

From the acoustic data collection to a labelled speech data bank of spoken Standard German^{*}

Klaus J. Kohler, Matthias Pätzold and Adrian P. Simpson

1 Introduction

1.1 Global aims for corpus-based acoustic speech data processing

Large computer-accessible speech data bases have become a prerequisite for modern studies of phonetic (segmental and prosodic) regularities in individual languages, their dialectal varieties and their different speaking style realizations. This requirement applies to basic phonetic and phonological research as well as to its applications, e.g. in speech technology, where extensive data for training speech recognition systems, and statistically representative data bases for, e.g., the implementation of pronunciation rules in automatic speech output (text-to-speech synthesis) are necessary. However, the speech wave recordings themselves are not sufficient for phonetic and linguistic investigations; for the data to be retrievable they have to be annotated. In its simplest form this may be an orthographic representation of the spoken words (**transliteration**) and their alignment to the time scale of the related speech signal (**segmentation and labelling**).

In the case of read speech the orthographic form is given for data acquisition, and the task of a simple annotation consists in the spelling-to-signal alignment of words. In the case of spontaneous speech a human observer has to provide an orthographic transliteration, subsequent to the speech recording, by applying the lexical spelling and the punctuation rules of the language, on the one hand, and by symbolizing ad-

^{*}The work reported here was funded partly by the German Federal Ministries of Research and Technology (BMFT)/ Education, Science, Research and Technology (BMBF) under ASL and VERBMOBIL contracts 01 IV103 B4 and 01 IV101 M7 and partly by the University of Kiel. The responsibility for the contents lies with the IPDS Kiel.

ditional, non-lexical phenomena, such as pauses, breathing, hesitations, truncations and resumptions, etc., as well as dialectal and stylistic variants of words, on the other hand. These systematic transliterations thus require codified sets of symbols and rules of application.

As the orthographic representations of words are not always phonetically unique they need to be transformed into phonological/phonetic symbolizations of lexical citation form pronunciations as listed in a dictionary in context-free form for each word (**canonical transcription**). These phonetic notations, however, do not tell us yet how the individual words are actually pronounced in the particular contexts of the speech recording. In order to capture this information symbolically the speech waves have to be segmented into sound units and provided with phonetic labels (**transcription of phonetic variants**). On the basis of phonetic corpora processed in this way we can ask, and find answers to, the question: “How are the words of the language under investigation pronounced?”

In the pursuit of this goal we need

- to work out phonetic transcription conventions, over and above the orthographic transliteration code, for the adequate phonetic representation of read and of a broad spectrum of spontaneous speech
- to integrate the speech wave and corresponding spelling and label files in a structured data bank.

A wide array of pronunciation issues can then be studied by setting up a library of search routines to be applied to such a phonetic data bank, including the analysis and statistical evaluation of phonetic manifestations of different speaking styles in the symbol and signal domains.

1.2 Aims of the ASL and VERBMOBIL projects

The task outlined in 1.1 was part of two successive long-term programs funded by the German Federal Ministry of Research and Technology (BMFT) and its successor, the German Federal Ministry for Education, Science, Research and Technology (BMBF), respectively: ASL (Architecture for Speech and Language) and VERBMOBIL (Mobile Translation System for Face-to-Face Dialogue) (Karger and Wahlster 1995). Both had the automatic recognition of continuous speech as their research goal. In ASL the speech environment was defined by the scenario of a train information system in the reading mode. In VERBMOBIL spontaneous speech of an appointment scheduling scenario within a specified technical recording setup became the standard, and automatic translation and synthetic speech output were added as further work areas. For the tasks in both projects, setting up a speech signal data bank of spoken German was of fundamental importance.

It took its point of departure from the PHONDAT concept developed in a previous project (Kohler 1992b, 1992e). Therefore, the collection and preprocessing of speech

data in both ASL and VERBMOBIL were incorporated into a sub-project PHONDAT (Hess et al. 1995). IPDS Kiel contributed to the PHONDAT work packages within a consortium consisting also of the phonetics institutes in Bonn and Munich as well as the Department of Computer Science in Karlsruhe. In VERBMOBIL it was furthermore responsible for the package ‘Pronunciation Variants’ of the sub-project ‘Speech Synthesis’, which was closely linked to PHONDAT. In a series of interrelated steps IPDS Kiel proceeded from the acoustic data collection to the compilation of a representative list of phonetic variants of words for application in speech recognition and speech synthesis.

1.3 Tasks and work plan of IPDS Kiel in ASL-PHONDAT and VERBMOBIL-PHONDAT

Within the framework outlined in 1.2, IPDS Kiel completed the following tasks in a logically dependent sequence of steps.

- Step 1: Acoustic data collection in ASL/VERBMOBIL

In both projects, IPDS contributed continually to the acoustic data acquisition besides the phonetic institutes at the Universities of Bonn and Munich. In VERBMOBIL the Department of Computer Science at Karlsruhe University was a further partner for this task (Hess et al. 1995; Kohler 1992e; Thon 1992; Thon and Dommelen 1992).

- Step 2: Handbook for the transliteration of dialogues in VERBMOBIL

Within VERBMOBIL IPDS was given the initial task of compiling a handbook for the orthographic transliteration of spontaneous dialogues, which

- defined the objects of symbolization
- provided a systematic inventory of symbols for it
- and laid down the conventions for their use.

This handbook (Kohler et al. 1994) became the basis for the transliteration (Step 3) of the speech data that were collected at Bonn, Karlsruhe, Kiel and Munich and distributed on CD-ROM (Step 1).

- Step 3: Orthographic transliteration of dialogues in VERBMOBIL

For the orthographic transliteration of speech data IPDS created a user-friendly platform on workstations which provides

- the graphic display of speech signals
- acoustic output

- parallel word processing for fast generation of transliteration files on the basis of the IPDS transliteration handbook
- automatic spelling and consistency checks.

- Step 4: Generation of canonical phonetic transcription from orthographic texts in ASL/VERBMOBIL

The grapheme-to-phoneme converter of the RULSYS/INFOVOX TTS synthesis system for German (Carlson et al. 1990; Kohler 1992a) was used to automatically generate canonical transcriptions from the orthographic texts in ASL or from the output of Step 3 in VERBMOBIL. The phonetic transcriptions were checked manually and corrected at a rate of approximately 3% of running text. This automatic grapheme-to-phoneme converter was first developed at IPDS outside PHONDAT for orthographic standard text and was therefore applicable to the text corpus in ASL without further modification. In VERBMOBIL, however, it had to be adapted to the special demands of transliterated spontaneous text input and at the same time take the phonetic symbolization conventions (Step 6) into account.

- Step 5: Automatic generation of canonical lexica from canonical transcriptions in ASL/VERBMOBIL From the canonical transcription texts of Step 4 canonical pronunciation lexica with word frequencies were automatically derived for each partial as well as, cumulatively, for the total corpus by applying specially developed data bank search tools (Pätzold 1997).

- Step 6: Handbook of phonetic transcription

A handbook had to be provided to lay down the phonetic symbolization code for canonical transcriptions (Step 4), as well as the segmental and prosodic labelling conventions in relation to canonical transcriptions (Step 7) (Dommelen 1992a, 1992b; Kohler, Pätzold, and Simpson 1994, 1995).

- Step 7: Manual label files in ASL/VERBMOBIL

Subsequent to the generation of canonical transcriptions (Step 4) IPDS produced segmental and prosodic label files for speech data, collected at Kiel, in interactive computer processing. This task presupposed segmental and prosodic labelling conventions in relation to canonical transcriptions (Step 6). The software platform that was used at the outset was the MIX program for Apollo workstations (Carlson and Granström 1985), imported from KTH Stockholm and adapted to the Kiel data processing environment. For the development of a new software package see 1.4. The manual label files are the input to the generation of variants lexica and lists of variants (Steps 8,9), which in turn provide the basis for setting up contextual pronunciation rules (reductions and elaborations) in Step 10.

- Step 8: Automatic generation of variants lexica in ASL/VERBMOBIL

From the segmental label files (Step 7) variants lexica were generated for the processed read (Kohler 1994c) or spontaneous speech corpus and continually updated, with the help of fully automatic search tools, developed at IPDS (see Pätzold 1997). These variants lexica are a necessary complement to the canonical lexica: they contain the complete inventory of word forms of the processed speech files, and give orthographic, canonical phonetic and labelled phonetic representations together with information on frequencies of occurrence of each word and its variants, as well as on corpus references of individual forms.

- Step 9: Generation of variants lists in ASL/VERBMOBIL

The complete variants lexicon (Step 8), as such, is not of much use to speech recognition and speech synthesis, but it is an important prerequisite for the derivation of variants lists according to statistical criteria. A variants list only contains a limited number of the most frequent basic variants above an empirically determined threshold. From the variants lexicon we may also deduce a small set of rules which can generate a large number of frequently occurring variants from very few base forms. Base variants and rules together constitute the variants list, which still only represents a section of the total variants lexicon: many forms of low probability of occurrence are excluded right away. The outcome is not only essential for speech recognition and speech synthesis, but is, moreover, an important basis for the development of automatic segmentation and labelling techniques.

- Step 10: Corpus search, data analysis and rules for the phonetic realization of words

On the strength of the segmentally labelled data base (Step 7) it is also possible to set up rules of the phonetic realization of German words in read or spontaneous speech. These rules may then be the input to speech synthesis and speech recognition systems. In this connection of phonetic rule generation the development of a speech data bank and search tools is essential (see 1.4 and Pätzold 1997 for further work at IPDS towards this goal).

1.4 The Kiel Corpus of Read/Spontaneous Speech

In order to broaden the data base for the generation of phonetic word realization rules, and thus to strengthen the research results, IPDS labelled more data segmentally according to Step 7 than was prescribed by the work plans in the two projects; project-external resources (institute staff and university funding) were used to complete this work. Within ASL the entire PHONDAT data base recorded at Kiel was provided with segmental transcriptions, and from the label files a variants lexicon of read speech was derived (IPDS 1994; Kohler 1994c). The labelled data base and the variants lexicon

are now large enough to carry out investigations into the sound patterns of read speech at the sentence level.

This expansion of the labelled data base was still more urgent in VERBMOBIL because phonetic rule generation (Step 10) was defined as IPDS's task in the sub-project 'Speech Synthesis', and presupposed the provision of sufficient amounts of segmentally transcribed data (Step 7), which the other partners in the data acquisition consortium, however, did not supply. About two thirds of the VERBMOBIL data recorded at Kiel have been labelled segmentally, and a variants lexicon for spontaneous speech has been compiled cumulatively (IPDS 1995, 1996, 1997a). The labelled PHONDAT and VERBMOBIL read and spontaneous speech corpora are now of a comparable size (over 30,000 words each) so that phonetic comparisons of the two speaking styles become possible as well.

Step 7 within the VERBMOBIL tasks also included prosodic labelling: on the basis of the conventions laid down in the handbook of phonetic transcription (Step 6), 31 dialogues (about one quarter of the segmentally labelled corpus) have also been provided with prosody markings, and this labelling is being continued. It has also been extended to the PHONDAT corpus: about one third has so far been labelled prosodically.

In order to make all the labelled IPDS data of read and spontaneous speech easily accessible the segmental label and corresponding signal files together with the respective canonical and variants lexica have been issued on CD-ROMs (with financial support from the University of Kiel) under the title 'The Kiel Corpus of Read/Spontaneous Speech' (IPDS 1994, 1995, 1996, 1997a). CD-ROM#1 contains the read data, CD-ROMs#2,3,4 the spontaneous data processed up to July 1997. This series of CD-ROMs of the *Kiel Corpus* will be continued. The additional prosodic label files will also be made available shortly.

It was mentioned in 1.3 under Step 7 that the phonetic labelling was first carried out on Apollo workstations within the imported MIX environment. As the MIX program did not run on any other computer system and the Apollo series was furthermore no longer supported by HP, and because of a bottleneck in the use of the available two stations by too large a number of labellers, IPDS has now developed its own very flexible labelling and analysis platform, outside the VERBMOBIL funding: the program package *xassp*— **A**dvanced **S**peech **S**ignal **P**rocessor under the *X Window System*, which is described in IPDS (1997b).

For the automatic compilation of word lists as well as canonical and variants lexica from the orthographic text and phonetic label files or for symbol-oriented data base searches a speech data bank was implemented to cope with a wide array of phonetic questions, among others the sound patterns of words in connected speech. This data bank allows quick access to symbolic chains, defined with reference to the phonetic problem under investigation; they can then be matched with the corresponding speech waves, which may in turn be analysed in *xassp*. The data bank is described in Pätzold (1997). It was developed at IPDS outside VERBMOBIL funding, but applied to a variety of VERBMOBIL tasks.

The labelled acoustic data base – *Kiel Corpus of Read/Spontaneous Speech*– the grapheme-to-phoneme converter for German orthographic text, the labelling and analysis tool kit for speech signals – *xassp*– and the Kiel speech data bank platform constitute the type of powerful integrated processing system for spoken language mentioned in 1.1 as a prerequisite to modern phonetic corpus studies. The following sections will summarize the results that have been obtained by Steps 1-10 within, and also outside, the ASL and VERBMOBIL projects.

2 From Recording to Transliteration

2.1 German Read Speech

The *Kiel Corpus of Read Speech*, Vol. I on CD-ROM#1 (IPDS 1994) comprises the following data:

- 598 sentences and 2 stories ('The North Wind and the Sun', 'The Butter Story')
- containing 1671 word types and 4932 word tokens
- from 53 speakers (27 male, 26 female) 2 of whom read the entire corpus, 3 a subset of 200 sentences, and 48 either subsets of 100 sentences or one of the two stories (Thon and Dommelen 1992; Thon 1992)
- with a total of 31,374 recorded words
- in speech and label files as well as in lexica of canonical citation forms and phonetic variants.

The data were recorded in a sound-treated room, separately for each speaker, using a condenser microphone Neumann U87. Subjects read the sentences, one at a time, from program-controlled monitor prompts in ordinary German spelling. The texts were read from boards; 'The North Wind and the Sun' was presented as a whole, 'The Butter Story' in three sections. For each read sentence and each read text (section), ASCII-coded and stored in individual text files, a separate digital signal file (16 bit/16kHz) was created. For further details concerning the recording and the computer preprocessing see Thon and Dommelen (1992), Thon (1992).

2.2 German Spontaneous Speech

Following the original VERBMOBIL guidelines (Karger and Wahlster 1995) recordings have been carried out with two dialogue partners communicating via headsets (Sennheiser HDM 410 or 414) between two separate quiet rooms. The recording environment ensures high quality speech and good channel separation. Speakers have to press a button whenever they wish to speak to their partner. Only when the button

is pressed, which is signalled by a green lamp lighting up, can a speaker be heard by the dialogue partner and recorded. The pressing of the button also blocks the other speaker's channel. This set-up leads to a strict delimitation of turns.

The speech signals are recorded directly to hard disk into a multiplex stereo file (2x16bit/16kHz), on a PC AT486/66 platform with about 500MB disk space, sufficient for recording sessions in excess of one hour. A backup on DAT is produced at the same time. A DSP (Loughborough LSI96002 board) controls the I/O channels. For each dialogue there is a single file which is subsequently demultiplexed and split automatically into two files, one for each channel. The automatic splitting of each of these into separate turn files is made possible by recording constant known signal markers onto the other channel while a speaker is holding the button pressed (Pätzold et al. 1995).

The data acquisition platform is kept so flexible as to allow the realization of other scenario constraints in the data recording and processing, such as overlapping dialogues (without button pressing). In order to get as much material as possible from a single speaker each dialogue session is divided into subsessions. The data included in the *Kiel Corpus of Spontaneous Speech* on the CD-ROMs#2,3,4 published so far were obtained in a scenario of eight subsessions, the first of which is used for familiarization with equipment and task and for setting the recording level, but is excluded from the corpus. In each subsession the subjects are given fresh sets of calendar sheets and an academic time table, with different shaded areas (representing unavailability) for each of the two dialogue partners. The task in each case is to make appointments of a pre-specified nature. For further details see Kohler, Pätzold, and Simpson (1995), Pätzold and Simpson (1994), Pätzold et al. (1995).

The *Kiel Corpus of Spontaneous Speech* on CD-ROMs#2,3,4 (IPDS 1995, 1996, 1997a) includes

- 117 dialogue subsessions
- from 21 speaker pairs, 24 male and 18 female speakers
- with 2085 word types and a total of 37,849 recorded words.

Whereas in the case of read speech the recording succeeds an orthographic model, the procedure is reversed in the acquisition of spontaneous speech: the speech recordings have to be transliterated orthographically post hoc. To reserve the term 'transcription' for phonetic symbolization, an orthographic rendering of speech through spelling conventions is referred to as 'transliteration'. With regard to lexical material and punctuation the spelling conventions of DUDEN (DUDEN 1991) are applied, but they have had to be extended to cover non-lexical and paralinguistic phenomena, such as hesitations, pauses, breathing, dysfluencies, external noises etc. The extended alphabet in ASCII coding as well as the rules for its use are set out in Kohler et al. (1994), Kohler, Pätzold, and Simpson (1995). The turns of one dialogue subsession are transliterated in a single text file, using the interactive computer environment mentioned in 1.3 Step 3.

3 From Canonical Transcription to Phonetic Variance

The further speech processing stages are identical, irrespective of read or spontaneous data.

3.1 Generating canonical transcriptions

In all cases the orthographic text files are automatically converted into segmental phonemic transcription using the grapheme-to-phoneme module and the pronunciation exceptions lexicon of the RULSYS/INFOVOX German TTS system (Kohler 1992a). To cope with the expanded symbolic repertoire of spontaneous speech transliterations the transformation rules, originally devised for standard orthographic text, have had to be supplemented. The alphabet used is modified and augmented SAMPA (Wells et al. 1989). For each orthographic text file input the module automatically generates a transcription file output. It is manually corrected and represents a lexical citation form pronunciation: **canonical transcription**.

The canonical word pronunciations are nevertheless generated from continuous text, not from derived word lists. The correct lexical interpretation of homographs, which is essential for further processing, is only possible via linguistic context information, e.g. for “aus” (function word in “aus Dresden” vs. adverbial verb particle in “wie sieht es bei Ihnen aus”). In such cases the automatic text-to-transcription conversion module carries out the correct separation on the basis of an incorporated rudimentary syntax component. Thus the Kiel processing system allows the automatic rule-governed derivation of canonical transcriptions from running orthographic text. At the Munich Phonetics Institute, on the other hand, canonical forms are retrieved from a canonical lexicon through the orthographic reference for each individual word. This is inadequate for several reasons:

- For the disambiguation of homographs and the attribution to different word classes with separate phonetic characteristics the syntactic context has to be considered which means that a simple orthographic-to-canonical correspondence list without morphological and syntactic information is not sufficient, but existing pronunciation dictionaries do not usually provide it.
- Words that are not yet represented in the canonical lexicon have to be transcribed manually and added to the dictionary ad hoc.
- The same applies to non-words, neologisms and word fragments of various origins.
- The orthographic transliteration of word dysfluencies (hesitation lengthening) as well as overlays of articulatory or non-articulatory noise on lexical units cannot be properly converted to phonetic transcription by the lexical look-up method.

- Non-lexical units, such as pauses, breathing, either have to be ignored in the canonical transcription or relegated to different tiers, which makes the investigation of segmental and prosodic interactions more difficult.

For these reasons a rule-governed canonical transcription system is to be preferred to the lexical look-up procedure. This also means that the relation between a canonical lexicon and canonical texts is reversed: the canonical lexicon is derived from canonical texts, not vice versa.

As a basis for subsequent manual phonetic labelling, so-called **prototype label files** are generated automatically from the canonical transcription files: a program combines corresponding orthographic and canonical files and, in the case of spontaneous data, brings them in line with the speech files, i.e. splits them up into turn size. Such a prototype label file contains the file name (corpus and speaker references), the orthographic text and the canonical transcription corresponding to a speech signal file, as well as the list of phonemic labels, one per line, taken from the canonical transcription and associated with sample values at an imaginary time point of 100 seconds (see Table 1 and Kohler, Pätzold, and Simpson 1995).

3.2 Phonetic segmentation and labelling

Phonetic labelling proceeds from a prototype label file and attributes its canonical symbols to the time scale of the related speech file. Thus a prototype label file is converted into a label file with real time values (see Table 2). In ASL and in the initial stages of VERBMOBIL this was achieved with the MIX program mentioned in 1.3 Step 7; it has been replaced by the *xassp* tool kit (see 1.4), which is described in detail in IPDS (1997b). In both cases a prototype label file and the corresponding speech signal file are linked ('mixed') to symbolize the actual pronunciation according to the following principles (see the example in Table 1):

- One phonemic symbol after another of the prototype label file is offered by the program for manual positioning in the speech signal.
- Through visual inspection of speech wave and spectrogram displays (and other possible signal representations, e.g. F0 curves), supported by auditory control via loudspeaker or headphones, the phonemic labels are manually aligned with the beginnings of the corresponding speech signal segments and are given their time (sample) values.
- A set label marks a speech segment up to the time of the next aligned label.
- The segmentation is thus strictly linear without overlap.
- All the canonical labels are kept, and may be augmented and modified. There are 4 possibilities of processing canonical labels:

- acceptance: **S**
- replacement: **S-S'**
- deletion: **S-**
- insertion: **-S'**

- In the case of deletion the next label is aligned to the same time mark as the label for the deleted segment, which thus receives zero duration. For further details on segmentation and labelling see Kohler, Pätzold, and Simpson (1995).

```
g071a008.slh
TIS008.
```

```
ja , aber immer . dann<Z> haben wir das +/auf je ma=/+
jedenfalls mal gekl"art .
oend
```

```
j 'a: , Q a: b 6+ Q 'I m 6 . d a n+ z: h a: b @ n+
v i: 6+ d a s+ Q aU f+ j e:+ m a: =/+
j 'e: d @ n #f "a l s m a: l+ g @ k l 'E: 6 t.
kend
```

```
hend
1600000 #c:
1600000 ##j
1600000 $'a:
1600000 #,
1600000 ##Q
1600000 $a:
1600000 $b
1600000 $6+
1600000 ##Q
1600000 $'I
1600000 $m
1600000 $6
1600000 #.
1600000 #c:
1600000 ##d
1600000 $a
1600000 $n+
1600000 $z: . . .
```

Table 1: Example of a prototype label file

```

g071a008.s1h
TIS008.
ja , aber immer . dann<Z> haben wir das +/auf je ma=/+
jedenfalls mal gekl"art .
oend
j 'a: , Q a: b 6+ Q 'I m 6 . d a n+ z: h a: b @ n+
v i: 6+ d a s+ Q aU f+ j e:+ m a: =/+
j 'e: d @ n #f "a l s m a: l+ g @ k l 'E: 6 t.
kend
c: -h: j 'a: , Q -q a: b 6+ Q- -q 'I m 6 . c:
-p: %d -h a-@ n+ z: -:k -p: h a: b-m @- n-+ v i:6+
d -h a s+ Q- -q aU f+ j e:+ m a: -l =/+ -p:
j 'e: d-n @- n- #f "a l s m a: l+ g -h @- k -h l 'E:6 t -h .
hend
1249 #c:
1249 #-h:
4050 ##j
5183 $'a:
7492 #,
7492 ##Q
8236 $-q
8236 $a:
9225 $b
9955 $6+
11181 ##Q-
11181 $-q
11181 $'I
12306 $m
13528 $6
16274 #.
16274 #c:
16274 #-p:
17456 ##%d
18994 $h
19503 $a-@
20142 $n+
26619 $z:
26619 #-:k
26619 $-p: . . .

```

Table 2: Example of a label file

4 From Segment to Prosody

The linear segmental phonemic frame allows the systematic and economical representation of lexical items in canonical citation forms and the labelling of actual pronun-

ciations with reference to them. This makes it possible to search large labelled data banks very efficiently for phonemic-type modifications, such as assimilations and elisions, or strong reductions in function words ('weak forms') in connected speech. But this linear segmental phonemic approach excludes important suprasegmental aspects of two types:

- articulatory features that can no longer receive exclusive temporal delimitation as is required in an otherwise linear segmentation
- utterance prosody: prosodic phrasing, stress, intonation, speech rate, register.

4.1 Non-linear components of articulation

Canonical segments may not be discernible as such in the actual speech signal and will therefore have to be marked as deleted in labelling. But traces may still be present as componential modifications of the remaining segment strings, referable to such processes as glottalization, nasalization, palatalization, velarization etc. A few examples from the spontaneous speech corpus are to illustrate this phenomenon. In each case the orthographic, canonical and labelled symbolizations are provided together with speech wave and spectrogram displays (further details in Kohler, Pätzold, and Simpson 1995):

- Example 1 (Ref. KAE g197a011): see Figure 1

könnten **k ɹ n t @ n+**
 k -h ' ʹ ~ n- t-q @- n+

- The first nasal consonant is deleted as a sequential element, but a residue of nasalization is still manifest in the preceding vowel as a componential feature.
- The plosive **t** is realized as glottalization somewhere in the sonorant context (vowel, nasal consonant), without a precise temporal and segmental alignment.
- In both cases the articulatory components require a non-linear symbolization, i.e. markers that do not receive durations:
 - * **-~** refers to nasalization
 - * **t-q** to glottalization;
 - * both are aligned to the same point in time as the following, non-deleted segment **n**,
 - * indexing phonetic parameters in the segmentally labelled environment (further details in Kohler, Pätzold, and Simpson 1995).

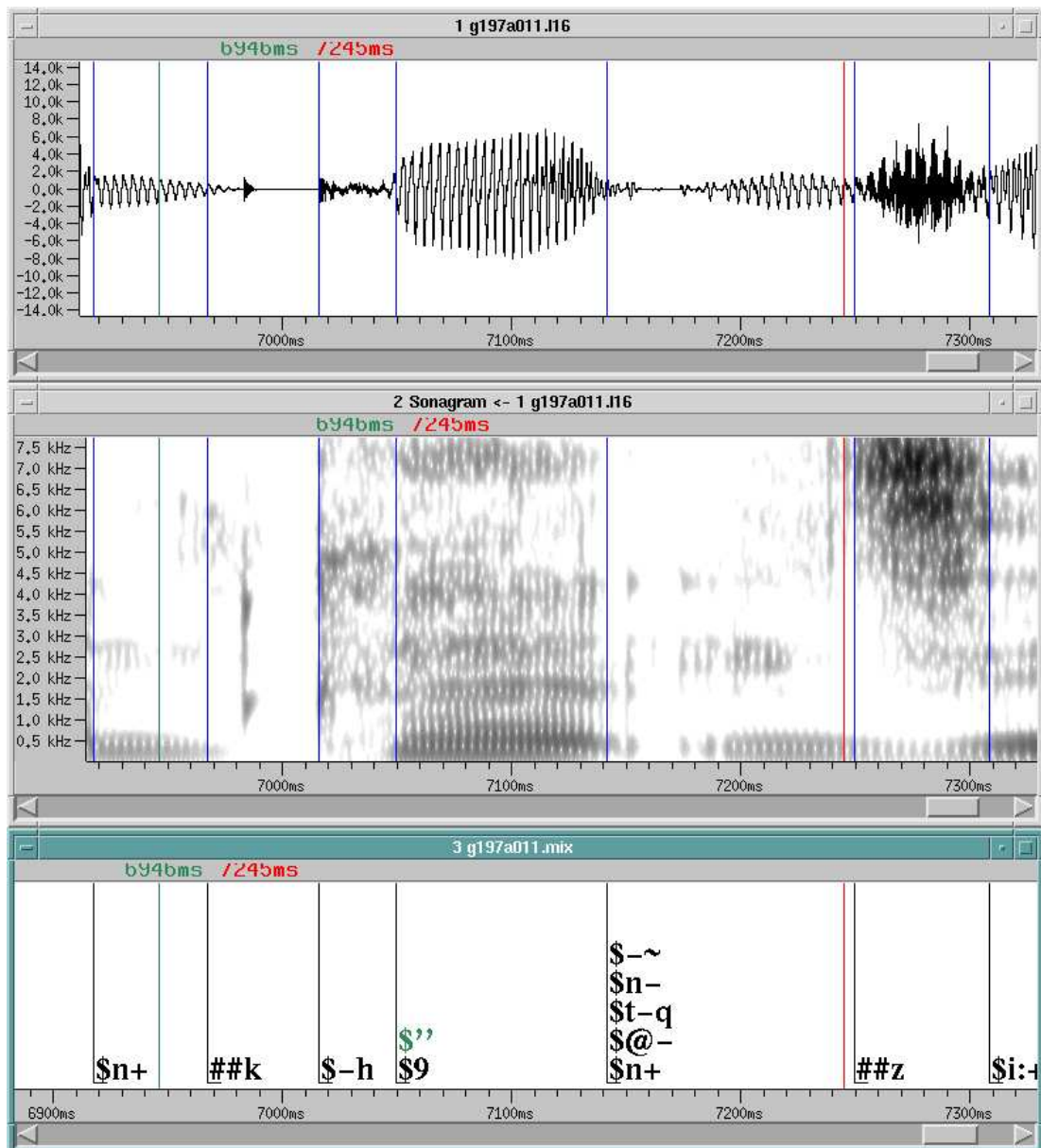


Figure 1: *xassp*— Speech wave, spectrogram and label sequence in “könnten” (Ex. 1)

In other cases the componential reflexes of deleted segments are more complex to specify phonetically and are symbolized by an inserted cover label **-MA**, i.e. a general “marker” preceding deleted symbols to indicate some phonetic residue (for details see Kohler, Pätzold, and Simpson 1995; Helgason and Kohler 1996). **-MA** again refers to some contrastive feature that distinguishes the pronunciation actually found from the one represented simply by deletions. In the signal the articulatory component is located to the left and/or the right of the marker.

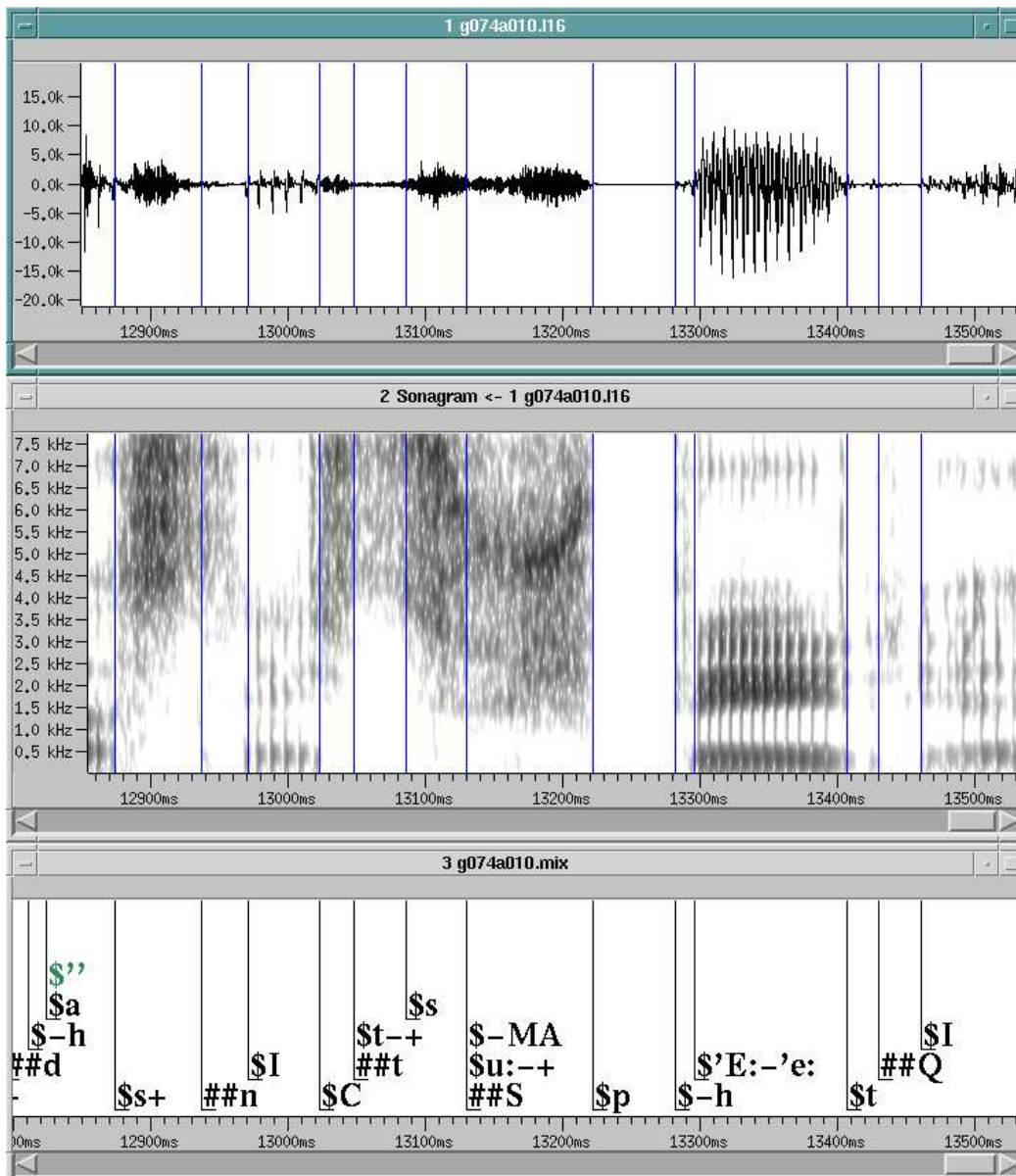


Figure 2: xassp— Speech wave, spectrogram and label sequence in “nicht zu spät” (Ex. 2)

- Example 2 (Ref. HAH g074a010): see Figure 2

nicht zu spät **n I C t+ t s u:+ S p 'E: t**
 n I C t-+ t s -MA u:-+ S p -h 'E:-'e: t

- The voiced vocalic stretch of **u:** is absent;
- its lip rounding remains as a componential residue in the surrounding fricatives.

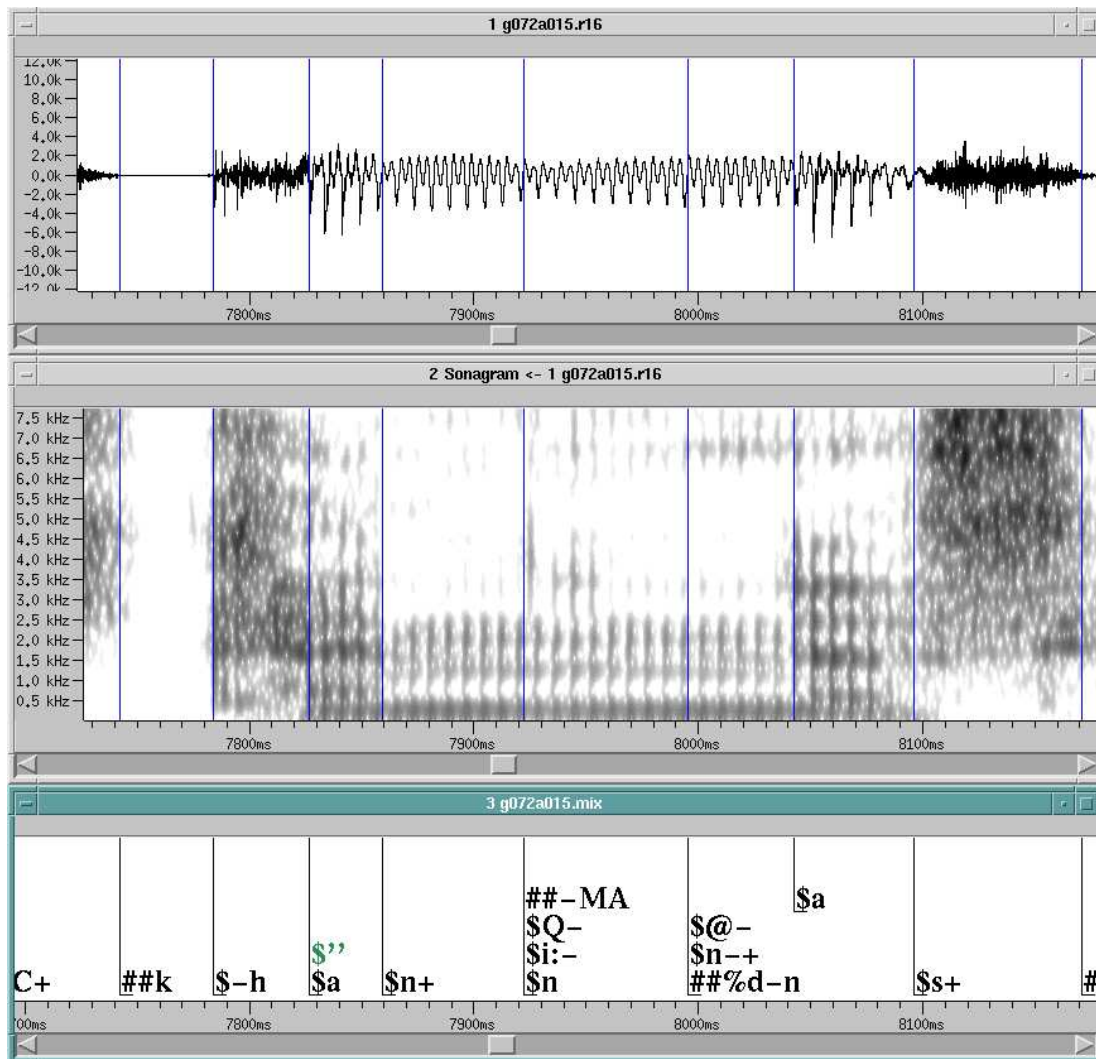


Figure 3: *xassp*— Speech wave, spectrogram and label sequence in “kann Ihnen das” (Ex. 3)

- Example 3 (Ref. TIS g072a015) see Figure 3

kann Ihnen das **k a n+ Q i: n @ n+ d a s+**
k -h ' 'a n+ -MA Q- i:- n @- n-+ %d-n a s

- The segment **i:** is deleted;
- its dorso-palatal tongue elevation remains as a componential residue of palatalization in the nasal consonants.

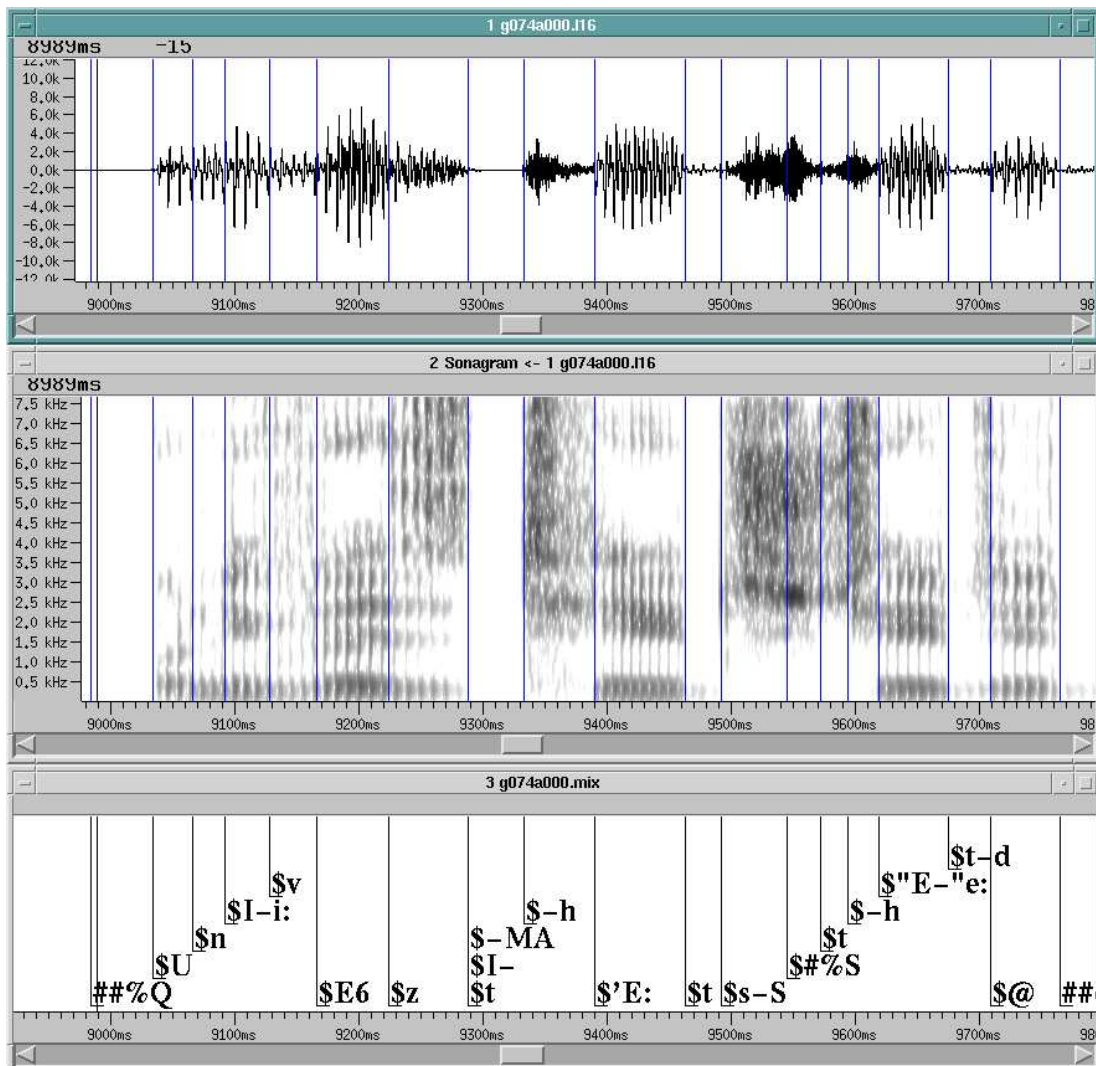


Figure 4: *xassp*— Speech wave, spectrogram and label sequence in “Universitätsstädte” (Ex. 4)

- Example 4 (Ref. HAH g074a000) see Figure 4

Universitätsstädte Q U n I v E6 z I t 'E: t s #S t "E t @

%Q U n I-i: v E6 z -MA I- t -h 'E: t s-S #%S t -h "E-"e: t-d @

- The segment **I** after the fricative **z** is deleted;
- its dorso-palatal elevation remains as a componential residue of palatization in **z**, which, moreover, keeps its voicing as in intervocalic position although it now occurs before **t**.

If in these instances only the segmental deletions were marked, without **-MA**, there would be a loss of contrastive phonological information because the signal contains

more relevant phonetic features. But the strictly linear segmental phonemic approach is not able to represent this distinctivity. So it needs supplementing with non-linear elements for an adequate phonological account of speech, and the labelling of data bases has to take this requirement into consideration. The *Kiel Corpus* is built on these principles, combining, in a ‘complementary phonology’ (Kohler 1994a), the advantages of linear segmental canonical base forms for lexical data bank searches with the need for contrastive phonetic adequacy.

4.2 Utterance prosody

Prosodic features in the traditional sense of the term, i.e. at the utterance level, need to be included in speech corpus labelling for two reasons

- as contextual frames at the segmental and componential articulatory levels
- as fields of study in their own right.

In the *Kiel Corpus*, prosodic labelling is based on an intonation model (**KIM**, see Kohler 1992c, 1996f, 1997a, 1997b) and on a symbolization system built on this model (**PROLAB**, see Kohler 1992d, 1995b, 1997a; Kohler, Pätzold, and Simpson 1995). Prosodic labels are indexed by the special marker **&** and are aligned to the speech wave on the same tier as all the other labels to make cross-references as easy and flexible as possible. The prosodic work platform is *xassp*; it makes phonetic information graphically available via linked windows, usually three for the display of the sound pressure wave, the F0 contour and the segmental labels, respectively; other sources of signal information, such as spectrogram and energy plot, may be added as well (Kohler, Pätzold, and Simpson 1995; IPDS 1997b). Parallel to these graphic displays of a speech file, acoustic output is possible through loudspeaker or headphones, either for the complete file or for any section marked in the display.

So far only part of the segmentally labelled data in the *Kiel Corpus of Read/Spontaneous Speech* have also been provided with prosodic markers. Among the read speech data on CD-ROM#1, those from the two speakers that recorded the whole corpus have also been processed prosodically. This amounts to just under 10,000 words of running text, i.e. about one third of the total *Kiel Corpus of Read Speech*. As to the spontaneous speech material, the 31 dialogues of CD-ROM#2 have likewise been completed. This is just over one quarter of the total *Kiel Corpus of Spontaneous Speech* to date. Table 3 gives an example of a prosodic label file in orthographic text (for further details see Kohler, Pätzold, and Simpson 1995; IPDS 1997b).

5 Data Bank Search and Speech Processing

The label files contain the following phonetic information about spoken texts:

g071a004.s1hTIS004:

#&2 <ähm> #&PGn #&2(D'ienstag #&0 würde+ #&0 mir+ #&0 g'ut #&0.
#&2) p'assen #, #&2. #&PGn

#&2 <ähm> #&PGn #&0 das+ #&2] h'eißt #, #&, #&PGn #p: #&2^ Mom'ent #, #&1.
#&PGn #&2('a\$z: llerdings #&0 erst\$z: #&0. #&PGn #&2(n'achm''ittags #h: #. #&2.
#&PGn

#&RP #&HP #&0 das+ #&0 wird+ #&0 dann+ #&2^ wahrsch'einlich #&0 'n+
#&0 b'ißchen #&1. #&2^ schw'ierig #. #&2. #&PGn

#&RM #&2^ D'ienstag #, #&0. #&2^ mittwochs\$z: #&1. #&PGn #&0 <äh> #&PGn
#p: #&0 is=/+ #&PG/ #&1^ s'ieht #&0 das+ #&0 bei+ #&0 m\$ir+\$z: #&0 sch=/+
#&2. #&PG/ #&2^ schw'ierig #&0 'aus #. #&2. #&PGn

#&0 da+ #&0 hab' #&0 ich+ #&2^ tags'über #&1. #&2^ Term'ine #. #&1. #&PGn

#h: #&2 <ähm> #&Pgn #&RP #&HP #&0 wie+ #&0 s'ieht #&0 das+ #&0 bei+
#&0 Ihnen+ #&0 am+ #&3^ D'onnerstag #&0 'aus #? #&2. #&PGn

Table 3: PROLAB labels in orthographic text of a dialogue turn

- corpus references: recording site, speaker, section of (read) corpus, sentence numbers and dialogue/turn numbers, respectively
- orthographic forms of words and non-lexical items
- canonical transcription transforms, including word boundaries and function word markers
- segmental and componential labels with their time marks
- time-aligned utterance prosody labels.

These symbolic label files, in conjunction with the acoustic data base of read and spontaneous speech files, enter into an annotated phonetic data bank, which, together with incorporated retrieval tools, allows corpus searches and speech processing of referenced signals for any phonetic research question in the phonological and acoustic domains (Kohler 1996e; Pätzold 1997), in particular

- the automatic generation of pronunciation lexica

ORTHOGRAPHY	CANONICAL	FREQUENCY
Anschluß	Q'an#SI''Us	4
April	Qapr'Il	11
Arbeit	Q'a6baIt	3
Arbeiten	Q'a6baIt@n	1
Arbeitsfrühstück	Q'a6baIts#fr''y:#St''Yk	5
Arbeitsfrühstücks	Q'a6baIts#fr''y:#St''Yks	1
Arbeitskreis	Q'a6baIts#kr''aIs	1
Arbeitssitzung	Q'a6baIts#z''ItsUN	2
Arbeitssitzungen	Q'a6baIts#z''ItsUN@n	2
Arbeitstreffen	Q'a6baIts#tr''Ef@n	8

Table 4: Excerpt from the canonical lexicon for CD-ROM#2 of the *Kiel Corpus of Spontaneous Speech*

- the systematic data bank search for segmental and prosodic phenomena of read/spontaneous speech (e.g., phonemes, phoneme combinations, their phonetic realizations, sound reduction, stress and intonation patterns), i.e. the generation of search files for specific phonetic aspects
- the acoustic analysis of the corresponding speech data retrieved with the help of the search files
- the formulation of rules for the phonetic realization of words in the symbol and signal domains.

5.1 The Canonical Lexicon

In the first instance a canonical pronunciation lexicon can be generated from the canonical transcription texts of the label files (see 3.1 and Table 2) for a particular corpus (of read or spontaneous speech). It lists all the word types in orthographic and in canonical phonological form, together with their frequencies of occurrence (IPDS 1994, 1995, 1996, 1997a; Kohler 1994c; Kohler, Pätzold, and Simpson 1995). Two sections are differentiated in such a lexicon: word forms and special forms (e.g. word fragments in spontaneous speech). In addition a reference list is provided that gives the corpus locations of lexical entries. Table 4 contains an excerpt from the canonical lexicon of CD-ROM#2.

5.2 The Variants Lexicon and the Variants List

Similarly, a variants lexicon can be derived for a particular data base, containing orthographic, canonical and label forms for each entry as well as frequencies of each lexical item and variant, together with the corpus references (dialogue session, turn

ORTH	CANON	VARIANT	LFR	VFR	CORPUS	REF
Arbeit	Q'a6baIt	Q-:q'a6baIt	3	1	G085A006	22
Arbeit	Q'a6baIt	Q-:q'a6baIt-:	3	1	G097A000	17
Arbeit	Q'a6baIt	Q=-z:'a6baIt-:	3	1	G085A008	13
Arbeiten	Q'a6baIt@n	Q-:q'a6baIt-q@-:n	1	1	G096A000	19

Table 5: Excerpt from the variants lexicon for CD-ROM#2 of the *Kiel Corpus of Spontaneous Speech*

and serial word number within the turn for spontaneous dialogues) (IPDS 1994, 1995, 1996, 1997a; Kohler 1994c; Kohler, Pätzold, and Simpson 1995). Table 5 presents an excerpt from the variants lexicon of CD-ROM#2.

The reduction of the complete variants lexicon to a statistically valued restricted variants list has been realised in the investigation of German function words. Besides a few base variants a narrowly delimited set of rules was devised for the generation of the more frequent phonetic variation observed empirically in these function words (Rehor 1996; Rehor and Pätzold 1996).

5.3 Data analyses

The Kiel data bank has also been used to carry out the following language and speech investigations in German:

- realization of schwa syllables (Kohler 1995a, 1996g, 1996h)
- glottal stop and glottalization (low-frequency irregular glottal pulsing) for word-initial vowel onset and as plosive realization (see 4.1) (Kohler 1994b, 1995c, 1996a, 1996c, 1996d, 1996g; Kohler and Rehor 1996)
- vowel deletion (high vowels [i:], [ɪ], [u:], [ʊ], when unstressed, especially in the environment of voiceless obstruents, frequently with an articulatory residue of palatalization or velarization or labialization (see 4.1) (Helgason and Kohler 1996; Kohler 1996a; Rodgers 1996, 1997; Rodgers et al. 1997)
- phonetic variants of function words (Dommelen 1992a; Rehor 1996; Rehor and Pätzold 1996)
- lenition (Helgason 1996)
- duration of vowels in lexical stress position (Kohler 1994d)
- acoustic analysis of the German vowel system (Pätzold and Simpson 1997)
- acoustic analysis of hesitation particles (Pätzold and Simpson 1995, 1996).

5.4 Automatic segmentation and labelling

The Kiel speech data bank of spoken German and the phonetic investigations that may be and have been carried out with it in the symbol and signal domains also make an important contribution to an at least partial automatization of manual segmentation and labelling, which are still quite time-consuming. The Munich Phonetics Institute is working on the development of such an automatic system (Kipp 1995; Wesenick 1995, 1996; Kipp and Wesenick 1996; Kipp et al. 1996; Wesenick and Schiel 1997): the ‘Munich AUtomatic Segmentation System’ (MAUS). Taking read speech as the point of departure a set of over 6000 symbolization rules for unlimited context has been extrapolated from a study of the literature and from generalized observations on a small amount of empirical transcription data (PHONRUL). These rules allow the symbolic generation of potential pronunciation variants (without statistical probabilities) from canonical word forms. A verification procedure then links the most probable transcription with the speech wave.

There is no doubt that these variants generation rules needed to be put on a more solid empirical basis to give a better chance for a successful automatic segmentation and labelling, especially of spontaneous speech. Since this empirical foundation can only be provided by manually labelled data, it is not surprising that MAUS is now also pursuing an alternative, using a set of data-driven ‘microrules’. They are derived automatically from manually segmented parts of a data base. They are thus corpus dependent and, contrary to PHONRUL, contain no a priori generalized phonetic knowledge. The size of the generated rule set is determined by the amount of labelled data used and by the pruning factor applied to exclude certain phonetic phenomena. It varies between 500 and 2000. The *Kiel Corpus of Spontaneous Speech* CD-ROMs#2,3 were used to derive about 1200 rules from 72 manually segmented dialogues. The development and the operation of this version of the automatic segmenting and labelling system MAUS thus presupposes a manually labelled data base, before a reliable phonetic label output can at all be expected, and the *Kiel Corpus* became an indispensable and valuable source.

The two versions of MAUS are opposites that oscillate between a dearth of empirical grounding and complete data dependence, between theory and data driven procedures. What is really needed is the derivation of empirically motivated, knowledge-based phonetic realization rules by applying search operations for an array of phonetic questions to the labelled data bank. This will drastically restrict the number of rules to be applied for automatic segmentation and labelling, because it will reduce the data to ‘significant points’ that may be generalized. And these rule generalizations should prove more successful.

Moreover, automatic labelling also ought to supplement the purely linear concept of the phonetic segment by the consideration of overlapping long articulatory/acoustic components, such as glottalization, nasalization etc. (articulatory prosodies), as has already been implemented in the manual labelling of the *Kiel Corpus* (see 4.1). Thus if MAUS labels a stretch of speech wave as **t@mi:l** (as in *das paßt mir terminlich*

schlecht) important non-segmental articulatory aspects of utterance are possibly being neglected, in which the nasal feature of the deleted nasal consonant **n** presumably lingers on in the nasalization of the vowel **i:** and the sonorant **l**, and where the word-final fricative **C** leaves its traces in the word-initial **S**. In such a case the Kiel labelling would insert \sim in the first, **-MA** in the second instance. An automatic transcription has to do likewise, because it is only then that the symbolized pronunciation becomes empirically plausible; **t@mi:l** is not. So in this respect manual labelling is again the indispensable foundation for its automatization, contributing through the latter to automatic speech recognition.

6 Conclusion and Outlook

The structural frame for a computer data bank of spoken German and its integration into basic and applied phonetic research has been developed at IPDS Kiel and is continually being filled with segmentally and prosodically labelled data from read and spontaneous speech recordings. In future, more diverse types of spontaneous interactions will be included in the *Kiel Corpus*. The range of phonetic analyses will be extended to include prosodic variables as independent topics of investigation and as contextual factors for segmental manifestations. Furthermore, comparison of speaking styles across a growing data base is a particularly interesting task, since our knowledge in this area is still very limited, even in well documented languages, such as German or English. The results of such comparative phonetic studies of situational speech varieties are very important for exhaustive descriptions of individual languages, as well as for practical applications, e.g. in speech technology.

Data bank analyses within such a broad frame provide statistically evaluated rules for speech processes in varying contexts of situation and speaking conditions. So they complement the traditional word-centred phonology with a long overdue phonetics of 'connected speech' (Simpson and Pätzold 1996). The scientific study of these rules will assist the development of automatic segmentation and labelling procedures, which will in return speed up and increase the extension of the labelled data base to the profit of basic research.

Finally, the Kiel phonetic data bank model can be extended to other languages to initiate large-scale multilingual phonetic investigations of comparable phonetic phenomena on the same theoretical and methodological basis (Kohler 1996b). The results of such multilingual comparisons will in turn lay the foundation for asking questions of phonetic and phonological typologies and universals at the sentence and utterance levels and for answering them competently.

References

- Carlson, R. and B. Granström (1985). Rule controlled data base search. *STL-QPSR* (4), 29–42.
- Carlson, R., B. Granström, and S. Hunnicut (1990). Multilingual text-to-speech development and applications. In W. A. Ainsworth (Ed.), *Advances in Speech, Hearing, and Language Processing*, pp. 269–296. London: JAI Press.
- Dommelen, W. v. (1992a). Die Erhebung und Verarbeitung spontansprachlichen Materials. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 97–110.
- Dommelen, W. v. (1992b). Segmentieren und Etikettieren im Kieler PHONDAT-Projekt. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 197–223.
- DUDEN (1991). *Der Duden Bd. 1, Rechtschreibung der deutschen Sprache* (20 ed.). Mannheim: Dudenverlag.
- Helgason, P. (1996). Lenition in German and Icelandic. In A. P. Simpson and M. Pätzold (Eds.), *Sound Patterns of Connected Speech: Description, Models, and Explanation*, AIPUK 31, pp. 211–218.
- Helgason, P. and K. J. Kohler (1996). Vowel deletion in the *Kiel Corpus of Spontaneous Speech*. In K. J. Kohler, C. Rehor, and A. P. Simpson (Eds.), *Sound Patterns in Spontaneous Speech*, AIPUK 30, pp. 115–157.
- Hess, W., K. J. Kohler, and H.-G. Tillmann (1995). The PHONDAT-VERBMOBIL speech corpus. In *Proc. of the 5th European Conference of Speech Communication and Technology*, Volume 1, Madrid, pp. 863–866.
- IPDS (1994). *The Kiel Corpus of Read Speech*, Volume 1, CD-ROM#1. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1995). *The Kiel Corpus of Spontaneous Speech*, Volume 1, CD-ROM#2. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1996). *The Kiel Corpus of Spontaneous Speech*, Volume 2, CD-ROM#3. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1997a). *The Kiel Corpus of Spontaneous Speech*, Volume 3, CD-ROM#4. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1997b). *xassp User's Manual* (Advanced Speech Signal Processor under the X Window System). In A. P. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech — Acoustic data base, processing tools and analysis results*, AIPUK 32, pp. 31–115.

- Karger, R. and W. Wahlster (1995). *Verbmobil Handbuch – Version 3*. Verbmobil Technisches Dokument Nr. 35. Saarbrücken: DFKI.
- Kipp, A. (1995). *Automatisches Segmentations- und Etikettiersystem*. Verbmobil Memo Nr. 95.
- Kipp, A. and M.-B. Wesenick (1996). Estimating the quality of phonetic transcriptions and segmentations of speech signals. In *Proc. ICSLP96*, Volume 1, Philadelphia, pp. 129–132.
- Kipp, A., M.-B. Wesenick, and F. Schiel (1996). Automatic detection and segmentation of pronunciation variants in German speech corpora. In *Proc. ICSLP96*, Volume 1, Philadelphia, pp. 106–109.
- Kohler, K. J. (1992a). Automatische Generierung der kanonischen Transkription und des Aussprachelexikons. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 175–196.
- Kohler, K. J. (1992b). Erstellen eines Textkorpus für eine phonetische Datenbank des Deutschen. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 11–39.
- Kohler, K. J. (1992c). A model of German intonation. In K. J. Kohler (Ed.), *Studies in German intonation*, AIPUK 25, pp. 295–360.
- Kohler, K. J. (1992d). Prosodisches Transkriptionssystem für die Etikettierung von Sprachsignalen. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 239–250.
- Kohler, K. J. (1992e). Sprachverarbeitung im Kieler PHONDAT-Projekt: Phonetische Grundlagen für ASL-Anwendungen. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 81–95.
- Kohler, K. J. (1994a). Complementary phonology: a theoretical frame for labelling an acoustic data base of dialogues. In *Proc. ICSLP94*, Volume 1, Yokohama, pp. 427–430.
- Kohler, K. J. (1994b). Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica* 51, 38–51.
- Kohler, K. J. (1994c). *Lexika of the Kiel PHONDAT Corpus Vol. I & II*. AIPUK 27 & 28.

- Kohler, K. J. (1994d). Stressed vowel duration in German. *Acta Linguistica Hafnien-sia* 27, 299–321.
- Kohler, K. J. (1995a). Articulatory reduction in different speaking styles. In *Proc. XIIIth ICPPhS*, Volume 2, Stockholm, pp. 12–19.
- Kohler, K. J. (1995b). PROLAB - the Kiel system of prosodic labelling. In *Proc. XIIIth ICPPhS*, Volume 3, Stockholm, pp. 162–165.
- Kohler, K. J. (1995c). The realization of plosives in nasal/lateral environments in spontaneous speech in German. In *Proc. XIIIth ICPPhS*, Volume 2, Stockholm, pp. 210–213.
- Kohler, K. J. (1996a). Articulatory reduction in German spontaneous speech. In *Proc. 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from control strategies to acoustics*, Autrans, pp. 1–4.
- Kohler, K. J. (1996b). Developing a research paradigm for sound patterns of connected speech in the languages of the world. In A. P. Simpson and M. Pätzold (Eds.), *Sound Patterns of Connected Speech: Description, Models, and Explanation*, AIPUK 31, pp. 227–233.
- Kohler, K. J. (1996c). Glottal stop and glottalization - A prosody in European languages. In K. J. Kohler, C. Rehor, and A. P. Simpson (Eds.), *Sound Patterns in Spontaneous Speech*, AIPUK 30, pp. 207–216.
- Kohler, K. J. (1996d). Glottalization across languages. In A. P. Simpson and M. Pätzold (Eds.), *Sound Patterns of Connected Speech: Description, Models, and Explanation*, AIPUK 31, pp. 207–210.
- Kohler, K. J. (1996e). Labelled data bank of spoken standard German - The *Kiel Corpus of Spontaneous Speech*. In *Proc. ICSLP96*, Volume 3, Philadelphia, pp. 1938–1941.
- Kohler, K. J. (1996f). Modellgesteuerte Prosodiegenerierung - Die Implementation des Kieler Intonationsmodells (KIM) in der TTS-Synthese für das Deutsche. In *Fortschritte der Akustik – DAGA96*, Bonn, pp. 90–91.
- Kohler, K. J. (1996g). Phonetic realization of German /ə/-syllables. In K. J. Kohler, C. Rehor, and A. P. Simpson (Eds.), *Sound Patterns in Spontaneous Speech*, AIPUK 30, pp. 159–194.
- Kohler, K. J. (1996h). The phonetic realization of /ə/ syllables in German. In A. P. Simpson and M. Pätzold (Eds.), *Sound Patterns of Connected Speech: Description, Models, and Explanation*, AIPUK 31, pp. 11–14.

- Kohler, K. J. (1997a). Modelling prosody in spontaneous speech. In Y. Sagisaka, N. Campbell, and N. Higuchi (Eds.), *Computing Prosody*, pp. 187–210. Berlin/Heidelberg/New York/Tokyo: Springer.
- Kohler, K. J. (1997b). Parametric control of prosodic variables by symbolic input in TTS synthesis. In J. P. H. v. Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (Eds.), *Progress in Speech Synthesis*, pp. 459–475. Berlin/Heidelberg/New York/Tokyo: Springer.
- Kohler, K. J., G. Lex, M. Pätzold, M. T. M. Scheffers, A. P. Simpson, and W. Thon (1994). *Handbuch zur Datenaufnahme und Transliteration in TP14 von Verbmobil - 3.0*. Verbmobil Technisches Dokument Nr. 11.
- Kohler, K. J., M. Pätzold, and A. P. Simpson (1994). *Handbuch zur Segmentation und Etikettierung von Spontansprache - 2.3*. Verbmobil Technisches Dokument Nr. 16.
- Kohler, K. J., M. Pätzold, and A. P. Simpson (1995). *From scenario to segment: the controlled elicitation, transcription, segmentation and labelling of spontaneous speech*. AIPUK 29.
- Kohler, K. J. and C. Rehor (1996). Glottalization across word and syllable boundaries. In K. J. Kohler, C. Rehor, and A. P. Simpson (Eds.), *Sound Patterns in Spontaneous Speech*, AIPUK 30, pp. 195–206.
- Pätzold, M. (1997). *KielDat - data bank utilities for the Kiel Corpus*. In A. P. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech - Acoustic data base, processing tools and analysis results*, AIPUK 32, pp. 117–126.
- Pätzold, M., M. Scheffers, A. P. Simpson, and W. Thon (1995). Controlled elicitation and processing of spontaneous speech in Verbmobil. In *Proc. XIIIth ICPHS*, Volume 3, Stockholm, pp. 314–317.
- Pätzold, M. and A. P. Simpson (1994). *Das Kieler Szenario zur Terminsprache*. Verbmobil Memo Nr. 53.
- Pätzold, M. and A. P. Simpson (1995). An acoustic analysis of hesitation particles in German. In *Proc. XIIIth ICPHS*, Volume 3, Stockholm, pp. 512–515.
- Pätzold, M. and A. P. Simpson (1996). *Phonetik und Phonologie von Häsitationspartikeln*. In *27. Jahrestagung der GAL*, Erfurt.
- Pätzold, M. and A. P. Simpson (1997). Acoustic analysis of German vowels in read speech. In A. P. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech - Acoustic data base, processing tools and analysis results*, AIPUK 32, pp. 215–247.

Rehor, C. (1996). Phonetische Realisierung von Funktionswörtern im Deutschen. In K. J. Kohler, C. Rehor, and A. P. Simpson (Eds.), *Sound Patterns in Spontaneous Speech*, AIPUK 30, pp. 1–113.

Rehor, C. and M. Pätzold (1996). The phonetic realization of function words in German spontaneous speech. In A. P. Simpson and M. Pätzold (Eds.), *Sound Patterns of Connected Speech: Description, Models, and Explanation*, AIPUK 31, pp. 5–10.

Rodgers, J. E. J. (1996). Vowel deletion/devoicing. In A. P. Simpson and M. Pätzold (Eds.), *Sound Patterns of Connected Speech: Description, Models, and Explanation*, AIPUK 31, pp. 211–218.

Rodgers, J. E. J. (1997). A comparison of vowel devoicing/deletion phenomena in English laboratory speech and German spontaneous speech. In A. P. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech — Acoustic data base, processing tools and analysis results*, AIPUK 32, pp. 197–214.

Rodgers, J. E. J., P. Helgason, and K. J. Kohler (1997). Segment deletion in the *Kiel Corpus of Spontaneous Speech*. In A. P. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech — Acoustic data base, processing tools and analysis results*, AIPUK 32, pp. 127–176.

Simpson, A. P. and M. Pätzold (Eds.) (1996). *Sound Patterns of Connected Speech*. AIPUK 31.

Thon, W. (1992). Struktur eines Datenverarbeitungssystems für das Kieler PHONDAT-Projekt: Von der Aufnahme ASL-PHONDAT 92 zur Datenanalyse. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 111–173.

Thon, W. and W. v. Dommelen (1992). PHONDAT90: Rechnerverarbeitbare Sprachaufnahmen eines umfangreichen Korpus des Deutschen. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 41–79.

Wells, J. C., W. J. Barry, and A. J. Fourcin (1989). Transcription, labelling and reference. In A. J. Fourcin, G. Harland, W. J. Barry, and V. Hazan (Eds.), *Speech Technology Assessment. Towards Standards and Methods for the EUROPEAN COMMUNITY*, pp. 141–159. Chichester: Ellis Horwood.

Wesenick, M.-B. (1995). *Regelsystem zur Generierung von Aussprachevarianten*. Verbmobil Memo Nr. 96.

Wesenick, M.-B. (1996). Automatic generation of German pronunciation variants. In *Proc. ICSLP96*, Volume 1, Philadelphia, pp. 125–128.

Wesenick, M.-B. and F. Schiel (1997). Pronunciation modeling applied to automatic segmentation of spontaneous speech. In *Proc. of the 6th European Conference of Speech Communication and Technology*, Rhodes.