TEXT-TO-SPEECH-SYSTEMS FOR GERMAN - A SHORT SURVEY

K. Wothke
Department of Computational Linguistics
Institut für Deutsche Sprache
D-6800 Mannheim

Abstract: A text-to-speech-system can usually be divided into two main components: The first component generates for an orthographic input text its phonetic transcription. The second component maps this phonetic transcription on a sequence of control symbols for a speech-synthesizer and transmits these symbols to the synthesizer, where they trigger the production of synthetic speech. This paper gives a short survey of the first component in several text-to-speech-systems for German which were developed during the last 10 years. First the linguistic problems are outlined. Then the different types of solutions for the realization of the first component are presented and compared. Subsequently some existing systems which transcribe German text are discussed. The last part of the paper deals with the most important unsolved linguistic problems in these systems.

Outline of the problem
The first main task to be performed in a text-to-speech-system is the fully automatic conversion of an orthographic input into its phonetic transcription. Suppose the input-sentence

"The vocal cords are vibrating at about 100 Hz now."

is fed into a text-to-speech-system for English. Then this system at first has to generate the transcription:

"əʊ ˈveɪkəl kɔːdə ə vaiˈbrei disp əˈbaut ˈəndrəd əhəːts nau"

This example demonstrates some of the problems to be solved during the automatic transcription: The system has to transcribe not only orthographic words but also abbreviations and (sequences of) digits. It must also recognize the word stress and mark it in the transcription. Furthermore there are some problems not illustrated by the example:

- transcription of special characters (%, $, § etc.),
- determination of pauses of different length,
- determination of sentence stress.

Types of solution
Most of the existing solutions for the problem of automatic transcription can be divided into three main steps. Especially the second step is realized in very different ways.
The *first step* is a preprocessing of the orthographic input. This preprocessing scans the text for the occurrence of digits, abbreviations and special characters, and it replaces them by the corresponding orthographic words. In the systems for German these orthographic words are usually determined:
- with a rule based method for the (sequences of) digits
- with a dictionary look-up for the abbreviations and the special characters.

The *second step* transcribes the preprocessed text on the segmental level (suprasegmentals are usually not considered in this step). Three alternative solutions exist for the transcription on the segmental level: The dictionary based method, the rule based method and a combination of these two methods.

Dictionary based systems use a dictionary whose entries have the structure

```
| orthographic word | phonetic transcription |
```

In order to transcribe an orthographic text these systems look up its words in the dictionary and replace them by the phonetic transcriptions found in the dictionary. This method guarantees that the system can always decide if a word is correctly transcribed, i.e. found in the dictionary. This is an advantage the rule based systems do not possess. A substantial drawback of the dictionary based method is that it can only transcribe those words of the text which are also entries in the dictionary. Thus this method does not take into account the fact that each natural language is an open system. The dictionary will never be complete. There are two further disadvantages of the dictionary based method:
- The creation of the dictionary requires a lot of work.
- The storage requirement does not permit implementation on a micro-computer.

A *rule based system* applies a set of transcription rules to the letters or letter sequences of the input text in order to transcribe them. Usually these rules are context sensitive, i.e. they transcribe the letters of the orthographic text dependent on their graphemic context. It is important to note that the term "rule" does not denote any linguistic regularities which exist between the orthographic representation and the phonetic transcription of a language. "Rule" denotes an instruction formulated in a formal language as to how to replace (strings of) letters by (strings of) phonetic symbols. The type of rule mostly used in the rule based systems for German is

```
u \leftrightarrow v \rightarrow x.
```

Read: If the letter (or string of letters) \(v\) occurs in a word on the right side of the letter/string of letters \(u\) and on the left side of the letter/string of letters \(w\), then replace \(v\) by the phonetic symbol(s) \(x\). \(u\) and \(w\) can also be missing.

Usually a set of rules is ordered: The system tries at first to apply the more specific rules to a given letter(-sequence). If these rules cannot be applied, the more general rules are applied.
- In contrast to the dictionary based method the rule based method takes into account that each natural language is an open system. The rule based method can potentially transcribe each word of an
input text. Further advantages are:
- A set of transcription rules can usually be created with less
effort than a dictionary.
- Rule-based systems usually need less storage than dictionary
based systems. Therefore they can also be run on micro-computers.

Rule-based systems have the drawback that they cannot decide
whether a word was correctly transcribed. An important subclassi-
fication of rule based systems can be made on the level of their
software architecture. The differences in the software architec-
ture, which lead to this subclassification, have strong effects on
the course of the development of a rule based system and on its
transparency and changeability. Rule based systems can be subclassi-
fied into so-called programmed systems, interpreters, and compi-
lers. In a programmed system the rules are formulated as statements
of a programming language. The rules are incorporated into a source
code. Interpreters read the rules from an external data set and
apply them to the orthographic text. The rules have been written to
the data set by a programmer or linguist in advance. For the nota-
tion of the rules the linguist can employ a problem oriented formal
language (the usual representation of the rules is: u \langle v \rangle \rightarrow w -> x).
Compilers first read the rules from an external data set where they
were written by the user. Then they generate a load module, which
can execute the transcription process described by the rules. When
this load module is started, it reads the orthographic text from a
data set and generates the corresponding phonetic transcription.

- Programmed systems are generally faster than interpreters. Since
the rules are formulated as statements of a programming language
they are often not very transparent and it is quite difficult to
alter them. Furthermore, each change of the rules requires a new
compilation of the source code. If an interpreter is used, it is
much easier to alter the rules because they are formulated in a
problem oriented notation and therefore more transparent. A change
of the rules does not require a change and a new compilation of the
interpreter program. An important advantage of an interpreter (and
also of a compiler) is that it can be used for the automatic tran-
scription of texts from different languages. The linguist must only
provide sets of rules for the particular languages in an external
data set. A compiler has the advantage that the rules can easily be
changed and that the generated load module usually is faster than
an interpreter. It may be considered as a drawback that each change
of the rules requires their new compilation.

The third type of solution for the transcription of orthographic
texts on the segmental level consists in a combination of the dic-
tionary based method and the rule based method. Some of these sys-
tems try to compensate the incompleteness of a dictionary based
system by adding a rule based component: Each orthographic word
which cannot be found in the dictionary is transcribed by the
application of the rules. Thus especially neologisms, which are
often not in the dictionary, can also be transcribed. On the other
hand, there are primarily rule based systems which try to prevent
the set of rules from being overloaded with descriptions of excep-
tions by using a dictionary which contains words with irregular
phonetic transcriptions.

The third step in the automatic transcription of orthographic text
deals with the suprasegmentals. This step tries to insert into the
text marks for pauses of different length, marks for the word
stress and marks for the sentence stress. In the existing text-to-
speech-systems for German this step shows the most problems. In
particular, the automatic identification of the sentence stress is
still very defective. An established paradigm of possible solutions does not exist for this step. A remarkable study on the treatment of suprasegmentals in text-to-speech-systems for German was carried out by Zingle (10).

Existing systems for the transcription of German text
In the following paragraphs some selected systems for the automatic transcription of German text will be presented.

A dictionary based system was developed by Chisholm (4). (This system was not used as a part of a text-to-speech-system, but for the transcription of large corpora of German verse as a preparation for a computer-assisted phonological analysis of German verse.) The dictionary comprises about 25,000 entries, i.e. orthographic words with their transcriptions (including word stress). The system correctly transcribes about 95% of the words (= tokens) of a text. Sentence stress and pauses are not taken into consideration. Further noteworthy dictionary based systems for German do not exist.

One of the first rule based systems for the automatic transcription of German text was developed in 1975/76 by Berry-Rogge (1) at the Institut für deutsche Sprache/Mannheim. Later, this system, named PHONOL, was improved by Breuer/Brustkern/Thyssen/Wille (3) and it became part of a text-to-speech-system. PHONOL is a programmed system. The system comprises:
- a preprocessing of the input (substitution of digits by the corresponding written words),
- the phonetic transcription on the segmental level,
- and the determination of the word stress and of pauses of different length.
Sentence stress is not considered by the system. PHONOL correctly transcribes 98% of the running words (= tokens) of a German text.

A rule based interpreter system was developed in the years 1977-1980 by Slaby and Spellmann (8, 9) at the University of Münster. The system consists of two steps: In the first step all digits in the text are replaced by the corresponding orthographic words. The second step transcribes the preprocessed text on the segmental level. For this process a very voluminous set of 13,370 rules is used. Suprasegmentals are not considered by the system. A statistical evaluation of the linguistic adequacy of the system was not carried out.

There exist several noteworthy compiler-systems which have been developed during the last years. Two of them are presented here.

At the Gesellschaft für Mathematik und Datenverarbeitung/Bonn B. S. Müller (5) developed a system called REDE. Compared to the other existing systems REDE takes into account most of the linguistic problems which must be considered during the automatic transcription of German text. The transcription is performed in eight steps:

1. Preprocessing of the orthographic text (abbreviations, digits, and special characters are replaced by the corresponding orthographic words).
2. Identification of unstressed words. This step is performed by looking up the words of the text in a list of words which are normally unstressed.
3. Identification of morpheme boundaries. The morpheme boundaries are recognized by scanning the words for sequences of letters which usually occur only at the end or at the beginning of a morpheme. Of all systems presented here, Müller’s system is the only rule based system which tries to recognize morpheme boundaries.
4. Identification of morphemes should be obligatory in a text-to-speech-system for German since in German words
many sequences of letters must be replaced by other phonetic symbols if they are part of one morpheme as opposed to occurring on the boundary of two morphemes (cf. 'Busch' vs 'Werks-chor').

4. Identification of prefixes and suffixes which exert an influence on the stress of the stem of a word (e.g. 'einsetzen' vs 'versetzen').

5. Marking of the word stress dependent on the results of steps 2 and 4.

6. and 7. Transcription of the vocalic letters.

8. Transcription of the consonantal letters.

As this sequence of steps illustrates, the processes involved in the automatic transcription of German text are not always executed in the order which was described under 'Types of solution': In the system REDE the suprasegmentals are treated before the transcription on the segmental level is carried out. An experimental version of Müller's system also included an additional component for the identification of sentence intonation. - Müller does not present any statistical evaluations of the linguistic capacity of his system.

Another compiler-system, named RULBOL, is being developed by Bierfert (2) and Stock (5) at the University of Bonn. Up to now the main interest in the development of this system has been the creation of a comfortable formal language for the notation of the transcription rules and the construction of a compiler for sets of rules written in this formal language. The existing set of rules for the transcription of German text is only a provisional one. In the past this set of rules has mainly been used to test the compiler during its development. The most interesting aspect of RULBOL is the formal language for the formulation of the rules. Some features of this language may be illustrated with examples from the provisional set of rules for German:

- \(<\text{Vb}\), 1, r, G \([<\text{Vl}\), L, R] -> g

Read: If the letter 'G' occurs on the right side of a phonetic vowel symbol ('<\text{Vb}>') or the phonetic symbols '13' or '7' and if immediately to the right of 'G' is a vocalic letter ('<\text{Vl}>) , the letter 'l' or the letter 'r' then substitute 'g' by the phonetic symbol '[g]'. - Note that this rule expects the left context of the letter 'G' already in a phonetic transcription.

- Signifiers for the sets of letters or sets of phonetic symbols (cf. '<\text{Vb}>' and '<\text{Vl}>' in the rule illustrated above) are defined by the user at the beginning of a set of rules. The signifier '<\text{Vl}>' used in the rule above is defined with the instruction

<br />

- G [\((<\text{KL}>; R, L), #] -> k

Read: Replace the letter 'G' by the phonetic symbol '[k]' if 'G' occurs on the left side of a consonantal letter ('<\text{KL}>) , except ('; 'r' or 'l', or if 'G' is immediately followed by a word boundary (#).

- [0] AB [-(ER)] -> 'ap

Read: Replace the letter sequence 'AB' by the sequence of phonetic symbols '[ap]' (+ symbolizes the glottal stop), if 'AB' is at the beginning of a word (#) and is not (-) followed by the
letter sequence 'ER'.

RULBOL seems to be an excellent tool for the development and automatic transcription of sets of rules for various transformations of strings.

Unsolved linguistic problems

A process which requires many improvements is the identification of word stress and of sentence stress. Solutions exist only for standard stress, emphases which deviate from the standard-stress cannot at present be recognized automatically. For their identification at least a rudimentary semantic and pragmatic analysis of the input text would be necessary.

- Problems also occur when (sequences of) letters must be transcribed which occur at morpheme boundaries. For their correct transcription it is often necessary to know where the morpheme boundaries are and therefore to segment the words of the orthographic text into their morphemes (cf. 'Buchstabe' vs 'Wachstum'). None of the existing text-to-speech-systems for German contains a component which performs the necessary segmentations in a sufficient way.

- Since none of the existing systems does an adequate syntactical and semantic analysis of sentences, words which have the same spelling but are differently pronounced cannot be differentiated (c.f. 'Montage' with the pronunciations ['mɔnta:ʒ], which means mounting, and ['mo:nəʒ], which means mondays).

Despite all these problems the quality of some existing text-to-speech-systems for German seems to be sufficient for several applications (e.g. reading-aid for blind secretaries/typists who work at an electronic typewriter, acoustic output at computer terminals).

References


Ankündigung für das LDV-FORUM 1/86
ERFAHRUNGSBERICHT ÜBER DAS KONKORDANZ-
PROGRAMM BYU-CONCORDANCE
FÜR ARBEITSPLATZRECHNER

J. Brustkern
Institut für Kommunikationsforschung
und Phonetik
D-5300 Bonn