

**The Kiel Corpus of Read Speech - Vol. I Rev. 1**  
**German: PHONDAT90 and PHONDAT92 (Kiel-CD #5)**  
*Editors Klaus J. Kohler and Michel Scheffers*

## **1 Introduction**

This CD-ROM contains a revised issue of the previously published Vol. I of the *Kiel Corpus of Read Speech* (PHONDAT90 and PHONDAT92: Kiel-CD #1, 1994), providing the audio and the time-aligned segmental label files, as well as label files in which prosodic labels have been added to the segmental ones. There are several reasons for this revision: First, at the time CD #1 was published there was no established audio file format. We could therefore assume that users of the CD would have access to tools to convert raw data to their preferred format. Nowadays the situation is almost reversed which is why we choose the RIFF-WAV format on this CD. Second, it turned out that over 50 recordings contained loud clicks due to numerical overflow in the post-processing. Most of these could be corrected automatically but 4 recordings had to be copied from DAT and their label files realigned. Third, apart from adding files with prosodic labels, years of working with the segmental label files had led to numerous corrections, including some changes in the labelling conventions (see Section 4.2).

The speech material consists of isolated sentences and two texts. For the recording, the speaker was seated in a sound-treated booth and read the sentences through a window from a monitor that was placed in the adjacent room, where the operator controlled the orthographic presentation of the sentences one by one. The texts were presented in two or three parts, respectively, on boards positioned inside the booth. A Neumann U87 condenser microphone (cardioid settings) was placed approximately 30 cm from the speaker's mouth. The microphone signals were amplified by a John Hardy M1 pre-amplifier and recorded on a SONY PCM 2500 DAT-recorder at a sampling rate of 44.1 kHz for PHONDAT90 and of 48 kHz for PHONDAT92, respectively, with 16 bit quantization. They were then digitally transferred to computer hard disk and downsampled to 16 kHz as well as high-pass filtered at 40 Hz.

Details concerning the make-up of the material, the recording and processing of the speech signals, as well as the old segmentation and labelling conventions applied to the data on Kiel-CD #1, can be found in Kohler (1992). The labelling conventions were subsequently further developed to be applicable also to spontaneous speech, cf. Kohler, Pätzold, Simpson (1995). Prosodic labelling (PROLAB) is based on the Kiel Intonation Model (KIM) and is described in Peters and Kohler (2004). Updates of both segmental and prosodic labelling are summarized in section 4.2.2 below. On the basis of canonical transcriptions, rule-generated from the orthographic text by Klaus J. Kohler, the segmental and prosodic labelling was carried out in *xassp* by phonetically trained students under the supervision of Klaus J. Kohler, Wim van Dommelen, Adrian Simpson, and Benno Peters. Cf. IPDS (1997); Kohler, Pätzold, and Simpson (1997).

Michel Scheffers serviced the speech processing software, integrated the data in the data bank and helped with updating Vol. I of the *Kiel Corpus of Read Speech* on this CD-ROM.

## 2 Directory structure

The CD-ROM has the following directory structure:

```
<CD-ROM>
+---ph90
| +---berlin
| | +---k01
| | | +---k01be001.s[12]h
| | | +---k01be001.wav
| | | .
| | | .
| | | +---k01be100.s[12]h
| | | +---k01be100.wav
| | +---k02
| | +---k03
| | +---k04
| | +---k05
| | +---k06
| | +---k61
| | +---k62
| | +---k63
| | +---k64
| | +---k65
| | +---k66
| +---marburg
| | +---k07
| | | +---k01mr001.s[12]h
| | | +---k01mr001.wav
| | | .
| | | .
| | | +---k01mr100.s[12]h
| | | +---k01mr100.wav
| | +---k08
| | +---k09
| | +---k10
| | +---k11
| | +---k12
| | +---k61
| | +---k62
| | +---k67
| | +---k68
| | +---k69
| | +---k70
| +---butter
| | +---k22
| | | +---k22butt1.s[12]h
| | | +---k22butt1.wav
| | | +---k22butt2.s[12]h
| | | +---k22butt2.wav
| | | +---k22butt3.s[12]h
```

| | | +---k22butt3.wav  
| | +---k23  
| | +---k24  
| | +---k25  
| | +---k26  
| | +---k27  
| | +---k28  
| | +---k29  
| | +---k30  
| | +---k61  
| | +---k62  
| | +---k76  
| | +---k77  
| | +---k78  
| | +---k79  
| | +---k80  
| +---nordwind  
| | +---k13  
| | | +---k13nord1.s[12]h  
| | | +---k13nord2.s[12]h  
| | | +---k13nord1.wav  
| | | +---k13nord2.wav  
| | +---k14  
| | +---k15  
| | +---k16  
| | +---k17  
| | +---k18  
| | +---k19  
| | +---k20  
| | +---k21  
| | +---k61  
| | +---k62  
| | +---k71  
| | +---k72  
| | +---k73  
| | +---k74  
| | +---k75  
| +---restkorp  
| | +---k61  
| | | +---k61cn001.s[12]h  
| | | +---k61cn001.wav  
.  
.  
| | | +---k61cn020.s[12]h  
| | | +---k61cn020.wav  
| | | +---k61ko001.s[12]h  
| | | +---k61ko001.wav  
.  
.  
| | | +---k61ko063.s[12]h  
| | | +---k61ko063.wav

```

| | | +---k61s1001.s[12]h
| | | +---k61s1001.wav
.
.
| | | +---k61s1045.s[12]h
| | | +---k61s1045.wav
| | | +---k61s2001.s[12]h
| | | +---k61s2001.wav
.
.
| | | +---k61s2025.s[12]h
| | | +---k61s2025.wav
| | | +---k61sr001.s[12]h
| | | +---k61sr001.wav
.
.
| | | +---k61sr015.s[12]h
| | | +---k61sr015.wav
| | | +---k61tk001.s[12]h
| | | +---k61tk001.wav
.
.
| | | +---k61tk030.s[12]h
| | | +---k61tk030.wav
| | +---k62
| +---addcorp not labelled, *.wav and *.s0 with orthographic text and canonical transcription
| | +---k13
...
| | +---k30
| | +---k71
...
| | +---k80
| +---kdb directories containing data banks for the labelled files of each speaker in PHONDAT90
| | +---k01.kdb
...
| | +---k80.kdb
+---ph92
| +---erlangen
| | +---dlm
| | | +---dlme001.s[12]h
| | | +---dlme001.wav
.
.
| | | +---dlme100.s[12]h
| | | +---dlme100.wav
| | +---hpt
| | +---kko
| | +---rtd
| | +---uga
| +---siemens
| | +---dlm

```

```

| | | +---dlms001.s[12]h
| | | +---dlms001.wav
.
.
| | | +---dlms100.s[12]h
| | | +---dlms100.wav
| | +---hpt
| | +---kko
| | +---rtd
| | +---uga
| +---kdb    directories containing data banks for each speaker in PHONDAT92

```

**Key**

```

ph90      PHONDAT90
ph92      PHONDAT92
berlin    100 Berlin sentences
marburg   100 Marburg sentences
butter    The Butter Story in 3 parts
nordwind  The North Wind and the Sun in 2 parts
restkorp  20 CNET sentences
           63 Kohler sentences
           45 SEL sentences
           25 additional SEL sentences
           15 Schiefer/Sommer sentences
           30 Tillmann/Kohler sentences
erlangen  100 Erlangen sentences
siemens   100 Siemens sentences

```

Each of the terminal nodes in the tree is a directory containing the signal and segmentation files of one speaker who was recorded speaking that particular part of the corpus. Two speakers (k61=kko and k62=rtd) produced the whole corpus.

**3 File naming conventions**

```

k05be005.s1h
^^^      speaker (PHONDAT90)
dlms001.s1h
^^^      speaker (PHONDAT92)

k05be005.s1h
^^       sentence/text type:
        be    Berlin sentences
        mr    Marburg sentences
        butt  The Butter Story
        nord  The North Wind and the Sun
        cn    CNET sentences,
        ko    Kohler sentences
        s1    SEL sentences
        s2    additional SEL sentences
        sr    Schiefer/Sommer sentences
        tk    Tillmann/Kohler sentences

dlms001.s1h
^        e    Erlangen sentences
        s    Siemens sentences

```

k05be005.s1h  
 ^^^ sentence number or text part ('The Butter Story' is in 3 parts,  
 e.g. k24butt2.s1h, 'The North Wind and the Sun' in 2 parts,  
 e.g. k74nord1.wav)

k05be005.s1h  
 ^^^ segmental label file; labels in SAMPA; plosive release (&  
 aspiration) included

k05be005.s2h  
 ^^^ segmental and prosodic label file; labels in SAMPA; plosive  
 release (& aspiration) included

k05be005.s0  
 ^^ not phonetically labelled, only orthographic text and canonical  
 transcription

k05be005.wav  
 ^^^ signal file; 16-bit/16kHz

## 4 File formats

### 4.1 Signal files

Signal files (\*.wav): 16 bit, 16 kHz, Microsoft RIFF-WAV format

### 4.2 Label files

7-bit ASCII files, DOS format (CR/LF at line end). Each label file \*.s1h or \*.s2h contains three sections: orthographic, canonical and phonetically labelled. *restcorp* has only been phonetically labelled for speakers k61 and k62. Other speakers read varying sections of *restcorp*, for which only \*.wav and \*.s0 with orthographic text and canonical transcription are provided in the sub-directory *addcorp*.

#### 4.2.1 Conventions in orthographic section

The orthographic text follows DUDEN before the spelling reform. Word-initial small letters are kept sentence-initially. T<sub>E</sub>X conventions are used to represent umlaut and "ß": "a, "A = ä, Ä; "o, "O = ö, Ö; "u, "U = ü, Ü; "s = ß. Because of the T<sub>E</sub>X notation of umlaut no double inverted commas are allowed for quotations.

#### 4.2.2 Conventions in label section

The segmental transcription is modified SAMPA, the prosodic annotation follows PROLAB: cf. Kohler et al. (1995) for segmental labelling; Peters and Kohler (2004) for PROLAB labelling. Both segmental and prosodic labelling conventions for the *Kiel Corpus of Read Speech* Vol. I are updated and summarized in this section. From orthographic representations of the sentences and texts, canonical, citation form word transcriptions were automatically generated in modified SAMPA, including punctuation marks. Subsequently, segmental and prosodic labelling was carried out in *xassp* (IPDS (1997); Kohler, Pätzold, and Simpson (1997)). This procedure produced a label file \*.s[12]h for each sentence or text section from the canonical transcription by allocating one canonical symbol after another to a point on the time scale of the acoustic signal file, also modifying the series of segmental labels, as well as adding time-aligned prosodic labels. A segmental label referring to a signal portion is bounded by the time mark of the next label, i.e.

segmental labelling is strictly linear. Labels without duration are put on the same time point as the next label that is associated with a time extension.

#### 4.2.2.1 General principles for segmental and prosodic labels

- (1) All segmental and prosodic labels (including syntactic punctuation) are placed on one tier.
  - Segmental labels in lexical items are prefixed by **##** at word boundaries, e.g. **##f** in *fort, fahren*  
**\$#** at word boundaries inside compounds, e.g. **\$#f** in *fortfahren*  
**\$** in all other cases, e.g. **\$t** in *fort*;  
 the prefixes are in turn followed by **%** if the segment boundary is uncertain.
  - **#** prefix of non-verbal labels and syntactic punctuation marks, e.g. **#p:** (pause), **#h:** (breathing), **#c:** (beginning of sentence), **#.** (period)
  - **&** indicates prefixes of prosodic, and only prosodic labels, and is put after **#** in the default case, e.g. **#&2^** (accent level 2 with medial peak)  
**\$** inside lexical items, e.g. in the case of stress shift, see PEAK Examples in 4.2.2.3 (4).
- (2) Label prefixes **##**, **#** and **\$** are obligatory in the label lines of the \*.slh files. In the running transcription section of these files, **##** and **#** are replaced by two blanks, **\$** by one (see 4.2.3).
- (3) Prosodic labels (as well as punctuation marks) are linked to a point in time, but have no duration. They receive the same time stamp as the subsequent label. So if there are several such labels in a row they are all put on the same time as the next label that is associated with a time extension. There are ordering conventions for prosodic labels, see section 4.2.2.3 B.
- (5) Uncertainty as to the prosodic category to be labelled is marked by **%**, placed after the prosodic index **&**, e.g. **&%PGn** (uncertain prosodic phrase boundary). In the case of two prosodic features in one label, each may be marked by **%**, e.g. **&%2^** or **&2^%** or **&%2^%**.

#### 4.2.2.2 Segmental labels: modified SAMPA

The canonical transcription is based on a phonemic system of German which recognizes 46 units  
 24 vowels      20 oral, and 4 nasal vowels, which may occur in loan words from French  
 22 consonants    3 (**x C Q**) have non-phonemic status, using them avoids the need to include morphological information to account adequately for the differences in the intervoallic dorsal fricative in pairs such as *Frauchen* [fʁaʊçən] and *rauchen* [ʁaʊxən] or the presence vs. absence of glottalization in pairs such as *vereisen* [fɛʔaɪzən] and *verreisen* [fɛʁaɪzən].

#### Vowels

SAMPA	IPA	Example
<b>a: a</b>	<b>a: a</b>	<b>k'a:m k'am</b> <i>kam</i> <i>Kamm</i>
<b>e: E</b>	<b>e: ε</b>	<b>b'e:t b'Et</b> <i>Beet</i> <i>Bett</i>
<b>i: I</b>	<b>i: ɪ</b>	<b>r'i:t r'It</b> <i>riet</i> <i>ritt</i>
<b>o: O</b>	<b>o: ɔ</b>	<b>b'o:k b'Ok</b> <i>bog</i> <i>Bock</i>
<b>u: U</b>	<b>u: ʊ</b>	<b>b'u:s@ b'Us@</b> <i>Buße</i> <i>Busse</i>
<b>y: Y</b>	<b>y: ʏ</b>	<b>h'y:t@ h'Yt@</b> <i>Hüte</i> <i>Hütte</i>

<b>E:</b>	<b>ɛ:</b>	<b>k'E:z@</b>	<i>Käse</i>
<b>2: 9</b>	<b>ø: œ</b>	<b>h'2:l@ h'9l@</b>	<i>Höhle Hölle</i>
<b>a~</b>	<b>ã</b>	<b>rEsto:r'a~</b>	<i>Restaurant</i>
<b>E~</b>	<b>ẽ</b>	<b>t'E~</b>	<i>Teint</i>
<b>O~</b>	<b>õ</b>	<b>se:z'O~</b>	<i>Saison</i>
<b>9~</b>	<b>œ</b>	<b>pa6f'9~</b>	<i>Parfum</i>
<b>aI</b>	<b>aɪ</b>	<b>tsv'aI</b>	<i>zwei</i>
<b>aU</b>	<b>aʊ</b>	<b>b'aUx</b>	<i>Bauch</i>
<b>OY</b>	<b>ɔʏ</b>	<b>n'Oyn</b>	<i>neun</i>
<b>@</b>	<b>ə</b>	<b>l'e:z@n</b>	<i>lesen</i>
<b>6</b>	<b>ɐ</b>	<b>l'e:z6</b>	<i>Leser</i>
		<b>h'e:6 h'E6</b>	<i>her Herr</i>

### Consonants

SAMPA	IPA	Example	
<b>b</b>	<b>b</b>	<b>b'aIn</b>	<i>Bein</i>
<b>d</b>	<b>d</b>	<b>d'aIn</b>	<i>dein</i>
<b>f</b>	<b>f</b>	<b>f'aIn</b>	<i>fein</i>
<b>g</b>	<b>g</b>	<b>g'ICt</b>	<i>Gicht</i>
<b>h</b>	<b>h</b>	<b>h'aIn</b>	<i>Hain</i>
<b>j</b>	<b>j</b>	<b>j'a:</b>	<i>ja</i>
<b>k</b>	<b>k</b>	<b>k'aIn</b>	<i>kein</i>
<b>l</b>	<b>l</b>	<b>l'ICt</b>	<i>Licht</i>
<b>m</b>	<b>m</b>	<b>m'aIn</b>	<i>mein</i>
<b>n</b>	<b>n</b>	<b>n'aIn</b>	<i>nein</i>
<b>p</b>	<b>p</b>	<b>p'aIn</b>	<i>Pein</i>
<b>r</b>	<b>r</b>	<b>r'aIn</b>	<i>rein</i>
<b>s</b>	<b>s</b>	<b>l'o:s</b>	<i>Los</i>
<b>t</b>	<b>t</b>	<b>l'o:t</b>	<i>Lot</i>
<b>v</b>	<b>v</b>	<b>v'aIn</b>	<i>Wein</i>
<b>z</b>	<b>z</b>	<b>z'aIn</b>	<i>sein</i>
<b>C</b>	<b>ç</b>	<b>z'ICt</b>	<i>Sicht</i>
<b>x</b>	<b>x</b>	<b>z'Uxt</b>	<i>Sucht</i>
<b>S</b>	<b>ʃ</b>	<b>S'aIn</b>	<i>Schein</i>
<b>Z</b>	<b>ʒ</b>	<b>l'o:Z@</b>	<i>Loge</i>
<b>N</b>	<b>ŋ</b>	<b>z'INt</b>	<i>singt</i>
<b>Q</b>	<b>ʔ</b>	<b>Q'aIns</b>	<i>eins</i>

## Label Modifications

In segmental labelling, one of four decisions has to be taken:

- (1) the canonical symbol is aligned without modification
- (2) the canonical symbol has no corresponding signal portion, it is marked as **deleted**, and is put on the same mark as the next non-deleted label: e.g. \$@-
- (3) the canonical symbol does not represent the signal portion adequately, it is marked as **replaced** by a phonetically more appropriate symbol from the canonical list: e.g. \$n-m
- (4) there is no canonical symbol on offer for the signal portion to be labelled, a label is **inserted**: e.g. \$-t .

Biphonemic canonical symbol sequences of a monophthongal vowel and the vowelised realization **6** of post-vocalic /r/ are amalgamated to a single label, which is aligned with the beginning of the entire vocalic portion of the signal, e.g.

Canonical	Label	Example
E 6	\$E6	<i>Herr</i>
e: 6	\$e:6	<i>Heer</i>
2: 6	\$2:6	<i>hört.</i>

This is done to avoid problems in attempting to segment such vocalic portions into two parts and to align the two symbols with the beginning of each. It follows from this procedure that in the canonical section of \*.s[12]h files the two symbols are separated by a blank, but they are on one label line (see 4.2.3). However, if the sequence is disyllabic, the label is split up and the second part is marked by = on a separate label line, e.g. ##f \$r \$'y: \$=6 *früher*

Replacements do not introduce symbols in addition to those contained in the canonical lists of vowels and consonants; they index changes to other categories within the vowel or the consonant system. In the case of vowels, only replacements by lax vowels, monophthongization or schwa are marked.

## Insertions

In addition to the insertion of symbols from the canonical list, such as \$-t in ##Q \$a \$l \$-t \$s , the following segment-related additional labels occur:

**\$-h** Plosive release phase: burst and any aspiration (this label has a duration)

**\$-q** Creaky voice

- (a) in addition to or instead of a glottal stop: ##Q \$-q ##Q- \$-q
- (b) to mark a morpheme boundary or the separation of vowels where a glottal stop is not symbolized in the canonical transcription:  
##n \$a \$x \$#-q \$aI \$n \$#-q \$a \$n \$d \$6 *nacheinander*  
##Q- \$-q \$E \$v \$E \$n \$t \$-h \$u: \$-q \$'E \$l *eventuell*
- (c) to replace lenis or fortis plosives in nasal or lateral contexts:  
##k \$9 \$n \$t-q \$-@ \$n+ *könnten*
- (d) to insert glottalization instead of a velar plosive that is not represented canonically  
##f \$6 \$b \$'I \$n \$d-n \$U \$N \$-kq ##n \$a: \$x+ *Verbindung nach*

In all these cases, the label has no duration and is put on the same time mark as the next non-deleted label.

**\$-~** Nasalization in the context of a deleted nasal consonant: placed after the latter if nasalization only affects the post-nasal sequence, or before it if it affects the preceding material:  
**##Q- \$U \$-~ \$-n \$s+** *uns*  
**##Q- \$E-@ \$s+ ##n- \$-~ \$I C \$t+** *es nicht.*

The label has no duration and is put on the same time mark as the next non-deleted label.

**\$-MA** Presence of phonetic correlates of deleted labels in preceding context:  
**##-MA ##v \$a:6-a \$S \$'aI \$n \$-MA \$I- \$I- \$C- ##Q- \$aI- \$n-m ##b \$'I \$s \$C \$@- \$n**  
*wahrscheinlich ein bißchen* [vaʃaɪnm bɪʃçn]  
 palatality of deleted syllable *-lich* transferred to preceding nasal

The label has no duration and is put at the beginning of the deleted sequence to indicate a phonetic residue in the segment(s) preceding it. The inserted label and all subsequent deleted symbols are put on the same time mark as the next non-deleted segmental label with a duration.

**#-p:** Insertion of pause or breathing, which are not in the orthographic text and are therefore not automatically generated in the canonical transcription, but are produced in reading; these labels have durations.  
**#-h:**

**\$-z:** Hesitational lengthening in reading, put after the lengthened segment on the same time mark as the next non-deleted segment.

**##-n:** “non-text”: deviates from orthographic text  
**##-n:** with duration of its own, marks a stretch of speech that is not represented orthographically and is not segmented and labelled, e.g. false starts

**k078butt1: ##-n: ##Q \$'aU \$s #f \$6 \$k \$-h \$'aU \$f \$t**  
 speaker erroneously reads **##f \$6** of *verkauft*, then corrects herself and reads the actual orthography *ausverkauft*; the false start is labelled as **##-n:**

**k027butt2: #-n: ##f \$'O6 \$d \$-h \$6 \$t \$-h \$@- \$n- ##-n: ##Q \$-q \$i: \$n+**  
 speaker erroneously replaces *forderten ihn* by *fordert ihn*, then corrects himself by repeating *ihn*; the canonical segments of *-en* are labelled as deleted and the false start between *fordert* and the repeated *ihn* as **##-n:** ; at the same time a deviant pronunciation of *forderten* is indicated by **##-n:** preceding the word, see below

**##-n:** marks a word as deviant in reading, e.g. slip of the tongue, and is put at the beginning of the misread item, which is segmentally labelled; it occurs on the same time mark as the first segment.

**k03be014 #-n: ##n-d \$-h \$e: \$b-m \$@- \$%n-m+ ##d-n \$e: \$m+ ##t \$-h \$'E \$I \$6**  
 in *neben dem Teller* the initial **n** is replaced by **d**, probably triggered by the subsequent initial oral **t** of *Teller* between two accented syllables; **##-n:** marks the deviant pronunciation of the following word

**k027butt2: #-n: ##f \$'O6 \$d \$-h \$6 \$t \$-h \$@- \$n- ##-n: ##Q \$-q \$i: \$n+**  
 see above

**k022butt2: #-n: ##Q- \$a- \$l- \$s-+ \$%S \$-ma \$o:- \$n+ ##-n:**

instead of the written sequence *als schon* the speaker produces *schon als*; the canonical labels of *als* are labelled as deleted, and the deviant reading is indicated by a preceding #-n: ; the signal stretch corresponding to the actually pronounced *als* after *schon* is labelled as ##-n: with no segmentation

#### 4.2.2.3 Prosodic labels: PROLAB

##### A Labels for prosodic categories

###### (1) PROSODIC PHRASE BOUNDARIES

- Syntagmatic prosodic phrasing is marked by phrase boundaries at the right edges of the prosodic chunks:  
**&PGn** in the default case
- In the main PROLAB system, prosodic phrase boundaries are not subclassified according to strength of separation. **n** is a place holder for subsequent finer differentiation.

###### (2) SENTENCE ACCENT

The lexical items within prosodic phrases receive an ACCENT specification on 4 levels, put before the segmental sequence

**&2** default accent

**&3** reinforced accent

**&1** partial deaccentuation

**&0** complete deaccentuation

e.g. **&2** *Max* **&0** *hat* **&0** *einen* **&3** *Brief* **&1** *geschrieben*. "Max has written a letter."

with narrow focus on *Brief* (answering "what did he write?"), complete deaccentuation of the function words "has" and "a", strong prominence on "Max", and weak prominence on "written".

###### (3) LEXICAL STRESS

Within words, one or more syllables are specified for STRESS.

- It is marked at the segmental level and linked to the symbol for the stressed vowel  
' primary stress  
" secondary stress  
e.g. **##f \$'O6 \$t \$f \$'a: \$r \$@ \$n** *fortfahren*
- Function words, identified by + postfixed to the last segment, by default do not get a lexical stress mark, e.g. **##m \$I \$t+** *mit*.
- If function words receive a sentence accent, \$' ' is inserted before the accented vowel, e.g. **##m \$' ' \$I \$t+** *'mit*.
- The same applies in the case of stress shift in relation to the canonical transcription, e.g. **##v \$ 'U \$n \$d \$6 \$b \$' ' \$a:6** *wunder'bar*.

###### (4) PITCH PATTERNS

Each ACCENT > **&0** receives a distinctive pitch pattern which is either

(rising-)falling – PEAK

or (falling-)rising – VALLEY

or FLAT

and is hooked to the environment of the stressed syllable of the accented word, more particularly to its stressed vowel.

It consists of two parts

- the SYNCHRONIZATION of the f0 maximum/minimum with the stressed vowel

- the CONTINUATION of the pattern beyond the synchronization point to concatenate with the next ACCENT pattern or to terminate at the next PHRASE BOUNDARY if there are no further accents in the prosodic phrase.

SYNCHRONIZATION is marked by an addition to the ACCENT label, CONTINUATION by the insertion of another prosodic label before the next ACCENT or PHRASE BOUNDARY label. In each case, the labels are specific for each of the three pitch patterns.

- PEAK

SYNCHRONIZATION

**&2)** EARLY synchronization of the peak maximum before the stressed vowel

**&2^** MEDIAL synchronization of the peak maximum in the stressed vowel

**&2(** LATE synchronization of the peak maximum late in, or after, the stressed vowel, similarly for **&1** and **&3**

CONTINUATION

**&2.** fall to low pitch, creating a large pitch dip between two accents or ending in the speaker's lowest voice range at the end of the prosodic phrase

**&1.** fall to medium pitch between two accents or at the end of the prosodic phrase

**&0.** no pitch dip between two accents (e.g. 'hat pattern' between two peaks) or ending level at the prosodic phrase boundary

**&2.,** or **&1.,** combined contours at phrase boundaries: fall is followed by low rise

**&2.?** or **&1.?** combined contours at phrase boundaries: fall is followed by high rise

**&2;** or **&1;** pseudoterminal contours at phrase boundaries: very slight rise after fall, softens terminality without removing it altogether.

### Examples

**#&0 am #&2^Samstag #&0 oder #&1. #&2( Sonntag #&2. #&PGn**

"on Saturday or Sunday" either – or alternatives

**#&0 am #&2^Samstag #&0 oder #&0. #&2) Sonntag #&2. #&PGn**

"on Saturday or Sunday" = "at the weekend"

**#&2^ tschüß #&1., #&PGn** "bye"

**#&2^ ##t \$-h \$E6 \$m \$'i: \$n \$&0. \$&2^ \$#v \$' '\$'Y \$n \$\$ \$&1. \$&2^ \$' '\$@ #&1.**

*Ter'min'wün'sche* "preferences for dates", pronounced with three accents: the first is hooked to the primary-stress syllable *min* of the canonical form and given a **#&** ACCENT label preceding the segmental string; the other two require primary lexical stress positions to be inserted into the canonical form by '\$' ', to which **\$&** ACCENT labels can then be linked inside the lexical item.

- VALLEY

SYNCHRONIZATION

**&2]** EARLY synchronization of the valley minimum before the stressed vowel

**&2[** LATE synchronization of the valley minimum in the stressed vowel

similarly for **&1** and **&3**

CONTINUATION

**&**, low rising

**&?** high rising, ending in the speaker's high voice range

### Examples

**&0 er &0 ist &0 in &2] Rome &? #&PGn** "he is in Rome?" asking for confirmation  
**&0 er &0 ist &0 in &3[ Rome &? #&PGn** "he is in Rome?" surprise question  
**&0 er &0 ist &0 in &2[ Rome &, #&PGn** "he is in Rome." friendly non-categorical statement

- FLAT pattern  
SYNCHRONIZATION  
**&2- &1- &3-**

CONTINUATION  
**&0. &1. &2.**

### Examples

**&2^ heute &0 ist &1. #&2( schönes &0. #&2- warmes &0. #&2) Sommerwetter  
&2. #&PGn**  
"today we have got beautiful warm summer weather"  
**&0 #&0 er hat &0 einen &2( Brief #&2. #&1- geschrieben #&2. #&PGn**  
"he has written a letter"

#### (5) DOWNSTEP and UPSTEP, RESET

In a sequence of peak contours of accent level 2 within a prosodic phrase, the maximum f0 is by default successively lowered. This is traditionally handled by DECLINATION over time. KIM treats it as structurally controlled DOWNSTEP from accent to accent, irrespective of the time that elapses between them. This default DOWNSTEP is not labelled. On the other hand, if the DOWNSTEP is interrupted within a prosodic phrase an UPSTEP is labelled by putting | in between the prosodic prefix and the ACCENT marker, e.g. **&|2^** .

Across PHRASE BOUNDARIES there is by default f0 RESET, i.e. DOWNSTEP starts afresh. If this is not the case and DOWNSTEP continues across the prosodic phrase boundary it is marked as **&=PGn** .

#### (6) PRE-HEAD

Unaccented syllables before the first PEAK ACCENT in a prosodic phrase are the PRE-HEAD of the contour and may be low or high in relation to the PEAK maximum. A low PRE-HEAD is taken to be default and not labelled. A HIGH PRE-HEAD is labelled as one of two different types  
**&HP1** flat f0 at level of PEAK maximum  
**&HP2** high f0 onset, gradually descending to level of PEAK maximum.

## B Ordering of prosodic labels

Punctuation marks, labels for CONTINUATION, for PHRASE BOUNDARIES, for sentence beginning, for HIGH PRE-HEAD, AND for type of ACCENT are put in this fixed order on the same time mark as the word boundary segment, also at word boundaries inside compound words.

<b>#[, . ? !]</b>	punctuation mark
<b>#&amp;[0. 1. 2. , ? etc.]</b>	CONTINUATION part of PITCH PATTERN
<b>#&amp;[PGn =PGn]</b>	PHRASE BOUNDARY
<b>#c:</b>	sentence beginning
<b>#-n:</b>	non-text
<b>#&amp;[HP1 HP2]</b>	HIGH PRE-HEAD

#& [1 2 3][(^) [ ]]	ACCENT, SYNCHRONIZATION of PITCH PATTERN
##	word boundary
\$&[0. 1. 2.]	CONTINUATION part of PITCH PATTERN inside compound
\$& [1 2 3][(^) [ ]]	ACCENT, SYNCHRONIZATION of PITCH PATTERN inside compound
\$#	boundary of compound element
\$' '	inserted lexical stress before vowel to receive ACCENT

### 4.2.3 Structure and generation of label files (\*.s[12]h)

Label files \*.s1[12]h have the following structure:

- a. file name
- b. 7-bit orthographic representation of the sentence bounded by 'oend' at the beginning of the next line
- c. canonical transcription of the sentence in SAMPA, bounded by 'kend' at the beginning of the next line
- d. phonetic transcription of the utterance: segmental (modified SAMPA) in \*.s1h, segmental and prosodic (PROLAB) in \*.s2h, constructed from the labels placed by the segmenter, and bounded by 'hend' at the beginning of the next line
- e. label lines with time stamp (sample number), label, time point in sec
  - the segmenter enters a prototype file \*.s0 into *xassp*, having the sections a., b. and c., and label lines with imaginary sample numbers and time points 100 sec in the future
  - by placing each label, one after the other as offered by the program, on the time scale of the signal file the imaginary values are changed to real ones
  - subsequently section d. of the label file is constructed
  - the \*.s0 files of the add(itional)corp(us), which has not been labelled, consequently do not contain section d.

### Examples

```

k01be004.s0
"Uber die Felder weht ein Wind .
oend
  Q y: b 6+ d i:+ f 'E l d 6 v 'e: t Q aI n+ v 'I n t .
kend
hend
1600000 #c:          99.9999375
1600000 ##Q         99.9999375
1600000 $y:          99.9999375
1600000 $b           99.9999375
1600000 $6+          99.9999375
1600000 ##d          99.9999375
1600000 $i:+         99.9999375
1600000 ##f          99.9999375
1600000 $'E          99.9999375
1600000 $l           99.9999375
1600000 $d           99.9999375
1600000 $6           99.9999375
1600000 ##v          99.9999375
1600000 $'e:         99.9999375
1600000 $t           99.9999375
1600000 ##Q          99.9999375
1600000 $aI          99.9999375

```

```

1600000 $n+          99.9999375
1600000 ##v         99.9999375
1600000 $'I        99.9999375
1600000 $n          99.9999375
1600000 $t          99.9999375
1600000 #.          99.9999375

```

k01be004.s1h

"Uber die Felder weht ein Wind .

oend

Q y: b 6+ d i:+ f 'E l d 6 v 'e: t Q aI n+ v 'I n t .

kend

c: Q- y: b 6+ d -h i:+ f 'E l d -h 6 v 'e: t -h Q- -q aI n+  
v 'I n t -h .

hend

```

3225 #c:          0.2015000
3225 ##Q-        0.2015000
3225 $y:          0.2015000
3847 $b          0.2403750
4741 $6+         0.2962500
5837 ##d         0.3647500
6472 $-h        0.4044375
6666 $i:+       0.4165625
7570 ##f         0.4730625
9175 $'E        0.5733750
10460 $l         0.6536875
11506 $d         0.7190625
11657 $-h       0.7285000
11926 $6         0.7453125
13715 ##v        0.8571250
14870 $'e:       0.9293125
16309 $t         1.0192500
16814 $-h       1.0508125
16958 ##Q-      1.0598125
16958 $-q       1.0598125
16958 $aI       1.0598125
18948 $n+       1.1841875
20015 ##v       1.2508750
21173 $'I       1.3232500
22161 $n        1.3850000
23749 $t        1.4842500
24322 $-h      1.5200625
24730 #.        1.5455625

```

k01be004.s2h

"Uber die Felder weht ein Wind.

oend

Q y: b 6+ d i:+ f 'E l d 6 v 'e: t Q aI n+ v 'I n t .

kend

c: &0 Q- y: b 6+ &0 d -h i:+ &2( f 'E l d -h 6 &0 v 'e: t -h  
&0 Q- -q aI n+ &0. &2) v 'I n t -h . &2. &PGn

hend

```

3225 #c:          0.2015000
3225 #&0         0.2015000
3225 ##Q-        0.2015000
3225 $y:          0.2015000
3847 $b          0.2403750
4741 $6+         0.2962500
5837 #&0        0.3647500
5837 ##d         0.3647500
6472 $-h        0.4044375
6666 $i:+       0.4165625
7570 #&2(       0.4730625

```

7570	##f	0.4730625
9175	\$'E	0.5733750
10460	\$l	0.6536875
11506	\$d	0.7190625
11657	\$-h	0.7285000
11926	\$6	0.7453125
13715	#&0	0.8571250
13715	##v	0.8571250
14870	\$'e:	0.9293125
16309	\$t	1.0192500
16814	\$-h	1.0508125
16958	#&0	1.0598125
16958	##Q-	1.0598125
16958	\$-q	1.0598125
16958	\$aI	1.0598125
18948	\$n+	1.1841875
20015	#&0.	1.2508750
20015	#&2)	1.2508750
20015	##v	1.2508750
21173	\$'I	1.3232500
22161	\$n	1.3850000
23749	\$t	1.4842500
24322	\$-h	1.5200625
24730	#.	1.5455625
24730	#&2.	1.5455625
24730	#&PGn	1.5455625

#### 4.2.4 Data banks (\*.kdb)

Each line of the data bank has the following field structure:

- 1 orthographic representation of lexical item
- 2 position of word in orthographic part of the label file
- 3 orthographic representation of an overlaying event, e.g. laughing
- 4 position of the overlaying event in orthographic part of the label file
- 5 canonical transcription of lexical item
- 6 transcription of pronunciation variant
- 7 date and time of label file inclusion in database
- 8 full filename of label file (including path)
- 9 basename of label file (i.e. no path suffix)
- 10 speaker and text file index
- 11 gender
- 12 format of labels (SAMPA notation)
- 13 sample frequency in Hz
- 14 start of word in seconds
- 15 end of word in seconds
- 16 (and all even-numbered fields following, i.e. 18, 20,...) start times of labels
- 17 (and all odd-numbered fields following, i.e. 19, 21,...) durations of labels

Elements of an entry are separated by tabs. Cf. Pätzold (1997).

## 5 Speaker information

The columns in the table below contain information about the speakers as follows:

1. Time of recording (month/year)
2. Speaker abbreviation

3. Gender (f= female; m = male)
4. Age
5. German spoken (classification made by speakers themselves):
  - SH – Schleswig-Holstein
  - NI – Lower Saxony (Niedersachsen)
  - ND – North German (Norddeutsch)
  - NW – North Rhine-Westfalia (Nordrhein-Westfalen)
  - HE – Hessen
  - RP – Rhineland-Palatinate (Rheinland-Pfalz)
  - MX – mixed
6. type of files
7. number of files
  - a: add(itional)corp(us), not labelled, only \*.wav and \*.s0 (orthographic and canonical)
8. duration of recording

09/90	k01	m	25	SH	berlin	100	4 min 35 sec
09/90	k02	f	25	SH	berlin	100	4 min 29 sec
09/90	k03	m	28	SH	berlin	100	4 min 04 sec
09/90	k04	f	26	SH	berlin	100	5 min 40 sec
09/90	k05	m	28	SH	berlin	100	5 min 42 sec
09/90	k06	f	23	RP	berlin	100	4 min 41 sec
09/90	k07	m	23	SH	marburg	100	5 min 04 sec
09/90	k08	f	24	SH	marburg	100	5 min 30 sec
09/90	k09	m	27	SH	marburg	100	4 min 55 sec
10/90	k10	f	24	SH	marburg	100	5 min 08 sec
09/90	k11	m	24	SH	marburg	100	5 min 20 sec
09/90	k12	f	22	SH	marburg	100	4 min 11 sec
09/90	k13	m	25	SH	nordwind	2	0 min 46 sec
					restkorp	a:53	3 min 23 sec
09/90	k14	f	26	SH	nordwind	2	0 min 39 sec
					restkorp	a:53	3 min 09 sec
09/90	k15	m	25	SH	nordwind	2	0 min 49 sec
					restkorp	a:53	3 min 14 sec
09/90	k16	f	23	SH	nordwind	2	0 min 43 sec
					restkorp	a:53	3 min 08 sec
09/90	k17	m	27	SH	nordwind	2	0 min 45 sec
					restkorp	a:53	2 min 57 sec
09/90	k18	f	28	NW	nordwind	2	0 min 43 sec
					restkorp	a:53	3 min 00 sec
09/90	k19	m	27	SH	nordwind	2	0 min 54 sec
					restkorp	a:58	4 min 17 sec
09/90	k20	f	24	NW	nordwind	2	0 min 47 sec
					restkorp	a:58	3 min 35 sec
09/90	k21	m	26	SH	nordwind	2	0 min 40 sec
					restkorp	a:58	3 min 02 sec
09/90	k22	f	26	SH	butter	3	1 min 17 sec
					restkorp	a:41	2 min 15 sec
09/90	k23	m	24	SH	butter	3	1 min 32 sec
					restkorp	a:41	2 min 29 sec
09/90	k24	f	20	SH	butter	3	1 min 17 sec
					restkorp	a:41	2 min 48 sec

10/90	k25	m	27	SH	butter	3	1 min 31 sec
					restkorp	a:41	2 min 51 sec
09/90	k26	f	23	NI	butter	3	1 min 17 sec
					restkorp	a:41	2 min 24 sec
10/90	k27	m	30	SH	butter	3	1 min 23 sec
					restkorp	a:41	2 min 31 sec
10/90	k28	f	25	NI	butter	3	1 min 15 sec
					restkorp	a:41	2 min 11 sec
09/90	k29	m	28	SH	butter	3	1 min 42 sec
					restkorp	a:41	2 min 53 sec
10/90	k30	m	24	NI	butter	3	1 min 23 sec
					restkorp	a:46	2 min 38 sec
09/90	k61	m	55	MX	compl.corp.	403	26 min 07 sec
09/90	k62	f	33	NI	compl.corp.	403	23 min 45 sec
09/90	k63	m	33	SH	berlin	100	4 min 30 sec
09/90	k64	f	50	ND	berlin	100	5 min 04 sec
09/90	k65	m	35	MX	berlin	100	4 min 34 sec
09/90	k66	f	39	MX	berlin	100	4 min 31 sec
09/90	k67	m	35	RP	marburg	100	5 min 11 sec
09/90	k68	f	50	NI	marburg	100	5 min 02 sec
09/90	k69	m	49	HE	marburg	100	4 min 07 sec
09/90	k70	f	36	SH	marburg	100	5 min 13 sec
09/90	k71	m	38	ND	nordwind	2	0 min 44 sec
					restkorp	a:53	2 min 51 sec
10/90	k72	f	46	NI	nordwind	2	0 min 52 sec
					restkorp	a:53	3 min 09 sec
09/90	k73	m	47	SH	nordwind	2	0 min 48 sec
					restkorp	a:53	3 min 29 sec
10/90	k74	f	56	SH	nordwind	2	1 min 03 sec
					restkorp	a:58	5 min 27 sec
09/90	k75	m	52	ND	nordwind	2	0 min 43 sec
					restkorp	a:58	3 min 54 sec
09/90	k76	f	30	SH	butter	3	1 min 36 sec
					restkorp	a:41	2 min 34 sec
09/90	k77	m	48	ND	butter	3	1 min 34 sec
					restkorp	a:41	2 min 39 sec
09/90	k78	f	47	SH	butter	3	1 min 18 sec
					restkorp	a:41	2 min 08 sec
09/90	k79	m	42	SH	butter	3	1 min 43 sec
					restkorp	a:41	3 min 37 sec
09/90	k80	f	48	SH	butter	3	1 min 45 sec
					restkorp	a:46	3 min 44 sec
04/92	d1m	m	21	SH	erlangen/siemens	200	13 min 13 sec
04/92	hpt	m	24	SH	erlangen/siemens	200	15 min 01 sec
04/92	kko	m	57	MX	erlangen/siemens	200	17 min 23 sec
	= k61						
04/92	rtd	f	35	NI	erlangen/siemens	200	17 min 07 sec
	= k62						
04/92	uga	f	23	SH	erlangen/siemens	200	13 min 29 sec

PhonDat90 labelled data : 178 min 52 sec  
PhonDat90 audio only : 86 min 16 sec  
PhonDat90 all audio : 265 min 08 sec  
PhonDat92 : 76 min 14 sec  
PhonDat labelled data : 255 min 06 sec  
PhonDat all audio : 341 min 23 sec

## 6 References

- IPDS (1997). xassp User's Manual (Advanced Speech Signal Processor under the X Window System. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, vol. 32 (Institut für Phonetik und digitale Sprachverarbeitung, Kiel), 31-115.
- Kohler, J. K. (Ed.) (1992), Phonetisch-Akustische Datenbasis des Hochdeutschen. Kieler Arbeiten zu den PHONDAT-Projekten 1989-1992. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 26 (Institut für Phonetik und digitale Sprachverarbeitung, Kiel).
- Kohler, K. J., Pätzold, M., and Simpson, A. (1995). From scenario to segment: the controlled elicitation, transliteration, segmentation and labelling of spontaneous speech. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, vol. 29 (Institut für Phonetik und digitale Sprachverarbeitung, Kiel).
- Kohler, K. J., Pätzold, M., and Simpson, A. (1997). From the acoustic data collection to a labelled speech data bank of spoken Standard German. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, vol. 32 (Institut für Phonetik und digitale Sprachverarbeitung, Kiel), 1-29.
- Pätzold, M. (1997). KielDat - Data bank utilities for the Kiel Corpus. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, vol. 32 (Institut für Phonetik und digitale Sprachverarbeitung, Kiel), 117-126.
- Peters, B., and Kohler, K. J. (2004). Trainingsmaterialien zur prosodischen Etikettierung mit dem Kieler Intonationsmodell KIM ("Training Material for Prosodic Labelling with the Kiel Intonation Model KIM"). Manuscript.  
URL: <http://www.ipds.uni-kiel.de/kjk/forschung/lautmuster.en.html>

## **Queries and comments**

Any queries or comments can be directed to:  
Dr Michel Scheffers  
Institut für Phonetik und digitale Sprachverarbeitung der  
Christian-Albrechts-Universität zu Kiel  
D - 24098 Kiel  
Germany  
Tel: ++49 431 880 3321  
Fax: ++49 431 880 1578  
e-mail: [ms@ipds.uni-kiel.de](mailto:ms@ipds.uni-kiel.de)  
www: <http://www.ipds.uni-kiel.de>

## **Limits of Liability**

Great care has been taken in the preparation of Kiel-CD #5, however the Editors are not responsible for any damage arising out of the use of this CD-ROM or the data contained on it.

## **Copyright notice**

The data on the CD-ROM are copyrighted. They are for research purposes only. No form of commercial use is allowed, nor may the data be sold or transferred, as a whole or in parts, to another party. The files on the CD-ROM or any part thereof may not be reproduced in any form without the written permission from

The Editors of Kiel-CD #5  
Institut für Phonetik und digitale Sprachverarbeitung der  
Christian-Albrechts-Universität zu Kiel  
D - 24098 Kiel  
Germany  
© September 2007