

# Description of the German SpeechDat Telephone Speech Data Collection

Christoph Draxler  
Institute of Phonetics and Speech Communication  
University of Munich  
Schellingstr. 3  
D 80799 Munich  
+49 +89 2866 9968  
draxler@phonetik.uni-muenchen.de

26. April 1996

## 1 Working Specifications

### 1.1 Recording Specifications

The recording platform used is a 486 DX-2 66 MHz Intel PC with 32 MB main memory, an internal 1 GB hard disk and an external 4 GB hard disk running SCO UNIX (required by the NMS board).

The ISDN primary rate interface is an Aculab ISDN board, and the audio processing is performed on an AG-30 Voice Board of NMS (Natural Micro Systems). This board contains six TMS 320C51 DSPs of Texas Instruments.

- 8 kHz Sampling rate
- Data recorded with digital line (ISDN), resulting in alaw format (8 bit).
- Header/file formats: SAM.
- Compression with `gzip -S "Z"` (under UNIX, the option `-S` allows customizing the file suffix).
- File names are 8 characters plus 3 characters extension and follow the specifications in SpeechDat deliverable D.1.4.1.

DDNMNNTT.CCX

where:

DD : Database identifier, e.g. **A0** for the fixed telephone database

NNNN : Session number (0000..9999)

TT : Type of utterance, e.g. **S1** for the first read phonetically rich sentence

CC : country code, e.g. **DE** for Ger

- [S] for speaker noise, e.g. breathing, swallowing, coughing, and [G] for non-speaker noise, e.g. telephone ringing, dial-tones, background noise, cross-talk. The transcribers were instructed to set these markers only if the speech signal was severely impaired.
- A full stop . denotes a longer silence (approx. .3s) within an utterance.
- <no speech> indicates that there is a recording file, but no audible speech.
- Hesitations were denoted by [äh], [ähm], [hm], if the speech signal contained some indication of some vowel, a vowel and a nasal, or only a nasal respectively, and the generic form [häs] in all other cases. Note that by these conventions, the utterances "öhänd äh", and "äh" are all transcribed by the same marker [äh].
- The asterisk "\*" is used to indicate unusual pronunciation if it is placed at the beginning of a word, and an incomplete word if it is placed immediately after the word fragment. Note that due to the word composition possibilities in German, word fragments very often could also be words on their own, e.g. *acht* is a word on its own, but also the beginning of *achtzehn*, *achtundzwanzig*, etc. It was marked with an asterisk whenever it was clear from the signal that the utterance had not been completed, e.g. due to some repair by the speaker, e.g. *acht\* neunzehn*.
- A double asterisk "\*\*" is used to denote incomprehensible parts of an utterance.
- The tilde ("~ ") is used to denote a truncated signal; it may only be used as the first or the last character in a transcription. It is placed immediately before the first or after the last word without a separating blank. In case a word is incomplete due to a truncated signal, the tilde is set without an asterisk. Also, if a signal begins or ends with some speaker or non-speaker noise, the tilde is not set.

The transaction time for each speaker and call including time for setup and breaks, etc. was approximately 30 minutes for novice operators, 15 minutes for experienced operators (using the Macintosh based validation software developed at the University of Munich. This software is described in a manual which can be found in the DOC directory of the CD-ROM).

## 1.4 Lexicon

The lexicon was provided by Siemens AG as a two-column table which in the first column contained the orthographic form of a word (following the LaTeX conventions for German umlauts and ßs, the second column contained one standard high German reference pronunciation in the SPICOS format (no variants were allowed). The lexicon contained all words that occurred at least twice in the corpus of 17.000 sentences taken from the economics pages of the *Süddeutsche Zeitung* of Mai 1994. The lexicon covered approximately 12.000 sentences.

For SpeechDat, the pronunciation was converted automatically to the Verbmobil/SAM phonetic alphabet (*VM/SAM-PA*. *VM/SAM-PA* deviates from standard SAM-PA in the

symbol used for the glottal stop and the accent markers. Because VM/SAM-PA is a superset of SPICOS, some VM/SAM-PA symbols were never used by the automatic conversion. This lexicon was used to generate the phonetically balanced sentences and the other items to be read by the speaker. It contains the orthographic form of a word in its first column, a VM/SAM-PA reference pronunciation in the second column, and a word count in the third column.

Allianz-Versicherung    a l i : a n t s f E r z I C @ r U N    31

The transcription of the spontaneous speech added new items to the dictionary. Any item not already in the dictionary was manually assigned a reference pronunciation. This reference pronunciation takes advantage of the full set of SAM-PA symbols.

Croissants    k r 0 6 s 0 ~ s    32

The lexicon contains approx. 3600 entries, 1100 of which were added to the lexicon by the spontaneous speech transcription.

## 1.5 Validation

The transcriptions were done by 5 students (studying Phonetics, German literature and theatre). All transcriptions were checked automatically for spelling errors and semi-automatically for errors in the use of the asterisk and the tilde. A final check of the transcriptions with audio output was performed only on a small fraction of the transcriptions (approx. 5

## 2 Speaker Recruitment

The recruitment was carried out internally by Siemens Munich. Prompt sheets (provided by IPK Munich) were sent to group leaders etc. to recruit 10 employees each. If a certain number of recordings in this group could be verified (by signature), the whole group received phone cards (value DM 6.-) as a reward. To balance the gender distribution, additional female speakers were recruited at Munich university.

The data collection on CD comprises 1000 speakers, balanced with respect to gender (500 male, 500 female). The regional distribution is based on the division of Germany into states plus the neighbouring countries Switzerland and Austria. This is only a rough approximation of the actual language regions, but the data can easily be obtained from the speakers.

Region	Count	Region	Count
BB	7	RP	36
BE	46	SH	26
BW	69	SL	8
BY	372	SN	28
HB	10	ST	10
HE	38	TH	11
HH	27	A	11
MV	4	D	15
NI	79	others	20
NW	163	unknown	20

Clearly, Bavaria (*BY*) is over-represented (this is due to the fact that Siemens AG is a Bavarian company and due the University recruitment in Munich); the other states of West Germany and Berlin (*BE*) are represented according to their population, whereas the states of East Germany (*BB*, *MV*, *SN*, *ST*, *TH*) are under-represented.

The age distribution of the callers is as follows:

Age	Count
up to 20	37
21 - 30	219
31 - 40	320
41 - 50	213
51 - 60	161
61 - 70	30
71 - 80	2
unknown	18

Speakers called from any location (home, office, with wired/cordless phone).

The speaker data is held in a table in file `SPEAKER.TBL`.

### 3 Prompt Specs

At the Garmisch-meeting of the industrial partners of the SPEECHDAT project in June 27-28, 1994, the following utterances to be gathered were specified:

- 1 isolated digit
- 3 connected digits
  - 4 digit id/sheet number
  - 10 digit telephone number
  - 12 digit credit card number
- 3 natural numbers
- 2 money amounts

- 1 large amount (main currency units only)
- 1 small amount
- 3 spelled words
- 1 time of day (spontaneous)
- 1 time phrase (prompted, word style)
- 1 date (spontaneous, e.g. person's birthday, typically numerical)
- 2 dates (prompted, word style)
- 3 yes/no questions (e.g. mobile/non mobile; male/female)
- 1 city of call/birth
- 6 common application words out of a vocabulary of 50 words
- 3 application word phrases using embedded words in application vocabulary
- 9 sentences (read) for phonetic coverage

For the German SpeechDat recordings, 112 different prompt sheets were created. A total of 7500 sheets (i.e. 70 copies of each sheet) was distributed to the potential callers.

The prompt sheet contains the following items (ordering according to specification above. For the actual organization of prompt sheets see the sample sheets given on the CD-ROMs):

### 3.1 1 Isolated Digit

This item is read.

Presentation example:

(regular face text represents the computer prompt text, **bold face** text represents what the speaker is asked to speak; the bullet represents the prompt beep):

- **1**

Digit vocabulary:

- *eins, zwei, zwo, drei, vier, fünf, sechs, sieben, acht, neun, null*

### 3.2 3 Connected Digits

These items are presented as digits (or groups of digits) and they are read from the prompt sheet.

- One n-digit string (prompt sheet number; max. 3 digits followed by a hyphen and a check sum digit including the letter "X")

- One 16-digit string (credit card)
- One 14-digit string (phone number)

Presentation example:

- **449-6**
- **1724 4011 5814 6467**
- **80 07 72 30 70 54 68**

Connected digit vocabulary:

- digits: see above.
- the letter *X*
- the words *und*, *ein*, and *Strich* (i.e. *hyphen*).

### 3.3 3 Natural Numbers

These items are presented as numbers and they are read from the prompt sheet.

The numbers are a random selection from zero to ten thousand (exclusive).

Presentation example:

- **65**
- **438**
- **7647**

Natural number vocabulary:

- digits (see above) plus *ein*
- ‘teens’: *elf*, *zwölf*, *dreizehn*, *vierzehn*, *fünfzehn*, *sechzehn*, *siebzehn*, *achtzehn*, *neunzehn*
- ‘tys’: *zehn*, *zwanzig*, *dreissig*, *vierzig*, *fünfzig*, *sechzig*, *siebziger*, *achtzig*, *neunzig*
- *hundert*, *tausend*
- the word *und*

Numbers less than 100 are written as one word; larger numbers are written as more than one word to reduce the size of the dictionary. Thus, instead of ‘siebentausendsechshundertsiebenundvierzig’ the orthographic transliteration is ‘sieben tausend sechs hundert siebenundvierzig’.

### 3.4 2 Money Amounts

These items are presented in numerical format (as used on price tags) and they are read from the prompt sheet.

One small (fraction of 1 DM) and one large (full DM) money amount. Decimals (only in the small amount) are restricted to: 'fünfzig' (XX.50 DM) and 'neunundneunzig' (XX.99 DM) with an optional currency fraction item ('Pfennig') e.g. 'zwölf Mark fünfzig', 'sechsdreissig Mark', 'drei Mark neunundneunzig', 'sechzehn Mark und fünfzig Pfennig'.

Presentation example:

- **3.70**
- **6310 Deutsche Mark**

Money amount vocabulary:

- digits (see above) plus *ein*
- 'teens' (see above)
- 'tys' (see above)
- *hundert, tausend*
- *und, Mark, Deutsche Mark, D-Mark, DM, Pfennig, Komma*

### 3.5 3 Spellings

Two items are presented as capital letters with a blank between each letter and are read, the third is a prompt for the spelling of the speaker's first name.

The first prompted word to be spelled was taken from the pronunciation dictionary (which may include the hyphen "-" which is common in compound words and names), the second is a random selection of 8 letters (to cover all German letters equally). There will be no additional special spellings, e.g. 'apostrophe' and no difference between lower case and capital letters (with the exception of the  $\beta$  for which there is no capital letter representation).

Presentation example:

- **R E N N E R**
- **Ü Ü W B O J V ß**

Spelling vocabulary:

- *A-Z, Ä, Ö, Ü, ß, Strich, Bindestrich*

### 3.6 1 Time of Day (spontaneous)

The speaker is prompted for the current time of day.

**Note:** The speaker is warned in the introduction to the prompt sheet that he/she will have to produce a date and a time of day. Otherwise, the results are unpredictable ('I have no watch', 'don't know').

Prompt example:

Wie sp

- weekdays: *Montag, Dienstag, Mittwoch, Donnerstag, Freitag, Samstag, Sonnabend, Sonntag*
- months: *Januar, Februar, März, April, Mai, Juni, Juli, August, September, Oktober, November, Dezember*
- ordinal numbers: *erste, zweite, dritte ... einunddreissigste* or the other cases respectively

### 3.9 2 dates (prompted)

These items will be read.

All dates are later than 1994 and before 2010. The format includes both the long and the short version of the year number, i.e. '1995' and '95'.

Presentation example:

- **nächsten Monat**
- **20. Juni 98**

Vocabulary:

- same as for the spontaneous dates.

### 3.10 3 Confirmations/Rejection

These items are prompted.

In German there are many 'allowed' words for confirmation or rejection, e.g. 'ja', 'jawohl', 'ok', 'nein', 'richtig', 'falsch', which are often embedded in phrases: 'das ist richtig', 'nein, falsch', etc. Speakers are thus prompted to answer with 'ja' or 'nein', but the evaluation should also cover most of the other single word responses.

Prompting example:

- Ist heute Donnerstag? •
- Waren Sie letzte Woche im Kino? •
- Rufen Sie von zuhause an? •

Confirmation/rejection vocabulary:

- *ja, jawohl, nein, richtig, falsch*

### 3.11 1 City Name

This item is prompted.

Prompt example:

- Aus welcher Stadt rufen Sie an? •

### 3.12 6 Com

- möchten Sie die Aufnahme abhören
- möchten Sie die Nachricht anhören
- speichern Sie den Anruf
- löschen Sie die Seite
- drücken Sie halt
- Ende der Ansage
- wir verbinden Sie weiter
- Bitte die Ansage auslösen
- die Nachricht wiedergeben
- bestätigen Sie die Eingabe
- bestätigen Sie den Anruf
- Bitte den Anruf weiterleiten
- aktivieren Sie den Dienst
- Ihr Anruf wurde gespeichert
- die Ansage ist gelöscht
- die Umleitung aufheben
- die Auskunft ist zur Zeit nicht verfügbar
- aktivieren Sie den Beantworter
- der Dienst ist zur Zeit nicht verfügbar
- der Anruf ist extern
- hinterlassen Sie eine Nachricht
- der Anruf ist intern
- die Konferenz ist beendet
- wählen Sie aus dem Menü
- Ende der Nachricht
- wählen Sie die Nummer
- programmieren Sie die Nummer
- drücken Sie Rücklauf

- aktivieren Sie den Rückrufwunsch
- Schutz zuschalten
- Drücken Sie Stop
- Gehen Sie ans Telefon
- Übergabe bestätigen
- Umleitung aktivieren
- Vermittlung aktivieren
- drücken Sie Vorlauf
- bestätigen Sie die Wahl
- sprechen Sie weiter
- drücken Sie Wiedergabe
- Wiederholung der Nachricht
- zurück zum Menü
- Konferenz zuschalten

### 3.14 9 Phonetically Rich Sentences

Each sentence fits in one line of the prompt sheet. The only constraint in regard to the phoneme distribution is that in each set of 9 sentences to be printed on a prompt sheet each (context-independent) phoneme is represented at least two times.

The dictionary provided by Siemens represents the pronunciation in SPICOS format. This format has been translated to German SAM-PA. However, SPICOS does not contain all German SAM-PA phonemes: it does not contain stress marks ( ; " ), the glottal stop /?/, the vocalic r /6/ and /Z/ (as in *Garage*).

A newspaper corpus of > 10000 sentences with a total vocabulary of > 3000 words is used to select 112 mini corpora.

Presentation example:

- **Brüssel hilft Klöckner zu retten.**
- **Dollarkurs bleibt schwach - Franc auf Jahreshoch.**
- **USA verhindern eine Kapitalerhöhung.**
- **Ausland bestellt wieder mehr in Ostdeutschland.**
- **Joint venture in Brünn für Motoren errichtet.**
- **entsprechend locker ist der Umgang mit den Resultaten.**
- **höhere steuern können Kapitalbildung beeinträchtigen.**
- **Frühjahrs-Belebung im westlichen Einzelhandel.**
- **die deutsche Werkzeugindustrie hat die Talsohle noch nicht erreicht.**

For the complete set of sentences, see the file "TABLE/PRMPTSHT.TXT" on the CD-ROMs.

### 3.15 Prompt Sheet ID, Sex and Region

These items are spontaneous.

Three additional questions are asked to obtain the information about the individual prompt sheet, and the gender and region of the speaker.

Prompt example:

Bitte lesen Sie die Fragebogennummer vor. • 449-2

In welcher Stadt haben Sie die Volksschule besucht? •

Sind Sie weiblich oder männlich ? •

### 3.16 Spontaneous Speech

The last item of the prompt sheet is a request for some spontaneous speech: "if you still have some time, please tell me what you prefer to eat for breakfast ". A maximum of 20 seconds of speech is recorded.

### 3.17 Mapping of File Types to Contents

The exact mapping of the two-letter file types to the type of speech please see the appendix.

## A File Types and Contents

I1 isolated digit  
C1 3-5 digit id/sheet nr  
C2 16 digits in quadruples  
C3 14 digits in pairs  
N1 natural numbers (<100)  
N2 natural numbers (<1000)  
N3 natural number (<10000)  
M1 large money amount  
M2 small money amount  
L1 spelled word (spontaneous, first name)  
L2 spelled word (from dictionary)  
L3 spelled word (letter sequence, non-word)  
T1 time of day (spontaneous)  
T2 time phrase (prompted, word style)  
D1 date (spontaneous)  
D2 date (prompted, word style)  
D3 date (prompted, word style)  
Q1 yes/no question  
Q2 yes/no question  
Q3 yes/no question

P1 state of entry to school  
A1 application word  
A2 application word  
A3 application word  
A4 application word  
A5 application word  
A6 application word  
E1 command phrase with application word  
E2 command phrase with application word  
E3 command phrase with application word  
S1 phonetic rich sentence  
S2 phonetic rich sentence  
S3 phonetic rich sentence  
S4 phonetic rich sentence  
S5 phonetic rich sentence  
S6 phonetic rich sentence  
S7 phonetic rich sentence  
S8 phonetic rich sentence  
S9 phonetic rich sentence  
S0 spontaneous speech  
\end{verbatim}

\section{Sample SAM Header}

\begin{verbatim}

LHD: SAM, 5.00

DBN: SpeechDat(M)\_German

VOL: FIXEDODE\_01

SES: 0520

SHT: 449-6

REG: BY

SEX: M

AGE: 33

DIR: \FIXEDODE\CD01\BLOCK05\SES0520

SRC: A00520S9.DEZ

CCD: S9

REP: Institute of Phonetics, University of Munich, Germany

RED: 12.10.95

RET:

BEG: 0

END: 80000

SAM: 8000

SNB: 1

SBF:

SSB: 8

QNT: A-LAW

CMP: GZIP

LBD:

LBR: 0,80000, , , , die deutsche Werkzeugindustrie hat die Talsohle noch nicht err  
 LBO: 0,40000,80000, die deutsche Wert\* Werkzeugindustrie hat die Talsohle noch nich  
 ELF:

## B Regions of German

The language regions of German are mapped to the states of Germany and the neighbouring countries Switzerland and Austria.

State Name	Abbreviation
Baden-Württemberg	BW
Bayern	BY
Berlin	BE
Brandenburg	BB
Bremen	HB
Hamburg	HH
Hessen	HE
Mecklenburg-Vorpommern	MV
Niedersachsen	NI
Nordrhein-Westfalen	NW
Rheinland-Pfalz	RP
Saarland	SL
Sachsen	SN
Sachsen-Anhalt	ST
Schleswig-Holstein	SH
Thüringen	TH
Deutschland	D
Österreich	A
Schweiz	CH
andere	