The "Tiepstem" : an experimental Dutch keyboard-to-speech system for the speech impaired

The " Tiepstem": an experimental Dutch keyboard-to-speech system for the speech impaired

René J.H. Deliege

The "Tiepstem": an experimental Dutch keyboard-to-speech system for the speech impaired.

Proefschrift

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof. ir. M. Tels, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op dinsdag 31 oktober 1989 te 16.00 uur

door

RENÉ JOHANNES HUBERTUS DELIEGE

geboren te Eindhoven

Dit proefschrift is goedgekeurd door de promotoren:

Prof. Dr. H. Bouma

en

Prof. Dr. S.G. Nooteboom

Prologue

This thesis describes work carried out at the Institute for Perception Research (IPO, Eindhoven) from 1983 until 1987, on speech communication aids for the speech impaired. The work was carried out in the "Hearing and speech" research group and the "Communication aids for the handicapped" working group at IPO. The research was financially supported by the Eindhoven University of Technology. The Institute for Rehabilitation Research (IRV, Hoensbroek) also participated in the project and contributed to the rehabilitation aspects and the field evaluations.

Parallel to this project another project was carried out at IPO [Waterham, 1989], which investigated the use of synthetic speech for the speech impaired. The difference between the projects was that our project focused on unlimited vocabulary speech synthesis, while the other focused on user aspects and used a limited number of speech messages. Because both projects applied speech technology and both projects had the speech impaired as a target group, close cooperation was realized from the start. This led to an identical project strategy and the use of similar techniques and electronic designs. Both projects were also able to benefit from the cooperation with the IRV in that evaluations were carried out in a similar way. The cooperation furthermore resulted in a combination of the devices developed in the two projects, combining the strong properties of each separate device.

As a consequence of the correspondence of the two projects both theses show similarities. Each of the authors is responsible for his thesis as a whole, but a differentiation can be made according to the senior authorship of the various sections. Authors of chapter 1 are Waterham and Deliege, except part of section 1.1, which has been written by Deliege. Authors of chapter 2 are Deliege and Waterham, except section 2.4, which has been written by Deliege. Authors of chapter 3 are Waterham and Deliege, except part of section 3.5, which has been written

by Deliege. Author of chapter 4 is Deliege. Authors of chapter 5 are Deliege and Waterham except section 5.4.4, which has been written by Deliege. Author of chapter 6 is Deliege, except section 6.4, which has been written by Deliege and Waterham.

Part of chapter 4 has already been published in the international journal "Speech Communication" [Deliege, 1989].

Contents

APPENDICES

Chapter 1

Preface

1.1 Introduction

Recent developments in the field of speech technology have led to speech synthesis techniques that offer a quality sufficient to be used in practical applications. These developments, together with developments in electrical engineering, have created the possibility to apply synthetic speech in small portable devices. This has opened the way to various practical applications. One class of applications is in speech communication aids as a replacement of natural speech for people who have lost the ability to speak. This offers interesting possibilities to contribute to alleviating the consequences of serious speech handicaps. Such a handicap can be caused by a language disorder or by a malfunction of the speech organs. Speech can be considered to be an essential communication channel in human life. In the Netherlands in total about 48,000 persons [CBS, 1974] have a functional speech disorder. Of this group about two thirds have additional disorders such as motor or cognitive disorders. Because of the serious consequences of a speech impairment and because of the number of people involved, it is worth attempting to apply synthetic speech in communication aids for the speech impaired. In this project we therefore investigate whether a Dutch speech communication aid based on speech synthesis can be constructed in such a way that it can be used with success by speech-impaired persons for diminishing their handicap.

In the field of aids for the handicapped human factors are as important as technical aspects. Our investigation therefore pays attention to both aspects. Also our project set-up allows for the application of these different fields of expertise in our investigation. Because this investiga-

tion was carried out at the Institute of Perception Research (IPO) in cooperation with an interfaculty university working group, the required multi-disciplinary (technological and ergonomic) knowledge and experience was available. In addition we solicited the help of the Institute of Rehabilitation Research (IRV) with respect to the technique of field evaluation.

We started with some basic requirements available from literature and from our contacts with therapists. Basic requirements are [Talbott, 1984]: intelligible speech to make communication possible, natural speech to make it socially acceptable, ease of operation and a large enough vocabulary to be of value to the potential user, portability so the user can use it wherever he wishes, and low price to make it affordable. Ease of operation and a large enough vocabulary do not easily go together so a choice had to be made based on what is technically possible and of practical use.

In this project the focus is on a sufficiently large vocabulary. In order to achieve this we use the diphone concatenation technique of speech synthesis, which offers an unlimited vocabulary. Diphones are speech fragments, running from some point in the steady-state portion of one speech sound to some point in the steady-state portion of the next speech sound, in this way containing the transitions between speech sounds in precompiled form. A set of Dutch diphones was available from another project carried out at IPO in which the possibilities of Dutch diphones for speech synthesis purposes were investigated [Elsendoorn, 1984]. To improve the naturalness and intelligibility of this speech we use available Dutch intonation rules ['t Hart and Collier, 1975]. Although an aid with an unlimited vocabulary is necessarily somewhat complex to operate, we tried to make its operation as easy and fast as possible. For the input we used a normal keyboard. Some extra features such as input editing and memory facilities were implemented to facilitate the input process. The whole system is battery-operated and as compact as possible. We called this device the "Tiepstem", i.e. Typing-voice.

An evaluation by potential users was carried out to reveal how adequate our basic requirements were and how well we had achieved them. Next, we updated these requirements to be used to design and make an improved device. Generally, such a process comes to an end when a de-

vice is realized which can be used with success or when we know why it is not yet possible to do so.

In a parallel project [Waterham, 1989] the focus is on ease of operation, which led to a restricted vocabulary and the use of precompiled speech messages. This resulted in an aid called the "Pocketstem". We also tried to combine the results of both projects by realizing a combination of the aids developed in these projects.

1.2 Research on aids for the handicapped

Developments in the field of speech technology have opened new possibilities for the realization of aids for the handicapped. Relevant fields of speech technology are speech recognition and speech production. In this project speech production is applied. Speech production, be it by synthesis or by resynthesis, is the more developed technique and is therefore already being applied in aids for the handicapped. The current state of speech production technology offers a speech quality sufficient to be used in practical applications. Although much research on the application of speech recognition is being carried out, e.g., in voice input environmental control or typewriter operation, this technique is not yet widely used in aids for the handicapped.

The developments in the field of micro-electronics provided us with powerful microprocessors and microcomputers. Their properties are also very useful for aids for the handicapped [Mariani, 1984]. Another result of technological developments is the availability of CMOS technology, enabling components with low power consumption to be produced. These results combined offer the possibility to realize complex functions in a small, battery-powered system.

Besides these technical considerations the results of an investigation into the needs for the application of synthetic speech for the handicapped were available from a project carried out by Kroon [Kroon, 1986]. The main need found in this investigation was for a speech communication aid. Among some other needs noticed were a talking typewriter and a reading machine. In that project the research was directed towards the development of a talking typewriter. The application of synthetic speech where human speech fails or is absent seems a natural one. It might par-

tially replace failing human speech by other kinds of speech, speech being the main communication channel between men. Speech has numerous advantages over other ways of communication. These advantages became clear from experiences with an early experimental communication aid using synthetic speech in Sweden [Carlson, Galyas, Granström, Petterson and Zachrisson, 1980]. The advantages mentioned included participation in group activities and discussions, communication with children and use of the telephone. This leads to the conclusion that developing aids for the speech impaired employing synthetic speech can be useful.

How are other aids for the handicapped being developed? An often encountered, practical situation is the realization of an aid with a very narrow focus to address the special need of a particular individual [Gordon and Zabo, 1984]. Usually this is done by someone who works in a technical service group in a rehabilitation centre, a nursing home or a hospital, or by a friendly neighbour or relative. The advantage is that the aid can be optimal for that user, provided that the designer is adequately skilled. The disadvantages are that the aid is seldom useful for other handicapped persons and one has to find someone who is able and willing to realize the aid. A further disadvantage is that the development does not take place in a professional environment, so the aid is probably not provided with up-to-date, professional, components, knowledge and techniques. Important stages in making aids for the handicapped available to those who need them are: research, development, production and marketing. In the field of consumer products the industry covers all the forementioned aspects. But in the field of aids for the handicapped the case is somewhat different because of the following considerations:

- Size of the market. The word "handicapped" stands for all kinds of different handicaps. In practice an aid may only be useful for a specific kind of handicap. This restricts the market size. Taken into account that the cost of research and development is rather independent of the size of the market, it is obvious that the price of the product will be higher if the market is small. Even so, the market is more difficult to reach and develop. This makes it more difficult to sell the product. These aspects make research and development for such a relatively small market very risky.

- Diversity of the group of handicapped. Even if we focus on a spe-

cific kind of handicap (or a specific aid) there will be a problem in describing the general abilities of that group. Because of the fact that these abilities differ, an aid is usually only useful for a part of the target group, or it will be more expensive because of individual adjustments needed.

- Complexity of the market. The speech impaired are often represented by a therapist or a relative and so are seldom speaking for themselves. The handicapped are often less capable of expressing their needs and demands. Because of this, demands of the potential users are hard to estimate.

- Expensive research and development. In comparison with consumer products of similar technical complexity, the ergonomic demands on aids for the handicapped are more severe because of the restricted perceptual and/or motor skills. Appropriate treatment of these ergonomic demands certainly requires more time and often calls for a multi-disciplinary approach to the problem. These aspects make research and development on aids for the handicapped more expensive.

These considerations make it clear why the industry is generally not very keen on doing research and development (R & D) on aids for the handicapped. Because of this, alternatives should be looked for. A party which can be interested in research and development in this field is a university, because R & D will enlarge knowledge, and the acquisition, generation and spreading of knowledge is one of a university's main tasks.

The above-mentioned complex and multi-disciplinary character of research and development on aids for the handicapped provides another reason why this work can very well be performed at a university, if the R & D fits in the university's research programme. The possibilities of realizing this R & D on aids for the handicapped are increased by recent stimulation and financing by both university and government.

A disadvantage of R & D on aids for the handicapped at a university is that a university is not the place to produce and market a product. So we are confronted with a different problem, which is how to transfer knowledge, ideas, experimental models and research results to an industry that is willing to undertake the production and marketing of

the product. This transfer will encounter some difficulties due to the following considerations [Ring, 1983]:

- Research done at a university is, in general, not always popular in the industry because there is little guarantee about the speed with which the work is carried out.
- Industry has problems with the fact that a market analysis in this case is notoriously difficult because of the fragmented nature of the field in which marketing will be performed.
- Grant-awarding bodies, or other sponsoring agencies, often underrate the cost of effectively exploiting a useful development.
- Potential manufacturing and marketing organizations are not introduced to the product at a sufficiently early stage in development.

When these difficulties are recognized and adequate efforts are made to overcome them, research and development of aids for the handicapped in a university research programme seems a suitable way to use up-to-date knowledge and techniques in the best interests of the handicapped. At our university the results of some projects have already been transferred to an industry, e.g., the artificial larynx with semi-automatic pitch control [Schuurmann and Mélotte, 1982], the Reflotalk [Waterham, 1983] and the Monoselector [Leliveld, Bosch, Mathijssen and Ossevoort, 1988].

1.3 Scope of the study

The previous section showed that several factors obstruct research on aids for the handicapped. To make sure that all these factors get proper attention we adopted in our project an explicit research strategy, which has proved successful in a number of projects in the field of aids for the handicapped [Collins, 1974; Damper, Burnett, Gray, Straus and Symes, 1987; Galyas and Liljencrants, 1987; Kroon, 1986; Maling, 1974; Sadare, 1984]. In this research strategy all points of interest are arranged in a coherent order, so that [Klip, 1982]: **a)** A good overview is obtained of all these aspects. **b)** There is less likelihood that some aspects will be overlooked. **c)** The effectiveness of the work increases. **d)** There is a greater possibility of the resulting design being useful.

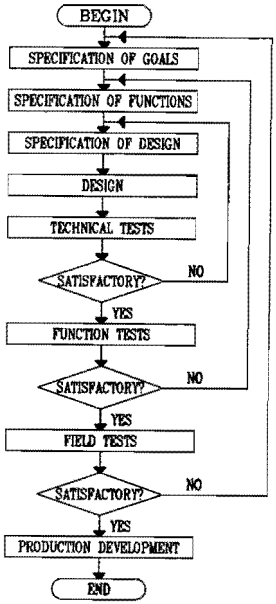

Figure 1.1: Flowchart of the research strategy [Soede, 1980].

This research strategy is formalized in figure 1.1 [Soede, 1980]. The start is an orientation on the problem which is done by studying the available literature and by interviewing members of the target group or people who have experience in the field (therapists, physicians etc.). From this orientation some elementary requirements are derived for the aid to be developed. On the basis of these elementary requirements and the available techniques an experimental model is made. The model is then evaluated to test its usefulness and in the light of the results of this evaluation the requirements on the aid are adjusted. It is widely acknowledged that the most useful evaluation is a full exposure of the model to the user population and monitoring its performance [Ring, 1983]. One should avoid including only a few persons in this evaluation because in that case there

is a chance that only those persons' needs will be met in the aid [Gordon and Zabo, 1984]. On the other hand sufficient attention should be paid to the particular needs of an individual user [Vanderheiden, 1982]. There is another aspect of the evaluation that should be taken into consideration: care should be taken to avoid giving participants in an evaluation too high expectations with regard to the time they are allowed to use the model or with regard to the time it will take for the aid to become available on the market. If this procedure of defining user requirements, creating an experimental model and evaluating this model is repeated two or more times, there is a fair chance of ending up with a useful aid. Before we discuss the actual work of the project two general aspects deserve attention: (1) Available synthetic speech techniques and the technique used in this project, discussed in chapter 2; (2) Ergonomics implies that we have to know what kinds of speech impairment can occur and what their implications are. Such aspects of communication and communication disorders are discussed in chapter 3.

The actual development and realization of the Tiepstem and its evaluation are discussed in chapter 4. A combination of the Tiepstem with the Pocketstem is discussed in chapter 5. Finally in chapter 6 a survey of the project, its results and its outlook are discussed.

Chapter 2

Speech storage and production

2.1 Introduction

This chapter gives an overview of some relevant aspects of synthetic speech and the possible storage and coding techniques [Witten, 1982; O'Shaughnessy, 1987; Holmes, 1988]. It will explain why we opted for the use of a formant synthesizer in combination with digital storage of speech data in an integrated circuit memory (EPROM).

Section 2.2 starts with a comparison of our choice of speech data storage in EPROM with other possibilities of speech data storage. We chose a digital storage medium because of its robustness. In order to reduce the cost of digital storage, we used a coding technique to lower the data rate. The technique available at our institute for coding speech in formant parameters, which combines both data rate reduction and the possibility of easy manipulation of the speech data, is compared with other coding techniques. The hardware synthesizer chip used, based on this coding technique, is described.

Section 2.3 discusses various techniques used to generate utterances and the specific properties of those techniques. The speech synthesis technique used in this project is discussed together with other techniques for synthesis or resynthesis of speech, in order to explain specific advantages and disadvantages.

Section 2.4 finally discusses the specific aspects and consequences of our choice of speech synthesis, because it forms the basis of our speech output system and determines a number of conditions for the project.

2.2 Speech coding and reproduction

In this project we opted for speech data storage in an integrated circuit read-only memory (ROM). In order to explain our choice we shall first discuss several storage techniques and their properties. Our demands on the storage technique (and medium) are: robustness, fast access, no severe deterioration of the speech quality, compactness and affordability. The robustness demand stems from the use of our aid in activities of daily life, where it is likely to undergo some mechanical shocks. The demand for fast access stems from the use of the aid in communication situations in order to adequately react to the environment. As a rule of thumb we want to have access to the stored speech data within one second. The demand for good, at least intelligible, speech also stems from the use of the aid in daily life communication situations. Furthermore the speech storage should not take up so much space that the aid would have to be considerably larger than necessary for other components such as speaker, battery and keyboard. Last of all the speech storage should not make the aid excessively expensive.

For speech storage, analog storage of speech is the most straightforward technique. None of the existing analog speech storage techniques, however, fulfill all our requirements. Mostly the mechanical robustness is insufficient (e.g., record) and when it is satisfactory, access speed is intolerably slow (e.g., magnetic tape, cassette recording). A less straightforward, but commercially widely available way to store speech is to convert the analog signal into a digital signal and store this digital information. Playback of this information always starts with a conversion into an analog signal, which in turn is made audible. Again, most of the available techniques do not meet the requirements of both robustness and fast access. For instance a CD player or floppy disc is not (yet) capable of tolerating shocks and a digital tape recorder does not have the required fast access. The storage of speech data in ROM has some advantages. A ROM (or other memory Integrated Circuits) is robust (no moving parts) and allows immediate access. The other three requirements, however, lead to a compromise. If we want to store a considerable amount of speech of good (or perfect) quality, the storage will be both expensive and extensive.

Apart from price and dimensions when a reasonable capacity is wanted, the storage of speech in ROM is perfect for situations where mechanical robustness is a major requirement. In the future, when large capacity ROMs become available at low cost, the storage of speech data in ROM through Analog-Digital conversion and the reproduction through Digital-Analog conversion is likely to be widely used. Until then ROMs can economically be used when the data rate of the speech signal is lowered. Several coding techniques have been developed for this purpose [Witten, 1982; O'Shaughnessy, 1987; Holmes, 1988].

A first class of coding techniques comprises the waveform coders. Waveform coders, as their name implies, attempt to copy the actual shape of the speech signal. The simplest form of waveform coding, Pulse Code Modulation (PCM), is normally not used for bulk storage of speech in simple systems, because the required bit rate for acceptable quality is too high. The necessary bit rate can be reduced by exploiting redundancies in the speech signal (e.g., Delta Modulation) or properties of the human hearing (e.g., Mozer coding).

Another class of coding techniques uses the restrictions imposed by the process of human speech production to further reduce the necessary bit rate. For this purpose they assume a speech production model. To this class belong the techniques of Linear Predictive Coding (LPC) and formant coding. In addition to a rather low bit rate these techniques have the additional advantage that special-purpose chips for LPC or formant synthesis are available and relatively inexpensive. For these reasons one of these chips is a good choice for our purpose. The combination with storage of the speech data in ROM fulfills satisfactorily our requirements of robustness, fast access, no severe deterioration of the speech quality, appropriate dimensions and affordability.

For the storage we use widely available commercial EPROMs (Erasable Programmable Read Only Memory). These allow easy programming at low cost. Our choice of the speech synthesizer chip is mainly influenced by the availability of the analysis software that is necessary to code the speech into parameters for the speech synthesizer used. At our institute the LVS software package was available [Vogten, 1983], which allows for analysis, manipulation and synthesis of speech and can code speech data as formant parameters suitable for the Philips MEA8000 for-

mant synthesizer (succeeded in 1988 by the PCF8200). For this reason we chose this chip for use in our project. It is based on the source-filter model for speech production [Fant, 1960]. In this model we distinguish a sound source, which produces the sound, and a filter, which shapes the spectrum of the produced sound. In this spectrum a number of peaks can be found, called formants.

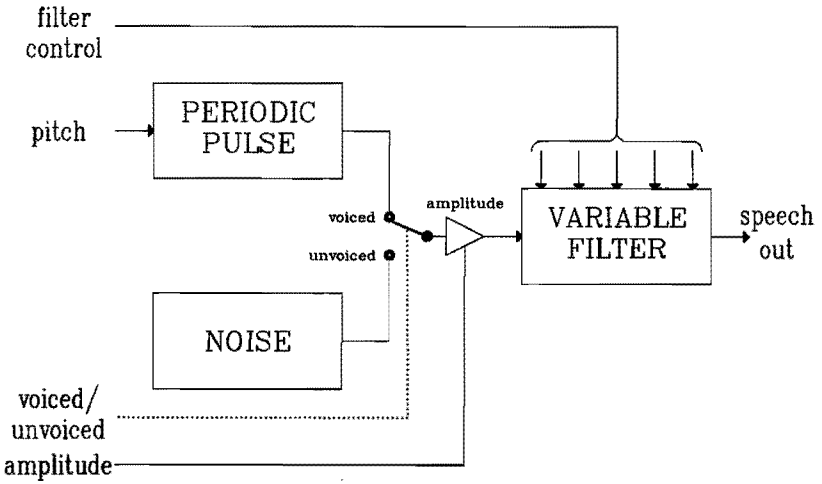

Figure 2.1: Simple electronic model of the human speech production mechanism.

In human speech production the sound source consists of the vibrations of the vocal folds, the air turbulence that is caused when air is forced through a constriction in the vocal tract, or a sudden release of built-up air pressure. The filter in human speech production is formed by the part of the vocal tract between the sound source and the free air. In the MEA8000 synthesizer chip a simplified electronic model of the human speech production mechanism is implemented (figure 2.1). A periodic signal, representing the pitch of the original voiced signals, or an aperiodic signal, representing the unvoiced sound in the speech, is fed to a variable filter comprising four resonators, via an amplifier that

controls the amplitude of the synthesized sound. The resonators model the sound in accordance with the formants in the original speech. Each resonator is controlled by two parameters, one for the resonant frequency and one for the bandwidth. Thus the information required to control the synthesizer comprises:

pitch
voiced/unvoiced sound selector
amplitude
filter settings.

A detailed description of the MEA8000 synthesizer and the data format used will be found in appendix A.

2.3 Speech synthesis and resynthesis

In the previous section we described various ways to store, code and decode speech. The speech material is of course application-dependent. Some applications only need a restricted number of utterances while others need the whole vocabulary of a certain language. In order to give an overview of the different techniques for realizing these varying demands on the speech vocabulary we will start with the most straightforward technique, that is to store the utterances that are needed. This technique is known as "speech resynthesis" or as "stored speech": an actual human utterance is recorded, perhaps processed to lower the data rate, and stored for subsequent regeneration when required [Witten, 1982]. The main advantage of this technique is that all aspects of the original speech, such as prosody and rhythm, are preserved, although the speech quality may be degraded depending on the coding/decoding technique used. A disadvantage can be that all utterances needed have to be known and recorded beforehand. This becomes a real problem when an extensive set of utterances is needed. In this case preparation time and memory requirements become unacceptable. The method is very useful, however, in applications that need natural speech and only a limited vocabulary. Examples are alarm message systems and communication aids with a limited vocabulary. If the number of utterances becomes too large or if all utterances are not known beforehand, this technique cannot be applied.

An approach that tries to solve the forementioned problem is the technique of "speech synthesis": the machine produces its own individual utterances, which are not based on recordings of the same utterance by a human speaker [Witten, 1982]. The way a machine generates such utterances is by using speech units as building blocks. The most obvious unit seems to be the word. Word concatenation was (and perhaps still is) the most widely-used synthesis method [Witten, 1982]. The main advantage of this method is that a great variety of utterances can be created with a limited set of words. Disadvantages can be that although prosody at word level is preserved, sentence prosody is lost, and in applications where naturalness is a major demand some coarticulation rules are necessary [O'Shaughnessy, 1987]. This method is useful, however, in applications where the number of words needed is limited and in applications where words can be used in carrier phrases, as for example in applications producing spoken output of the value of a numerical display (e.g., talking calculators, watches, multimeters). If the number of words becomes too large or if all words are not known beforehand this technique cannot be applied either. This will occur most notably when a machine should be able to speak the whole vocabulary of a certain language.

One obvious way of overcoming this difficulty is to select building blocks from which, by means of concatenation, words can be constructed. Building blocks can be units of phoneme size. A phoneme is the smallest unit in speech where substitution of one unit for another might make a distinction of meaning [Fischer-Jørgensen, 1956]. The actual sound manifestations of a single phoneme show a wide variation in their acoustic properties. This is partly due to the effects of co-articulation. Therefore the problem with phoneme-like units for speech synthesis is that their pronunciation depends heavily on phonetic context. This requires smoothing and adjustment processes, and reconstruction of acoustic transitions from one unit to the next, by relatively complex rules, to achieve intelligible and natural speech.

Another way of using small units, while still achieving natural coarticulation, is to make the units include the transition regions. Many speech sounds contain an approximately steady-state region, where the acoustic properties are not greatly influenced by the neighbouring sounds. Thus concatenation of small units can be improved if each

unit contains the transition from one phoneme-size segment to the next, rather than a single segment in isolation. Storing the transition regions requires the number of units to be of the order of the square of the number of the individual phonemes of the language. The number required for Dutch is about 1600. This makes it possible to achieve an unlimited vocabulary. As the individual units are quite short the storage disadvantage is not too serious. Such units have variously been described as diphones, dyads or demisyllables. The general principles of all three are similar, but there are differences in detail between techniques developed by different research groups. We used a set of diphones that was available at our institute as a result of an earlier project [Elsendoorn and 't Hart, 1982]. Diphones are speech fragments, running from some point in the steady-state portion of one speech sound to some point in the steady-state portion of the next speech sound, in this way containing the transitions between speech sounds in precompiled form. Diphones are excised from analysed speech and stored in terms of sequences of analysis frames, each frame containing the momentary parameter settings for a speech synthesizer. A disadvantage of this method, however, common to all techniques using small speech fragments, is the necessity to generate a pitch contour. The original pitch of the small speech units depends amongst other things on the position of that unit in the sentence, and therefore cannot be used in constructing new, artificial utterances.

2.4 Our application

As was mentioned in the previous chapter, we want to realize a communication aid with an unlimited vocabulary. Additional requirements are to fulfill this demand with as little memory as possible (because of power consumption, cost and dimensions), a high speech quality (intelligible and natural-sounding speech) and preferably without complex software (to keep the system simple).

In a previous section we explained our choice of a special-purpose formant speech synthesizer. This kind of synthesizer offers reduction of the data rate (which reduces memory requirements) and allows manipulation of the speech parameters such as the fundamental frequency.

The speech synthesis technique we use is diphone concatenation, as

mentioned in the previous section. This technique allows us to fulfill the demand for an unlimited vocabulary combined with a speech quality that is believed to be high enough for this kind of application. The use of diphone concatenation, however, imposes two problems. It requires the generation of a pitch contour, as was mentioned in the previous section, and some kind of input translation. For the generation of an artificial pitch contour we made use of the results of earlier investigations carried out at our institute ['t Hart and Cohen, 1973].

The necessity for some kind of input translation is due to the fact that there is no one-to-one correspondence between the spelling of an utterance and the sequence of diphones that has to be concatenated to produce the same utterance. Diphones consist of the transition between the sound manifestation of two subsequent phonemes, so there is a one-to-one correspondence between the sequence of phonemes and the sequence of diphones for a certain utterance (figure 2.2).

```
Diphones:      #K KA AT T#
Phonemes:      #  K  A  T  #
```

Figure 2.2: Example of diphone-phoneme relationship (# is silence).

In most languages the correspondence between the spelling (graphemes) and speech sounds (phonemes) is not one-to-one. Therefore the input translation, in general called grapheme-to-phoneme conversion, has to transform the spelling of the input to the corresponding phoneme sequence. If this correspondence is to a large extent regular, it is possible to develop algorithms performing this task. Although for some languages rather successful algorithms have been developed [Allen, Hunnicutt, Carlson and Granström, 1979; Hertz, 1982; Klatt, 1982; O'Shaughnessy, 1984; Hunnicutt, 1980], for Dutch work was still in progress at the start of this project. Therefore we chose to use a pseudo-phonetic input notation for the Tiepstem instead of normal spelling. At the moment at least one satisfactory Dutch grapheme-to-phoneme conversion algorithm is available [Kerkhoff, Wester and Boves, 1984], which will be used in the successor of the Tiepstem (appendix C).

Chapter 3

Speech disorders and communication aids

3.1 Introduction

This chapter deals with speech communication, communication handicaps and communication aids. Section 3.2 discusses speech disorders and their causes, which is necessary to understand classifications of the group of speech impaired made in the following chapters. The consequences of these causes related to additional disorders are discussed because these disorders may seriously affect cognitive and motor abilities, which in turn have an important bearing on the design requirements of a communication aid. Section 3.3 discusses the use of communication aids to compensate for or to restore disturbed speech communication compared with the use of therapy and the use of alternative communication. Furthermore several objectives of a speech communication aid are discussed, because the aid developed in our project, intended as a speech-replacing aid, turned out to be useful as a therapy-supporting aid as well. In section 3.4 a survey of available speech communication aids in the Netherlands is presented in order to show that no speech communication aid with speech output was available at the time and to give some insight into the properties of the aids already in use by the Dutch speech impaired. The application of synthetic speech in a practical device, however, is not new. In section 3.5 therefore we take a look at existing applications suited to be used as a speech communication aid.

3.2 Communication handicaps

A Dutch dictionary [Geerts and Heestermans, 1984] defines communication (latin: communicare, to make something common, to share, to inform) as the (opportunity to) exchange (of) thoughts, to have mental interaction. A more technical description [Steehouder, Jansen and Staak, 1984] states that communication occurs when someone lets another one know something. The first one is called the sender and the other is called the receiver. The object that the sender transfers to the receiver is called a message. If the sender uses spoken language, the interaction is called verbal communication. If the sender uses means other than spoken language (e.g., gestures, mimes, signs or pictures) the interaction is called non-verbal communication. We can combine the distinctions between sender and receiver and between verbal and non-verbal communication to form four communication modes [Verniers and Verpoorten, 1986]:

1. verbal, expressive: a person expresses thoughts by means of spoken language.
2. non-verbal, expressive: a person expresses thoughts by other means than spoken language.
3. verbal, receptive: a person receives messages through spoken language and interprets them.
4. non-verbal, receptive: a person receives messages through other means than spoken language and interprets them.

Verbal communication is a fast way of communication with a communication rate up to 175 words/minute [Foulds, 1986]. Apart from its speed, verbal communication is important because it is the primary means for interacting, for expressing feelings and ideas, for venting anxieties and frustrations, for effecting change and for enabling one to find out what another is perceiving and thinking [Weiss and Lillywhite, 1981]. We focus on disorders in verbal communication because they create a serious handicap which can be diminished by speech synthesis devices. We further narrow our focus to dysfunctions of the verbal-expressive channel, because, as already mentioned in the previous chapters, it is this channel

we want to replace with synthetic speech. A receptive dysfunction can however cause an expressive dysfunction and is in that case still of interest to us. Because both language and speech are necessary for spoken language we can divide verbal-expressive communication disorders into speech and language disorders. The two are closely related but we might make a distinction inasmuch as language has to do with the creation of a message and speech is the actual signal that carries the message from sender to receiver [Dudley, 1940].

We shall first discuss some aspects of speech and speech disorders and then do the same for language and language disorders. After the survey of disorders we shall sum up possible causes of the disorders mentioned. This survey of disorders and their causes is given because in medical health care the group of speech impaired is partly classified by their disorder and partly by the cause of their disorder.

For speech a good voice and articulation are necessary, which implies well-functioning speech organs. These speech organs are [Nooteboom and Cohen, 1984]:

lungs	→	airflow
larynx	→	sound generation
mouth, throat, nose cavities	→	resonance
lips, velum, tongue	→	articulation

In addition the brain and the central nervous system play an important role because of their control and coordination. With these elements of speech production in mind we can distinguish the following categories of speech disorder [Mumenthaler, 1977]:

1 Psychogenic: dysphonic or aphonic (disturbed or total absence of voiced sounds). For instance caused by schizophrenia.

2 Laryngologic: dysphonic or aphonic. For instance caused by larynx removal, changes in the vocal chords, split palate, tongue removal, trauma or face muscle paralysis.

3 Neurologic: dysarthric, anarthric, dyspraxic or apraxic (disturbed or total absence of control of the speech organs due to dysfunction of

the nervous system). For instance caused by central or peripheral injury.

Apart from these speech production organs, hearing is important because of the necessary feedback [Vincent, 1987]. A hearing deficiency can therefore also cause a speech disorder.

Language disorders are mostly caused by cerebral disorders or disorders of parts of the central nervous system [Tervoort, Geest and Hubers, 1976]. We can distinguish the following types of language disorder [Mumenthaler, 1977]:

1 dysphasic or aphasic (difficulty or inability to use language as a result of cerebral damage).

   a Expressive aphasia (Broca aphasia): inability to produce language (understanding is not affected).

   b Receptive aphasia (Wernicke aphasia): inability to understand language (resulting in a disturbed language production).

   c total aphasia (Déjérine aphasia): total incapability of either understanding or to producing language.

2 disorders in language development, caused among other things by:

   - autism.
   - minimal brain dysfunction.
   - congenital hearing or sight deficiency.

Both speech and language disorders can have various causes which can be congenital or acquired. Examples of congenital causes are hearing deficiency, reduced mental capabilities or mental defectiveness. Examples of acquired causes are:

a. chronic diseases (multiple sclerosis, Parkinson's disease, Amyotrophic Lateral Sclerosis)
b. traumata (Cerebral Contusion and (pseudo) bulbar lesion)
c. vascular accidents (Cerebral Vascular Accident (CVA) and Transient

Ischemic Attack)

d. intoxications.

The speech and language disorders are not one-to-one related to their causes. For instance a CVA can cause both dys-/anarthria or aphasia, depending on the severeness and the location of the CVA.

It is obvious that a considerable number of the abovementioned causes not only affect speech or language but also result in additional disorders. A bulbar lesion which paralyses the speech organs, for instance, will often result in a dysfunction of other parts of the body. A cerebral contusion is not limited to specific parts of the brain, so it will equally likely result in other disorders than only a speech disorder. These aspects are confirmed by a statistical investigation into the handicapped population in the Netherlands carried out in 1974 [CBS, 1974]. The figures of the speech impaired related to additional disorders are shown in table 3.1. A speech-impaired person is taken as anyone who has a functional speech disorder, at least to a certain degree ranging from moderate (can speak but is difficult to understand in a group) to severe (cannot speak).

Table 3.1: Figures of the speech impaired in the Netherlands related to additional disorders.

Function disorder	Estimated number in the Netherlands	Percentage of all speech handicapped
speech (sp.) alone	13,400	31.6
sp. & walking	1,300	3.1
sp. & arm/hand	900	2.0
sp. & sight	200	0.5
sp. & hearing	6,700	15.8
sp. & stamina	2,600	6.1
sp. & remaining disorders	1,300	3.1
sp. & walking & arm/hand	10,600	25.0
sp. & walking & other	3,900	9.2
sp. & arm/hand & other	900	2.0
sp. & 2 others	600	1.5

From these figures we can see that approximately one third of the population of people with a speech disorder have no other handicaps. Approximately another third have a second handicap (often a hearing disorder), the remaining third consisting of speech impaired with two or more additional handicaps (often including a hand-function disorder). The important conclusion from these figures is that a speech communication handicap is in most cases accompanied by one or more other handicaps. These additional handicaps have to be seriously considered when designing an aid. For instance arm/hand function disorders diminish the ability to operate a device (keyboard, switch etc.) and a sight disorder may make a visual display less useful.

One of the aspects that is not shown in table 3.1 is the level of cognition. If we combine the two aspects, motor disorders and cognitive disorders, we can conclude that the group of speech-impaired persons ranges from people who only have lost speech to people who have little cognitive abilities left, combined with severe motor disabilities.

The fact that about one-third of speech-impaired persons only has a speech disorder provides the reason to do research on a speech communication aid with an unlimited vocabulary. This unlimited vocabulary implies that in principle all communication demands can be satisfied, although not all speech-impaired persons will be able to operate the aid because of their lack of motor and/or cognitive skills. If the remaining motor and/or cognitive skill is not sufficient, an easy-to-operate aid with a limited vocabulary is wanted. The latter aspects are dealt with in a thesis about the Pocketstem [Waterham, 1989].

3.3 Communication aids

In the previous section communication and some of its aspects were discussed. Verbal-expressive communication, which from here on we will call speech communication, is a vital part of a fast and powerful communication process. It is even valid to state that it is this power of speech that distinguishes men from other living creatures [Weiss and Lillywhite, 1981]. It is this very importance of speech communication for human life that makes the loss of speech so unbearable and calls for ways to overcome it and to restore effective communication.

To compensate for such a loss three approaches can be adopted. The first approach is, if possible, to restore original speech communication, for instance through therapy in the case of a light form of aphasia or by learning oesophageal speech.

The second approach is to use an alternative communication channel, e.g., lip reading, sign-language. This approach has the drawback that extensive training is necessary. Furthermore such an alternative channel is not normally used by non-handicapped people, so that the same training may also be necessary for the communication partner (e.g., sign language).

The third approach is to use a speech communication aid. The advantage of a speech communication aid is that only the user has to learn to operate it. The disadvantage, however, can be that a speech communication aid reduces communication speed or the vocabulary and is practically or socially less acceptable.

The existence of these three different approaches already suggests that none of them is the perfect solution in all cases. In the context of this study we concentrate on the third approach.

The objective of a speech communication aid can be one of the following three [Ring, 1983]:

Therapy-supporting: the aid is not intended to be used for actual communication, but for communication training (e.g., Laryngograph).

Speech-supporting: the aid is used as a support for speech production when part of the normal speech production mechanism is still intact (e.g., speech amplifier [Leliveld, Ossevoort and Severs, 1979], electrolarynx [v. Geel, 1983]).

Speech-replacing: the aid is used for communication without the use of the original human speech (e.g., Canon communicator, Multi-talk [Galyas and Liljecrants, 1987]).

In this project we are concerned with aids of the last category. Although practice showed that our aid can also be used in therapy, this was not our primary goal.

We will now try to give shape to some basic requirements for speech-replacing aids. Although an ideal speech-replacing aid would restore all aspects of natural speech, this is virtually impossible to realize in practice. For instance a large vocabulary implies (at least up to now)

complex operation, which makes the aid both slower and harder to learn to operate. If not all aspects of natural speech can be restored, at least some essential aspects have to be. Aspirations as to what to say can vary greatly amongst individuals, but some elementary aspirations, such as attracting attention, interrupting a discussion and addressing several people at once will almost always be present. Furthermore the user may want to use speech to communicate without eye-contact or at a distance and to use a telephone. Preferably, the realization of these aspirations should not impose an additional load on the user, neither during training nor during use, in order to make the aid effective.

One aspect of natural speech, namely its high communication rate, is very difficult to restore with a speech communication aid. Communication rates slower than three words per minute are found to be intolerable [Soede, 1986], so a speech communication aid has to be fast enough in operation, processing and output at least to meet this minimum rate. But although a speech communication aid may slow down the communication rate, it may still be useful because something is better than nothing. Last but not least, the aid should be available at an affordable price.

3.4 Available speech-replacing aids

In order to show where our aid is an addition to other aids and where it is unique, we give an overview of the available speech-replacing aids in the Netherlands. No aid with speech output was commercially available in 1988. We present the list of available aids together with a short description of them.

First of all we mention pen and paper which can be used to communicate. This is generally a practical and inexpensive solution provided that the user is capable of writing at an acceptable speed. Other inexpensive solutions are communication aids which can be self-constructed. For instance communication-cards or maps made by personnel of a hospital, clinic or institute such as therapists, relatives etc. can prove to be of great use although communication speed is low. The use of these aids is simple; the user points to a letter, word or symbol on the aid, which can be observed by the communication partner. Another example of this

group of self-constructed aids are those that use eye-communication. For instance a "look-through frame" is based on the principle that looking at a certain point (letter, word or symbol) of the frame can be observed by a communication partner who is opposite the user.

Commercially available aids in the Netherlands can all be regarded as a replacement for speech, although none of these aids uses artificial speech as a replacement for the original speech. Properties of the available aids such as a price indication, the size of the vocabulary and the function necessary to operate the aid, are summarized in table 3.2. Communication speed varies from fast (pointing at sentences),through slow (typing or looking up a sentence), to very slow (scanning or eye-pointing to produce letters or words).

Table 3.2: Commercially available communication aids in the Netherlands.

Name of the aid	price[1]	voc.[2]	speed[3]	function needed
Communicatie-klappers	1	fixed	f/s	hand
Symbocom	2	fixed	s	single switch operation
Electronote	2	16	s	single switch operation
Zygo 100 (16)	2	100 (16)	s	single switch operation
Canon communicator	3	∞	s	language, hand
Pocketcomputers	1	∞	s	language, hand
One-function Canon	3	∞	vs	language single switch operation
Prisma communicator	2	∞	vs	language, eye
Lichtvlekaanwijzer	2	∞	vs	language, head

[1] 1=under fl.100, 2=from fl.100 to fl.1000, 3=over fl.1000

[2] voc.=vocabulary

[3] f=fast, s=slow, vs=very slow

A brief description of the aids mentioned in table 3.2:

- The "communicatie-klappers" are commercially available versions of the earlier mentioned communication-maps.
- The "Electronote" is an aid which indicates a message by means of a LED (Light Emitting Diode). The indicating LED scans the 16

messages and scanning can be halted by a switch. The messages are easily changeable.

- The "Zygo 100" is an aid (in an attaché case) which indicates 100 small squares on which a symbol or word can be placed. The messages are easily changeable. The aid has several scanning possibilities (for instance repetition of series of messages) and offers one-touch operation. The "Zygo 16" is similar in size and operation, apart from the number of squares.
- The "Symbocom" is a portable aid similar to the "Zygo".
- The "Canon communicator" is a small portable aid which prints characters on a slip of paper. The aid offers many features and options such as an optional connection to a printer, typewriter or personal computer. The one-touch-operation version of the communicator includes scanning functions.
- Pocket calculators and portable computers (provided with an alphanumeric keyboard and an LCD display) can also serve as a communication aid although this is not their intended use.
- The "Prisma communicator" is an aid based on eye-communication. It facilitates communication by means of easy recognition of the spot on the aid where the user is looking.
- The "Lichtvlekaanwijzer" makes use of a red light source which can easily be fixed to a pair of glasses. The user communicates by pointing to a spot (letter, word, symbol or message) by moving his head.

The listed speech-replacing aids do not restore communication to a normal level, because they do not result in a normal communication rate and they do not have the advantages of speech mentioned in section 3.3. The reduced communication rate is not easy to overcome because it is due not only to the loss of speech but also to the difficulty in operating an aid, caused by a disorder. The advantages of speech are restored if the aid is provided with synthetic speech, provided of course that the synthetic speech is intelligible and socially acceptable. We can also see in table 3.2 that there are aids which offer an unlimited vocabulary

(e.g., Canon communicator, pocket computer) and aids which offer only a limited vocabulary combined with ease of operation (e.g., Electronote, symbol-chart etc.).

An investigation done in the Netherlands [Kroon, 1986] indicated the need for speech communication aids with speech output. In other countries we can see that speech communication aids with speech output are already successfully used [Peterson, 1982; Carlson, Galyas, Granström, Petterson and Zachrisson, 1980]. This leads to the conclusion that for the Dutch situation the availability of speech communication aids with speech output is desirable. The presence of aids either with a limited or with an unlimited vocabulary suggests that both categories are useful for aids with speech output.

3.5 Aids with synthetic speech

In this section we discuss existing speech communication systems and devices (aids) producing synthetic speech. In principle all systems and aids that are meant to be or can be used as a speech communication aid are of interest, but systems and aids of which evaluation data have been published are of special interest.

We first divide the field into some categories relating to potential user groups. The presented aids and systems can be looked upon as illustrations of these categories. Appendix B gives a list of aids together with more detailed information and references.

The major division we can make between these systems is according to the vocabulary being limited (1) or unlimited (2).

1 Systems which are able to produce a fixed vocabulary of utterances. We also include in this category all systems that have the possibility to program these utterances. For instance if the device is capable of creating or recording utterances, but its primary function is to reproduce these utterances, we still consider them to belong to this category.

In this category we can make a further division.

a Systems or devices that use preprogrammed sentences, words, letters or other speech fragments are for example Vocaid (phrases), Falck 3310 (phrases), Bliss-stem (words/phrases) and the Namcon Talkin'Aid (Japanese speech fragments).

b Systems that can be programmed by the user are for example Alltalk, the Zygo Parrot and the Prentke Römich Introtalker (recording of speech by digital storage of the AD-converted signal and playback through DA-conversion) and systems like the Handivoice, the Vòis, and the Touch-/Lighttalker, which can be programmed through the synthesizer incorporated.

2 Systems which have an unlimited vocabulary. Some of these systems also have the possibility to store generated utterances.

In this category we can again make a further division.

a Systems that use some kind of keyboard-input (e.g., text, Bliss) are for example: Multi-Talk, Sahara.

b Systems that use ASCII input to convert it into speech are for example: Dectalk, Prose 2000/3000.
Note that these systems are not necessarily useful communication aids, because they need some kind of input-to-ASCII converter and they are mostly not equipped with special facilities to make the complete system user-friendly.

c Software synthesizer systems that consist of a software package for a general-purpose microcomputer (pc or home computer) and need only an analog output to generate speech (e.g., a DA-converter present in the computer). We also include pc accessories in this category. An example of such systems is the Software Automatic Mouth (SAM).

Some of these systems are what we call "laboratory systems". These systems appear in literature (by some sort of description), but many of them do not become commercially available, although usually one or more systems have been realized and evaluated. It is uncertain therefore what the status of these devices is or will be. Examples of these laboratory

systems are: "Sadare's speech system", Psytalk, French Text-to-speech System.

As was remarked in section 3.4, some speech communication aids available in the Netherlands offer a limited vocabulary, while others offer an unlimited vocabulary. The same is noticed for aids abroad that offer speech output. Both categories are amply represented, and it is therefore interesting to investigate both varieties of a speech communication aid with speech output. In the available aids we find sometimes that the fixed vocabulary systems are programmable and that the unlimited vocabulary systems offer a storage and recall function Because we are developing an unlimited vocabulary system and in another project a smaller, easy-to-operate aid with a fixed vocabulary is being developed [Waterham, 1989], we can realize a form of an easily reprogrammable aid by combining both devices, using our system for programming and the other as the communication aid.

Of the aids and systems with an unlimited vocabulary none can produce Dutch and only the systems with the Epson HX-20 and speech extension combined in a suitcase (e.g., Multi-Talk) are portable.

In chapter 2 we motivated our choice of a hardware formant synthesizer and the diphone concatenation method. These techniques are not used in any of the forementioned systems. Additionally the requirement for a system that produces Dutch speech with an unlimited vocabulary led us to the decision to develop a new speech communication aid in this project. For the linguistic knowledge necessary we have to rely on existing and available knowledge, because linguistic research is not incorporated in this project.

Because the kind of aid we are developing is a complete device on its own, including the input facilities, it belongs to category 2a. So we can take a look at the devices in this category and see what general ideas we can learn from them. Most of them use a normal QWERTY keyboard for input. Because this is the most straightforward way to enter a message we also use this kind of input. This input may exclude some potential users (e.g., persons with additional motor handicaps). To facilitate the text entry by the keyboard, correction and storage facilities are implemented in almost all devices. This feature is worth implementing in our device too. As far as the synthetic speech is concerned, quite often low-quality

speech is used (Votrax chip [Greene, Logan and Pisoni, 1986]) and is apparently accepted in practice. This gives good hope that our diphone speech quality which is probably better, will be useful for this kind of application.

As far as the evaluation of the available aids is concerned, little information has been published. If available it consisted mainly of a description of use by a few users without quantitative data.

Chapter 4

An experimental model of the Tiepstem

4.1 Introduction

This chapter describes our first attempt to realize a successful speech communication aid with an unlimited vocabulary. We call it the experimental model. We did not expect that it would be a perfect device at once, if only because some necessary techniques (e.g., correct orthographic input processing) were not yet available. The main purpose of this model was that it could be used in practice, so we could check and update our initial requirements. Despite the lack of some necessary techniques such an evaluation should reveal much practical information. Because the application of speech synthesis is rather new and not widely experienced it is much more difficult if not impossible to collect this information by other means such as literature, questionnaires and interviews.

In the previous chapters we already mentioned the basic requirements that can be found in literature: intelligibility and naturalness of the synthetic speech, large enough vocabulary, ease of operation, portability and low price. We also mentioned that a compromise has to be made between large vocabulary and ease of operation. This project would focus on a large vocabulary and therefore we already explained our choice of the technique of diphone concatenation, offering an unlimited vocabulary. Two problems associated with the use of small speech fragments (e.g., diphones) are the need for some input processing and a natural sounding intonation contour.

The next section (4.2) discusses the choices we have made to over-

come these problems and other design specifications for this experimental model. The third section (4.3) describes the realized experimental model in detail. It describes construction, functioning from a user's point of view, the hardware and software, and the realization of a suitable diphone set. In the fourth section (4.4) the evaluation of this device is presented and we compare the results of this evaluation with our initial requirements in order to find out how far we got in our first attempt. In this way we can update our initial requirements and locate the weak points in our first design in order to arrive at the requirements for a second model.

4.2 Design specifications

The design specifications for this experimental model can be divided into three categories: one related to the synthetic speech, one related to the features available for operation of the device and one related to the physical properties of the device.

For the synthetic speech we chose the diphone concatenation method. This choice left us with two problems: what input is used and how do we achieve a natural-sounding sentence melody? For the input we chose a normal (typewriter style) keyboard because it is the most straightforward input medium for an unlimited vocabulary system. Because the input to the device is the most time consuming process in the production of a message a simple and fast input will result in a faster message generation. The most straightforward input spelling would then be normal orthography. This, however, was not yet possible for this device because it would require a grapheme-to-phoneme conversion system for Dutch, which was not available at the time. Therefore a temporary compromise was made. We chose to use a pseudo-phonetic notation that resembles normal orthography but is closer to pronunciation. For the generation of a natural-sounding intonation contour we could make use of pitch movements as specified in the grammar of Dutch intonation ['t Hart and Collier, 1975]. The placement of these movements in a sentence depends on the syntax and meaning of the sentence. Because the necessary syntactic and semantic analysis could not yet be done automatically we had to decide to let the user provide accent information in the input.

The main purpose of the necessary features is to make input easier and faster. This is desirable for two reasons. One reason has to do with speed of communication. The faster communication is possible with this aid, the more successful it can be as a speech-replacing device. The other reason has to do with the user's motor abilities that are required to operate the system. Because a speech-impaired person often suffers from additional motor handicaps, an easy and fast input will allow more people to use the device. We tried to make the input process simple and fast by means of the following features: a normal sized keyboard; a display that shows what has been typed; editing facilities to correct what has been typed; an option to go back to a previously typed sentence and use it again, possibly after editing; an option to repeat a message without retyping; an option to use digits as inputs instead of spelling out numbers; an option to store often used sentences or parts of sentences in a memory. This memory should retain its information even if the device is switched off.

The specifications of the physical properties of this device have mainly to do with the requirement of portability. This imposes limits on the size and weight of the device. We also want it to be battery-operated so the user is not limited to some fixed locations. This implies that the power consumption of the whole circuit should be low enough to make battery operation possible. To reduce power consumption the device should automatically switch off when it has not been used for some time. For transportation purposes a manual on/off switch should be present. The only adjustment the user needs to make is to the output volume control. A control with a clear visual indication is preferable because the user can select the volume before the device actually speaks. The maximum output volume should be sufficient even for noisy environments.

4.3 Realization of the experimental model

4.3.1 General construction

The device is housed in a plastic case measuring 380 x 240 x 60 mm^3. The weight of the device is 2.5 kg. The case used is that of the Acorn Atom homecomputer. This was an available case (together with keyboard) that satisfied our needs. The use of a commercially available case is

preferable because only a small number of devices have to be made and they will be used for only a short time. The keyboard that comes with this case consists of 60 (mechanical) keys of 13 x 13 mm^2 and with a centre-to-centre distance of 17 mm. Some of the keys were given a new label and keys that have no function in our device were made black. Some components have been added to the case: an LCD display (Epson EA-Y40080AT), a loudspeaker (Philips AD3371/Y8), a potentiometer (volume control), an on/off switch and a battery-charger socket. Figure 4.1 shows the complete device and figure 4.2 gives an overview of all the forementioned keys and components.

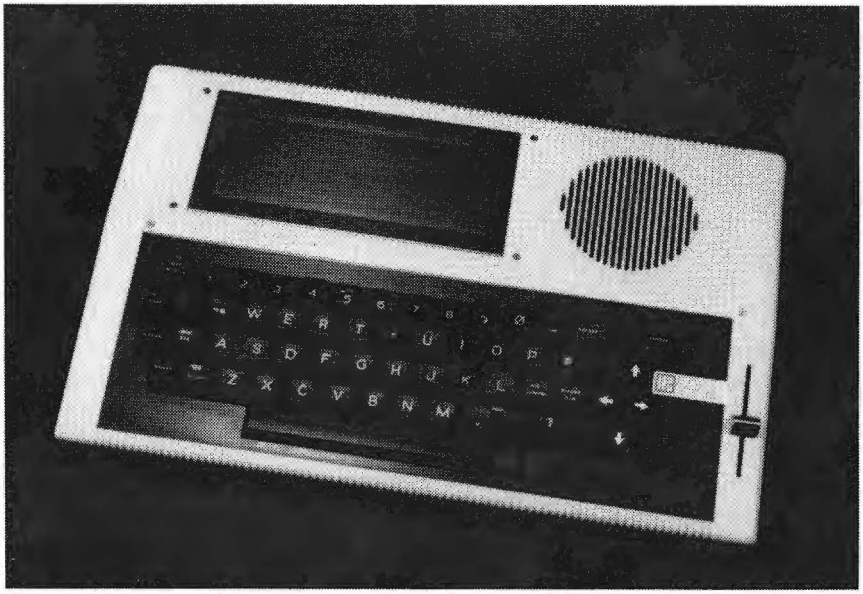

Figure 4.1: The Tiepstem.

The electronic circuitry of our system is placed on two printed circuit boards: a speech synthesis board and an input/output board (see section 4.3.3). The power source consists of eight NiCd batteries (size AA, 1.2 V, 500 mAh). Figure 4.3 shows the locations of the various parts inside the device.

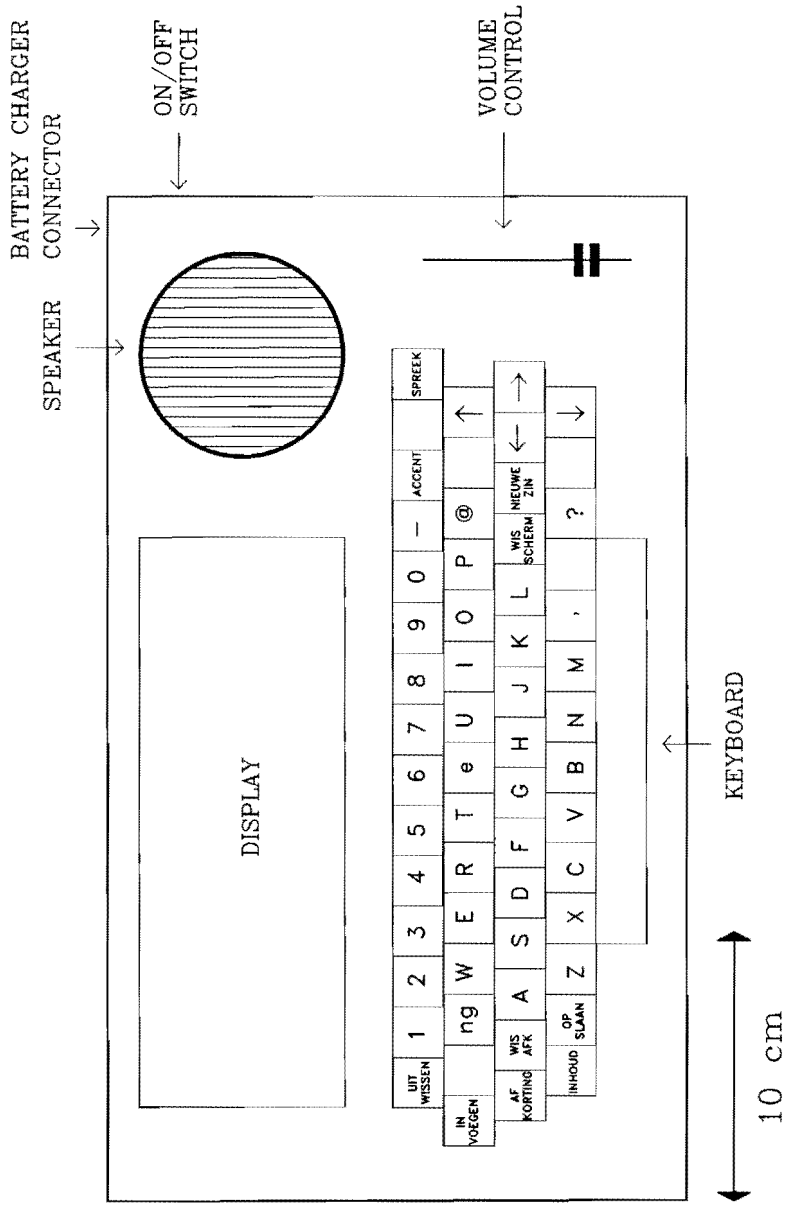

Figure 4.2: Overview of the Tiepstem.

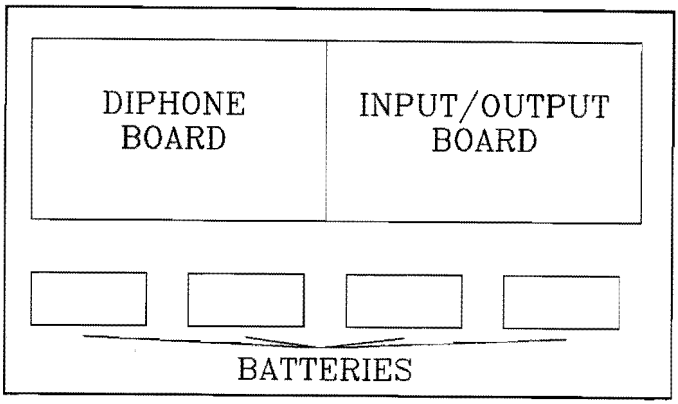

Figure 4.3: Internal components.

4.3.2 Functional description

The device is switched on by means of the on/off switch. The appearance of the cursor on the display is the indication that the device is on. After not being used for about two minutes, the device automatically switches off. Hitting any key on the keyboard switches it on again. The separate on/off switch is provided to switch the device off, for instance for transportation, when a key could be pressed accidently.

When the device is on, a message can be typed. For entering text the alphanumeric character keys are available. As already mentioned, the input has to be in a pseudo-phonetic notation, which is explained in section 4.3.4. Number keys are available to enter numbers. The comma, question mark and accent key can be used to create a correct sentence melody. The silence key (-) and glottal stop (@) can be used to further improve pronunciation.

The keyboard input is echoed on the display. The entered sentence is spoken when the key labelled *spreek (speak)* is pressed. The same sentence can be spoken again by again pressing the speak key. After a sentence has been spoken, the cursor is moved to the next line when

new text entries are made. When the bottom of the screen is reached, automatic scrolling takes place. Over and above these automatic cursor movements the cursor can be located by the user anywhere on the screen through four cursor direction keys (labelled ⇑⇓⇐⇒). At the current cursor position new text can be typed in, replacing whatever was present, or text can be inserted or deleted (keys *invoegen (insert)* and *uitwissen (delete)*). In this way misspelled messages can be corrected. The whole screen can be cleared by the *wis scherm (clear screen)* key and the cursor can be moved to a new line by the *nieuwe zin (new sentence)* key. The latter possibility can be useful because the automatic movement to a new line only takes place when the previous key was the *spreek* key.

Because typing a complete message is rather time-consuming, time can be saved by retrieving pre-stored messages from memory instead of typing them in. For this purpose a non-volatile memory is present; its content is saved even if power is switched off. An alphanumeric code is used as index to these messages. This allows the user to choose the index system he likes most, e.g., numbers, abbreviations, mnemonics, keywords. Sentences or parts of sentences on the screen can be stored in this memory by pressing the key labelled *opslaan (store message)*. The system then asks for the code under which this message is to be stored. The user can freely choose this code, using all alphanumeric characters available. So a sentence *I have to go to the toilet* could be stored under code *WC* or code *1* for example. When the key *afkorting (abbreviation)* is pressed, the system asks for the code of the message to be retrieved. The retrieved message appears on the screen, just as if it was typed and can be used as a whole sentence, as the beginning of a sentence that has to be completed, as a part within a sentence typed by the user etc. So the previous example could be realized by the four-key sequence *afkorting, W, C, spreek*. It is also possible to delete a message or to clear the whole memory (*wis afk (delete abbreviation)*). Pressing the key labelled *inhoud (contents)* gives an overview of the stored messages and their codes.

4.3.3 Hardware

The whole system is implemented as two microprocessor systems. One system, the input/output board, takes care of the user interface of the system; the second, the speech board, is a general speech synthesis sys-

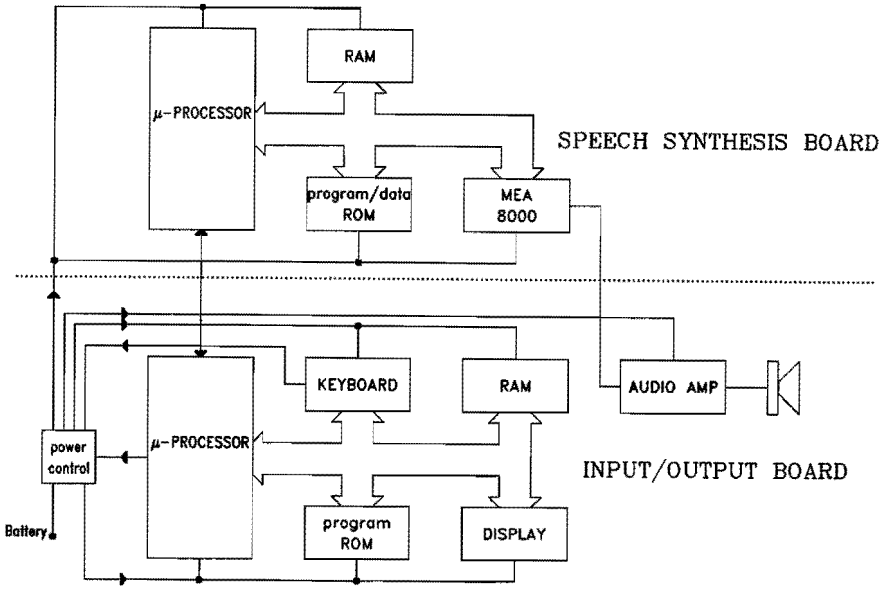

Figure 4.4: Block diagram of the Tiepstem.

tem. Because it turned out during the development of our system that there were potential applications for the speech synthesis system apart from the application in a speech communication aid, the hardware development and software implementation for this part have been done at the CAB department of Philips [Falzoni, 1985]. This resulted in a commercially available board (OM8002), which is used in our system with a few modifications (mainly to meet our requirement for low power consumption). Because it was impossible to add to this speech board the user-interface we had in mind, these functions are realized on a separate board. A block diagram of the complete system is presented in figure 4.4.

The main components for each system are the microprocessor, the program memory and the input/output. The input to the speech board is

ASCII coded text. Starting from this input text, speech data are fetched from a data memory that contains the diphone inventory. These data are then fed to the speech synthesizer, which produces the analog speech signal. The input/output board receives its input from the keyboard. Input consisting of text or editing commands etc. results in output to the display. When the input is the *spreek* key the text of the current sentence is transmitted to the speech board. From this board an analog speech signal is then received, which is then amplified and fed to the loudspeaker to produce the audio output. The input/output board can switch itself and/or the speech board on and off.

We will now describe each of these blocks in more detail, starting with the speech board.

First a choice has to be made of the processor to be used. As it was clear from the beginning that the storage requirements for the diphone inventory are about 80 kbytes, normal 8-bit microprocessors with a 16-bit address bus, giving an address range of 64 kbyte, would create some problems. Because a 16-bit microprocessor was too big for this application in terms of price, size and power consumption, another solution was looked for. This was found in the Intel MCS51 microcontroller series, which can handle a separate program memory and data memory of 64 kbytes each [Intel, 1985]. In addition this processor allows a simple, small and cheap system design because much of the additional hardware needed is integrated in this chip. Because so much memory is needed the version with internal program memory (4 kbyte) is not used, but one without internal memory, the 80C31. Because the data storage exceeds 64 kbyte the data are partially placed in program memory space. This processor uses a multiplexed data/address bus so an address latch is necessary. For the program and data storage six EPROMs (Erasable Programmable Read Only Memory) of the type 27128 (16 kbyte) are used: 4 in program and 2 in data memory space. As temporary storage 2 kbyte of RAM (Random Access Memory) is used (6116). Input/output is performed through the built-in serial port of the processor and an RS232 level converter or through a parallel port with some handshake logic. The MEA8000 speech synthesizer is mapped in data memory space and its data request line is connected to one of the processor's interrupts. To make stand-alone operation of this board possible it is equipped with

a small integrated audio amplifier (TDA7050, max. 150 mW). Because this audio output does not meet our requirement for sufficient output power this amplifier is not used in our application and the amplifier chip is removed to reduce power consumption. As the speech board was not designed for using low power CMOS components some more changes have been made. The parallel logic and the RS232 level converter are removed because they are not used by us. The processor is replaced by its CMOS equivalent (80C31) and the six EPROMs are replaced by four CMOS EPROMs of type 27C256. All these changes together reduced the power consumption of this board from 400 mA to 80mA.

The input/output board uses the same processor (80C31). For program storage one EPROM of the type 27C64 (8 kbyte) is used. One RAM of type 6264 (8 kbyte) serves for the scratch-pad memory and the non-volatile memory. To make this memory non-volatile it is permanently powered. This board is connected to the speech board through the serial port of the processor. Input and output are performed through the keyboard and the LCD display. The keyboard switches are connected in a matrix configuration. This matrix is scanned and decoded by the processor instead of by means of a separate keyboard controller to reduce component cost and power consumption. The auxiliary components needed are a latch and a buffer. Both are mapped in data memory space. We can now write an 8-bit pattern into the latch. The output of this latch is connected to the keyboard matrix. The eight output lines of this matrix are led to a buffer that can be read by the processor. The combination of the data written to the latch and the buffer data identifies up to 64 keys. In order to be able to activate the device the processor has to write FF(hexadecimal) to the latch before going power-down. The signal to activate the device again is derived from the eight output lines of the matrix by means of eight diodes which perform a logical OR function. This signal is both fed to an interrupt input of the processor and to the power control circuitry. When the system is on, the processor notices the key-action through the interrupt. When the system has been automatically switched off, a key-action switches it on again.

For the LCD display the Epson EA-Y40087AT was chosen. This display is large enough (8 lines of 40 characters) to give a good overview of what has been typed. In addition it only needs a single power supply

and has a built-in character generator. This display is also mapped in data memory space. The audio amplifier is a bridge amplifier using two LM388 integrated audio amplifiers. This circuit delivers 2.6 Watt of output power at a supply voltage of 9.6 Volt. It is switched on only when the system actually speaks because it is a major power consumer. To avoid audible effects due to this switching the loudspeaker is connected through a relay-contact. This relay, normally closed, is activated for 100 ms during the switching on or off of the amplifier. This board also contains the power control circuitry. The manual on/off switch is the main power switch. Independent of this switch, backup power is provided for the keyboard latch and the RAM. Behind this switch an electronic switch is placed by which the device is switched off (by the processor) or on (by the keyboard). A second electronic switch switches the power for the speech board and the audio amplifier, which are powered only when the system speaks. The power circuit also contains a low-voltage detection circuit, which informs the processor when the batteries need recharging. The power consumption of the whole system is:

system manually switched off	80 μA
system automatically switched off	300 μA
system on, no key pressed	12 mA
system on, key pressed	30 mA
speaking (zero audio volume)	120 mA

Power consumption reaches a maximum when a message is spoken and depends mostly on the audio output. In this case peaks up to 500 mA can occur. When we compare these power consumption values with the battery capacity (500 mAh), it is clear that the system can be used for a whole day without recharging. With an average active daytime of 12 hours, about 150 mAh (12 hours $\times$ 12 mA) will be consumed if the system is active all the time (which is a worst case assumption), so there remains enough capacity for speaking, which will be done only a fraction of the time.

4.3.4 Software

The keyboard-to-speech process can be divided into four functional modules. The first is an input module that provides the user interface. The linguistic module next performs certain operations on the input and

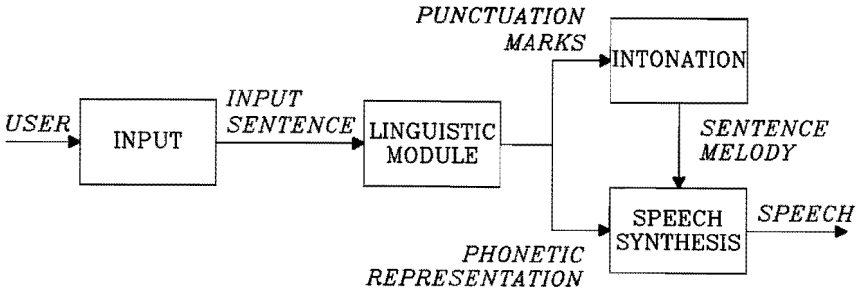

Figure 4.5: Software structure.

generates a phonetic output sequence. The third module is the speech synthesis module that generates the actual speech. Finally, a sentence melody for this speech is generated by the intonation module. This division is mainly a functional one; the actual software is not quite structured as suggested here; especially the distinction between the speech synthesis module and the intonation module is not that clear. The functional division is illustrated in figure 4.5. The input module runs on the input/output board, the other modules run on the speech board.

All software in the system has been written in assembly language, because at the time no high level language compiler was available for the processor used. The input module was directly written in assembly language because this module is strongly hardware-dependent. The only exception is the number names algorithm, which has been translated from Algol. The other modules were all developed on a VAX 11/780 computer in Pascal and then rewritten in assembly language.

Input module

The purpose of the input module is to provide a user-friendly user-interface. In addition this module looks after the hardware power control. This is performed by the execution of the following loop. When there is nothing to do (no key pressed and no speech output in progress) the

program execution halts by putting the processor into the idle mode. In this mode power consumption is reduced. The system is still powered, so the display is still showing its information. Program execution resumes when an interrupt occurs. If the interrupt is caused by the auto-power-off timer in the processor, the system is switched off. If the interrupt was caused by the keyboard, idle mode is terminated and the keyboard is scanned to determine which key was pressed. If the speak key was pressed, the speech board is switched on and the current sentence is sent to it. The program then waits until the speech board finishes speaking, switches it off and goes into idle mode again. The other keys result in the corresponding character being put into the input buffer for the alphanumeric keys or in the appropriate action being performed for the function keys. The actions performed can be edit operations on the input buffer or storage functions operating on the memory.

The input module also takes care of number translation. Numbers (up to six digits) are converted into number names before being sent to the speech board. This conversion is an implementation of the algorithm by Brandt Corstius [1965], which in turn is based on Van Katwijk's grammar for Dutch number names [Van Katwijk, 1965].

Linguistic module

Ideally, the linguistic module should translate normal, written text into its symbolic phonetic representation. This written text could include abbreviations, symbols like +,-,=,%,&, numbers, etc. Our linguistic processor does not reach this ideal, however. The reason for this is that full translation of written text of a certain language requires the development of a grapheme-to-phoneme conversion system for that language. For some languages this is a relatively simple task because of the more or less one-to-one correspondence between orthography and pronunciation. More difficult are other languages such as English and Dutch. Whereas various rule sets have been developed for English, work for Dutch is still going on. Although satisfactory Dutch pronunciation rules have become available lately, no rules were available at the time the present system was designed.

As a temporary compromise we created an input notation that resembles normal orthography but is closer to pronunciation. This notation

uses symbols normally available on a keyboard and was considered easy to learn. The first step in implementing this notation was to assign to each phoneme a character (or character combination) that represents this phoneme. These characters were chosen in such a way that their pronunciation is supposedly straightforward to Dutchmen. Table 4.1 gives a complete list of the phonemes in their representation in our system. Two symbols in this list (silence, glottal stop) are not phonemes, but are added because they are treated in this and other modules in the same way as phonemes.

With these phoneme symbols, words and sentences can be built by giving the proper sequence of symbols using the character keys on the keyboard. These keys all generate uppercase characters but the keys that normally give the **Q** and **Y** now yield the **g** and **e** respectively to make the input resemble normal orthography more closely. Also some additional keys are available. The glottal stop can be given by the key labelled "@". The "-" key generates a pause of 260 ms. The same pause is generated by the "," key. The latter also controls the sentence melody (via the intonation module). A question intonation can be obtained by ending the sentence with a "?".

To make the input somewhat closer to normal orthography, some pronunciation rules are added. These are straightforward rules, which apply without exceptions. An example is the application of the Dutch phonological rule that doubles the duration of certain word-final vowel characters (A, O, U). Another rule takes care for example of different notations for the same pronunciation (AU and OU, EI and IJ). So these rules are quite simple to implement and add to user-friendliness. Figure 4.6 gives an example of how an input sentence is translated by the linguistic module. The implemented rules are listed in figure 4.7.

```
WANNEER WORDT MIJN BUURO SCHOONGeMAAKT?

# W A N ER R W O R T M EI N B UU R OO S G OO N G e M AA K T
? #
```

Figure 4.6: Example sentence. First line is the input to the linguistic module, the lower line shows its output.

Table 4.1: Dutch phonemes (and some of their allophones) and their character representation inside the system.

character representation	example word	character representation	example word
#	*silence*	M	*m*eer
@	*glottal stop*	N	*n*eer
A	m*a*t	NN	ma*n*dje
AA	m*aa*t	O	r*o*t
AI	det*ai*l	OE	r*oe*t
AJ	m*aai*t	OJ	h*ooi*t
AU	k*au*d	OO	r*oo*d
AW	k*auw*	OR	w*oor*d
B	*b*as	P	*p*as
D	*d*as	g	ba*ng*
DJ	*dj*atiehout	R	*r*ok
E	l*e*s	S	*s*ok
EE	l*ee*s	SJ	*sj*aak
EI	r*ei*s	T	*t*as
ER	b*eer*	TJ	*tj*alk
EU	k*eu*s	U	p*u*t
EW	l*eeuw*	UI	m*ui*s
F	*f*ok	UJ	r*oei*t
G	*g*ok	UU	m*uur*
H	*h*ok	V	*v*uur
I	p*i*t	W	*w*eer
IE	l*ie*p	X	*g*oal
IW	k*ieuw*	e	d*e*
J	*j*an	Z	*z*eer
K	*k*an	ZJ	*j*ournaal
L	*l*ang		

```
B <at word-end> -> P
D <at word-end> -> T
DT <at word-end>-> T
I <at word-end> -> IE
A <at word-end> -> AA
O <at word-end> -> OO
U <at word-end> -> UU
SCH <before space, comma, period or e> -> S
CH -> G
IJ -> EI
OU -> AU
```

Figure 4.7: Implemented rules.

Speech synthesis module

The output of the linguistic module is fed into the speech synthesis module. This module receives input in the form of a sequence of phonemes, each phoneme coded as two ASCII-coded characters according to Table 4.1. For phonemes coded with one character the second character is a space in order to get a uniform length. From this phoneme sequence the module has to generate a sequence of diphones. With a few exceptions (e.g. the h-triphones), this task consists of combining all neighbouring phonemes. As a result, each diphone is coded as its two constituting phonemes (still in ASCII form). From this representation the speech data for that diphone have to be found.

This process is illustrated in figure 4.8 for the diphone *mɛ*. First, both phonemes are coded as a number instead of as ASCII characters. This is done by going through a table that contains all ASCII-coded phonemes. The index in this table is used as the number code. The two numbers of the phonemes forming this diphone are then joined to form a unique diphone code. With this code the diphone information table is searched, which contains the diphone code for each diphone, its start-address and length in the speech data table and the phoneme boundary for use in the intonation module. When the information block of this diphone is found,

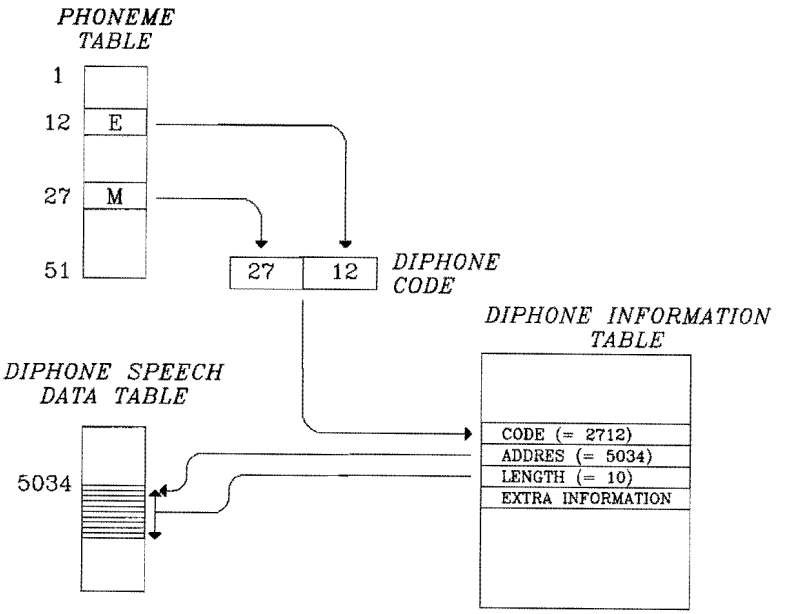

Figure 4.8: Speech data access.

we fetch the data for this diphone from the speech data table. These speech data are then sent to the synthesizer, frame by frame. Because the fundamental frequency was not stored in the diphones (see 4.3.5) an intonation module is present that generates a fundamental frequency for each frame. This frequency value is inserted in the speech data frames before they are sent to the synthesizer.

Intonation module

Prosody has a number of aspects of which the fundamental frequency (intonation) is generally considered the most important. This implies that in order to come up with natural-sounding speech we have to generate at least a pitch contour which is a description of the fundamental frequency for the whole utterance. Because this contour is based on the structure

of the sentence as a whole, it is useless to take the original pitch from the diphones. Instead a pitch contour is to be generated separately, based on the syntax and semantic contents of the utterance. This requires complex rules and knowledge to develop these rules or extra information from the input. In this project the choice is made to use extra input information, from which a contour has to be generated. Basic elements for this contour are generally considered to be rises and falls of the pitch. The specification of these rises and falls of the pitch requires comprehensive analysis of natural speech. This is language-dependent. An analysis of Dutch intonation had already been performed at our institute ['t Hart and Cohen, 1973]. In this analysis the natural pitch contours were replaced by stylized contours that were perceptually equivalent to the original ones. In terms of such stylized contours an "intonation grammar" was formulated ['t Hart and Collier, 1975], consisting of an inventory of discrete pitch movements plus a set of rules combining these movements into complete pitch contours, thus generating acceptable Dutch sentence melodies. These patterns consist of rises and falls of the fundamental frequency superimposed on a declination line.

Declination is the name of the effect of a gradually falling frequency throughout an utterance [Cohen, Collier and 't Hart, 1982]. For long utterances (> 5 s) the difference between start and end frequency is constant, for shorter utterances it depends on the length of the sentence. This behaviour can be expressed in the following formula (D is the declination in semitones/second, t is the sentence length in seconds):

$$D = \frac{-11}{t+1.5} \quad t \leq 5\ s$$

$$D = \frac{-8.5}{t} \quad t > 5\ s$$

It was found that all utterances of one speaker end more or less at the same frequency. For male voices 75 Hz is a good approximation for the end frequency of utterances without a final rise.

From all the patterns described in the grammar of Dutch intonation, six basic patterns were chosen as building blocks for our intonation contour. The specifications of these patterns are as follows.

All rises and falls occur between the basic declination line and a parallel line 6 semitones above this basic line. The accent-lending rises start 70 ms before vowel onset and have a duration of 120 ms. Accent-

lending falls also have a duration of 120 ms. They start 80 ms after vowel onset when used in one syllable in combination with a preceding accent-lending rise and 30 ms after vowel onset when not immediately following an accent-lending rise. Final rises and continuation rises start 120 ms before end of voicing (beginning of unvoiced part or silence) and have a duration of 120 ms.

In the intonation module synthetic pitch contours are generated from these basic movements. Selection of the movements is controlled by special markers in the input in the following way.

As previously mentioned, the input to our system is enriched with extra symbols for the intonation module. This extra information consists of punctuation marks: a special accent symbol, comma and a question mark.

The accent symbol is used to indicate the syllable that should be provided with an accent-lending pitch movement. Since this movement is related to the vowel position within the syllable, it is realized on the first vowel following the accent symbol. When an accent symbol is detected in the linguistic module, the position of the onset of the first vowel to be processed is stored as an accent position. This position is stored as a time interval from the beginning of the utterance to this vowel onset. This time interval can easily be calculated because from all preceding diphones the duration and the phoneme boundary (to find the vowel onset position) are stored in the diphone info table. The accent symbol is stored together with this time interval.

The comma generates a continuation rise pitch movement (after which the pitch has to fall back to the basic declination line); the question mark at the end of a sentence yields a rising pitch at the end of the sentence. Since this pitch rise has the same form as the continuation rise, the comma and the question mark are treated in the same way in the intonation module. When a comma or question mark is detected in the linguistic module, current and preceding diphones are searched backwards until a voiced speech frame is found, which becomes the end of the pitch rise. This position is stored together with the comma symbol, for at this point there is no need anymore to distinguish between comma and question mark.

In this way the input for the intonation module consists of a number

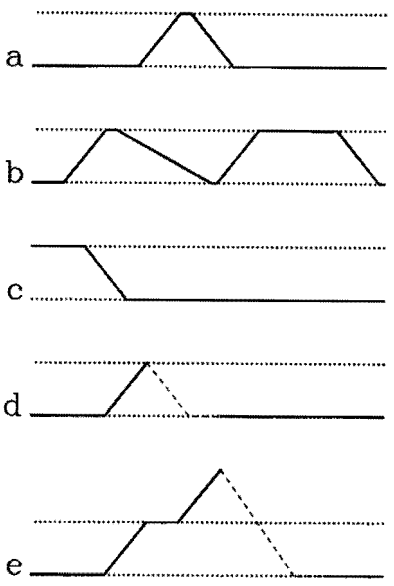

Figure 4.9: Basic intonation patterns. The declination effect is omitted.

of accent marks with their position in the sequence of speech frames to be synthesized. From this information selections are made from the six basic intonation building blocks using the following algorithm. Selection of the pitch movements depends first on the number of accent symbols in the sentence or part of a sentence, boundaries being the beginning and end of the sentence and the comma. One accent symbol leads to an accent-lending rise and fall (Fig. 4.9a). Two or more accent symbols lead to an accent-lending rise followed by a slow fall for each accent symbol except for the last two symbols. The last but one accent symbol leads to an accent-lending rise and the last one to an accent-lending fall (Fig. 4.9b). One exception to this selection procedure occurs when the first accent position in the sentence is within the first 70 ms. In this case the pitch contour starts at the higher declination line and an accent-lending fall is generated for this accent symbol (Fig. 4.9c). The

comma and question mark lead to a continuation rise (Fig. 4.9d); the fall following the continuation rise lies in the voiceless speech frames or silence following it. If the time interval between the last accent position in a sentence part and the following comma or question mark is less than 250 ms, two rises are generated as an accent-lending rise and a continuation rise (Fig. 4.9e), thereby rising above the higher declination line. The remaining accent symbol(s) in this sentence part are treated as described before.

In this way a piecewise linear pitch contour (on a logarithmic frequency scale) is generated. The starting time and slope of each linear segment are stored. The slope of the declination line is then combined with the pitch contour. With this information the fundamental frequency for each frame is calculated before it is sent to the synthesizer. Figure 4.10 gives an example of an input sentence and the corresponding generated intonation contour.

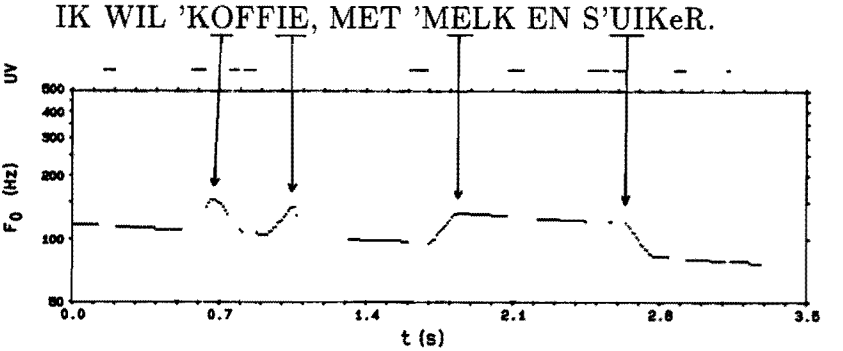

Figure 4.10: Example of an intonation contour.

Sentence prosody does not consist of intonation alone. Another aspect is temporal organization: in natural speech vowel duration and to a lesser extent consonant duration vary depending on their location in the word and sentence [Elsendoorn, 1985]. These kinds of variation in duration are not implemented in this system for two reasons: no satisfactory duration

rules were yet available and durations cannot be adjusted in the speech synthesizer used (MEA8000).

4.3.5 Speech data acquisition

Speech output in the speech synthesis module is generated by means of diphone concatenation. This requires the presence of a complete diphone inventory in the system. So first of all we have to prepare this diphone inventory and then store it in the system in such a way that it can be accessed by the speech synthesis module. This diphone inventory was available from another project [Elsendoorn and 't Hart, 1982, Elsendoorn, 1984]; only the second step was necessary for this special purpose.

The whole process of creating a diphone inventory (as done by Elsendoorn) went as follows:
All CV and VC diphones (C: consonant, V: vowel) were segmented from trisyllabic nonsense words. These words were of the general C∂CVC∂ type where the second syllable was stressed and the three consonants were identical. These words were spoken in a carrier phrase to get a natural speech rate and carried sentence accent. Subsequently these sentences were digitized with a sampling frequency of 10 kHz and stored in a VAX computer. The nonsense words were segmented from the carrier phrase and then analysed by 10th order LPC (see section 2.2). The LPC parameters were transformed into formant frequencies and bandwidths. For every frame (10 ms) 12 parameters were stored: voicing (V/UV), amplitude (G) and five formants with corresponding bandwidths ($F_1 - -F_5, B_1 - -B_5$). The pitch (F_0) was discarded because the original pitch is of no use when we create new utterances by concatenating the diphones. The resulting bit rate in the VAX computer is then approximately 12 kbit/s. From the stressed syllable two diphones were segmented, a CV diphone and a VC diphone; Figure 4.11 gives an example of the diphone $m\varepsilon$.

For the current inventory the segmentation was done by hand, but later an automatic segmentation program became available [Van Hemert, 1985]. The consonants were cut in the middle, whereas in the vowel part the truncation point was chosen in such a way that for a given vowel the vowel part of all CV diphones was identical in duration. This means that

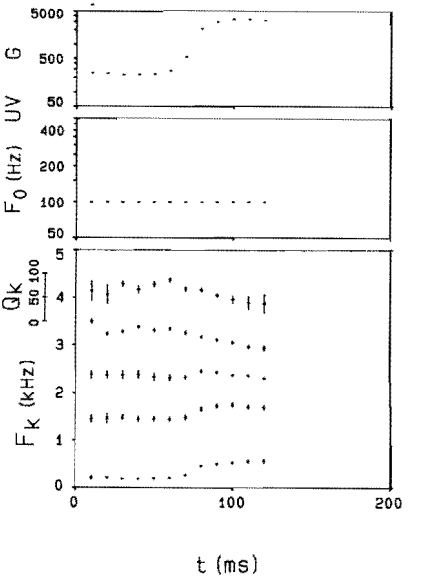

Figure 4.11: The diphone *mɛ*.

the total duration for a given vowel automatically varies, depending on the following consonant, as is the case in natural speech [Peterson and Lehiste, 1960; Elsendoorn, 1984]. For segments involving the /h/, segments were used consisting of three instead of two basic elements (vowel or consonant part, complete /h/ and vowel part), because it was found impossible to generate acceptable speech with diphones incorporating the /h/. These segments were called triphones. In addition to the CV and VC types, diphones including the glottal stop and VV diphones were segmented. Initial and final diphones were segmented from monosyllabic nonsense words spoken in isolation. Existing words were used for CC diphones containing consonant combinations. Also unstressed /*ə*/ diphones were included. After this each diphone was inspected (e.g. for voicing errors) and corrected where necessary. The parameters of the corrected diphones were then quantized for use with the Philips MEA8000 formant synthesizer chip, thereby reducing the bit rate to 2 kbit/s.

At this point we had a complete diphone inventory, consisting of a

number of coded speech frames for each diphone. To store this inventory in our system, the speech data for all diphones were put together in one table (68 kbytes). A second table (12 kbytes) was created containing all the information necessary to locate each diphone in the first table. These tables were then stored in EPROM in a form suitable for the speech synthesis module.

4.4 Evaluation

4.4.1 Introduction

The purpose of the Tiepstem was to get an impression of the usefulness in Dutch of a speech communication aid using speech synthesis. We wanted to find out whether speech synthesis by diphone concatenation is socially acceptable for this purpose and we wanted to learn more about the input aspects (spelling, extra input symbols such as accents, time-saving facilities such as an editor and memory). Finally we wanted to get some idea of the categories of speech-impaired persons for whom this kind of aid can be useful.

For this reason two copies of the Tiepstem were built and subsequently used in practice by potential users. This evaluation was carried out in cooperation with the Institute for Rehabilitation Research (IRV, Hoensbroek). After about half a year the results so far were written down by IRV in a report [Speth-Lemmens, 1987]. After this half-year evaluation period the evaluation was continued whenever potential users were available.

4.4.2 Procedure

We carried out the evaluation in the following way. The device was introduced to the user and/or his therapist by the author and the responsible therapist from the IRV. In nearly all cases all features of the device were demonstrated. A user manual was left for later practice and reference. Only the users with good cognitive and motor skills were able to operate the device on their own from this moment. The other users were helped and/or trained by their therapist. During the evaluation period it was always possible to contact IPO or IRV for further explanation or when

problems occurred. At the end of the evaluation period the use of the device was discussed with the user and/or his therapist, again by both the author and the IRV therapist. A questionnaire was used as a guide during these discussions [Speth and Oostinjen, 1986].

4.4.3 Participants

The participants in the evaluation were expected to have the minimal abilities required to operate the device. A survey of all the participants at the evaluation is given in Table 4.2. Numbers 1 – 9 are participants during the initial half-year evaluation period, numbers 10 – 11 participated after this period.

4.4.4 Results

Because of the limited number of users and the short period they used the device the results of this evaluation do not allow generally valid conclusions can be drawn from them. Although in all cases the questionnaire was answered, knowledge of the user, the environment etc. is necessary to interpret the answers. This interpretation is presented, in the next section where we present the results of the evaluation as can be concluded from the answers and comments of users.

The results of the evaluation can be divided into four categories:

1. Usefulness:
   1-Technical usefulness (e.g., malfunctions, battery capacity)
   2-Practical usefulness (e.g., operation, intelligibility, portability)
   3-Personal usefulness (e.g., communication speed, frequency of use)

2. Selection criteria of the users (e.g., physical abilities, cognition)

3. Therapy aspects (e.g., training phase, determination of the vocabulary)

4. Environmental reactions (e.g., reactions to speech output)

The results of the evaluation are discussed under the forementioned categories.

Table 4.2: Survey of participants in the evaluation and the respective periods of use (up to 1-1-1989).

no.	sex	age	diagnosis and limitations in functionality and motor abilities	communi-cative skills or aid	uses wheel-chair	weeks of use
1	f	22	cerebral contusion reasonable arm/hand function concentration and memory problems	Canon communicator, gestures	y	1
2	m	24	cerebral contusion limited arm/hand function slow and without initiative	Canon communicator, gestures	y	1
3	m	19	cerebral contusion limited arm/hand function slow		y	2
4	m	57	CVA limited arm/hand function	gestures, some words, writing,	y	4
5	f	43	laryngectomy	whispering, Servox, writing	n	1
6	m	34	cerebral contusion limited arm/hand function	Sharp	y	8
7	m	43	laryngectomy with tongue removal	home computer	n	8
8	m	58	laryngectomy (post operative)	gestures, writing	n	1
9	f	70	laryngectomy (post operative)	gestures, writing	n	1
10	m	40+	laryngectomy with tongue removal	writing	n	
11	f	40+	amyotrophic lateral sclerosis limited arm/hand function	Canon communicator	y	20

1.1 Technical usefulness.

- The only serious technical problem that occurred during the evaluation was an occasional loss of the memory contents. This (hardware) error was corrected during the evaluation.
- Some glued connections were broken during rough handling.

1.2 Practical usefulness.
The practical usefulness includes aspects such as the intelligibility and the social acceptability of the speech and the ability of the users to operate the aid.

- Usefulness of the device was mostly limited by its input. Very often this input was considered to be too complex. The pseudo-phonetic notation often caused trouble, which resulted in misspelled input and consequently less intelligible speech. The use of glottal stops was far too difficult for all users so they were not used at all.
- Another problem was the placement of accent markers. Although this was not too difficult for some users, even they often did not use this feature because of the extra time needed for correct placement.
- The edit facilities, however, were generally used and appreciated; the storage facility was only used by users with good cognitive skills.
- The users with a limited arm/hand function had problems with operating the keyboard.
- The speech quality was generally judged moderate. Often the wish was expressed for a more personal voice (male/female, dialect). Because of the problems with the speech quality the addressed people often used the display to read the message.
- As far as the appearance of the device was concerned it was often judged to be too big or heavy for easy transportation. But this problem could be largely relieved by some extra features such as a carrying handle and mounting facilities (e.g., for wheelchair or bed).
- The battery capacity and the audio volume were judged to be sufficient.
- The screen was judged big enough and the users were satisfied with the presented information and the legibility.
- The auto power-off facility made it difficult to prepare for a conversation because then information on the screen is lost when the device switches off.

- According to one user the loudspeaker should be directed towards the listener.

1.3 Personal usefulness.

In order to describe the personal usefulness (i.e. does the user benefit from the use of the aid) we have to look into intensity of use, situations of use and personal feelings towards use.

The frequency of use varied from more than 40 times a day to only during therapy sessions. All users found communication through this device not fast enough and preferred the method they were used to (pen and paper, Canon or Sharp communicator).

2. Selection criteria for users.

The number of users was too limited to draw general conclusions but some general remarks can be made. The group of users with cerebral contusion showed that this kind of aid requires more motor and cognitive skill than is generally present in this category. The aid was refused in a rehabilitation centre for children because its pseudo-phonetic input would interfere with normal language instruction.

3. Therapy aspects (training phase etc.).

From our own experience it became clear that introduction of the device can be a very time-consuming process. For most users, everything had to be explained step by step or continuous help was necessary.

4. Reactions of the environment.

Reactions of the environment were (especially in the beginning) generally positive. Because of the problems with the speech quality this enthusiasm decreased somewhat after time.

4.4.5 Discussion

The main conclusion from the evaluation is that the pseudo-phonetic input spelling that we used in this model is not suitable for easy use by naive users. This input spelling not only caused problems during input but also influenced the speech quality.

The speech quality was judged intelligible by the more successful users. In this respect our initial idea to use the diphone concatenation technique for speech synthesis still holds. For a better evaluation of the speech quality the input problems have to be solved first because it not clear yet to what extent these input problems contribute to the speech quality.

As we had expected, communication rate turned out to be an important factor. The result of this was among other things that users stuck to the straightforward method of entering a message and pronouncing it. The use of extra features (accents, glottal stop, memory) was a possible cause of mistakes and delay and these features were therefore not used. The problems with accent mark and glottal stop have been discussed in the previous section. The memory facility was too difficult for some of the users considering their problems with other input aspects. The other users were in most cases not sufficiently motivated to use the memory. This could be due to the problems mentioned before, which made it difficult to communicate using the device. Users were then so busy trying to get communication working that they did not bother about additional features not related to the basic operation of the aid. The use of the memory was also hampered during the beginning of the evaluation by some technical problems.

The above mentioned memory malfunction was the only technical problem that was encountered. For the rest the device functioned without problems. The battery capacity, audio volume and screen size were all satisfactory. The case also functioned well, apart from some glued connections breaking down. The users, however, handled the device with extra care because they considered it to be a special and expensive item (which indeed it was). They asked for a more robust case with more transport and fixing facilities and suitable for outdoor use. If possible they would like it to be smaller.

The user manual was not much used. All users were given a personal introduction to the device and started from this information. The manual

was perhaps too complex for them. The motivation of the user (and his therapist) played an important role in the use of the manual.

Because of the great variety found in the user's responses (ranging from "worthless" to "I cannot do without it") the question arises if all the participants in the evaluation can be considered to be potential users for this kind of aid. It is not clear yet whether the operation of this kind of aid is too difficult for some of them or whether this particular aid is not user-friendly enough. In the evaluation of a successor of this aid we have to be careful, therefore, that participants have sufficient motor and cognitive abilities. Otherwise not all available features will be used and we might wrongly conclude that these features are useless for all users.

After all we can conclude that this evaluation provided us with a first impression about the use of this kind of aid. Although we expected ease of operation to be an important factor for practical use, the evaluation made it perfectly clear how easy this operation should be. In this respect we may even conclude that it is impossible to come up with valid design requirements without this practical knowledge. So despite the compromise we had to make (pseudo-phonetic input) the evaluation was indispensable in this project.

From the results of the current evaluation we can conclude that a successor model should meet the following requirements. Priority number one is an easier input notation, preferably normal orthography. Automatic accent placement would come in very handy. The second point for improvement concerns the memory facility. This facility should be easier to operate than the current one. Thirdly a good, self-explanatory user manual should be made. Further points for attention are the case of the device and the speech quality. A choice from more than one voice (both male and female) and the possibility of personal adjustments (speed, pitch) would be appreciated.

Chapter 5

Combination with the Pocketstem

5.1 Introduction

As briefly mentioned in chapter 1, the possibilities for synthetic speech in a speech communication aid, the Pocketstem, are also being investigated in a parallel project. In that project, however, the focus is on ease of operation, which led to a restricted vocabulary and the use of precompiled speech messages. This chapter describes a combination of this aid with our Tiepstem. The goal of this combination is to create the possibility to load the Pocketstem with messages that are prepared on the Tiepstem. This removes some of the (temporary) limitations of the devices developed in the two projects. This in turn allows us to gain additional experience in practice.

The removal of some of the limitations will also extend the potential group of users. The Tiepstem is only useful for a certain part of the whole group of speech-impaired persons (chapter 3). This is due to the extra handicaps of many speech-impaired persons. In order to visualize this division of the user group, we present our target group in a two-dimensional diagram (see figure 5.1). In this diagram we set out the most important selection criteria along the axes, namely the motor and cognitive abilities. Although such a diagram is not very exact, it helps to indicate the target group for the Tiepstem and for other aids within the complete group of speech impaired.

For operating the Tiepstem a rather high level of both motor and cognitive abilities is necessary. The required motor skill is a hand or other motor function for operating a QWERTY keyboard at acceptable speed. The required cognitive skill is the ability to formulate and spell

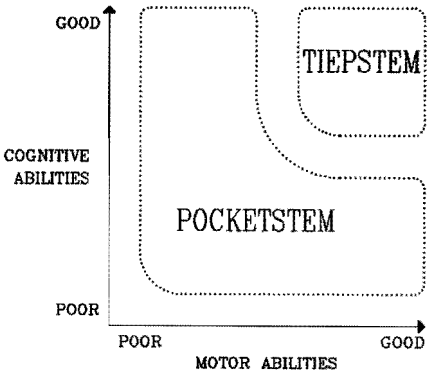

Figure 5.1: Possible usefulness of the Tiepstem and Pocketstem in relation to remaining abilities.

sentences and some linguistic knowledge to use the pseudo-phonetic input correctly and place accents at the right places. So the target group for the Tiepstem is found in the upper right corner of the diagram.

Another part of the whole group of speech impaired are potential users for the Pocketstem [Waterham, 1989], which focuses on ease of operation. Ease of operation is achieved by restricting the vocabulary of the Pocketstem to a fixed set of 28 messages. For its operation both the required motor and cognitive skills are less than those required for operation of the Tiepstem. The required motor ability is that the user is able to select and press one of the keys. The required cognitive skill is the ability to make the link between an idea and the corresponding pictogram. The potential users of the Pocketstem are found in the lower left corner of the diagram. The figure offers a good visualization of the fact that there are still speech-impaired persons who cannot immediately be considered to be potential users for one of these two kinds of speech communication aids. Use of the Tiepstem may be practically impossible because of a lack of motor and/or cognitive skills. In both cases operation of the device may still be possible, but at such a low rate that it is

unacceptable for communication purposes. The Pocketstem is very easy to operate and can also be used by people with more motor and/or cognitive skills than necessary. The problem in this case is that these users may feel soon limited by its restricted vocabulary.

We tried to create communication possibilities for those speech impaired who cannot be considered to be potential users for either the Tiepstem or the Pocketstem by combining these two aids. This combination makes use of the unlimited vocabulary system as a message-preparation system and of the easy-to-operate device as the actual communication aid. There are three potential applications for this combination: **1)** For users who find the operation of the unlimited vocabulary system too difficult in actual communication but can manage it when given enough time. **2)** For users of the fixed message device who have rapidly changing demands for the set of messages. Both groups consist of users who in our diagram lie between the two indicated potential user groups. **3)** A third application is the use of one Pocketstem by a group of users. In this case the message set can easily be updated for each user. **Ad 1)** These users do not have the required abilities to operate the Tiepstem in actual communication situations. This could be due to lack of sufficient motor abilities (the communication is too slow or too strenuous), or cognitive abilities (too many errors occur or too much time is needed to find the correct keys). These users could be able, however, to create a correct message when given enough time. This message could then be stored for later retrieval using the storage facility. Retrieving this message, however, requires at least three keystrokes. The operation of the Pocketstem on the contrary requires only one keystroke of one pictogram-labelled key, so it can still be used when the operation of the Tiepstem is too difficult. **Ad 2)** These users have more than the required cognitive abilities for operating the Pocketstem. In this case they will soon find the number of 28 messages too limiting for their communication purposes. Reprogramming of the message set, however, can at the moment only be done at our institute. This makes it rather difficult to change the message set frequently and almost impossible to adjust it to daily changing needs. If the reprogramming, however, can be done by someone in the neighbourhood of the user (e.g., relative, therapist) it is possible for the user to have his message set easily updated.

For this combination we will use the same project strategy as we did for the Tiepstem, that is we will evaluate this new combination in practice and use the results to update our initial requirements. An additional advantage of the evaluation of this combination in practice is that it enables us to form a better judgement of the quality of our diphone speech. Because in this case the Tiepstem is operated by a trained person (therapist) and not by the user, we expect that fewer input errors will be made. This allows us to validate our statement that many of the complaints about the speech quality were caused by input errors. Because the Pocketstem originally uses precompiled speech messages (speech resynthesis) we can compare diphone-based speech synthesis with speech resynthesis quality.

Section 5.2 describes the Pocketstem. The design specifications and the realization of the combination are presented in sections 5.3 and 5.4. An initial evaluation of this combination is presented and discussed in section 5.5.

5.2 The Pocketstem

The Pocketstem [Waterham, 1989] is contained in a standard case measuring 155 × 92 × 33 mm^3 and weighs 450 grams (figure 5.2).

On top of the case is a membrane keyboard, size 148 × 84 × 1 mm^3, containing 28 keys of 17 × 17 mm^2. The keyboard has an exchangeable card on which symbols can be attached. The audio volume can be altered by means of a (5-position) switch, which is located at the bottom of the case. The speaker is mounted in such a way that the output is directed to both one side and the bottom of the case. In this way the output is not hampered for instance by placing the Pocketstem on a table and is directed towards the communication partner.

Operation of the Pocketstem consists of selecting the desired message. The message selected is spoken by pressing a key, labelled with a symbol. These symbols are pictograms, specially designed for this purpose. Every message is stored in two different versions. When a key is pressed twice, the message is spoken differently the second time, in order to improve intelligibility and to avoid monotonous repetition. The Pocketstem is

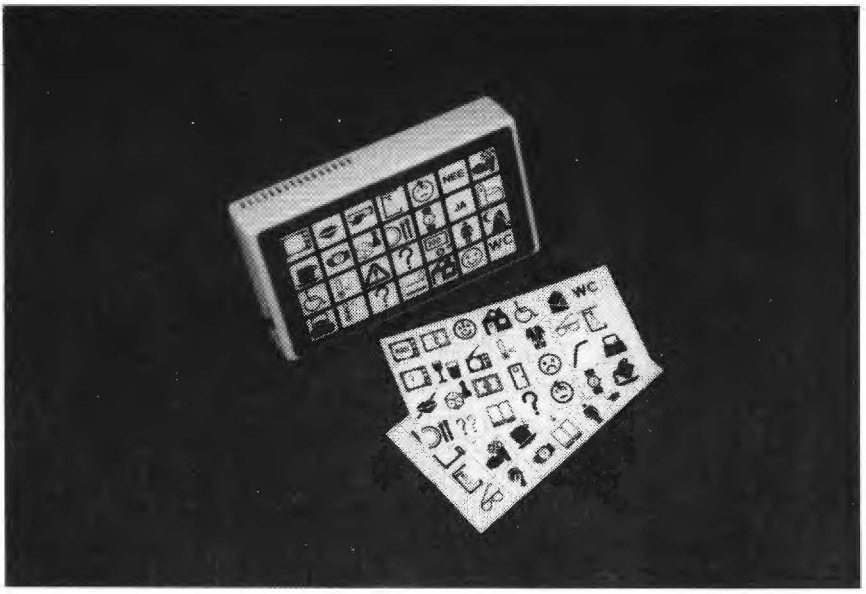

Figure 5.2. The Pocketstem

switched on by means of the on/off switch. When switched on, the Pocketstem immediately responds when a key is pressed by pronouncing the message stored under the key. The device has only to be switched off to prevent accidental talking, for instance during transport in a pocket or bag. For normal use the Pocketstem can be left on all the time, even when the battery is being recharged. A second function of the on/off switch is that it resets the microprocessor of the Pocketstem. Next to the on/off switch a battery low indication LED is situated. The audio volume of the Pocketstem can be selected with the five-position switch located at the bottom of the device. The Pocketstem has a vocabulary capacity of 28 messages. The messages were spoken by someone, recorded and subsequently coded (to lower the data-rate) and stored. Participants in the evaluation can select these messages in advance from a present set of about 200 messages, if necessary supplemented with personal messages.

5.3 Design specifications

The goal of the combination of the Tiepstem and the Pocketstem is to create the possibility to load the Pocketstem with messages created on the Tiepstem. The design specifications for this combination have mainly to do with this loading process. Because this is implemented as an extra feature in two existing devices we take these devices and their operation as a starting point.

First of all we have to define the possibilities this combination has to offer. The most straightforward one is the loading of a sentence into the Pocketstem. The version of the Pocketstem used offers the feature of two different versions of the same message spoken alternatively and we will preserve this feature and allow the programming of two messages under the same Pocketstem key. This can also be useful because intelligibility and naturalness of the diphone speech may be inferior to that of resynthesized speech. In addition to the possibility to program one sentence at a time it will be possible to program more than one (up to 28) sentences in one operation. This can be used when the same set of messages, stored in the Tiepstem, is programmed several times.

Because the Tiepstem has a display and the Pocketstem has not, we will use the Tiepstem display for displaying Pocketstem information when both devices are connected. For evaluation purposes the Pocketstem already keeps track of the number of times each key is used. Information about the usage of each key can also be useful for the user in the case of loading new messages. It allows him for example to replace the message that is least used. Another feature that will be implemented is to display the text of a certain sentence in the Pocketstem. In this way the user can recall how this particular message was entered into the Tiepstem, which is expected to be useful considering the present pseudo-phonetic input used for the Tiepstem.

Another group of design specifications has to do with the operations needed to activate the features mentioned. The major requirement for operation is that it is kept simple. In order to operate the combination of both aids the user should already be able to operate them separately. The extra knowledge needed to operate the combination should be as little as possible and in line with the operation of the separate devices. The

whole operation should be error-proof: wrong actions should not result in unwanted effects (e.g., malfunction). The data transport from Tiepstem to Pocketstem to reprogram a sentence should be realized within a reasonable time, of the order of magnitude of the time needed for preparing and speaking a message on the Tiepstem. All operations should be explained in a user manual. Because the evaluation of the Tiepstem showed that the user manual was not as good as needed, special attention has to be paid to the user manual of the combination.

As a last group of design specifications we can formulate some technical requirements. The changes in the original hardware should be kept to a minimum, especially for the Pocketstem where limited space is left for additional hardware. Consequently the connection and the connector should be small. The changes should not spoil the power-saving features of both devices.

5.4 Realization

5.4.1 Functional description

The version of the Pocketstem for use in combination with the Tiepstem contains a set of 28 messages in EPROM in the same way as the original Pocketstem. These messages, however, can be replaced by new utterances loaded from the Tiepstem and stored in RAM. The programming mode is entered automatically in both devices when the connection between them is made. Without this connection both devices act normally. In order to program a sentence into the Pocketstem the following operations should be performed [Storm, 1987]:

1. Connect both devices.
2. Prepare the message on the Tiepstem. This message can be edited, listened to etc. as normally done on the Tiepstem.
3. Press the key on the Pocketstem under which the message has to be programmed. Pressing a key once has no special effect, so it is possible to check various keys for their current message.
4. Press the same key for a second time to start programming. The message is first spoken by the Tiepstem, transmitted and then spo-

ken by the Pocketstem. The Tiepstem display indicates that transmission is in progress.

Because the Pocketstem can store two versions of each message, each key can be programmed twice. Therefore each odd attempt to program a key automatically results in clearing the previous stored sentence (both versions). Each even attempt to program a key results in a second version of a sentence being programmed under the same key. The total length of the 56 sentences in the Pocketstem should not exceed 120 seconds, otherwise an error message appears.

The protocol described so far is the basic operation for programming a sentence. In addition there are some more sophisticated features available. To activate these features, three pseudo-abbreviations are available on the Tiepstem: **ALLES, INFO** and **RESET**. The effects of these commands are:

No command One sentence is programmed.

ALLES All sentences stored in the Tiepstem under abbreviations **E01 - E28** (first intonation) and **T01 - T28** (second intonation) are programmed under the corresponding Pocketstem keys (1 - 28) in one operation when any Pocketstem key is pressed twice.

INFO No programming is done, but the Tiepstem display shows the text of the message and the number of times it has been used for each Pocketstem key that is pressed twice.

INFO, ALLES No programming is done, but the Tiepstem display shows the usage per message for all Pocketstem messages if any Pocketstem key is pressed twice.

RESET All reprogrammed sentences of the Pocketstem are cleared and the usage counters are reset when a Pocketstem key is pressed twice. The sentences originally stored in the Pocketstem EPROM are now back under the Pocketstem keys.

In addition to these three commands there is a fourth pseudo-abbreviation: the question mark. When it is entered in the abbreviation mode, the Tiepstem display shows the communication option currently in effect, i.e. what will happen when a Pocketstem key is pressed twice.

The **INFO** and **ALLES** commands are toggle functions: entering these abbreviations alternately set and reset the corresponding option. The question mark command can be used to display the current status. All options can be reset by pressing the Tiepstem's *wis scherm* key.

5.4.2 Hardware

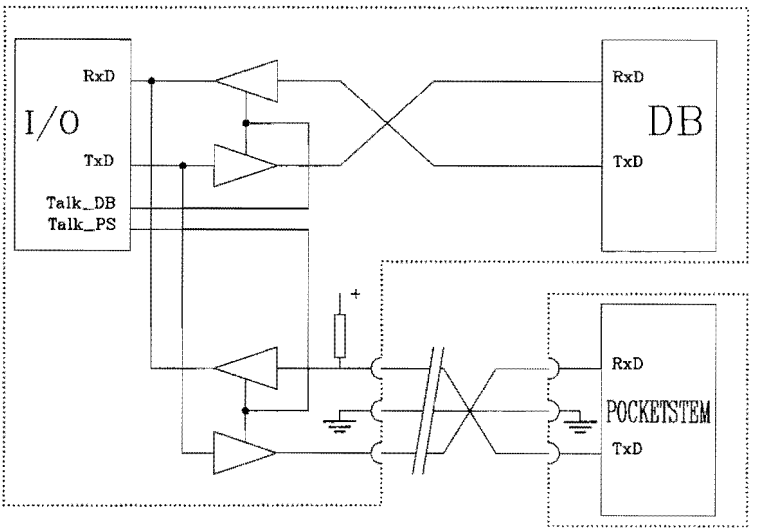

Figure 5.3: Adapted hardware of the Tiepstem

Two changes had to be made to the original hardware. The first was the realization of a communication channel to the Pocketstem. To keep this communication channel simple, serial communication has been chosen. The two processors in the Tiepstem both have a built-in serial port, but they already use this port to communicate with each other. There are three possibilities to establish a serial channel to the serial port of the processor of the Pocketstem:

1. use of an additional serial chip

2. switching of one of the available ports

3. use of parallel communication between the two processors in the Tiepstem.

The second possibility was chosen because it requires the least additional hardware and modifications of the original hardware. Figure 5.3 shows the realization.

Another change had to be made to the original Tiepstem that might require hardware modifications. To program the Pocketstem with speech data, these data have to be available in the Tiepstem. This is not the case in the original design. To make these data available three possibilities exist:

1. The data can be sent to the parallel port of the speech board processor which is not used in our application.

2. The data can be sent to the serial port of the speech board processor, during the time when there is no transport of text through this same port.

3. With some additional hardware the data can be retrieved when being sent to the speech synthesizer chip.

The first two possibilities require changes in the software of the speech board. Because this would create some practical problems, the third method was adopted. Each data byte sent to the synthesizer is latched and the input/output processor is interrupted (figure 5.4). The input/output processor can than read this byte and store it in its memory. When a complete utterance is stored in this way it can be sent to the Pocketstem.

5.4.3 Communication protocol

General

The communication between the Tiepstem and the Pocketstem uses only 7 bit ASCII codes. All ASCII control codes are reserved for commands, while the alphanumeric codes are used for data transport. Each data byte is split into two nibbles (one nibble is four bits) which in turn are

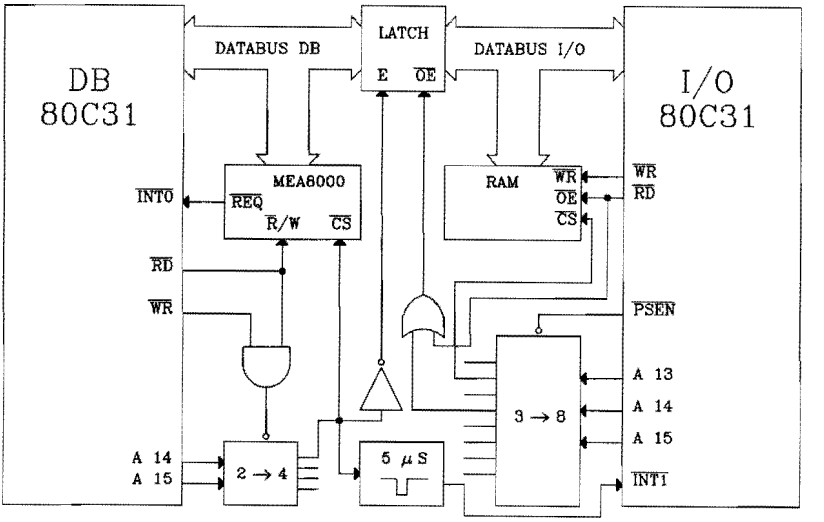

Figure 5.4: Speech data retrieval from synthesizer.

coded as two characters. For example the data byte with hexadecimal value 1A is coded as the characters "1" (hexadecimal value 49) and "A" (hexadecimal value 65) successively.

Communication from Tiepstem to Pocketstem

The communication is started when the same Pocketstem key is pressed twice. In this case the Pocketstem sends the code $< ENQ >$ and waits for the Tiepstem to respond with $< ACK >$. When there is no response within 20 ms, the Pocketstem assumes that there is no connection to the Tiepstem and speaks the "second intonation" of the selected message (the original action when a key is pressed twice). When the Tiepstem responds, communication takes place according to the diagram in figure 5.5. The meaning of the various control codes used is:

FF "ALLES" mode selected: one operation for a set of messages.

STX The Tiepstem starts speaking a sentence.

CAN The Tiepstem cannot speak this sentence.

SI/SO A first/second intonation is transmitted.

ETX End of speech data.

DLE Tiepstem asks for information from the Pocketstem (about usage of message(s)).

SUB Pocketstem has to go back to the original (EPROM) sentences.

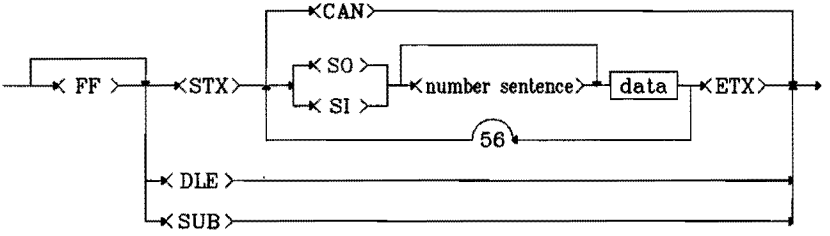

Figure 5.5: Communication protocol from Tiepstem to Pocketstem.

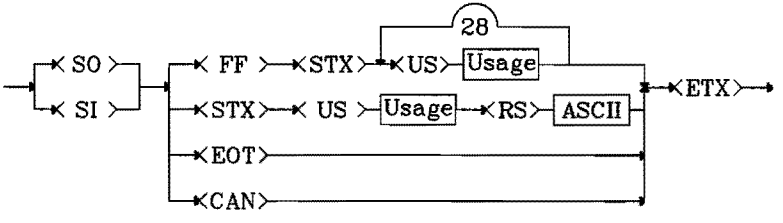

Figure 5.6: Communication protocol from Pocketstem to Tiepstem.

Communication from Pocketstem to Tiepstem

When communication has been started through the $< ENQ >< ACK >$ sequence as described in the previous section, the Pocketstem transmits $< SO >$ or $< SI >$, indicating whether a first or second intonation is needed. This depends on whether there are already zero, one or two messages stored under the key pressed. The Tiepstem responds with speech data or a command ($< DLE >$ or $< SUB >$, see previous section). The Pocketstem responds according to the diagram in figure 5.6. The meaning of the various control codes used is:

FF Usage data of all messages follows.

STX Start of usage data.

ETX End of usage data.

EOT Speech data received, end of communication.

CAN Communication cancelled by the Pocketstem.

5.4.4 Software

The main change in the Tiepstem software is the addition of a second speak routine, which (in addition to the functions of the original speak routine) takes care of the capture of the speech data sent to the synthesizer. This data capture and storage is done by the input/output board processor in an interrupt routine. Care has to be taken that this routine is fast enough to keep pace with the data transport to the synthesizer. When a whole sentence is stored, it is converted to ASCII-coded nibbles and sent to the Pocketstem using the protocol described above.

To switch the serial port of the input/output board processor two routines are available. These routines select the appropriate buffers (see section 5.3.2) and select the correct baud rate. The baud rate is 9600 baud for communication with the speech board and 1225 baud for communication with the Pocketstem. 1225 baud is the highest rate that can be achieved by both Pocketstem (clock 4 MHz) and Tiepstem (clock 11.059 MHz).

5.5 First impressions of practical use

The combination that has been made of the two aids has to be evaluated in practice. In connection with this work we want to find an answer to two questions: **1)** what new communication possibilities does it offer? and **2)** how does the combination work in practice? The first question deals with the new possibilities we foresaw for this easily reprogrammable aid. Through an evaluation we want to find out if this new combination indeed offers the forementioned (or other) new communication possibilities, or communication possibilities for new user groups. The second question deals with the actual realization of this combination and how it functions in practice. We want to find out if enough features are present and if they are easy enough to operate.

This new combination can be evaluated in the same way as the separate aids, but we have to make a choice between an extensive evaluation while we know that the Tiepstem will be succeeded by a more user-friendly one, and a short evaluation which can be used to check in a preliminary way the usefulness and user-friendliness of the programming. We chose a short evaluation with only one set because an extensive evaluation would cost a lot of time and work while we know beforehand that a second combination is likely to be available in the near future and would probably interfere with work on the successor of the Tiepstem. Furthermore the results of an extensive evaluation would only be available at the time the successor should be ready, so that limited benefits can be expected from such an evaluation.

The evaluation place chosen was a school class of mentally and motorically handicapped children. This was the best location available and although it was not one of the main target groups it seemed also a possible application for this new combination. The results of this evaluation are the following.

The aids were mainly used during speech training and as a "toy". Both the Tiepstem and the Pocketstem were used. The programming was done by the teaching staff and was considered easy enough by them. The user manual, which explained everything well enough, was needed when the programming had not been done recently. It would be appreciated if one could store more than one message set for the Pocketstem

in the Tiepstem. This would make it possible to use the reprogrammed Pocketstem in more situations.

About the operation of the Tiepstem for this purpose a few practical suggestions were noted. It would help the overview of the stored messages if they were presented in numeric sequence (E01, E02.....T28) instead of in the order in which they were entered, as is the case now. The process of altering a message in memory was judged too cumbersome. In this version this has to be done by retrieving a message, editing it, deleting it from memory and storing the edited version. Children judged the speech quality as good, adults as moderate but after a period of habituation as sufficient.

We can conclude that this evaluation was not extensive enough to come up with results about the part of the group of speech impaired for which it is suited. Nor did we get an answer to our first question: what new possibilities does it offer? We did however get an impression of how the combination works in practice. For therapists it appeared to be easy enough to program the Pocketstem with the Tiepstem. Because our initial requirements are not enough checked, and because a new model of the Tiepstem will become available, it will be worth while to carry out a second loop of the project set-up. In this loop a second evaluation has to be carried out. In order to get an answer to the question as to what new possibilities this combination offers, it should be evaluated (together with the Tiepstem alone) on a larger scale. Therefore this possibility will be implemented in the new version of the Tiepstem.

Chapter 6

Discussion and outlook

6.1 Introduction

In this chapter we discuss the main aspects of our project. Our basic question was to find out whether Dutch speech synthesis can successfully be used in a speech communication aid for the speech impaired (see chapter 1.1). We tried to find an answer to this question by, after an initial orientation, carrying out a loop of development, practical use and evaluation of such an aid. First we discuss in section 6.2 the general answer to the question. In section 6.3 we discuss the project strategy we chose, especially how it worked out for our project. During execution of the project we found that publicity aimed at therapists in the field of speech therapy and ergotherapy, and at others working in the field of aids for the handicapped, was essential to obtain enthusiastic cooperation from therapists and potential users. We discuss some aspects of publicity and our contacts in the therapy environment in 6.4. We look at the project results and compare them with other work done in this field in section 6.5. For the realization of our experimental model and for the experimental model and the prototype of the parallel project, circuit designs and software were realized, which are also useful in other applications of synthetic speech. We mention some of these applications in section 6.6. Finally, in section 6.7, we discuss remaining problems that require continuation of this project.

6.2 Project results

In chapter 1 we formulated our main problem as follows: can we construct a speech communication aid, based on speech synthesis for Dutch, such that it can be used with success by speech-impaired persons for diminishing the consequences of the handicap? At this point we attempt to answer that question. To do so we have to go through the project and see how the decisions, realizations and results contribute to our answer.

First, we found a problem in combining ease of operation of a possible aid with a large vocabulary. This fact, combined with the observation that our target group of speech impaired shows a great variety in remaining abilities and skills, led us to a choice: we should focus either on ease of operation or on a large vocabulary. Because the starting point in this project was the use of speech synthesis (by diphone concatenation), we focused on the large vocabulary of the aid. Although we did not neglect ease of operation, this decision limited the group of potential users to those that have the possibilities, both motor and linguistic to create messages by typing them on a keyboard. This is a limitation of the user group found as part of the answer to our problem.

A second problem is concerned with the acceptability of synthetic speech for this purpose. Acceptability is influenced by various aspects of the speech, such as intelligibility, naturalness, intonation, speech rate, individuality, etc. Some users considered the current speech quality acceptable for their purpose, but more users considered it not intelligible and natural enough. As already discussed, however, this result is probably influenced by the difficulties of the input notation used for this current model. This notation requires a good language skill in order to produce correct speech. When the input is not correct at the beginning, the resulting speech will certainly be less acceptable. These problems and the fact that some users find the speech acceptable led us to the conclusion that it is worthwhile continuing the investigation into the use of this kind of synthetic speech for this application.

A third problem has to do with the strategy of our investigation. We started our project with speech synthesis by diphone concatenation, but the realization of a complete Dutch text-to-speech system was not possible at that time, mainly because of the lack of Dutch grapheme-

to-phoneme conversion. Despite this incompleteness of the necessary techniques we decided to start our investigation and not wait until everything we could probably use was available. We can now look back and evaluate this decision. First we can conclude that we did not succeed in creating a generally useful communication aid, due to problems created by the lack of some techniques. On the other hand, by starting this investigation and creating a device with all the techniques available at the time we did gain much knowledge and practical experience. By creating such a device and evaluating it in practice we gained a better idea of the design requirements for this kind of aid. We also got a better insight into the user group that can be expected to be potential users for this kind of aid and into their demands. All this is especially useful because little is found in the literature about these aspects. From a more practical point of view we got experience in evaluating this kind of device and gained entrance to the potential user group and their therapists for evaluation purposes. Through this evaluation we also let possible users and their environment know what is technically possible, so they can get used to the idea of using synthetic speech and perhaps come up with better suggestions and requirements for such an aid.

Generally we can conclude that there is good hope that this kind of synthetic speech can be applied usefully in a speech communication aid for speech-impaired persons with reasonably good cognitive and motor skills. Although the aid developed during this project is already useful to some users, it should be redesigned to make it easier to use before it can become useful to a larger group of users. Further research, therefore, is necessary to perform this redesign and to look for new problems that may arise or were previously hidden behind more serious problems.

6.3 Project strategy

In chapter 1 we motivated our choice to adopt a project strategy that is commonly used in this field. The essential aspect of this strategy is that it consists of loops of design, realization and evaluation. Our main reason for this strategy was that we thought it to be a good way to incorporate the essential participation of the potential user group. We also considered this participation desirable because little information is found in the

literature about user requirements and we believed that the human factors are as important as the technical aspects. Our expectations about this project strategy were confirmed. In our project we executed only one loop, but this one loop already showed the benefits of this strategy. Especially the evaluation by full exposure of our experimental model to some potential users worked well and offered enough suggestions for us to start a second project loop.

In chapter 1 we also indicated the multi-disciplinary character of this research. Because we did not have all the required disciplines available at our institute, we looked for cooperation in an attempt to get a wider knowledge of the field of communication aids. For this reason we contacted the Institute for Rehabilitation Affairs (IRV, Hoensbroek) because of their knowledge of the rehabilitation situation, their contacts with therapists and their experience with evaluation in this field. This cooperation worked out well. It helped us to perform the evaluation in a uniform way and brought us contacts with potential users and therapists easier and faster than we could have achieved ourselves.

As already mentioned, a second, related, project was carried out more or less at the same time, focusing on ease of operation. Because less technical effort was necessary to achieve this goal compared to our project, in this project two project loops were carried out. This resulted in an experimental model and next a prototype model of the Pocketstem. Simultaneous running of these two projects offered several advantages:

- Parts of the electronic design could be used in both projects.
- The evaluations served in part both projects (especially concerning insight into the diversity of the user group and its specific requirements).
- As the evaluation strategy was the same, more experience and knowledge about the evaluation itself could be gathered and mutual contacts could be used, which meant that the purpose of the evaluation was known and the therapist already had experience with the evaluation the second time.
- Not only could experience be combined, but also the aids themselves could be combined relatively easily. This combination could prove

to be a good alternative for users who find the Pocketstem too limited, but the use of the Tiepstem is too complicated to realize an acceptable communication speed.

6.4 Contacts and publicity

Publicity, both international and national, was an aspect of our work [Deliege and Waterham, 1986a,b; Deliege and Waterham, 1988; Deliege and Waterham, 1989; Deliege, Speth-Lemmens and Waterham, 1988a,b; Deliege, Speth-Lemmens and Waterham, 1989; Deliege, 1989].

The benefit of international publicity lies in the fact that there is little information available in literature about other aids and projects. Getting into contact with colleagues working in this field created some possibilities for discussion and also response to our ideas.

National publicity, through presentations, articles and press interviews, served somewhat more practical purposes. It brought us into touch with potential users, therapists and others who are interested in this work (e.g., relatives, social services). These contacts were important for the evaluation, because through them we were able to select volunteer users with a variety of medical diagnoses. Furthermore these contacts gave us an impression of the actual situation and needs of the different groups of speech impaired. Apart from that, publicity informed those potential users and therapists who were not directly involved in the evaluation. National publicity also incorporates the informing of speech therapists and ergotherapists in their training. This training does not as yet contain much information about the technological status of synthetic speech or the use of micro-electronics in general, for the benefit of patients or for therapy.

Publicity has, however, a side effect that is both positive and negative. As the technical possibilities of such aids become familiar to therapists and potential users, their first question is where can these devices be purchased or when will they become available. This has the positive effect that we can show a potential manufacturer that there is a demand for such aids. The negative effect is that people become disappointed when they are unable to acquire what would be of great help to them, and which they know to be technically feasible. The chances of this

negative effect are reduced when a therapist is used as an intermediary between us and the patient. The therapist knows the user and can judge his attitude to a temporary use of the aid and has the authority to make its temporary nature clear. If after all it appeared undesirable to take the aid back, we let the user keep it on loan for an indefinite time.

We can conclude that well controlled publicity overall had a positive effect on the project. It is, however, advisable to handle publicity with great care, especially when it is publicity in the press. The fact that generally not enough time is allowed to discuss the subject intensively too often results in a demonstration in which it is not made clear that the aid is under development, what its merits and limitations are, and if and when it will become available.

6.5 Comparison with other work

The idea of a speech communication aid with keyboard input and offering an unlimited vocabulary is not new. As mentioned in chapter 3, various devices of this kind have been developed and some are even commercially available.

The main difference between these devices and our work is the language used. Because speech synthesis is mostly language-dependent, it is not possible to turn the available devices into a Dutch-speaking one. In our work Dutch speech building blocks and extensive knowledge of Dutch intonation are used to create a Dutch speaking device.

A second difference between these devices and our project lies in the speech synthesis. The synthesis techniques found in other systems are (in order of frequency): phoneme synthesis (by hardware phoneme synthesizer), synthesis by rule (by signal processor) and diphone synthesis. When we look at our technique of formant-coded diphones and a hardware formant synthesizer, we can conclude that the hardware is as simple as the hardware phoneme synthesizer, only the memory requirement is larger. The software complexity is also comparable with this technique. Speech quality, however, is much higher than with phoneme synthesis.

Another great difference is the extensive cooperation of potential users and therapists during the formulation of the design requirements and during the evaluation of the realized device. Documented evaluations

are seldom found for the other devices, with a few exceptions: Multi-Talk [Schildt and Sterner, 1986] and "Sadare's speech system" [Sadare, 1984]. These evaluations, however, are only a description of the use of the device by some (selected) users. In these evaluations the positive aspects of the device are highlighted, but points for improvement are seldom mentioned. Therefore it is often difficult to judge the success of these devices. From our evaluation we can conclude that ease of operation, and consequently, communication speed are the most important conditions for successful use. Although we implemented several facilities to make operation easy and fast (e.g., storage facility, large screen with edit facilities) the evaluation showed that operation of this device was still too complex. We may therefore expect that other devices with less attention paid to ease of operation and communication speed will be useful only to users with very good cognitive and motor skills.

The kind of (pseudo-phonetic) input notation we used is not found in other devices. Most of them use normal orthographic input and sometimes pure phonetic input. This is due to the fact that for many languages (especially English) rather well functioning grapheme-to-phoneme conversion systems are available. Little is known, however, about the functioning of these systems in practice, for example the number of mispronounced words during normal conversation or the effect of typing errors.

The idea of combining an unlimited vocabulary system with an easy to operate input system is also not completely new. We mentioned various devices of this kind in chapter 3 (as category 1b). The main difference, however, between them and our combination is that they combine both functions into one device while we use two separate devices that are complete, useful communication aids on their own. The main advantage of this approach is that the actual aid used for communication is much smaller than it would be if it had to incorporate the programming part. When necessary, the user does not even have to be aware of the presence of the programming facilities when these are too complex to be handled by himself. This avoids confusion by the user. Another advantage is that the programming device is not only useful as such, but also as a communication aid on its own.

6.6 Spin-off

The techniques and circuits developed within this project have also been used outside our project. The commercial availability of the speech synthesis board made it easy to use this board in other applications.

One of these applications is as a speech output system to a personal computer for visually impaired users, sold under the name "Inspect" [Nottroth and Van Jole, 1988]. This system consists of our speech synthesis board, which is put in a case with power supply, and some additional software for the personal computer. This system allows the user to have each input character spoken, or to listen to a line on the computer screen or to the whole screen. Text on the screen is sent untranslated to the speech board. This will in general result in an incorrect pronunciation of this text. For this application, however, the listener is always the same person (speech is used as an alternative communication channel). He can therefore familiarize himself with the pronunciation and learn to grasp the spelling from the spoken output. Use of the system by a number of speech-impaired persons is necessary to validate this idea. When the speech board refuses to speak a certain word because it results in an invalid phoneme combination (for which no diphone is present), the Inspect software switches to the spelling mode. In this mode, which can also be entered manually, words are spelled out using a spelling alphabet.

The speech board is also used as a speech output device for an existing communication aid for the speech impaired. This system ("Schrijfblok") was developed in Belgium and consists of a word selection system operated by two switches. To add speech output to this aid an investigation has been carried out, resulting in the application of the speech synthesis board [Konings, 1986]. In order to facilitate the input process, some effort has been put into the development of an input-processing program. This program does some limited grapheme-to-phoneme conversion.

Various parts of our circuit can be used for other systems generating synthetic speech. The general idea of using an inexpensive but powerful microcontroller in combination with a speech synthesizer was already known and applied. But realizing the circuitry in CMOS technology and thus saving power consumption is and has proven to be a universal basis for battery-powered speech-generating applications. The power

control circuit (or parts of it) can also be used in various designs. The same holds for the keyboard-scanning and the switched audio amplifier. Projects that have made use of these circuits include an update of the Typophone (an experimental model of a typewriter for the blind with spoken feedback) [Kroon, 1986; Neerven, 1987], a speaking household balance [Alberti, 1986], a talking version of the Possum communication aid [Krol, 1988], a speaking clock and an update of the Reflo-talk (speech output for a glucose measurement system) [Waterham, 1983; Steeksma, 1988]. The power-control circuitry and the switched amplifier are used in a general speech output device for a personal computer [Deliege, 1987]. This device, which uses no microprocessor, is connected to the printer port of a personal computer.

Parts of the software set-up (intonation contour calculation and storage, diphone access and storage) are used in other programs for diphone synthesis. Part of it is formalized in a diphone-standardization document [Scheffers and Ten Have, 1986]. The aim of this standardization is that software developed in the future can be easily maintained and that parts of it can easily be updated.

Apparently our development was not only useful for our application for the speech impaired but also for some other applications where it could be directly applied or used as a working example to start from.

6.7 Future

In a previous chapter we formulated the initial design requirements. These turned out to be not such a bad choice after all: the evaluation did not reveal requirements that were not really relevant. However, the priority of the various requirements has become much clearer now. It became obvious that ease of operation is a major condition for possible use. If operation is too difficult for certain users, they are so in trouble trying to get things working that communication fails to be achieved. Therefore our main attention should go to the operation and input of the aid. However, as we indicated before, ease of operation and technical possibilities (unlimited vocabulary) do not go together very well. To achieve easier operation than with our previous model we have two possibilities. We can reduce the number of available functions or we can

try to make the functions easier to operate. To make this choice we have to look at the functions we implemented in the previous model and how they worked out in practice.

We can divide the functions incorporated in the previous model into four categories. **1:** the input of a message from scratch. **2:** functions adding editing facilities to this input process. **3:** the function to store and later retrieve messages. **4:** physical operation: on/off switching, volume control etc. We will take a closer look at these categories and see to what extent they can be omitted or made easier.

1: the actual input of a message. It is obvious that this function will always be necessary, because it is the most basic expressive function. The most difficult aspect related to this function happened to be the spelling used that had to be learned by all users. For this problem a solution is available: the availability of Dutch grapheme-to-phoneme conversion software and CMOS 16-bit microprocessors make it possible to use normal Dutch input spelling for a new aid. This input will then also no longer use the special symbols needed in the previous version (pause, glottal stop). One problem remains: typing and spelling errors will still cause improper pronunciation. The evaluation of the new device has to show to what extent typing errors occur and what their effect on the speech output is. The grapheme-to-phoneme conversion system also includes automatic accent placement. Although this placement is not perfect, the trade-off between ease of operation and a better speech quality may favour the use of this automatic placement. This is a point of interest for the evaluation. A number-translation algorithm was also incorporated. It seems worth while to keep this function.

2: the editing functions. These functions are not necessary for the basic operation of the aid. If a message is entered correctly and then spoken, there is no need for editing. This is just the ideal situation, however. In practice typing and/or spelling errors are often made. The editing functions can then be used to correct the message before it is spoken when the user is aware that he made a mistake, or to correct it after it is spoken and apparently not understood. The user can then correct possible mistakes or change the message somewhat. The previous evaluation showed that the functions available in the previous model were understood and used by most users. There is no direct need, therefore, to

omit or change these features. However, because the previous evaluation showed the importance of the input facilities, it may be worth trying to make these edit facilities even more userfriendly. The current editor behaves like a computer screen-editor, which is probably not the most userfriendly one for users unfamiliar with this kind of editor.

3: the storage function. This function is also not necessary for the basic operation of the aid. A practical storage facility, however, can add a great deal to communication speed. The way this facility is implemented in the previous model appears to be too cumbersome to make it useful. If we are to keep this function, which seems worth while considering its increase of communication speed, it should be easier to operate. The development of an optimal user-interface , however, is not that easy and techniques to perform such a development systematically are not widely known. In our application things are even more difficult because we aim at the speech impaired, who can have other handicaps that influence their ability to operate an aid. Therefore this problem is being studied separately [Verhoeven, 1989], in an investigation which can thus focus completely on this aspect. The input process can be simulated (e.g., on a personal computer) and a short evaluation will show user performance of various storage-operating techniques. Due to the complexity of the grapheme-to-phoneme conversion software our new model will have to be programmed in a high-level language (Pascal in this case), so the developed storage software can easily be implemented in our new model when it is written in the same language.

4: the physical properties. The evaluation of the previous model showed a demand for a more protective and smaller case (both in use and during transportation). A case that meets these requirements is the kind of case found for laptop computers: an upper part with the display in it that can be set upright when in use or put over the lower part (thus protecting both the display and the keyboard). This kind of case is rather difficult to construct, however, and it is questionable whether an available case of this kind can be found. Whatever the possibilities may be, it is certainly necessary to pay more attention to the physical properties of the new model.

As far as the speech quality is concerned, there are some points for improvement. Some of these points found during the previous evaluation can be improved, e.g., naturalness of the speech is still improving [Collier and Houtsma, 1988]. The MEA8000 speech synthesizer chip we used is succeeded by now by the PCF8200, which has a better specification of the speech parameters (bitrate approx. 3 kbit/s) and the possibility to produce a female voice. Other points, especially those dealing with the personality of the voice, require further fundamental research. At the moment for each other voice (e.g., female, child) a complete new diphone set has to be made. Because speech research is still going on, improvements in the field of speech synthesis may become available in the near future. Because most of these research results (e.g., durational control, more lively intonation) will improve the naturalness of the speech, it is worth while following this research and implementing the results when appropriate.

Interesting results may also arise from fields outside speech research. Input prediction systems, for instance, may be a solution to the slow typing process.

Because we expect that this redesign will be useful for a number of users, attention should be paid to the aspects involved in commercialising the device. As we mentioned in chapter 1, this is not always an easy task. To facilitate this process, industrial contacts have been made at an early stage and not just after a successful evaluation.

Because this continuation involves many things to do (as least as much as for our first model) this is being done in a new project. This thesis ends therefore at this point with the starting conditions for the new project. One thing to be done, however, has already been carried out, namely the combination of the already mentioned grapheme-to-phoneme conversion with our diphone synthesis and the implementation in a microprocessor system. This work is described in Appendix C.

References

Unpublished literature is marked with an *.
It may be retrieved from the institute mentioned.

Alberti J.A.M. (1986)*, "Ontwikkeling van een sprekende weegschaal t.b.v. blinden en slechtzienden", *Afstudeerverslag Technische Universiteit Eindhoven, vakgroep EME.*

Allen J., Hunnicutt S., Carlson R. and Granström B. (1979), "MITalk-79: The 1979 MIT text-to-speech system", in *ASA-50 Speech Communication Papers*, ed. by Wolf J.J. and Klatt D.H. (The Acoustical Society of America), pp. 507 - 510.

Brandt Corstius H. (1965), "Automatic translation of numbers into Dutch", *Foundations of Language*, January 1965, pp. 59 - 62.

Carlson R., Galyas K., Granström B., Petterson M. and Zachrisson G. (1980), "Speech synthesis for the non-vocal in training and communication", *STL-QPSR*, 1/1980 (Royal Institute of Technology, Stockholm), pp. 13 - 27.

CBS (1974), "Gehandicapten wel geteld: lichamelijk gehandicapten 1971/1972." *Deel 1: Kerncijfers, Staatsuitgeverij, 's Gravenhage.*

Cohen A., Collier R. and 't Hart J. (1982), "Declination: construct or intrinsic feature of speech pitch ?", *Phonetica* 39, pp. 254-273.

Collier R. and Houtsma A.J.M. (1988), "Developments", *IPO Annual Progress Report* 23, pp. 13 - 14.

Collins D.W. (1974), "Patient initiated light operated telecontrol (PILOT)", in *Aids for the Severely Handicapped*, ed. by Copeland K. (Sector Publishing Ltd., London), pp. 31 - 41.

Damper R.I., Burnett J.W., Gray P.W., Straus L.P. and Symes R.A. (1987), "Hand-held text-to-speech device for the non-vocal disabled", *Journal of Biomedical Engineering*, October 1987, Volume 9, pp. 332 - 340.

Deliege R.J.H. (1987)*, "A speech output system using the Centronics interface." *Ontwerp Specificatie no. 41, Instituut voor Perceptie Onderzoek, Eindhoven.*

Deliege R.J.H. (1989), "An experimental Dutch keyboard-to-speech system for the speech impaired.", *Speech Communication* no. 8, 1989, pp. 81 - 89.

Deliege R.J.H., Speth-Lemmens I.M.A.F. and Waterham R.P. (1988a), "Ontwikkeling en evaluatie van twee communicatiehulpmiddelen met spraakuitvoer.", *Nederlands Tijdschrift voor Ergotherapie*, jaargang 16 no. 2, April 1988, pp 37 - 40.

Deliege R.J.H., Speth-Lemmens I.M.A.F. and Waterham R.P. (1988b), "Ontwikkeling en evaluatie van twee communicatiehulpmiddelen met spraakuitvoer.", *Tijdschrift voor Logopedie en Foniatrie*, jaargang 60 no. 7/8, Juli/Aug. 1988, pp 220 - 224.

Deliege R.J.H., Speth-Lemmens I.M.A.F. and Waterham R.P. (1989), "Ontwikkeling en evaluatie van twee communicatiehulpmiddelen met spraakuitvoer.", *Journal of Medical Engineering & Technology*, Volume 13 no. 1/2, Jan/Apr. 1989, pp 18 - 22.

Deliege R.J.H. and Waterham R.P. (1986a), "Communicatiehulpmiddelen voor spraakgehandicapten: twee sprekende voorbeelden van communicatiehulpmiddelen.", *Communication Aids for the Physically Handi-*

capped, ed. by Smeets, J.W. and Alexander, J.J., Conference proceedings "Handicap en Communicatie", Technological University, Delft, pp. 49 - 54.

Deliege R.J.H. and Waterham R.P. (1986b), "Application of speech synthesis and resynthesis in two speech communication aids.", *IPO Annual Progress Report*, 21, pp. 110 - 115.

Deliege R.J.H. and Waterham R.P. (1988), "Realization and evaluation of two speech communication aids.", *IPO Annual Progress Report*, 23, pp. 115 - 120.

Deliege R.J.H. and Waterham R.P. (1989), "Toepassing van synthetische spraak in communicatiehulpmiddelen voor spraakgehandicapten.", *to be published in Informatie.*

Dudley H. (1940), "The carrier nature of speech", *Bell System Tech. J.*, vol. 19, pp. 495 - 515.

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

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

Elsendoorn B.A.G. (1985), "Acceptability of temporal variations in synthetic speech: a preliminary investigation", *IPO Annual Progress Report* 20, pp. 33 - 42.

Falzoni E. (1985)*, "Development of hardware and software for the diphone speech board", *PII report.*

Fant G. (1960), *Acoustic Theory of Speech Production*, (Mouton, The Hague).

Fischer-Jørgensen E. (1956), "The commutation test and its application to phonemic analysis", in *For Roman Jakobson*, ed. by Halle M., Lunt H.G., Mclean H. and Schooneveld C.H. van, (Mouton, The Hague), pp. 140 - 151.

Foulds R.A. (1986), "Towards improved human communication: A review of communication aid research in North America", in *Handicap en communicatie*, ed. by Smeets J.W. and Alexander J.J., (Delftse Universitaire Pers).

Galyas K. and Liljencrants J. (1987), "Multi-talk, a new portable multilingual speech output communication aid", *European Conference on Speech Technology, Edinburgh, September 1987*, ed. by Laver J. and Jack M.A. (CEP Consultants Ltd, Edinburgh), pp. 357 - 360.

Geel R. van (1983), "Pitch inflection in electrolaryngeal speech", *Doctoral thesis, Eindhoven University of Technology.*

Geerts, G. and Heestermans, H. (eds.) (1984), "Van Dale Groot Woordenboek der Nederlandse Taal.", *Elfde herziene druk, Van Dale Lexicografie, Utrecht-Antwerpen.*

Gordon D. and Zabo D. (1984), "Technical solutions for people with communication impairments", *Proceedings of the International Congress on Technology & Technology Exchange, Pittsburgh USA, 8 - 10 October 1984*, pp. 42 - 43.

Greene B.G., Logan J.S. and Pisoni D.B. (1986), "Perception of synthetic speech produced automatically by rule: intelligibility of eight text-to-speech systems", *Behavior Research Methods, Instruments & Computers*, 18 (2), pp. 100 - 107.

't Hart J. and Cohen A. (1973), "Intonation by rule: a perceptual quest", *Journal of Phonetics* 1, pp. 309 - 327.

't Hart J. and Collier R. (1975), "Integrating different levels of intonation

analysis", *Journal of Phonetics* 3, pp. 235 - 255.

Hemert J.P. van (1985), "Automatic diphone preparation", *IPO Annual Progress Report* 20, pp. 23 - 32.

Hertz S.R. (1982), "From text to speech with SRS", *J. Acoust. Soc. Am.*, Vol. 72, pp. 1155 - 1170.

Holmes J.N. (1988), *Speech Synthesis and Resynthesis*, (Van Nostrand Reinhold).

Hunnicutt S. (1980), "Grapheme-to-phoneme rules: a review", *STL-QPSR*, 1/1980 (Royal Institute of Technology, Stockholm), pp. 13 - 27.

Intel (1985), "Micro-controller Handbook", *User's manual, Intel Cooperation, Santa Clara.*

Katwijk A. van (1965), "A grammar of Dutch number names", *Foundations of language*, January 1965, pp. 51 - 58.

Kerkhoff J., Wester J. and Boves L. (1984), "A compiler for implementing the linguistic phase of a text-to-speech conversion system", *Linguistics in the Netherlands*, pp. 111 - 117.

Klatt D.H. (1982), "The Klattalk text-to-speech conversion system", *Proceedings ICASSP 1982*, pp. 1589 - 1592.

Klip E.J. (1982), "Van vraagstelling tot verkrijgbaar produkt. 2 Ontwerpmethodiek", *Handicap en techniek. Verslag van het congres handicap en techniek, Rotterdam, 27 en 28 Augustus 1981*, ed. by Bangma B.D. and de Vlieger M. (Bohn, Scheltema & Holkema, Utrecht/Antwerpen), pp. 19 - 23.

Konings M. (1986)*, "Een prototype van het praatblok, uitgevoerd m.b.v. een microcomputer", *Eindwerk Katholieke Industriële Hogeschool*

De Nayer, St. Katelijne-Waver.

Krol R.C.P. van der (1988)*, "Spraakmodule voor de Possum; een communicatiehulpmiddel voor meervoudig gehandicapten.", *Stageverslag Technische Universiteit Eindhoven, vakgroep EME.*

Kroon J.N. (1986), "The Typophone: a talking typewriter", *Doctoral thesis, Eindhoven University of Technology.*

Leliveld W.H., Ossevoort H.J.M. and Severs J. (1979)*, "Bouwaanwijzing voor een spraakversterker met laadapparaat t.b.v. o.a. gelaryngectomeerden", *Intern rapport Technische Universiteit Eindhoven, vakgroep Meten en Regelen, leerstoel Medische Electrotechniek.*

Leliveld W.H., Bosch J.G., Mathijssen R.W.M. and Ossevoort H.J.M. (1988), "The Monoselector: an electronic aid for environmental remote control of electrical appliances for paralyzed persons.", *Medical Progress through Technology 13: 165-170.*

Maling R. (1974), "Control systems - concept and development (POSSUM)", in *Aids for the Severely Handicapped*, ed. by Copeland K. (Sector Publishing Ltd., London), pp. 22 - 30.

Mariani L. (1984), "Gli ausili tecnici per l'autonomia della persona disabile: risultati recenti, problemi e prospettive", *Elettrotechnica*, vol. LXXI, no. 11, November 1984, pp. 1007 - 1029.

Mumenthaler M. (1977), *Neurology*, (Georg Thieme Publishers, Stuttgart).

Neerven R.F.L. van (1987)*, "De Typofoon: herzien ontwerp van een sprekende schrijfmachine", *Afstudeerverslag technische Universiteit Eindhoven, vakgroep EME.*

Nooteboom S.G. and Cohen A. (1984), "Spreken en verstaan: Een nieuwe inleiding tot de experimentele fonetiek", *Van Gorcum, Assen.*

Nottroth M. and Van Jole F. (1988), "Spraakcomputer: revolutie voor blinden", *SURF 2*, no. 4, pp. 31 - 32.

O'Shaughnessy D. (1984), "Design of a real-time French Text-to-Speech system", *Speech Communication* 3, pp. 233 - 243.

O'Shaughnessy D. (1987), *Speech Communication, Human and Machine*, (Addison-Wesley Publ. Comp.)

Peterson C.D. (1982), "The Assisted Communicator: A Monograph on the use of the Phonic Mirror Handivoice.", *Phonic Ear Inc., California.*

Peterson G.E. and Lehiste I. (1960), "Duration of syllable nuclei in English", *J. Acoust. Soc. Am.*, Vol. 32, pp. 693 - 703.

Ring N. (1983), "General considerations of the way ahead", in *High Technology Aids for the Disabled*, ed. by Perkins W.J. (Butterworths, London), pp. 207 - 210.

Sadare A.D. (1984)*, "An investigation into the design requirements for a synthetic speech system which incorporates user defined speech parameters for use by the disabled", *Doctoral thesis, King's College London.*

Scheffers M. and Ten Have M. (1986)*, "Diphone software standardization document (version 1.1)", *Laboratory report DPE86133, Philips Central Application Laboratory, Eindhoven.*

Schildt A. and Sterner M. (1986), *Talsyntes som talhjälpmedel: en utvärdering*, (Handikappinstitutet, Bromma).

Schuurmann, P.L.H. and Mélotte, H.E.M. (1982), "An artificial larynx with semi-automatic pitch control",*IPO Annual Progress Report* 17, pp. 156 - 160.

Soede M. (1980), "Development and evaluation of complex aids: a case

study", in *The Use of Technology in the Care of the Elderly and the Disabled*, ed. by Bray J. and Wright S. (Frances Pinter Ltd., London), pp. 93 - 99.

Soede M. (1986), "Invoermethoden", in *Communication aids for the Physically Handicapped*, ed. by Smeets, J.W. and Alexander, J.J., Conference proceedings "Handicap en Communicatie", Technological University, Delft, pp 41 - 47.

Speth-Lemmens I.M.A.F. and Oostinjen E. (1986)*, "Evaluatie Tiepstem", *Evaluatielijsten, Instituut voor Rehabilitatie Vraagstukken, Hoensbroek.*

Speth-Lemmens I.M.A.F. (1987)*, "Evaluatie Tiepstem", *Intern rapport IRV/1 doc.(87), Instituut voor Revalidatie Vraagstukken, Hoensbroek.*

Steehouder M.F., Jansen C.J.M. en Staak J.L.C. van der (1984), "Leren communiceren: procedures voor mondelinge en schriftelijke communicatie", *Wolters-Noordhof, Groningen.*

Steeksma C.K.J. (1988)*, "De sprekende bloedsuikermeter: een hulpmiddel voor visueel gehandicapte diabetici", *Afstudeerverslag Technische Universiteit Eindhoven, vakgroep EME.*

Storm A. (1987)*, "Koppeling van twee hulpmiddelen voor spraakgehandicapten", *Rapport no. 625, Instituut voor Perceptie Onderzoek, Eindhoven.*

Taalzakboek (1983), *Boek- en Leermiddelen Centrale Arnhem B.V.*

Talbott M. (1984), "A cookbook of application ideas", in *Electronic Speech Synthesis*, ed. by Bristow G. (Granada, London), pp. 303 - 319.

Tervoort B.T., Geest A.J.M. van der, and Hubers G.A.C. (1976), "Psycholinguistiek", *Het Spectrum, Utrecht/Antwerpen.*

Vanderheiden G. (1982), "Computers can play a dual role for disabled individuals", *Byte*, September 1982, pp. 136 - 162.

Verhoeven M.W.C. (1989)*, "Een systeemopzet voor de Tiepstem-2", *Rapport no. 686, Instituut voor Perceptie Onderzoek, Eindhoven.*

Verniers R. en Verpoorten W. (1986)*, "Logopedie in de zwakzinnigenzorg", *Syllabus Logopedische Opleiding, Eindhoven.*

Vincent M.-A. (1987)*, "Communicatie-hulpmiddelen, computers & logopedie en hun onderlinge verbanden", *Scriptie Logopedische Opleiding, Eindhoven.*

Vogten L.L.M. (1983), "Analyse, zuinige codering en resynthese van spraakgeluid", *Doctoral thesis, Eindhoven University of Technology.*

Waterham R.P. (1983)*, "Aanpassingen aan bloedsuikermeetapparatuur t.b.v. visueel gehandicapte diabetici", *Afstudeerverslag Technische Universiteit Eindhoven, vakgroep EME.*

Waterham R.P. (1989), "The "Pocketstem": an easy-to-use speech communication aid for the vocally handicapped", *Doctoral thesis, Eindhoven University of Technology.*

Weiss E. and Lillywhite H.S. (1981), "Communicative Disorders", *The C. V. Mosby Company, St. Louis.*

Witten I.H. (1982), *Principles of Computer Speech*, (Academic Press).

Summary

This thesis deals with the application of synthetic speech in a communication aid. Recent developments in the field of speech technology and electronics have led to the possibility to apply synthetic speech with a quality sufficient to be used in practical applications, in small, portable devices. One class of applications is in speech communication aids, i.e. as a replacement for natural speech for people who have lost the ability to speak. Such a handicap can be caused by a language disorder or by a malfunction of the speech organs. A speech handicap can be considered to be a serious handicap, because speech is the most essential communication channel in human life, essential for attracting attention, taking part in group activities and for use in communication means such as a telephone. Because of the seriousness of a speech impairment and because of the number of people suffering from it, it seems worth attempting to apply synthetic speech in the field of communication aids for the speech impaired. In this project we therefore investigate whether a speech communication aid (primarily Dutch) can be constructed in such a way that it can be used with success by speech-impaired persons for diminishing their handicap.

Basic requirements for such an aid are intelligible and natural-sounding speech, ease of operation, a large enough vocabulary, portability, and availability, preferably at a low price. In this case ease of operation and a large enough vocabulary are more or less contradictory. We focus on a large vocabulary by using the technique of diphone concatenation to generate synthetic speech, which offers an unlimited vocabulary. In a parallel project the focus is on ease of operation, which resulted in an aid (Pocketstem) that uses a small keyboard (28 keys), each provided with a symbol and representing one message. Because both projects were carried out in a similar way, knowledge could be shared and results could be compared.

We designed, made and evaluated an experimental model of a speech communication aid. This resulted in a small, portable, and battery-powered device. This aid, called the Tiepstem (Typevoice), uses a QWERTY keyboard for input; a liquid crystal display showing what is typed. The messages can be spoken by pressing a "speak key". Several correcting and editing functions are available, as is a storage facility that allows for storage and retrieval of complete messages or parts of them. Normal Dutch orthographic input spelling was not possible at the time, because no Dutch grapheme-to-phoneme conversion was available. Therefore a pseudo-phonetic input spelling is used, as a temporary compromise. A synthetic pitch contour is generated by use of basic intonation patterns, available from earlier research. Selection of these patterns is triggered by punctuation marks and by a special input symbol that the user has to provide in the input. For the actual speech generation, the diphone concatenation technique and a hardware formant synthesizer chip are used. This technique, which is based on the concatenation of small speech fragments, allows for an unlimited vocabulary and offers a quality that is believed to be high enough for this kind of application. The evaluation showed possibilities for such a device, but indicated also some problems, the main problem being the pseudo-phonetic input.

In order to overcome this input problem and to create new possibilities for a new group of potential users a combination of the Pocketstem and the Tiepstem has been realized. In this setup the Tiepstem is used to create messages that can be stored in the Pocketstem for actual use. The field evaluation carried out with this combination was too limited to allow general conclusions to be drawn yet.

Because of the recent developments in linguistics and electronics some of the problems of our first model can be solved by now. Especially the availability of a Dutch grapheme-to-phoneme conversion and powerful CMOS microprocessors allow for normal Dutch orthography as input spelling. Therefore a successor of our device will be designed, made and evaluated again.

Samenvatting

Dit proefschrift beschrijft een onderzoek naar de toepassing van synthetische spraak in een communicatiehulpmiddel. Recente ontwikkelingen op het gebied van de spraaktechnologie en de micro-elektronica hebben geleid tot de mogelijkheid om synthetische spraak, met voldoende kwaliteit voor praktische toepassingen, toe te passen in kleine, draagbare apparaten. Eén van de toepassingen is in een spraakhulpmiddel, als vervanging van natuurlijke spraak voor mensen die hun spraakvermogen kwijt zijn. Deze handicap kan veroorzaakt worden door een taalstoornis of door niet functioneren van het spraakorgaan. Een spraakstoornis is een ernstige handicap omdat spraak het meest gebruikte communicatiekanaal tussen mensen is, essentieel is om aandacht te trekken, deel te nemen aan groepsactiviteiten en voor het gebruik van communicatiemiddelen zoals de telefoon. Vanwege de ernst van deze handicap en het aantal betroffenen lijkt het zinvol om synthetische spraak proberen toe te passen op het gebied van communicatiehulpmiddelen voor spraakgehandicapten. Daarom onderzoeken we in dit project of we een (Nederlandstalig) spraakhulpmiddel kunnen maken dat met succes gebruikt kan worden door spraakgehandicapten om hun handicap te verlichten.

De elementaire eisen voor zo'n hulpmiddel zijn verstaanbare en natuurlijk klinkende spraak, makkelijke bediening, een voldoende groot vocabulair, draagbaarheid en lage prijs. Makkelijke bediening en een voldoend groot vocabulair zijn echter tegenstrijdige eisen. In dit project richten we ons op een voldoende groot vocabulair door middel van de techniek van difoon concatenatie, die een onbeperkte vocabulair mogelijk maakt. In een parallel project is de aandacht gericht op de makkelijke bediening, wat resulteerde in een hulpmiddel genaamd Pocketstem. Dit hulpmiddel bevat een toetsenbord (28 toetsen), elk voorzien van een symbool en gekoppeld aan één boodschap. Aangezien beide projecten op een soortgelijke manier uitgevoerd werden kon kennis gedeeld en resultaten

vergeleken worden.

In dit project is een experimenteel model van een spraak communicatiehulpmiddel ontworpen, gebouwd en geevalueerd. Dit is een klein, draagbaar batterij-gevoed apparaat, genaamd de Tiepstem. Het apparaat gebruikt een QWERTY toetsenbord voor invoer; een Liquid Crystal Display toont wat ingevoerd wordt. Het spreken gebeurt met een toets "spreek". Er zijn verschillende correctiemogelijkheden beschikbaar, evenals een geheugenfaciliteit voor opslag en ophalen van boodschappen of delen daarvan. Invoer in normale Nederlandse spelling was op dat moment nog niet mogelijk, omdat nog geen Nederlandse grafeem-foneem omzetter beschikbaar was. Daarom is, als een tijdelijk compromis, gekozen voor een pseudo-fonetische invoer. Er wordt een kunstmatige intonatie-contour gegenereerd op basis van intonatiepatronen, die uit eerder onderzoek beschikbaar waren. De keuze van deze patronen vind plaats op basis van leestekens en een speciaal accent symbool dat de gebruiker moet invoeren. De spraak wordt gegenereerd door middel van difoon concatenatie en een hardware formant synthesizer. Deze techniek, gebaseerd op de concatenatie van kleine spraakfragmenten, maakt een onbeperkte vocabulair mogelijk met een spraakkwaliteit, hoog genoeg voor deze toepassing. De evaluatie toonde de mogelijkheden voor zulk een apparaat, maar signaleerde ook verscheidene problemen, voornamelijk met de pseudo-fonetische invoer.

Als mogelijke oplossing voor dit invoerprobleem en om de groep potentiële gebruikers uit te breiden is ook een combinatie gemaakt van de Tiepstem met de Pocketstem. In deze combinatie wordt de Tiepstem gebruikt om de boodschappen te maken, die dan in de Pocketstem opgeslagen worden. De uitgevoerde evaluatie was echter te beperkt om conclusie te trekken.

Ten gevolge van nieuwe ontwikkelingen in de taalwetenschappen en de electronica kunnen sommige van de gesignaleerde problemen in ons apparaat nu opgelost worden. Vooral de ontwikkeling van een Nederlandse grafeem-foneem omzetter in combinatie met krachtige CMOS microprocessoren maken het gebruik van normale Nederlandse spelling als invoer mogelijk. Daarom zal een opvolger van ons apparaat ontworpen, gebouwd en geëvalueerd worden.

Appendix A

MEA8000 speech synthesizer

A.1 Introduction

In this appendix we describe the MEA8000 synthesizer and its features. The MEA8000 speech synthesizer incorporates an electronic implementation of the source filter model of human speech production. A periodic signal, representing the pitch of voiced speech, or an aperiodic signal, representing the unvoiced speech, is fed to a variable filter comprising four resonators, via an amplifier that controls the amplitude of the synthesized sound (figure A.1). The resonators model the sound in accordance with the formants in the original speech. Each resonator is controlled by two parameters, one for the resonant frequency and one for the bandwidth. The information required to control the synthesizer is:

pitch
amplitude
voiced/unvoiced source selector
filter settings.

A good replica of the original speech is obtained by periodic updating of this control information.

The features of the MEA8000 synthesizer are [Van Brück and Teuling, 1982]:

- 4 kHz bandwidth.
- bit rate 500 - 4000 bits/sec.
- four formants (the frequency of the 4th is fixed).

- NMOS technology.
- 8-bit digital-to-analog converter.

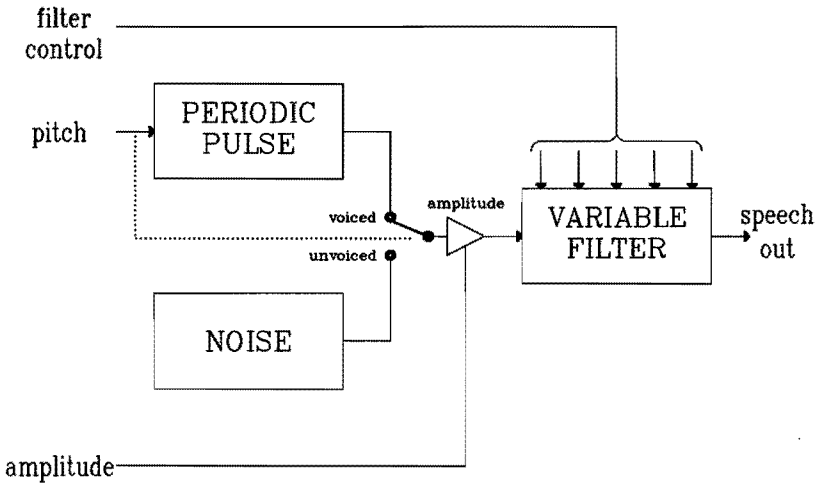

Figure A.1: Model of the MEA8000 speech synthesizer.

A.2 Speech code format

Since the human vocal tract is a mechanical system, its characteristics change quite slowly during the formation of voice sounds. It has been found that the speech synthesizer control parameters can be adequately represented if they are updated once every few tens of milliseconds with linear interpolation during the intervals to ensure a smooth changeover from one set of parameters to the next. In the MEA8000, the updating period (called a speech frame) can be set to 8, 16, 32 or 64 ms.

During voice output, the speech codes from a microcomputer or external ROM are transmitted on an 8-bit databus to the MEA8000 in blocks of four bytes, each block characterizing a speech frame; see figure A.2 and table A.1. Byte four contains a 5-bit pitch increment code,

which can be positive or negative. However, when the synthesizer starts to talk, a preliminary byte containing the full starting pitch code must be transmitted. This byte goes directly to the pitch-generating circuitry. This method of encoding pitch contributes to a lower bit rate. After the starting pitch code, the codes of each speech frame are shifted into a four-byte input buffer. The parameter interpolation logic calculates the difference between consecutive parameters and interpolates linearly between them to smooth the parameter transients. The interpolation interval is decoded using the two frame duration (FD) bits in each speech frame.

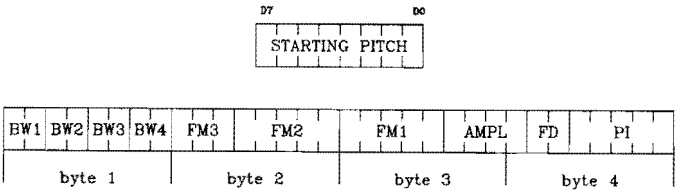

Figure A.2: Frame format.

Table A.1: Frame bit allocation.

code	bits	parameter
PI	5	pitch increment or noise selection
FD	2	speech frame duration
AMPL	4	amplitude
FM1	5	frequency of 1st formant
FM2	5	frequency of 2nd formant
FM3	3	frequency of 3rd formant
BW1	2	bandwidth of 1st formant
BW2	2	bandwidth of 2nd formant
BW3	2	bandwidth of 3rd formant
BW4	2	bandwidth of 4th formant
FM4, the frequency of the fourth formant, is fixed.		

A.3 Hardware

The MEA8000 is realized in NMOS technology and consumes 30 mA from one 5 V supply. It has a crystal-controlled internal oscillator or it can be driven by an external clock (4 MHz). The sample rate of the built-in digital-to-analog converter is 64 kHz, which is far above the audible frequency range, allowing the use of a simple external audio filter. The request signal for a new byte of speech code, or the starting pitch byte if the synthesizer is stopped, is available as a hardware signal or as a bit in the status register of the synthesizer.

A.4 Interface protocol

The interface protocol, using the request signal REQ, is illustrated in figure A.3. It is not usually necessary to check the status of REQ after each byte since a new request occurs within 3 μs of receiving each byte.

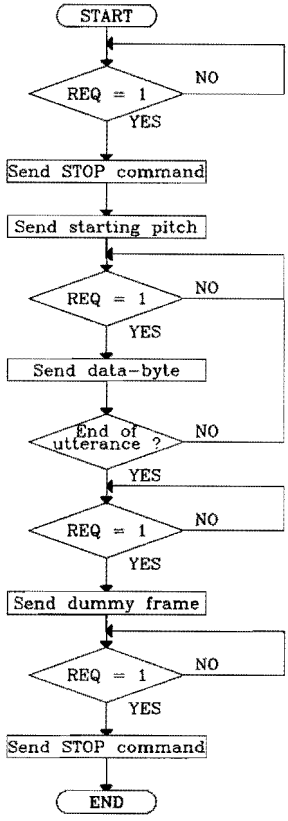

Figure A.3: Interface protocol.

A.5 References

Philips (1983), *MEA8000 voice synthesizer: principles and interfacing*, Philips Technical Publication 101.

Van Brück H.E. and Teuling D.J.A. (1982), "Integrated voice synthesizer", *Electronic Components and Applications*, vol. 4, no. 2, pp. 72-79.

Appendix B

Available synthetic speech devices

B.1 Available systems

This appendix lists existing speech communication aids and systems that can be used as such. An explanation will be found in section 3.5.

Name	category	input[1]	language[2]	synthesis technique
Form-a-phrase	1a	keys (s & n)	e.	?
Scan & Spell	1a	1 or 2 keys	e.	lpc
Speak & Spell	1a	keyboard	e.	lpc
Vocaid	1a	36 keys	e.	lpc
Falck 3310	1a	keys n	all	?
Bliss-stem	1a	378 keys	d.	lpc
Handivoice	1b	keys n	e.	phonemes
Minspeak I	1b	60 keys s	e	phonemes
Vòis 130	1b	keys s	e.	phonemes
Vòis 140	1b	keys n	e.	phonemes
Vòis 150	1b	keys n /js /ss	e.	phonemes
Alltalk	1b	128 keys	all	AD/DA
Namcon Talkin'Aid	1b	keys ?	j.	?
Touchtalker	1b	128 keys	e.	phonemes
Lighttalker	1b	128 keys lo	e.	phonemes
VPSP[6]	2a	keyboard	e.	phonemes
SAL[7]	2a	Blissboard	e.	phonemes
KTH comm. system	2a	Blissboard 500 keys / keyboard	m.	synth.-by-rules
Sahara II	2a	Blissboard 486 keys	f.	diphone
Sevoca [8]	2a	keyboard	sp./e.	phonemes
Say-It-All	2a	keyboard	e.	?
Say-It-All +	2a	keyboard	e.	?
Words + II	2a	keyboard/ ss	e.	?
Special friend	2a	keyboard	e.	?
The Speechaid	2a	keyboard	e.	phonemes
SpeechPac	2a	keyboard/ ss	e.	phonemes
Smalltalk	2a	keyboard	e.	?
Talking 100	2a	keyboard/ ss	e.	?
Sadare's speech system	2a	keyboard	e.	phonemes
Psytalk	2a	keyboard	e.	phonemes
Multi-talk	2a	keyboard	m.	synth.-by-rules
Kurzweil Talking Terminal	2b	ASCII	e.	?
Microvox	2b	ASCII	e.	phonemes
Dectalk	2b	ASCII	e.	synth.-by-rules
Infovox SA101	2b	ASCII	m.	synth.-by-rules
Prose 2000/3000	2b	ASCII	e.	synth.-by-rules
FTS [9]	2b	ASCII	f.	synth.-by-rules
ICO 85	2b	ASCII	f.	diphone
Type'n'Talk	2b	ASCII	e.	phonemes
Echo speech system	2b	ASCII	e.	phonemes
SAM [10]	2c	pc/hc	e.	phonemes
Radio Shack voice synth.	2c	pc/hc	e.	phonemes

Name	vocabulary (size)[3]	status[4]	literature[5]
Form-a-phrase	128 w.	a:$ 650	c.o.
Scan & Spell	alphabet	a:$ 475	c.o.
Speak & Spell	alphabet	a:$ 50	c.o.
Vocaid	36 utterances (4x)	a:$ 148	c.o.
Falck 3310	max. 162 utt.	a: ?	catalogue
Bliss-stem	1400 utt.	project	catalogue
Handivoice	500/900 w.	a:$ 1995	c.o.
Minspeak I	60 utt.	a: ?	c.o.
Vòis 130	352 w./19 ph./alphabet	a: ?	c.o.
Vòis 140	891 w./19 ph./alphabet	a: ?	c.o.
Vòis 150	891 w./19 ph./alphabet	a: ?	c.o.
Alltalk	128 utt.	a:$ 4995	c.o.
Namcon Talkin'Aid	?	a: ?	catalogue
Touchtalker	128 utt.	a: ?	catalogue
Lighttalker	128 utt.	a: ?	catalogue
VPSP[6]	∞	project	c.o.
SAL[7]	∞	project	c.o.
KTH comm system	∞	project	[Carlson *et al.*, 1981]
Sahara II	∞	project	[Emerard *et al.*, 1980]
Sevoca [8]	∞	project	c.o.
Say-It-All	∞	a:$ 1200	c.o.
Say-It-All +	∞ + 128 utt.	a:$ 1600	c.o.
Words + II	∞	a:$ 1775-2075	c.o.
Special friend	∞	a:$ 1295	c.o.
The Speechaid	∞	a: ?	c.o.
SpeechPac	∞	a:$2195	c.o.
Smalltalk	∞	a:$ 1995	c.o.
Talking 100	∞	a: ?	c.o.
Sadare's speech system	∞	project	[Sadare, 1984]
Psytalk	∞	project	[Damper *et al.*, 1987]
Multi-talk	∞	a: ?	[Galyas & Liljecrants, 1987]
Kurzweil Talking Terminal	∞	a:$ 3650	c.o.
Microvox	∞	a: ?	c.o.
Dectalk	∞	a:$ 2195	[Bruckert, 1984]
Infovox SA101	∞	a: ?	[Magnusson *et al.*, 1984]
Prose 2000/3000	∞	a:$ 4800	[Groner *et al.*, 1982]
FTS [9]	∞	project	[O'Shaugnessy, 1984]
ICO 85	∞	a:$ 1000	[Gauvain and Gangolf, 1983]
Type'n'Talk	∞	a: ?	[Greene *et al.*, 1986]
Echo speech system	∞	a: ?	[Greene *et al.*, 1986]
em SAM [10]	∞	a:$ 60-125	c.o.
Radio Shack voice synth.	∞	a:$ 400	c.o.

1. s = symbol
   n = numeric
   ss = single switch
   js = joystick
   lo = light operated
   pc = personal computer
   hc = home computer

2. e = English
   d = Dutch
   j = Japanese
   m = multi-lingual (English, French, Spanish, German, Swedish, Italian, Norwegian)
   f = French
   sp = Spanish

3. w = words
   ph = phonemes
   utt = utterances

4. a = available

5. c.o. = *"Communication Outlook"*, Artificial Language Lab, Michigan State University, Lansing.

6. Versatile Portable Speech Prosthesis (children's hospital Stanford)

7. Semantically Accessible Language board (artificial language laboratory, Michigan State University)

8. Spanish-English Voice Output Communication Aid (children's hospital Stanford)

9. French Text-to-speech System

10. Software Automatic Mouth

B.2 References

Bruckert E. (1984), "A new text-to-speech product produces dynamic human-quality voice", *Speech Technology*, Jan./Feb. 1984, pp. 114-119.

Carlson R., Galyas K., Granström, Hunnicutt S., Larsson B. and Neovius L. (1981), "A multi-language, portable text-to-speech system for the disabled", *STL-QPSR*, 2-3/1981 (Royal Institute of Technology, Stockholm), pp. 8 - 16.

Damper R.I., Burnett J.W., Gray P.W., Straus L.P. and Symes R.A. (1987), "Hand-held text-to-speech device for the non-vocal disabled", *Journal of Biomedical Engineering*, October 1987, Volume 9, pp. 332 - 340.

Emerard F., Graillot P., Cyne G. and Lucas J.J. (1980), "Protheses de parole destinées a la communication des handicapés moteurs deficients de la parole", *Recherches / Acoustique CNET, Lannion III*, pp. 133 - 144.

Galyas K. and Liljencrants J. (1987), "Multi-talk, a new portable multilingual speech output communication aid", *European Conference on Speech Technology, Edinburgh, September 1987*, ed. by Laver J. and Jack M.A. (CEP Consultants Ltd, Edinburgh), pp. 357 - 360.

Gauvain J.L. and Gangolf J.J. (1983), "Terminal integrates speech recognition and text-to-speech synthesis", *Speech Technology*, Sept./Oct. 1983, pp. 25 - 38.

Greene B.G., Logan J.S. and Pisoni D.B. (1986), "Perception of synthetic speech produced automatically by rule: intelligibility of eight text-to-speech systems", *Behavior Research Methods, Instruments & Computers*, 18 (2), pp. 100 - 107.

Groner G.F., Bernstein J., Ingber E., Pearlman J. and Toal T. (1982), "A real-time text-to-speech converter", *Speech Technology*, April 1982, pp. 73 - 76.

Magnusson L., Blomberg M., Carlson R., Elenius K. and Granström B. (1984), "Swedish speech researchers team up with electronic venture capitalists", *Speech Technology*, Jan./Feb. 1984, pp. 15 - 24.

Sadare A.D. (1984), "An investigation into the design requirements for a synthetic speech system which incorporates user defined speech parameters for use by the disabled", *Doctoral thesis, King's College London.*

O'Shaughnessy D. (1984), "Design of a real-time French Text-to-Speech system", *Speech Communication* 3, pp. 233 - 243.

Appendix C

Grapheme-to-phoneme conversion

C.1 Introduction

For some time research has been going on at various institutes to come up with (among other things) a grapheme-to-phoneme conversion system for Dutch. At least one of these systems has reached the point where its performance is good enough for practical use and which can easily be implemented in a microprocessor system. This system [Kerkhoff, Wester and Boves, 1984] has been developed at the Institute of Phonetics of the Nijmegen University. In the context of the cooperation of all Dutch speech and phonetic research groups under the national SPIN program "Analysis and Synthesis of Speech (ASSP)" we could make use of this system for research purposes. Before we can use this system in our application together with the diphone concatenation some changes had to be made.

This section describes this grapheme-to-phoneme conversion system as it became available to us, the changes made to it and some implementation points.

C.2 The original system

The system is rule-based, i.e. the grapheme-to-phoneme correspondences are formulated as a set of linguistic rules. This set of rules can be compiled into a Pascal program by means of a rule-compiler program that was developed for this purpose (called FONPARS). This process is illustrated in figure C.1.

The format in which the rules are written is analogous to the notation

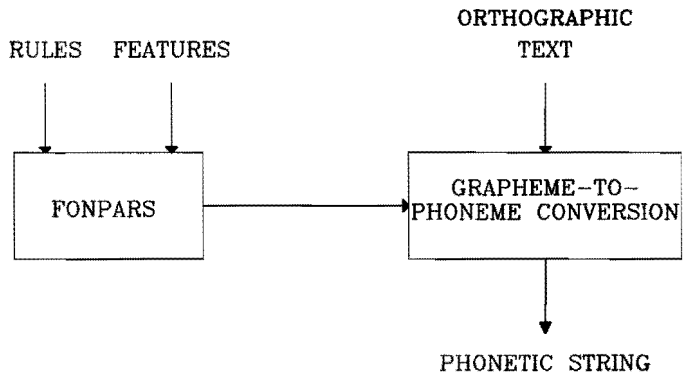

Figure C.1: Use of FONPARS.

described in the Sound Pattern of English (SPE) [Chomsky and Halle, 1968]. The general format of a rule is:

```
F -> C / L --- R
```

where: F = focus, C = change, L = left context, R = right context.

The capital letters in the rule format can represent so-called "phonological features". Phonological features are characteristics of sound segments, with which we can address groups of sound segments as a set. We can, for instance, address the vowels and consonants as sets via obvious features like [+voc] and [+cons] respectively. FONPARS needs to have access to a feature table in which the characteristics of sound segments are presented.

The complete rule format is described in Kerkhoff and Wester (1987). It allows among other things for: insertions, deletions, exchanges, feature specifications, optional elements, or-or statements and negations. Figure C.2 gives an example of some rules. The first rule specifies the pronunciation of the graphemes **au** when preceded by the graphemes **rest** or **ch** and followed by **f** or **r**. The second rule specifies a phonetic variant of **n** when followed by zero or more symbols with feature **[-seg]** followed by a **r**, **g** or **x**.

```
au -> oo / {rest/ch} --- {f/r}
n -> ~ / --- [-seg]0 {r/g/x}
```

Figure C.2: Example rules.

All rules are successively applied to the input string. For each rule the focus F moves from left to right through the string. Whenever the focus matches, the left and right contexts are evaluated. When these also match, the rule applies and the focus is replaced by the change C.
For grapheme-to-phoneme conversion a complete set of rules has been developed [Kerkhoff, Wester and Boves, 1984]. In this rule set we can distinguish various groups of rules. These groups are shown in Table C.1.

Table C.1.: Grapheme-to-phoneme rules.

Function	Approx. number
roman numbers	10
arabic numbers	25
abbreviations	5
uppercase words	5
sentence accent	25
grapheme-to-phoneme	100
word accent	50
assimilation	20
syllable boundaries	10

For text-to-speech application this system is followed by a speech synthesis part. This synthesis part generates allophones by rules. These rules are written in a format close to that of the grapheme-to-phoneme rules.

C.3 The modified system

In order to use this grapheme-to-phoneme conversion together with the diphone-based speech synthesis, some modifications had to be made, for

the following reasons:

1. The phonetic output of this system differs from our phonetic notation, both in coding and in the allophonic variations used.
2. The phonetic output is meant as input for an allophone synthesizer, rather than a diphone-based one. This sometimes results in other phoneme sequences.
3. The intonation part of the system already selects basic intonation patterns while we have our own selection system that expects the accent positions and punctuation marks as input.
4. The system, which is rule-based, does not support an exception lexicon. Such a lexicon is necessary, however, in our application (e.g., for names).

Ad 1). The phonetic output of this conversion (Table C.2) had to be translated to the notation as used with the diphones (Table C.3).

Table C.2: Dutch phonemes (and some of their allophones) and their character representation for use with the allophone synthesis.

character representation	**example word**	**character representation**	**example word**
A	b*a*d	p	*p*ut
E	b*e*d	b	*b*ad
I	b*i*d	t	*t*ak
O	b*o*d	d	*d*ak
U	p*u*t	k	*k*at
a	b*aa*t	G	*g*oal
e	b*ee*t	f	*f*iets
o	b*oo*t	v	*v*at
i	b*ie*t	s	*s*ap
y	b*oe*k	z	*z*at
u	b*uu*rt	C	po*tj*e
@	b*eu*k	S	wa*sj*e
E:	f*air*	Z	*j*aquet
EI	b*ij*t	x	la*ch*en
UI	b*ui*t	g	la*g*en
AU	b*ou*t	m	*m*at
&	d*e*	n	*n*at
!	a*n*jer	l	*l*at
N	la*ng*	r	*r*at
~	i*ng*aan	j	*j*at
*	aa*n*was	w	*w*at
h	*h*ad		

Table C.3: Dutch phonemes (and some of their allophones) and their character representation for use with the diphone synthesis.

character representation	example word	character representation	example word
SI	< *stilte* >	EW	l*eeuw*
GS	< *glottalstop* >	IW	k*ieuw*
II	l*ie*p	YW	d*uw*
I	p*i*t	P	*p*as
EE	l*ee*s	T	*t*as
E	l*e*s	K	*k*as
EH	mayon*ai*se	B	*b*as
AA	m*aa*t	D	*d*as
A	m*a*t	G	*g*oal
OO	r*oo*d	S	*s*ok
O	r*o*t	F	*f*ok
OH	z*o*ne	X	*g*ok
U	r*oe*t	Z	*z*eer
Y	f*uu*t	V	*v*eer
CC	p*u*t	M	*m*eer
C	d*e*	N	*n*eer
UH	fr*eu*le	NN	man*d*je
AU	k*ou*d	Q	ba*ng*
UI	m*ui*s	L	*l*ang
OE	k*eu*s	LL	da*l*
EI	r*ei*s	R	*r*ang
AI	det*ai*l	W	*w*ang
AJ	m*aai*t	J	*j*an
OI	h*oi*	H	*h*ang
OJ	h*ooi*t	PJ	boom*pj*e
UJ	r*oei*t	TJ	*tj*olk
ER	b*eer*	SJ	*sj*aak
OR	w*oord*	DJ	*dj*atiehout
CR	k*eur*	ZJ	*j*ournaal
AW	k*auw*	DZ	mana*g*er

Because there is no one-to-one correspondence between these two notations a simple table conversion is not possible. In addition there are some ambiguities, for instance E, I and EI are all valid phonemes. Therefore some kind of interface software is necessary. The rule format already used for the grapheme-to-phoneme conversion is perfectly suited to this task, so this translation is written as an additional set of rules. These rules are listed in figure C.3.

```
(* Conversion of KUN phonetic symbols to IPO notation   *)
(*=====================================================*)
(* Remove double spaces *)
(* Remove spaces before a comma *)
# -> $ / --- {# / ,}
(* Remove space between accent marker and word *)
# -> $ / + ---
(* Use \ as word separator instead of space *)
# -> \
(* Add a space between all phonemes *)
$ -> # / ^#^ ---
E # I -> EI
U # I -> UI
A # U -> AU
(* Conversion KUN to IPO notatie *)
C -> TJ
U -> CC / # --- #
a -> AA
e -> EE
o -> OO
i -> II
y -> U
u -> Y
@ -> OE
E # : -> EH
& -> C
N -> Q
*! -> NN
~ -> N
```

```
** -> N
h -> H
p # " # j -> PJ
p -> P
b -> B
t # " # j -> TJ
t -> T
d -> D
k -> K
f -> F
v -> V
S -> SJ
s # " # j -> SJ
s -> S
Z -> ZJ
z -> Z
x -> X
g -> X
m -> M
n -> N
l -> L
r -> R
j -> J
w -> W
" # -> $
(* Remedy against aankomst -> AAQKOMST etc. *)
Q -> N / [+voc] [+voc] # ---
```

Figure C.3: Conversion rules from KUN to IPO phoneme notation.

Ad 2). Some rules have to be added after this translation because a number of phonetic variations used by the diphone synthesis are not covered by the phonetic output of the original system. These variations are the silence phoneme, glottal stop, thick L and some diphthongs. These rules are listed in figure C.4.

```
(* Allophonic variants *)
AA # {I/J} -> AJ / --- [-ipo_seg]0 {[+ipo_cons] / \}
OO # {I/J} -> OJ / --- [-ipo_seg]0 {[+ipo_cons] / \}
O # {I/J} -> OI / --- [-ipo_seg]0 {[+ipo_cons] / \}
U # {I/J} -> UJ / --- [-ipo_seg]0 {[+ipo_cons] / \}
A # I -> AI / --- [-ipo_seg]0 {[+ipo_cons] / \}
AU # W -> AW / --- # [-ipo_seg]0 {[+ipo_cons] / \}
EE # W -> EW / --- # [-ipo_seg]0 {[+ipo_cons] / \}
II # W -> IW / --- # [-ipo_seg]0 {[+ipo_cons] / \}
Y # W -> YW
D # J -> DJ
J # -> $ / NN # [-ipo_seg]0 ---
# N # -> # NN # / --- [-ipo_seg]0 {SJ/TJ}
S -> Z / EH # ---
(* Thick L rules *)
# L # -> # LL # / --- [-ipo_seg]0 \ [-ipo_seg]0 ^L^
# L # -> # LL # / --- {S/T/D}
# L # -> # LL # / [+ipo_seg] --- [-ipo_seg]0 [+ipo_cons]
                                    [-ipo_seg]0 [+ipo_voc]
(* Add silence after a comma *)
$ -> # *# # *# / , ---
(* Add glottal stop between words *)
# -> # GS # / [+ipo_seg] # \ -L- [-ipo_seg]0 [+ipo_voc]
(* Add glottal stop in words (geopend, geacht) *)
# -> # GS # / C -L- [-ipo_seg]0 [+ipo_voc]
(* Allophonic variants *)
OO -> OR / --- # [-ipo_seg]0 R
EE -> ER / --- # [-ipo_seg]0 R
OE -> CR / --- # [-ipo_seg]0 R
(* Translate silence symbol *)
*# -> SI
```

Figure C.4: Rules for allophonic variants.

In addition, some rules of the existing set had to be modified. These rules take care of effects that are already incorporated in the diphones (e.g., glide insertion).

Ad 3). The original system comes up with two types of accent information: the sentence accent (a pattern number before a word to be accented) and word accent (the accent position within a word). The software is changed in such a way that the sentence accents are removed and the word accents are kept in those words that had this sentence accent.

Ad 4). The existing system uses no exception lexicon. For practical applications this is a serious omission. However good the grapheme-to-phoneme rules will be, there will always be irregularities not covered by these rules in a practical application (e.g., jargon, proper names). These irregularities can be covered by putting them together with their correct phonetic translation in a lexicon. In order to obtain maximum benefit from this lexicon, the user (or his helpers) should preferably be able to change it. In this case he can tune the lexicon to his own application and vocabulary. Attention has to be paid to the translation of the input string, because the rules are designed to operate on the complete sentence, while the lexicon operates on words.

To incorporate these changes, the rule set (plus the additional self written rules) has been divided into four groups as given in table C.4.

Table C.4.: Grouped grapheme-to-phoneme rules.

Name	Rules
Front	roman numbers
	arabic numbers
	abbreviations
	uppercase words
Accent	sentence accent
Grafon	grapheme-to-phoneme
	word accent
	assimilation
End	allophonic variations

A new software framework has been developed using these four groups of rules as illustrated in figure C.5. This translation strategy differs in two points from the straightforward translation by the rules. These two points are the possibility to skip the sentence (or sentence and word)

accent rules and the presence of an exception lexicon.

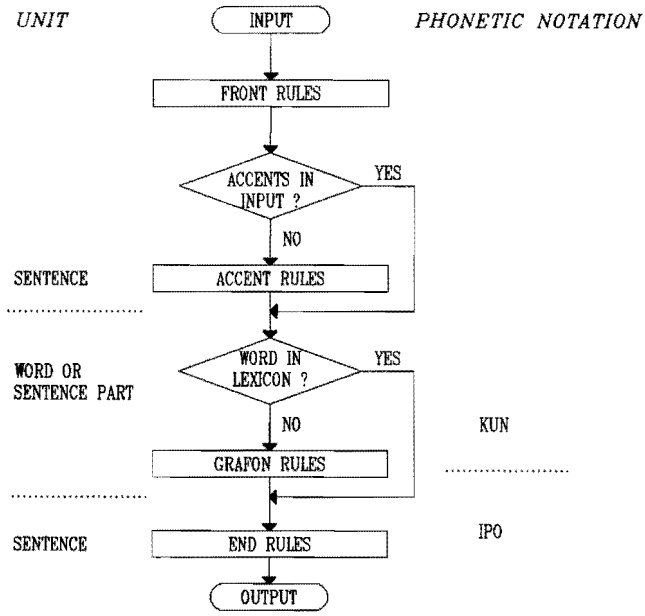

Figure C.5: Grapheme-to-phoneme conversion.

The possibility to skip the accent rules is implemented because the accent rules often come up with accent positions that are far from ideal. This is due to the fact that proper intonation is dependent on the meaning of the sentence, which a machine cannot know without extensive semantic analysis, and on the complexity of the other factors that influence intonation, such as syntax. It is worth while, therefore, creating a possibility to let the user provide the accent positions. On the other hand, when the user does not want or is not able to provide this information, the accent positions generated by the rules are presumably better than no accents at all. The choice whether to use the accent rules has therefore to be made by the user and is implemented as follows. If the user supplies accents in the input, these accents are used, otherwise the

rule-generated accents are used. Because the word accent rules (accent position within a word) function considerably better than the sentence accent rules, these word accent rules can be used most of the time. When the user places an accent symbol before a word, he indicates that this word has to be accented and the word accent rules determine the exact accent position. For cases where this goes wrong, he can place the accent symbol within the word, indicating the syllable to be stressed. This feature, however, was not foreseen during the development of the rules and therefore this placement of an additional symbol within a word, may cause the grapheme-to-phoneme rules to work incorrectly. Because this feature is only meant to be used in "emergency cases" no better solution is looked for.

As mentioned before, the exception lexicon operates on words, while the rules work on complete sentences. The rules are designed to work on complete sentences because 1) for calculating sentence accent positions the whole sentence is needed and 2) neighbouring graphemes influence each other in the grapheme-to-phoneme conversion. Therefore the conversion by the lexicon and by the group of grafon rules is carried out as follows. Each word in the input string is searched for in the lexicon, going from left to right through the string. This process continues until a word matches or the end of the string is reached. At that point the part of the string preceding the matching word (or the whole string in case the end was reached) is translated by the rules and the matching word is translated by the lexicon. The same process then starts again with the remaining part of the sentence. In this way the problems with the lexicon operating on words and the rules operating on sentences are minimised. The calculation of the accent positions (1) is still done on the whole sentence. The grapheme-to-phoneme conversion rules operate now on sentence parts (2). The only places where these rules miss the neighbouring graphemes is where the sentence is interrupted by words present in the lexicon. In this way this missing information is kept to a minimum. The rules in the group of end rules often work across word boundaries, so this group of rules is processed separately from the grafon rules. In this way these rules always work on a complete sentence.

C.4 Implementation

The described software is implemented on a 68000 microprocessor. Almost all of the software is written in Pascal. Two often-used routines (VGL and CORLENGT) are rewritten in assembly language to make the grapheme-to-phoneme conversion faster. The routines that access the exception lexicon are also written in assembly language in order to allow fast search for it.

C.5 References

Chomsky N. and Halle M. (1968), *The sound pattern of English*, (Harper & Row, New York).

Kerkhoff J., Wester J. and Boves L. (1984), "A compiler for implementing the linguistic phase of a text-to-speech conversion system", *Linguistics in the Netherlands*, pp. 111-117.

Kerkhoff J. and Wester J. (1987), *Fonpars1 user manual, Part I: rule format*, internal publication Institute of Phonetics, Nijmegen University.

Curriculum vitae

30 september 1957	Geboren te Eindhoven.
aug. 1970 – juni 1976	Bisschoppelijk College Roermond. Gymnasium β.
sep. 1976 – dec. 1984	TU Eindhoven, Elektrotechniek. Afstudeerrichting: Digitale Systemen.
maart 1985 – dec. 1987	Wetenschappelijk onderzoeker TU Eindhoven, gedetacheerd op het IPO voor het verrichten van onderzoek naar de toepasbaarheid van spraaksynthese in een communicatiehulpmiddel.
jan. 1988 – heden	Toegevoegd onderzoeker in dienst van het SPIN programma "Analyse en synthese van spraak", via de TU Eindhoven gedetacheerd op het IPO. Onderwerp: realisatie van een "stand-alone" tekst-naar-spraak systeem.

Stellingen

1. Ondanks de waarschuwing van Gordon en Zabo worden hulpmiddelen voor gehandicapten helaas vaak op maat gemaakt voor één individu.

   Gordon D. and Zabo D. (1984), "Technical solutions for people with communication impairments", *Proceedings of the International Congress on Technology & Technology Exchange, Pittsburgh USA, 8 - 10 October 1984*, pp. 42 - 43.

2. Terecht stelt Soede, reeds in 1980, dat publiciteit over onderzoek naar hulpmiddelen voor gehandicapten vóór het tijdstip van produktie beperkt moet blijven tot wetenschappelijke en professionele literatuur.

   Soede M. (1980), "Development and evaluation of complex aids: a case study", in *The Use of Technology in the Care of the Elderly and the Disabled*, ed. by Bray J. and Wright S. (Frances Pinter Ltd, London), pp. 93 - 99.

3. Een van de grootste problemen van gebruikers van communicatiehulpmiddelen voor gehandicapten, gesignaleerd door Newell, is het gebrek aan snelheid waarmee zij deze kunnen bedienen.

   Newell A.F. (1986), "Speech communication technology - lessons from the disabled", *Electronics & Power*, September 1986, pp. 661 - 664.

4. Een tekst-naar-spraak systeem alleen is nog geen communicatiehulpmiddel voor gehandicapten.

   Deliege R.J.H. (1989), "An experimental Dutch keyboard-to-speech system for the speech impaired.", *Speech Communication* no. 8, 1989, pp. 81 - 89.

5. Het niet in de beschouwing betrekken van de basis-declinatielijn door sommige auteurs levert een onvolledige beschrijving op van het verschijnsel declinatie.

   Liberman M. and Pierrehumbert J. (1984), "Intonational invariance under changes in pitch range and length", in *Language Sound and Structure*, ed. by Aronoff and Oehrle (MIT press, Cambridge), pp. 157 - 233.

6. De kwaliteit van de huidige spraaksynthese is goed genoeg voor toepassing in spraakvervangende communicatie-apparatuur.

7. Bij het in de praktijk evalueren van hulpmiddelen voor gehandicapten bestaat grote kans op het wekken van onvervulbare verwachtingen ten aanzien van gebruiksduur van het evaluatiemodel en beschikbaarheid van een uiteindelijk ontwerp.

8. Bij evaluatie op het gebied van hulpmiddelen voor gehandicapten is persoonlijke begeleiding en observatie door de onderzoeker noodzakelijk.

9. In de opleiding tot ingenieur wordt te weinig aandacht besteed aan de ontwikkeling van de schrijfvaardigheid.

10. Het branden van de kantoorverlichting in veel kantoren op momenten dat dit niet nodig is, bewijst dat het ontsteken van deze verlichting voor velen een te automatische handeling is.

René J.H. Deliege
Eindhoven, 31 oktober 1989