Ludwig Maximilians Universität München

INSTITUT FÜR PHONETIK UND SPRACHVERARBEITUNG

ESP: Research at IPS

Empirical Speech Data Tools

Florian Schiel
Bavarian Archive for Speech Signals
Institute of Phonetics and Speech Processing
Ludwig-Maximilians-Universität München, Germany

Empirical Speech Data

- Primary: speech signals (*time functions*) + annotation (*symbols*)
  speech signal = *oscillogram, laryngogram, electropalatogram, EMA data, myogram, video of lips/face/body...*
  annotation = *orthographic/phonetic transcript, event labeling, prosody, POS, noise, segmentation of ...*

- Meta: speaker information, recording protocol, (documentation)
  preferably machine readable (e.g. XML or line based)

- Secondary: statistical models, rules, grammars derived from primary speech data, e.g.
  general statistics about events/durations/relations
  Hidden Markov Models, rule sets etc.

Empirical Speech Data

- Primary: speech signals (*time functions*) + annotation (*symbols*)
  speech signal = *oscillogram, laryngogram, electropalatogram, EMA data, myogram, video of lips/face/body...*
  annotation = *orthographic/phonetic transcript, event labeling, prosody, POS, noise, segmentation of ...*

- Meta: speaker information, recording protocol, (documentation)
  preferably machine readable (e.g. XML or line based)

- Secondary: statistical models, rules, grammars derived from primary speech data, e.g.
  general statistics about events/durations/relations
  Hidden Markov Models, rule sets etc.

Empirical Speech Data

Speech Signal + Annotation + Meta = Corpus

Most projects in ESP will probably start with a corpus.

Speech corpus collection needs a lot of effort and is very time-consumig.

Careful design of the corpus can save time (and money).

Corpus Design : www.phonetik.uni-muenchen.de/forschung/BITS/TP1/Cookbook
oder: Chr. Draxler. Korpusbasierte Sprachverarbeitung.
Narr Studienbücher, Gunter Narr Verlag Tübingen, 2008.

If possible, re-use existing speech corpora!

BAS Corpora (1)

www.bas.uni-muenchen.de/Bas/BasKorporaeng.html

All BAS corpora are freely available to ESP members.

- PhonDat 1 : read, phonetically balanced, MAUS
- PhonDat 2 : read, part. phon. segmented, MAUS
- Kielkorpus : read/spont., phon. segmented
- SC1, SC10: read, L2 speakers
- Verbmobil: spont., human-human, MAUS, prosody, German/English/Japanese
- SMARTKOM : spont., human-maschine, MAUS, video, gesture, prosody, emotion
- WebCommand : read, UK Engl./French, MAUS
- RVG-1, RVG-J : read/spont., dialectal, MAUS
- SVC : human-human-maschine, video
- PHATT : read/spont., adolescents, MAUS
- ALC : alkoholized speech, MAUS
- BITS : synthesis speech, prosody, laryngo, phon. segmented
- SIGNUM : German sign language, video
- HEMPEL : monologues, 4000 speakers
- SMC, SC2: real noise, extreme acoustics
- SI100,SI1000 : dictation speech, MAUS
- TAXI : human-human, bilingual
- SI1000P : synthesis speech, laryngo

BAS Corpora (2)

Not in BAS catalogue but also available to ESP:

- VERIDAT : read, speaker verification
- EMA data : (Phil Hoole)
- MRI data : (Phil Hoole)
- HOESI : spont., human-human, Lombard effects
- EPG data : (Marianne Pouplier)
- ERBA : read, train information queries

Other sources:

- ELRA : European Language Resources, catalog.elra.info, universal.elra.info
- LDC : Linguistic Data Consortium, www.ldc.upenn.edu

If you are looking for data, first speak with us!

BAS Corpora (3)

What do you need to do as a ESP member to access BAS?

- sign a form to accept the copyrights of BAS (*Florian Schiel*)
- get an access account at IPS (*Klaus Jänsch*)
- consult the catalogue or talk to IPS staff to find your corpus, then:

```
cd /bmnt/BAS
ls
```

```
ALC             README              RVG1_LQ_7           TASKFORM            VM23.1                  VM39.1  VM59.0
ALC_1.2_1       RVG1_by_labels      RVG1_LQ_8           TAXI                VM24.0                  VM40.0  VM59.1
ALC_1.2_2       RVG1_DVD_1          RVG1_LQ_9           TEMP                VM24.1                  VM40.1  VM60.0
ALC_1.2_3       RVG1_DVD_2          RVG1_LQ_by_dialects tmp                 VM25.0                  VM4.1   VM60.1
ALC_1.2_4       RVG1_DVD_3          RVG1_LQ_by_numbers  TOOLS               VM25.1                  VM41.0  VM6.1
ALC_1.2_5       RVG1_DVD_4          RVG-J_1             VM1.1               VM26.0                  VM41.1  VM61.0
BITS            RVG1_DVD_5          RVG-J_2             VM12.1              VM26.1                  VM42.0  VM61.1
BITS_LG_1       RVG1_HQ_1           RVG-J_3             VM13.1              VM27.0                  VM42.1  VM62.0
BITS_LG_2       RVG1_HQ_10          RVG-J_4             VM14.1              VM27.1                  VM43.0  VM62.1
BITS_LG_3       RVG1_HQ_11          SC1                 VM15.0              VM28.0                  VM43.1  VM63.0
BITS_LG_4       RVG1_HQ_12          SC10_1              VM15.1              VM28.1                  VM44.0  VM64.0
BITS_US_1       RVG1_HQ_2           SC10_2              VM16.0              VM29.0                  VM44.1  VM65.0
BITS_US_2       RVG1_HQ_3           SC2                 VM16.1              VM29.1                  VM45.0  VM7.1
BITS_US_3       RVG1_HQ_4           SI1000_1            VM16.1_1            VM2_english_by_dialogs  VM45.1  VM8.1
BITS_US_4       RVG1_HQ_5           SI1000_2            VM16.1_2            VM2_german_by_dialogs   VM46.0  VMBONUS
DATA            RVG1_HQ_6           SI1000_3            VM17.0              VM2_japanese_by_dialogs VM46.1  VM_DVD_1
...........
```

```
ls HEMPEL
COPYRIGH.TXT   DATA   DOC   INDEX   README   TABLE
```

- Look for `readme` or subdir `doc` or `docu` and RTFM

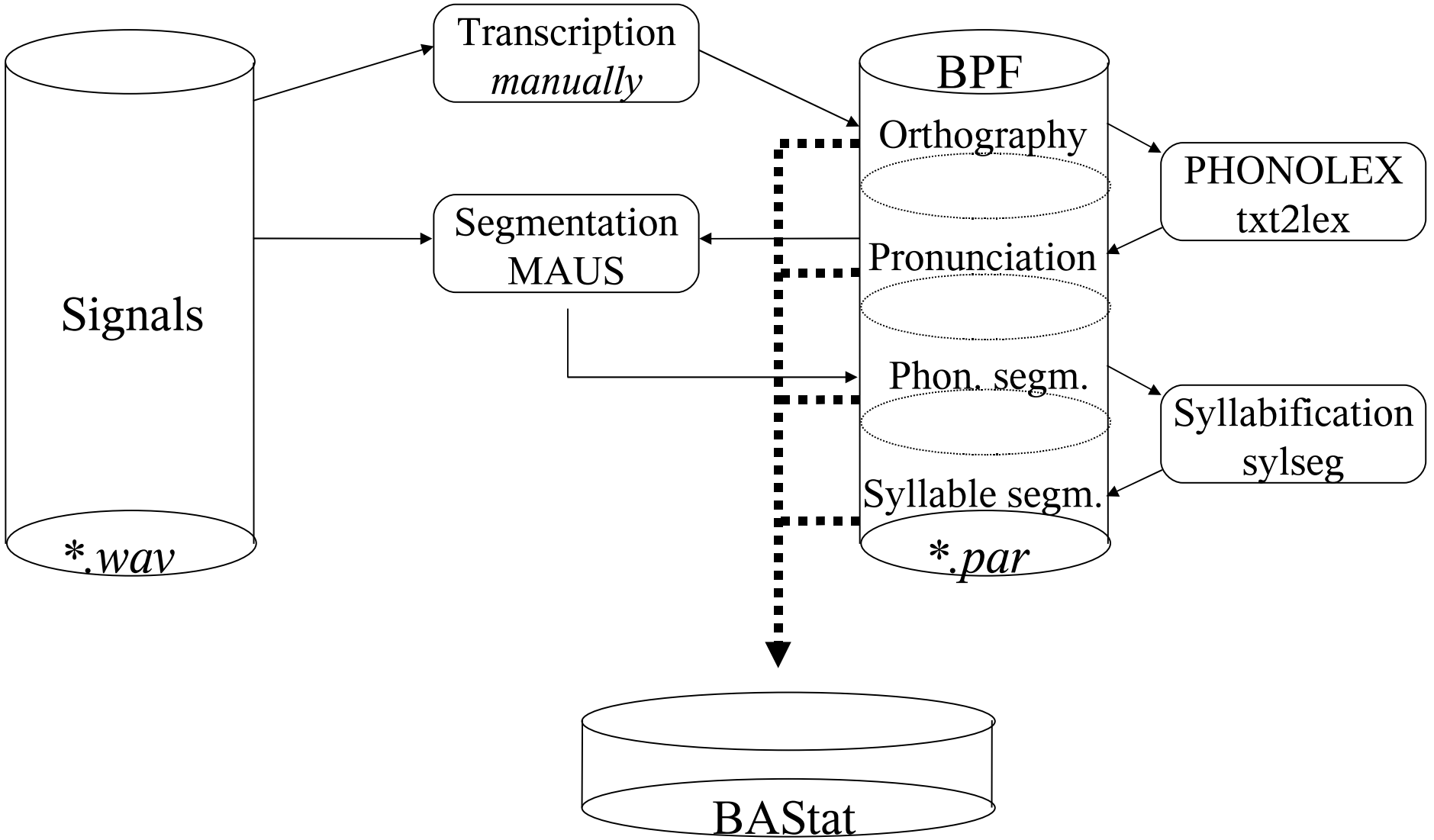
Empirical Speech Data

- Primary: speech signals (*time functions*) + annotation (*symbols*)
  speech signal = *oscillogram, laryngogram, electropalatogram, EMA data, myogram, video of lips/face/body...*
  annotation = *orthographic/phonetic transcript, event labeling, prosody, POS, noise, segmentation of ...*

- Meta: speaker information, recording protocol, (documentation)
  preferably machine readable (e.g. XML or line based)

- Secondary: statistical models, rules, grammars derived from primary speech data, e.g.
  general statistics about events/durations/relations
  Hidden Markov Models, rule sets etc.

BAS Statistics (1)



BAS Statistics (2)

www.bas.uni-muenchen.de/Bas/BasPHONSTATeng.html

BAStat : statistical data based on speech corpora

- phone occurrence (general, word-initial,word-final, ...)
- phone duration statistics
- phone bigrams
- probability of phone sequences (estimates)
- syllable occurrence (general, word-initial,word-final, single-syllable word, ...)
- syllable duration statistics
- syllable bigrams
- word occurrence
- word duration statistics
- word bigrams
- word pronunciation statistics

Ludwig Maximilians Universität München

BAS Statistics – Examples (1)

INSTITUT FÜR PHONETIK UND SPRACHVERARBEITUNG

For an experiment I need to know whether the word onsets /par/ and /kal/ are of equal probability.

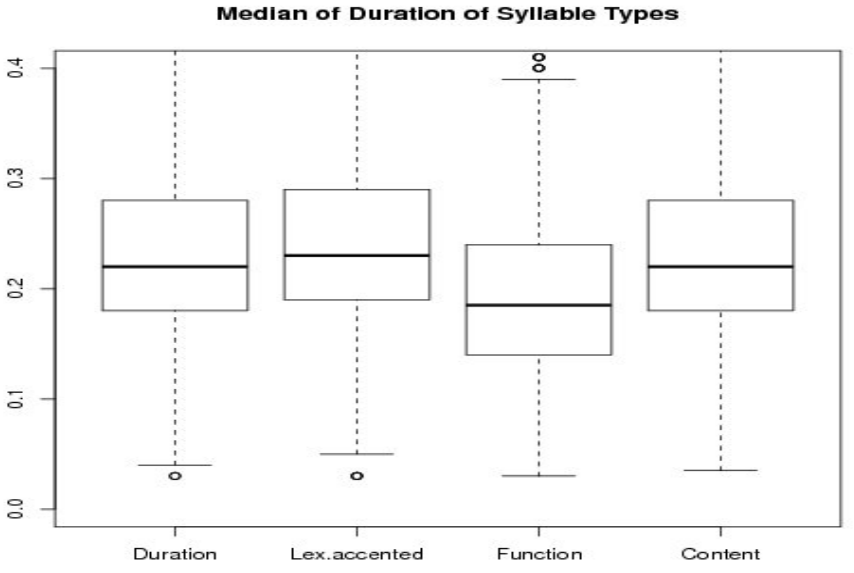
-> www.phonetik.uni-muenchen.de/forschung/Bas/BasPHONSTAT/TOTAL.syl.monogram

P(/par/) * P(WI|/par/) = 1.649544e-04 * 2.176471e-01 = 3.590e-05
P(/kal/) * P(WI|/kal/) = 1.523402e-04 * 9.490446e-01 = 1.445e-04

The probabilities deviate by the factor 4!

My model predicts that function words are produced faster than content words.
-> www.phonetik.uni-muenchen.de/forschung/Bas/BasPHONSTAT/TOTAL.syl.monogram

Syllable durations are significantly lower in function words.



BAS Statistics – Examples (2)

What is the probability of a word-initial /p/ with right-context /r/ (e.g. 'prüfen')?

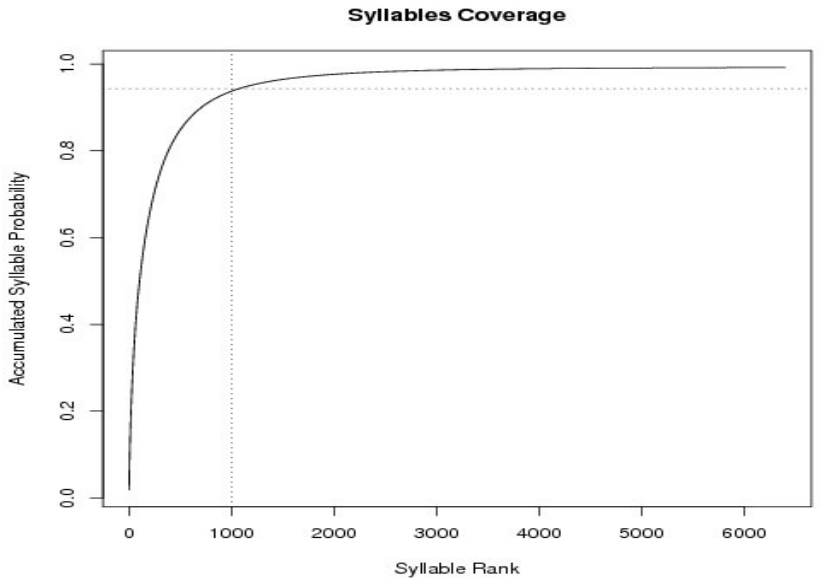
-> www.phonetik.uni-muenchen.de/forschung/Bas/BasPHONSTAT/TOTAL.monogram
www.phonetik.uni-muenchen.de/forschung/Bas/BasPHONSTAT/TOTAL.bigram

*P(col=/r/|row=/p/) * P(/p/) * P(word-initial|/p/) =*
*= 2.249895e-01 * 1.968239e-02 * 2.466251e-01 = 0.00109213*

How many syllables are needed to cover 90% of the language?

-> www.phonetik.uni-muenchen.de/forschung/Bas/BasPHONSTAT/TOTAL.syl.base.rnk

707 syllables



Tools

At IPS several tools are developed and maintained.

Members of ESP are free to use these tools for their projects.

- Emu Speech Database System `Emu` emu.sourceforge.net
- SpeechRecorder www.phonetik.uni-muenchen.de/forschung/Bas/software/speechrecorder
- Munich AUtomatic Segmentation `maus`
- Text-to-Phoneme `txt2lex.pl`
- Speech Detection `wav2trn`
- Munich Speaker Verification System MAVS
- AudioEditorUI www.phonetik.uni-muenchen.de/forschung/Bas/software/audioeditor
- Transcriber Tool WebTranscribe (Chr. Draxler)

Tools Examples (1)

Speech Detector wav2trn

```
% wav2trn wav=file.wav
6615    112015
```

detects begin and duration of speech (in samples) in file.wav

```
% wav2trn wav=file.wav par=file.par
TRN:    6615    112015  0,1,2,3,4    file
```

creates TRN tier

```
% wav2trn mult=1 wav=file.wav
TRN:    13671   266363  -1      0
TRN:    307818  178163  -1      1
TRN:    525231  1450448 -1      2
TRN:    2007873 482894  -1      3
TRN:    2524725 174194  -1      4
```

segments into speech parts

Tools Examples (2)

Text-to-Phonem txt2lex.pl

```
txt2lex.pl -cg -i Arbeitsminister

Arbeitsminister QarbaItsminIst6
```

- produces canonical pronunciation in SAM-PA single words, lists, text
- predicts lexical accentuation
- syllabification phonetic + orthographic
- morph segmentation
- Parts-of-Speech tagging (POS)

Tools Examples (3)

Process a list of words:

```
txt2lex.pl -fgl -i file.txt  >  file.lex

heute    hOYt@
ist      QIst
schönes  S2:n@s
Frühlingswetter fry:lINsvEt6
```

Tools Examples (4)

Münchner AUtomatische Segmentierung maus

```
maus SIGNAL=file.wav KANSTR="QarbaItsminIst6"
   OUT=file.mau
```

produces `file.mau` with phonetic segmentation and labeling

```
MAU: 1324 567  0  Q
MAU: 1892 1255 0  a
MAU: 3148 877  0  r
...
```

Tools Examples (5)

Read canonical pronunciation from BPF file:

```
maus SIGNAL=file.wav BPF=file.par
   OUT=file.mau
```

Create output in praat TextGrid:

```
maus SIGNAL=file.wav BPF=file.par
   OUT=file.TextGrid OUTFORMAT=TextGrid
   INSORTTEXTGRID=yes INSKANTEXTGRID=yes
```

Process a whole speech corpus:

```
maus.corpus SLIST=Corpus.slist BPFDIR=dir
```

Questions?