BAVARIAN ARCHIVE FOR SPEECH SIGNALS (BAS) STATUS REPORT 1995-2000

Florian Schiel, Christoph Draxler, Phil Hoole

Bavarian Archive for Speech Signals (BAS)
University of Munich, Germany
[bas@phonetik.uni-muenchen.de]

Outline of this Report

The *Bavarian Archive for Speech Signals* (BAS) is a joint initiative of the Bavarian State and the Ludwig Maximilians Universität München. It is located at the host organisation *Institut für Phonetik und Sprachliche Kommunikation* and collects, evaluates, produces and disseminates speech based resources to the scientific community. Our focus is the German language covering a large geographical part of central Europe.

This report covers BAS related activities in the period of Jan 1995 to Feb 2000. The report is organised as follows.

The first part after the general introduction deals with our experiences during the last five years to produce new, highly re-usable speech resources in close cooperation with industrial partners. We will explain our BAS policies that ensure that valuable resources do not only satisfy short-term needs of certain industrial applications but are useful for other research or engineering activities as well. Some example projects will highlight these points.

The second part lists projects and/or cooperations that have been conducted or started during 1995 to 2000. Furthermore, associated projects where BAS is not directly funded but agreed to act as a focal point or distribution/evaluation center are given here.

The third section gives an overview about the administration and dissemination methods currently used by BAS. We also will give numbers about how many resources were distributed in the reporting period to how many users as well as details about our cooperation with the European Language Resources Association (ELRA).

Section 4 gives a brief overview of some special scientific multi-modal data collected at BAS, while the final section 5 is reserved for minor BAS activities and contacts.

In appendix A you will find a concise description of all German speech and speech-related resources that are currently available at BAS, while in appendix B you will find a brief description of the annotation tool WWWTranscribe developed and distributed by BAS.

Introduction / Background

The *Bavarian Archive for Speech Signals* (BAS) is a public institution hosted by the *University of Munich*. It was founded in January 1995 with the aim of making corpora of current spoken German available to both the basic research and the speech technology communities via a maximally comprehensive digital speech-signal database. The basic speech material should be structured in a manner allowing flexible and precise access, with acoustic-phonetic and linguistic-phonetic evaluation forming an integral part of it. Furthermore, we seek to promote scientific progress in the new field of Computational Phonetics by applying new techniques of speech processing to large corpora. The

outcome of these activities will hopefully influence the performance of ASR systems as well as Speech Synthesis systems.

BAS is mainly funded by the *Bavarian State* (scientific staff) and the *Ludwig Maximilians Universität München* (infra-structure, administrative staff).

The last few years have seen an abrupt increase in the demand for large speech signal data collections, both on the part of academic investigators carrying out basic research as well as on the part of engineers from industry working in the new integrated field of speech and information technology. There are many reasons for this. Primarily, the sudden increase in demand must be attributed to the breakneck pace of hardware and software development in speech signal processing. The increasing number of techniques for acoustic-phonetic signal processing, and the increasing amount of speech data that can be efficiently handled and processed together generate an accompanying demand not only for linguistically interesting text material (which of course emerges automatically from the modern printing industry) but also for reliably acquired and phonetically validated spoken language material. A number of national and international initiatives (such as BDSON, PHONDAT, LDC, SPEX or COCOSDA) have already resulted in the collection and distribution of large speech corpora. However, they exhibit a variety of formats, corresponding to the variety in the aims pursued. For German, a central institution was clearly lacking that could carry out such tasks within a long-term perspective.
(http://www.bas.uni-muenchen.de/Bas/).

1 The BAS Policies Outlined

The production of a new speech resource is always twofold: Usually expensive resources are not produced without a reason. The most common reasons are that the resources will be necessary for a publicly funded research project or that the resources are needed for the development of a new product. In both cases it's in the interest of the funding part of the project not to spend more money than is needed to fulfil the basic needs at hand. On the other hand we all know that resources which were designed and produced with the sole purpose of a single project or product will not be very useful for other tasks in the future. In other words: there always will be the trade-off between expenses and re-usability.

Our work at BAS during the last 3 years has shown that in most cases *it is possible* to combine these antagonistic positions in a way that satisfies both sides perfectly. Let us give two examples:

The *Regional Variants of German 1* (RVG1) corpus that was collected in close cooperation with *AT&T* and *Lucent Technologies* contains speech data for short-term needs (digits, phone numbers, commands) as well as phonetically rich sentences and spontaneous speech for research and future application purposes (see [Burger & Schiel, 1998]). Furthermore, the data were collected not only with standard low cost recording technique but with two additional high quality channels.

The first EU funded *SpeechDat M* project (1000 speakers per language recorded over public telephone lines, see [Draxler, 1996]) contained phonetically rich sentences and - in some languages - spontaneous speech as well.

In both examples the monetary effort was mainly due to the recruiting and recording techniques. The additional cost for the extra recordings were marginal. But these extra recordings already begin to play a major role in the exploitation of these resources. (This can be seen for instance in the fact that in SpeechDat II the phonetically rich words and sentences got a prominent role.)

The last three years have also shown that there is a growing demand by Speech Engineering for highly specialised speech corpora. Examples are:

- speech under real life noise conditions

2

- speech in the running car

- speech for specific dictation domains

- speech of children

- speech with special microphones / head-sets / built-in-mics

- dialectal speech

- speech with non-prompted utterances

At the moment we do not expect this tendency to come to an end. Thus, the production of re-usable resources should be a major issue for all speech resource producing sites in Europe.

Consequently, our main policy issues at BAS are:

1. BAS offers to co-produce specialised corpora together with industrial or scientific partners, but – whenever possible – BAS will make sure that the properties of the resulting resource will be usable for more than one very specialised application.

2. BAS will negotiate a certain time span for the resource, during which the funding partner(s) will have the exclusive rights to use the resource for his/their commercial application. After that 'blocking period' (usually 1 year) the resource will be available to other members of the scientific community from BAS (a possible license fee is negotiated with the industrial partner as well).

3. BAS takes care that the produced resource satisfies a minimum of internationally accepted standards of technique and quality, thus again improving the re-usability of the resource by others than the directly involved partners.

Following these three basic policies we hope that in a few years there will be a vast amount of re-usable speech and speech-related resources available at BAS.

We strongly recommend other speech focus centres in Europe to follow similar policies while producing speech resources in other European languages. We would be very interested to discuss these points with colleagues from other European countries.

2 Cooperations and Projects

TERABACK - Leibniz Rechenzentrum München

In 1994 a pilot study within the 'Bayern On-line' initiative was started by the 'Leibniz Rechenzentrum München' (LRZ) in close cooperation with BAS. In this pilot - called TERABACK - BAS acted as a 'real world tester' for the newly installed ADSM backup and archive facilities of the LRZ. The pilot ended after one year and resulted in a permanent use of the LRZ ADSM backup and archive facilities for the BAS archive. Since then BAS has not only occupied more than 350 GByte of double-backup archive space but also used the LRZ server machines for fast access to cached BAS master volumes. For technical details see section 'Archive and Dissemination'.

RVG1 - Lucent Technologies / ATT Bell Labs

In March 1995 BAS started a long-term collection financed by AT&T Germany (later by Lucent Technologies and AT&T Labs) called 'Regional Variants of German' (RVG). The aim of this project was to collect speech of speakers originating from all German speaking areas in Europe (that is Germany, Austria, Italy and Switzerland). The speech samples were collected 'in-field' with 4 different microphones into a standard PC. The read speech consists of application oriented commands and number items (telephone, banking, dates, etc), phonetically balanced sentences and 1 minute of spontaneous monologue. Mode details can be found in appendix A, section RVG 1.
The first phase of the project ended in 1998 with 500 recorded speakers. The first two

microphone channels (low quality) were published after one year of blocking period. The other two channels were recorded on high quality DAT tapes and were processed during 1999. The latter was solely funded by the royalties earned by distributing the RVG corpus. The high quality channels will be distributed for the first time in April 2000.

The industrial partners aggreed to use their license royalties from this corpus to process and publish the corpus via BAS.

SpeechDat - EU and industry

In the SpeechDat projects, large databases for voice-operated teleservices are created according to common specifications for many European (and now also non-European) languages.

BAS has been involved in the SpeechDat project since its initiation: in 1994, it joined SpeechDat(M) as an associate partner to collect 1000 German speakers via the fixed network telephone; this collection was carried out in collaboration with SIEMENS AG, Germany.

SpeechDat(M) defined a common core of approx. 45 items: digits, numbers, digit strings, spellings, command words and phrases, and phonetically rich material; all speech is recorded via the telephone. Eight databases of 1000 speakers each were collected for Danish, English, French, German, Italian, Portuguese, Spanish, and Swiss French; additionally, a database of 300 Italian speakers via the mobile phone was recorded.

In **SpeechDat(II)**, the BAS was a subcontractor to SIEMENS AG for the collection of 4000 German speakers via the fixed telephone network, and to Vocalis Ltd, UK for 1000 speakers via the mobile telephone network. Both the fixed and the mobile network database were annotated by BAS; additionally, 500 German speakers were annotated under a contract by Lernout & Hauspie, Belgium. Both the mobile and the fixed network German databases have been validated successfully by SPEX; for the fixed network German database the error rate for speech was found to be 3.4%, the error rate for non-speech (i.e. noise) was found to be 2.0% (both values well below the limit of 5.0% and 20.0%). For the Luxemburg database the corresponding values are 4.7% and 0.6%; in the mobile data 4.7% and 0.5%.

In **SpeechDat-Car**, BAS is responsible for the collection of the German database under a contract with Robert Bosch GmbH and BMW AG, Germany. In this project, 600 sessions will be recorded in nine languages; the recordings are carried out both in a car and synchronously via a GSM phone. In the car, four high-bandwidth channels are recorded; the vocabulary consists of application words and phrases for vehicle control, teleservices, and telecommunication commands (60%), and the standard SpeechDat material (40%) [11]. Finally, the BAS maintains the SpeechDat WWW server: www.speechdat.org from which all SpeechDat projects can be accessed.

The SpeechDat have spawned a number of similar projects:

- SpeechDat(E): SpeechDat in Eastern Europe, and

- SALA: SpeechDat Across Latin America

There are three main reasons for the success of the SpeechDat idea: the databases meet the short and mid-term requirements of SLP industry and research, there is an external validation of all databases to guarantee a high quality standard for all databases, and all specifications and formats were made publicly available.

For
the annotation of the SpeechDat databases an innovative WWW-based tool (WWWTranscribe) was developed and presented successfully at international conferences (Eurospeech 97, LREC 98, MATISSE 99).

PHONOLEX - Universität Leipzig, Deutsches Forschungszentrum für künstliche Intelligenz (DFKI)

Since nowadays SLP application in German are still designed on a word based structure, there is always a general need for a very large pronunciation dictionary of contemporary German. In 1996 BAS started an initiative called PHONOLEX to produce a very large (500.000 and up) reproducable pronunciation dictionary for this purpose. As a first partner the 'Deutsches Forschungszentrum für künstliche Intelligenz' (DFKI) joined the initiative and delivered a first fully inflected base list of German (650.000 entries). At BAS a text-to-phoneme system kindly provided by the University of Bonn was ported to C and adapted to the PHONOLEX task. A first version (1.4) of PHONOLEX was published end of 1996. In 1998 the University of Leipzig joined the group and added an empirically based word list of contemporary written German of 1.000.000 entries to the project. Furthermore, end of 1998 the MAUS group at BAS added their pronunciation analysis of the Verbmobil I corpus to the dictionary. Currently the PHONOLEX dictionary contains the following items per entry: orthographic string in LaTeX, linguistic word class (optional), genus (optional), origin, canonical pronunciation, empirically detected pronunciations. The list in its current version comprises over 1.6 Million entries.

SC2 - Siemens AG München

End of 1997 a small collection was performed by BAS in cooperation with the newly founded 'Speech Center' at Siemens AG, Munich. The corpus - listed as 'SC2 - Noises' in the 'Strange Corpora' series of BAS - was recorded in-field in two car maintenance halls in Munich. The 8.000 speech samples collected from 10 male speakers consist of 3-7 word commands in the domain of car damage diagnosis. The signals are enriched by real world technical noise of different origins. All signals were validated according to SpeechDat-like protocols and technical noises were marked and labelled in the validation strings. The SC2 corpus will be available to the public beginning of 2001.

SI1000P - Siemens AG München

In 1996 this corpus was recorded in cooperation with Siemens AG ZT, Munich. The aim of this project was to produce high quality speech samples for PSOLA-like concatenative speech synthesis. Two professional speakers read the sentence list of the SI1000 corpus (newspaper texts) in a echo cancelled environment. The speech signal as well as the laryngographic signal were recorded. The latter was analysed for the location glottal pulses in the time stream. Furthermore the recordings of one speaker were labelled prosodically, that is boundary markers B3, B2 and B9 (GTobi) and sentences stress (primary, secondary and contrastive) were marked in the transcript of the recording.
The SI1000P corpus will be distributed the first time in April 2000.

Verbmobil I & II - Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMB&F)

Immediately after its foundation in 1995 BAS agreed to act as an informal partner to the **Verbmobil I** consortium in respect to speech resources resulting from this project. The main task for BAS was (and still is) to ensure a proper publication and dissemination of empiric data to the speech science community after the blocking period of one year. Furthermore, BAS re-edits all Verbmobil speech corpora after one year and includes all segmental or other symbolic information available over the time of the project. BAS also acts as a contact point to the European ELRA in Paris with regard to Verbmobil resources.
Verbmobil II started in 1997 with a reduced consortium. BAS aggreed to continue its services to the end of the project in 2000. Currently (Feb 2000) Verbmobil II is in its final phase.
BAS continues to maintain the freely available parts of the collected data. In contrast

to Verbmobil I the scenarios of the recorded situations were extended and the format was re-defined for a better parseability as well as a better handling of the English and Japanese parts of the corpus. The Verbmobil I (VMI) corpus containing recorded dialogues in the Verbmobil scheduling task (8 German, 3 English CDROMs) is currently extended by the following data (estimates):

- 4 CDROMs of Japanese VM I dialogues

- approx. 10 CDROMs of German VM II dialogues

- approx. 4 CDROMs of English VM II dialogues

- approx. 5 CDROMs of Japanese VM II dialogues

Furthermore, the speaker and recording database was re-defined and standardized for all data. Pronunciation lexica for the three languages, phonemic segmentations of the German part and other linguistic resources (such as dialog act labeling, prosodic labeling, tree banks, parts of speech tagging) will be included in the final corpus.

SmartKom - Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMB&F)

The SmartKom project started in Sept 1999 and aims to develop new intelligent human - machine interfaces in the area of personal computing. The IPSK is one of 10 industrial and academic partners of the SmartKom consortium directed by the Deutsches Forschungszentrum für künstliche Intelligenz (DFKI) in Saarbrücken, Germany. All empiric investigations and data collection will be carried out at IPSK in Munich. The BAS agreed to maintain and disseminate the collected mult-modal data to the scientific community after a blocking period of one year. The multi-modal data consist of audio recordings (10 channels), video streams (3 channels), gesture (pointing) recognition and graphic

tableau output. The target interface of SmartKom combines multi-channel speech input under severe conditions (public, car, home environment), hand gestures (2-dimensional) and facial expressions with an intelligent dialogue handling and multi-modal output channels.
More details can be found in http://smartkom.dfki.de.

VeriDat - German Telekom, T-Nova

The VeriDat project started in Nov 1999 and aims to produce the first very large speaker verification database via public phone lines for German. 150 speakers are recorded in 20 different sessions with 40 read items each. The environment and type of phone line is controlled as well as a concise speaker profile.
Since the database will be SpeechDat compatible (that is, compatible to the Speech-Dat specifications regarding speaker verification data bases), the BAS in concerned with the evaluation and validation of the corpus. More detailed, BAS is responsible for the pre-validation test, the transcription and validation of all recordings and for general consulting to ensure SpeechDat quality. The final validation of the corpus will be carried out by SPEX (Netherlands) to ensure an independent objective test set.

3 Archive and Dissemination

Basic Archive

In most cases speech resources can be divided into two fundamental parts: the physical signals and the symbolic information that accompany the signals, such as orthographic transcripts, segmentations and labeling, documentation, speaker information, etc. The first part usually consumes much more memory than the second which basically consists of simple text files.

Our work with speech resources during the last 10 years has shown that the symbolic information of a resource is much more often updated or extended, while the physical signals itself very seldom need a further pro-

cessing after the first release (exceptions are of course possible). We therefore decided to sperate these two parts of each resource in different archive locations, where the signals are stored on a hardware with maximum security and practically infinite space but slow access, while the symbolic information is kept online on a standard UNIX file system. Using 'Andrew File System' (AFS) and symbolic linking under UNIX it was possible to administer a virtual file space of about 100 Giga-Byte containing all archived Master Volumes (MV) together with former versions and security backups of the currently available BAS resources.

More in detail the virtual master is organised as follows: The storage hardware consists of two robots handling 8 high speed tape devices and a silo of several hundred tapes each. The silo units are connected to a so called 'cache' system of approx. 12 GigaByte via fast optical fiber connections. The cache area is divided into 16 Master Volume partitions that each can store the amount of one CDROM (650 MegaByte each). The cache is linked into an AFS file system on the Internet that allows every workstation at BAS to access the data of the 15 Master volumes. Simple archive and retrieve commands download or save existing Master Volumes from tape archive to cache and vice versa. Simultaneously the contents of the cache, which also acts as a processing space are backupped every night to a different section of the tape archive. All the symbolic information of each Master Volume is stored at BAS and linked to the master volume in the cache. That way the symbolic parts are always online and can even be retrieved by users via the World Wide Web.

This concept was made possible by a close cooperation with the *Leibniz Rechenzentrum München* (LRZ) which provided BAS with about 2 TeraByte (later 6 TeraByte) of effective storage space in their IBM based ADSM backup and storage system.

In 1999 we added a local 140 GByte cache to simplify the access to the most used mas-

ter volumes. This enables us to perform cross corpora searches on the BAS archive as well as special customized editions of sub-corpora.

Dissemination via CDROM

The standard way to disseminate a BAS resource is by ISO 9660 CD-R. Since there is no market for a mass production (press), each distributed CD-R is pressed after ordering. This has the advantage that the user always receives the latest version of the resource. The disadvantage is, of course, that the 'press on demand' is more time-comsuming and possible more prone to errors. Therefore every CD-R is test-mounted and checked for technical errors before shipping. The time to produce 1 CD-R including complete download from the tape archive (see previous section), writing and checking is about 1h20m (of course most steps are done automatically).

CD-Rs have a lower life time than pressed CDROMs. We therefore offer a licence free refresh service to users that want a newly written CD-R after 5 years or have problems to read their originally received media.

CD-Rs are shipped via public mail or express services if required by the user.

Dissemination via DVD

Since the speech corpora produced recently tend to grow in dimensions that do not allow a dissemination via CDROM any more (corpora larger than 20GByte like the RVG1 corpus), we started to investigate the possibility of distribution via DVD 5 media. Since Dec 1999 BAS has the possibility to burn on DVD 5 media, which allows about 4,7 GByte to be stored. However, we still have problems to find users who are willing to accept DVD instead CDROM as well as technical problems with several OS that do not accept file systems greater than 2 GByte.

Dissemination via Internet

Speech Resouces containing only symbolic information like PHONOLEX or PHONRUL as

well as update service packs are distributed via the Internet.

Update Service

All users of BAS resources are entitled to the BAS Update Service. The Update Service includes only the symbolic parts of the resource. For instance if the speech corpus X receives a newly calculated phonemic segmentation, the data are offered freely on the BAS ftp server. Each user is informed via email about the new updated resource.

Distribution Numbers

Within the reporting period approx. 400 CD-Rs containing 205 different resources were distributed to 86 users in the scientific community. Most of the users ordered more than one resource from BAS.

European Language Resources Association (ELRA)

The EU funded European Language Resources Association together with its distribution agency ELDA have signed a contract with BAS that regulates cooperation and brokership procedures between BAS and ELDA. BAS entitles ELDA to list all of its resources in the official ELRA catalogue and acts as a broker between BAS and the end user. Until July 1998 the commission for ELDA was 10% of the published basic BAS CD-R fee and 0% on additional license fees (claimed by the copyright holder).

Since 1st of Aug 1998 the commission is raised to 30% for the basic CD-R fee and to 20% on the additional licence fees. At the same time the basic CD-R fee was raised to DM 500,- to compensate the production costs at BAS.

4 Multi-Modal Data

Three key categories of articulatory data have been targeted for inclusion in BAS: Measurements of fleshpoint data, NMRI data of the vocal tract, digitized video.

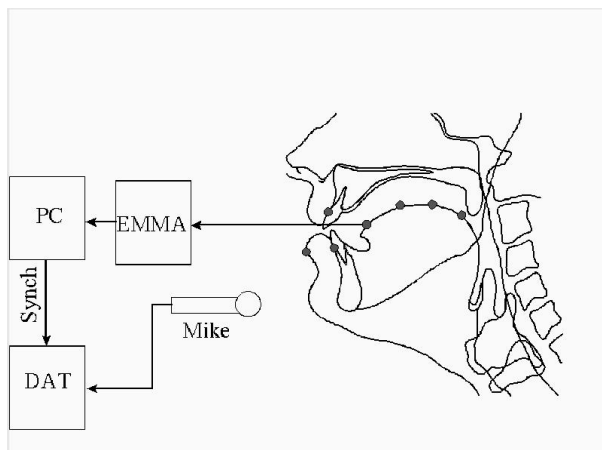**Fleshpoint data** The first set of data avail-

Figure 1: Typical arrangement of sensors for EMMA experiment

able in this category was originally acquired for a project on vowel articulation in German. Movement data was acquired by means of electromagnetic midsagittal articulography for lower lip, jaw and four points on the tongue (see Fig. 1). High quality synchronized audio data was also acquired (16bit, 16kHz). Seven speakers spoke the following corpora:

a) a pseudo-word corpus (in a carrier phrase) of multiple repetitions of the target vowels in three consonant contexts (/p,t,k/). The corpus was recorded at normal and at fast speech rates.

b) a corpus of 105 meaningful sentences (approx. 15 syllables each) containing each target vowel in 15 different contexts. For all corpora a manual segmentation and labeling of the target vowels is available. In addition, for the sentence material a MAUS-based segmentation ([Kipp et al, 1997]) of the complete speech material has been performed. Further details of recording techniques and articulatory pre-processing can be found in [Hoole, 1996]. As an example of analysis results, Fig. 2 shows a speaker-independent articulatory representation of the German vowel space based on PARAFAC factor analysis of the recorded tongue configurations (further details in [Hoole, 1998]). Further multi-speaker corpora from more recent projects will be available in due course.

German vowels in the factor space
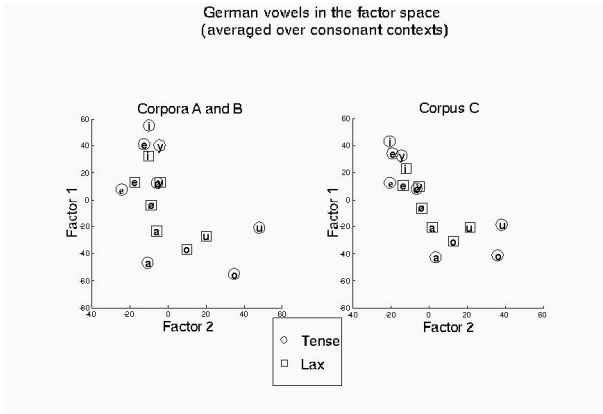(averaged over consonant contexts)

Figure 2: Distribution of German vowels in the factor space determined from PARAFAC analysis of tongue configurations. Corpora A and B are the pseudo-word corpora (normal, fast). Corpus C is the sentence corpus. VOW-ELS

NMRI data of the vocal tract

As part of an ongoing study of alveolar consonant production, and as a supplement to the earlier vowel project just mentioned (wherever possible with the same speakers), NMRI scans have been carried out on, to date, 6 speakers producing the consonants /s/, /SH/, /l/, /n/, /t/, and the vowels /a/, /e/, /i/, /o/, /u/, /y/, /oe/. For each sound 23 slices in each of 3 planes (coronal, axial, sagittal) were collected using a T1-weighted FLASH sequence. Basic MATLAB software for aligning the data from the different image planes and for determining vocal tract cross-sections has been developed. Coupled with EMMA data of the kind outlined above this image data should provide a substantial resource for modelling articulatory-acoustic relationships.

Digitized video

We have implemented procedures for acquiring high-quality digitized video sequences in real time, together with synchronized audio (and EMMA if desired). These techniques have been used, for example, to analyze vertical larynx position in vowel production ([Hoole et al, 1998], and of course are also suitable for facial movements (e.g simultaneous frontal and profile filming using a mir-

ror). We are currently investigating the most suitable means of preparing such very high bandwidth data (approx. 9mb/s) for further distribution.

5 Miscellaneous

BAS Web Services

The BAS maintains several Web server for documentation and data retrieval. The main server address is: www.bas.uni-muenchen.de/Bas where you will find information about the currently available speech resources, on-going projects as well as documentation of file formats, etc.
Currently, approx. 200 documents are maintained on the primary server alone. Other Web servers are used for the offline annotation of speech corpora using the software tool WWWTranscribe developed by Chr. Draxler.

European Union

Since the very foundation of the 'European Language Resources Association' (ELRA) in Paris the BAS is in close cooperation with this institution. BAS has agreed that all language resources (LR) administered and distributed by BAS are included into the official ELRA catalogue (see section 3 for details). Furthermore, F. Schiel was appointed in 1999 by the EC to act as an expert in the field of LR production and dissemination.

Linguistic Data Consortium (LDC)

BAS is in close contact with Linguistic Data Consortium with regards to standardisation of symbolic information accompanying speech signals. A very detailed overview about the methodology and possibles solutions was published by LDC in 1999; this also includes the BAS Partitur Format used for all LRs at BAS ([Bird & Liberman, 1999]). Furthermore, the LDC provides a very detailed web page about existing annotation schemes and annotation software (morph.ldc.upenn.edu/annotation/)

XLDB Workshop

The BAS acts as the organizer of the first in-

ternational XLDB workshop 2000 in Athens (XLDB = 'x-tra large data bases'). This workshop will take place as a satellite event of the 2nd LREC 2000 in June 2000.

Corpus Gesproken Netherlands

In Nov 1999 an expert hearing took place in Tilburg, Netherlands for the new CGN speech data base. The BAS reported in an invited paper about our experiences with regards to specification, annotation, validation and distribution of large speech data bases.

Summer School Leuven

In July 1997 the 'Elsnet 5th European Summer School' took place in Leuven, Belgium. Main topic of this summer school was the development of lexica in the fields of SLP and NLP. The summer school was organized by Frank van Einde and Dafydd Gibbon. Chr. Draxler has given a 7-day course about "speech data bases": specification, recording, validation, etc. A book with extended material of this summer school is in print.

International Conference on Phonetic Sciences (ICPhS)

The BAS reported in an invited paper about the MAUS technique to annotate large corpora of recorded speech into phonemic units ([Schiel, 1999]).

For other publications and conference papers please refer to the reference list in this report. Most papers may be downloaded from the following URL:
www.phonetik.uni-muenchen.de/Publications/
If you prefer a hard copy, please send an request to bas@bas.uni-muenchen.de together with your postal address or fax number.

References

[Beringer et al, 1998] Beringer, N., Schiel, F., Brietzmann, P. (1998). German Regional Variants - A Problem for Automatic Speech Recognition?; in: Proceedings of the ICSLP 1998. Sydney, Vol. 2, pp. 85ff, Dec. 1998.

[Beringer & Schiel, 1999] Beringer, N.; Schiel, F. (1999). Independent Automatic Segmentation of Speech by Pronunciation Modeling. Proc. of the ICPhS 1999. San Francisco. August 1999. pp. 1653-1656.

[Bird & Liberman, 1999] Steven Bird and Mark Liberman (1999). A Formal Framework for Linguistic Annotation Technical Report MS-CIS-99-01 - Department of Computer and Information Science, University of Pennsylvania (expanded from version presented at ICSLP-98, Sydney).

[Burger & Schiel, 1998] S. Burger & F. Schiel (1998). RVG1 - A Database for Regional Variants of Contemporary German. Proceedings of the First International Conference on Language Resources and Evaluation 1998, Granada 1998.

[Burger & Draxler, 1998] S. Burger, Ch. Draxler (1998). Identifying Dialects of German from Digit Strings; Proceedings of the FIRST INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION 1998, Granada, Spain.

[Burger, 1995] S. Burger (1995). Transliterationslexikon (Verbmobil-TechDok 36-95). University of Munich, October 1995. (Online version in English: http://www.phonetik.uni-muenchen.de/VMTraLexeng.html)

[Burger, 1997] S. Burger (1997). Transliteration spontansprachlicher Daten - Lexikon der Transliterationskonventionen - Verbmobil II (Verbmobil-TechDok 56-97), University of Munich, April 1997. (On-

line version: http://www.phonetik.uni-muenchen.de/VMtrlex2d.html)

[Draxler, 1996] Draxler, Chr. (1996). The German SpeechDat Telephone Speech Corpus - Overview and Experiences. Speech Science and Technology 1996 Conference, Adelaide, Australia.

[Draxler, 1997] Draxler, Chr. (1997). WWW-Transcribe - A Modular Transcription System Based on the World Wide Web; Proc. of Eurospeech 97 Conference, Rhodos, 1997.

[Draxler, 1998] Draxler, Chr. (1998). WWWSigTranscribe - A Java Extension of the WWWTranscribe Toolbox; Proc. of 1st Int'l Conference on Language Resources and Evaluation (LREC), Granada, 1998.

[Draxler, 1999] Draxler, Chr. (1999). WWWSigTranscribe - Annotation via the WWW; Proc. of Method and Tool Innovations for Speech Science Education (MATISSE), London, 1999, to appear.

[Geumann et al, 1997] Geumann, A., D. Oppermann, F. Schaeffler (1997). The conventions for phonetic transcription and segmentation of German used for the Munich Verbmobil corpus; Verbmobil Memo No. 129-97, Institut fr Phonetik und Sprachliche Kommunikation, Universitt Mnchen, Dec 1997.

[Grice et al., 1995] Grice, Martine and Ralf Benzmueller (1995). Transcription of German Intonation using ToBI tones; The Saarbruecken System. Paper presented at Tutorial Workshop on Discourse and Dialogue Prosody, Stuttgart, February 1995, modified version also in Phonus 1, University of the Saarland, pp33-51.

[Heid et al, 1995] S. Heid, M.-B. Wesenick, Chr. Draxler (1995). Phonetic Analysis of Vowel Segments in the PhonDat Database of Spoken German , Proceedings of the ICPhS 1995, Stockholm, Sweden, 4:416-419.

[Hoole, 1996] P. Hoole (1996). Issues in the acquisition, processing, reduction and parameterization of articulographic data. in: FIPKM, 34, 158-173.

[Hoole, 1998] P. Hoole (1998). Modelling tongue configuration in German vowel production. Proc. 5th Int. Conf. Spoken Lang. Processing, 5, pp. 1863-1866.

[Hoole et al, 1998] P. Hoole, C. Kroos (1998). Control of larynx height in vowel production. in: Proc. 5th Int. Conf. Spoken Lang. Processing, 2, pp. 531-534.

[Kipp et al., 1996] A. Kipp, M.-B. Wesenick, F. Schiel (1996). Automatic Detection and Segmentation of Pronunciation Variants in German Speech Corpora. Proceedings of the ICSLP 1996. Philadelphia, pp. 106-109, Oct 1996. A. Kipp, M.-B. Wesenick, F. Schiel (1997): Pronunciation Modeling Applied to Automatic Segmentation of Spontaneous Speech. in: Proceedings of the EUROSPEECH, Sept 1997, Rhodos, Greece, pp. 1023-1026.

[Kipp et al, 1997] A. Kipp, M.-B. Wesenick, F. Schiel (1997). Pronunciation Modeling Applied to Automatic Segmentation of Spontaneous Speech. in: Proceedings of the EUROSPEECH, Sept 1997, Rhodos, Greece, pp. 1023-1026.

[SAM, 1989] http://www.phon.ucl.ac.uk/home/sampa/home.htm

[Schiel et al, 1997] F. Schiel, Ch. Draxler, H.G. Tillmann (1997). The Bavarian Archive for Speech Signals: Resources for the Speech Community. Proceedings of the EUROSPEECH 1997, Rhodos, Greece, pp. 1687-1690.

[Schiel, 1997] F. Schiel (1997). Probabilistic analysis of pronunciation with MAUS. The ELRA Newsletter, December 1997, pp. 6-9.

[Tillmann et al, 1995] H.G. Tillmann, Chr. Draxler, K. Kotten, F. Schiel (1995). The Phonetic Goals of the new Bavarian

Archive for Speech Signals. Proceedings of the ICPhS 1995 Stockholm, pp. 4:550-553.

[Schiel et al, 1998] F. Schiel, A. Kipp, H.G. Tillmann (1998). Statistical Modeling of Pronunciation: It's not the Model, it's the data; Proceedings of the ESCA Tutorial and Research Workshop on 'Modeling Pronunciation Variation for Automatic Speech Recognition', May 1998, Kerkrade/Netherlands, pp. 31-36.

[Schiel, 1998] F. Schiel (1998). Speech and Speech-Related Resources at BAS; Proceedings of the First International Conference on Language Resources and Evaluation 1998, Granada, Spain, pp. 343-349.

[Schiel et al, 1998] F. Schiel, S. Burger, A. Geumann, K. Weilhammer (1998). The Partitur Format at BAS; Proceedings of the First International Conference on Language Resources and Evaluation 1998, pp. 1295-1301, Granada, Spain.

[Schiel, 1999] Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech Proc. of the ICPhS 1999. San Francisco, August 1999. pp. 607-610.

[Schiel et al, 1999] Schiel, F.; Draxler, Chr.; Hoole, Ph.; Tillmann, H.G. (1999). New Resources at BAS: Acoustic, Multimodal, Linguistic, Proc. of the Eurospeech 1999, Budpest, Hungary, pp. 2271-2274.

[Tillmann & Tischer, 1995] Tillmann, H.G. & Tischer, B. (1995). Collection and exploitation of spontaneous speech produced in negotiation dialogues. Proceedings of the ESCA Workshop on Spoken Dialogue Systems (Vigso, Denmark, May 30 - June 2, 1995). Aalborg, 1995: ESCA and Center for PersonKommunikation, Aalborg University. S. 217-220.

[Tillmann et al, 1996] H.G. Tillmann, A. Zierdt, P. Hoole (1996). Towards a Three Dimensional Articulographic System. Third Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, Hawaii. JASA vol.100, no.4, pt.2, 2662.

[Wesenick, 1996] M.-B. Wesenick (1996). Automatic Generation of German Pronunciation Variants; in: Proceedings of the ICSLP 1996. Philadelphia, pp. 125-128, Oct 1996.

[Wesenick & Kipp, 1996] M.-B. Wesenick, A. Kipp (1996). Estimating the Quality of Phonetic Transcriptions and Segmentations of Speech Signals; in: Proceedings of the ICSLP 1996. Philadelphia, pp. 129-132, Oct 1996.

A Resources at BAS

The following is a short listing of all speech resources currently available at BAS. Please note that in-depth information to each resource can be found on our Web server. In most cases even online access to the original documentation is possible via WWW. Also note that resources that are not available for the next year starting with April 1998 are listed in a separate section at the end of the paper.

Strange Corpora

The 'Strange Corpora' series was motivated to facilitate the investigation of certain well known problems in speech engineering as well as in the speech sciences. Such fields of investigation are:

- Speaker characteristics (speaker adaptation / normalisation)

- Pathological speech

- Speech of children or the elderly

- Speech in real life noise (Lombard effects, robustness)

- Prompted and non-prompted speech (intonation)

- Typical spontaneous speech effects as hesitations, repairs, breaks

- Accents

- Dialects

The SC series is a collection of smaller corpora (compared to nowadays collections like the SpeechDat project!) which give well documented reference data (bench marks) in the above mentioned topics. Researchers as well a speech engineers might use these corpora to verify their algorithms or applications under controlled and reproducible conditions.

Currently available:

SC1 - Accents
Type: read speech
Format: 14 Bit / 16 kHz / PhonDat
Environment: studio
Recording sites: 1
Speakers: 88
Transcript: orthographic
Segmentation: phonemic
Total: 88 stories (111 words each)
Medium: 1 CDROM
Description:
The corpus contains the same text read by 16 native German speakers and 72 speakers from other cultures/countries. The reference speakers (native Germans) were manually segmented and labelled into SAM-PA phonemic segments ([SAM, 1989]).
Original Purpose of Recording:
Scientific investigation of foreign accents; forensic classification of unknown voices.
Other Usabilities:
Automatic accent detection; adaptation to foreign accents; robust ASR; forensic applications; speaker verification.

Available in Jan 2001:

SC2 - Noises
Type: read speech
Format: 16 Bit / 16 kHz / NIST
Environment: car maintenance hall
Recording sites: 1
Speakers: 10
Transcript: orthographic
Segmentation: noise markers
Total: 8000 utterances (average 4.6 words)
Medium: 1 CDROM
Description:

10 speakers from a car diagnosis firm were asked to read 800 'automobil diagnosis phrases' from a corpus of 100 different phrases (each phrase read 8 times). The recording took place in a car maintenence hall with up to 6 active car lines. The speech was prompted via screen; the speech signals were recorded via a DECT phone system directly to a portable IBM compatible PC. Backround noise of all kinds were classified during the validation process.

Original Purpose of Recording:
Speech recognition for car diagnosis.

Other Usabilities:
Robust ASR under heavy noise conditions; Lombard effects; noise cancelling techniques.

Read Speech Corpora

The following speech corpora contain different types of read speech, whole utterances, commands and single words.

PD1
Type: read speech
Format: 16 Bit / 16 kHz / PhonDat
Environment: studio
Recording sites: 4
Speakers: 201
Transcript: orthographic
Segmentation: phonemic
Total: 21681 utterances (average 7.3 words)
Medium: 4 CDROMs
Description:
The corpus contains carefully read speech of 201 speakers recorded in a echo cancelled studio environment. The speech corpus was selected to cover all possible di-phone combinations in the German standard language (without foreign words). The text corpus consists of 450 different sentence equivalents (including alphanumericals and two shorter passages of prose text) and is not domain specific.

Original Purpose of Recording:
Diphone based ASR.

Other Usabilities:
Bootstrapping ASR; concatenative speech synthesis.

PD2

Type: read speech
Format: 16 Bit / 16 kHz / PhonDat
Environment: studio
Recording sites: 4
Speakers: 16
Transcript: orthographic
Segmentation: phonemic,prosodic,words
Total: 3200 utterances (average 12.3 words)
Medium: 1 CDROM
Description:
The corpus contains 16 x 200 sentences from the train inquiry task fluently read by 16 native German speakers. A subcorpus of 64 sentences per speaker was manually segmented and labelled into SAM-PA segments. The whole corpus was automatically segmented by MAUS. The data of 8 speakers (8000 utterances) were annotated and segmented prosodically.

Original Purpose of Recording:
ASR for a train inquiry system.

Other Usabilities:
Bootstrapping ASR; phonetic investigations; prosodic investigations; ASR using prosodic features.

ERBA
Type: read speech
Format: 14 Bit / 16 kHz / RAW
Environment: office
Recording sites: 4
Speakers: 106
Transcript: orthographic
Segmentation: none
Total: 11100 utterances (average 13.1 words)
Medium: 4 CDROMs
Description:
The corpus contains 101 x 100 (training) and 5 x 200 (test) sentences read by native German speakers. Sentences are unique (with some exceptions) and produced by a stochastic sentence generator (grammar). (Therefore, some sentences are somewhat unusual.) Availability is limited to scientific usage.

Original Purpose of Recording:
ASR for a train inquiry system.

Other Usabilities:
General ASR; speaker adaptation; speaker identification.

SPINA

Type: read speech (mostly single words)
Format: 16 Bit / 16 kHz / RAW
Environment: studio
Recording sites: 2
Speakers: 22
Transcript: orthographic
Segmentation: phonemic,word
Total: 10810 utterances (average 1.2 words)
Medium: 1 CDROM
Description:
The corpus contains very specific commands to control an industrial robot. The text corpus consists of 10 robot command sentences and 62 robot command words. Each speaker has read the entire text corpus at least 5 times. Small parts of the corpus are segmented and labelled into SAM-PA and word units.
Original Purpose of Recording:
ASR for robot control.
Other Usabilities:
General ASR.

RVG 1

Type: screen prompted speech
Format: 16 Bit / 22.05 kHz / NIST
Environment: office
Recording sites: 6
Speakers: 498
Transcript: orthographic (with linguistic markers and noise class)
Segmentation: none
Total: 42000 utterances (average 6 words)
Medium: 30 CDROMs / 5 DVD
Description:
The *Regional Variants of German* (RVG1) corpus contains 85 screen prompted utterances (digits, phone numbers, computer command phrases, phonetically rich sentences) randomly selected from a group of larger text corpora. The 498 speakers were selected according to demographic densities in Germany, Austria, parts of Switzerland and Italy. Speakers were asked to speak informally but not dialectally. Recordings were done with 4 different microphones (from high quality studio mi to low cost desk top mic). See [Burger & Schiel, 1998] for an in-depth description of this resource.

Original Purpose of Recording:
ASR training material with broad regional coverage.
Other Usabilities:
General ASR; research of pronunciation variants; prosody; phonetic investigations.

Dictation Speech Corpora

The following speech corpora contain read speech recorded in a dictation task. The spoken texts were derived from a German newspaper corpus.

SI1000

Type: dictated speech
Format: 16 Bit / 16 kHz / PhonDat
Environment: studio
Recording sites: 1
Speakers: 10
Transcript: orthographic
Segmentation: prosodic (in text corpus)
Total: 10000 utterances (average 25.1 words)
Medium: 5 CDROMs (compressed)
Description:
The corpus contains 1000 sentences from a newspaper corpus read by 10 native German speakers in a dictation task (puntuations are spoken). The text corpus is segmented into phrase boundaries B2, B3 and B9 (GTobi, [Grice et al., 1995]) and words are marked with accent labels PA, NA and EK.
Original Purpose of Recording:
ASR for dictation.
Other Usabilities:
Prosodic segmentation; speaker adaptation; speaker verification.

SI100

Type: dictated speech
Format: 16 Bit / 16 kHz / PhonDat
Environment: studio
Recording sites: 1
Speakers: 101
Transcript: orthographic
Segmentation: none
Total: 10100 utterances (average 23.4 words)
Medium: 7 CDROMs / 1 DVD
Description:
The corpus contains 101 x 100 sentences se-

lected from two different newspaper text corpora (544 + 483 sentences) read by 101 native German speakers in a dictation task (punctuations are spoken).
Original Purpose of Recording:
ASR for dictation.
Other Usabilities:
General ASR; speaker adaptation; speaker identification.

Spontaneous Speech Corpora

The following gives an overview about speech corpora at BAS containing or consisting entirely of spontaneous elicited speech. The term *spontaneous* as it is used in this paper does not imply a totally unaware recording. The correct terminus technicus would be *unscripted speech*. However, since the term *spontaneous speech* is used in many publications and documentations, we'll stick to it.

German VM I
Type: dialogues
Format: 16 Bit / 16 kHz / PhonDat
Environment: studio/office
Recording sites: 4
Speakers: 779
Transcript: VM I + VM II transliteration ([Burger, 1995], [Burger, 1997])
Segmentation:
phonemic,word,prosodic,dialogact
Total: 13910 utterances (average 22.8 words)
Medium: 9 CDROMs / 2 DVDs
Description:
The German Verbmobil I corpus contains 1956 dialog recordings of 779 different speakers. In each dialog both speakers had to negotiate up to 4 business appointments. The whole corpus was segmented and labelled by MAUS into SAM-PA segments; parts of the corpus were segmented manually with regard to phonology, prosody and dialogacts.
Original Purpose of Recording:
ASR for online translation German to English / Japanese.
Other Usabilities:
General ASR; dialog systems; research of elicited spontaneous speech; prosody; phonetic investigations.

German VM II
Type: dialogues
Format: 16 Bit / 16 kHz / NIST
Environment: office
Recording sites: 2
Speakers: not known yet
Transcript: VM II transliteration ([Burger, 1995], [Burger, 1997])
Segmentation:
phonemic,word,prosodic,dialogact
Total: not known yet
Medium: 10 CDROMs / 2 DVDs (estimates)
Description:
In contrast to Verbmobil I the scenarios of the recorded situations were extended and the format was re-defined for a better parseability as well as a better handling of the English and Japanese parts of the corpus. Furthermore, the speaker and recording database was re-defined and standardized for all data. Pronunciation lexica for the three languages, phonemic segmentations of the German part ([Kipp et al, 1997]) and other linguistic resources (such as dialog act labeling, prosodic labeling, tree banks, parts of speech tagging) will be included into the final corpus.
Original Purpose of Recording:
ASR for online translation German to English / Japanese.
Other Usabilities:
General ASR; dialog systems; research of elicited spontaneous speech; prosody; phonetic investigations.

RVG 1
Type: monologues
Format: 16 Bit / 22.05 kHz / NIST
Environment: office
Recording sites: 6
Speakers: 500
Transcript: VM II transliteration ([Burger, 1997])
Segmentation: none
Total: 500 x 1 minute monologue
Medium: (18 CDROMs)
Description:
The *Regional Variants of German* (RVG1) corpus contains – besides the prompted speech (see above) – 1 minute of free monologue of

each speaker. The speakers were asked to talk about their activities of the last week. The data are transcribed according to the Verbmobil transliteration standard ([Burger, 1997]).
Original Purpose of Recording:
Empiric investigations of dialectal variation within Standard German.
Other Usabilities:
General ASR; dialog systems; research of elicited spontaneous speech; prosody; phonetic investigations.

Non-German Corpora

Although BAS is dedicated to the spoken German language there exist a few speech corpora of other languages mostly linked to some German BAS resources.

American VM I
Type: dialogues (American English and 'Denglish')
Format: 16 Bit / 16 kHz / PhonDat
Environment: office
Recording sites: 3
Speakers: 256
Transcript: VM I + VM II transliteration
Segmentation: none
Total: 4029 utterances (average 27.5 words)
Medium: 3 CDROMs
Description:
This corpus contains Verbmobil I style recordings done at Carnegie Mellon University, USA, University of Karlsruhe and Bonn University, Germany. The vast majority of the recordings are done with native American speakers; a small subcorpus was spoken by native Germans with average knowledge of the English language ('Denglish').
Original Purpose of Recording:
ASR for online translation German to English / Japanese.
Other Usabilities:
General ASR; dialog systems; research of elicited spontaneous speech; prosody; foreign accents; phonetic investigations.

American VM II
Type: dialogues (American English)
Format: 16 Bit / 16 kHz / NIST

Environment: office
Recording sites: 1
Speakers: not known yet
Transcript: VM II transliteration
Segmentation: none
Total: not known yet
Medium: 6 CDROMs (estimates)
Description:
This corpus contains Verbmobil II style recordings done at Carnegie Mellon University, USA.
Original Purpose of Recording:
ASR for online translation German to English / Japanese.
Other Usabilities:
General ASR; dialog systems; research of elicited spontaneous speech; prosody; foreign accents; phonetic investigations.

Japanese VM I
Type: dialogues
Format: 16 Bit / 16 kHz / RAW (compressed)
Environment: office
Recording sites: 1
Transcript: adapted VM I transliteration
Segmentation: none
Total: 800 dialogues
Medium: 4 CDROMs
Description:
This corpus contains Verbmobil I style recordings done at Tokyo University, Japan. This corpus has not been validated by BAS and is distributed 'as is'.
Original Purpose of Recording:
ASR for online translation German to English / Japanese.
Other Usabilities:
General ASR; dialog systems; research of elicited spontaneous speech; prosody; phonetic investigations.

Japanese VM II
Type: dialogues
Format: 16 Bit / 16 kHz / NIST
Environment: office
Recording sites: 1
Transcript: VM II transliteration
Segmentation: none
Total: not known yet
Medium: 8 CDROMs (estimates)

Description:
This corpus contains Verbmobil II style recordings done at Tokyo University, Japan.
Original Purpose of Recording:
ASR for online translation German to English / Japanese.
Other Usabilities:
General ASR; dialog systems; research of elicited spontaneous speech; prosody; phonetic investigations.

Other Resources

EMA1
Type: screen prompted speech
Format: Audio: 16 Bit / 16 kHz. EMA: 6+2 sensors, X,Y + velocity + tilt, 250 Hz / 16 Bit
Environment: studio
Recording sites: 1
Speakers: 7 (6 male, 1 female)
Transcript: Orthographic
Segmentation: phonemic, vowel onsets, context consonants, etc.
Total: 3906 utterances
Medium: 1 CDROM
Description:
The corpus contains recording of the movement of the main articulators in the midsagittal plane together with the speech signal. The data include the speech signal, X/Y position, X/Y velocity and tilt factor (reliability) of 6 sensors and 2 reference sensors. Four fifth of the text corpus consists of carrier phrases with all German vowels embedded into changing consonantal context spoken with normal and fast speed (2 x 225); one fifth (108) consists of real sentences with the same vowels contained.
Original Purpose of Recording:
Investigation of the vowel production in German.
Other Usabilities:
Modelling of the vocal tract; improving ASR with articulatory parameters; phonetic investigations.

PHONOLEX
Type: Pronunciation Dictionary for German
Format: ASCII
Total: approx. 1.600.000 entries
Medium: 1 CDROM or FTP
Description:
The PHONOLEX dictionary contains a fully inflected list of the most common German words together with their canonical pronunciation in SAM-PA.
Original Purpose of Recording:
Lexicon lookup for automatic phonemic segmentation with MAUS.
Other Usabilities:
ASR; Speech Synthesis.

PHONRUL
Type: Pronunciation Rule Set
Format: ASCII
Total: approx. 5.000 rules
Medium: Floppy Disk or FTP
Description:
PHONRUL is a collection of simple re-write rules for German pronunciation. Starting with a canonical representation of the utterance in SAM-PA the rule set can be used to create the most likely pronunciation variants expected in Standard German (no dialectal variation).
Original Purpose of Recording:
Calculating pronunciation hypothesis for automatic phonemic segmentation with MAUS.
Other Usabilities:
ASR; Speech Synthesis.

Future editions at BAS

The following resources are currently produced at BAS and will be available in the near future.

ZipTel
Part of the German SpeechDat(M) corpus (1000 speakers) that is not included into the official SpeechDat data set. Contains phone number, ZIP codes and German street names read via telephone line by approx. 1000 speakers. Release end of 2000.

FormTask
Part of the German SpeechDat(II) corpus (4000 speakers) that is not included into the official SpeechDat data set. Contains the an-

swers of each speaker to 7 questions about a form in the data sheet. Release end of 2000.

Hempel's Sofa
Part of the German SpeechDat(II) corpus (4000 speakers) that is not included into the official SpeechDat data set. Contains max. 60 sec of spontaneous monolue by each speaker Release end of 2000.

Japanese Verbmobil II
Final edition after blocking period in Sept 2001.

American Verbmobil II
Final edition after blocking period in Sept 2001.

VeriDat
Probably in 2002 or 2003 (still ongoing negotiations about the release date).

Speech in the running car
Currently three corpora with recordings in the running car are or have been produced at the Department of Phonetics, University of Munich:
CSDC2: 238 speakers, 50 utterances, available not before June 2000.
CSDC1: 155 speakers, 92 utterances, available not before June 2000.
CSDC4: 105 speakers, 154 utterances, available not before May 2003.

For more details about BAS resources and ordering information please refer to our WWW documentation:

www.phonetik.uni-muenchen.de/Bas