Automatic Analysis of Speech Corpora

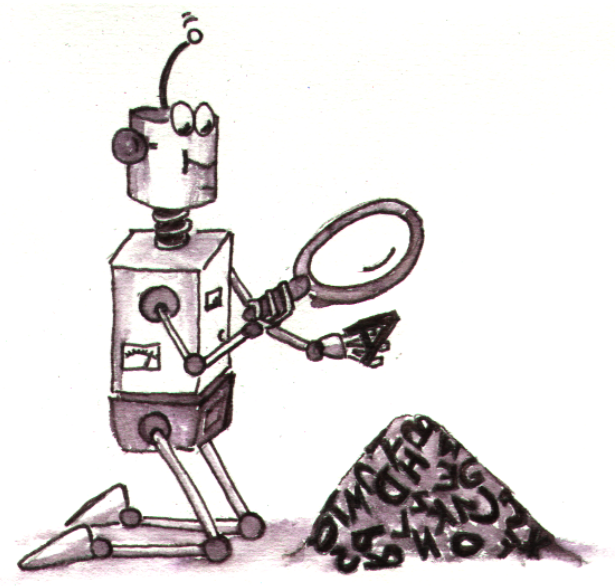
Uwe Koloska

Institut für Technische Akustik
Technische Universität Dresden, D-01062 Dresden
mailto:koloska@rcs.urz.tu-dresden.de

1 What and Why?

Automatic Analysis means having a black box, feeding it with some speech data and got the analysed data, for example an inventory for speech synthesis. Also it means to have all the data in a form readable by a computer and no manual intervention is necessary to use the data for training learnable algorithms or examining and improving rules for rule-based systems.

Look at this robot: He ist *filled* with our knowlegde about analysing und using speech data. Now he can get a mass of speech data and analyse and use it without us.



This paper does not present a complete solution. It presents an overview, first experiences and a way to go. Furthermore this is a working title and rough description of the subject for my doctoral thesis.

2 The first step: CABSI

CABSI means *Computer Aided Building of Speech Inventories* and was first realized with SuSaSch. It helps in making a speech inventory by providing a database with all the needed units and the material from which the units are taken. Because making an inventory means a lot of stupid (and therefore errorprone) tasks, SuSaSch can help to make it consistent and speed up the whole process.

This concept has been proved to be useful by making an inventory for concatenative speech synthesis with 1449 diphones.

3 And now: CAASC

CAASC *(Computer Aided Analysis of Speech Corpora)* is the intermediate step to *Automatic Analysis*. Here we have to manually make availlable all informations, that cannot be provided automatically (Phone- and Phraselabels). Later on we will fill the gaps with automatic procedures.

We have used this step in analysing two speech corpora for training a neural net and extracting rules for a rule-based system. Both algorithms generate a rythmic structure for synthetic speech and because the underlying data is based on the same speech corpus as used by the sythesis system, we get some *personality* for this voices.

4 The future

The most important step from CAASC to *Automatic Analysing* is **building an aligner** that is *precise enough (?)* for making synthesis units.

Also we want to make the data structure and programming library useful in general, so that we can easily build new tools if they are needed and focus on research not programming.

A great part of our work should be spent on building the database(s) and not on developing tools that assist in this task. So we think about making this tools freely available. And hopefully we will receive help and others can benefit – for the sake of research!