have to check boundary and accent tones, correct their quality, remove or insert some of them.
- When the initial prosodic labeling is finished, the labeler saves the work and shuts PRAAT. The labeler is asked by „erst-seg“, whether the work is really finished (“yes”) or not (“no”) or whether the file is defective (“defect”). If the work is successfully finished, a commentary about specifics of the speech data can be written.
- At this step segmentation and prosodic labeling are checked by a parser that detects impossible segmentation labels or prosodic labels. Concerning prosodic labels it detects three possible sources of faults:

1.: If one word¹ bears a word accent label „+“ AND a phrase accent label, both labels must be on the same segment.

2.: A standard accent must follow a linking accent.

3.: Only the following labels are allowed: H*L, L*H, H*, ..L, L*, ..H, H*L?, L*H?, *?, -, %, H%, -?, %?, H%?.

If the parser detects any faults, the labeler has to restart erst-seg with the formerly used wav- and TextGrid-files.

- If everything is finished, the initially labeled files are moved to a folder that contains all files for the second step of segmentation (./data/data45/BITS_DATA/SEGMENTIERUNG/erst_fertig/BLOCK10/SES100[1-4]/“).

2. Second segmentation

- In this step, segmenters check the segmentation by using the script „zweit-seg“, if the files is roughly segmented data (“MAU”). “SAP”-files must be checked if they are not signed as being corrected (this information is given in the commentary files at “/data/data45/BITS_DATA/RECS/BLOCK10/SES100[1-4]/US100[1-4]?????.txt”). If this is the case, they may be moved immediately, if not, they must be corrected, as well as the “MAU”-files. All files are finally moved to ./data/data45/BITS_DATA/SEGMENTIERUNG/zweit_fertig/BLOCK10/SES100[1-4]“. Again, a commentary may be appended to the commentary file.

1. Second prosodic labeling

- The second prosodic labelers use the script „abgeschlossen-seg“ to check and eventually change exclusively the prosodic labeling. The labelers can use labels of uncertainty (to be described later).
- Then the files are moved to a folder containing all finished files (./data/data45/BITS_DATA/SEGMENTIERUNG/abgeschlossen/BLOCK10/SES100[1-4]“). Again, a commentary may be written.
- The second labelers give feedback to the initial labelers, as appropriate.

1 A „word“ is here a group of segments, grouped by brackets, e.g. „(hOYt@)“.

B) Fixture of tones and their placement

A division must be made between the terms „boundary“ and „boundary tones“ and a definition of the term „phrase accent“ must be given. Boundaries as well as phrase accent in the broader sense are produced by several prosodic means. But the prosodic labelers just rate the intonatory features, so they label *boundary tones* and also just tonal features of *phrase accents*. So, they listen to the data and get an impression of where boundaries are, and therefore, where the phrases are. In these phrases they seek the positions of phrase accents, i.e. the most prominent syllables in the most prominent words. The labels are placed at these positions, as well as at the positions of the boundaries; the labels are affected by the tonal features at these positions. This does not necessarily mean that the tonal features are the main features that influence the labeler's decision.

See *Mayer/Schmidt* resp. *Mayer* for the rest of the document:

1. Set of tones

- 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 can contain intermediate phrases, an ip contains at least one pitch accent.

Fixture:	-	Intermediate phrase boundary (ip)
	%	General phrase boundary or intonation phrase boundary (IP)
	H%	High boundary tone IP
	x?	Uncertainty about the presence/quality of a boundary:
	-?:	presence of “-”?
	%?:	“%” or “-”?
	H%?:	“H%” or “%”?

- (Phrase) Accents

Fixture:	Standard:	H*L	fall
		L*H	rise
	Linking accents:	H*	high target on accented syllable
		..L	low trail tone
		L*	low target on accented syllable
		..H	high trail tone
	Additional labels of uncertainty:	x?	Uncertainty about accent type
		*?	uncertainty about accentuation

See B.)3.) for an explanation of “target” and “trail tone”.

The labels of uncertainty are exclusively used by the second prosodic labelers.

2. Placing of the boundary tones

- The labeler listens to the utterance and decides where s/he perceives boundaries. The labels for these boundaries are set at the last segmental boundary of the last word of the phrase.
- When the position of boundaries is clear, a decision must be made as to what kind of label has to be set. If there is a clear perception of a pause, the boundary tone is an intonation phrase boundary tone. It is more important that the labeler has the perception of a pause, less important whether there is a pause in the physical signal or not. Depending on the last syllable's pitch contour, its label is % (falling or progressive contour) or H% (rising contour).
- The intonation phrases may be perceived as being structured. The boundaries of these structures are the so called intermediate phrase boundaries. There is only one possible label: -.
- Within very long words (e.g.: “Neunzehnhundertdreiundneunzig” (= 1993)), compounds (e.g.: “Innenstaatssekretär” (= Under-Secretary of State of the Interior)) or double-barreled names (like “Randzio-Plath”) only very clear boundaries may be labeled. Intermediate phrase boundaries are not allowed here.

3. Placing of phrase accents

- In the phrases, the labeler tries now to recognize the most prominent words. The labels are set at the nucleus of the word's most prominent syllable, i.e. a vowel in most cases. The exact setting is at the segment's midpoint. Although the labels should be set because of the word's prominence within the whole utterance, its kind is dependent on the intonation phrases and intermediate phrases (=intonation units), as it may depend on the position of the boundaries (see below). A rough rule of thumb is to set at least (!) two accents between two boundaries, i.e. within one phrase, whenever it seems to be possible.
- Every word is a possible candidate for accentuation. There are two types of words within our lexicon, those with lexical accent (or word accent), represented by a + on the most prominent syllable's nucleus, and those without such a lexical accent (mostly function words). The phrase accent may be on a different syllable from the one given in the lexicon. So, also the realized accentuation of the word changes. Therefore, it may be necessary to change a +'s position to the actually accentuated syllable.
For the data of the speakers US1002 and US1004, the rule applied, that one word could be the carrier of only one phrase accent. We changed this rule for the data of the speakers US1001 and US1003, as we wanted to deliver as many auditively perceived accents as possible: very long words (e.g.:

“Neunzehnhundertdreiundneunzig” (= 1993)) or compounds (e.g.: “Innenstaatssekretär”) or double-barreled names (e.g.: “Randzio-Plath”) may carry two accents, if the secondary accent is especially prominent. The main accent is the one on the syllable, that is additionally given the “+”-symbol.

- The “carriers” of the accents are the most prominent syllables. The corresponding labels are positioned at the midpoint of the syllable's nucleus (usually a vowel).
- The labeler's decision is influenced by perception, but also by the F0 contour extracted by PRAAT. If the prominent syllable is the last syllable, followed by a boundary, only this syllable's F0 contour is decisive. In other cases, the contour on the following syllables is also decisive, so the decision in the latter case is more relative. In fact, boundaries may shorten the decisive domain of observation. But both cases have one thing in common. An accent e.g. of the form H*L consists of the “target” H* and the “trail tone” L. The target is the position of the extremum of the F0 contour, while the trail tone is the following context on the way to the target's tonal opposition. It has been explained earlier that it is of no importance whether the trail tone can be found on the same syllable where the target can be found, or also on following syllables, including the syllable being the antecessor of the following accentuated syllable.

4. The phrase accents in detail

- The labelers judge the F0 contour with the help of a construction called *top* and *bottom line*. These two lines are thought of as the envelope of the possible F0 values during a speaker's utterance. The top line starts at the extreme high end of the speaker's pitch range and falls in a more extreme way than the bottom line does, which starts approximately shortly below the midpoint of the range and ends in the lowest regions. So, during an utterance F0 falls generally speaking, and the range of possibilities to vary the F0 contour gets smaller and smaller, as the lines do not fall in parallel. So, a H*L at the beginning of an utterance may look quite different from a H*L near the end of an utterance.
*H*L/L*H* or *H*(..L)/L*(..H)* generally means a high/ low target on the accented syllable followed by a falling/ rising pitch. As top and bottom line are generally falling, a progradient pitch must be interpreted as a rising one.
- *Standard accent1:*
If the accented syllable is the last syllable of an intonation unit (intonation phrase/ intermediate phrase), the high/ low target and the fall/ rise are realized on one syllable, namely the accented syllable, i.e. Target and trail tone are on one syllable. The standard accent1 is the only possible accent.
- *Standard accent2:*
If there are syllables following the accented one within the same intonation unit, the high/ low target is reached on the accented syllable followed by the first part of the fall/ rise, which is continued on the next syllable. If there is not a following standard accent, the only possibility is to make use of a standard accent. If another standard accent does follow, the labeler has to decide, whether a standard accent2

or a linking accent is the best way to describe the accented syllable.

- There may be *cases of doubt*, whether H^*L or L^*H may be the right label, if F0's extremum (maximum respectively minimum) can be found exactly on the segmental boundary between the accented syllable and the following (trail tone carrying) syllable.

These cases are treated in two different ways, as it turned out, that the first way, which was used for the speakers US1002 and US1004, could not satisfy our pretensions to deliver a consistent labeling, as to many F0-contours seemed to contradict this guideline. So, the guideline was kept for the rest of the data of US1002 and US1004, but changed for the still unprocessed data of the speakers US1001 and US1003. For US1002 and US1004 we used the guideline that the extremum is part of the trail tone, i.e. in the case of a maximum, the label was L^*H . For the speakers US1001 and US1003 we use the guideline that the extremum is part of the target. So, a maximum on the boundary leads to the label H^*L .

- *Details of linking accents $H^*(..L)/L^*(..H)$:*

Linking rules 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 ($H^* ..L/ L^* ..H$)) or even completely omitted (i.e. complete linking (H^*/ L^*)).

H^* is a high target on the accented syllable that is not followed by an immediate fall and L^* is a low target on the accented syllable not followed by an immediate rise. The course of the F0 contour after the target depends on the type of linking. Its tendency is to fall/ rise constantly over at least two unaccented syllables. As standardaccent2's trail tone may also fall/ rise over two following unaccented syllables, standardaccent2 and linking accents may be hard to distinguish: the only distinctive feature is the slope of the F0 contour which is more abrupt in cases of standardaccent2.

- In case of *complete linking* the F0 course depends on the next accent. If H^* is followed by H^*L , F0 runs between the accented syllables in the upper third of the speaker's range. If H^* is followed by L^*H , the contour falls between the two accented syllables.

These rules apply accordingly to L^* .

- In case of *partial linking* the contour should roughly be interpolated between the starred tone (the high target) and the partial linked trail tone $..L$, which is associated with the syllable just before the next accented syllable. This results in a smooth fall starting on the accented syllable and ending on the next preaccental syllable.

These rules apply accordingly to $L^* ..H$.

Lit.:

Mayer/Schmidt: J.Mayer, accretions by S.Schmidt: „Transcription of German Intonation. The Munich System.“, in-house document, Munich, 2005.

Mayer: J.Mayer: “Transcription of German Intonation. The Stuttgart System.” IMS. Stuttgart, 1995.