A stand-alone text-to-speech system

R.J.H. Deliege

Abstract

A prototype of a general-purpose stand-alone text-to-speech system has been built that converts text in normal Dutch orthography into speech. This paper briefly describes the features of this system.

Introduction

In many places, research is going on in the field of text-to-speech conversion, either to tackle a fundamental research topic or to develop a potential product. This project is in the last category and aims to realize a stand-alone text-to-speech system. Work on this project is supported by the national research programme 'Analysis and Synthesis of Speech'. One of the aims of the programme is to transfer available knowledge to industry by constructing prototypes. The aim of the project, therefore, is to combine various sources of expertise into a working product. Available knowledge does not necessarily come from IPO. For the system described here, the grapheme-to-phoneme conversion and the accentuation rules are provided by Nijmegen University (Kerkhof, Wester & Boves, 1984). The speech synthesis is based on the IPO diphone-concatenation method (Elsendoorn & 't Hart, 1982; Elsendoorn, 1984).

The system is quite similar to the multilingual text-to-speech system (Van Rijnsoever, 1988) that operates on the VAX computer at IPO. Besides the hardware, the main differences are that the stand-alone system only works for Dutch texts and that it does not contain parallel routines for the same task. It rather contains an optimal path through the VAX-based system in the sense that it uses those routines that are most developed or best suited to implementation in a microprocessor system.

This system can be used anywhere where spoken output is wanted. Applications include output for persons who are disabled, such as the blind or the speech impaired, or for persons who are too busy to look at a message, for instance when driving a car.

Description of the system

The basic structure of the system is illustrated in Figure 1. The orthographic input in ASCII format is fed into the system through a serial connection, either with a computer terminal or with a text-generating system. A second serial connection is available for two special functions. The first is the incorporation of the system into an existing terminal link to a computer, thus providing the last named with spoken output. The second is the possibility of generating speech data that can be stored externally, for example in a computer or a memory chip. In this way, the system can be used as a

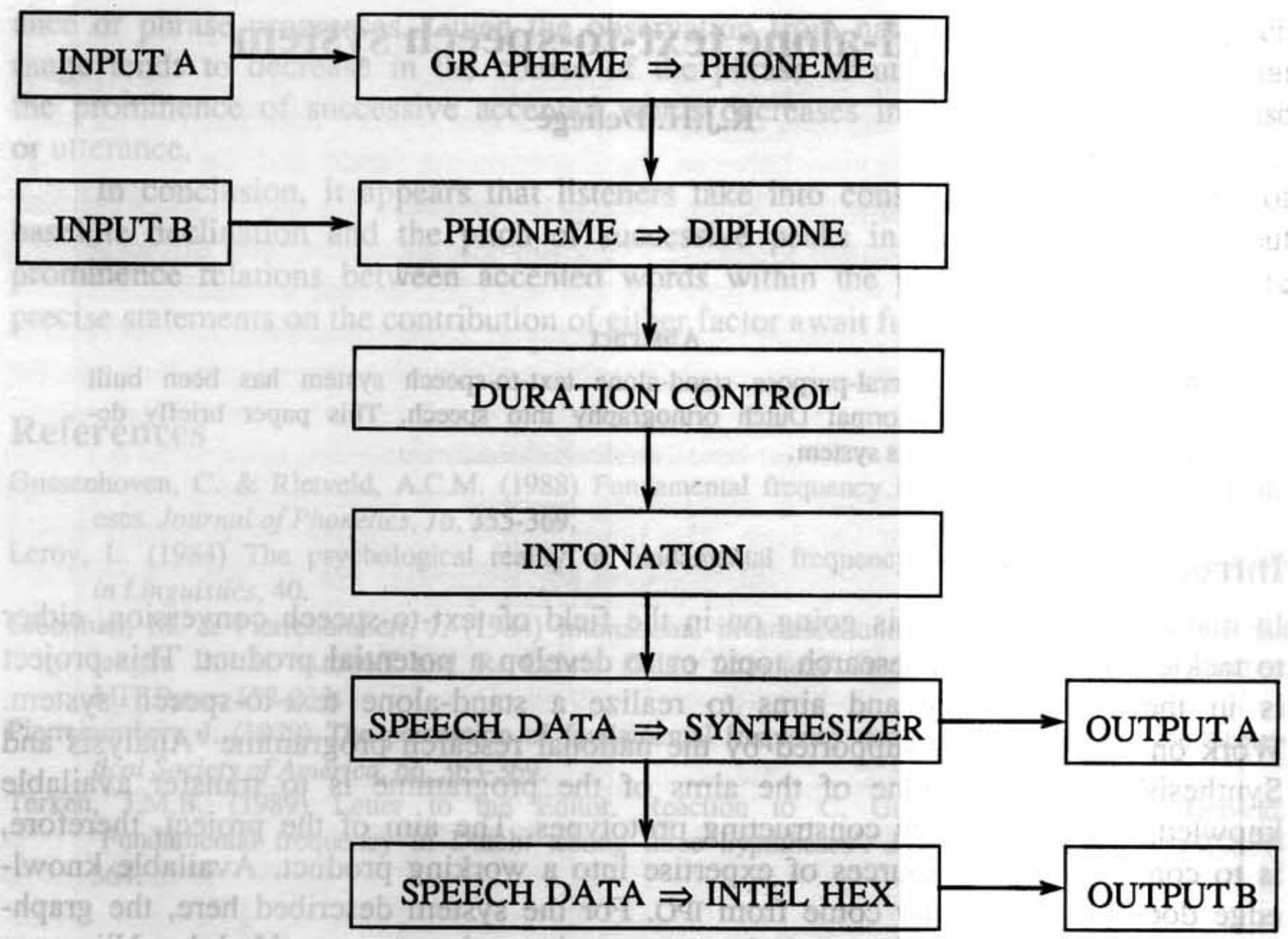

Figure 1: The text-to-speech conversion.

speech-programming device. In a later version, the system will be equipped with its own keyboard and visual display, in order to become really 'stand-alone'.

The input can consist of either text or commands. Commands are preceded by the special symbol /.

The input text can be interpreted in two ways, viz. as orthographic or as phonetic spelling. This interpretation can be selected by means of appropriate commands. Thanks to the phonetic-input option, it is possible to use the system in combination with other letter-to-sound converters than the one incorporated into the system.

The orthographic input can be enriched by user-supplied accentuation markers, '. In the absence of such markers, the system itself provides word stress and sentence accents. Sentence melody (intonation) is entirely computed by rule. A few duration rules are included, e.g. the lengthening of clause final syllables.

The system has an exceptions lexicon that can contain the correct pronunciation of irregular forms, e.g. foreign names.

The following facilities are currently available to the user:

- a memory for the last nine sentences;
- a permanent memory for nine sentences;
- a screen editor;

- an editor for the exceptions lexicon;
- upload and download facilities for the exceptions lexicon;
- a software-adjustable audio volume;
- an adjustable talking speed;
- the simulation of a female voice.

Hardware implementation

The hardware of the system consists of a 68000 microprocessor, 512 kbyte EPROM, 64 kbyte RAM with battery backup, a Philips speech synthesizer (PCF8200), a double RS232 interface, all in CMOS, and a simple audio amplifier (TDA7052). Figure 2 shows a prototype of the system.

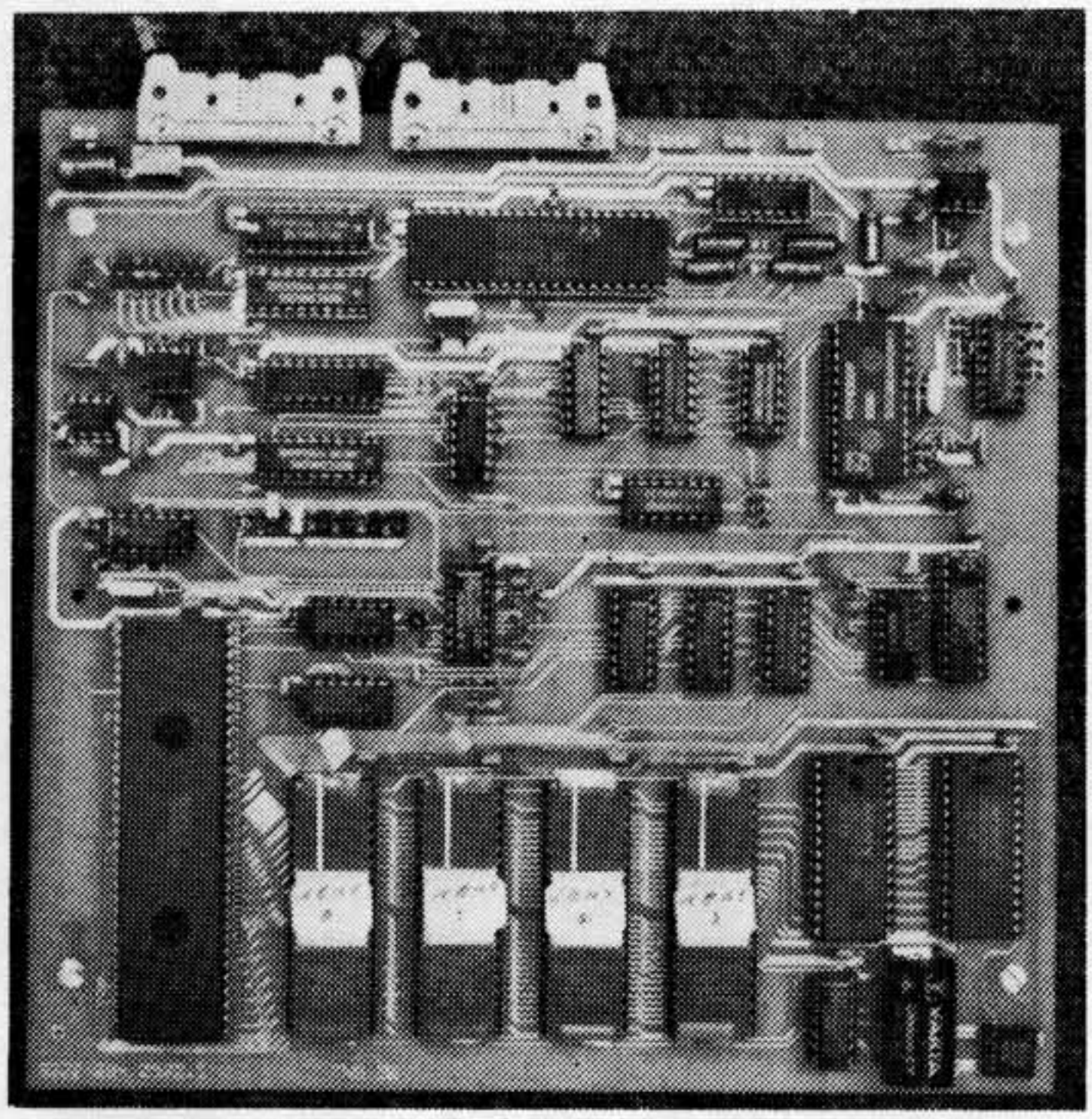

Figure 2: The stand-alone text-to-speech system.

Future plans

A printed-circuit board is being developed for this system, so that more copies can be built easily. In the future we will investigate the implementation of better prosody algorithms that have recently become available (Quené & Kager, 1989). Finally, there are plans to investigate the implementation of languages other than Dutch, namely British English and German, for which expertise is available at IPO.

References

Elsendoorn, B.A.G. & Hart, J. 't (1982) Exploring the possibilities of speech synthesis with Dutch diphones. *IPO Annual Progress Report, 17*, 63-65.

Elsendoorn, B.A.G. (1984) Heading for a diphone speech synthesis system for Dutch. *IPO Annual Progress Report, 19*, 32-35.

Kerkhof, J., Wester, J. & Boves, L. (1984) A compiler for implementing the linguistic phase of a text-to-speech conversion system. In: H. Bennis and W.U.S. van Lessen Kloeke (Eds): *Linguistics in the Netherlands*. Dordrecht: Foris Publications, 111-117.

Rijnsoever, P.A. van (1988) A multilingual text-to-speech system. *IPO Annual Progress Report, 23*, 34-40.

Quené H. & Kager R. (1989) Automatic accentuation and prosodic phrasing for a Dutch text-to-speech system. In: J.P. Tubach and J.J. Mariani (Eds): *Proceedings Eurospeech 89, Paris*, 214-217.