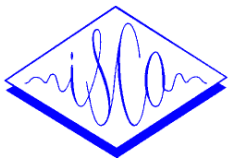

The ESCA Workshop
on Speech Synthesis
Autrans, France
September 25-28, 1990

FROM ORTHOGRAPHY TO PHONETIC TRANSCRIPTION IN THE GERMAN TEXT-TO-SPEECH SYSTEM TETOS

Klaus Wothke

IBM Scientific Center
Heidelberg, West Germany

ABSTRACT

A text-to-speech system was developed which reads aloud unrestricted German texts. It converts orthographic input text into synthetic speech in 3 main steps: First the words of the text are preprocessed and abbreviations, special characters, and digits are replaced by their full orthographic correlates. Then the preprocessed words are phonetically transcribed by means of about 1,370 letter-to-phone rules. Finally the phonetic transcriptions are mapped on sequences of control codes for a speech synthesizer. The rate of incorrect conversions of written words to phonetic transcriptions is near 2% of the running words of a text.

1. INTRODUCTION

This paper deals with some results of a project carried out between 1985 and 1988. The aims were:

- Development of a text-to-speech system which reads aloud unrestricted German text with good acoustic quality.
- Design of a system architecture which allows an easy modification of the linguistic knowledge.
- Separation of the software from the language specific data, in order to permit the use of the software for other languages.

The outcome of the project is the TETOS system.

2. ENVIRONMENT OF TETOS

The environment of TETOS is illustrated in fig. 1. TETOS has 4 input files:

- A dictionary containing abbreviations and special characters with their full orthographic correlates.
- A file with letter-to-phone rules.
- A file with mappings of phone symbols on control codes for the synthesizer.
- An orthographic German text to be converted into speech.

By means of the linguistic knowledge in the first 3 files TETOS maps an orthographic German text on a sequence of control codes for the commercial speech synthesizer SynPhonix 200[1].

Each of these files exists in 2 representations: An external representation, which permits an easy modification of the linguistic knowledge by means of an editor, and an internal representation which is no longer human readable, but which is more

1 SynPhonix 200 is a trademark of Artic Technologies.

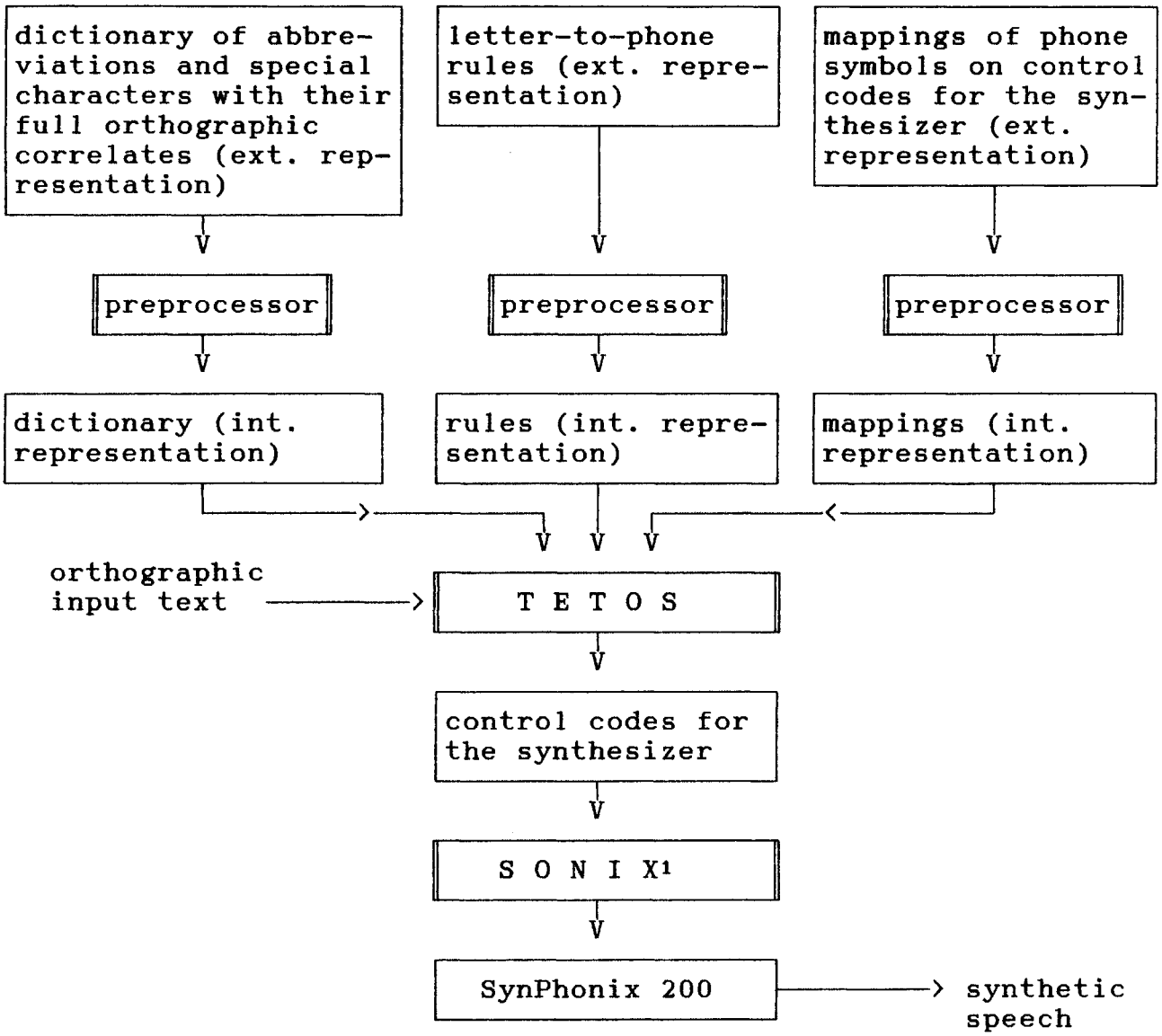

Fig. 1. Environment of the Text-to-Speech System TETOS.

suitable for the use of the linguistic knowledge by TETOS. The internal representations are generated from the external representations by preprocessors, which also check the external representations for formal errors.

Almost all language specific data are in the first 3 input files. Only the knowledge necessary for the conversion of digits to their full orthographic representation had to be integrated into program code.

Each time TETOS generates a sequence of control codes for an orthographic word, it sends this sequence to SONIX, which is an interface between the synthesizer SynPhonix 200 and TETOS.

TETOS and the preprocessors are written in the programming language Turbo C[2]. The system runs under MS DOS[3].

1 SONIX is a trademark of Artic Technologies.

2 Turbo C is a trademark of Borland International, Inc.

3 MS DOS is a trademark of the Microsoft Corporation.

3 LANGUAGE SPECIFIC KNOWLEDGE FOR GERMAN

3.1 Abbreviations and Special Characters

Before abbreviations and special characters are phonetically transcribed and mapped on control codes for the synthesizer, they are replaced by their full orthographic correlates. For this process TETOS uses a dictionary of about 1.800 abbreviations and special characters with their orthographic correlates.

3.2 Digits

Also digits are first replaced by their full orthographic correlates. They are subclassified into the 4 classes of whole cardinal numbers, real numbers, dates, and others. Depending on the classification the full orthographic correlate is generated.

3.3 Letter-to-Phone Rules

By means of the letter-to-phone rules TETOS maps all orthographic strings on their phonetic transcription. The letter-to-phone rules are applied to orthographic strings which either occur literally in the orginal text or which are the full orthographic correlates of abbreviations, special characters, or digits.

Currently about 1,370 letter-to-phone rules are used. They were created in the following way: We started with a set of about 200 rules created by A. Günther [published in Bierfert, 1985: pp. 316]. This initial set was tested against the entries in the German pronunciation dictionary "Groβes Wörterbuch der deutschen Aussprache" [Krech, 1982] and iteratively improved. A pronunciation dictionary does not contain any inflected words; therefore the transcription of inflectional endings was tested by applying the rules to German texts. This resulted in a further improvement of the rules.

The formalism for the formulation of the rules (= external representation) consists of 2 main elements:

Set Definitions

A collection of letter strings which are often used together in the rules may be combined into a set which is designated by a set identifier. Instead of the concrete letter strings the set identifier may then be used in the rules. - A set definition has the general format

+SET X={x_1 x_2 ... x_n} + ...

where X is the set identifier (a capital letter), each x_i is a string of lower case letters, and '+ ...' is a comment.

Rules

A rule has the general format

X[Y]Z=>W + ...

Read: Replace the letter string Y by the phone string W, if Y occurs right of the letter pattern X and left of the letter pattern Z ('+ ...' is a comment).

Some important conventions:

- X and Z may contain strings as well as set identifiers.
- In X and Z the characters @ and # may be used: @ represents the set of all characters. # represents the word boundary.
- A set identifier may be followed by the restricted Kleene operator. Its format is *n ($1 \leq n \leq 5$). For arbitrary n A*n always implies A*0. Example: 'A*2' means 'AA', 'A', and ''. Thus the following 2 rule formulations are functionally equivalent:

		bAA[a]B=>a:
bA*2[a]B=>a:	<====>	bA[a]B=>a:
		b[a]=>a:

- The rules are ordered: For a given letter their application is tried sequentially according to their order in the rule file.
- Object language strings are in lower case letters. Capital letters are set identifiers.

To illustrate how letter-to-phone rules can be expressed in the formalism introduced above, an excerpt from the German phonetic-ization rules for the letter 't' is listed in fig. 2.

```
+SET B={a o u}                      + Letters for back vowels
+SET C={b c ch d f g ... z β}       + Consonantal letters
+SET E={e em en es er ern ns}       + Nominal endings

[tie]llE*1#=>tsjЄ                  + 'essentiell' etc.
k[t]ie=>ts                          + only: '-aktie/n-'
@[t]ien=>ts                         + 'kroatien', 'helvetien' etc.
[t]ii=>ts                           + 'initiieren', 'negotiieren' etc.
[ti]B=>tsj                          + 'exerzitium' etc.
[t]her#=>t                          + 'dorther', 'seither' etc.
[th]=>t                             + STANDARD
C[tt]=>tt                           + 'bartträger' etc.
[tth]=>t                            + 'matthias' etc.
[tt]=>t                             + STANDARD
[tz]=>ts                            + STANDARD
[t]=>t                              + STANDARD
```

Fig. 2. Example for Letter-to-Phone Rules

3.4 Mappings of Phone Symbols on Control Codes for the Synthesizer

When the phonetic transcription of an orthographic word has been generated, it is mapped on control codes for the synthesizer. For this task TETOS accesses a file containing the control codes for each phone symbol. Currently the file comprises 76 mappings, which are partially context sensitive.

4 EVALUATION

TETOS converts unrestricted German texts into intelligible synthetic speech. About 6% of the different words and about 2% of the running words of a text are incorrectly phonetically transcribed, but most of these erroneous transcriptions do not lead to a prominent deviation in the synthetic speech.

The transcription errors are mainly a consequence of the fact that the system does not have any information of the morph structure of the words. In German the phonetic transcription of a letter is not only dependent on the letter itself and its orthographic context, but also on its morphological context. The absence of information about the morph structure causes e.g. errors of the following type: When the letters 'b', 'd', 'g' occur in a compound word at the end of a morph and in front of a lexical morph or a prefix beginning with a vocalic letter, they are incorrectly mapped on the voiced plosives [b], [d], [g] instead of [p], [t], [k], respectively. Furthermore, the glottal stop symbol is missing when the initial letter of the following morph is a vocalic letter. Example: 'Mittagessen' (engl.: lunch) was incorrectly transcribed as ['mita:gЄsn]* instead ['mita:k?Єsn]. A German text-to-speech system which takes into account morph boundaries is RELEXIS [cf. Wolf/Strecker/Fries, 1989].

A second reason for transcription errors is that ordinal numbers are not recognized as such. They are pronounced literally. E.g. '3' in 'der 3. Advent' (engl.: the 3rd Advent) is mapped on the orthographic correlate 'drei' (engl.: three) instead of 'dritte' (engl.: third).

A further restriction of TETOS is that it does not currently take into account the prosodic features.

Two important features of TETOS result from the fact that most of the linguistic knowledge is separate from the software and that it is represented in mnemotechnically favourable formalisms: The linguistic knowledge can be modified easily without changing the software, and TETOS can be adapted to other languages almost without any modifications of the software.

REFERENCES

Bierfert, H. (1985): Automatische Spracherkennung - von der Theorie zur Praxis. Tübingen.

Krech, E.-M. et al. (eds.) (1982): Großes Wörterbuch der deutschen Aussprache. Leipzig.

Wolf, H.E./Strecker, F./Fries, G. (1989): Text-Sprache-Umsetzung für Anwendungen bei automatischen Informations- und Transaktionssystemen. In: Informationstechnik 31(1989)5. PP. 334-341.