



## Automatic Analysis of Speech Corpora

Overview, experiences and roadmap

*Uwe Koloska*

- ✘ What is and why do we need automated analysis?
- ✘ First steps:
  - CABSI (Computer Aided Building of Speech Inventories)
  - Application: Building an english inventory
  - Correcting the automatic segmentation of an HMM-Labeler
- ✘ CAASC (Computer Aided Analysis of Speech Corpora)
- ✘ Gathering data for two algorithms to manipulate the rythm of speech
- ✘ Extending the tools and improving the knowlegde



## What is and why do we need *Automatic Analysis*?

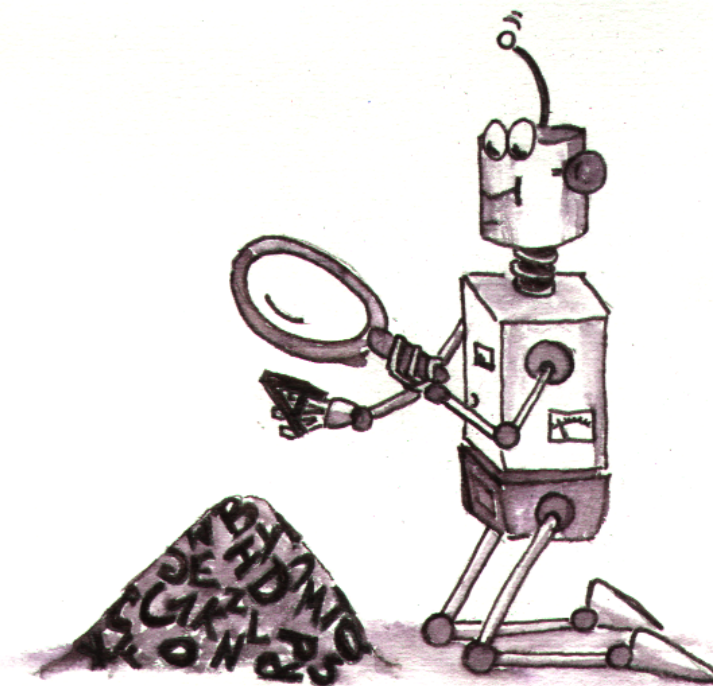
✗ ideal case:

Speech Corpus → Inventory

✗ today: some steps are handmade

✗ speech technology needs data  
many data, more data – everytime

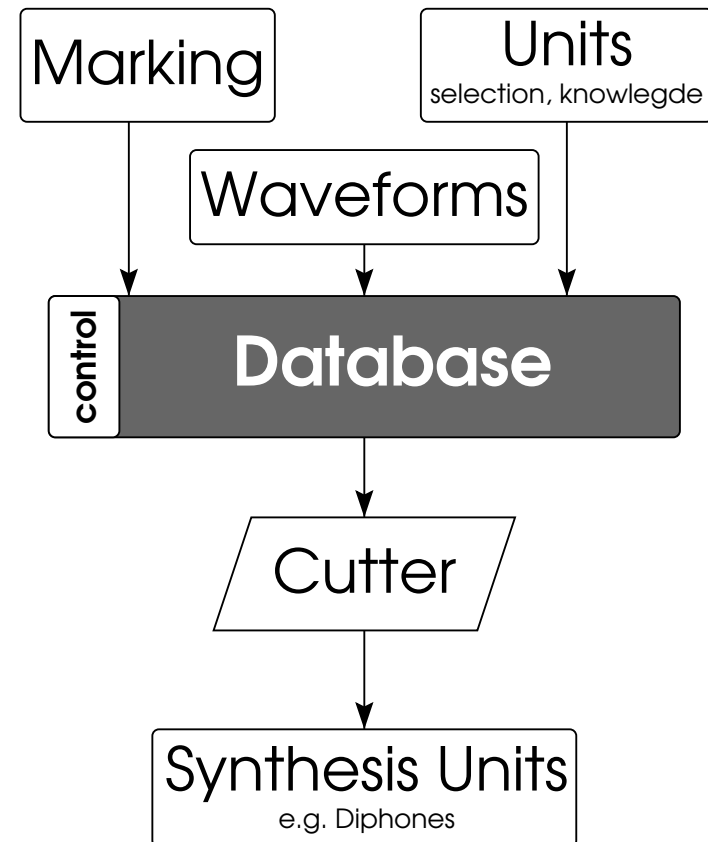
- concatenative speech synthesis needs building blocks
- speech recognition has to be trained
- learnable algorithms need training data





## CABSI (Computer Aided Building of Speech Inventories)

- ✗ SuSaSch helps with the *mechanic work*
- ✗ important property: a fundamental database as basis for all steps
- ✗ marking the waveforms manually
- ✗ but: making the inventory automatically



## Building an english inventory

- ✘ first real world application with **SuSaSch**
- ✘ Input:
  - list of needed units (1449 diphones)
  - recorded speech: one waveform for each unit
- ✘ Problems:
  - input data was not consistent (transcription, list of units)
  - the wanted unit is not included in the speech waveform (because of wrong transcription or pronunciation)
  - some carrier sentences contain a few units but are present more than once

## Correcting some automatic segmentation

### ✘ Input:

- list of needed units: 8521 phones with left and right context
- recorded speech: 936 clippings of radio news

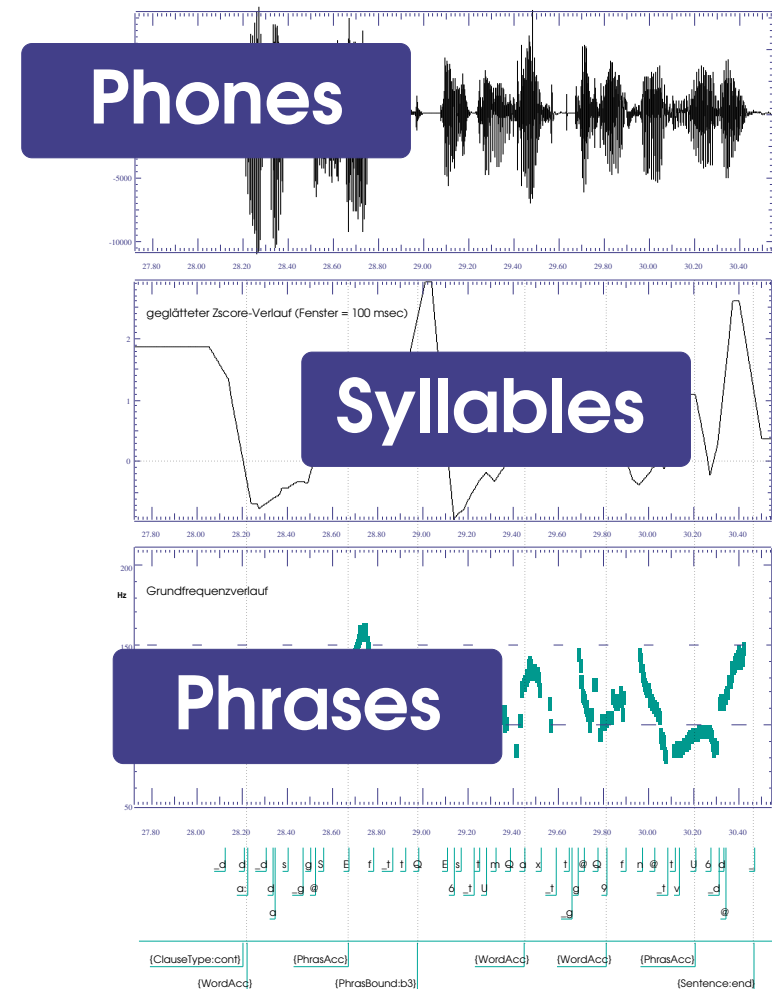
### ✘ Problems:

- foreign words were transcribed natively
- a few words were transcribed as spoken in some dialect but spoken in the standard language



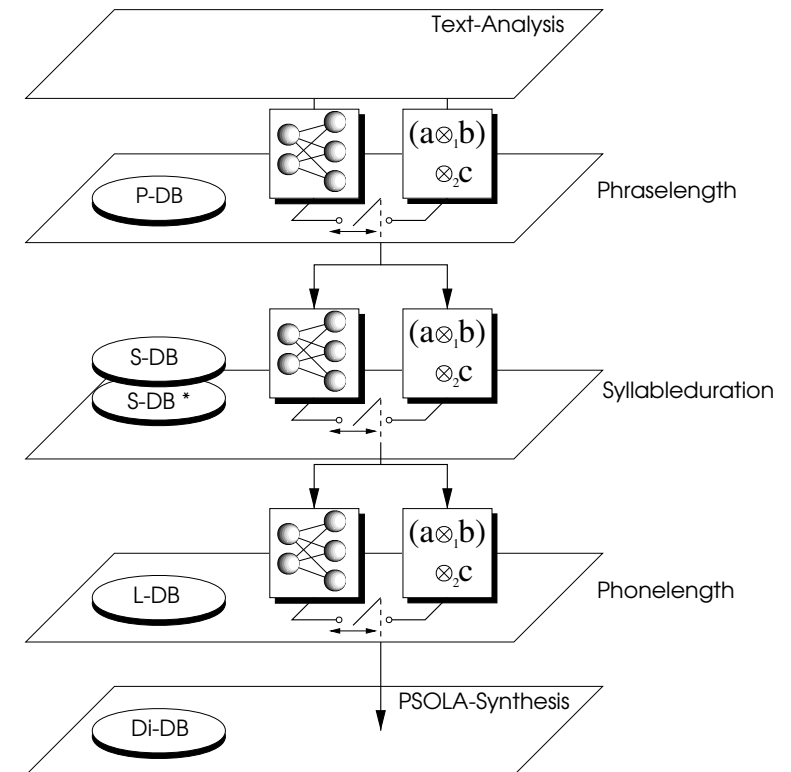
## CAASC (Computer Aided Analysis of Speech Corpora)

- ✗ for now the database consists of three layers: Phones, Syllables and Phrases
  - the phones are labelled manually
  - syllable markers are created automatically
  - phrases are marked semi-automatically
- ✗ compute derived data automatically like: zscore, syllable- and phon-duration statistics



## Gathering data for two learnable algorithms

- ✘ determine the rhythm of synthetic speech by her own speakers characteristics
- ✘ cause the database is in a state of flux, changed data has to be propagated through all layers and to all applications





## Extending the tools and improving the knowlegde

- ✗ the most important task:

**Build an aligner!**

- ✗ improve the data formats
- ✗ work on data structures and build a programming library for gathering and analysing speech data
- ✗ we think about making this library freely available

