# RVG 1 - A Database for Regional Variants of Contemporary German

**Susanne Burger & Florian Schiel**
Institute of Phonetics and Speech Communication
University of Munich
80799 Munich, GERMANY
[burger@phonetik.uni-muenchen.de]

**Abstract**

Regional speaker variability is a major problem in today's state-of-the-art speech recognition systems. Therefore, a major point in the creation of speech resources is the regional coverage of data within one language. At the beginning of 1996 we started to collect data for the RVG1 (Regional Variants of German) corpus. This project was established in cooperation between the American telephone company AT&T, Lucent Technologies and the Bavarian Archive for Speech Signals – BAS, Munich (Schiel, 1997). It can be seen as a first small database of regionally covered recordings of German representing the most common dialectal regions or at least all those regions which could be important for categorising regional variants into broader classes. RVG1 contains read numbers, phonetically rich sentences and computer commands as well as spontaneous speech. The speech signals are recorded in parallel in four different sound qualities. This paper documents the tasks, the criteria for speaker selection, the recording procedure, the technical recording set-up and the labelling procedure. A special issue will be the labelling of pronunciation variants on the orthographic level which is of great importance for the first analysis of dialectal or regional variants.

## Introduction

Recent years have shown a great need for large, segmented and labelled speech databases for speech synthesis, speech recognition and phonetic research. Special attention has been directed to databases containing collections of spontaneous speech. Another and partly new aspect of these spontaneous speech databases is the concentration on corpora with regional coverage within one language. These recordings allow a more detailed analysis of regional variants of the standard language and can help in developing individual ‘submodels’ within language models. This may have a favourable effect on speech recognition and improve the performance and acceptance of speech synthesis. Additionally, such a database of regional variants can be used for analysis of changes in the usage of dialect and standard language over time.
The RVG 1 data collection is a pilot database for the task of collecting regional variants of German. We will use this database to verify whether the clustering is appropriate for an even coverage of a fixed amount of speakers. We hope to establish an empirically based clustering of the area in Europe where the Standard German is spoken. The result should be an appropriate scheme for further recordings in the same way but with a lot more speakers than the 500 of the pilot study and appropriate tools and structures to speed up the recordings.

## Tasks of the Project

The primary task of the project was,
- to collect 500 speakers of German (including Swiss, Austrian and northern Italian speakers)
- to select speakers according to demographic density
- to collect read as well as spontaneous speech
- to record with a range of different qualities of recording technologies
- to collect detailed information on the speakers in an additional database
- to annotate and verify the recorded speech signals

## Criteria for Speaker Selection

With regard to the main task of collecting current spoken German we determined by means of population density how many speakers of each German-speaking region to record. Thus, more speech data is collected from conurbations than from regions with sparse population.
For the distribution of the 500 speakers over the German speaking regions we decided to separate these regions in a kind of grid.
However, there exist different possibilities to cluster into regions:
1. clustering into ‘Bundesländer’/’Kantone’/’Provinzen’ (political entities, states)
2. clustering into dialectal regions
3. clustering into pieces with similar numbers of inhabitants
4. clustering into pieces of equal geographical extension

Regarding case 1 there is the problem that some dialectal variants extend over ‘Bundesland’ borders and some German ‘Bundesländer’ show more than one strong variant.
Case 2 seems to be a practical solution, but on the other hand we are searching for regional variants of standard German, not for dialects. Additionally, the maps for the subdivision of dialects found in the literature are quite a bit out-of-date for our task: Some dialects aren’t spoken anymore, some changed or merged with others. The borders between dialectal regions had shifted over time. Also, there exist very small dialect regions only spoken by a few hundred people.
Case 3 and 4 show similar problems to case 1: Some regions might not be recorded in favour of regions with more inhabitants.

Finally, a clustering that was closely based to the dialectal subdivision introduced by König (1978) was chosen for RVG 1. All dialects not spoken anymore were deleted, regions with small numbers of inhabitants or very small dialect regions were assigned to adjacent regions, while larger regions were separated into subclasses of dialects and regions with very high numbers of inhabitants were separated into smaller pieces. Borders of these clusters are always aligned to 'Landkreis' (county), 'Bundesland' or 'Kanton' (state) borders to simplify the speaker clustering. For each cluster we calculated the number of inhabitants and distributed these numbers over the total number of 500 speakers.
Table 1 shows the identifications of the used clusters, the names of the resulting 9 main dialect regions, the 36 distinct regions and the percentage of the 500 speakers per region.

| **RVG-cluster** | **name of dialect region** | **Inhabitants/ percent** |
|---|---|---|
| *A* | *1.Niederfränkisch* | |
| A2 | Niederrheinisch | 7,95 |
| *B* | *2.Westniederdeutsch* | |
| B1 | Schleswigisch | 0,98 |
| B2 | Holsteinisch | 3,67 |
| B3 | Nordniedersächsisch | 4,31 |
| B4 | Westfälisch | 4,77 |
| B5 | Ostfälisch | 4,22 |
| *C* | *3.Ostniederdeutsch* | |
| C1 | Mecklenburgisch | 1,93 |
| C2 | Märkisch,Nordmärkisch Mittelmärkisch, Südmärkisch | 1,87 |
| C3 | Brandenburgisch | 3,66 |
| *D* | *4.Westmitteldeutsch* | |
| D1 | Mittelfränkisch | 2,43 |
| D2 | Moselfränkisch | 1,11 |
| D3 | Rheinfränkisch | 1,14 |
| D4 | Hessisch | 6,30 |
| D5 | Pfälzisch | 2,70 |
| D6 | Ripuarisch | 4,98 |
| *E* | *5.Ostmitteldeutsch* | |
| E1 | Thüringisch | 3,64 |
| E2 | Obersächsisch | 7,86 |
| *F* | *6.Alemannisch* | |
| F2 | Niederalemannisch | 2,31 |
| F3 | Hochalemannisch | 2,84 |
| F4 | Höchstalemannisch | 1,90 |
| F5 | Schwäbisch | 5,62 |
| *G* | *7.Ostfränkisch* | |
| G1 | Ostfränkisch | 5,21 |
| *H* | *8.Südfränkisch* | |
| H1 | Südfränkisch | 2,86 |
| *I* | *9.Bairisch-Österreichisch* | |
| I1 | Nordbairisch | 1,53 |
| I2 | Mittelbairisch Nordösterreichisch | 3,83 |
| I3 | Südbairisch Südösterreichisch | 8,42 |
| I4 | Tirolisch | 1,95 |

Table 1: Speaker distribution in percent

The map in figure 1 shows the clusters distributed over the German speaking regions (edged with fat lines) and the German 'Bundesländer', Austria and German Switzerland in the shaded fields .

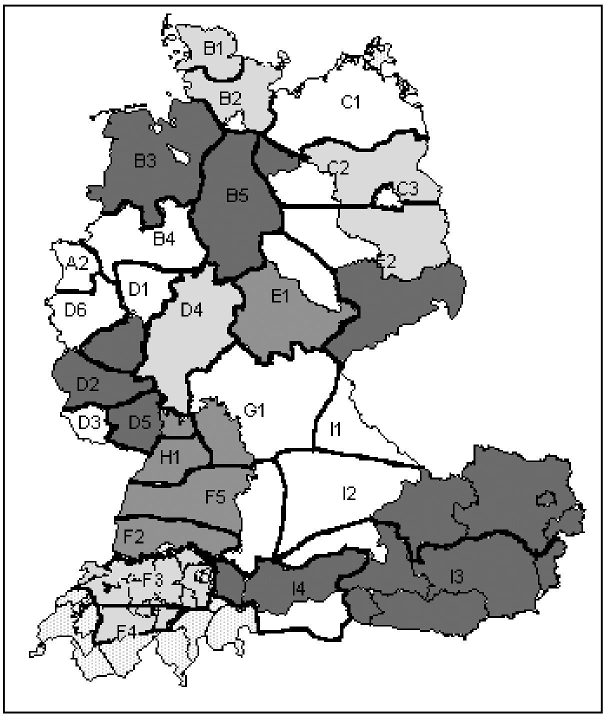


Figure 1: Map of German speaking regions; the shadowed fields show the Bundesländer/States; the fields edged with fat boarders show the clusters

## Recording Procedure

### Recorded utterances

The corpus consists of single digits, connected digits, phone numbers, phonetically balanced sentences, computer command phrases and spontaneous speech. Each speaker read a sub-corpus of 85 items.
Table 2 shows the different prompt classes as well as the number of selected prompts per speaker. The selection of prompts was not random, but an ordered scheme to provide uniform distribution of prompts over speakers. There was no controlled relationship between dialect region and selected prompts, because the dialectal classification was done after the recording. The 1 minute spontaneous speech was prompted by the invitation to speak about the work of the last week. Here the speakers were asked to imagine talking to a person from the speakers home region.

| Utterance ID | selected prompts | prompts in total | description |
|---|---|---|---|
| iso | 11 | 11 | single digits (0 - 9 + 'zwo') |
| is2 | 19 | 19 | 11-19, 20 - 100 (in steps of ten) |
| phr | 12 | 32 | computer command phrases |
| st1 | 30 | 398 | phonetically balanced sentences |
| t6l | 5 | 20 | 6-digit phone numbers |
| t7l | 5 | 20 | 7-digit phone numbers |
| std | 2 | 20 | phone numbers with area code |
| sp1(sp2) | | | 1 minute spontaneous speech |

Table 2: Prompted utterance classes

### Recording situation

The speaker is placed in front of a standard IBM-compatible PC in normal office environment. The background noise is limited to the usual noise in office environment, e.g. door slam, background cross talk, phone ringing, paper rustle, PC noise, etc. The head of the speaker is in a range between 50 cm – 1 m to the screen, 30 cm – 60 cm from the desktop microphones. The speaker is not forced into a special position.
The speaker wears a Sennheiser HD 410 (microphone 1 inch to the left and 1 inch down from left mouth corner) and is free to use the keyboard or the mouse in front of him/her.
The placement of the three desktop microphones Sennheiser MD 441 U, Telex (Soundblaster) and Talk Back (AT&T) is as follows:

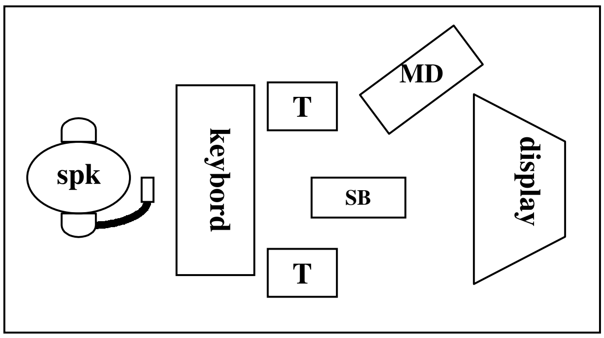


Figure 2: spk=speaker with headset, T=Talkback, SB=Soundblaster microphone, MD=Sennheiser microphone

### Signal file format

The resolution of the speech data is 16 Bit; the byte order is HiLo (Motorola). The sampling frequency is 22.050 kHz (except for the speakers 001 to 036 which were recorded at 11.025 kHz). Each microphone channel is stored into a separate file.

## Speaker database

Speakers were questioned extensively about their regional background and additional issues of general interest, like age or weight. These data are recorded in the speaker database and support the classification of the dialect or regional pronunciation of a speaker.
Table 3 shows the data a speaker is asked about after a recording session.

| Nr | Description |
|---|---|
| 1 | age of speaker |
| 2 | sex (m = male, w = female) |
| 3 | size in cm |
| 4 | weight in kg |
| 5 | place of residence during the first years of elementary school |
| 6 | place of residence during the longest period of life |
| 7 | dialect relationship to parents; opinion of speaker (dE : same dialect as both parents<br>dM : same dialect as mother<br>dV : same dialect as father<br>dA : other dialect than both parents) |
| 8 | place of origin mother |
| 9 | place of origin father |
| 10 | educational level |
| 11 | profession |
| 12 | self-defined dialect |

Table 3: Speaker data

## Verification and Annotation

### Verification of the read speech

All read speech signals were checked for their quality. Mis-readings or special pronunciation variants were recorded as well as background noise or any interruptions by means of VERBMOBIL-type markers (Burger, 1997). The validation was carried out by trained German students of phonetics. The technique was a WWW-based system that allowed simultaneous access via the network (Draxler, 1997). The acoustic judgements were done by listening only to one channel of the recorded microphones (AT&T Talk Back). As result of the verification, each recorded utterance is accompanied by a validation file which stores information about the validation process. This file contains the name of the validation person, the platform of the validation (Mac, Linux), the speaker identification number and the utterance identification. If the speaker inserted or changed words from the prompt, these changes are recorded accordingly. If there were hesitations in the utterance, they are shown in the text as follows:
*<aeh>* pure vowel hesitation
*<aehm>* vowel + nasal hesitation
*<hm>* nasal hesitation
*<hes>* other filled pauses
If the speaker did a word break, the word is completed in the text (to avoid non-words). Same is valid for technically caused interruptions of words. In the latter

case one of the buttons 'frontcut' or 'endcut' must be marked (see below).
Special characters:
'+/.../+' denotes a phrase which is repeated or corrected afterwards
'-/.../-' denotes a sentence break, that is a break not caused by a technical break and not followed by a repetition or correction.
Via 28 buttons the validation person marks special noises like rustle, door slam or phone ring, variations between the prompted text and the actually spoken utterance like slips, dialectal variants or repetitions/corrections, unusual voice qualities like hoarse voice or throat clearing or technical disturbances. If a button is pressed, the positional entry in the validation file is a keyword; if the button is not pressed the entry is '-'.

Example of a validation file:

*dieter linux 071 phr00023 +/Rechner/+ Rechner, aufpassen!*
*- - knock - - - - - - - -*
*- - - - - - repair - - -*
*- - - - - frontcut -*

In the case of the example above, the validation person ‘dieter’ working on a Linux platform annotated for the computer phrase number 23 spoken from speaker 71 a repetition in the prompted text. He pressed the button ‘knock’ for a background noise, the button ‘repair’ for the repetition and the button ‘frontcut’ to indicate that the signal file was cut at the beginning of the utterance.

## Annotation of the spontaneous speech

The spontaneous monologues were transliterated on an orthographic level together with special symbols for typical spontaneous phenomena like lengthening, hesitations, non-grammatical phrases and background noise. The transliteration conventions follow the standard for transliteration of spontaneous speech as defined in VERBMOBIL (Burger, 1997). Special attention was directed to the annotation of pronunciation variants. Variants appears with a higher frequency than in the VERBMOBIL corpus; accordingly, the rules for pronunciation comments as defined for the VERBMOBIL transliteration (Burger, Kachelrieß, 1996) had to be improved for this task. The transliteration conventions give only a restricted information about regional variants; because of the limitation of the orthographic alphabet in a lot of cases it is not possible to describe variants of speech sounds accordingly. But the rules for pronunciation comments allow a first classification of variants which may give a broad overview over the differences between regions and lists utterances which are worth of further analysis.
The rules of transliteration of pronunciation comments are as follows:
Dialectal pronunciations are transliterated in standard language, e.g. in German according to Duden (1991). After that the validation person tries to give a written version of the diverging pronunciation in comment brackets. This is done by remaining as orthographic as possible and marking elisions with apostrophes. By doing this a first idea of what the divergence looks like is presented and an indication for further analysis is given where these phenomena can be found.
The comment on pronunciation is separated from the commented element by one white space. The actual comment has the prefix ‘<!’, a number which indicates how many elements are commented followed by another white space. The end of a comment is marked by the closing bracket ‘>’.
When in doubt orthographic conventions (also any rules on capitalisation) are exercised within the parenthesis, as long as the actual pronunciation produced by the speaker is not affected. On the other hand, orthography may be used to demonstrate certain divergences (e.g. in German a long /i:/ instead of a short /i/ may be transliterated as "ie").
Example: *gewinnen <!1 gewiennen>*
Sound elisions are indicated with an apostrophe. If more than only one sound represented by one letter is not uttered, only one apostrophe is applied as marker. In case of missing final sounds of the first word and initial sounds of the following, only one apostrophe without any further white spaces shows the position of the elision.
Example: *und dann <!2 un'a'>*
If one or several sounds are substituted in a variant or sounds are added, then the transliteration will try to represent the diverging pronunciation without using any apostrophes. The variant will be transliterated (in the comment) as if it were a new word.
Example: *Donnerstag <!1 Donnaschag>*
Enclitics and strongly merged words should be transliterated as one word, but the number at the beginning of the comment always indicates the concerned number of lexical units.
Example: *haben wir <!2 hamma>*

# Current State of the RVG1 Corpus

Currently the entire corpus consists of 16 ISO 9660 CD-ROMs, each volume containing about 600 - 650 MB of uncompressed data. The total of recorded speakers is 533 on DAT, 491 of these recording sessions are also available as PC WAV signals together with the accompanying evaluation and transliteration files. The recording process will be finished at the end of April 1998.

Analytical results of the progress are presented below:

- Evaluation:
  92% of the recorded data have been evaluated, the remaining 8% aren’t recorded on PC for the following reasons:
  a. not all items have been read or the spontaneous monologue is missing
  b. too much noise or electrical hum present in the recordings (most of the time caused by using a laptop as recording medium)
  c. speaker performed a dialect which is obviously not his/her own
- Sex:
  57% of the recorded and validated speakers are male, 43% female.
- Age:
  Table 4 shows the distribution of age in 10 bins.

| age | distribution/percent |
|---|---|
| 9 – 20 | 8% |
| 21 – 25 | 37% |
| 26 – 30 | 29% |
| 31 – 35 | 11% |
| 36 – 40 | 4% |
| 41 – 45 | 2% |
| 46 – 50 | 2% |
| 50 – 55 | 3% |
| 56- 60 | 2% |
| Over 60 | 2% |

Table 4: Distribution of age in percent

- Educational Level:
  Table 5 shows the distribution of educational level across the recorded speakers.

| educational level | percent/speaker |
|---|---|
| Abitur (high school) | 86% |
| Fachabitur | 3% |
| Mittlere Reife | 7% |
| Hauptschulabschluß | 2% |
| Volksschule (elementary school) | 2% |

Table 5: Educational level in percent

The distribution shows that 89% of the recorded speakers finished their education with the high school degree (‘Abitur/Fachabitur’). This is caused by the recruiting of speakers which in most cases took place at universities or scientific sites. This also led to a high percentage of students, scientists and professors (a total of 63%) looking at the professions of the speakers.

- Self-defined dialect:
  Another interesting result is the self assessment of dialect. 11% described their language as High German, 9% described their accent as either northern or southern High German, 80% of the speakers characterised their accent as a dialect.

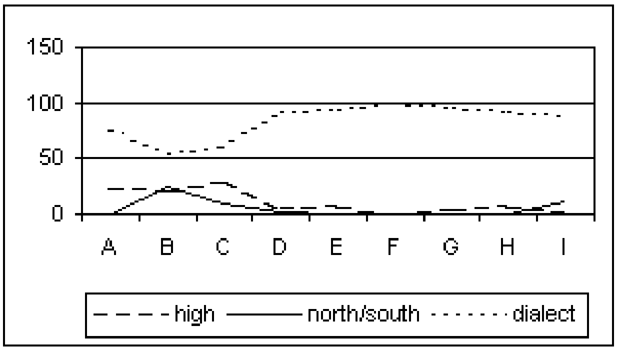


Figure 3: Dialectal self assessment, distribution over the main regions

Figure 3 shows the distribution of the speakers assessing themselves as speaking High German (high), north or south accent speaking (north/south) and speaking a dialect (dialect). Speakers of regions B and C, both in the north of Germany describe themselves to a larger extend as High German speaking while most of the speakers of the regions D, E, F, G, H and I characterise themselves as dialect speakers. Region A is situated in the north-west of Germany and again shows a higher self assessment to speak dialect.

## Current Research and Future Work

As a first experiment with the RVG1 database, we are using the recorded digits to repeat an experiment we presented at EUROSPEECH 97 in Rhode (Draxler, Burger, 1997), but now with data of a higher quality and a precise dialect determination (the first version of the experiment was done with telephone speech and without certain knowledge of the speakers' origins) (Burger, Draxler, 1998). In this experiment the test persons have to determine from where a speaker originates only by listening to the recorded digits. A second analysis will show whether the information of the pronunciation comments fulfils the requirements for assigning the data to certain regions. Furthermore, we are planing to automatically segment the data into phonemic units and label the entire material prosodically.

## References


Burger, S. & Kachelrieß, E. (1996). Aussprachevarianten in der Verbmobil-Transliteration - Regeln zur konsistenteren Verschriftung. München.*Verbmobil Memo-111-96*.

Burger, S. (1997). Transliteration spontansprachlicher Daten - Lexikon der Transliterationskonventionen - VERBMOBIL II. München. *Verbmobil TechDok-56-97*

Burger, S. & Draxler, Ch. (1998). Identifying Dialects of German from Digit Strings. *Proc. of LREC 1998.* Granada

König, W. (1978). *Dtv-Atlas zur deutschen Sprache.* München: Dtv-Verlag.

Draxler, Ch. (1997). WWWTranscribe – A Modular Transcription System Based on the World Wide Web. *Proc. of Eurospeech 1997.* (pp. 1691--1694). Rhodos.

Draxler, Ch. & Burger, S. (1997). Identification of Regional Variants of High German from Digit Sequences in German Telephone Speech. *Eurospeech 1997*. (pp. 747--750). Rhodos

*Duden - Rechtschreibung der deutschen Sprache* (1991). 20., neu bearb. und erw. Aufl. Mannheim, Wien, Zürich: *Dudenverlag*.

Schiel, F. (1997). The Bavarian Archive for Speech Signals: Resources for the Speech Community. *Eurospeech ’97*. (pp. 1687--1690). Rhodos