The Partitur Format at BAS

Florian Schiel¹ & Susanne Burger² & Anja Geumann² & Karl Weilhammer²

¹Bavarian Archive for Speech Signals (BAS), Munich, Germany
²Department of Phonetics and Speech Communication University of Munich, Germany
[bas@phonetik.uni-muenchen.de]

Abstract

Most spoken language resources are produced and disseminated together with symbolic information relating to the speech signal. These are for instance orthographic transcriptions, labelling and segmentation on the phonologic, phonetic, prosodic, phrasal level. Most of the known formats for these symbolic data are defined in a 'closed form' that is not flexible enough to allow simple and platform-independent processing and easy extensions.

At the Bavarian Archive for Speech Signals (BAS) a new format has been developed and used over the last few years that shows some significant advantages over other existing formats. This paper describes the basic principles behind this format, discusses briefly the advantages and gives detailed definitions of the description levels used so far. Furthermore, we will give some examples for easy processing of the format and distributed work on the same data.

In the future all corpora produced and disseminated by BAS will be distributed with the new BAS Partitur Format if they contain segmental information of any kind. The formerly used formats will be retained but not further updated.

General Overview

Most file formats containing segmental information on speech signals have the disadvantage that

- they are not easy to extend (without rewriting software that uses the existing format).
- they are not easy to process with UNIX standard tools.
- they mix different description levels (which leads to technical and conceptual problems)
- they were defined as ad-hoc solutions for very specialised problems and are not capable of being re-used in a different setup.

Therefore a new open format based on the SAM Label Format was developed, which circumvents most of the mentioned problems. In this format all levels of description may be annotated independently but are time aligned like the individual tiers of a score. Hence this format was called 'BAS Partitur Format' ('Partitur' = German for 'score').

In the future all BAS corpora will be distributed with the new BAS Partitur Format if they contain segmental information of any kind. The formerly used formats will be retained but not further updated.

A first draft of the BAS Partitur Format was published in [Atmanspacher et al., 1995].

The BAS Partitur Format has the following features:

- SAM compatible header structure
- Easy to extend and to process by simple UNIX commands
- Open format; extensions to the format can be implemented without alterations to the software that reads the older format
- Time-aligned independent description of a virtually unlimited number of different levels of the speech signal (see examples later in this paper).
- Symbolic links between the independent levels allow logical assignments aside from the physical time scale. These links are based on the word units of the utterance.

Definition (Version 1.2)

A Partitur file name has the same prefix as the corresponding signal file (8 Bytes for Iso 9000 compatibility) but the extension .par. All contents are in 7-bit ASCII exclusively (to guarantee portability to all platforms). Each line starts with a three-byte label followed by a colon; this label defines synopsis and semantics of the ensuing line. The following units of the line are separated by 'white spaces' (blank, tab).

The Partitur file is structured into a header and a body (like SAM description files). The header stretches from the beginning of the file to the label LBD; the body from the label LBD; to the end of file where the last line has to be closed by a 'new line' or a 'CR + LF' (the final SAM label ELF was omitted.
for the BAS Partitur Format since it prevents effective processing of the Partitur files).

The header contains SAM-compatible lines of general information. The following entries are compulsory:

- **LHD**: Partitur file version
- **REP**: Place of recording
- **SNB**: Number of Bytes per Sample
- **SAM**: Sampling Frequency in Hz
- **SBF**: Byteorder (Intel 01, Motorola 10)
- **SSB**: Bit Resolution
- **NCH**: Number of Channels
- **SPN**: Speaker ID
- **LBD**:

Example:

```
LHD: Partitur 1.2
REP: Muenchen
SNB: 2
SAM: 16000
SBF: 01
SSB: 16
NCH: 1
SPN: PS1
LBD:
```

The following entries are optional; apart from these, other entries are tolerated as long as they do not conflict with compulsory and optional entries:

- **FIL**: SAM File Type
- **TYP**: Type of SAM Label File
- **DBN**: Corpus Name
- **VOL**: Number of Volume
- **DIR**: Directory in Volume
- **SRC**: Name of speech file
- **BEG**: Begin of labelling sequence
- **END**: End of labelling sequence
- **RED**: Date of Recording
- **RET**: Duration
- **RCC**: Recording Conditions
- **CMT**: Comment
- **SPI**: Speaker Information
- **EXP**: Name of Segmenter
- **SYS**: Labelling System
- **DAT**: Date of Labelling
- **SPA**: SAM-PA Version

In the following sections the five basic classes of tiers are defined.

**Tiers with symbolic relation (class 1)**

A line of this tier contains:

- the tier label
- a comma-separated list of integers *(symbolic links)*
- a string with the labelling information

The symbolic links refer to a reference tier which numbers the word units beginning with zero. The label string has an internal synopsis which is defined in the tier definition.

Example:

```
TRL: 6,7 mit´m
```

In this example the word events 6 and 7 of an utterance are transliterated.

**Tiers with time-consuming events (class 2)**

A line of this tier contains:

- the tier label
- two integers denoting the begin and duration of the event.
- a string containing the labelling information

The semantics of the integers is defined by the tier definition (possible are samples, milliseconds, etc.)

Example:

```
PHN: 13456 3450 aU
```

In this example a phonemic segment labelled /aU/ stretches from sample 13456 for the next 3450 samples.

**Tiers with non time-consuming events (class 3)**

A line of this tier contains:

- the tier label
- an integer denoting the time position of the event
- a string containing the labelling information

Example:

```
PRØ: 13456 TÖN: P*; FUN: PÅ
```
In this example the prosodic event labelled TON: P*; FUN: PA (GTobi, see [Grice et al., 1995]) takes place at sample 13456 of the utterance.

**Tiers with time and symbolic relation, time-consuming (class 4)**

A line of this tier contains:

- the tier label
- two integers denoting the start and duration of the event.
- a comma separated list of integers (symbolic links)
- a string containing the labelling information

Example:

SAP: 13456 3450 9 au

In addition to the example above this tier not only gives the starting point and the duration of the phonemic segment but also a pointer to the word unit where it belongs (word 9).

**Tiers with time and symbolic relation, not time-consuming (class 5)**

A line of this tier contains:

- the tier label
- an integer denoting the time position of the event.
- a comma separated list of integers (symbolic links)
- a string containing the labelling information

Example:

PRB: 13456 9 TON: P*; FUN: PA

Again, in this example the prosodic event is not only placed in time but also assigned to a word of the utterance (word 9).

**Remarks**

- If the symbolic link in a tier is not (or not yet) known, the symbolic link is set to -1 (e.g., noises from other sources than the recorded speaker).
- The same symbolic relation may occur in different lines of a tier (for example if more than one event can be assigned to the same word of an utterance).

**Definition of Tiers**

The following sections give an overview of the currently defined tiers in the BAS Partitur Format (version 1.2.2). Please keep in mind that this is an open list in the sense that new tiers can be defined whenever there is a need for it. If somebody would like to work with speech resources from BAS and to define a new tier for his or her specific problem, please contact the BAS to get a new tier label assigned. By doing this we can keep up a consistent documentation of the format and avoid conflicts between matching labels. The version of the BAS Partitur Format is incremented by one on the third digit whenever a new tier definition is added to it. In accordance to the basic principle this does not imply that any software has to be changed.

**Canonical Pronunciation**

- **Tier label:** KAN
- **Class:** 1
- **Synopsis:** (symbolic links) (transcript)

This tier is the reference tier for all other tiers that use symbolic links. It contains a list of the spoken words within the utterance annotated in extended German SAM-PA (see [SAM, 1989] for a general definition of the SAM-PA and [SAM, 1990] for a special description of the extended German SAM-PA as used in several German projects). Note that these forms are the phonologically expected citation forms, not the actually spoken form.

The segmentation of the whole utterance is done into word units, where everything counts as a word that is produced by the articulatory organs of the speaker and can be seen as speech. Following this definition hesitations are words, whereas laughing, coughs, etc., are not. This separation isn't always clear, but on the other hand the selection of word units is arbitrary as well. The main point here is a unique reference tier for symbolic relations in other tiers. Another problem is the reduction of words that are annotated in the orthographic form, e.g., "mit'm". In these cases the reduction is restituted (in this example /mIt de:m/). The reason for this lies in the fact that some of these reductions should later be automatically accessible.
Example:

KAN: 0 j' a:
KAN: 1 Qalzo:
KAN: 2 QE: m
KAN: 3 h'Ot\v/t:
KAN: 4 Qo: d\v/
KAN: 5 m': O\v/g\v/\v/n

Orthography

- **Tier label:** ORT
- **Class:** 1
- **Synopsis:** (symbolic links) (lexical orthography)

The tier orthography contains the orthographic (lexical) strings corresponding to the units in the tier canonical form. Words are not capitalised at the beginning of an utterance or sentence within an utterance (except nouns of course). German 'Umlauts' and other letters not included within 7 Bit ASCII are written in LaTeX notation. This tier is used for easy lexical access; therefore no additional markers except lexical words are allowed. There is no punctuation in this tier. Lexical words include items that are contained in the KAN tier (e.g. hesitations, repairs, word fragments, etc.). This tier can be used to access customised pronunciation dictionaries, to create unique word frequency lists, etc.

Example:

ORT: 0 ja
ORT: 1 also
ORT: 2 <<ahm>>
ORT: 3 heute
ORT: 4 oder
ORT: 5 morgen

Verbmobil Transliteration - VM I

- **Tier label:** TRL
- **Class:** 1
- **Synopsis:** (symbolic links) (transliteration)

The tier transliteration VM I contains the orthographic transcript of the utterance according to the VM I conventions 3.0. The transliteration is segmented into the units of the tier canonical pronunciation. Therefore multiple references may occur (e.g. if a reduced form of two words is written as one unit in the transliteration). Although especially defined for the German Verbmobil I project, this format has been used in many other resources of spontaneous speech as well. See [Kohler et al., 1994] (German only) or online [Burger, 1995] for a detailed description of the VM I Transliteration format.

Example:

TRL: 0 <A>
TRL: 0 ja ,
TRL: 1 also
TRL: 2 <<ahm>>
TRL: 3 <<#Klicken> heute:
TRL: 4 oder
TRL: 5 morgen .

Verbmobil Transliteration - VM II

- **Tier label:** TR2
- **Class:** 1
- **Synopsis:** (symbolic links) (transliteration)

The tier transliteration VM II contains the orthographic transcript of the utterance according to the VM II conventions. A detailed definition of this format can be found in [Burger, 1997] (German only). In contrast to the VM I format this new updated definition has the advantage of being fully parsable. Furthermore, with this format multi-party and multi-lingual dialogues may be transliterated (because compatible definitions for the languages English and Japanese do exist). To denote overlapping speech parts between different speakers in a dialog, a new tier SUP was defined (see below).

Superimposed Speech - VM II

- **Tier label:** SUP
- **Class:** 1
- **Synopsis:** (symbolic links) (transliteration)

This is a very specialised tier to denote overlapping speech in multi-party recordings. The synopsis of the turn marker and the transliteration is defined for the VM II transliteration format (see above). The speech annotated in this tier stems from a different speaker who actively superimposes his speech on the speech of this Partitur file. See [Burger, 1997] (German only) for a detailed description of superimposed speech in the VM II format.

Example:

TR2: 0 ich
TR2: 1 w"urde
TR2: 2 vorschlagen ,
In this example the utterance of another speaker (AAK, utterance "ja") is superimposed on the 4th and 5th word of the Partitur file (utterance "wir dann").

**Broad Phonetic Segmentation - PhonDat**

- **Tier label:** PHO
- **Class:** 4
- **Synopsis:** (integer) (integer) (list of symbolic links) (label string)

This tier contains a totally time-consuming segmentation into broad phonetic units (extended German SAM-PA). The first number denotes the beginning of the segment in samples counted from the beginning of the speech file; the second number the duration of the segment in samples. The label string contains an additional relation to the canonical pronunciation (aside from the symbolic links to the tier canonical form). The '−' sign denotes differences to the expected canonical pronunciation on a segmental level: a leading '−' sign means the following segment was inserted (e.g. /−a:/); a trailing '−' sign means the segment was deleted (e.g. /−a:/); a '−' sign between segment labels means that the canonical expected segment was replaced (e.g. /−E:/). This tier also contains prosodic and phrasal labelling and segmentation. The full conventions of labelling and segmentation for German are briefly described in [Pompino, 1992] or online in [PHO, 1995].

Example:

<table>
<thead>
<tr>
<th>PHO:</th>
<th>6637</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHO:</td>
<td>6637</td>
<td>916</td>
<td>0</td>
</tr>
<tr>
<td>PHO:</td>
<td>7553</td>
<td>820</td>
<td>0</td>
</tr>
<tr>
<td>PHO:</td>
<td>8373</td>
<td>919</td>
<td>0</td>
</tr>
<tr>
<td>PHO:</td>
<td>9292</td>
<td>870</td>
<td>1</td>
</tr>
<tr>
<td>PHO:</td>
<td>10162</td>
<td>1424</td>
<td>1</td>
</tr>
<tr>
<td>PHO:</td>
<td>11586</td>
<td>724</td>
<td>1</td>
</tr>
<tr>
<td>PHO:</td>
<td>12310</td>
<td>966</td>
<td>1</td>
</tr>
<tr>
<td>PHO:</td>
<td>13276</td>
<td>689</td>
<td>1</td>
</tr>
<tr>
<td>PHO:</td>
<td>13965</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

**Automatic Broad Phonetic Segmentation by MAUS**

- **Tier label:** MAU

A detailed description of the SAP labelling conventions can be found in [Geumann et al., 1997].
This tier contains an automatically generated broad phonetic segmentation in units of German SAM-PA. The segmentation is done fully automatically by the MAUS system ([Kipp et al., 1996]). The segmentation is totally time-consuming and the labelling has no direct relation to the tier canonical form as done in the tier SAP. However, there are symbolic links to the words. The units are labelled in extended German SAM-PA as in the definition of the SAP tier. Additional labels are <nib> (non-speech event) and <p>: (pause). These labels always get the symbolic link -1 (no link).

Example:

MAU/: 0 676 -1 <p:>
MAU/: 677 7861 -1 <nib>
MAU/: 8539 450 0 g
MAU/: 8990 2436 0 u:
MAU/: 11427 1740 0 t
MAU/: 13168 968 1 d
MAU/: 14127 1298 1 a
MAU/: 15426 3820 1 n
MAU/: 19247 303 2 n
MAU/: 19551 1785 2 e:
MAU/: 21337 624 2 m
MAU/: 21962 636 2 n
MAU/: 22699 501 3 v

Dialog Act Segmentation

This tier contains a segmentation in dialog acts according to the ongoing work of the 'Deutsches Forschungszentrum für künstliche Intelligenz' (DFKI), Saarbrücken, Germany. Each marker covers a portion of the speech signal that is denoted by the symbolic links to the reference tier canonical form. A description of the format can be found in [Jekat et al., 1995] or online in [DAS, 1996].

Example:

DAS/: 0,1,2,3,4,5 @m(REJECT_DATE) @m(GIVE_REASON)
DAS/: 6,7,8,9 @(SUGGEST_SUPPORT_DATE)
DAS/: 10,11,12,13,14 @(REQUEST_SUGGEST_DATE)

Prosodic Segmentation - GTobi

This tier contains the prosodic segmentation (by hand) according to GTobi defined by the Technical University of Braunschweig, Germany. A detailed description of the GTobi labelling format can be found in [Grice et al., 1995] or online in [PRB, 1996] (German only).

Example:

PRB/: 54212 5 TÖN: H*; FUN: NA
PRB/: 63269 7 TÖN: L+H*; FUN: EK
PRB/: 76371 8 BRE: B3; TÖN: L-L%
PRB/: 79967 8 TÖN: L+H; FUN: PA

Easy Processing and Distributed Work

Since the Bas Partitur Format is strictly line structured, allows only 7-Bit ASCII and the order within a file does have no semantic meaning, it is very easy to use standard UNIX text processing tools like gawk,
grep or sed to work with data stored in this format. For example the following lines of GAWK code will analyse a stream of piped-in BAS Partitur files for a certain phoneme, capture the total length and summarise into a mean value:

```
/MaU:.*aU$/ {count ++
totallength += $3}
END
{print "Mean Duration for /aU/:");
print totallength/count }
```

In the same manner single BAS Partitur tiers may be selected, updated or filtered using grep, tiers can be easily transformed into format suitable for different kinds of visualising tools (for instance the public domain software package SFS by University College London).

The German Verbmobil project gives a good example for the benefits of using the BAS Partitur format at different sites on the same data. For instance the tiers DAS and PRB were defined by partners at DFKI, Saarbrücken and University of Braunschweig respectively. Since such an extension does not require any basic software to be re-written, these cooperations using the same physical data went very smoothly.

References


[DAS, 1996] http://www.phonetik.uni-muenchen.de/Bas/BasDialogaktDok/vm report-for-partitur_1.html

[PHO, 1995] http://www.phonetik.uni-muenchen.de/Bas/BasFormatsPHOdeu.html

[PRB, 1996] http://www.phonetik.uni-muenchen.de/Bas/BasProsoodie.html


[SAM, 1996] http://www.phonetik.uni-muenchen.de/Bas/BasSAMP A