An experimental DSP-based tactile hearing aid : a feasibility study
Mathijssen, R.W.M.

DOI:
10.6100/IR357357

Published: 01/01/1991

Citation for published version (APA):
An Experimental DSP-Based Tactile Hearing Aid

a feasibility study

Roland W. M. Mathijssen
About the cover photo

Although the cover photo has strictly speaking no direct relation to the tactile hearing aid discussed in this thesis, it was considered that there were enough parallels to make its use relevant. The picture shows the image of a hand, impressed in a so-called image captor, a rectangle full of rods that can follow the shape of an object pressed into it. It relates to this thesis in the following ways:

1) the image is that of a hand: in the experiments, the tactile aid presented the information onto the hand (or more precise, the fingertip);

2) the image of the hand is formed using a matrix of rods; the tactile aid also offers shapes using a matrix of (vibrating) rods;

3) the image shows a (universal) sign, derived from the hand alphabet (finger-spelling) for the deaf, sometimes used to support speech-reading: the tactile aid is intended to support speech-reading;

4) by means of the rods from the image captor, the image is digitized in the space domain; the tactile aid uses a digital signal processor, that uses digitized sound signals;

5) some people do not recognize the image immediately—once it has been recognized it usually becomes a clear image; tactile patterns offered on the skin are not always recognized at once by most people—when one is used to perceiving the patterns, they can be felt clearly in most cases.

The sign made by the hand is (in finger-spelling [Janssen, 1986]) a combination of the signs for the letters I, L and Y, an abbreviation of: ‘I Love You’. 
An Experimental DSP-Based Tactile Hearing Aid

A Feasibility Study

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof. dr. J.H. van Lint, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op vrijdag 20 september 1991 om 16.00 uur

door
Rolandus Wilhelmus Maria Mathijssen

geboren op 26 augustus 1962 te Roosendaal

druk: Koninklijke van Poll, Roosendaal
Dit proefschrift is goedgekeurd door de promotoren

Prof. dr. ir. J.E.W. Beneken
en
Prof. dr. H. Bouma

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Mathijssen, Rolandus Wilhelmus Maria

An Experimental DSP-Based Tactile Hearing Aid: a feasibility study
Rolandus Wilhelmus Maria Mathijssen. - [S.l. : s.n.]. - Ill., fig., tab.
ISBN 90-9004216-4
NUGI 832
Trefw.: doofheid / hoorprothesen; tastzin.
Preface

This thesis describes the work carried out at the division of Medical Electrical Engineering of Eindhoven University of Technology, from September 1986 until December 1990 on tactile transfer of information extracted from speech and sound for the profoundly deaf.

This research incorporates information from various fields, such as electronics, digital signal-processing, speech-analysis, phonetics, linguistics, biology, medicine, psychology, and speech-therapy. Also information about several types of disabilities and handicaps—such as blindness and deafness—and information typically concerned with these disabilities was gathered from various sources.

Much was learned from discussions with people connected in different ways with the above mentioned fields. Although most information is available in written form, this could not always be located. It was especially difficult to acquire suitable literature from those areas of research not commonly related to electronics or physical science. This situation also applies when for example a knowledge base is compiled for medical artificial intelligence systems [Blom, 1990].

The research project “Tactile Information Transfer for the Profoundly Deaf” was originally started with two researchers. The project was divided into two parts. One part was the technical realization of the tactile hearing aid, while the other part was to develop optimum coding strategies to present the extracted information onto the skin. Unfortunately this last part was stopped before any useful results had been obtained. The split research project was therefore combined again. Consequently the attention of the remaining researcher could not be equally divided between both parts, which means that in practice the strategy of coding the extracted information is now based on know-
Preface

ledge obtained from literature and from deaf people and their hearing colleagues. A certain amount of intuition obtained during the development of the tactile hearing aid and some preliminary experiments on coding [Wang, 1990] has also been of importance.

To complete this thesis and the work that led to its being written, a lot of people have given their assistance in one way or another. I want to express my thanks to everyone who has been of help. Especially to ir. W. Leliveld and Prof. dr. ir. J. Beneken without whom this thesis would never have been possible. Thanks to prof. dr. H. Bouma, prof. dr. ir. v. Bokhoven, ir. F. Jorritsma, dr. F. Coninx, dr. ir. L. Vogten, prof. R. Plomp, dr. de Kousemaeker, N.C.B. Durrant and Fa. Tieman for their advice and cooperation on several points. Also thanks for reasons they will know, to: Herman Ossevoort, Hans Bosch, Anton van Uitert, Sjef Couwenberg, Ronald Waterham and Xue Wang; to Ronald Mies, Jaap Heffels, Harm van Eijnsbergen and Guido Lemeer; to Ankie van Turnhout; to Hein Panken, Patrick Deykers and of course Xandra Docters. Without them the completion of this thesis would have been much more difficult and less pleasant. Thanks are further due to everyone not mentioned here, but definitely not forgotten, who has supported this work in all manner of ways. Finally my thanks go to my parents for enabling me to do this Ph.D. work.

Roosendaal, April 1991.
Summary

The only way in which profoundly deaf people can understand speech communication in everyday situations is by using speech reading (also called lip reading). This means that a speaker whose mouth either moves less clearly (people who articulate poorly or speak too fast), or can be seen less clearly, is difficult to understand by someone who is deaf. But even in optimum conditions, i.e. when someone speaks clearly, not too fast, and in good lighting conditions, speech-reading is not easy and is definitely strenuous, because a number of vocal and articulatory movements are not visible externally.

Furthermore deaf people cannot hear the everyday sounds that now and then carry important information, such as car horns, shouting, or a train approaching.

It would help deaf people if they could obtain some information about speech and other sounds. At least the knowledge that there is a (loud) sound can give important information. In this thesis, an experimental system will be described that is able to inform that sounds are present, and that can offer supplementary information to speech reading.

Since this information cannot be received acoustically, an alternative way to present it will be needed. We have decided not to offer the information visually since deaf people use (and need) their eyes at least as much as hearing persons, such as for speech reading and seeing things that hearing persons can hear. The skin, or more precisely the tactile sense, might be a useful alternative, which is the basis of the system described here; hence its name: Tactile Hearing Aid.

Since it is not yet certain which speech information is most suitable for the tactile sense, the system is so designed that it is in principle possible to extract various sorts of information from the speech signal or any acoustic signal. This is realized by using a so-called Digital
Summary

Signal Processor (DSP), a micro-processor especially designed to effect computations on (e.g. audio) signals.

In this thesis several algorithms will be described for the tactile hearing aid, that can extract the pitch, or the formant frequencies from a speech signal, or the frequency spectrum or the amplitude of any (acoustic) signal, or a combination of these. All these algorithms work close to real-time and offer the tactual information within ca. 50 milliseconds.

One method, using an algorithm that extracts formant information and energy from speech sounds, has been tried out in a preliminary test with both deaf and hearing subjects. It appeared that all subjects had higher recognition scores in speech reading when the tactile supplementary information was offered to a fingertip. The training needed for a significant rise in recognition score was relatively short: depending on the subject it varied from about 5 to 10 hours. For the deaf subjects, an average relative rise of 50% was observed in the recognition scores when tactile supplemental information is offered.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>iii</td>
</tr>
<tr>
<td>Summary</td>
<td>v</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>- Tactile communication aids</td>
<td>3</td>
</tr>
<tr>
<td>- Commercially available tactile aids</td>
<td>4</td>
</tr>
<tr>
<td>Scope of the study</td>
<td>5</td>
</tr>
<tr>
<td>Research environment</td>
<td>6</td>
</tr>
<tr>
<td>2 Requirements for a tactile aid</td>
<td>7</td>
</tr>
<tr>
<td>- Target group</td>
<td>7</td>
</tr>
<tr>
<td>- Offering tactual information</td>
<td>11</td>
</tr>
<tr>
<td>- Speech-reading information</td>
<td>12</td>
</tr>
<tr>
<td>- Recognizable phonemes</td>
<td>12</td>
</tr>
<tr>
<td>- Offering speech information</td>
<td>14</td>
</tr>
<tr>
<td>Summary</td>
<td>19</td>
</tr>
<tr>
<td>- Basic technical requirements for a tactile hearing aid</td>
<td>20</td>
</tr>
<tr>
<td>- Signal processing</td>
<td>21</td>
</tr>
<tr>
<td>- Mechanical demands</td>
<td>22</td>
</tr>
<tr>
<td>- Electrical demands</td>
<td>23</td>
</tr>
<tr>
<td>3 Signal processing</td>
<td>25</td>
</tr>
<tr>
<td>- Introduction</td>
<td>25</td>
</tr>
<tr>
<td>- Presentable information</td>
<td>27</td>
</tr>
<tr>
<td>- Common aspects</td>
<td>30</td>
</tr>
<tr>
<td>- Formant analysis</td>
<td>31</td>
</tr>
<tr>
<td>4 Signal processors</td>
<td>33</td>
</tr>
<tr>
<td>- Analog processors</td>
<td>33</td>
</tr>
<tr>
<td>- Pure analog processors</td>
<td>34</td>
</tr>
<tr>
<td>- Combined analog/digital processors (hybrid processors)</td>
<td>34</td>
</tr>
<tr>
<td>- Neural network computers</td>
<td>35</td>
</tr>
</tbody>
</table>
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital processors.</td>
<td>37</td>
</tr>
<tr>
<td>Choice of the signal processor</td>
<td>38</td>
</tr>
<tr>
<td>Choice</td>
<td>42</td>
</tr>
<tr>
<td>Which DSP?</td>
<td>44</td>
</tr>
<tr>
<td><strong>5 Displays</strong></td>
<td></td>
</tr>
<tr>
<td>Perceiving the information</td>
<td></td>
</tr>
<tr>
<td>The ear</td>
<td>45</td>
</tr>
<tr>
<td>The nose / the tongue</td>
<td>46</td>
</tr>
<tr>
<td>The eye</td>
<td>47</td>
</tr>
<tr>
<td>The skin (i.e. the tactile sense)</td>
<td>48</td>
</tr>
<tr>
<td>Tactile displays</td>
<td>51</td>
</tr>
<tr>
<td>Stimulating the skin</td>
<td>53</td>
</tr>
<tr>
<td>The sense of touch</td>
<td>55</td>
</tr>
<tr>
<td>Mechanical stimulation versus electrical stimulation</td>
<td>57</td>
</tr>
<tr>
<td>Laser stimulation</td>
<td>60</td>
</tr>
<tr>
<td>The display used for the experiments</td>
<td>61</td>
</tr>
<tr>
<td>A new vibrotactile display</td>
<td>62</td>
</tr>
<tr>
<td>Coding strategies</td>
<td>62</td>
</tr>
<tr>
<td>Presenting formant information</td>
<td>63</td>
</tr>
<tr>
<td>Presenting pitch information</td>
<td>66</td>
</tr>
<tr>
<td>Summary</td>
<td>68</td>
</tr>
<tr>
<td><strong>6 Technical description</strong></td>
<td></td>
</tr>
<tr>
<td>The hardware of the system</td>
<td>70</td>
</tr>
<tr>
<td>The digital signal processor</td>
<td>73</td>
</tr>
<tr>
<td>External memory</td>
<td>78</td>
</tr>
<tr>
<td>Peripherals</td>
<td>83</td>
</tr>
<tr>
<td>Analog components</td>
<td>87</td>
</tr>
<tr>
<td>Power supply / voltage reference</td>
<td>91</td>
</tr>
<tr>
<td>Total scheme</td>
<td>91</td>
</tr>
<tr>
<td>The software of the system</td>
<td>94</td>
</tr>
<tr>
<td>Architecture of the software</td>
<td>94</td>
</tr>
<tr>
<td>Timing of the software</td>
<td>99</td>
</tr>
<tr>
<td>The signal processing algorithms</td>
<td>101</td>
</tr>
<tr>
<td>The algorithms in practice</td>
<td>104</td>
</tr>
<tr>
<td>Suggestions for the future</td>
<td>106</td>
</tr>
<tr>
<td>Reducing the power consumption</td>
<td>106</td>
</tr>
<tr>
<td>Reducing the size of the system</td>
<td>109</td>
</tr>
<tr>
<td>Hardware suggestions</td>
<td>111</td>
</tr>
</tbody>
</table>
A great part of the everyday communication between people takes place by means of speech, or more generally, by means of sound. Most people take this act of acoustically transferring information for granted. For a small part of the population however, perceiving sounds or speech is not self-evident. It is estimated that 3 to 4% of the human population suffers from impaired hearing, which makes the understanding of speech more difficult [Plomp, 1978]. Most people who have a hearing impediment are ‘hard of hearing’. This means that their average hearing loss is approximately between 35 dB and 90 dB. The use of a conventional hearing aid, which amplifies the (speech) sound, is normally enough to enable them to understand speech in a quiet surrounding. With the help of some speech-reading (sometimes called lip-reading), their ability to understand speech can increase further.

People who have an average hearing loss of about 90 dB or more (about 0.1% of the population [Sacks, 1990; Van Cleve, 1987; Breed & Swaans-Joha, 1986]), are generally not able to understand speech solely by the acoustic information that they might be able to perceive. These people are called ‘deaf’, and they can only perceive speech by means of speech-reading, sometimes partly aided by the little hearing they have.

Speechreading is the art of understanding speech by looking at the articulatory movements, combined with the facial expression and the movements of the speaker, and other information that might be obtained from the surroundings. Although almost everyone who is not born visually impaired has some experience with speech-reading, as it is learned more or less automatically as an infant [Dodd & Campbell, 1987], it is still a very difficult task. As will be explained further in the next chapter, only a small part of a spoken message can be perceived by mere speech-reading. The missing parts have to be filled in by a
Introduction

mixture of guesswork and context information; much more than is the case with hearing people. This is generally below the conscious level.

To facilitate speech-reading it is possible to offer extra or supplemental information. When talking with one another, deaf or severely hard of hearing people often use certain hand and arm movements, called signs, to make speech-reading easier. Sometimes deaf people do not use speech at all to communicate, but use a sign language instead [Sacks, 1990]. However, sign language alone does not allow for communication with most hearing people, because they do not understand this language. In fact, most hearing people hardly know the correct signs to facilitate speech-reading, other than certain generally used symbolic signs.

It will be clear that deaf people, when talking with hearing people, can usually obtain only limited supplementary information. Therefore it might help the deaf if a communication aid could be developed that automatically offers some form of supplementary information for speech-reading.

There are several ways of offering supplementary information to the deaf. Where possible it might be offered acoustically; however, not all deaf people have enough hearing ability for this purpose. Another possibility is to supply the information at the acoustic nerve—a technique used in cochlear implants, e.g. [Pickett, 1987a; Peeters, 1990]. When the hearing deficit is situated distal (i.e. further away from the brain) to the acoustic nerve, this technique can be used quite successfully. When the deficit is situated at, or proximal to, the acoustic nerve, this technique cannot be used.

The remaining possibilities for supplying the supplementary information are the visual channel (the eye) and the tactile channel (the skin). Both channels have been tried for offering supplementary information [Pickett, 1987b]. These alternative channels are further discussed in chapter 5. For now it will suffice to state that the visual channel is less often used. The main reason for this is probably that the visual channel is too much occupied with speech-reading. In this thesis we shall focus on the tactile channel. Of course this channel has its limitations too. The main limitation is the limited transfer rate of this channel. It is probably not possible to transfer enough information for perceiving

- 2 -
speech in real-time without the need for other information, such as from speech-reading. Further, the tactile sense is most sensitive to frequencies between 200 to 400Hz. To perceive other frequencies higher vibration amplitudes are needed. This means that a communication aid that uses the tactile sense can probably only offer supplemental information for speech-reading.

Tactile communication aids

The earliest recorded use of the tactile channel for speech communication was in 1924: [Gault, 1924]. Gault used a 14-foot long tube to conduct the "vibrations of the experimenter's vocal apparatus" to the palm of a subject's hand. One end of the tube had a small inlet for speech sounds, while the (hearing) subject held his hand at the other end. Sound insulation assured that the subject could hear no sounds. After extensive training the subject was able to distinguish between several words.

It was quite a number of years before other investigations into tactile communication were published. In the 1960's, however, the interest in tactile aids started to rise again [Risberg, 1983; Sherrick, 1984]. The tactile aids that were developed can be roughly divided in 4 groups, according to the way the acoustic signal is processed before it is fed to the skin:

- no special coding: the acoustic signal is supplied almost directly to the skin. The only forms of processing can be filtering and amplification. Usually a low pass or band pass filter is used, that filters out frequencies above 500 to 1000 Hz. In this way the only frequencies offered are those to which the tactile sense is most sensitive and which contain the fundamental frequency of speech. The (filtered) signal is transmitted to just one place on the skin. Gault's tube can be considered as such a tactile aid.

- special filtering of the signal: with a special filter, usually a band-pass filter combined with a digital circuit, a single (speech) parameter is extracted from the acoustic signal. This parameter is often the fundamental frequency. Since this frequency is available in a
Introduction

digital form (as a number), the frequency can easily be transformed into spatial information or into frequency information.

- by dividing the signal into a number of frequency bands with a filterbank, information is available about the energy in a number of frequency bands. Each frequency band has its own ‘display’ of one or more excitators, thus forming a one-dimensional or two-dimensional array. The frequency information is presented spatially; the energy information per band is displayed either as energy (amplitude differences of the excitators; one display per band) [Özdamar et al., 1987; Leysieffer, 1986] or as spatial information (at higher energies either more excitators are active, or another excitator becomes active) [Sparks et al., 1978].

- by using special techniques to extract other information from the acoustic channel. This includes in fact anything that does not fit into one of the other groups. One could consider a tactile aid that displays information about phonemes (e.g. plosives, voiced signal, nasals, etc.). Also aids that discriminate between certain environmental sounds fit in this group [Miyazaki & Ishida, 1987].

Commercially available tactile aids

At present several tactile hearing aids are commercially available [Franklin, 1988]. Most of the available aids are single-channel aids. The acoustic signal is supplied to the skin without much processing. Some aids split up the signal into two channels. One channel displays the overall energy of the signal, while the other channel displays frequency information (usually the fundamental frequency). The few commercially available multiple-channel devices all work according to the filterbank principle.

Most of the aids that have been tested show an improvement in speech-reading scores or seem to enable the users to recognize words without the need of speech-reading [Sherrick, 1984; Risberg, 1983]. When the aid is used as a speech-reading aid, it is usually used for a couple of weeks to a couple of months. After that period the users often show an increased speech-reading ability, without the need to use the device further. At this stage they often stop using the device, since it
does not further improve the speech-reading. In this case the device functions mainly as a training aid for speech-reading. The literature records that users can learn to recognize up to about 200 words or sentences when the device is used independently of speech-reading [Goldstein & Proctor, 1985]. Further results that might show better scores are not found in literature. There is also no good comparison between the various tactile aids, that shows what kind of information should be offered in which way.

Scope of the study

In the previous paragraphs we have seen that a communication aid for the profoundly deaf could be useful. A possible method for supporting speech-reading is tactile presentation of the speech signal. Various tactile aids have been developed, but only a few of them are currently available. Every available aid uses only a limited form of signal processing. Since the skin has a low channel capacity, the correct data need to be presented to the tactile sense. By processing speech properly, it is hoped to design a tactile hearing aid that offers better information to support speech-reading than the currently available ones.

The aim of this study is to develop an experimental tactile hearing aid that can perform various signal processing algorithms. With this system it should be possible to compare the improvement in speech-reading scores using various processing techniques using the same starting conditions (the same kind of device and the same kind of tactile display). This way we hope that a reliable comparison can be made.

Since we want the tactile aid to use software-based processing techniques, it is also intended for evaluating novel algorithms that can be compared reliably with existing algorithms.

In this thesis the hardware and software for such an experimental tactile hearing aid will be described. Details of a preliminary experiment with a limited number of subjects will also be presented. Owing to the limited time available no comparison between signal processing algorithms and display methods has been made yet. The hardware and the software for such an experiment is available, however.
Introduction

In chapter 2 the design requirements for a tactile hearing aid, and more specifically for the tactile hearing aid discussed in this thesis, are discussed. Various methods and algorithms for processing speech and sounds and available components for signal processing are the topic of chapter 3 and 4. Chapter 5 will describe the different ways of presenting acoustic data, especially for auditory impaired people. This is followed by some data about tactile stimulation and the resulting design requirements for a tactile display. Possible ways of presenting the tactile stimuli conclude that chapter. The technical description of the developed tactile aid will be described in chapter 6, together with several facts and suggestions for future research on the tactile hearing aid. Preliminary experiments with the tactile hearing aid and their results are presented in chapter 7. Finally the conclusions of this research can be found in chapter 8.

Research environment

At Eindhoven University of Technology about 20 research groups are working on research and education on Biomedical and Health Care Technology (BMT, or BMGT). One group is the Division of Medical Electrical Engineering (EME). The research activities of EME cover the following areas:
- Servo-anesthesia
- Ultrasound imaging techniques
- Instrumentation for the disabled

A number of the projects from the last research activity are carried out within the Interfaculty Working Group Communication Aids for the Handicapped, consisting of members of EME and the Institute for Perception Research (IPO) in Eindhoven. The tactile hearing aid, described in this thesis is one of these projects.
In the previous chapter we have suggested that a tactile aid might be a useful communication aid for the severely auditory impaired. In this chapter we shall try to explain when such a hearing aid can be useful. Further we shall discuss the requirements for such an aid:

- what kind of information should be presented to the skin;
- how can this information technically be extracted from the sound signal;
- what are the electrical and mechanical requirements for a tactile hearing aid.

Target group

The first question when developing a tactile hearing aid is how the group of possible users is composed.

In chapter 5 we shall see that the tactile sense has only a limited ‘channel capacity’, i.e. only a small number of data, presented as tactually perceivable patterns, can be transferred per second. This information rate is much lower than the information rate that can be found in normal running speech. Even when the speech information is represented as characters, as in written language, the information rate is still too high, assuming that one speaks at an average speed.

When the acoustic channel still retains some function, its channel capacity is usually higher than the skin’s capacity, presenting correctly matched information for each sense. This means that only people who have virtually no hearing left (i.e. who are profoundly deaf) can be considered as possible users of a tactile hearing aid. In this case ‘deaf’ is defined as having a hearing loss of more than 90 dB [Breed & Swaans-Joha, 1986; Franklin, 1988], or as having so little hearing left,
Requirements for a tactile aid

that it is not possible to understand speech, even with the help of the most advanced (acoustic) hearing aids, other than by using non-acoustical information (such as speech-reading) [CBS, 1976].

In 1976 it was estimated that for the Dutch population (at that time of about $13.10^6$ people), about 12,000 people were deaf [CBS, 1976]. More recent estimates vary from between 8,000 and 14,000 [Breed & Swaans-Joha, 1986] to about 28,000 [Tervoort, 1987]. These differences might be caused by differences in the used definitions. A rule of thumb is that 0.1% of the population is born deaf [Sacks, 1990; Van Cleve, 1987], which translates to about 15,000 prelingual deaf for the Netherlands.

Hard of hearing people, i.e. with a hearing loss of less than 90 dB, who are not helped with a conventional hearing aid, and some deaf people who nevertheless retain some useful hearing ability might be helped when the acoustic information is coded or processed properly. Nowadays hearing aids are becoming available that can perform some form of signal processing e.g. in order to reduce noise: [van Dijkhuizen et al., 1987; Stein et al., 1989].

For the auditory impaired who cannot be helped with an hearing aid which offers its information acoustically, there is the possibility that a cochlear implant can be used [Pickett, 1987a; Tye-Murray & Tyler, 1989; Tye-Murray et al., 1990]. The amount of information that can be offered by a cochlear implant is—at present—usually less than can be offered purely acoustically as long as some hearing is intact. This is partly a result of the technique currently used in cochlear implants, which cannot yet be fully optimized owing to technical limitations, and partly because of the limited number of electrodes that can be implanted and the ‘cross-talk’ between the electrodes in the cochlea (especially when electrodes are implanted close together).

Since the auditory nerve is stimulated directly, and the (artificial) information arrives directly in the acoustic centre of the brain, the information transfer capacity (the quantity of data, e.g. in bits, that can be transferred per unit of time) of a cochlear implant can be superior to that of a tactile hearing aid. As we shall see in chapter 5, the amount of information that can be transmitted per second via the tactile sense is rather low, compared to the auditory and the visual sense.
Therefore the following question arises:
if a cochlear implant is able to transfer more information about the
acoustic signal than a tactile aid, why then still perform research on
tactile hearing aids?

A cochlear implant uses the auditory nerve as its information channel.
But this is only possible when the hearing impairment is caused by a
‘malfunction’ situated in the acoustic path before the nerve (distal): the
complete route from the beginning of the auditory nerve (near the
cochlea) to the auditory centres in the brain, and of course the auditory
centres themselves, have to be intact. The ‘malfunction’ is in the inner
ear (fig. 2-1).

Even when the above conditions are fulfilled, it might not always be
preferable to implant a cochlear hearing aid. The use of such an aid
always requires surgery. Furthermore, there is sometimes a theoretical
Requirements for a tactile aid

chance of a spontaneous improvement of the hearing ability. Once a cochlear implant has been brought in, this small chance is definitely eliminated. For this reason the criteria for implanting a cochlear implant are rather strict [Pickett, 1987a].

The use of a tactile communication aid does not require surgery, has no effect on (possibly returning) residual hearing functions and can be stopped whenever desired, while no artificial components remain in or on the user. The basic requirement for a tactile aid is that the tactile sense is not impaired.

Thus we come to the category of people who are possible users of a tactile hearing aid:

- those with a hearing loss of above 90 dB;
- those whose tactual sense is not impaired;
- those for whom implantation of a cochlear aid is not meaningful\(^1\);
- and
- those for whom there is a possibility of spontaneous hearing improvement.

Considering the age of possible users one can distinguish three categories: children, adults younger than about 50, and adults older than about 50 years. It is assumed that the latter group has more difficulty in learning to perceive tactile patterns. (It may be noted that braille is most frequently successfully learned before the age of about 50. It is not known if a parallel can be assumed for users of a tactile hearing aid). As yet there is no reason for not starting to apply a tactile hearing aid when one is over 50. In our tests (see chapter 7), the one subject that was close to the assumed critical age had no difficulty at all using the aid.

Children are often used as subjects in experiments with a tactile hearing aid [Brooks et al., 1987; Goldstein & Proctor, 1985], and they seem to

\(^1\) *It should be noted that a cochlear implant does not exclude the use of a tactile hearing aid. A combination of both aids has been suggested [Pickett, 1987a].*
be perfect candidates. Not because they are too young for a cochlear implant (cochlear aids can be implanted in very young children [Berliner et al., 1989]), but the acquisition of (spoken) language and speech is easier when one can perceive at least some information about speech sounds. A tactile aid can be a means of providing this information. Only when one chooses to raise a deaf child with sign language as the first language could the above mentioned reason for using a tactile aid be dropped.

Finally adults who are younger than 50 can use a tactile aid too. It can be used for improving speech-reading, but also as a signalling device for the presence of sound. In fact, all age groups can use a tactile hearing aid for this latter purpose.

When looking for subjects for evaluating a tactile hearing aid, it is probably best not to use deaf children. It is possible that the experiments ask too much time from the child, which can negatively influence its performance at school. Also the child might get used to the aid too much, although he can use it only during a fixed period of time. When he can no longer use the aid, it might be difficult to get used to the ‘old’ situation without the aid. We therefore conclude that the best deaf subjects for evaluating a tactile hearing aid are adults younger than 50.

Offering tactual information

Now we have an idea of who might use a tactile hearing aid, we can consider what kind of sound information a tactile hearing aid should offer. One choice is offering information about every sound in the vicinity. Another choice is offering information specifically about e.g. speech, or environmental (traffic) sounds [Miyazaki & Ishida, 1987].

The older tactile aids offer the tactile sense acoustical information about every sound. This has the advantage that the user always knows that there is a sound in the vicinity. Identification of the sound depends on the skill of the user.

Novel tactile aids sometimes offer specific information about one type of sound, usually speech [Risberg, 1983; Leysieffer, 1986]. Some of
Requirements for a tactile aid

these features can be very specific indeed, and are normally not properly detected in non-speech sound. Further investigation as to whether these specific features can offer useful information about environmental sounds too is needed.

The advantage of extracting and offering such specific information is that more information about that one specific sound can be transferred to the user. Especially for someone who is speech-reading, correctly chosen information can improve the recognition scores significantly [Breeuwer & Plomp, 1986; Sparks et al., 1978; Spens, 1981; Risberg, 1983]. The disadvantage of offering such specific information is that the user has hardly any information about other sounds.

It cannot be said in general what kind of information is preferable. For supporting lip-reading, the best choice seems to be offering specific information about the speech signal, rather than global information. For informing the user about the presence of sound, it is recommendable to offer at least the energy (amplitude) of the overall sound signal. In this thesis we shall mainly concentrate on the speech signal. Therefore the algorithms that have been developed for the tactile aid offer energy information too. This way the device will also have a signalling function.

Speech-reading information

An aid for supporting speech-reading is most useful when it offers to the tactile sense that kind of information that cannot be obtained by speech-reading alone, but which is useful for identifying different sounds. Knowing that only a limited amount of information per unit of time can be dealt with properly, several ‘sets of information’ can be found. Let us start by summarizing which sounds can be identified by speech-reading alone. Next, several possibilities for the exact information that can be presented will be discussed.

Recognizable phonemes

The Dutch language has about 40 different phonemes [Vogten, 1983; Nooteboom & Cohen, 1984]. A phoneme is defined as the smallest
distinctive unit of speech. For example the Dutch word “WOORD” consists of 4 phonemes: /W/ /OO/ /R/ /T/. Only a small number of sets of phonemes can be recognized by speech-reading. Depending both on the skill of the speech-reader and the visibility of the speaker’s mouth and the quality of articulation, roughly up to eight sets of phonemes can be discerned. Distinguishing between phonemes within one group is normally not possible [Corthals et al., 1986], e.g. the words ‘papa’ and ‘mama’ look precisely the same. Only very skilled speech-readers are sometimes able to recognize and discern more phonemes, when the speaker speaks clearly and is known to the speech-reader.

The Dutch phoneme sets that are normally recognizable are:

1. /A/, /AA/ (as in Dutch ‘bal’ and ‘baal’)
2. /IE/, /EE/, /I/ (as in Dutch ‘kiek’, ‘keel’ and ‘kil’)
3. /O/, /OO/ (as in Dutch ‘bot’ and ‘boot’)
4. /P/, /B/, /M/ (as in Dutch ‘pot’, ‘bot’ and ‘mot’)
5. /UU/, /OE/, /W/ (as in Dutch ‘buur’, ‘boer’, and ‘waar’)
6. /V/ (as in Dutch ‘fel’ and ‘vel’)
7. /E/ (as in Dutch ‘bel’)
8. /L/ (as in Dutch ‘look’)

The order of the above sets of phonemes roughly corresponds with the chance from high to low of recognizing a phoneme or set of phonemes. The /L/ for example can only be recognized by a skilled speech-reader and then only when the /L/ is articulated clearly.

As can be seen in the above list, most consonants cannot be recognized as such at all by speech-reading. This is because most consonants are formed in the back of the speech channel, as opposed to vowels, that are formed mostly in the front parts of the speech channel; the form of the lips plays an important role. When producing a consonant the form of the lips usually depends on the vowels proceeding and following.

---

2 For practical reasons the standard phoneme-symbols have not been used in this thesis. See appendix 2 for a list of Dutch phonemes and the symbols as used in this thesis.
Requirements for a tactile aid

For example, the /T/ in the words ‘tip’ and ‘top’ cannot visually be distinguished separately from the /I/ and the /O/ in those words.

Knowing that only such a small group of phonemes can be discerned visually, it will be clear that speech-reading is very difficult. A lot of guesswork is needed, with the help of context information and redundancy which are present in speech, to make sense from the information that can be perceived. Only with a lot of skill is it possible to perceive speech good enough for most everyday situations by mere speech-reading. Measured speech-reading scores from skilled speech-readers can go up to about 70% (or sometimes even higher) correctly perceived phonemes for words offered in short, meaningful sentences (see chapter 7).

Offering speech information

In this section various forms of information are presented that can possibly be offered as information additional to speech-reading. Both information that can already be extracted easily and information which cannot yet be extracted, except by very powerful computer systems, but which would probably be very useful will be discussed. The information offered does not necessarily have to have a direct relation to phonemes. Other types of information can also improve speech-reading [Breeuwer & Plomp, 1986; Sherrick, 1984; Kirman, 1982; Plant & Risberg, 1983, Risberg, 1983].

* Pitch information

When deaf people or their teachers are asked what kind of information it would be useful to offer, they often answer the ‘rhythm’ (or energy) of the speech signal or the ‘pitch’. Since we wanted the tactile aid to be a signalling device too, it presents amplitude information. In this section we shall focus on the pitch.

It appears that the perceived pitch of a (complex) signal is closely related to the pitch that can be perceived when a pure sinewave is offered with the frequency of the fundamental frequency (also called \( f_0 \)) of that (complex) signal [Goldstein, 1973; Gold & Rabiner, 1986].
In a speech signal, the pitch is related to the frequency of the vocal cords (during voiced speech).

In a frequency analysis of a speech fragment, the fundamental frequency is often visible as a series of equi-distant frequency peaks, see figure 2-4. The distance between those peaks (in figure 2-4 approximately 200 Hz) is this fundamental frequency.

Neither the pitch nor the energy of the speech signal can be perceived by speech-reading alone, yet both parameters contain useful information [Breeuwer & Plomp, 1986]. The first single-channel tactile aids offered information containing energy and/or fundamental frequency ($F_0$). The fundamental frequency was usually offered as a low-pass filtered speech signal, that also contained energy information [Risberg, 1983; Rothenberg et al., 1977].

* **Direct frequency information**

The first multiple-channel tactile aids offered direct frequency information to the skin. The idea behind this is that the ear—or to be more specific, the cochlea—separates the acoustic signal into small frequency bands. The ear is a frequency analyzer (a filterbank) that transmits frequency information to the hearing centre in the brain.

It seemed quite logical to offer frequency information to someone who is completely deaf. In this way the tactile hearing aid could imitate the cochlea. Tactile aids that worked this way usually offered from 6 to 32 frequency channels [Risberg, 1983; Sparks et al., 1978].

The problem with offering frequency information, however, is that the information transfer rate is too high. The reason why these tactile hearing aids do improve the speech-reading score is probably that the formant information is coded in the frequency information (see figure 2-4 and chapter 3).

* **Formants**

Other information closely related to certain phonemes are the formants [Vogten, 1983; Flanagan, 1972]. Formants result from resonances in
Requirements for a tactile aid

Fig 2-2: Sentence “Er was eens..” Amplitude vs. time

Fig 2-3: Fragment from fig. 2-2: “Er was eens”: vowel /A/

the oral and nasal cavity and appear in the spectrum as energy maxima. This can be illustrated by the following:

Figure 2-2 represents one second of speech. From this signal we take a short segment, called a ‘frame’, during which the signal can be considered stationary (e.g. the 20 msec segment from figure 2-3). The frequency content of the frame from figure 2-3 can be determined which results in figure 2-4. The frequency spectrum of the frame shows
a number of peaks. The distance between adjacent peaks is called the fundamental frequency (see the previous section), indicated by $F_0$. For formants we have to look at the overall spectrum, indicated by the dotted line in figure 2-4. Every (major) peak in the overall spectrum of the speech signal is called a formant frequency (there are usually 5 formants for male speech and 4 formants for female speech in a frequency band from 0 to 5 kHz [Vogten, 1983; Rabiner & Schafer, 1978; Willems, 1987]). The peak at the lowest frequency is called $F_1$, the next (higher) frequency is called $F_2$, and so forth.

The formants in a speech signal are caused by the shape of the vocal tract from the vocal cords to the lips [Flanagan, 1972]. To produce speech the vocal tract changes shape by the movement of the jaws, the lips and the tongue. The vocal tract can be seen as an acoustic filter that because of its shape either enhances or diminishes certain fre-
Requirements for a tactile aid

Frequency bands of the signal produced by the vocal folds or by other parts of the tract. The formant frequencies are the frequencies that are enhanced by this filter. In the field of speech analysis and synthesis this model is called the ‘source-filter model’ [Vogten, 1983; Waterham, 1989].

The formant frequencies contain information about phonemes. Especially $F_1$ and $F_2$ contain phoneme information. The higher-order formants contain less information about the phonemes, but nevertheless they determine the ‘colour’ of the speech.

Extracting formant information from a speech signal works more reliably and much faster than extracting phonetic features (see chapter 3). Breeuwer [Breeuwer & Plomp, 1985] showed that formant information ($F_1$ and $F_2$) — when offered acoustically to hearing persons — can improve speech-reading scores significantly.

* Phonemes

Instead of offering information which is closely related to phonemes, one could think about offering phonemes directly (for example coded on the skin as the shape of characters). Technically this would mean that the aid would have to perform speaker-independent speech recognition. Unfortunately this is not yet possible. But even if it were possible, it is doubtful whether this kind of information could be transferred via the tactile sense:

The maximum amount of data that can be transferred through the skin (coded as characters) appears to be approximately 35 to 40 bits per second [Craig & Sherrick, 1982]. Speech, spoken at a normal rate, contains after maximum reduction about 50 bits per second. Since speech coded into phonemes still contains redundancy, this type of information will have an even higher information rate than 50 bits per second. It is very likely that offering phonemes means offering too much information per unit of time to the skin. And indeed no tactile aids are as yet available that can properly offer information at normal speaking rates. Of course one could offer the information visually, but this creates the problem of finding a suitable method of displaying the information in most everyday situations (see also chapter 5). So, even if offering phonemes were to prove technically possible, it would still create quite a number of problems of realization.
* Phonetic features

One form of information that might be offered and which contains information that is closely related to phonemes are phonetic features [Rabiner & Schafer, 1978; Nooteboom & Cohen, 1984]. These phonetic features can be:

- **Plosive** (e.g. /P/, /B/, /K/, /T/, /D/)
- **Fricative** (e.g. /F/, /V/, /S/, /Z/, /G/)
- **Nasal** (e.g. /M/, /N/, /NG/)
- **Voiced** (e.g. all vowels and /B/, /M/, /V/, /Z/)
- **Unvoiced** (e.g. /P/, /S/, /F/, /T/)

Some of these phonetic features can be extracted from the speech signal (by a computer) reasonably well without too much difficulty, while most other phonetic features can as yet be extracted only with great difficulty. The above phonetic features cannot be perceived by speech-reading alone. Information about these phonetic features offers a speech-reader the possibility of recognizing those phonemes that normally cannot be detected by speech-reading alone (such as the /S/, /T/, and /K/) or of distinguishing between phonemes. Examples of the latter are the phonemes /P/, /B/, and /M/. These phonemes appear the same for a speech-reader (the lips close for a short time). If one could distinguish between a voiced and an unvoiced phoneme, it would help to distinguish between /P/ and {/B/, /M/}. Information about plosives would help to distinguish between /M/ and {/P/, /B/}.

**Summary**

We have tried to show that the information that can be offered to improve the speech-reading score can comprise the following:

- energy (amplitude)
- pitch
- frequency content (Filter bank)
- formants ($F_1$ and $F_2$)
- phonemes
- phonetic features
- combinations of the above
Requirements for a tactile aid

Although there have been attempts to compare some of the above sets of information [Risberg, 1983; Sherrick, 1984; Lemeer, 1990], the results to date do not permit any definitive conclusions. Therefore we have tried to develop an aid that is capable of extracting various of the above sets of information (other than phonetic features and phonemes). Using the same device it is then possible to compare reliably the effect of presenting different types of information. In the next section the basic demands for a tactile hearing aid that can offer one or more of the above sets of information will be discussed.

Basic technical requirements for a tactile hearing aid

In the previous section we have seen that a tactile hearing aid transforms the sound signal to a tactile signal. This results in a basic set-up that is valid for every tactile hearing aid (i.e. also for the experimental tactile aid). Figure 2-5 shows schematically the design for such an aid. The sound signal is picked up, usually by a microphone. Next the signal needs some conditioning (or preprocessing), such as amplification or filtering, before it can be processed (Signal processing). After the signal has been processed, it needs to be coded into certain patterns.

![Diagram](image.png)

*Fig 2-5: Basic diagram for a tactile hearing aid.*
and displayed onto the skin. This is described in the part ‘Tactile display’.

Signal processing

The tactile sense is not capable of recognizing – or even feeling – sound signals with a normal loudness without some artificial device. Very loud sounds might be felt, but speech under normal conditions is never loud enough to be felt (without touching a possible speaker or the face, throat or upper torso of the person who speaks). This is why the speech signal has to be converted (or at least amplified) in order to be perceived by the skin.

After Gault’s tube [Gault, 1924], which can be considered as a mechanical aid, the newer tactile aids were electronic ones. These electronic aids can be grouped according to the way the signal processing is realized, i.e. analog or digital (or both).

- Analog processing techniques make use of (analog) filters, amplifiers, or analog processors (the latter is in fact a combination of filters, amplifiers, integrators, and so forth).

- Digital processing techniques make use of one or more digital processors. Several types of digital processors are now available:
  - ‘Normal’ (von Neumann) micro processors;
  - Reduced Instruction Set Computers (RISC processors);
  - Digital Signal Processors (with a Harvard structure).

All of the above techniques and processors will be discussed in the next chapter.

Electronic aids have a microphone as input device. The signal conditioning is (usually) amplification and filtering of the signal. When digital techniques are used to process the signal, conditioning also covers the digitizing of the signal.

Earlier tactile hearing aids have an analog signal processing stage. Nowadays it is also possible to perform virtually any processing that used to be realized by means of analog techniques by using digital
Requirements for a tactile aid

circuitry. Signal processing can be anything from amplification and further low-pass filtering (either analog or digital) to automatic recognition of sounds. In the schematic drawing of figure 2-5 the tactile display performs coding of the output of the signal processing and it presents the information on the skin; either vibrotactile or electrotactile.

Mechanical demands

A tactile hearing aid should be used daily, both at home and outdoors. This can only be accomplished if the device can be worn without too much hindrance. In other words:

• the device should not be too big. The user must be able to wear the device more or less unobtrusively. Considering that for hearing people a walkman is quite acceptable, the device can have comparable dimensions (but preferably not larger);

• the device should not be too heavy. Considering that the device should be relatively small, this requirement should not pose many problems;

• the device must be able to work long enough without changing or recharging the batteries. The user must be able to rely on the device; it should not stop working while in use. On the other hand, only a limited number of small batteries should be used in order to fulfill the previous requirements.

• the tactile display should be easy to apply. Also it must not irritate the user, even after several hours of use.

\footnote{For evaluations under laboratory conditions, these demands are of course less strict. However, as soon as an experimental system needs to be evaluated in a real-life situation (i.e. outdoors), the demands for this system will be almost the same as for a final version of a tactile hearing aid, i.e. it needs to be small and also battery powered.}
Electrical demands

We have seen that a tactile hearing aid uses some form of signal processing. Although a great number of components can be used for signal processing, the components used by a tactile hearing aid (or any hearing aid) should be small and should not consume too much power. This follows from the mechanical demands.

The same holds for the tactile display: it should be so small that it does not bother the user. On the other hand, the tactile display cannot be too small, or it will not be possible to transmit the desired information (see chapter 5). The tactile display should consume as little energy as possible.

The tactile hearing aid we want to develop should be able to perform various processing techniques. This results in the following:
• the signal processing must operate ‘real-time’: the extracted information must be presented to the skin parallel to the visible information. Only a small time delay (of less than about 45 milliseconds\(^4\)) is permissible;

\(^{4}\) which is equal to the time delay due to a distance of about 15 meters.
Requirements for a tactile aid

- the signal processing part must be able to perform different processing techniques;
- the signal processing algorithm must be easy to modify: no hardware changes should be necessary, other than changing e.g. program ROMs;
- the signal processor must have sufficient processing power, without using too much energy.

For the tactile hearing aid described in this thesis, we have chosen for a Digital Signal Processor. In the next chapter we will discuss why we have chosen for a DSP and which DSP is chosen. With the above design requirements, we have come to the block diagram of figure 2-6. The signal processor performs various functions. Even the coding of the extracted information into tactile patterns is controlled by the DSP.

When we compare figure 2-6 with the general set-up from figure 2-5, we can see that the signal conditioning is realized by an amplifier, a low-pass filter and an analog to digital converter. In this device signal processing comprises:

- signal adjust, or preprocessing. This part is necessary for most processing stages. Preprocessing is usually the splitting of the signal into frames and windowing the signal (see chapter 3);
- processing. In this stage the signal is so processed that the data are ready for feature extraction. Processing is usually Fast Fourier Transform (FFT) Analysis, Linear Predictive Coding (LPC), auto correlation, or techniques derived from these three (see also chapter 3);
- feature extraction. Here the features are extracted from the processed signal. Features can be energy, pitch, formants, frequencies, etc.;
- pattern coding. This is where the extracted features are converted into patterns suitable for the tactile sense.

Finally, on the right of figure 2-6, we have the tactile display, in this case a two dimensional array of vibrators. The technical aspects of the tactile hearing aid will be further discussed in chapter 6.
Signal processing

In the previous chapter we considered the basic requirements for a tactile aid for sound perception. This chapter describes briefly some possible signal processing techniques for the aid. Once we know which techniques we want to implement in the tactile aid, we can decide what type of signal processing hardware we need to use. This choice will be made in chapter 4. Also we shall decide which kind of signal processing will be used for the tests that will be discussed in chapter 7.

Introduction

It is not yet clear what kind of information about the speech (or sound) signal should be offered to the tactile sense, and in which way this information should be best presented for supporting speech-reading or recognizing other sounds. It is even possible that different tactile aid users could be best served by different types of extracted information. To find out what can be offered best, we wanted to develop a (hardware) system that is able to extract different ‘characteristics’ from a signal without the need to re-design the hardware. If it appears that the type of information offered depends on the individual user, it should still be possible to use the same hardware for the final system. In that case the required information can be extracted simply by changing the software.

At present we are using a tactile display of 6 x 24 rods each of which can either vibrate or not vibrate, for use on the tip of a finger. As will be discussed in chapters 5 and 6, it is intended to use a different type of tactile display to that currently used. The current display presents its data on the tip of the finger, while in a final version of the aid the hands of the user should remain completely free. Also it might be possible that, in order to present the extracted information most favour-
ably, different types of displays need to be used. This not merely means displays with a varying number (or arrangement) of rods, but also the possibility of coding the information other than ‘on/off’, (see chapter 6) e.g. using amplitude modulation. Therefore it is best not to take into account the limitations of a display in the pure signal processing stage. The parameters of the display needed to code the extracted information to the best advantage on the display should be used in another (hardware or software) part that transfers this information into the tactile patterns. Figure 3-1 shows this technique in a flowchart. The part ‘Signal Processing’ has no links with the tactile display. The part ‘Display Driver’ on the other hand is fully display-dependent.

Although the part of the system that extracts information from the signal has to be independent of the actual tactile display, it should take

---

**Fig 3-1: The signal processing part should be independent of the display parameters. A special part (Display Driver) contains the display parameters and transforms the extracted information into usable tactile patterns on the display.**
into account that the tactile sense is capable of handling only a very limited amount of data (see chapters 2 and 5).

Presentable information

In the previous chapter we have already shown which information that could be useful for presenting to the tactile sense, might be extracted from sounds in general and speech in particular. How this information can be presented to the skin (i.e. as spatial, amplitude or frequency information, or a combination of these) will be discussed in chapter 5.

In the previous chapter the following features of speech signals were mentioned as possible candidates for supporting speech-reading:

- amplitude
- fundamental frequency
- frequency spectrum
- formants
- combinations of the above

In this chapter we shall consider these features. Some general techniques to extract them will be discussed. Once we know what mathematical techniques are necessary, we can try to find hardware that is able to extract all these features (though not necessarily simultaneously). The required flexibility has to be obtained by means of modifying (parts of) the software. In chapter 4 we shall try to find a (hardware) system that possesses this flexibility. To find this system, one needs some knowledge about the desired signal processing. A full discussion of these types of signal processing is beyond the scope of this chapter. We merely need to know the typical characteristics of the processing techniques.

* Amplitude

In order to present a user of the tactile aid with some basic information about sounds, we have decided to offer amplitude information in some form to the skin. To compute the amplitude, $A$, of a (sampled) signal, $s_i$, one usually computes the Root-Mean-Square (RMS) value: 

\[ A = \sqrt{\frac{1}{N} \sum_{i=1}^{N} s_i^2} \]
Signal processing

\[ A \equiv \sqrt{E} \quad \text{where} \quad E = \frac{1}{N} \sum_{i=0}^{N-1} s_i^2 \quad (3 - 1) \]

and \( N = \) number of samples

\( E \) is a measure for the energy of the signal. Instead of presenting amplitude information to the skin one can also offer the energy.

* **Fundamental frequency**

The fundamental frequency of a speech signal can be obtained in different ways, e.g. using the autocorrelation function, Linear Predictive Coding (LPC), or Cepstrum [Rabiner & Schafer, 1978]. The technique, described in chapter 6, to find the fundamental frequency, or the frequency that can be perceived as the pitch, is the harmonics sieve [Goldstein, 1973; Duifhuis & Willems, 1987]. For this technique the signal needs to be frequency analyzed, or fourier transformed first.

* **Frequency spectrum**

To obtain the frequency spectrum of a sound, one can use several techniques, see e.g. [Rabiner & Schafer, 1978]. Applying analog techniques, bandpass filters or low-pass filters are used, or a single bandpass filter with variable centre frequency. When applying digital techniques, one can use filters too, for example Finite Impulse Response, FIR, filters or the Discrete Fourier Transform (DFT) can be used. The DFT, \( F (k) \), of a digitized signal \( s_i \), consisting of \( N \) samples, can be written as [Baher, 1990]:

\[
F (k) = \sum_{i=0}^{N-1} s_i w^{-i k} \quad \text{where} \quad w = e^{j 2 \pi / N} \quad \text{and} \quad 0 \leq k < N
\]  

(3 - 2)

Equation (3-2) can be transformed into a Fast Fourier Transform (FFT) easily when \( N \) is a power of 2. The number of computations for an FFT
Chapter 3

is in general less than needed for a DFT [Baher, 1990]. To compute an FFT a special form of addressing to access the samples (and the partially processed samples) is required, called bit-reversed addressing [Baher, 1990; Analog, 1986].

The result from the Fourier transform can be used as the first stage for several signal processing algorithms, which compute other parameters from a signal (such as the fundamental frequency).

* **Formants**

A formant is specified by both its (centre) frequency and its bandwidth (since the formants are a measure for the acoustic filters formed by the vocal tract). We have decided not to present the bandwidths of the formants to the tactile sense. This decision will be discussed at the end of this chapter.

To obtain the formants from a speech signal, several techniques can be used. To estimate the formant frequencies, one can use a number of bandpass filters, with suitable centre frequencies, followed by a circuit to detect the maxima. The bandpass filters that have a local maximum output signal indicate where the formant frequencies are located. The accuracy of this technique depends (among other things) on the number of filters. This technique is applied in e.g. [Leysieffer, 1986].

When digital techniques are used, several other methods are possible too. Two techniques are closely related to each other: one technique uses LPC parameters. From the LPC parameters, it is possible to estimate the formant frequencies [Vogten, 1983].

Another technique, called the Robust Formant Analysis (RFA) [Willems, 1987], is derived from the LPC-analysis method, in such a way that one is ascertained that the required number of formant frequencies is always found, something which is not always certain when the previous method is used [Willems, 1987; Vogten, 1983]. The RFA finds the formant frequencies by computing so-called modified Line Spectrum Pairs (LSP's). The normal LSP's contain information for approximating the formant frequency and the bandwidth. The modified LSP's however contain a better approximation for the formant frequencies, but no information about their bandwidth. The original
Signal processing

RFA computes the bandwidths only after the formant frequencies have been determined [Willems, 1987; Willems, 1988].

Common aspects

We have now seen how we can obtain information about the sound or speech signal. When using digital techniques, one can say that (auto-) correlation, LPC analysis (or derived techniques), Fourier analysis and digital filtering are often used as a first stage or as the major stage when processing signals. A basic mathematical function which is used in these types of processing is the so-called sum of products:

\[ R_{n,m} = \sum_{i=k}^{l} x_i \cdot y_j \]  

where \( j = (n i + m) \)  

(3-3)

where \( k \) and \( l \) are the beginning and the end of a considered interval, \( x_i \) is a variable to be multiplied by \( y_j \), which is either a variable or a constant, and \( n \) and \( m \) are offsets that determine the step size and shift of \( j \) with respect to \( i \).

Let us consider how this sum of products occurs in the different processing techniques.

- When computing the amplitude of a signal, \( x_i = y_i \) are the samples \( s_i \); further \( n = 1, m = 0 \) and \( R / (l - k + 1) \) equals the energy (see (3-1)).
- For the DFT, \( x_i \) are the samples, while \( y{(n i)} \) are the so-called weights \( w^{-i n} \) (see (3-2), where \( n = k \)).
- For a FIR-filter, \( x_i \) are the samples at time \( l \) (i.e. \( x_l \) is the current sample), \( y_i \) are the filter parameters, and \( l - k + 1 \) is the length of the filter. \( R_n \) is the filtered sample at time \( l \).
- For the correlation, \( x_i \) are the samples which have to be multiplied by \( y_{i + m} \), from either the same signal \( x_i = y_i \), shifted in time (auto correlation), or from another signal shifted in time (cross correlation).
Formant analysis

Owing to time limitations, we could use only one type of extracted information for the tests discussed in chapter 7. As to which type of information, we have made the following decision.

In the first place we have chosen to offer information that is specifically useful for supplementing speech-reading. However, the information should not offer information merely about speech; it should also offer at least some information about sounds in general.

When reviewing the tactile aids described in literature, it appeared that the information mostly offered is:
- the sound signal after (band-pass) filtering;
- the frequency spectrum;
- a signal related to the fundamental frequency (or sometimes the fundamental frequency itself).

Only a few tactile aids offer information which consists mainly of the formant frequencies. In these few cases the formant frequencies are mostly extracted by means of a small filterbank, followed by peak detection (e.g. [Leysieffer, 1986]). Literature shows that the frequencies of the first two formants can be a useful supplement for speech-reading [Breeuwer & Plomp, 1985]. Breeuwer however offered the formant information acoustically to normally hearing subjects who had little experience with speech-reading. Offering formant information is not yet used in commercially available tactile aids.

As mentioned before, formants are defined by both their (centre) frequency and their bandwidth. In order to offer formant information tactually, one can follow several strategies. One possibility is to offer only the formant frequencies. Breeuwer offered the formant frequencies coded as pure sinewaves, during the periods that the signal was voiced. Neither the energy of the speech signal nor the bandwidths of the formants were offered [Breeuwer & Plomp, 1985].

Another possibility is to offer both the formant frequency and its bandwidth. However, the tactile sense has a limited channel capacity
Signal processing

(as explained in chapters 2 and 5). Therefore it is doubtful whether this extra information can be perceived properly, and can improve speech-reading scores significantly. Also in literature it is mostly only the frequencies that are offered. Any information about bandwidths is offered in a very rough form. Further studies on this matter will be necessary.

We wanted the device to offer some information about non-speech sounds too. Therefore we decided to offer the overall amplitude of the signal in addition to the two formant frequencies, so the energy of the speech signal within the bandwidth of the detected formants is not offered specifically. The offered formant frequencies are determined using the Robust Formant Analysis method [Willems, 1987]. Chapter 5 (Displays) gives details of how we have offered this information via the tactile display to the skin.
Signal processors

It was explained in chapter 2 that tactile hearing aids need some kind of signal processing to extract the desired features from the signal. Chapter 3 looked into various methods that are available for signal processing. This chapter discusses available components for signal processing. Both analog and digital components will be reviewed. At the end of this chapter we shall choose one of the available components for the tactile aid suggested by this thesis.

Analog processors

Early tactile hearing aids only used analog components. The reason for this was that digital components were either not yet available or were not able to perform real time signal processing. Another problem with digital components used to be that they were rather big and used relatively much energy (too much for battery operation during a couple of hours). This too made them not suitable for portable devices.

Although digital techniques have taken a big flight since the nineteen-sixties, they have not completely replaced analog processing techniques. Further, analog components are necessary for interfacing between the analog world and the digital processing.

Also it is now possible to combine analog and digital processing techniques on one chip [Intel, 1988]. Certain types of signal processing (for example filtering) are often much easier to accomplish with analog components, while other types of processing are easier using digital techniques (e.g. Fourier analysis).
Pure analog processors

In the existing tactile hearing aids, analog components are often used to build low-pass or band-pass filters. Low-pass filtering can for example be used to obtain a signal in which the fundamental frequency shows clearly. Band-pass filtering e.g. can be used to create a filter bank for frequency analysis. When the output of a filterbank is processed further, it is possible to extract information closely related to formant frequencies.

Also the analog computers, often used for process controlling functions, could be configured using filters, integrators and differentiators and (variable) resistors. Using such analog computers it is possible to process a signal according to a defined differential equation [Heetman, 1964; Johnson, 1963].

Although it is possible to use analog processors for the tactile hearing aid, the use of such processors has at least one major disadvantage; it is not possible to modify the processing algorithm easily. It takes a modification of the hardware wiring to modify the algorithm.

Combined analog/digital processors (hybrid processors)

Combining analog components with digital components to process the signal is used in a lot of earlier tactile hearing aids [Kirman, 1982], but this technique is still also applied nowadays [Yeung et.al, 1988]. The analog part can be used very well as e.g. low-pass or band-pass filter. The digital part is then used, either to process the signal further, e.g. zero crossing detection, or to prepare the extracted information for presentation on the skin. In those cases, the tactile display is usually a binary driven multiple-channel display. This means that a stimulator is either on or off [Sherrick, 1984; Risberg, 1983].

One example of an IC that combines analog and digital techniques and that can be used for speech processing is the NEC μPD7763 [NEC, 1986]. This component combines a 16-channel band-pass filter with an analog to digital converter. In combination with a microprocessor,
it becomes possible to perform further analysis on the frequency analyzed data.

Although hybrid processors are more flexible than analog processors, it is still not possible to change a part of the processing algorithm without modification of the hardware. The NEC chip for example only offers frequency analysis as preprocessing. When another analysis method is needed, this is not possible without exchanging the components.

**Neural network computers**

Neural networks [Lippmann, 1987; Souček & Souček, 1988; Aleksander & Morton, 1990], Neural Computers or Neural Network Computers are systems that have their foundation in models of the human neuron and brain. By using a large number of small systems, called neurons, after the biological neuron, that are interconnected on a large scale, varying from total interconnexion to interconnected in groups of neurons, one can make a system that is able to learn and classify.

The computational units—the neurons—are normally analog [Aleksander & Morton, 1990; Hudson, 1990], having an analog signal (voltage) as input, and an analog output as well. A model of a neuron can be seen in figure 4-1. Since neural networks are essentially analog, they are mentioned in this thesis after the analog computers and before the digital processors.

Opposed to digital computers, neural networks do not have memory units (such as RAM’s or ROM’s) for storing information. The only memory function that can be found are the weight-factors, which determine the influence of one neuron on another neuron. Information in a neural network is stored in these weight-factors, all over the network.

It is not possible to program a neural network, in the way digital computers are programmed. Where a digital computer has its program, consisting of a finite number of limited instructions, stored in a program memory, and executes the program mostly serially, the neural network has its ‘program’ hidden in the weight-factors and the struc-
Signal processors

Fig 4-1: Schematic model of a neuron
(after [Aleksander & Morton, 1990] and [Hudson, 1990])

For this reason it is practically impossible to program such a network in the traditional way. To let a neural network perform a desired operation, e.g. a recognition task, it has to be trained. By offering the network a number of input signals and the desired matching output signals, the network gradually learns that the output signals should correspond to a certain set of input signals. When a network is sufficiently trained it can even determine the correct output for an input signal that was until that moment unknown to the network.

Tasks that are very difficult to perform for a conventional digital computer, because they require a great number of operations that can only be performed sequentially, such as pattern or sound recognition, can sometimes be carried out much better and faster by a neural network.

Although the possibilities of neural networks seem promising, they do have certain disadvantages which might be solved in the future. One of these disadvantages is that large neural networks, working with
analog neurons, as would be required for a speech recognition task, are not yet commercially available. Only smaller networks are beginning to become available nowadays, while most research on these networks is done by simulating them on a digital computer [Aleksander & Morton, 1990].

As soon as large neural networks become available and indeed perform as they are believed to be able to, many new devices using recognition will become available. For the time being we will have to use the conventional analog and digital techniques.

**Digital processors.**

Since roughly the middle of the nineteen-sixties, digital components have been becoming available for commercial devices. The first single chip microprocessors (or microcomputers) appeared around the first half of the nineteen-seventies. With the appearance of the microprocessor, speech recognition devices also started to appear.

Tactile hearing aids started to include digital components for the processing of signals. One kind of signal processing that is often used is pitch extraction by means of determining the zero-crossings of a speech signal [Rabiner & Schafer, 1978]. But also more sophisticated processing methods became possible, such as LPC analysis, which offered information about the shape of the vocal tract, and more reliable pitch extractors, e.g. the Kanievsky device [Sensoraid, 1986].

Nowadays several kinds of digital processors can be used for signal processing. In chapter 2 we were introduced to several of those processors:

- standard microprocessors (μP) with a von Neumann architecture;
- reduced Instruction Set Computers (RISC);
- special Digital Signal Processors (DSP) with a Harvard architecture.

We can further add the following components to the list:

- microcontrollers (μC) which are in fact μP’s with several special functions (such as I/O, timers, memory, etc.) on chip;
- discrete processors. These are digital circuits made of a number of components to form a dedicated processor circuit;
Signal processors

- special-purpose processors. These can perform only one type of processing. A frequently found example is the processor that controls the keyboard in Personal Computers (P.C.'s). An example of a special-purpose processor for signal processing is the SP-1000, an LPC processor.

Choice of the signal processor

It was explained in chapter 2 that the tactile hearing aid as used for our experiments must be able to perform various processing techniques. Modifying the processing should be simple (to enable comparison of different signal analysis algorithms) and—if possible—without the necessity of hardware modifications. Table 4-1 shows the different processor techniques (analog, neural, digital) that are in principle available and their suitability for the tactile hearing aid.

In the above sections we have seen that modification of the processing technique or algorithm is rather difficult with analog processors. These processors are also often difficult or not available as CMOS (or equivalent) components, which means that analog components usually consume too much energy (for our purpose). This leaves us only the neural networks and the digital processors. Since no suitable neural networks are available yet, only digital processors satisfy our conditions (also taking into account other considerations than those mentioned here: see table 4-1).

It will be clear that from all the available digital processors, only the programmable processors can be considered useful for the tactile aid. We have to choose between a μP, a μC, a RISC and a DSP. The processor we shall use must consume as little energy as possible and yet have enough processing capacity. Owing to a phenomenon known as the Speed/Power Product (SPP) these are more or less conflicting conditions. The faster a processor (or in general a digital circuit) must operate, the more energy it needs. When using CMOS components, the speed/power ratio (that is, the speed—or clock frequency—divided by the required electrical current) is almost a constant [Philips, 1986]. In other words, we want a processor that uses as few clock cycles as possible per instruction, with instructions that are very powerful.
Further, to keep the circuitry as compact as possible, we want a processor that does not require many external components. On the other hand, to develop a system that might need some hardware changes the processor should not have too many components internally. Otherwise, when a small change might be needed in the system hardware, one would have to use a different processor with the required internal components. This again would mean that the system would have to be almost completely re-designed.

Table 4-1. Suitability of the different processor techniques for the experimental tactile hearing aid.

<table>
<thead>
<tr>
<th></th>
<th>Analog processors</th>
<th>Neural networks</th>
<th>Digital processors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexibility</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Power consumption</td>
<td>-/+</td>
<td>?</td>
<td>-/+</td>
</tr>
<tr>
<td>Size</td>
<td>-/+</td>
<td>? (-)</td>
<td>+/-</td>
</tr>
<tr>
<td>Availability</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Programming</td>
<td>- (adjusting)</td>
<td>- (learning)</td>
<td>+</td>
</tr>
<tr>
<td>Required components</td>
<td>-/+</td>
<td>?/-</td>
<td>+/-</td>
</tr>
</tbody>
</table>

*used symbols:*
- = less suitable
-/+ = generally less suitable
+/- = generally suitable
+ = suitable
? = unknown
Signal processors

Microcontrollers are in fact microprocessors combined with I/O ports, internal memory, and a combination of other devices, such as serial ports, timers, interrupt handlers, analog to digital converters and so forth. Since we want a processor uncluttered by so many special components, a µC is less attractive in use as an experimental system. Internal components might be useful, however, when the system becomes less experimental. In a system that needs no hardware and only now and then small software changes, internal components save on the total number of discrete components. This influences the size of the system directly: the fewer components, the smaller the system can be. Yet, at present we shall not use a microcontroller.

Microprocessors usually have a von Neumann architecture, which means that program code and data are accessed using a common data and address bus so the processor cannot simultaneously react to fetch a new instruction and access data.

On the other hand, most processors—especially the more recent ones—allow for a large instruction set; a lot of manipulations on data and conditional branch instructions are possible. This large instruction set also has its adverse effects. Since the instructions are often quite complicated, the processor needs some time to complete an instruction. The instruction frequency is often up to 15 times below the clock frequency at which the processor operates [Intel, 1988]. Recent processors have solved this problem partially by using a higher integration scale and pre-processing of the instruction codes (pipelining). The use of a cache-memory (a small but fast on-chip memory, containing the most recent instructions) can also help to speed up processing, since it can save instruction fetches during program loops.

One of the solutions to the long instruction time of µP’s is the RISC processor. This processor has a Reduced Instruction Set, which shortens the instruction time. RISC processors usually need only one or in special cases two clock cycles to complete an instruction. Although the instructions are less complex, it is possible to perform a number of instructions in the time in which a µP has completed only one.

For the field of signal processing, a special processor has become available, the Digital Signal Processor (DSP). This processor can be
seen as a mixture of a normal µP and a RISC processor, with some special hardware adaptations. The hardware of a DSP has a so-called Harvard structure. This means that program code and data are stored in separate memories. The program and the data memory have separate data and address buses, which means that the processor can fetch a new instruction at the same time that it accesses data. However, some data can be stored in the program memory too (see below).

To speed up the program, a DSP has a so-called cache memory [Analog, 1986], containing the most recent instruction codes. When performing a program loop, the processor can fetch its instructions from the cache memory, thus by-passing the program memory. It is therefore possible to access data from the program memory without disturbing the program run. This combined function possibility is one of the very strong points of a DSP.

In chapter 3 we saw that one of the expressions often used in digital signal processing is a repeated summation of products:

\[ R_{n,m} = \sum_{i=k}^{l} x_i \cdot y_j \quad \text{where} \ j = (n_i + m) \]  

(4 - 1)

where \( x_i \) and \( y_j \) are the variables (or constants for a DFT and FIR filter) to be processed, \( n \) a multiplication factor, as used at DFT’s, and \( m \) an offset, as used in correlation functions. The summation runs from \( k \) to \( l \). Most DSP’s have special instructions for this type of summation. As an example, we take at the ADSP-2100. In its assembly language, the summation translates to the program of figure 4-2.

In this example the memory pointers are initialized first. I0 and I4 are pointers to the variables to be processed. These pointers also contain the starting value \( k \) and offset \( m \) from equation (4-1). M0 and M4 are the incremental constants. Using equation (4-1), M0 equals 1, and M4 equals \( n \). A counter (CNTR) is loaded with the number of iterations. After this initialization, the summation loop is started. The special architecture of the DSP takes care of the counting down, without using
processing time. In other words, only the line labeled 'summation' will be repeated for \( l - k + 1 \) times. This single instruction multiplies \( x \) with \( y \) (\( MX0 \) and \( MY0 \)), then adds the result to the previous result (the summation). At the same time it loads new values for \( x \) and \( y \) from respectively the data and the program memory.

Most DSP’s also have other special instructions that are typically used in the field of signal processing, such as bit-reversed addressing, for FFT’s. Further, the average DSP has few special components, such as I/O ports, on chip. Only internal program and/or data memory can be found on a number of DSP’s. This is done to make the execution time shorter.

Choice

Knowing the differences between the various processors, we can now choose which processor we want to use (see also table 4-2). A \( \mu C \) will not be used, because the hardware is insufficiently flexible for our experimental system. The \( \mu P \) will also not be used. Although it will be possible to find a \( \mu P \) that is fast enough to perform real-time signal processing, such a processor will have to operate at a relatively high
Table 4-2. Suitability of the different digital processors for the experimental tactile hearing aid.

<table>
<thead>
<tr>
<th></th>
<th>µP</th>
<th>µC</th>
<th>RISC</th>
<th>DSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>internal components</td>
<td>few to moderate (ALU,</td>
<td>many (e.g. I/O, timers,</td>
<td>few to moderate (ALU’s,</td>
<td>moderate (ALU’s,</td>
</tr>
<tr>
<td>(without taking the number of transistors into consideration)</td>
<td>registers)</td>
<td>memory)</td>
<td>registers)</td>
<td>registers, (cache-) memory)</td>
</tr>
<tr>
<td>external components</td>
<td>memory, I/O, etc.</td>
<td>none or few.</td>
<td>memory, I/O, etc.</td>
<td>memory, I/O, etc.</td>
</tr>
<tr>
<td>required</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>instruction set</td>
<td>usually many, complex</td>
<td>usually many, complex</td>
<td>small set with single cycle</td>
<td>large set, complex, single</td>
</tr>
<tr>
<td></td>
<td>instructions</td>
<td>instructions</td>
<td>instr.</td>
<td>cycle instr.</td>
</tr>
<tr>
<td>architecture</td>
<td>von Neuman</td>
<td>von Neuman</td>
<td>von Neuman/Harvard</td>
<td>Harvard</td>
</tr>
<tr>
<td>especially suitable</td>
<td>general purpose</td>
<td>general purpose, control</td>
<td>general purpose, fast.</td>
<td>fast, signal processing</td>
</tr>
<tr>
<td>for:</td>
<td></td>
<td>functions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

clock frequency. This means that a µP will use relatively more power than a RISC or DSP.

The RISC is also excluded. RISC processors are fast enough for real-time signal processing; however, the RISC processors at the time of our choice had less powerful signal-processing instructions than the available DSP’s. This would mean that a RISC processor would have

- 43 -
Signal processors

to process more instructions than a DSP for the same signal-processing tasks, and would thus need a higher clock frequency.

So, our choice, made in the middle of 1987, was to use a Digital Signal Processor for our tasks.

Which DSP?

We then had to decide which DSP we should use. A number of DSP’s have been reviewed. It would go beyond this thesis to enter into details about what each DSP could and could not. Taking into account

- the accessible program and data memory,
- the internal memory,
- the maximal instruction speed
- the time required to perform e.g. a 1024-point FFT,
- the availability, both at that moment and during the next couple of years,
- the available software developing tools,
- the price (since it was an experimental system) of both the DSP and the required software, such as assemblers, simulators, etc.,
- future developments of the DSP (upward compatible types), and
- some other factors

we have chosen for a DSP from Analog Devices, the ADSP-2100 [Analog, 1986]. This processor has a relatively large external program and data memory that can be built up from regularly available components. Also a newer version of this DSP was announced that had program ROM, data memory and some other functions on the chip, which feature was not necessary for the experimental system, but would probably be very welcome for a more definite system. Finally the assembly language of the DSP is quite easy to read, which is of advantage in our educational situation. The chosen DSP will be discussed further in the next chapter.
In the previous chapter we discussed which information can be extracted and how this is performed. This chapter describes how the extracted information can be presented to the hearing impaired. Although the choice to use the tactile sense was in fact made before we started the study described in this thesis, it will be useful to review the available senses, and their suitability for transmitting information about sounds. This chapter concludes with possible ways of presenting the information to the tactile sense.

Perceiving the information

The majority of people can perceive acoustic information aurally. The system we have developed is meant to be used by people who cannot perceive sounds by the hearing organs, either directly or after amplification. For this reason another perception channel will have to transmit the information after it has been processed. We possess the following senses via their respective sense organs:

- the acoustic sense (the ear);
- the olfactory sense (the nose);
- the taste sense (the tongue);
- the visual sense (the eye);
- the tactual sense (the skin).

Although it might appear logical to exclude the first three senses in this list as a candidate for a hearing aid for the profoundly deaf, for reasons of completeness I would still like to discuss all five.
Displays

The ear

Although we are talking about a communication aid for the profoundly deaf, this does not necessarily mean that the ear cannot be used to transfer information. In chapter 2 we defined ‘deaf’ as ‘having a hearing loss of at least 90dB’, in other words ‘hearing so little, that it is not possible to understand normally spoken speech, even using conventional hearing aids, without another form of information transfer (such as speech-reading)’. This means that there are ‘deaf’ people who still hear a little. When talking about profoundly deaf, we mean deaf people who cannot hear anything at all.

When only a limited amount of information is offered to the ear, it is possible that these people can discriminate between the offered sounds. Some hearing aids have used this fact and perform data extraction comparable with existing tactile aids. One hearing aid, used for speech training, extracts the fundamental frequency and offers this to the ear (if necessary after frequency transformation of $F_0$) [Pickett, 1982].

However, the profoundly deaf will not be able to perceive any sounds by the ear. The only thing that might be perceived when acoustic information is presented loudly enough, is tactile information: one feels the sound on the skin, instead of hearing it via the ear.

To discover what information can be offered best to supplement speech-reading, Breeuwer [Breeuwer & Plomp, 1984, 1985, 1986] used hearing subjects, and offered speech information acoustically coded. Since perceiving acoustic information is something that hearing persons do all the time, the recognition of speech information which is coded acoustically (as a speech-like sound) appears to be learned very quickly. The subject usually has some ability to interpret the data correctly within a few minutes.

For our study we wanted to work with deaf subjects too. Therefore, presenting the acoustic information in an acoustically coded fashion, was not possible. For our research we had to use an alternative channel.
Chapter 5

The nose / the tongue

Both the nose and tongue are sense-organs, and might be used—in theory—to transfer information, even other than the ‘adequate’ information. However, since the information transfer rate of both senses is relatively low, and since artificially creating correct (adequate) stimuli for both senses still proves quite difficult, we shall pay no further attention to these organs.

The eye

Using the visual sense, it is possible to transfer information at a relatively high speed as in a film. Also linguistic information can usually be transferred at a speed higher than that of speech. Written messages can be read faster when they are not read aloud. Think for example of subtitles on a film: even when the subtitles contain literally the same information as the spoken message, it is for most people quite easy to stay ahead of the spoken message.

The fact that the visual system is a fast system, seems to offer good perspectives for using this sense as alternative channel for information about speech or sounds.

There are many methods for coding certain acoustic information into visual information. For the easier forms of information extraction, several visual codings are known in every-day life. Coding amplitude information is done in e.g. VU-meters, either analog or digital. Coding frequency information can be found in frequency analyzers as used in audio studios or the more highly developed audio equipment.

If one wants to present visual supplementary information for speech-reading, such as formant or pitch information, one has to display this in such a way that both the face of the speaker and the visual supplementary information can be seen clearly at the same time. One method to accomplish this would be to project the visual information on or near the speaker. For practical purposes this will be very difficult in situations other than laboratory conditions. Perhaps holographic
Displays

techniques might offer a solution in the future, when it appears to be possible to project images free of the user.

Another method of presenting visual information is to project it onto spectacles that the user would wear. The problem here is that the information cannot be offered as a sharp image (it is projected merely a few centimeters from the eye), which might result in images that can be perceived only with difficulty. A method which has been tried in practice [Pickett, 1982; Pickett, 1987b] is to project the image into the eye. Using a small projector mounted on the eyeglass frame, and a small reflecting area on one of the glasses, one can project the image directly onto the retina. This way the user can see the information at a certain distance. When the user looks at a speaker, the image can be seen close to the face of the speaker. These aids however never passed the experimental stage.

Although the visual sense might seem to be an ideal choice as an alternative channel (since it can perceive information rapidly), we have paid no further attention to it. Further, speech-reading is very strenuous and one can easily be distracted by other visual information. Presenting visual information that can be perceived simultaneously with speech-reading without too much difficulty for the user presents feasibility problems.

When automatic speech recognition (or phoneme recognition) of speaker-independent running speech becomes available, it is possible to present this information as written text. This would have the great advantage that hardly any training is needed for perceiving this information correctly (assuming the subject is able to read). The major difference with the previous ideas is that this kind of information is not necessarily supplementary to speech-reading, but it can be used instead of it. However, this will probably not be the case for a long time to come.

The skin (i.e. the tactile sense)

We have eliminated four of the five senses for use as an alternative channel. The only sense that is left is the tactile sense. Here we shall discuss if and how the tactile sense can be used.
It is known that the tactile sense can be used to transfer symbolic information, such as braille [Foulke, 1982]. Braille uses characters coded into a matrix of 6 or 8 dots that can be felt as reliefs on a sheet of paper. When written text is coded this way, it allows well-trained visually impaired people to read these texts. Figure 5-1 shows a visual representation of a short text in braille, using the 6 dot matrices. Experienced braille readers can reach speeds of about 100 words per minute, which is approximately half of normal speaking rates [Foulke, 1982].

A device that at a first glance seems to be derived from the braille principle is the Optacon, which allows well-trained visually impaired people to read normally printed text [Linvill & Bliss, 1966; Bliss et al., 1970]. Using a small camera, the Optacon translates the dark and light areas within the viewed window to a matrix of 6 times 24 rods, that either vibrate (for dark areas) or do not vibrate (for the light areas or vice versa). On the matrix one can feel the shape of the pattern under the camera. Figure 5-2 shows schematically how this principle works.
works. The letter ‘a’ is scanned under the Optacon camera, and appears on the tactile display as a number of rods that vibrate.

Usually reading rates with the Optacon are only about half the rates found with braille readers. Certain individuals can reach the same speed as braille readers, of about 100 words per minute [Foulke, 1982; Craig & Sherrick, 1982]. This gives an indication of the information transfer rate (or channel capacity) of the tactile sense for this type of information. It appears to be possible to offer usable tactual information –if properly chosen and after the proper training– at about at least half the speed of spoken language [Craig & Sherrick, 1982].

Tactile displays have not only been developed for the visually impaired. Also for the auditory impaired, various methods are known that use the tactile sense for presenting information about speech (or sound in general). In chapter 2 we noted the first experiments of Gault, who presented speech data on the hand using a long tube. Another method that is in use for the deaf-blind is the so-called Tadoma method [Alcorn, 1932; Norton et al., 1977]. This method, called after the first children using it –Tad and Oma– consists of placing the hand on the mouth, nose and throat of the speaker. The deaf-blind person then feels the movement of the jaws and lips, together with the airflow and the vibration of the larynx.
Other examples of tactile hearing aids also show that information can
be transferred using the tactile sense. Various sites on the body, various
types of displays and various ways of stimulation have been tried
[Risberg, 1983; Lemeer, 1990; Leysieffer, 1986]; however a proper
investigation into what kind of display to place and where to place it
has not been found in literature.

What can be found in literature, is that the tactile sense can be used
reasonably well as an alternative channel for presenting information
derived from the speech signal. In the next section we shall look at
what kinds of display can be used, and which area of the skin is most
sensitive for the tactile presentation of data. Also we shall discuss
which areas might be useful, but will not be used, because of cosmetic
or other social and psychological reasons.

Tactile displays

[Lemeer, 1990] refers to several tactile displays that have been de-
scribed in literature. The main distinction is between single-channel
devices and multiple-channel devices. What type of display is used is
usually closely related to the signal processing and analysis used. For
example the very first tactile aids employed a single-channel vibrator,
mainly because it was still quite difficult to extract the proper signals
for driving a multiple-channel display.

When more sophisticated processing and analysis became possible, the
first multiple-channel displays appeared [Risberg, 1983; Sherrick,
1984], using different coding strategies. We take the filterbank analysis
as an example to illustrate several coding strategies, such as could be
implemented on an Optacon or a comparable tactile display. The output
of a filterbank consists of a number of energy values for the available
frequency bands (see figure 5-3).

(a) Conventional coding of the result of filterbank analysis, similar to
the use in present audio equipment (where a visual instead of a
tactile display is used). Every frequency band has a number of
vibrators. The energy contained within a frequency band is directly
proportional to the number of vibrators that are operational. Figure
Fig 5-3: Examples of different multiple-channel displays. A = Amplitude (or Energy), f = frequency, t = time.
5-3a illustrates how this display might function when a signal is offered.

(b) Here only the topmost rod of every column vibrates, to indicate the energy level. This technique can be compared with the settings of an equalizer in audio equipment.

(c) Here we have a one-dimensional display, instead of the two-dimensional display of the previous methods. Frequency is still encoded along the array, but the energy in a frequency band is coded as the vibration energy of each stimulator (using amplitude modulation). [Franklin, 1988]

(d) This is again a two-dimensional array. However, the momentary information is offered in a single dimension, as in the previous display. The other dimension is used as history. The information from moment $t - 1$ is shifted on moment $t$ to the next row, and so forth, until it is displayed in the last row (the bottom row), where it leaves the display when new data become available.

The most common signal-processing algorithms for tactile hearing aids offer information that can be expressed as frequency and power. This information can be coded in a way comparable to the filterbank analysis method.

**Stimulating the skin**

In the previous sections we have seen that we can use the tactile sense for offering information extracted from sound. In this section we shall review how the skin can be stimulated. We shall look at both adequate and inadequate stimuli. First we shall look briefly at the upper layer of the skin with its senses.

Figure 5-4 shows schematically which neural structures can be found in the human skin [Becker et al., 1971; Bernards & Bouman, 1974]. Several receptors are present. Some are sensitive to pressure (such as the Pacinian corpuscles and Meissner’s corpuscles), some perceive warmth or cold (Rufinian endings and Krause’s end bulbs), and some are sensitive to light touch (Peritrichal network). There are also free nerve-endings, subserving the pain sense. Depending on the type of stimulus used, one or more of these structures will be activated.
Fig 5-4 Sensory receptors in the skin (Schematic, after [Becker et al., 1971] and [Bernards & Bouman, 1974])

Literature shows two possible ways of stimulating the skin in order to transfer information as obtained by a tactile aid [Sparks et al., 1978; Leysieffer, 1986]:
- mechanical vibration;
- electrical current.

The human skin has no receptors specifically for perceiving an externally applied electrical current. For this reason this type of stimulation is called inadequate [Sherrick & Cholewiak, 1986]. Depending on the strength of the current, various receptors can be activated, resulting in perception corresponding to those receptors.

At the end of this section we shall look briefly at a third possible method for stimulating the skin using inadequate stimuli, not yet applied in tactile aids, namely stimulation by means of LASER light.
Chapter 5

The sense of touch

An adequate method to stimulate the skin is by means of a mechanical stimulus. When applying such stimuli, one encounters certain properties of the tactile sense. The most important are listed below [Sherrick & Cholewiak, 1986; Loomis & Lederman, 1986].

• Thresholds for touch: the skin is not uniformly sensitive to stimuli. The absolute threshold for pressure or vibration varies over the body. Places which are most sensitive to pressure are e.g. the nose, the lips, the cheeks and the forehead (requiring forces equivalent to about 0.5 to 1.5 mg [Weinstein, 1968]. The index finger (which is usually used for the Optacon display) has a threshold of about 11 1/2 mg.

• Threshold for vibrations as a function of frequency: the threshold for vibrations depends on the frequency applied. A minimum amplitude is required for frequencies between about 200 and 400 Hz. Both at higher and lower frequencies the tactile sense is less sensitive so higher amplitudes are needed to generate a sensation. Where vibrations of about 0.2 μm can be perceived at 200 Hz, vibrations of about 5 to 10 μm are needed at frequencies of about 40 Hz. Thresholds for vibration have little correlation with thresholds for pressure. The fingers, which are less sensitive to pressure, are very sensitive to vibration: at about 200 Hz, vibrations of about 0.2 μm can be perceived [Weinstein, 1968]. Other places which are relatively sensitive for vibrations are e.g. the forearm and the sole of the foot (about 0.4 μm). The thigh has a threshold of about 1.8 μm (all thresholds at about 200 Hz) [Sherrick & Cholewiak, 1986; Weinstein, 1968].

• Spatial resolution: just as the threshold for touch varies over the body, the spatial resolution varies too. Resolutions vary from about 3 mm (for the fingers) to about 45 mm for the upper arm and calf. When two stimuli are applied simultaneously, they can only then be discerned when the distance between them is above the spatial resolution. Otherwise one perceives the two stimuli as one stimulus where the amplitude is about the sum of the two amplitudes.

• Spatial masking: two simultaneous stimuli not only interfere when they are applied close together. Also when the distance between them is greater than the spatial threshold, e.g. one stimulus at the index finger and the other stimulus at the little finger of the same
hand, interference can occur. Although both stimuli can be perceived separately, the perceived intensity can be different from the perceived intensity when only one stimulus is offered. Also differences in intensity are perceived less well when two stimuli are offered simultaneously.

- Temporal resolution: in order to discriminate between two stimulation impulses (of 1 msec), the pulses should be separated by about 10 msec in time. Otherwise the stimuli are perceived as a single stimulus.

- Pressure adaptation: when the tactile sense is exposed to a static pressure, this is felt best at the moment when the pressure is first applied. After that the sensation gradually disappears. The speed of adaptation depends on the intensity and the stimulated area. Owing to this adaptation, it is not recommended to offer the supplementary information as (more or less) static patterns. One needs to use vibrating excitators.

- Vibrotactile adaptation: not only static pressures are subject to adaptation, periodic pressures show this behaviour too. It appears that the threshold for vibrations increases with about 5 dB after one minute of continuous exposure to a vibration which is 34 dB above the (starting) threshold to about 17 dB after 20 minutes.

- Temporal masking: this effect occurs when a second stimulus is applied a short time after the first. It appears that both stimuli can interfere with each other when the time between both stimuli is shorter than about 200 msec [Sherrick & Cholewiak, 1986; Wang, 1990].

- Amplitude discrimination: the difference threshold for the amplitude of vibrotactile stimuli is around 1.5 dB for vibrations of 160 Hz: the amplitude has to change by about 20% in order to be perceived as a different amplitude.

- Frequency discrimination: the frequency discrimination of the skin is poor. Around 300 Hz, reported noticeable differences are 60 to 75 Hz. Instead of perceiving a different frequency, the skin often perceives a different magnitude of vibration when the frequency is changed.
Mechanical stimulation versus electrical stimulation

It was stated in the previous section that the fingers are most sensitive to vibrotactile stimulation, while their spatial resolution is high too. Therefore one could conclude that the fingers are the best place for receiving tactile information. Tactile aids for the auditory impaired usually use another site on the skin. Tactile hearing aids are aids that need to be used all day long, so it would be bothersome to interfere with the use of one or both hands. Also, a number of deaf people use their hands to communicate or to support communication. When one or both hands are required for a tactile display, this function would no longer be available. For this reason most tactile aids present their information on other parts of the skin. The abdomen, the sternum the back and the lower arm or wrist are favourite sites [Risberg, 1983;

\[\text{Fig 5-5: Schematic representation of electrodes and electric field}\]
Displays

Sherrick, 1984; Franklin, 1988; Lemeer, 1990]. However, some (ex­
peridential) aids also use the hand or fingers [Leysieffer, 1986].

Since electro-mechanical vibrators tend to use relatively much energy
—usually too much for battery-operation during several hours of use—
electrical stimulation is sometimes applied. Displays that operate this
way drive a small current through the skin, thus enervating various
senses. The senses in the top layer of the skin especially are very
sensitive to these currents, since the current density is highest in the
outer region of the skin. Figure 5-5 illustrates schematically how the
current flows through the skin, assuming that so-called concentric
electrodes are used, i.e. electrodes, surrounded by an annular ground­
ing electrode. When the current is kept low (about 0.25 to 1 mA) the
sensation generated is usually perceived as a vibrotactile stimulation
[Sherrick & Cholewiak, 1986].

When electrodes are used they should be current-driven, because the
electrode-skin impedance can vary considerably, and has a non-linear
behaviour. Consequently a much higher voltage is sometimes needed
to start the required current through the skin; once the current is
flowing, the voltage should be much lower.

So, considering merely the energy consumption, one would choose for
electrical stimulation, although the required electronics might be rather
complex. However, when the physiological effects are considered, the
advantage of electrical stimulation is not so clear. Unfortunately, not
much is known about the preferences of users of a tactile aid for either
stimulation method. No comparative studies of long-term use of both
methods could be found in the literature.

The frequency range of vibrotactile stimulation is a little smaller than
the range of electrotactile stimulation. The application of the excitators
to the skin is easier with vibrotactile stimulators than with electrotactile
stimulators. The latter need a low electrode-skin impedance. When the
impedance is too high, the voltage required to generate a sensation is
too high either for the electronic circuit (resulting in currents that are
too low and cannot be felt) or for the skin: burning spots might be the
result.
Finally we have to consider the long-term effects of both methods of stimulation. The problem is that very little can be found in literature about this effect [van Eijnsbergen, 1990]. Most information can be obtained from people working with these methods of stimulation, such as physiotherapists. It appears that both stimulation methods have comparable long-term effects. For example, both mechanical and electrical stimulation can cause callosity where the skin is stimulated. Also the sensitivity for the stimulation can increase, as a result of learning effects (in the brain) and possible changes in the connexions of neurons in the sense organs.

The only known long-term adverse effect that can occur, specifically with electrotactile stimulation, is the migration of ions from the electrodes into the skin. Especially when the electrodes are not made of the correct materials, and when the mean value of the electric stimulation current is not equal to zero, this effect is likely to occur [Sherrick & Cholewiak, 1986]. When too many ions from the electrodes or the gel have entered the skin, irritation or other adverse effects can occur. When the mean value of the current is equal to zero, the migrated ions can remigrate into the electrode or in the conducting gel. Owing to the slow speed at which the ions move, they will hardly have a chance to enter the human skin when an alternating current (with no D.C. component) is applied.

Even with shorter periods of use irritation can occur since the skin is constantly stimulated. Also the conducting gel, used with electrostimulation, can cause irritation or allergic reactions.

Although little research into the long-term effects of either method of stimulation has been done, researchers usually prefer to use vibrotactile stimulation. The thought behind this is that vibrotactile stimulation is an adequate stimulus, whereas electrotactile stimulation is an inadequate stimulus. The latter might result in adverse effects in the long run such as (minor) changes in or damages to the nervous system.

Considering all the known effects of both stimulation methods (and keeping the uncertain factors in mind), we have chosen to use a vibrotactile display for our system. Once the long-term effects of both stimulation methods are fully known, and the positive effect of the
Displays

tactile hearing aid on speech-reading is clear, it may be worthwhile to re-evaluate the different stimulation methods. The problem concerning power consumption of the vibrotactile stimulators can be partially solved by using modern piezo-electric stimulators.

Laser stimulation

A further possibility for skin-stimulation that might also be useful for presenting rapidly changing patterns, is by means of LASER-light. This stimulus is (just as the electrical stimulus) an inadequate stimulus. Literature shows that it is possible to generate small pain effects – for example as used in Evoked Potential measurements [Kakigi et al., 1989] – by stimulating the skin with short laser pulses with sufficiently high intensity [Mor & Carmon, 1975; Bromm & Treede, 1984; Bromm & Treede, 1987]. The duration of the laser pulses should be very short indeed since with longer pulses we run the risk of causing burning. There are, however, no adverse side effects recorded.

To generate a sensation on the skin, no special measures, such as conducting gel are necessary. Direct contact with the skin is not necessary when the laser beams are narrow enough. This prevents the irritation that could be caused by electro- or vibrotactile displays.

Since this method of stimulation appears to have a lot in common with electrostimulation, it might be worthwhile to investigate further into the possibilities of laser light. Because modern laser diodes can be made very small and very little energy is needed to drive the laser diodes, this type of display seems to offer especially good perspectives for the future. However, as laser light has not yet been used for presenting patterns to the skin, it could not be found in the available literature whether such patterns can be perceived effectively by the tactile sense.

In short, it can at present be suggested that the use of laser light offers the advantages of both electrical and mechanical stimulation without most of the disadvantages, but further studies are necessary.
The display used for the experiments

Since it is not that easy to design a new vibrotactile display, we started by using an existing display, slightly modified so it could be connected to the DSP-system. We chose an Optacon, with the camera replaced by an interface circuit\(^2\). The stimulus offered by the Optacon has been well tested in other experiments [Craig & Sherrick, 1982; Sherrick & Craig, 1982]. Furthermore, the Optacon has been used in other experimental tactile hearing aids and in measurements concerning the recognition of processed sounds offered to the skin [Kirman, 1982].

\(^2\) The Optacon we have used was a gift from Fa. Tieman b.v., Moolhoek 5, Rockanje, Netherlands
Displays

A new vibrotactile display

Once it was clear that the tactile hearing aid provided useful information to a number of both deaf and hearing subjects, we started developing a new tactile display [Lemeer, 1990] that could present data on other parts of the skin. As alternatives for the fingertip, we have chosen for the lower arm and the upper leg. As vibrating elements so-called bimorphs are used. These bimorphs are made of piezo-electric ceramic material. When a voltage is applied to this element, it will start to bend. The degree of bending is in direct relation to the applied voltage. Owing to the nature of piezo-electric materials (the cross-effects) bending of the bimorph generates an electric voltage across the bimorph. This effect is not used in the display.

Although the Optacon had an array of 6 times 24 rods, the new display has fewer vibrators. To evaluate this novel display, we considered an array of 4 times 6 vibrators enough; it enables a two-dimensional presentation of the data, while the mechanical and electrical construction is kept relatively simple. Results of this display will be discussed in chapter 7. The construction of the new display can be found in appendix 3 and in [Lemeer, 1990].

Coding strategies

In the section ‘Tactile displays’ we have seen some forms of coding for the extracted data. Using the filterbank data, four different displays were presented. The question arises, which type of coding gives the highest recognition scores. Unfortunately literature gives little information about this. The little information that is available deals mostly with filterbank analysis. It appears that a contour display (fig. 5-3b) gives better results than the filled display (fig. 5-3a) [Green et al., 1983], probably because of the effects of spatial masking.

When a filled array is used, it is easier to perceive the energy in a frequency-band (more vibrating stimulators give the impression that the column displaying a certain frequency band vibrates with a greater amplitude). However, when some stimulators are active in every
column, this effect will be largely eliminated, because it is very
difficult to distinguish between adjoining columns: two vibrating
columns can be felt as one column that vibrates at a higher energy level.
One might think that the bottom row in a filled display makes perception
of the patterns easier, because it could function as a reference,
while contour displays seem to have no reference (for the energy
display). However, contour displays have a virtual reference. Since
—within a relatively short time— almost every stimulator in a contour
display will be activated at least once (owing to the random properties
of most sounds), one is able to form a mental reference of the display.
In this way the stimulators that are active in a filled display hardly
contain any useful information.

Also, in some tests the display which contains the history (fig. 5- 3d,
the time-swept display) has been used [Clements et al., 1982; Ifukube,
1982; Spens, 1981]; however it could not be proven that its results were
better than when a display without history was used.

There are no suitable comparative tests available when information
other than that obtained by the filterbank method is offered. To create
such a display one has to apply some amount of common sense and
intuition. In this section some alternatives will be presented for offering
formant and pitch information. Since we used an Optacon for the tactile
display, the described coding methods are based on this display of 6
times 24 vibrators unless noted otherwise.

Presenting formant information

The easiest method for presenting formant information is related to the
frequency analysis methods. Since in the system described in this thesis
the formant information is presented as the frequencies of the extracted
formants combined with the energy of the (overall) speech signal, we
present the frequency along one axis of the display, and energy along
the other axis. The energy can offer information about the rhythm of
the speech and the presence of sounds in general. All four methods
shown in figure 5-3 can be used to present the formants too: at the
extracted formant frequency the energy is presented along the other
(orthogonal) axis; at frequencies other than the extracted formants, no
information is presented. The method from fig. 5-3a, depicted for
Displays

presenting two formants in figure 5-7, has been used for the experiments presented in chapter 7.

![Formant display as used in experiments](image)

Although the method described above might seem the most logical, this is not necessarily so. Other methods can be used that might give equally good, or perhaps even better results. A method to represent the first two formants, \( F_1 \) and \( F_2 \) visually to deaf people during speech training [Fletcher, 1982] is the so-called vowel triangle (fig. 5-8) [Rabiner & Schafer, 1978; Vogten, 1983]. Here the \( F_2 \) is plotted along the x-axis, the \( F_1 \) along the y-axis. The area in which the combination \((F_1, F_2)\) falls has a high correlation with the uttered vowel. Thus the auditory impaired person can see almost directly whether his vowel pronunciation is correct. This method of presenting the formants \( F_1 \) and \( F_2 \) visually can easily be transferred to a tactile display. When a triangular display is used, the only step required is presenting the frequencies on a discrete matrix. The energy can be presented as the amplitude with which the stimulator is driven. A reference for the position of the stimulators will be formed mentally by the user during the use of the display.
Fig 5-8: Vowel triangle with a number of (Dutch) vowels (after [Rabiner & Schafer, 1978] and [Vogten, 1983]).

Since the latter type of coding requires a specially designed tactile display, it has not been tried. The Optacon display has a resolution of 24 times 6 vibrators. Although the vertical resolution (24 rods) seems high enough, the horizontal resolution of 6 rods is probably too small. A slight error in the calculation of one of the formants can result in displacing the frequency by one rod. When a resolution of 24 places is used, the result of this error will be a small shift in place. This shift will probably be hardly noticeable. With a resolution of only 6 places, the result of this error is a relatively large spatial shift, resulting possibly in an erroneous interpretation.

It is also possible to code the formant frequencies not in place, but in vibration frequency. However, since the frequency resolution of the
Displays

skin is very poor, this coding strategy will probably give less good results than the previously mentioned techniques.

Presenting pitch information

Literature shows several methods for representing pitch. Since the pitch can be expressed as a frequency, it can be presented in the same way as the formant frequencies: the frequency can be presented along one axis, the energy along the other (or coded in amplitude information).

Pitch is represented as only one variable frequency. To offer this one frequency, it is often tactually presented as vibration frequency again. When next the amplitude is coded as amplitude we get a single-channel vibrator. Since the fundamental frequency falls within the frequency range for which the skin is most sensitive, it can be presented almost directly (i.e. without transforming the fundamental frequency to another frequency). A drawback of this method is that the resolution of the skin for perceiving different frequencies is poor, especially when other parameters (such as amplitude) are also variable.

It is also possible to code a function derived from the pitch instead of presenting the pitch onto the skin (either coded as frequency, or as place information). The absolute value of the fundamental frequency appears to be less important for understanding speech than the changes in this frequency [Boothroyd et al., 1988], that offer information about the intonation (e.g. when one doubles $F_0$, speech can sound like ‘Donald Duck’, yet it can still be understood). Therefore, we thought it could be useful to include information whether $F_0$ is rising, steady or falling.

When only the pitch (and basic information about the derivative of the pitch) is presented, the user has no information about the intensity or other qualities of sounds. To present the pitch, the changes in the pitch and information about the energy of a sound to the user of the aid, the following coding strategy has been developed (and implemented, though for the reasons mentioned in chapter 3, not used in the experiments as described in the chapter 7) (figure 5-9). First the display is divided into two parts: one part to represent the pitch and the change
in pitch, the other part to represent energy information (in our case the energy of the signal divided over six frequency bands).

The frequency of the pitch is presented in the first (left) part of the display along one axis. Thus the user is offered information about the intonation of the speech. The changes in the pitch can be perceived along the other axis. The energy of the signal, coded as a small frequency spectrum is presented in the second (smaller) part of the display. Normally this spectrum will only be used for perceiving information about the intensity of a sound. When non-speech sounds are present, the frequency information contained in the spectrum can be useful for obtaining some information about these sounds.

Since the system is primarily intended for supporting speech-reading, the pitch-area fills the major part of the display. In this way $f_0$ is represented in 16 frequency bands [Hnath-Chisolm & Medwetsky, 1988].

The result of this strategy can be seen in figure 5-9. The left side of the display shows a relatively low fundamental frequency that is falling. On the right side of the display we can see that the signal contains relatively high energy in the lower frequency bands.

---

**Fig 5-9: Example of the used pitch display**
Displays

Since pitch contains information about the intonation of speech, it can be useful to keep a (small) record of the history of the pitch. For this purpose we can use the idea from figure 5-3d. When only the change of the pitch is presented, it is possible to present an even longer history, when the horizontal axis (with 24 vibrators) is used as time axis. This ‘time-swept’ display has not (yet) been implemented.

Summary

We have discussed various methods of stimulating the skin. For our experiments we shall use mechanical stimulation, since this method seems to be the safest, and the easiest to accomplish. Because an Optacon device was available, the display of this device will be used to present the extracted information during the first experiments. Meanwhile a new vibrotactile display is being developed, which can be used to present data on places on the skin other than the hand (see chapter 6).

The extracted information is presented completely in the space domain. Frequency coding or amplitude coding will not yet be used. Finally it appears worthwhile to investigate further into laser stimulation for use in a future display.
In the previous chapters we have discussed the specifications of the tactile hearing aid. It must be small, portable and battery-operated. For evaluating the experimental system we have aimed at an operating time of at least 4 hours without changing or recharging batteries. Further, the system must able to perform in real-time various types of signal processing, which should be easy to select and modify.

In this chapter we shall discuss how these requirements can be met technically. First we shall review the hardware of the system, starting with the Digital Signal Processor (DSP), then we shall discuss the input circuit, and finally the tactile display. Next we shall consider the software for the system. This chapter ends with some technical results and suggestions.

When developing a system with both hardware and software, it is not easy to predict the precise hardware requirements (such as memory size and processor speed) that the various algorithms will require for running real-time. On the other hand, to write optimum algorithms one needs to know the exact hardware configuration. An iterative development will be required to develop a system that approaches the optimum design as closely as possible. Figure 6-1 shows this process in a flowchart [Waterham, 1989; Deliege, 1989]. In our case the hardware has only once been redeveloped. Only minor changes were necessary to make the system optimal for our purpose. In this chapter we shall describe both systems. The first system will be called THA-I (Tactile Hearing Aid - I), the second, THA-II [Mathijssen, 1988; Mathijssen & Leliveld, 1989].
The hardware of the system

The hardware of the system comprises three major parts:
- the input, where the audio signal, either from a microphone or from an optional audio input, enters the system and is prepared for
- the processing stage, where the input signal is processed and where information is (or features are) extracted, and
- the output, where the extracted information is presented on the skin. Next to the tactile display a visual display can be added for training purposes.

Table 6-1 gives a conspectus of the hardware of both systems. The parts mentioned in this table will be discussed further in this chapter.
Table 6-1. Synopsis of the hardware of THA-I and THA-II.

<table>
<thead>
<tr>
<th></th>
<th>THA-I</th>
<th>THA-II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DSP</strong></td>
<td>ADSP-2100 (KG)</td>
<td>ADSP-2100 (KG)</td>
</tr>
<tr>
<td><strong>Program memory</strong></td>
<td>3 x 32kByte EPROM:</td>
<td>3 x 32kByte EPROM:</td>
</tr>
<tr>
<td></td>
<td>16kWord program /</td>
<td>2 x 32kByte RAM:</td>
</tr>
<tr>
<td></td>
<td>16kWord constants</td>
<td>16 kWord program /</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 (16) kWord constants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 (0) kWord data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ratio 8/8 or 16/0 adjustable)</td>
</tr>
<tr>
<td><strong>Data memory</strong></td>
<td>2 x 32kByte RAM:</td>
<td>2 x 32kByte RAM:</td>
</tr>
<tr>
<td></td>
<td>14 kWord data</td>
<td>14 kWord data</td>
</tr>
<tr>
<td><strong>Clock / timers</strong></td>
<td>12 MHz clock. Derived:</td>
<td>16 MHz DSP clock.</td>
</tr>
<tr>
<td></td>
<td>12 MHz (DSP)</td>
<td>2.56 MHz Timer clock.</td>
</tr>
<tr>
<td></td>
<td>1.5 MHz (ADC-clock)</td>
<td>Derived:</td>
</tr>
<tr>
<td></td>
<td>11.72 kHz (ADC)</td>
<td>2.56 MHz (ADC clock)</td>
</tr>
<tr>
<td></td>
<td>2.93 kHz (display int.)</td>
<td>10 kHz (ADC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5 kHz (display int.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>320 kHz (S.C. filter)</td>
</tr>
<tr>
<td><strong>Power supply</strong></td>
<td>Universal power supply</td>
<td>MAX 7663 stabilizer</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>2 Latches: LED display</td>
<td>2 Latches: LED display</td>
</tr>
<tr>
<td></td>
<td>Mem.mapped: Optacon</td>
<td>and tactile display (Optacon)</td>
</tr>
<tr>
<td><strong>AD converter</strong></td>
<td>AD7580 (10 bit)</td>
<td>AD7580 (10 bit) + buffer</td>
</tr>
<tr>
<td><strong>Filter</strong></td>
<td>XR-1016 (cut-off frequency variable)</td>
<td>XR-1016 + filtering from amplifier</td>
</tr>
<tr>
<td><strong>Amplifier</strong></td>
<td>None (laboratory)</td>
<td>TLC27L4: Microphone-and audio-amplifier</td>
</tr>
</tbody>
</table>
Fig 6-2: Block diagram of the THA-I

Fig 6-3: Block diagram of the THA-II
Chapter 6

Block diagrams of both systems can be found in figure 6-2 and figure 6-3.

First we shall look at the heart of the system. This is the part where the signal is processed and prepared for presentation on the skin.

The digital signal processor

It was seen in chapter 4 that the best method for processing (speech) signals is with the use of a Digital Signal Processor (DSP). We have chosen to use the ADSP-2100 (-KG) from Analog Devices. The ADSP-2100 is a CMOS DSP that can operate at a maximum instruction rate of 12.5 MHz (or 12.5 Mega Instructions Per Second, MIPS). In this case the oscillator frequency (or clock frequency) of the DSP is 50 MHz. When in this chapter an operating frequency is mentioned, it will be the instruction frequency unless otherwise specified.

* Internal structure

The ADSP-2100 [Analog, 1986] is a programmable single-chip microprocessor, optimized for digital signal-processing and other high-speed numeric processing applications. It comprises three full-function and independent computational units: an Arithmetic/Logic Unit (ALU, as most processors have), a multiplier/accumulator and a barrel shifter. All three units process 16-bit data directly. Figure 6-4 shows schematically the internal architecture.

The data memory of the DSP consists of at most 16k data words of 16 bits length. Peripherals (such as AD-converters of other I/O-ports) should be memory mapped. The program memory of the DSP consists of a maximum of 16k program instructions of 24 bits length (all instructions are one word instructions of 24 bit length) and 16k data code (either fixed data or variable data) of 16 or 24 bits length. The DSP has a special output line to indicate whether program code or data is fetched from the program memory.

The Harvard structure, typical for DSP's, is obvious in this figure (fig. 6-4): the data buses and address buses for the program and the data memory are completely separate. This means that the processor can...
fetch a new program instruction and a data word in parallel. When the instruction code is already available in cache memory (which will be discussed later), the DSP can fetch two data words (each of 16 bits length) simultaneously, one from the program memory and one from the data memory. This fact can be used to speed up processing considerably, as will be shown later. This structure that enables the storage of data in the program memory is called the modified Harvard structure.

Every time the DSP fetches its next instruction, it first checks the CACHE-memory. This memory contains up to 16 instructions, which are the ones most recently fetched from the program memory. Under most conditions, it does not make much difference whether an instruction is fetched from the program memory or from the cache, other than that in the latter case the program memory need not be accessed (which might save some energy). However, when data has to be written to or
read from the program memory, cache memory is very useful. Since the DSP can fetch either an instruction or data from the program memory (but not both at the same time), it will normally need an extra instruction cycle to access program data. First the instruction is fetched, and the data can then be accessed in the next cycle. However, when the instruction is in cache, the DSP can immediately access the program data, thus saving one instruction cycle.

The two Data Address Generators (DAG) each consist of four powerful pointers that can point to the data and program memory. Using these DAG's one can construct cyclic arrays with hardly any software overhead. Once the DAG is set correctly, it is possible to read a cyclic array word for word, without using software to increment the address pointers or to check the boundaries. It is in fact even possible to read an array backwards, or with increments greater than one. One DAG is equipped with the possibility of Bit-Reverse Addressing, which makes a Fast Fourier Transform easy to implement.

Only one of the three computational units can be accessed during program flow. Most instructions that access these units can be combined with a memory or register access. This means that it is possible to set up new data while the ‘old’ data are processed. Also the output registers of the computational units can be used directly as input for either of the three units.

The counter function of the ADSP-2100 is also worth mentioning. Using the internal counter it is possible to create program loops without any software overhead, other than initializing the counter. Counter loops can be nested up to four levels deep. The possibility of nesting counter loops facilitates e.g. matrix manipulation and repeated summation.

The ADSP-2100 has 4 internal stacks. The PC-stack, which is 16 words long, is used for storing the program counter when calling a subroutine or an interrupt routine. The status stack (4 words) is used to store the status when an interrupt routine is executed. The counter and loop stack (each 4 words) enable the use of nested counter functions and program loops.
Technical description

When the software is discussed, we shall see most of these features again and how they can be used successfully.

* Using the DSP

One of the system requirements is that the THA should use as little energy as possible, to enable an operating time of at least 4 hours. Normally this requires that CMOS components be used [Waterham, 1989], or that special measures have to be taken when using non-CMOS components. However when CMOS components are used at a high operating-frequency, they still use relatively much energy. A typical maximum value for the energy used by CMOS memory components is 40 mW/Mhz (clock frequency) [Nec, 1987; Intel, 1988]. With an operating voltage of 5 Volts, this results in 8 mA/Mhz. This means that the lower the operating frequency, the lower the energy consumed by the CMOS components. When the frequency is too low however, it will no longer be possible to perform real-time signal-processing. Therefore one has to find a minimum operating frequency at which most desired algorithms will run real-time. Of course these algorithms will have to be optimized for shortest processing time too.

From the literature available about the ADSP-2100 (e.g. [Nell & Fine, 1986]) it was decided to let the first system (Tactile Hearing Aid no.1: THA-I) operate on an instruction frequency of 3 MHz. Since all registers in the DSP are static memories, the DSP can run at every frequency from 0 to its maximum frequency of 12.5 MHz. Our choice of 3 MHz proved to be sufficient for most basic signal-processing algorithms to run in real-time. This frequency, which is relatively low for the DSP, also had some other advantages. Normally a DSP requires special memory IC's, that can be accessed very fast (access times of less than 40 ns). However, these fast memories were not yet available in CMOS versions when the system was developed. The fastest CMOS memories available at that time had an access time of about 100 ns. The instruction cycle of the DSP having been lowered to 3 MHz, it was possible to use normal external CMOS memory components.

Table 6-2 shows the measured current required by several parts of the system THA-I, which will be described more extensively in the next sections. These were measured under different operating modes of the
Chapter 6

Table 6-2. Current (in mA) used by various parts of the tactile aid (THA-I). Measured values versus theoretical maximum values (according to the datasheets, corrected for the operating speed).

<table>
<thead>
<tr>
<th></th>
<th>DSP</th>
<th>Memory</th>
<th>ADC</th>
<th>Gates</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode 1</td>
<td>0.48</td>
<td>4</td>
<td>0.01</td>
<td>0.05</td>
<td>4.5</td>
</tr>
<tr>
<td>Mode 2</td>
<td>8.0</td>
<td>?</td>
<td>10</td>
<td>30</td>
<td>4.5</td>
</tr>
<tr>
<td>Mode 3</td>
<td>12.2</td>
<td>32</td>
<td>55</td>
<td>100</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Notes:
- **Mode 1**: DSP in HALT-mode
- **Mode 2**: DSP running, no cache access, little DM access
- **Mode 3**: DSP running, no cache access, with complex arithmetic, read and write instructions (worst case situation)

DSP. Mode 1 is the DSP in HALT-mode, i.e. the program is stopped (no instructions are executed), but the oscillator still runs. In mode 2 the DSP accesses the program memory continuously (no cache operation), while the data memory is accessed for 5% of the time or less. In mode 2 the DSP performs “complex arithmetic, read and write instructions” [Analog, 1986], i.e. in one instruction an arithmetic operation is performed and two data bytes are moved (one from the program memory and one from the data memory) without using the cache. This mode can be considered as as a worst case situation, which only occurs in the implemented algorithms during short periods of time.

The components that used the most energy are the memories, when they are being accessed each instruction cycle (reading program memory, and writing to data memory). In this case five memory chips are accessed with 3 MHz. The maximum specified current that can be consumed (according to the product data) at this frequency is about
Table 6-3 shows the total current used by the THA-II measured for the different algorithms. The several test points for measuring the current on the THA-I were not considered necessary for the THA-II. The maximum average current of 75 mA means that the system can operate for at least 6 hours when rechargeable penlight batteries with a capacity of 500 mAh are used.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Average current used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filterbank</td>
<td>70 mA</td>
</tr>
<tr>
<td>Pitch detection + filterbank</td>
<td>60 mA (unvoiced) / 75 mA (voiced)</td>
</tr>
<tr>
<td>Formant analysis</td>
<td>75 mA</td>
</tr>
</tbody>
</table>

External memory

The ADSP-2100 has no internal memory, apart from its cache, stacks and registers. For this reason all memory must be applied externally. When the system is less experimental it will be possible to replace the ADSP-2100 with a compatible DSP, that has internal program- and data- memory. For the experimental system we preferred to use external memory, because it facilitates measurements on and modification of the system. With external memory it is (theoretically) possible to enlarge the memory area as far as needed. Internal memory cannot be enlarged when necessary. Only with special software and hardware might it be possible to increase memory capacity in this case.
Chapter 6

* Program memory

The program code and program data for the DSP can be stored in several ways. Let us first concentrate on the program code (in short: program). The program can be stored either in volatile or non-volatile memory. Development systems often use volatile memory (RAM) for program storage. Together with a controller (e.g. a microcontroller, or a small start-up program for the processor) and a (serial) interface, it is possible to download the program from a computer (e.g. a P.C.) into the program memory. Modification of the program is easy, and the program code can be changed fast. A disadvantage of this method is that the program memory will be erased when the system is switched off. This is rather bothersome for a portable system (unless special measures are taken, such as constant battery back-up). Also the system will be larger owing to the extra hardware required.

When non-volatile memories (such as EPROM’s) are used, the program will not be erased, even when the system is disconnected from the power supply (i.e. the program memory is also without power). Special hardware can also be omitted when non-volatile memory is used. The disadvantage of this method is, however, that modification of the program is less easy. To change the program, one has to replace the memories (erase the EPROM’s and program them again). It is possible to compensate for this disadvantage by using multiple sets of EPROM’s.

Next we have to focus on the program data, i.e. data which is stored in the program memory. Here we have to answer the following questions:

- Do we want data in the program memory?
  - If so,
  - Do we want the data strictly separated from the program?
  - Do we want fixed data (constants) or variables?
  - Do we allow extra components for data in program memory?
  - Do we need normal word-length (of 16-bit) or longer words (of 24-bit)?

When we look at the algorithms that should be implemented (see chapter 3), we can see that a number of these, such as (auto-) correla-
Technical description

...tion, discrete fourier transform (DFT), and the often used FIR filters, consist of a sum of products:

\[ R_{n,m} = \sum_{i=k}^{l} x_i \cdot y_j \quad \text{where} \quad j = (n \cdot i + m) \quad (6 - 1) \]

To compute a DFT or a FIR filter one has to multiply variables (the input samples) by constants. In order to compute the auto-correlation, variables need to be multiplied by variables.

The ADSP-2100 has special instructions to fetch two data bytes in one instruction cycle, when one data byte is in data memory and the second is in program memory. For this reason it will be very useful to have room for data (variables or constants) in the program memory area. For this reason the first question will be answered in the affirmative.

The next question is whether the data and program area (in the program memory) need to be separated. This depends largely on the amount of program and data we want. The ADSP-2100 can access 16 kWWord program memory directly. When program code and data together will never exceed 16 kWWord, it is possible to mix both (i.e. data and program can be anywhere in the 16k memory space). To indicate that the ADSP-2100 fetches data, it has a special signal: Program Memory Data Access (PMDA). This signal can be used to separate physically program and data. This also enables the use of 16 kWWord true program space and 16 kWWord true data space in the program memory.

Since there was no reason to store data and program code in the same memory area, we have chosen for the last option. In the first system (THA-I), three 32 kByte EPROM's were used, where the lower 16 kByte contained the program, while the higher 16 kByte contained the (fixed) data. The signal PMDA is used as the highest address bit and selects either the lower 16 k (for program code) or the higher 16 k (for data).
Chapter 6

The next questions (fixed/variable data and extra components) have in fact already been answered for THA-I. To minimize the number of components, the first choice was to use no extra memory components, and store the (program memory) data in the same EPROM's as the program code. This also meant that no variable data could be stored in the program memory.

However, when developing the programs for the THA-I, this proved to be not very efficient for certain algorithms. Especially where correlation functions were concerned, where two variables have to be processed, the chosen frequency proved to be too low. A considerable reduction in the required processing time could be realized when variables could be stored in program memory. This meant however that RAM had to be added to the system. Since the minimum required processing-time could be reduced by nearly 50%, it was thought preferable to add some components rather than increase the processing speed. When the speed is increased the power consumption of the digital components also rises proportionally. When two memories are added (so the total number of memories rises from 5 to 7), while the operating frequency remains unchanged, the power consumption will increase at most by 40%, as opposed to 100% when the frequency is doubled. Owing to the fact that the RAM in the program memory will not be accessed every instruction cycle, the increase will be even less than 40%. For this reason RAM memory was added to the program memory in the THA-II.

To enable the storage of constants, the data area in the program memory of the THA-II is divided into two parts: one part of 8 kWord for constants (in EPROM) and one part of 8 kWord for variables (in RAM). For compatibility with the first system (THA-I) the constants are placed in the lower area. It is also possible to disable the RAM, and use the remaining 8 kWord of EPROM instead (since EPROM’s are not available as 24 kByte types, but as 16 kByte or 32 kByte, three 32kByte EPROM’s are used). Since the ADSP-2100 normally processes only 16-bit words, we have made no use of the option to use 24-bit data words in its program memory. Although 24-bit constants are still possible (since the program memory is 24-bit wide, and the constants are stored in the same IC’s), variables can only be stored as 16-bit words. The use of 24-bit variables—which will almost never
Technical description

occur—would also require an extra RAM IC. In the program memory the same type of RAM’s are used as in the data memory: 32 kByte RAM’s. Of course it is possible to use 8 kByte RAM’s (16 kByte RAM’s are quite unusual). However, the size and energy consumption of these components is about the same as of 32 kByte RAM’s. Also 32 kByte RAM’s were already used as data memory, so we have used the same type of RAM for the program memory.

* Data memory

The maximum amount of data memory that the ADSP-2100 can manage directly is 16 kWord (of 16-bit each). In the data memory also some area should be reserved for mapping the peripherals (these are memory mapped). In both systems (THA-I and the updated version, the THA-II) the data memory and the peripherals are combined in such a way that the maximum of accessible data memory is present with the minimum of components. It was possible to have 14 kWord of data memory, while 8 I/O ports can be used, with a minimum of hardware overhead.

To select between RAM and peripheral and to select a peripheral, standard gates are used (see figure 6-5: a triple three input NAND-gate—G6, G7, G8— for RAM/peripheral selection, and a 3 to 8 de-multiplexer—HC138— for peripheral selection). Gate G6 selects between the lower 14kWord and the upper 2kWord. When the DSP selects its data memory, line DMS becomes active. This signal is gated (via G7, which operates as an inverter and G8, which operates as an OR-gate) to the RAM only when the selected address (A0..A13) is in the lower 14 kWord. The 3 to 8 de-multiplexer becomes active only when DMS is active and the output from G6 is low, i.e. in the upper 2kWord range. When the de-multiplexer is activated it makes one of its 8 output lines (Y0 .. Y7) active, depending on the lower three address lines (A0, A1 and A2). Only two peripherals are momentarily used (connected to Y0 and Y1).

Instead of using standard gates one could think of applying programmable logic. Although these devices give a slightly greater freedom (because their functioning can be reprogrammed and because fewer components are needed), we have chosen not to use them. First of all it would substitute at most 4 small (14- or 16-pins) IC’s with one larger
Fig 6-5: Address select logic

(24- or 40-pins) IC, which is virtually no gain in required space. Further most programmable devices use more energy than standard gates, in order to have short delay times; the operating current at 1 MHz is about 10 mA [Intel, 1987], where High-speed CMOS (HC)-gates use about 0.4 mA [Philips, 1986]. And finally the delay (or propagation) time of a programmable device is often longer (about 50 nsec versus 10 nsec for an HC gate). When operating at a clock frequency of 16 MHz (and using memories with an access time of 100 nsec), these longer delays will make operation critical.

Peripherals

The tactile hearing aid has two major peripherals. One is the input device –for the DSP the analog to digital converter– and the other is the output device—in this case a buffer for the tactile display. The buffer consists of two 8-bit wide latches. For the DSP these latches can be considered as a single write only memory location.
The input device—the ADC—needs some more attention. It will appear that there are some conflicting requirements for the ADC. In the first place we would like to have as high an accuracy as possible (with a maximum of 16 bit, since that is the DSP’s standard accuracy), in order to obtain optimum signal to noise ratio and dynamic range. On the other hand, we also want to keep energy consumption low. As will be discussed in the section about the analog part, we wanted to use a unipolar supply voltage, which means that the ADC and its reference voltage need to be unipolar. The supply voltage should be approximately 5 volts. Finally the ADC needs to be fast enough for a sample frequency of about 10 kHz, and it should be possible to interface it easily with the ADSP-2100.

Most CMOS ADC’s that can operate on a single supply voltage have an accuracy of 8 bit. However, for speech processing a resolution of 8 bit is rather low; 12 bit is more common (e.g. [Vogten, 1983; Rabiner & Schafer, 1978]). Unfortunately 12-bit CMOS ADC’s that operate on 5 volts were not yet available when the THA-I and THA-II were designed ①. The available ADC’s with the highest accuracy that matched our requirements had an accuracy of 10 bit. The AD converter used in the tactile aid, the AD7580, operates on a single 5-volt power supply, has an internal sample and hold amplifier and has an accuracy of 10 bit. Its reference voltage is 2.5 volts and the maximum sample frequency is 50 kHz. At the end of a conversion, the ADC generates an interrupt for the DSP, so that the DSP knows that it can read the new sample. In the THA-I the ADC is connected directly to the DSP. In the THA-II a buffer is added between the ADC and the DSP. In this way no fast changing digital signals are present on the output port of the ADC, which could cause noise (owing to cross-talk) on the input signal of the ADC.

The interrupt routine to read the input sample will be discussed in the software section of this chapter.

① Only recently, Burr-Brown has introduced a 12-bit CMOS-ADC with a single operating voltage of 5 volts [Burr-Brown, 1990].
**Timing of the peripherals**

When sampling a signal, accurate sample intervals are advantageous [Oppenheim & Schafer, 1989]. The ADSP-2100 has no timer functions, which means that generating accurate intervals with the DSP is difficult. An external timer circuit will offer intervals that are much more accurate. In the first system, the sample intervals are derived from the system clock for the DSP, using a binary counter. This resulted in a sample frequency close to 12 kHz. For THA II the ADC-Clock is a separate oscillator circuit based on a 2.56 MHz quartz crystal, with the appropriate counters to realize a 10 kHz timer. This timer governs starting the AD conversion and is a clock for the ADC. At the end of a conversion, the ADC generates an interrupt, which should be handled by the DSP within a fixed time. Otherwise the obtained sample will be lost, owing to a new conversion.

The external timer is also used for controlling an (optional) visual display, which will be discussed below. Roughly every 40 ms this LED display requires new data, which means a 2.5 kHz interrupt for the DSP. In a short interrupt routine the DSP takes care of updating the display. In the THA-I this frequency is derived from the DSP clock, in the THA-II from the ADC-clock.

---

**Fig 6-6: The clock circuit for the ADSP-2100 in THA-II**
The tactile display is fed its data every time the DSP finishes processing a frame of speech. The DSP in THA-II obtains its clock signal from a 16 MHz oscillator, which has already been designed for combining with a microcontroller that can stop the oscillator (this will be discussed later). Using high speed CMOS NAND gates (74ACT00) and a 16 MHz crystal we have an oscillator that can be stopped externally (see figure 6-6). The first system had a comparable oscillator, using HCT inverter gates, running at 12 MHz, which could not be stopped externally. All the other timing signals were derived from this 12 MHz clock.

Summarizing, we have seen that the THA-I had one 12 MHz oscillator. The clock signals for the DSP, the ADC and the display interrupt were derived from this oscillator. The THA-II has two oscillators. One 16 MHz oscillator for the DSP and one 2.56 MHz oscillator, from which the clocks for the ADC, the display interrupt and the switched capacitor filter which will be discussed in the section “Analog components” are derived.

* The tactile and visual display

The output device of the tactile hearing aid is a tactile display. We used the display from the Optacon during the development and the experiments. Since the Optacon was originally intended as a reading aid for the visually impaired, it had to be modified slightly to suit into our application. Using an interface card, connected to the wires which originally were connected to the camera, it is possible to drive each rod of the optacon individually.

For training purposes a visual display has been added, consisting of 8 times 32 Light Emitting Diodes (LEDs). The LED display is controlled almost directly by the DSP. For the processor, it is possible to select one row of 8 LEDs and control which LEDs should be on or off. By selecting each row during a short time, it is possible to emulate a fully controlled display (multiplexing). This multiplexing is performed at a frequency of 2.5 kHz, which means that the display oscillates at a frequency of about 80 Hz. Under normal circumstances the display will appear to be lit continuously.
Both the visual display and the tactile display are connected to the same output port. The Most Significant Bits (MSBs) of the port control whether the data are channelled to the visual or the tactile display.

Analog components

Since the tactile aid picks up analog signals, it will be clear that at least some part of the system needs to be analog. The signal is picked up by a microphone; for experimental purposes, a second audio input was required for prerecorded signals. Before the analog signal can enter the analog to digital converter, it needs to be amplified and filtered [Heffels, 1989]. One can only obtain the highest resolution when the input signal can reach an amplitude close to 5 volt (peak-peak) during loud sounds, since the input range of the ADC is 5 volts. Filtering is necessary to prevent aliasing (the Nyquist theorem). When sampling a signal at frequency $f_s$, the maximum input frequency must be below $f_s/2$.

Our design of the analog circuit took the following desiderata into account:

- use of a uni-polar voltage;
- a supply voltage of maximum 9 volts;
- analog circuit that consumes little energy (because of the necessity of portability);
- easy-to-reproduce analog circuitry.

Before we look at the realization of the analog circuit, we shall discuss the requirements. We wanted the supply voltage to be uni-polar to keep the system as compact as possible. A bipolar supply voltage necessitates either a double set of batteries or a circuit to convert a positive voltage to a negative one. The first solution means that the system will be heavier and larger, even when the battery for the negative supply can be smaller than the one for the positive supply. Also recharging of the batteries (when nickel-cadmium batteries are used) will be more complex. The second solution—converting the voltage—needs an extra circuit. It is difficult to design a small circuit that uses little energy that accomplishes this task. For these reasons we have decided to use a uni-polar supply.
The same reasons led to the requirement that the supply voltage should be no higher than 9 volts. We already need 5-volt supply for the digital components. To stabilize the power supply at 5 volts one needs to start with about 7 volts (see also the next section). However, the battery voltage should not be much higher either, since this would increase the power consumed (the current remains unchanged while the voltage increases, just as the consumed power). Battery packs or a combination of ‘penlight’ batteries often provide 9 volts when the batteries are new, which is only slightly higher than the 7 volts required for obtaining a stable 5-volt supply. To increase the voltage, either larger or more batteries or a voltage converter are needed.

It will be clear that the analog components should consume as little energy as possible. It would be possible for the digital components to be switched off automatically when no relevant information can be offered to the user. In this case the analog circuit will normally have to remain on continuously so the analog circuitry should still be designed to consume as little power as possible, otherwise switching off the digital components is of little benefit.

The last point is that the analog circuit should be easy to reproduce. We shall see that the anti-aliasing filter must have a relatively high quality factor. Owing to the inaccuracy of the available components, it is not easy to design a circuit that is reproducible with a very high degree of accuracy under worst case situations when this filter is designed using operational amplifiers, capacitors and resistors. This can be realized with an integrated low-pass filter, based on switched capacitor techniques [Hegt, 1988; Allen & Sanchez-Sinencio, 1984; Moschytz, 1984]

Bearing these points in mind we have come to the following input circuit (fig 6-7): The pre-amplifier (IC1a, IC1b) and amplifier (IC1c) are made using low-power CMOS operational amplifiers (TLC27L4-CN). These OpAmps consume about 20 μA per amplifier (i.e. 80 μA per IC). It is not necessary to switch off these amplifiers when the (digital) system is switched off to save energy. Because of the low power consumed by these amplifiers their unity-gain bandwidth is rather low. Where the average operational amplifier (that uses about 1 mA) has a unity-gain bandwidth of at least 1 MHz, the maximum
unity-gain bandwidth of these CMOS amplifiers is 50 to 100 kHz (the TLC27L4-CN has a typical unity-gain bandwidth of 85 kHz at 25°C). This means that the maximum amplification factor that can be obtained with one amplifier is between 10 to 20 (with a bandwidth of 5 kHz). Using 3 amplifiers (IC1a ... IC1b) it is possible to bring the input signal from the microphone level (which is about 1 mV) to the full input swing of the AD-converter (i.e. 5 volts peak-peak).

The input signal needs to be filtered to prevent aliasing. Using a 10-bit AD-converter that takes samples at 10 kHz, the amplitude of signal components above 5 kHz needs to be less than $5V / 2^{11}$, which is about 2.5 mVpp (0.88 mVe), or 66 dB attenuation. As the input signal will have most of its energy in the lower frequencies, and the input signal will reach its maximum input value of 5 Vpp only very rarely (to prevent clipping), the attenuation at 5 kHz can be slightly less than the theoretical value of 66 dB without causing aliasing.

The main low-pass filter to accomplish the attenuation is a seventh order switched capacitor filter (SCF), the XR-1016. The -3dB point of
this filter is set at 3.2 kHz. Since the switched capacitor filter has an elliptic filter characteristic, frequencies at 5 kHz will be attenuated by about 60 dB [Exar, 1987; Heffels, 1988; Moschytz & Horn, 1981]. The RC filter at the output of the SCF (see fig 6-5) eliminates the sample frequency of the SCF. To buffer the filtered signal, before it enters the AD converter, IC1d is added.

Because the switched capacitor filter needs to sample the signal in order to filter it, the input signal of this filter needs to be low-pass filtered too, to prevent aliasing. This filtering has already taken place in the amplifier section (IC1c and IC1a, IC1b) by the frequency characteristic of these amplifiers owing to their low unity-gain bandwidth. Since the sample frequency of the digital filter is 50 or 100 times as high as the desired cut-off frequency, the anti-alias filter for the switched capacitor filter does not need to be as good as the filter needed for the AD converter. The filtering realized by the operational amplifiers will filter the input signal sufficiently.

In this way all frequencies above 5 kHz will be decreased sufficiently (approximately 60 dB at 5 kHz). The fall-off (or corner) frequency of the switched capacitor filter (the XR-1016 from EXAR) can be set with its clock frequency at 1/50 or 1/100 of the clock frequency. In the THA-II we have chosen for the latter ratio.

As mentioned before a maximum supply voltage of 9 volts was allowed, though 5 volts would be ideal. It appeared that the analog circuit could operate on a single 5-volt power supply, while the output signal could still reach amplitudes near 5 $V_{pp}$· So there was no need to increase the supply voltage above 5 volts.

The AD converter requires an input signal between 0 volts and 5 volts. In other words, the mean value of the input signal should be around 2.5 volts. Since the amplifiers only have a positive power supply of 5 volts, the signal needs an offset of about 2.5 volts, otherwise clipping will occur. To adjust the offset, a variable resistor (of 500 kΩ) is added (see fig. 6-7). Because of the offset voltages of the operational amplifiers, it will not always be possible to ensure a very stable offset voltage of 2.5 volts. For this reason a software offset adjust has been added, after digitization of the signal. This method will be described when the software is discussed (fig. 6-11).
Power supply / voltage reference

The first system, the THA-I, obtained its supply voltage from a laboratory power supply. The THA-II could be powered from batteries and had a special circuit for the power supply. The final system will be powered by batteries e.g. rechargeable nickel-cadmium batteries. For a stable supply of 5 volts (needed by both the analog and digital components), a battery voltage of about 7 volts is needed, followed by a voltage regulator. Since NiCad's produce 1.2 volts nominally, at least six batteries will be needed, resulting in a voltage of 7.2 volts.

To produce an energy-economic system, the circuit that stabilizes the supply voltage needs to be low-power. For this purpose special components are available. In our system a MAX-7663 is used, a stabilizing IC that has a quiescent current of approximately 5 $\mu$A [Maxim, 1989]. This IC can deliver 40 mA directly, i.e. without special external components, which is too low when the DSP performs arithmetic-intensive algorithms. A transistor, operating as emitter follower is used to enlarge the current that can be used. The maximum average current used by the second system (THA-II) is about 65 mA by most digital components and 10 mA by the combination of the analog components, the AD-converter and the timing circuit for the ADC and the tactile display (see also table 6-3).

The AD-converter needs a very stable reference voltage of 2.5 volts, with a low-output impedance. Most voltage references (also the one used in THA-I) need at least some tenths of a milli-ampere for producing a stable voltage. A true low-power and accurate voltage reference can be made with a MAX-7663 too. Although this component is slightly larger than most references, its other features (such as its reference voltage that can be set at any value desired and its very low quiescent current) compensate for this disadvantage.

Total scheme

The diagrams of the digital part of the first version of the tactile hearing aid (THA-I) and the second version (THA-II) can be seen in figure 6-8.
Fig 6.8: Diagram of the tactile hearing aid, THA1.
Technical description

... and 6-9. The amplifier circuit (figure 6-7) and the power supply circuit are not included. For clarity no pin numbers have been added.

The software of the system

A theoretical description of the algorithms used in the tactile hearing aid has been given in chapter 3. In this section we shall review how these algorithms were implemented, which software techniques were used, and we shall discuss the input and output routines, which are normally interrupt-driven.

Architecture of the software

* Software tools

A number of software developing tools are available for the ADSP-2100 [Analog, 1987]. We shall briefly discuss which tools were used for developing the software described in this chapter.

The system builder describes, checks and defines the memory, when the hardware description is given. Once the hardware has been defined, using the system builder, one can define variable names for memory areas, without having to bother about precisely where the memory is located. The programming seems to become more like a higher programming language than an assembly language.

The cross-assembler assembles the source code. The use of both local and global variables and constants is possible. Also the exact memory location of variables and of linear and circular arrays need not be defined. The assembler combined with the other software developing tools take care of this memory mapping. The aforementioned features allow the use of so-called modules. A program can be written using several modules, e.g. for the interrupt routines, for initialization and for several other routines.

The linker combines the assembled modules and maps the variables and arrays on the available memory. The output of the linker can be
used directly for running it on the DSP system, or for simulation on a personal computer or mainframe.

To verify the correctness of a written program (that has no syntax errors, of course), one can use a simulator or an emulator. Both are available for the ADSP-2100. The emulator is a combined hardware and software tool, used in combination with the developed hardware system, and works real-time. The simulator is a software tool, that runs on a PC or mainframe, and does not work real-time. For developing the software for the tactile hearing aid only the simulator has been used. The simulator worked well for this purpose; once a program ran error-free on the simulator, no errors were observed in the real system.

* **Block diagram of the software**

The layout of the programs written for the tactile aid can all be described by the same block diagram (fig. 6-10). We shall describe this diagram, followed by a discussion of each of the blocks. It will appear

![Block diagram of the software for the THA](image)

**Fig 6-10: Block diagram of the software for the THA**
that only the blocks where the signal is processed and where the desired information is extracted from the processed data will differ for the different processing algorithms. The other blocks will remain the same (except for some possible minor changes such as length of arrays etc.).

When the system is reset (e.g. after switching on the power), it will start by initializing the necessary registers, counters, variables, arrays and so forth. Also the displays will be reset, i.e. every LED and vibrator will be turned off. Once the initialization is finished, the DSP will start reading the samples from the AD-converter.

The samples from the ADC are read on an interrupt basis. Every time the ADC has completed a conversion, it generates an interrupt, forcing the ADSP-2100 to run a small routine for reading the sample, compensating for the offset voltage, storing the sample and checking whether the offset voltage has changed. This routine is called upon every 100 $\mu$s and normally uses 11 instruction cycles (2.75 $\mu$s), including the interrupt handling. When the software detects that the offset voltage has changed, it takes 4 to 5 instructions extra to compensate the offset by 1 LSB value.

Compensating for the offset voltage is done as follows (see also figure 6-11):

normally the mean value of the input voltage for the ADC should be 2.5 volts, which corresponds to a digitized value of 512. The ADSP-2100 operates internally with twos-complement data. This means that the mean voltage (which is in fact the zero-volt level) should be transformed from 512 to 0. This is done by subtracting the offset voltage (512) from the sample. However, when the zero-volt level has changed, the samples will to have an average value that is not equal to zero. For some signal-processing algorithms, this has adverse side effects, such as less accuracy for the obtained results. For this reason we compensate for this change in offset voltage by means of a small software routine. In this routine we add every sample read from the ADC to a 16-bit register. When the offset voltage is identical to the voltage subtracted from the input sample, the result from this addition will always be below a certain limit, with an average value of zero (since the input signal does not contain D.C. signals, or frequencies very close to zero). However, when the offset voltage differs from the internal offset, the addition will cause an overflow or underflow, since
the average sample is no longer equal to zero. When this over- or underflow occurs, the internal offset value is either increased or decreased by one bit. Thus changes in the offset voltage (which will normally be very slow) are compensated for very slowly. During one frame (20 to 30 ms), this compensation will occur no more than once or twice (under normal conditions, i.e. only slow and small changes in the offset). The processing will hardly be influenced by this small adjustment in the signal, while an error in the offset voltage of 100 mV will be compensated within about 10 seconds.

The samples are stored in a circular buffer that is (at least) twice as long as the number of samples used in the signal processing. When the DSP is not processing data, it is waiting for the right amount of new data (or 'waiting for the buffer to be filled'). The DSP checks how many new samples have been read, and when the required number of samples is available, it starts the signal processing (see also figure 6-10). During signal processing, the storing of new samples continues.
**Technical description**

When the signal processing is complete and the extracted information displayed, the DSP will wait again until a new frame of samples is ready to be processed.

The signal processing consists of the implementation of the algorithms described in chapter 3. Since the ADSP-2100 is a 16-bit integer based signal-processor, most algorithms have been implemented using the integer arithmetic. However some algorithms (such as robust formant analysis) require floating-point arithmetic (FPA). Using two integers per real number and short arithmetic floating-point subroutines, it is possible to implement floating-point algorithms. The major drawback to FPA is, however, that it requires more instruction cycles for one computation.

When the speech (or sound) data have been processed, the DSP needs to extract the desired information for displaying on the skin. This is necessary, since the processing itself still leaves too much information for presenting it to the skin. Since we wanted to be able to change the tactile display easily, the signal-processing algorithms are written without using specifications of a display. The extracted information is adapted for the used display only after signal processing. When considering formant analysis, for example, it is necessary to compute all five formants (when using the RFA method). To display the formant information, we use only the first two formants. Also the formant frequencies, as computed, are far too accurate for displaying. For this reason we need to compress the information obtained.

Once we have got the precise information we want to display, it needs to be coded for displaying via the visual and tactile display. For each display, an array has been defined in which the display data has to be stored. The data are sent to the tactile display when the array is filled. The data for the visual display are continuously sent to the display. Every 0.4 ms a new row is accessed on the display. Every 0.4 ms the DSP gets an external interrupt, starting the display routine, which uses seven instruction cycles (including interrupt handling). It is possible to keep the interrupt routine so short, because the display data are stored together with the place information where the row has to be displayed. This information is of course available as the place in the array, however to combine this information in the interrupt routine would
cost extra processing-time. Now the DSP merely has to fetch the data from the array and send it to the correct address.

Timing of the software

As mentioned before, the tactile hearing aid needs to process the data in real time. Information extracted from the speech signal needs to be offered within fractions of a second, otherwise a noticeable and irritating delay will occur. For the tactile hearing aid we have allowed delays of about 50 ms between the picking up of a sound and the presentation onto the skin. This delay is comparable with the delay that occurs when listening to a speaker at a distance of about 15 meters. It is not known whether the time delay introduced by the tactile sense has an influence on the usefulness of the aid.

The available time needs to be split up in two parts: one part where the data are gathered and one part where the data are processed. Although algorithms are possible where the data are already being processed while gathered, it is not always the easiest and fastest way for an algorithm. Therefore we have chosen to gather the data first and to process them only when all samples are available. In this way it is possible to process a frame while the next frame is being loaded. The available processing time for one frame is –using this method– at most equal to the time needed to gather one sample. So, when only 50 ms are available for sampling and processing, one can use about 25 ms for sampling (resulting in a frame of 256 samples) and 25 ms for processing. In this way the maximum delay will be 50 ms, while the average delay (since we are in fact ‘averaging’ over a framelen) is 37.5 ms ($\frac{25}{2} + 25$).

The net time available for the processing in this period of 25 ms is a little less, owing to the two processes running parallel to the major processing. During these 25 ms, 250 ADC interrupts and 62.5 display
Technical description

Interrupts are handled\(^2\), which is equal to \((11 \times 250) + (7 \times 62.5) = 3187.5\) instructions or about 0.8 ms. The processing of 256 samples should take at most 24 ms or 96,000 instructions.

If one sets the average delay time at 50 ms, and if the processing time needed is small enough, it is possible to use another sequence to process the signals. This method, that is implemented in the filterbank analysis, uses framelengths of 50 ms. The time to process this frame of 50 ms is still less than 25 ms. Under these conditions, it is possible to process a 50 ms frame every 25 ms, creating overlaps of half the framelength. In this way the average delay time is 50 ms \((50/2 + 25)\). The advantage of this method is that the information at the edges of a frame is not lost, thanks to the required windowing. Figure 6-12 shows this idea schematically. Unfortunately this method can only be used when the

Frame nr. (256 samples each):

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\end{array}
\]

Processing frame:

\[
\begin{array}{ccccccc}
1 & + & 2 \\
2 & + & 3 \\
3 & + & 4 \\
4 & + & 5 \\
5 & + & 6 \\
6 & + & 7 \\
7 & + & 8 \\
\end{array}
\]

Fig 6-12: Overlap of the processed frames

\(^2\) The display interrupts are only used during training and evaluation purposes. In a final system these interrupts will not be used.
processing time of a frame is short enough (as is the case with the filterbank analysis algorithm).

There is also another circumstance where the average delay needs to be less than 50 ms. Some algorithms, such as pitch detection, need longer frames than the ‘standard’ 25 ms, and also slightly more processing time. When the average delay needs to be at most 50 ms, we can use frames and processing times of 33 ms at most. The frames used for pitch detection are 32 ms long.

The signal processing algorithms

For the tactile aid the following algorithms have been fully implemented:
- filterbank analysis
- pitch detection
- pitch detection + small filterbank
- formant analysis

These algorithms are written to be independent from the tactile display that will be used. After the signal-processing algorithms a transformation or coding takes place to display the information on the used tactile display.

* Filterbank analysis

The filterbank analysis is based on the Fast Fourier Transform. Since the ADSP-2100 has special features for computing a FFT, it needs relatively little time for processing. A 512 point FFT uses about 11 ms. This means that it is possible to use overlapping frames of 50 ms each. The result of the 512 point FFT is a frequency spectrum of 256 small frequency bands, each with a bandwidth of about 20 Hz. Of course this many frequency bands cannot be just as accurately displayed on a 24 times 6 display. The 256 bands are compressed to 24 bands, while the computed resolution of 16 bit needs to be compressed to 6 bit.

In the module where the data to be displayed are generated, this compression is performed linearly, logarithmically, or in any other way that one considers appropriate. For the tactile aid the frequency bands
are compressed linearly, resulting in bandwidths of about 200 Hz, while the energy in each band is compressed logarithmically.

* **Pitch detection**

The pitch detection algorithm uses frames of 32 ms. These longer frames are chosen for retaining more information about the lower frequencies in the processed window. Since the pitch detection algorithm needs more processing time than \(\frac{32}{2} = 16\) ms, it is not possible to use the technique of overlapping frames, unless the frames overlap e.g. only for \(\frac{1}{3}\). The algorithm needs approximately 80,000 cycles, which is equal to 20 ms. The result from the pitch detection algorithm, is information whether the speech signal is voiced or unvoiced, and if it is voiced, it gives the pitch frequency. The pitch frequency is denoted as an integer from 1 to 64, indicating a frequency range in which the true pitch will lie. This integer is transformed to an integer from 1 to 24 when only pitch is presented. When some frequency information is also offered (see next section), it is compressed to an area of 1 to 16.

* **Pitch detection + small filterbank**

As stated in the previous section, the pitch detection algorithm does not use all the time available within one frame. The spare time can be used to perform some other type of processing, whether related to pitch detection or not. In our case we have added some information about the spectrum. It appeared possible to perform a fourier transform in the remaining time. Using the technique explained under filterbank analysis, the two results are displayed together. The two algorithms combined use approximately 100,000 instruction cycles for voiced speech, which is equal to 25 ms. For unvoiced speech the algorithms need approximately 70,000 instructions or 17.5 ms for 32 ms of speech, because the program does not need to find the pitch frequency.

* **Formant analysis**

The formant analysis implemented in the tactile aid is based on the Robust Formant Analysis method [Willems, 1987] (see chapter 3). This technique works best, and is the easiest to implement, with floating-point arithmetic (FPA). In this algorithm single-precision, two-word floating-point numbers are used. One word denotes the
mantissa (or fraction) and the sign of the number, while the other word denotes the (signed) exponent. Since only 16-bit words are used, the greatest number that can be denoted is still (slightly less than) 32768, which is just as much as can be denoted in a 16-bit signed integer. The smallest number greater than zero that can be denoted is $\frac{1}{65536}$. The advantage of using FPA is mainly the higher accuracy, the possibility of denoting fractions easily and the greater (logarithmic) range for denoting numbers.

The time needed for the RFA varies slightly, depending on the number of iterations needed before the formants are found. Measured values for the processing time are between 15 and 19 ms.

The fact that RFA uses an approximation algorithm, that finishes only when a certain (residual) value ($X_r$) is small enough, and that the floating-point numbers, used in the approximation algorithm, have a finite precision, means that one has to ensure that the processing uses only a finite time. It is possible that because of the numeric imprecision the value $X_r$ cannot be reached. This problem can be solved by using floating-point numbers with a greater accuracy, yet this would decrease the speed of operation. Another way to solve this problem is by incorporating extra stopping criteria. One criterion is that the approximation should stop when the approximation value has become stable. In this case it is of no further use to continue, since the program has reached its maximum accuracy. Another criterion is to stop when a fixed number of iterations has been performed. It appeared that iterations that find their end-value reach this value after 3 to 8 iterations. For this reason, we have decided to stop the approximation after 10 iterations, assuming that the end-value will be reached as accurately as possible, but that the approximation value still 'wanders' around in a small range around the end value. In this way it is assured that the program is finite, and that it will end within the specified processing time.

The original Robust Formant Analysis algorithm also computes the bandwidth of each formant. It appeared that the original RFA program needs about 4 times as much computing time for computing the bandwidths as for finding the formant frequencies. Since bandwidth will probably not add very much extra information to a possible user of the tactile aid (and in literature no clear examples of tactile aids were
found where the bandwidth is presented), and since computing the bandwidth is an algorithm that uses much processing time, this has not been implemented.

The algorithms in practice

All four implemented algorithms have been tested in three ways. First the algorithm is checked to be free of syntax errors. This means that the simulator generates no errors or warnings of any kind during the simulation of the program. Also it is checked that the program on the THA-system never enters an infinite loop. The program has to generate new data for the tactile display constantly.

Next the program is tested on the simulator. On the simulator it is very easy to access all registers of the DSP and the output of the algorithms. Signals with known parameters are used as input, while the output is checked. For filterbank analysis e.g. (co-)sinewaves and blockwaves give a fast indication about the correctness of the algorithm. For the robust formant analysis test-values were available for checking the correctness [Willems, 1988]. The discrepancy between the obtained output parameters and the given parameters was less than 0.05%.

Finally the program is run on the THA and checked again (mainly using the output on the visual display), while known signals are fed to the system.

Only one algorithm—namely the robust formant analysis—could be used on subjects, for validating the device. Figure 6-13 shows 4 seconds of speech, where the formants F1 and F2 are plotted, computed both by the IPO analysis program and the THA-II system. The latter values are obtained just before they will be coded into the tactile patterns by the algorithm. This figure shows that the results from the THA-II matches the results from the IPO software quite nicely. The differences that are present can be explained by the fact that the THA-II uses 10-bit samples, while the IPO-software uses 12-bit samples. Also the internal accuracy of the real-variables in the THA-II is smaller and the time between the computation of new formants is longer.
Fig 6-13: The result of the analysis of 4 seconds of
speech. Shown are the formants $F_1$ and $F_2$ from both the IPO
analysis software (the dashed lines) and the THA-II software (the
continuous lines). The upper curve represents the energy of the
speech. It can be noticed that the resulting formant frequencies from
the IPO program are set to zero for silent periods, while the THA-II
software does not (see for $t < 0.2$ sec and $t = 1.9$ sec).
(Sentence: “Er was eens een moerasvolkje, dat woonde op de...”)

The differences that are currently present—not considering the periods
with low energy— are at most about 5% and will result in variations of
at most two places on the tactile display, including the inherent
inaccuracy of digitizing (the bars presenting the formant frequencies
can be shifted up or down two rows). The inherent inaccuracy of
digitizing is one ‘digit’, which is in this case one row. It is not known
whether this shift of two rows will have much influence on the
usefulness of the tactile supplement. Probably this influence will be
very small. For example: a variation of about 5% in the formant
frequencies does not usually result in the (auditory) perception of
different phonemes [Rabiner & Schafer, 1978].

Suggestions for the future

We have just seen how the experimental tactile aid (THA-II) has been
designed. In this section we shall discuss possible improvements
concerning the size, the energy consumption and the performance of
the tactile aid. Some of them were encountered when working with the
system. Other suggested improvements were already known when the
system was built. However, owing to the experimental nature of the
system we have chosen not to implement these improvements yet,
since they could make the system more expensive and less flexible.
These modifications would not improve the performance of the system
for the experiments we wanted to perform, yet they can improve the
performance when the system has to be evaluated ‘in the field’.

First some methods to reduce the power consumption will be dis­
cussed. Next techniques to reduce the size of the system will be
presented. And finally we shall focus on suggestions for various parts
of the system.

Reducing the power consumption

The system described in this thesis is designed with a minimal power
consumption in mind. It was not the prime goal to minimize fully the
power dissipation at this stage. However, since it might prove difficult
to reduce it later, and since an evaluation ‘in the field’ requires a system
that is more or less portable, it did put some restrictions on the system.
Thus only CMOS components were used, since they only consume
energy when the internal switches (transistors) change state. It was also
decided to operate the system on the lowest possible working fre­
quency. The lower the frequency with which the transistors in the
CMOS components switch, the less the energy consumed. Of course
the system has to operate at a certain minimum frequency, in order to
complete the signal-analysis fast enough for the system to work real-time.

It will prove to be difficult to reduce the energy consumption further by reducing the clock frequency. It might even be necessary to increase this frequency somewhat when the algorithms are changed and require more processing time. To reduce the energy consumption, we shall have to concentrate on the parts that use most power.

It appeared that under normal conditions no specific part could be found that used most energy. Only the RAM used a relatively high current when data were being continuously written to it. The only way to reduce this current is by using components with a higher maximum working frequency (since their internal capacitors are smaller). The same holds for all other digital components.

* Stopping the system

The best method to reduce power is to switch the system off. This might seem an obvious statement; however, a system that is able to switch itself off when it is not in use (i.e. when no information can be offered, e.g. owing to signals with a very low input level), does save energy. In various communication aids for the disabled this possibility proved to be very useful [Waterham, 1989], [Deliege, 1989], [van der Krol, 1988], [Leliveld, 1989]. Most of these aids were devices that produce, rather than analyze, speech. Yet, with some special adaptations and under special circumstances, it is possible to switch off the tactile hearing aid.

This idea is relatively easy to implement, when one leaves the pre-amplifier switched on. It is possible to add a comparator that monitors the signal from this amplifier. Whenever the output signal is below a defined threshold for a certain time (one does not want the system to switch off every time the signal goes below the threshold, which happens every time the signal crosses the zero-axis), this comparator tells the system to go to sleep (or power-down). When the signal amplitude rises above the second threshold the comparator reactivates the system.
To perform this switching as reliably as possible, it is wise to let a processor monitor the task. In this way it is possible to add some intelligence to the system, e.g. to decide when a ‘silent’ period is real silence, or to prevent switching on and off continuously. To add this task to the DSP’s would be rather difficult, in the first place because the processor itself needs its time to analyze the signals, and secondly because we want to switch off this processor, to save energy. For this reason we have thought about adding a microcontroller to perform this task [Mies, 1988]. Figure 6-14 shows the basic scheme of this set-up. The microcontroller is able to control both the complete DSP with peripheral components (including e.g. the ADC) and the microcontroller itself. This means that the microcontroller is able to put itself in a state where it uses only very little current (appr. 0.5mA when the controller operates at 4 MHz.) [Intel, 1988], while it can start working again almost immediately, either by external or internal interrupts or timers.

Also it is possible to let the microcontroller communicate with the DSP. By doing this several things can be accomplished. To start with it is possible to let both the microcontroller and the DSP decide when
the system can go into the power-down mode. It might be possible that
the microcontroller decides that the incoming signal is below a pre-
defined threshold, while the DSP can still detect some useful informa-
tion, even though the amplitude of the signal is low. In this case the
DSP should be able to tell the microcontroller that going power-down
is not correct. Next it is possible to let the microcontroller perform
some of the tasks that are done now by the DSP, such as coding the
extracted information into usable patterns for the tactile display. When
the DSP is relieved of this task, more time becomes available for its
ture task: processing signals. Finally some sort of error detection can
be implemented, where one processor checks the other processor now
and again. In this way (software) lock-ups can be corrected by the
system itself. One could then say that each processor is the watch-dog
of the other.

Although the method described before for stopping the system has
been built and partly tested [Mies, 1988], we have decided not to
implement the extra microcontroller yet. The main reason for this, was
that it would make the system more complicated, while not much
would be gained for use in our experiments. The system as we have
used it was an experimental system for evaluating the signal analysis
and pattern coding on the skin and was not designed to save as much
energy as possible. Another reason for not yet implementing the
microcontroller is to limit the size of the system, which will be
described in the next section.

Reducing the size of the system

It will be clear that a device designed to be worn almost all day should
not be unnecessarily clumsy. For this reason it will be necessary to
reduce the size and weight of the system (both for a prototype and for
a model for evaluation ‘in the field’). One method of reducing the size
and weight has been seen in the previous section. By reducing the
power consumption, and using a single-voltage power supply (instead
of a bipolar supply), it is possible to reduce the number and size of the
batteries, that supply the system’s power. Batteries have a relatively
high weight per volume, because of the dense materials of which they
are made. So, saving on batteries means saving on weight.
Once we have reduced the power consumption, we can try to reduce the size of the system, by using fewer components, smaller components, or placing the components closer together.

* Using fewer components

The system as described in this thesis contains as few components as we thought possible (using standard components that were available at that time). However, since new components become available virtually every day, it should prove to be possible to reduce the number of components. One possibility is the use of a DSP that has integrated more functions, such as program and data memory. Recent DSP’s from the ADSP-21XX family [Analog, 1988] have these features incorporated. Even a DSP with a built in 16-bit AD-converter will be available soon. This means that almost half the components of the current system can be integrated in one IC.

A second possibility to reduce the number of components is the use of dedicated IC’s, such as ASIC’s (Application Specified Integrated Circuits). The only disadvantage of these components is their price. Only when 1.000 or more IC’s are ordered, these techniques become financially attractive. For an experimental system the prices of ASIC’s are highly unattractive.

* Using smaller components

Much space can be saved when the current components are replaced by smaller equivalents. For experimental purposes standard IC’s (so-called DIL or Dual In Line IC’s) and components (such as resistors and capacitors) were used, to facilitate measurements in the system or changing parts of the system. When the design has become more permanent, it is possible to use SMD (Surface Mounted Devices) components. Almost every component used in the system is also available in SMD technology which can reduce the volume to a tenth or less of the prototype. When SMD components are used, it is possible to make much smaller printed circuit boards (PCB’s).
* Placing components closer together

A last step in reducing the size of the system is to group the components more compactly on the PCB. In the current system this has not been done, for the same reason that SMD’s were not used. When a PCB is designed where all components are placed as near to each other as possible, it will be necessary to use a multi-layer PCB. Attention should also be paid to cross-talk between the digital and the analog signals when the tracks on the PCB are close together.

Combining all three possibilities to reduce the size of the system, it should be feasible to produce a system the size of a modern walkman. The only problem is the power supply. The batteries needed to operate the system would be at least as large as the system itself (assuming that the power dissipation cannot be reduced much further).

This may still seem to be a large system, compared to modern (audio amplifying) hearing aids, but it is roughly the size of the first electronic hearing aids. Nowadays millions of people seem unperturbed by a piece of portable audio equipment of this size. It is even considered normal.

Hardware suggestions

We have seen how the size of the system and its energy consumption might be reduced. In this section several parts or components of the tactile aid will be discussed that could be improved. We shall walk through the system from the input to the output.

* The microphone

At present we use a simple dynamic microphone to pick up the (speech) signal. An audio input is present parallel to the microphone input, e.g. to use recorded sounds as an alternative input signal for the system. The separate audio input proved to be very useful for training purposes, and should be maintained in a future system. The microphone can be replaced by a better type; for training purposes a unidirectional microphone (which is sensitive for sounds from only one direction), while
Technical description

for use in everyday situations a microphone that is slightly less directional can be used. In a normal everyday situation it will be useful if the system gives an indication that a sound is present, no matter where the sound originates from. However when someone is talking to the user, every sound from the environment will make the result from the analysis of the speech-sounds correlate less with the (desired) speech signal. This again will make it more difficult for the deaf user of the tactile aid to follow the conversation.

The user cannot carry around a hand-held microphone. Thus a small microphone needs to be used that can be worn e.g. on one’s clothes. Another possibility is to use the same kind of microphone as in conventional hearing aids. This would have several advantages:

• the microphone (put in a case that can be placed behind the ear) is commercially available, and can be used with no or only small modifications;
• the microphone is so small that a user will hardly notice that he or she is wearing it;
• the microphone is directed to sounds in a ‘natural’ way. By looking at the origin of the sound, the microphone picks up the desired sounds under the most favourable conditions; sounds from other directions will be slightly suppressed;
• if people see this type of microphone, they will know that the user has a hearing problem. This can have the advantage that people speak more clearly, articulate better and look at the listener.

* The (pre-)amplifier

Since the energy level in sound signals can vary considerably, it can be worthwhile to use an amplifier with automatic gain control. In this way it is possible to input sounds with a low and a high amplitude to the DSP with about the same (relative) accuracy. To compensate for possible signal-processing errors, caused by step-wise changes in the gain, the DSP needs to know the gain factor. It also has to know the gain factor for presenting the amplitude of the sound signal correctly to the user. To increase the dynamic range of the input signal, it is also possible to modify the analog to digital conversion. This will be discussed in the next section.
* The analog to digital converter

The current system uses a 10-bit A-to-D converter. The main reason for applying this converter was the voltage required by the available converters. The highest resolution available for ADC’s that worked on a single voltage was 10 bit. For our purposes this resolution gave good results. When the tactile aid is used in situations where more noise is present and the signals have a greater dynamic range (i.e. ‘in the field’), it will be useful to enlarge the dynamic range of the ADC. One method of doing this is the use of an automatic gain amplifier, as discussed above. A second method, which seems easier to implement (also in software) is to use an ADC with a higher resolution. Converters often used in speech-analysis have a resolution of 12 to 14 bits. This results in an increase in dynamic range of 12 to 24 dB, compared to the currently used 10-bit ADC [Oppenheim, 1978]. A third method is to use a logarithmic ADC (or logarithmic compression of the signal before it is converted) [Rabiner & Schafer, 1978]. In this way it is also possible to increase the dynamic range by about 12 dB. In the software for the DSP, this logarithmic coding needs to be compensated for.

* The digital signal processor

The ADSP-2100 performed well in the tactile aid. We believe that there is no direct need for using another type of processor yet. When the system needs to become smaller, it is possible to use a new version of the ADSP-2100, as discussed before in the section ‘Reducing the size of the system’. Owing to the power and size restrictions, it will be difficult to use DSP’s which are more powerful, such as 32-bit processors, or processors that can operate on a higher frequency. Only processors that have more powerful instructions might give better results, since they would need less instructions to complete a task, which would mean that they could operate on a lower frequency.

* Suggestions for the tactile display

As mentioned in chapter 5, the tactile display that was used for the experiments cannot be the definitive version of the tactile hearing aid. The display needs to be worn on a place other than the finger or hand. Places which are most sensitive for vibrotactile stimulation are the
Technical description

finger tips, the lips, the nose and the hand [Schiff & Foulke, 1982; Sherrick & Cholewiak, 1986; Loomis & Lederman, 1986]. However, we have chosen not to use the hand or fingertips, and sites on the face are quite impracticable too. This means that we need to choose another site, which will be less sensitive. Possible places on the skin can be the arm, the upper leg, the abdomen and the back. For an experimental display we have chosen the upper leg and the lower arm as possible candidates.

The new vibrotactile display is made of piezo-electric vibrators; in our case so-called bimorphs. These are strips of a thin conducting layer with on both sides a layer of ceramic material. When a voltage is applied to the bimorph, it will bend either up or down, depending on the polarity of the applied voltage. One end of the bimorph is fastened on the base of the display, while on the other end a small rod is glued,

Fig 6-15: Bimorph on the skin (schematic).
When not active, the rod rests lightly on the skin. The ring around the rod has two functions: 1) to prevent cross-talk between vibrators, and 2) to prevent the user from pressing too hard on the rod, which would diminish the vibrotactile function. In the experimental display, the bottom plate (on which the vibrators are attached) realizes the function of ring.
which is in contact with the skin, through a hole in the base (see also appendix 3). Figure 6-15 gives a schematic representation of such a bimorph, touching the skin. The ring around the rod represents the base or chassis of the tactile display. The new experimental vibrotactile display consists of 24 bimorphs, arranged in a matrix of 6 times 4 rods [Lemeer, 1990]. The construction drawings, the sizes and photos of this display can be found in appendix 3.

This array has been tested on its functionality on both the upper leg and the lower arm [Lemeer, 1990]. No apparent differences were found in the capability of perceiving tactile patterns and localizing stimuli on these extremities.

The method of making a tactile display with bimorphs proved to work quite well [Lemeer, 1990]. As long as no smaller mechanical stimulators can be found, bimorphs appear to be the best solution. To make the system wearable, this tactile display needs to be improved. The energy consumption, the size and the weight of the display described are still not as good as might be. Other methods, such as electro-stimulation or laser stimulation, can make the display much smaller, but they have certain disadvantages as we saw in chapter 5.
Technical description
Preliminary tests

This chapter presents details of the first experiments performed with the tactile hearing aid. The results of the first measurements, comparing recognition scores of a number of (Dutch) sentences, both with and without tactile aid, will be presented. It was necessary to train the subjects before these results could be obtained. For this purpose a three-stage training was used. Recognition scores after a very short period of training and after a longer period of training were obtained.

Both hearing and deaf subjects were used for the experiments. In all cases the use of the tactile aid appeared to influence the recognition scores positively.

The system used for the experiments

Before we begin to describe the experiments, we shall briefly review the system as used during the experiments: see also figure 7-1 (a more extensive description is given in the previous chapter). The hardware consists of a system based on a DSP. The input for the DSP is a signal from a microphone or audio-input (for using 'standard- speech', e.g. from a recorder). The audio-signal is low-pass filtered at about 3500 Hz with a 7th-order elliptic switched-capacitor filter. Both amplification and offset can be adjusted by variable resistors.

Once the analog signal has been filtered, it is digitized by a 10-bit sampling analog to digital converter at 10 kHz (10,000 samples per second). It then enters the DSP, where the signal is processed. The results from the signal-processing are coded for the tactile display. For training purposes, a visual display shows the data that can be felt on the tactile display. This display could be used for feedback by the subjects in cases where they are not sure what they are expected to feel.
Preliminary tests

Fig 7-1: Schematic diagram of the tactile aid

Fig 7-2: Formant display as used in the experiments
During measurements the display could not be seen by the subjects. For the measurements the first two formants were presented, together with amplitude information (see also fig. 7-2).

Methods and materials

Our experiments were preliminary, as we wanted to get an impression as to whether the system really can help to improve the perception scores. Finally we wanted to find out how quickly one can learn to make (at least some) use of the offered information.

Subjects

For the experiments we had the cooperation of six subjects. Three subjects were postlingually deaf (i.e. their hearing loss of over 90 dB occurred after acquisition of a spoken language). The other three subjects have normal hearing.

There are several reasons for using both hearing and deaf subjects.

- Hearing subjects were used because of:
  - their ready availability and
  - their presumed lack of speech-reading ability (which can give the researchers an insight into the effectiveness of the device for those with hardly any speech-reading experience who are suddenly afflicted with deafness).

- Deaf subjects were used because:
  - the final device is to be used by deaf people. It therefore seems logical to evaluate the system on possible future users;
  - it seemed useful to evaluate the device with people who have experience with speech-reading; the deaf subjects were presumed to have this experience;
  - it is difficult to predict fully the demands of possible future users, other than by asking them. However, to be able to answer questions about how the system should be, it is necessary to have some ideas about such a system. During and after the experi-
Preliminary tests

ments the deaf subjects were asked for comments on the current system, e.g. whether certain parts could be improved or changed (still bearing in mind that the current system was only for experimental purposes).

Table 7-1 gives some information about the hearing subjects. No auditory or visual defects (either with or without glasses) were known. Also no deficiencies in the tactile sense were known or detected during the experiments. One of the subjects had a little experience with lip-reading and with tactile perception of information, partly because of prior experiments.

Table 7-2 gives some information on the deaf subjects. Since the

<table>
<thead>
<tr>
<th>Table 7-1. Hearing subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Subject H1: Male 27</td>
</tr>
<tr>
<td>Subject H2: Male 23</td>
</tr>
<tr>
<td>Subject H3: Male 23</td>
</tr>
</tbody>
</table>

subjects had to repeat the sentences offered during training and measurements, we preferred deaf subjects with a good quality of speech. This facilitates checking of the given responses, as doubts about what is replied will be less frequent. An extra advantage of a good quality of speech is that communication with the subjects will be much easier (and faster). Unfortunately the speech quality of the prelingual deaf is
Table 7-2. Deaf subjects.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Remaining hearing ability</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject D1: Female</td>
<td>28</td>
<td>appr. -100dB, hearing aid</td>
<td>formerly hard of hearing</td>
</tr>
<tr>
<td>Subject D2: Male</td>
<td>25</td>
<td>appr. -90dB, no hearing aid</td>
<td>formerly normal hearing. Mountaineer</td>
</tr>
<tr>
<td>Subject D3: Male</td>
<td>46</td>
<td>appr. -100dB, hearing aid</td>
<td>formerly hard of hearing</td>
</tr>
</tbody>
</table>

frequently poor [Maassen, 1986]. For this reason we preferred to use postlingual deaf subjects.

One subject had normal hearing until his 20th year, the other subjects were hard of hearing since birth and have attended special schools for the hard of hearing. All subjects had been deaf for at least 2 years. Two subjects used a hearing aid (which did not allow them to understand speech, nor to hear sounds below a certain level). One subject practised mountaineering as a hobby, which made his hands rather callous.

Material: sentences

Short sentences were used for training and obtaining the recognition score (see appendix 1). This speech material (selected by R. Plomp, [Plomp & Mimpren, 1979]), comprised twenty lists of thirteen short sentences (of eight or nine syllables), avoiding words of more than three syllables. The sentences represent conversational ‘everyday’ speech, without too much redundancy (as in proverbs: these are not
Preliminary tests

used in the lists). Each list contains about 300 phonemes, and these phonemes are roughly the same for each list [Plomp & Mimpen, 1979].

We selected these sentences, because:
• they have been statistically evaluated [Plomp & Mimpen, 1979; Breeuwer, 1985];
• they had used before, for a study to determine what supplementary information to speech-reading could be offered best (in this study the extracted information was presented acoustically) [Breeuwer & Plomp, 1984, 1985, 1986]. In this way both investigations are easier to compare;
• they were available on video tape spoken clearly and correctly articulated by a (female) speech-trainer;
• they are grammatically correct
• no information about the sentences can be extracted from other sentences in the same set.

The last two reasons were considered to be important for a correct interpretation of the use of the aid. The aid should be used in an everyday situation, where (most of the time) sentences will be used that are more or less grammatically correct. This means that some of the information can be obtained using the redundancy which is always present in grammatically correct speech. The tactile aid should give extra information beyond the ever-present information that is available owing to the redundancy in normal speech.

On the other hand we did not want to overplay this redundancy. This is why no a priori knowledge about the sentences should exist. Translated into everyday terms this can be compared to changing the topic of discussion, or an inconsequent sentence brought into a conversation.

Furthermore these sentences hardly contain any information besides the actual spoken message. This means that the face of the speaker gives away virtually no clues about the message, nor are the hands or rest of the body visible, which might convey extra information. The only useful information that the subjects can perceive from the video tape is the movement of the lips, jaws, cheeks and throat.
Method

The experiment was set up as a number of sessions of about one hour each. Concentrating on the tactile sense is something one is not accustomed to doing and poses problems in concentration. Further, the tactile sense shows adaptation effects for (vibro-)tactile stimulation (see chapter 5). Also speech-reading itself is quite intensive. For this reason the volunteers did not work under pressure and time was reserved for coffee or other drinks, for some small talk, or for suggestions about the experiment or the system, both before and during an experiment. The effective time for the experiments (measurement or training) turned out to be between 30 and 40 minutes.

* Set-up

The set-up for the experiments is sketched in figure 7-3. Here the video recorder, monitor, microphone and a tape-deck are added to the system. The tape-deck is used to record the responses from the deaf subjects.
Preliminary tests

during the tests, or to mask the sound from the tactile display — during both training and measurements — for the hearing subjects. It proved that the sound from the tactile display could also pass on information. Masking of the sound from the tactile display is effected by offering noise to the (hearing) subjects via headphones. Since the sound that the tactile display produces will obviously be related to the speech signal, it could conceivably offer more information to the hearing subjects — who are used to perceiving speech data acoustically — than the tactile information itself⁠¹, especially when one is not yet experienced in perceiving tactile information.

Fig 7-4: Global synopsis of the scheme for training and measurements. Depending on the training speed of a subject, this scheme could be altered slightly.

---

⁠¹ During the tests the experimenter (who had seen the sentences before) was able to recognize almost every sentence using only the visual information combined with the acoustic information from the tactile display. Without the acoustic information he was able to recognize only about half as much.
* Training

We give below a conspectus of the methods used to train the subjects (see also figure 7-4).

Since most people are not accustomed to perceiving (artificial) information via the tactile sense, it was necessary to train the tactile sense first. During prior experiments investigating the tactile sense [Wang, 1990] and in literature about tactile perception, e.g. [Katz, 1925; Spens, 1981; Schiff & Foulke, 1982; Sherrick, 1984; Loomis & Lederman, 1986], it can be observed that at first it is rather difficult to perceive artificial tactual information. However, once the tactile sense has been trained to perceive this information, it is much easier to perceive other artificial tactile information too [Spens, 1981]. Although it will still be necessary to train the tactile sense to perceive novel information, the training could be much shorter, since one is usually able to recognize the pattern that is offered to the skin.

For this reason we started by presenting patterns to the skin that are already familiar to the subject (although not as tactile information). The patterns used in this training were ten capital letters [Wang, 1990] viz:


Fig 7-5: The ten characters used for training the tactile sense of the subjects. The fat dots represent vibrating rods on the tactile display from the Optacon, while the light dots represent rods that do not vibrate.
Preliminary tests

It appeared that these letters could be discerned without too much difficulty [Kume & Ohzu, 1986; Wang, 1990]. Figure 7-5 shows the way these patterns were presented on the tactile display.

For training purposes it was possible to present these characters in two ways. One method was presenting the complete character on the display during a certain period of time (in our case during 0.5 second); the other method was to shift the pattern into the display step by step, and when the character was completely on the display, to shift it out again [Wang, 1990] (in our case the time between these steps was 0.1 second). Figure 7-6 shows an example of the two methods, both presenting the character S. All subjects preferred to train using the shifting mode of presentation. Although this letter training has no connexion to the presentation of formants, the shifting mode is prob-

![Diagram](image)

**Fig 7-6:** Graphic representation of two possible methods of presentation of the characters on the tactile display: a): static presentation. Each character is offered for 0.5 second; b): shifting presentation. Every 0.1 sec. the character is shifted one position (one column) to the left.
ably closer to the formant representation than the static method. Both
the formant and the shifting letter presentation are built up of changing
patterns.

At the beginning of the experiments this training was performed every
session. After four sessions, the training was repeated every other
session. At the end of the experiments the character recognition score
was determined, to see if a correlation might exist between this ability
and an increase in the speech-reading score with tactile supplement.

To train these characters the ‘simpler’ characters (I, L and O) were
offered first. Once the subject could distinguish these three characters
well enough (i.e. about three correct responses per four tests), the other
characters would be offered as well. The subject was informed each
time he was to be offered a new character (i.e. a character not pre-
viously felt). The new character was first offered tactually. Next the
character would be shown (on the LED-display) and offered tactually
to the subject again. From then on the character would be used in the
recognition-training.

To train a subject, a random character was offered (from the set of
characters that were already known to the subject). The subject was
asked to name the character. If the answer was correct, the subject was
told so and another character would be offered. If the answer was
incorrect, the same character would be offered again. If the subject still
could not identify the character, the correct answer would be told (or
shown). The character (which was now known to the subject) would
then be offered a third time.

At the start of a new session with recognition-training, each character
would be offered to the subject both tactually and visually, to ‘refresh’
his memory. To measure the character-recognition score at the last day
of the experiments, the ten characters were offered in random order.
In total each character would be offered three times, placed in a random
sequence of 30 characters. The subject had to give an answer every
time a character was offered. We shall call this recognition score: $R_c$.

When the subject had obtained some experience in the recognition-
training, he was exposed to perceiving formant information. During
two or three sessions (which were in fact the second and the third
Preliminary tests

session, see fig. 7-4) only three phonemes (the /OE/, /UU/ and /AA/ ) were presented on the tactile display (without visual information: neither the speaker's face nor the optional visual display were visible), while the subject was asked to distinguish between the phonemes. Training was largely similar to the character recognition-training: a phoneme was offered (as formant information) and the subject had to learn to identify it. When correct, another phoneme was offered. When the response was incorrect, the formant information would be offered again. If the response was still incorrect, the correct phoneme would be revealed, and offered again.

Since the recognition of these pure phonemes is a relatively easy task (the tactile patterns are quite distinct), the training took only about a quarter of an hour, spread out over two or three sessions. After this time the recognition of these phonemes was usually at least three out of four correct. It was decided not to train more phonemes. The main reason for this is that in running speech hardly any single phonemes will be found. The perception of single phonemes is vastly different from the perception of the same phonemes when they are presented for a much shorter period of time and when they are not constant while they are being presented. Although a phoneme might be thought to be constant, it is clearly not. Depending on the phonemes before and after it, the acoustical form of a phoneme can change. A clear example is the phoneme /H/: in the word 'hello' the /H/ is acoustically quite different from the /H/ in the word 'how' (in fact the mouth has already completely shaped to form the next phoneme, at the moment when the /H/ sound has only just started).

The next step in the training is the presentation of running speech together with the extracted formant information. Here too the information is presented on both the tactile and the visual display, but the subjects could use only the tactile display for the supplementary information to speech-reading. The visual display could be used by the experimenter, e.g. to verify the correct functioning of the equipment.

During this stage of the training, the subject has to train himself to find out which information represents what. To enable the subject detecting the correlation between what is felt and what is said, the subject could see a speaker speaking a sentence on a TV monitor. The processed speech is presented simultaneously to the skin. When the sentence was
spoken, the subject had to repeat what he thought he had seen/felt. When the response was not completely right, the sentence would be shown again on the monitor. The only feedback would be that the sentence was either completely wrong, or partially wrong.

After showing the sentence a second time, the subject was asked to reproduce the sentence again. If it was still not recognized correctly, the experimenter would tell (or if necessary show on paper) the subject the correct sentence. Once the correct sentence was known, the spoken sentence would be shown on the monitor (and presented on the tactile display) again. If the subject still could not recognize the sentence, even though he knew what had been said (e.g. because he thought the sentence was slightly different), the sentence would be shown again.

This method of showing the sentences, asking to repeat them, and when necessary correcting the responses, appeared to work well, as can be seen when the results are discussed. Possible improvements for the method of training will be discussed at the end of this chapter.

To obtain a more or less objective number expressing the degree of correctness of a perceived sentence, the number of correctly perceived phonemes was counted. For example, the Dutch sentence “De bal vloog over de schutting” has the following 21 phonemes:

\[
/\text{D}/ /\text{e}/ /\text{B}/ /\text{A}/ /\text{L}/ /\text{N}/ /\text{O}/ /\text{G}/ /\text{V}/ /\text{O}/ /\text{e}/ /\text{R}/ /\text{D}/ /\text{e}/ /\text{S}/ /\text{G}/ /\text{U}/ /\text{T}/ /\text{I}/ /\text{ng}/
\]

If a subject thought to have perceived “De paal stond boven de schuur”, this reply was split into phonemes:

\[
/\text{D}/ /\text{e}/ /\text{P}/ /\text{A}/ /\text{L}/ /\text{S}/ /\text{O}/ /\text{N}/ /\text{B}/ /\text{O}/ /\text{e}/ /\text{N}/ /\text{D}/ /\text{e}/ /\text{S}/ /\text{G}/ /\text{U}/ /\text{R}/
\]

Next the number of correctly perceived phonemes was noted in the original sentence (paying attention to the sequence of the phonemes in the original sentence and the repeated sentence):

\[
/\text{D}/ /\text{e}/ /\text{B}/ /\text{A}/ /\text{L}/ /\text{N}/ /\text{O}/ /\text{G}/ /\text{O}/ /\text{e}/ /\text{R}/ /\text{D}/ /\text{e}/ /\text{S}/ /\text{G}/ /\text{U}/ /\text{T}/ /\text{I}/ /\text{ng}/
\]

In this sentence 10 phonemes were perceived correctly. For this sentence we would say that the recognition score is:

\[
\frac{10}{21} \times 100\% = 48\%.
\]

No special attention is paid to the number of incorrect phonemes in the replied sentence. This means that a reply such as “De ballon vloog over de schutting” is considered as 100%
Preliminary tests

correct too, although two extra phonemes had been incorporated into the reply:
/D/ /e/ /B/ /A/ /L/ /O/ /N/ /V/ /L/ /O0/ /G/ /O0/ /V/ /e/ /R/ /D/ /e/ /S/ /G/ /U/ /T/ /I/ /ng/

During the experiments the phenomenon that more phonemes were found than were present in the original sentence hardly occurred. Almost every phoneme in the replies could be fitted into the original sentence either as correct or as wrong. Missed phonemes were also considered to be wrong.

The recognition scores of the speech-reading experiment will be called \( R_s \) (for speech-reading alone) and \( R_{st} \) (for speech-reading supplemented with tactile information).

* Measurements

During the training sessions the recognition scores were not collected objectively, as opposed to the measurement sessions. However, improvements in the speech-reading ability with tactile supplement were observed (subjectively) by the experimenter. When the speech-reading score with tactile supplement started to rise, one could normally also notice an increase in the number of sentences that were recognized (almost) correctly. To obtain objective recognition scores two measurements were included during the period of training. The first measurement was done when the subject had only very little experience with the tactile aid (usually after about two or three training sessions during which the sentences were presented). The second (and final) measurement was made after at least ten training sessions. It was sometimes decided to add a few training sessions (at most 4 extra sessions, which would be inserted in the scheme of figure 7-4 between ‘day 11’ and ‘day 12’), especially when more time was spent on the character-recognition and less on the sentence-training (subject D2).

The sentences that were used for the experiments can be found in appendix 1. Unless noted otherwise, sentence set 6 was used for obtaining the speech-reading score \( R_s \) and sentence set 7 for obtaining the recognition score of the speech-reading supplemented with tactile information \( R_{st} \) for the first experiment. For the second experiment sentence set 12 was used for \( R_s \) and set 13 for obtaining \( R_{st} \). These sets were of course not yet known to the subjects; for the training other
sets had been used. For various reasons it was not possible to stay with the same sentences for every subject. The main reason e.g. for using a different set of sentences for subject H1 was that he had seen some of the sentences before. The sentences used for the measurement sessions were not known to the subject.

To get an impression of how the tactile sense was able to recognize certain patterns, the final recognition scores of the presentation of the ten characters was used. This measurement was done on the same day that the last set of ‘sentence-recognition-scores’ was obtained. Figure 7-4 gives a (global) breakdown of when which part of the training took place and when the measurements were done.

Results

The results of the experiments are presented in table 7-3, with $R_s$ and $R_{st}$ after a short period of training (about 1 hour with the sentences and tactile supplement) and after a longer period of training (from about 5 to 10 hours). Finally this table shows $R_c$, obtained during the second measurement. This table also shows that after training all subjects showed a rise in speech-reading recognition score when tactile supplementary information was offered.

Statistical evaluation

It should be noted that only a small number of subjects was used for the experiments. The group we used was so small that the assumptions, normally underlying statistical analysis cannot be satisfied. One of these assumptions is that the number of subjects is large enough and is more or less representative of the total population [Ferguson, 1981; Hays, 1988]. With only three subjects representing the deaf population and three subjects representing the hearing population, it will be clear that this assumption is not fulfilled. One of the consequences of this is that the obtained standard deviation and confidence interval are rather large.

However, since we are concerned with a preliminary experiment to validate the functioning of the system, we believe it was still worth-
Preliminary tests

while to evaluate the results statistically; even though the resulting confidence intervals are large, they can give us useful information. In the following paragraphs these symbols are used:

\[ R_s \] : the speech-reading score (% correct) without tactile supplement

\[ R_{st} \] : the speech-reading score (% correct) with tactile supplement

\[ R_c \] : the character recognition score (% correct)

\[ R_{dx} = R_{st} - R_s \], for measurement x

\[ R_{rdx} = \left( \frac{R_{st} - R_s}{R_s} \right) \times 100\% \], for measurement x

Table 7-3. results of the measurements

<table>
<thead>
<tr>
<th>Subject</th>
<th>First measurement</th>
<th>Second measurement</th>
<th>Char.</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rs</td>
<td>Rst</td>
<td>notes</td>
<td>Rs</td>
</tr>
<tr>
<td>Subject D1</td>
<td>77%</td>
<td>80%</td>
<td></td>
<td>69%</td>
</tr>
<tr>
<td>Subject D2</td>
<td>15%</td>
<td>24%</td>
<td>1)</td>
<td>19%</td>
</tr>
<tr>
<td>Subject D3</td>
<td>74%</td>
<td>86%</td>
<td></td>
<td>51%</td>
</tr>
<tr>
<td>Subject H1</td>
<td>58%</td>
<td>59%</td>
<td>2)</td>
<td>47%</td>
</tr>
<tr>
<td>Subject H2</td>
<td>48%</td>
<td>45%</td>
<td></td>
<td>44%</td>
</tr>
<tr>
<td>Subject H3</td>
<td>37%</td>
<td>36%</td>
<td></td>
<td>48%</td>
</tr>
</tbody>
</table>

Notes:
1) sentence set 4 and 6
2) sentence set 12 and 13
3) sentence set 16 and 17
**Non-parametrical test**

There are several statistical methods for testing whether two samples (or sets of observations) are taken from the same population. When some statistical parameters of the original population are known, it is possible to apply a so-called parametric test. For example, when the population has a normal distribution, it is possible to use standard parametric significance tests. When the distribution of the population is not known, it is still possible to say something about the observations, using so-called distribution-independent or non-parametric tests.

The non-parametric tests can always be applied to test an hypothesis, even when dealing with a known distribution. The disadvantage of applying a non-parametric test is that the confidence levels to accept or reject an hypothesis are inferior to those for the proper parametric test. When however an hypothesis can be accepted or rejected with a non-parametric test, one can be sure that the proper parametric test will do so too, with a confidence level that is usually better (but at least the same) [Mack, 1975].

The hypothesis that we want to test for the results in table 7-3 is that there is no difference between the recognition scores of pure speech-reading and speech-reading plus tactile supplement. This is the so-called null-hypothesis. When we can reject the null-hypothesis with enough confidence, we can say that there is a good reason to assume that the speech-reading results are influenced by the tactile supplement.

A non-parametric test that can be used to test this null-hypothesis is the Wilcoxon Matched Pair Signed Rank test. To perform this test, we have to rank the differences from small to large, without regard to the sign. Next the sign of the difference is attached to the ranks, and the sum of the positive ranks, $T_+$, is computed. Using statistical tables (e.g. [Owen, 1962] or [Mack, 1975]) one can find the confidence level to accept or reject the null-hypothesis.

It appears that for the first measurement, where $T_6 = 16$, there is no reason yet to reject the null-hypothesis. For the second measurement however, $T_6 = 21$ (the maximum value), we can reject the null-hypothesis with at least 95% confidence. In other words:
Preliminary tests

after little or no training, there is not enough reason to say that offering the tactile supplementary information has a noticeable effect; after some training however, one can say —with 95% certainty— that offering tactile supplementary information has had a (positive) influence on the speech-reading score.

* Parametrical tests

We have just seen that, using a non-parametrical test, we can show statistically that offering tactile information can be considered useful. When a parametrical test can be applied we might get more information from the results. When the population has a normal distribution, statistical analysis would we easiest, especially with the available statistical programs.

It appears that samples taken from a large population tend to have a normal distribution [Mack, 1975]. This means that when the results are treated as if they are taken from a population with normal distribution, the results will be close to the true values. The recognition scores $R_{s}$ and $R_{st}$ are taken from a large population (of about 300 phonemes).

Table 7-4. Statistical evaluation of the results from the measurements. (Summary from SPSS analysis)

<table>
<thead>
<tr>
<th>Mean value</th>
<th>95% conf. int. for mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{d1}$ (D1..3)</td>
<td>8.0%</td>
</tr>
<tr>
<td>$R_{d1}$ (H1..3)</td>
<td>-1.0%</td>
</tr>
<tr>
<td>$R_{d1}$ (both)</td>
<td>3.5%</td>
</tr>
<tr>
<td>$R_{d2}$ (D1..3)</td>
<td>22.0%</td>
</tr>
<tr>
<td>$R_{d2}$ (H1..3)</td>
<td>7.7%</td>
</tr>
<tr>
<td>$R_{d2}$ (both)</td>
<td>14.8%</td>
</tr>
</tbody>
</table>
Table 7-4 gives the mean values for the rise in recognition score, together with the 95% confidence interval, computed with the statistical program SPSS-x [SPSS, 1988]. This interval indicates the area in which the mean value of the total population can be found, with 95% certainty. So, there is only a 5% chance, that the true mean value is outside the given interval. The confidence intervals show that after training the rise in recognition score is positive (with at least 95% confidence) for the combination of deaf and hearing subjects. When the relative rise is considered (see table 7-5), the 95% confidence interval is positive for both the deaf subjects and the combined deaf and hearing subjects.

Table 7-5. Statistical evaluation of the results from the measurements. (Summary from SPSS-analysis)

<table>
<thead>
<tr>
<th></th>
<th>Mean value</th>
<th>95% conf. int. for mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{R}_{rd1}$ (D1..3)</td>
<td>26.7%</td>
<td>-46.6%</td>
</tr>
<tr>
<td>$\mathcal{R}_{rd1}$ (H1..3)</td>
<td>-2.3%</td>
<td>-12.4%</td>
</tr>
<tr>
<td>$\mathcal{R}_{rd1}$ (both)</td>
<td>12.2%</td>
<td>-13.7%</td>
</tr>
<tr>
<td>$\mathcal{R}_{rd2}$ (D1..3)</td>
<td>52.3%</td>
<td>10.7%</td>
</tr>
<tr>
<td>$\mathcal{R}_{rd2}$ (H1..3)</td>
<td>16.7%</td>
<td>-25.6%</td>
</tr>
<tr>
<td>$\mathcal{R}_{rd2}$ (both)</td>
<td>34.5%</td>
<td>8.6%</td>
</tr>
</tbody>
</table>

Discussion

From the results of the experiments, we can draw several conclusions. However, we have to be careful in interpreting these results, since only a limited number of subjects has been used during a relatively short period of time.
Preliminary tests

General remarks about the data

Let us start with a review of the results from the measurements. Without even looking at the statistical evaluation, we can see some remarkable facts in table 7-3. It is easy to see that subject D2 shows a rather low speech-reading score. He is even below the results obtained for the hearing subjects. The main reason for his low score is (most probably) that this subject—as opposed to the two other subjects—had normal hearing for about 20 years. After that he became suddenly deaf. Where the two subjects D1 and D3, who were hard of hearing from birth, had speech-reading experience from a very early age, subject D2 had not. Why the speech-reading score for subject D2 is below the scores of the hearing subjects is not clear. In fact the hearing subjects all show a remarkably good ability to read speech.

Results from the first measurement

The first measurement shows mainly that the speech-reading recognition scores do not improve relevantly when tactile supplementary information is offered. Looking at the mean values, the deaf subjects show a rise in recognition score, while the hearing subjects show a slight decrease. The effect of the small number of subjects shows clearly.

Results from the second measurement

The second measurement does show a relevant rise in the speech-reading scores of the subjects. Even a non-parametrical test shows that the rise is statistically relevant (with 95% accuracy). The parametrical test also indicates that the rise $R_d$ is significantly greater than zero for the deaf subjects alone. For the deaf subjects the mean (or predicted) value of the relative rise is above 50%.

The idea that a correlation might exist between character recognition and improved speech-reading scores does not show in these results. In fact, the results appear to indicate the opposite. This result does not
indicate whether previous character recognition training leads to faster perception of the tactile supplement or not. To check this another experiment is needed.

Discussion of the results

The results from the experiments appear to show that deaf subjects are a lot faster in learning to use the supplementary tactile information. It is clear that at the stage of the second experiment the deaf subjects do make use of the supplement, while the hearing subjects appear to use this information less efficiently. At the time of the first experiment the mean values of $R_d$ show a positive value for the deaf subjects, and a slightly negative value for the hearing subjects. There is good reason to assume that the deaf subjects use the tactile supplemental information faster. Also, all deaf subjects indicated (both during training and after the last experiment) that they truly had the feeling that the tactile information did improve their speech-reading ability. We believe that the difference between the results from deaf and hearing subjects can be explained by the fact that deaf people have in general more experience with speech-reading than hearing people. When tactile information is offered to deaf people, it might distract them a little from speech-reading. But since most of them are quite used to speech-reading, this ‘tickling’ will hardly influence negatively the recognition of what is said. Also experienced speech-readers are used to using all kinds of supplementary information (such as movements made by the speaker). So, one can expect that other forms of supplementary information, which is in this case tactually offered, will also be used easily.

\footnote{In fact, one (deaf) subject once said he had the impression that he heard some of the tactually offered speech, rather than felt it. Physically it was not possible for him to hear either the speech sound—since this sound was off—or the sound from the tactile display—since it was too weak and could not be heard by the subject, even when every rod vibrated and his ear (hearing aid) was placed almost touching the display.}
Preliminary tests

Hearing people on the contrary are generally inexperienced in speech-reading. Anything that distracts an unpractised speech-reader from his task will decrease his chance of recognizing what has been said. Furthermore inexperienced speech-readers will have more difficulty in using any supplementary information (also visual); it will mean that one has to concentrate on two relatively new and therefore unfamiliar tasks. So the tactile information will at first distract the hearing subjects from speech-reading. Only when the hearing subjects start getting used to speech-reading, can they try to extract some information from the tactile supplement, something the deaf subjects could do from the very beginning of the experiments.

This means that it seems only logical that hearing subjects appear to perform less during this experiment than their deaf colleagues. Now the question may remain, why the hearing subjects in Breeuwer’s research [Breeuwer, 1985; Breeuwer & Plomp, 1986] showed an increase in recognition score so very quickly (virtually no training was needed). We suggest the following reason: since in Breeuwer’s experiment the supplementary information is offered acoustically, the nature of the ‘supplement’ is known to the (hearing!) subjects very well. The subjects hardly know any differently than that speech is perceived acoustically (though theoretically they may know that some people use speech-reading). So when one sees someone speaking and a sound can be heard that very much resembles speech, it is very likely that one tries to extract maximum information from the acoustic signal (perhaps without being consciously aware of it). The visual information will be used as supplementary information, instead of the acoustical information. One could compare this with people who are hard of hearing. They too understand speech much better when speech-reading is used as supplement (one says that “they hear much better when they can see the face of the speaker clearly”).

* Review

One goal of the experiment was to obtain some knowledge and experience for setting up further experiments for evaluating the tactile hearing aid and comparing different extraction and display methods for supplementary information. Although the method we have used for
this experiment worked well enough for obtaining these initial results, we believe that some improvements are possible.

Owing to the limited time available, it was possible to do just this one experiment with a small number of subjects. Future experiments, where various forms of supplementary information have to be compared, need to be set up with more subjects. According to the results obtained from this experiment, one could say that it is preferable to use deaf subjects. First because they have speech-reading experience, secondly because they will be the possible future users of a tactile aid and thirdly because they show a fast rise in recognition scores. However, the use of deaf subjects can carry a risk, especially when the tactile aid is used ‘in the field’ as a wearable device. It might be that the deaf subjects become dependent on the device, while it can be used only temporarily. Also changing the offered information for comparing the different methods might be very difficult once the deaf subject is used to a certain method. For a valid method of comparison, this might mean that for every new type of supplementary information new deaf subjects have to be found.

Seemingly the only disadvantage of using hearing subjects is that it is difficult to find a sufficient number with sufficient speech-reading experience. The hearing of these subjects, once found, has to be temporarily reduced to near-zero level in order to reproduce a comparable situation to postlingual deafness. Finally this experiment indicates that hearing subjects need a longer period of training before the offered supplementary information can be used by them properly. Before starting an experiment for comparing various techniques, this problem about the choice of subjects has to be solved properly.

The method of training the subjects appeared to work well (an average relative rise in the speech-reading scores of about 50% could be observed for the deaf subjects after merely 5 to 10 hours of training). For a future evaluation of the device and a comparison between different types of supplementary information, it is probably useful to develop a more systematic training, rather than offering sentences, while the subject has to learn to perceive the tactile information more or less by himself. We believe that parallels exist between a tactile aid and a cochlear implant. It will be worthwhile to investigate the training methods developed in the latter field, or methods derived from them.
Preliminary tests

Unfortunately these training methods are not as yet often described in literature.

The method of measuring the speech-reading score with the chosen sentence sets worked well. However, the sentences that we used were originally intended for testing the speech reception threshold for sentences [Plomp & Mimpen, 1979], which is an aural (i.e. acoustical) test. To measure speech-reading scores (e.g. [Breeuwer, 1985]) a video tape was made on which these sentences were spoken by a (female) speech-therapist. This tape is also being used by some speech-therapists for determining the speech-reading ability of deaf people. During the experiment we heard and we noticed that some sentences appear to be easier to recognize by means of speech-reading than other sentences. It is not known whether this fact influences the results of a complete set of sentences noticeably. During the training it was not observed that certain sets (of 13 sentences) were clearly easier to recognize than other sets. When these sentences are used for future experiments, one should try to eliminate the differences between sentence sets (if these exist), e.g. by using different sets (instead of one set) for one measurement.

* Remarks

When working with the tactile aid and with subjects, it will be clear that ideas or facts, other than those which show up statistically, emerge too. While setting up the experiments, ideas and assumptions were encountered, which sometimes appeared to be proven during the experiments, yet sometimes seemed to be contradicted. Because of the limited number of subjects most of these facts we noted can hardly be proven, yet it would be a pity not to mention them.

- One of the assumptions we heard was that learning to use the Optacon device by blind people was generally considered very difficult for people older than about 50 years. Since one of our subjects is close to 50, we thought that he might have difficulty in learning to perceive the tactile information. During the experiments the opposite seemed true since he was able to perceive the tactile information rather well after a short time of training (his character-recognition score was the highest for the deaf subjects, while the rise in his recognition score was the highest of all).
During the experiments it often happened that sentences were either recognized (almost) completely correctly, or recognized hardly at all. Sentences were less often recognized partly correctly (e.g. only the first half or the last half correct). It appeared that most subjects, and especially the deaf subjects, tended to construct new sentences that made some sort of sense and were grammatically correct. Nonsensical sentences hardly occurred. A nice example of this phenomenon is the following:

one of the presented sentences was: “De grote stad trok hem wel aan” (English: the big town rather attracted him). Since the words ‘stad’ and ‘trok’ can be recognized only with difficulty, one of the replies was: “De grote kat kon hem wel aan” (English: The big cat could surely beat him).

An effect which is probably similar to the previous one is that the good speech-readers sometimes said they recognized nothing of a sentence, although other sentences in a session were recognized completely correctly. One would expect that good speech-readers would at least recognize some clear sounds, such as /AA/ or /OO/. But even these basic mouth patterns did not occur in the replies, although the subjects knew that responding merely with sounds was just as good as responding with whole words.

**Conclusions**

Looking at the results of the experiment, we believe we can say that offering the tactile supplementary information that we have chosen helps to improve speech-reading. At the second measurement all subjects show a rise in recognition score when tactile supplementary information is offered. For the deaf subjects and the deaf and hearing subjects combined the confidence interval displays this rise too. To observe this higher recognition score, a training of 5 to 10 hours was needed. After this (relatively short) training the deaf subjects show a mean relative rise, $R_d$, of about 50%.

It seems that after 5 to 10 hours of training, deaf subjects are better equipped to use the tactile information than hearing subjects. Whether this means that deaf people should be used to compare different analysis and display methods is doubtful.
A supposed correlation between the ability to recognize tactile characters ($R_c$) and better speech-reading when tactile information is offered ($R_d$) could not be found.

The method chosen for training the subjects and measuring the speech-reading scores appeared to work properly for this preliminary experiment. More extensive experiments might need another method for training and measuring recognition scores.
Conclusions and remarks

After developing, testing, working with and evaluating the tactile hearing aid, we believe we can draw the following conclusions.

Hardware

After developing the first model of the tactile hearing aid (THA-I), some minor changes were needed. Some algorithms could not run real time in the first system, but after the speed was slightly increased and Random Access Memory was added to the program memory, these limitations were removed. The second version (THA-II, see also chapter 6 and appendix 4) proved to work well. No problems occurred during the use of the system, and every algorithm that we wanted to implement could be implemented.

The hardware used relatively little power. Although the system did not yet operate on batteries (simply because this was not yet needed), it would in principle have been possible to use them. Using six penlight batteries (to obtain the correct operating voltage) it will be possible to use the system for at least a number of hours (2 to 10 hours, depending on the type of batteries: rechargeable or alkaline).

The system can be made portable relatively easy. As mentioned before, the system was not yet fully portable, yet it was built in a small briefcase and could easily be carried from place to place.

The choice of a DSP proved to be a good one. With the use of a DSP we were able to perform several speech-processing algorithms real time, on a small system. By using the DSP on a frequency that is about 1/3 of its maximum operating frequency, this could be realized without using much energy.
Conclusions and remarks

Software

Algorithms for filterbank-analysis, for pitch-detection and for formant-analysis were implemented and tested and performed well. All these algorithms could be run real-time. It was even possible to combine filterbank-analysis and pitch-detection while still operating real-time. To validate the use of the tactile hearing aid, only the formant analysis algorithm was used in a preliminary experiment.

Displays

During the experiments we worked with a tactile display on the tip of a finger. The tactile display was an Optacon, modified in such a way, that it was possible to present patterns by means of the tactile hearing aid. The Optacon was very useful for a preliminary evaluation of the system, not merely because we could use an existing display, but also because the finger tip is normally more sensitive for perceiving various forms of tactile information.

A prototype of a novel display, that is better adapted for everyday use, has been developed. Using bimorphs it will be possible to design a tactile display that can be used on e.g. the upper leg, or the arm. In this way the (future) user can use his hands without any obstruction from the tactile display.

Experiments

The preliminary experiments with the tactile aid were especially useful since they gave us some insight into the way to conduct future experiments. These future experiments will be necessary to determine which information can best be offered in what way. Not much is known about this at present.

The results from the experiments showed that offering the first two formants, combined with the energy of the signal, on a vibrotactile display of 6 times 24 rods on the tip of the finger is useful. Both deaf
and hearing subjects showed, after a couple of hours of training, an improvement of the speech-reading recognition score when tactile supplement was offered.

A rise in the recognition score could be noted after a relatively short period of training. Within 5 to 10 hours of training, all subjects were able to recognize the offered sentences better. Especially the deaf subjects had the feeling that the tactile supplement was useful for the perception of speech. The mean rise in speech-reading scores with tactile supplement for deaf subjects was about 50% (with a 95% confidence interval from 11% to 94%). The hearing subjects showed a mean rise of about 17%, however the 95% confidence interval for this value leaves room for doubt (from 25% decrease to 59% increase).

**Overall conclusions**

We believe that the system as described in this thesis is a good tool for comparing different methods for extracting and offering information from speech that can be used as tactile supplement for speech-reading. With this system, it is possible to set up a study where every part of the tactile hearing aid can be modified, while the other parts can be left unchanged. For example, one can substitute the formant-analysis algorithm by a pitch-detection algorithm, while the display remains the same, and the extracted information can be coded into comparable tactile patterns. One can also modify only the pattern coding, and thus find an optimum coding for a certain extracted feature. And finally it is possible to work out an optimum display, matching the extracted information and pattern coding best, without the need to modify the system that extracts and codes the information. It will be clear that a future study is needed for this.

**Remarks**

**Use of the system in other applications**

The tactile hearing aid described in this thesis is designed essentially as a universal system. It can be used to present several types of
Conclusions and remarks

information via a tactile display. The method of presentation can be altered easily too. The tactile display is connected via a 16-bit wide-output bus, which means that changing the display can be relatively easy. At present both the visual display and the tactile display are connected to the same output port, without interfering with one another.

During the last couple of years the (acoustic, amplifying) hearing aids have also been given signal-processing capacities. Hearing aids that automatically adapt the gain, or the frequency response, or that try to suppress noise are beginning to become available (e.g. [Peterson et al., 1990], [Tyler & Kuk, 1989], [van Dijkhuizen et al., 1987]). In order to perform flexible signal processing (for research), one can use a DSP-based system.

The hardware system (and certain parts of the software) described in this thesis would also be suitable for this purpose. Of course, the output device needs to be different, namely a DA-converter with the necessary electronics and a (hearing aid-) speaker. The part that takes care of the signal processing could be left unchanged.

A last important application of the system is as a training aid. It can be used both to improve the quality of speech of the deaf, since it offers feedback about their own speech, and to improve speech-reading ability.
References

Alcorn, S.
The Tadoma method.

Aleksander, I. and H. Morton.
An introduction to neural computing.

Allen, P.E. and E. Sanchez-Sinencio.
Switched capacitor circuits.
Van Nostrand Reinhold electrical - computer science and engineering series.

Analog Devices
ADSP-2100 User’s Manual. 2nd ed.
Order number E971a-10-11/86.

Analog Devices
Order number E972b-5-4/87.

Analog Devices
ADSP-2101/2102 DSP Microcomputer. Datasheet.
Order number P1187-8-4/87.

* References indicated with an asterix (*) refer to unpublished literature. This can be obtained from the institute mentioned.
References

Baher, H.
Analog & digital signal processing.

The anatomical basis of medical practice.
Baltimore: Williams & Wilkins, 1971.

Open-set speech recognition in children with a single-channel cochlear implant.

Bernards, J.A. and L.N. Bouman.
Fysiologie van de mens (in Dutch).

Optical-to-tactile image conversion for the blind.

Blom, J.A.
The simplexys experiment : real time expert systems in patient monitoring.

Voice fundamental frequency as an auditory supplement to the speechreading of sentences.

Doven in Nederland : een exploratief onderzoek naar de leefsituatie van volwassen dove mensen in relatie tot opvoeding en onderwijs (in Dutch).

Breeuwer, M.
Speechreading supplemented with auditory information.

Speechreading supplemented with frequency-selective sound-pressure information.
References

Speechreading supplemented with formant-frequency information from voiced speech.

Speechreading supplemented with auditorily presented speech parameters.

Brigham, E.O.
The Fast Fourier Transform.

Bromm, B. and R.B. Treede.
Nerve fibre discharges, cerebral potentials and sensations induced by CO₂ laser stimulation.

Bromm, B. and R.B. Treede.
Human cerebral potentials evoked by CO₂ laser stimuli causing pain.

Word and feature identification by profoundly deaf teenagers using the Queen's university tactile vocoder.

Burr-Brown
ADC7802: Autocalibrating, 4-channel, 12-bit analog-to-digital converter.
Data-sheet.
Order number: PDS-1050.

CBS
English summary of “Gehandicapten wel geteld” parts 1, 2, 3.

Tactile communication of speech: comparison of two spectral displays in a vowel discrimination task.

Corthals, P. and F. Loncke, R. Hulpiau.
Luister met je ogen (in Dutch).
Brussel: BRT-Instructieve Omroep (Belgian Broadcast), 1986.
References

Craig, J.C. and C.E. Sherrick.
Dynamic tactile displays.

Deliege, R.J.H.
The "Tiepstem": an experimental Dutch keyboard-to-speech system for the speech impaired.

Dodd, B. and R. Campbell (Ed.).
Hearing by eye: the psychology of lip-reading.


Exar
Seventh order elliptic low pass filters. Datasheet of XR-1015/1016.

Ferguson, G.A.
Statistical analysis in psychology and education. 5th ed.
McGraw-Hill series in psychology.

Flanagan, J.L.
Speech analysis: synthesis and perception. 2nd expanded ed.
Kommunikation und Kybernetik in Einzeldarstellungen, Vol. 3.

Fletcher, S.G.
Seeing speech in real time.

Foulke, E.
Reading Braille.

Franklin, D.
Tactile aids: what are they?
References

Gault, R.H.  
**Progress in experiments on tactual interpretation of oral speech.**  

**Parallel processing techniques for estimating pitch periods of speech in the time domain.**  

Goldstein, J.L.  
**An optimum processor theory for the central formation of the pitch of complex tones.**  

Goldstein, M.H. Jr. and A. Proctor.  
**Tactile aids for profoundly deaf children.**  

**Vibrotactile identification of vowels.**  

Hays, W.L.  
**Statistics.** 4th ed.  

Heetman, A.  
**Analoge rekenmachines** (in Dutch).  
*Collegedictaat (lecture notes) No. 535.*

Heffels, J.J.M.  
**Het ingangsfilter van een Tactile Hearing Aid** (in Dutch).  
Eindhoven: Division of Medical Electrical Engineering, Faculty of Electrical Engineering, Eindhoven University of Technology, 1988.  
*Stageverslag (Practical Work Report) No. 5397.*

Hegt, J.A.  
**Contributions to switched capacitor filter synthesis.**  

Hess, W.  
**Pitch determination of speech signals : algorithms and devices.**  
*Springer series in information sciences, Vol. 3.*
References

Hnath-Chisolm, T. and L. Medwetsky.
Perception of frequency contours via temporal and spatial tactile transforms.

Hudson, P.T.W.
Neural Networks.
In: Neurale Netwerken : Symposiumverslag.
P. 7-28.

Ifukube, T.
A cued tactual vocoder.

Intel
Programmable logic handbook.
Order number 296083-002.

Intel
Embedded controller handbook : volume I 8-Bit.
Order number 210918-006.

Janssen, R; in cooperation with de Nederlandse Dovenraad and A. Bijlsma, H. Buter, K. Schippers.
Het Handalfabet : officieel Nederlands handalfabet van doven (in Dutch).

Johnson, C.L.
Analog computer techniques. 2nd ed.

Pain-related somatosensory evoked potentials following CO2 laser stimulation in man.
Katz, D.
**Der Aufbau der Tastwelt** (in German).
Zeitschrift für Psychologie und Physiologie der Sinnesorgane, Ergänzungsband 11 (Supplement 11), 1925, p. 1-270.
Leipzig: Johann Ambrosius Barth, 1925.

Kirman, J.H.
**Current developments in tactile communication of speech.**
In: **Tactual perception: a sourcebook.** Ed. by W. Schiff and E. Foulke.

Kume, Y. and H. Ohzu.
**Perception of dot matrix characters through skin sensation by electrical stimulation.**

Leeliveld, W.H.
**The Monoselector: a new actuator for an environmental control system.**

Lemeer, G.T.C.
**Realisatie van een tactiel display ten behoeve van auditiief gehandicapten** (in Dutch).
Eindhoven: Division of Medical Electrical Engineering, Faculty of Electrical Engineering, Eindhoven University of Technology, 1990.
**Afstudeerverslag (M.Sc. Thesis)** No. 5617.

Leysieffer, H.
**A wearable multi-channel auditory prosthesis with vibrotactile skin stimulation.**

Linvill, J.G. and J.C. Bliss.
**A direct translation reading aid for the blind.**

Lippmann, R.P.
**An introduction to computing with neural nets.**

**Tactual perception.**
References

Maassen, B.
Marking word boundaries to improve the intelligibility of the speech of the deaf.

Mack, C.
Essentials of statistics for scientists and technologists.

Mathijssen, R.W.M.
Technical report about the digital part of the DSP-circuit for the tactile hearing aid (THA.I).
Eindhoven: Division of Medical Electrical Engineering, Faculty of Electrical Engineering, Eindhoven University of Technology, 1988.
Internal report No. 88EME02. *

Proposed use of a digital signal processor in an experimental tactile hearing aid for the profoundly deaf: preliminary communication.
Journal of Medical Engineering & Technology, Vol. 13 (1989), No. 1/2, p. 84-86.

Maxim
Order number DM012/REV000.

Mies, R.J.M.
Ontwerp van een energie-zuinige besturing voor het tactile hearing aid (in Dutch).
Eindhoven: Division of Medical Electrical Engineering, Faculty of Electrical Engineering, Eindhoven University of Technology, 1990.
Stageverslag (Practical Work Report) No. 4957. *

Miyazaki, S. and A. Ishida.
Traffic-alarm monitor for aurally handicapped drivers.

Mor, J. and A. Carmon.
Laser emitted radiant heat for pain research.

Moschytz, G.S.
MOS switched-capacitor filters: analysis and design.
IEEE Press selected reprint series.
References

Moschytz, G.S and P. Horn.
Active filter design handbook : for use with programmable pocket calculators and mini-computers.

NEC
Düsseldorf: NEC Electronics (Europe), 1986.

NEC

Fast, flexible CMOS single-chip μP for digital signal processing.

Nooteboom, S.G. and A. Cohen.
Spreken en verstaan : een nieuwe inleiding tot de experimentele fonetiek (in Dutch). 2nd (revised) ed.

Analytic study of the Tadoma method : background and preliminary results.

Oppenheim, A.V. (ed.)
Applications of Digital Signal Processing.
Prentice-Hall signal processing series.

Discrete-time Signal Processing.
Prentice-Hall signal processing series.

Owen, D.B.
Handbook of statistical tables.
Addison-Wesley series in statistics.

Tactile vocoders for the deaf.
References

Peeters, S.
The LAURA cochlear prosthesis: technical aspects.

Peterson, M.E. and M.P. Feeney, P.A. Yantis.
The effect of automatic gain control in hearing-impaired listeners with different dynamic ranges.

Philips
High-speed CMOS PC74HC/HCT/HCU logic family.
Book IC06N. Order number 9398 138 60011.

Pickett, J.M.
Hearing through sight and feeling.

Pickett, J.M. (a)
Cochlear implants.

Pickett, J.M. (b)
Sensory aids.

Plant, G. and A. Risberg.
The transmission of fundamental frequency variations via a single channel vibrotactile aid.
Speech Transmission Laboratory, Quarterly Progress and Status Report, No. 2-3, 1983.
Stockholm: Dept. of speech communication & music acoustics, Speech Transmission Laboratory, Royal Institute of Technology (KTH), 1983. P. 61-84.

Plomp, R.
Auditory handicap of hearing impairment and the limited benefit of hearing aids.
Plomp, R. and A.M. Mimpen.  
**Improving the reliability of testing the speech reception threshold for sentences.**  

**Digital processing of speech signals.**  
*Prentice-Hall signal processing series.*

Risberg, A.  
**Speech coding in aids for the deaf: an overview of research from 1924 to 1982.**  
Speech Transmission Laboratory, Quarterly Progress and Status Report, No. 4, 1982.  

**Vibrotactile frequency for encoding a speech parameter.**  

Sacks, O.W.  
**Seeing voices: a journey into the world of the Deaf.**  

Schiff, W. and E. Foulke (ed.)  
**Tactual perception: a sourcebook.**  

Sensoraid  
**Kanievski Sound: A radically new portable aid for the totally deaf.**  

Sherrick, C.E.  
**Basic and applied research on tactile aids for the deaf people: progress and prospects.**  

Sherrick, C.E. and Cholewiak R.W.  
**Cutaneous sensitivity.**  
References

Sherrick, C.E. and J.C. Craig.
The psychophysics of touch.

Soucek, B. and M. Soucek.
Neural and massively parallel computers: the sixth generation.

Investigating the MESA (Multipoint Electrotactile Speech Aid): the
transmission of segmental features of speech.

Spens, K.E.
Tactile speech communication aids for the deaf: a comparison.
Speech Transmission Laboratory, Quarterly Progress and Status Report, No. 4, 1980.
Stockholm: Dept. of speech communication & music acoustics, Speech
Transmission Laboratory, Royal Institute of Technology (KTH), 1981. P. 23-38.

SPSS
SPSS-X user’s guide. 3rd ed.

Stein, L. and T. McGee, P. Lewis.
Speech recognition measures with noise suppression hearing aids using a
single-subject experimental design.
Ear and Hearing, Vol. 10 (1989), No. 6, p. 375-381.

Tervoort, B.T.
Sign languages: Dutch.
In: Gallaudet Encyclopedia of deaf people and deafness, Vol. 3. Ed. by
John V. Van Cleve.

Tye-Murray, N. and M. Lowder, R.S. Tyler.
Comparison of the F0F2 and F0F1F2 processing strategies for the
corporation cochlear implant.

Tye-Murray, N. and R.S. Tyler.
Auditory consonant and word recognition skills of cochlear implant users.
Tyler, R.S. and F.K. Kuk.  
The effects of "noise suppression" hearing aids on consonant recognition in speech-babble and low-frequency noise.  
Ear and Hearing, Vol. 10 (1989), No. 4, p. 243-249.

Van Cleve, J.V. (ed.).  

Van der Krol, R.C.P.  
Sprakmodule voor de Possum : een communicatiehulpmiddel voor meervoudig gehandicapten (in Dutch).  
Eindhoven: Division of Medical Electrical Engineering, Faculty of Electrical Engineering, Eindhoven University of Technology, 1988.  
Stageverslag (Practical Work Report) No. 5454.*

Van Dijkhuizen, J.N. and P.C. Anema, R. Plomp.  
The effect of varying the slope of the amplitude frequency response on the masked speech perception threshold of sentences.  

Van Eijnsbergen, H.H.W.M.  
Electro-tactiele stimulatie als hulpmiddel voor auditief gehandicapten (in Dutch).  
Eindhoven: Division of Medical Electrical Engineering, Faculty of Electrical Engineering, Eindhoven University of Technology, 1990.  
Stageverslag (Practical Work Report). *

Vogten, L.L.M.  
Analyse, zuinige codering en resynthese van spraakgeluid (in Dutch).  

Wang, X.  
Preliminary investigations on tactile perception of graphical patterns.  
EUT Report 90-E-239.

Waterham, R.P.  
The "Pocketstem": an easy-to-use communication aid for the vocally handicapped.  
References

Weinstein, S.
Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality.
In: The skin senses. Ed. by R. Kenshalo.

Willems, L.F.
Robust formant analysis for speech synthesis applications.

Willems L.F.
Program implementing the robust formant analysis.
Internal IPO report, No. 632. *

Yeung, E. and A. Boothroyd, C. Redmond.
A wearable multichannel tactile display of voice fundamental frequency.
Appendix 1

list of sentences used in the experiments [Plomp & Mimpen, 1979]

(sentences used with kind permission of prof. Plomp, VU University-Amsterdam)
List of sentences

sentence set 1

1 De bal vloog over de schutting
Morgen wil ik maar één liter melk
Deze kerk moet gesloopt worden
De spoortrein was al gauw kapot
De nieuwe fiets is gestolen
Zijn manier van werken ligt mij niet
Het slot van de deur is kapot
Dat hotel heeft een slechte naam
De jongen werd stevig aangepakt
Het natte hout sist in het vuur
Zijn fantasie kent geen grenzen
De aardappels liggen in de schuur
Alle prijzen waren verhoogd

---

2 Zijn leeftijd ligt boven de dertig
Het dak moet nodig hersteld worden
De kachel is nog steeds niet aan
Van de viool is een snaar kapot
De tuinman heeft het gras gemaaid
De appels aan de boom zijn rijp
Voor het eerst was er nieuwe haring
Het loket bleef lang gesloten
Er werd een diepe kuil gegraven
Zijn gezicht heeft een rode kleur
Het begon vroeg donker te worden
Het gras was helemaal verdroogd
Spoedig kwam er een einde aan

---

3 Ieder half uur komt hier een bus langs
De bel van de voordeur is kapot
De wind waait vandaag uit het westen
De slang bewoog zich door het gras
De kamer rook naar sigaren
De appel had een zure smaak
De trein kwam met een schok tot stilstand
De koeien werden juist gemolken
Het duurt niet langer dan een minuut
De grijze lucht voorspelt regen
Hij kon de hamer nergens vinden
Deze berg is nog niet beklimmen
De bel van mijn fiets is kapot

---

4 De auto heeft een lekké band
Het moeilijke werk bleef liggen
Het vliegtuig vertrekt over een uur
Die jongens vechten de hele dag
De schoenen moeten verzaaid worden
In de krant staat vandaag niet veel nieuws
Door de neus ademen is beter
Het kind was niet in staat te spreken
De witte zwaan dook onder water
Hij nam het pak onder zijn arm
Gelukkig sloeg de motor niet af
De leraar gaf hem een laag cijfer
Het huis brandde tot de grond toe af

---

5 De foto is mooi ingelijst
Mijn broer gaat elke dag fietsen
Een kopje koffie zal goed smaken
De schrijver van dit boek is dood
Zij heeft haar proefwerk slecht gemaakt
De sigaar ligt in de asbak
De appelboom stond in volle bloei
Er wordt in dit land geen rijst verbouwd
Hij kan er nu eenmaal niets aan doen
De kleren waren niet gewassen
Het gedicht werd voorgelezen
Haar gezicht was zwart van het vuil
De letters stonden op hun kop

---

6 De groene appels waren erg zuur
In het gebouw waren vier liften
Lopen is gezonder dan fietsen
Het lawaai maakte hem wakker
Mijn buurman heeft een auto gekocht
Als het flink vriest kunnen we schaatsen
De kast was een meter verschoven
Oude meubels zijn zeer in trek
De portier ging met vakantie
De lantaarn gaf niet veel licht meer
Door zijn snelheid vloog hij uit de bocht
Het is hier nog steeds veel te koud
De oude man was kaal geworden
De bomen waren helemaal kaal
Rijden onder invloed is strafbaar
Onze bank geeft vijf procent rente
Het verslag in de krant is kort
In de vijver zwemmen veel vissen
Honden mogen niet in het gebouw
Een flinke borrel zal mij goed doen
Gisteren waaide het nog harder
Het meisje stond lang te wachten
De volgende dag kwam hij ook niet
Het geschreeuw is duidelijk hoorbaar
Eindelijk kwam de trein op gang
De grote stad trok hem wel aan

---

De bus is vandaag niet op tijd
Onze dochter speelt goed blokfluit
Ook in de zomer is het hier koel
Zij moesten vier uur hard werken
Niemand kan de Fransman verstaan
Eiken balken zijn erg kostbaar
Het aantal was moeilijk te schatten
Er waaide een stevig briesje
De vis sprong een eind uit het water
Iedereen genoot van het uitzicht
Het regent al de hele dag
Het tempo was voor hem veel te hoog
In juni zijn de dagen het langst

---

De bakkers bezorgen vandaag niet
Het licht in de gang brandt nog steeds
De wagen reed snel de berg af
Lawaai maakt je op den duur doof
In de kerk wordt mooi orgel gespeeld
De schaatsen zijn in het vet gezet
Toch lijkt me dat een goed voorstel
Hij probeerde het nog een keer
De zak zat vol oude rommel
Zij werd misselijk van het rijden
Door zijn haast maakten hij veel fouten
De nieuwe zaak is pas geopend
Dat is voor hem een bittere pil

Appendix 1

Op het gras mag men niet lopen
Steile trappen zijn gevaarlijk
De zon gaat in het westen onder
De hond blafte de hele nacht
De kat van de buren is weg
De trein vertrekt over twee uur
Het was stil in de duinen
Hij rookte zijn sigaret op
De rivier trad buiten haar oevers
De jongens gingen er gauw vandoor
Moeizaam klom de man naar boven
De biefstuk is vandaag erg mals
De kat likt het schoteltje leeg
Het water is nog steeds erg koud
Deze krant wordt niet meer bezorgd
Het eten was dit keer erg lekker
De peren waren nog niet rijp
De voordeur bij de buren klemt
Men kan op het ijs nog niet schaatsen
Het gras moet nodig gemaaid worden
De bomen staan weer in volle bloei
Ik krijg dit boek vandaag niet uit
Het regent al de hele middag
Zij mag van de dokter niet roken
Op de markt zijn de bloemen goedkoop
De schilder heeft zijn huis verkocht

Gisteren is hier iemand beroofd
Onze auto wilde niet starten
Het wordt morgen erg druk in de zaak
De jongen kreeg een nieuwe voetbal
Het is fijn weer om te zeilen
De lange steeg was slecht verlicht
Dat geld heb ik van de bank gehaald
Hij liet zijn broek opnieuw persen
Paling is erg duur geworden
Dit smalle pad voert naar het dorp
Deze tekst is bijna onleesbaar
De kip legde geen eieren meer
De weg slingerde door het dal

In maart heerste er nog strenge vorst
De volgende dag sneeuwde het
Zijn woorden waren goed uitgekozen
De melk in die emmer is zuur
We zullen ons nog moeten haasten
De leesboeken liggen op tafel
Ik heb nieuwe schoenen gekocht
Het zou vandaag zeer rustig worden
Onze bakker moet erg vroeg opstaan
De radio stond veel te hard
We zagen een hert in het bos
Het plan is een week uitgesteld
Wij rijden samen naar het werk

De schoorsteen is door de storm vernield
Ik heb de krant nog niet gelezen
De reis is lang en saai geweest
Het was een erg leuke avond
Haar laatste rapport was niet zo best
Ik zal nog eens thee inschenken
Het meisje glimlachte verleid
Zei heeft al koffie gedronken
De snijbonen smaken uitstekend
Deze leunstoel zit wel lekker
De voetballer maakte een schuiver
Die winkel is vanavond open
Zijn gedrag levert gevaar op
17  De foto's zijn weer goed gelukt
Het water in de sloot is vuil
Mijn buurman heeft de rozen gesnoeid
Op woensdag komt de bakker niet
Hij schrijft een brief aan zijn vriendin
Die struiken bloeien in de zomer
De vrachtboot voer de haven binnen
Hij zeilde ook met harde wind
Onze kip heeft een ei gelegd
Geen mens kan lang zonder voedsel
Die kleuters gaan om acht uur naar bed
In het voorjaar bloeien de tulpen
De vloer van de zaal is erg glad

---

18  Mijn moeder is kleiner dan ik
Vroeger waren er weinig auto's
De man gooide de deur hard dicht
De boeken vielen uit de kast
Hij krijgt een grote taart cadeau
Mijn zoon is van de fiets gevallen
Alle deuren zijn rood geverfd
Wij wandelen veel in dit park
Er zijn veel mensen met vakantie
De lichten gingen vanzelf aan
Op reis leer je veel mensen kennen
De omslag van dit boek is van leer
Hij heeft zijn rekeningen niet betaald

---

19  Onze gast blijft vannacht slapen
De zon gaat schuil achter de wolken
De schrijver zit achter zijn bureau
De schipper hees de rode vlag
Er zitten veel mensen in de zaal
De accu moet vervangen worden
Veel oude bomen zijn omgewaaid
De lak van die auto is slecht
Harde muziek bevalt hem niet
Die kleine vogel is zeldzaam
De poes van mijn tante is ziek
Deze stof heeft een vale kleur
Die planten hebben zonlicht nodig
List of sentences
Appendix 2

list of Dutch phonemes: original (or standard) symbols, symbols used in this thesis, and an example of their pronunciation [Nootenboom & Cohen, 1984].

(Note: this list contains the ‘basic’ phonemes. Some lists contain more phonemes (and allophones), depending on the accuracy with which one wants to describe the pronunciation. For this reason some lists have more than twice as many phonemes as this list.)
List of Dutch phonemes
## List of Dutch phonemes

<table>
<thead>
<tr>
<th>Dutch vowels</th>
<th>Dutch consonants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phoneme</strong></td>
<td><strong>Example</strong></td>
</tr>
<tr>
<td><strong>standard</strong></td>
<td><strong>thesis</strong></td>
</tr>
<tr>
<td>/i/</td>
<td>IE/</td>
</tr>
<tr>
<td>/y/</td>
<td>UU/</td>
</tr>
<tr>
<td>/u/</td>
<td>OE/</td>
</tr>
<tr>
<td>/i/</td>
<td>I/</td>
</tr>
<tr>
<td>/e/</td>
<td>EE/</td>
</tr>
<tr>
<td>/e/</td>
<td>E/</td>
</tr>
<tr>
<td>/a/</td>
<td>AA/</td>
</tr>
<tr>
<td>/ø/</td>
<td>EU/</td>
</tr>
<tr>
<td>/ø/</td>
<td>U/</td>
</tr>
<tr>
<td>/o/</td>
<td>OO/</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>O/</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>A/</td>
</tr>
<tr>
<td>/ɛ:/</td>
<td>AI/</td>
</tr>
<tr>
<td>/œ:/</td>
<td>EU/</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>OO/</td>
</tr>
<tr>
<td>/ɛi/</td>
<td>EI/</td>
</tr>
<tr>
<td>/øy/</td>
<td>U1/</td>
</tr>
<tr>
<td>/œu/</td>
<td>AU/</td>
</tr>
<tr>
<td>/ə/</td>
<td>e/</td>
</tr>
<tr>
<td>/h/</td>
<td>H/</td>
</tr>
</tbody>
</table>
Appendix 3

Drawings and photos of the novel experimental vibrotactile display.
A new experimental tactile display
Appendix 3

Experimental tactile display

Fig A.3-1: Schematic representation of the tactile display [Lemeer, 1990].

Fig A.3-2: Schematic drawing of the bottom plate of the novel tactile display [Lemeer, 1990].

Sizes in mm
A new experimental tactile display

Photo: stafgroep reproductie en fotografie, T.U.E.

Photo of the new experimental tactile display.
Photo of the new experimental tactile display.
A new experimental tactile display
Samenvatting

De enige manier waarop totaal dove mensen in een alledaagse situatie spraak kunnen verstaan is door gebruik te maken van spraakafzien (ook wel liplezen genoemd). Dit betekent echter, dat mensen bij wie de mond minder duidelijk beweegt (bijvoorbeeld door slechte articulatie of te vlug spreken), of bij wie de mond minder goed te zien is (door te weinig licht, of tegenlicht), slecht te verstaan zijn door iemand die doof is. Maar zelfs onder optimale omstandigheden, d.w.z. wanneer iemand duidelijk en niet te vlug spreekt bij een correcte verlichting, is spraakafzien nog niet eenvoudig –omdat een aantal bewegingen van het spraakkanaal van buiten af niet zichtbaar is– en zeker vermoeiend.

Verder kunnen dove mensen allerlei alledaagse geluiden niet horen, terwijl deze soms wel belangrijke informatie kunnen bevatten. Denk hierbij aan claxons, aan roepen, aan een aanstormende trein, enzovoort.

Als nu enige vorm van informatie over spraakgeluiden, of geluiden in het algemeen, aangeboden kon worden, zou dit iemand die doof is aanmerkelijk kunnen helpen. Op zich is het weten dat er een (luid) geluid is vaak al voldoende informatie. Door te kijken kan men dan te weten komen wat dat geluid veroorzaakt kan hebben. Een experimenteel systeem, dat in staat is om aan te geven dat er geluid is, en dat voor het spraakafzien ondersteunende informatie aan kan bieden, wordt in dit proefschrift besproken.

Aangezien deze informatie niet akoestisch kan worden aangeboden, zal dit systeem gebruik moeten maken van een alternatieve manier om zijn informatie aan te bieden. We hebben er voor gekozen om deze informatie niet visueel aan te bieden. Immers dove mensen hebben hun ogen maar al te zeer nodig om te kunnen spraakafzien, en om te letten op zaken, waarop horende mensen door geluid attent worden gemaakt. De huid –of meer exact, het gevoels- en tast orgaan (het tactiele orgaan)– kan wel een goed alternatief zijn om informatie aan te bieden.
Samenvatting

Het systeem dat in dit proefschrift wordt beschreven probeert daar dan ook gebruik van te maken; vandaar zijn naam: een Tactiel Hoor Hulpmiddel.

Aangezien het nog niet duidelijk is welke (spraak) informatie het meest geschikt is om aan de huid aan te bieden, is het systeem zodanig opgezet, dat het in principe mogelijk is om verschillende soorten informatie te halen uit het spraaksignaal, of elk willekeurig geluidssignaal. Dit is gerealiseerd door gebruik te maken van een zogenaamde Digitale Signaal Processor (DSP). Dit is een micro-processor die speciaal ontworpen is om te rekenen aan (geluids-)signalen.

In dit proefschrift wordt een aantal programma’s voor het tactiele hoor hulpmiddel besproken, welke de toonhoogte (of grondfrequentie), of de formant frequenties uit het spraaksignaal kunnen afleiden, en het frequentie spectrum of de energie inhoud van een geluidssignaal kunnen bepalen. Al deze programma’s werken praktisch ‘real-time’ (d.w.z. ze werken zo snel dat ze een stukje geluidssignaal kunnen bewerken, binnen de tijdsduur van dat stukje geluid), en kunnen de onttrokken informatie binnen ca. 50 milliseconde presenteren.

Eén methode, waarbij informatie wordt aangeboden over de formant frequenties en energie inhoud van spraaksignalen, is geëvalueerd in een inleidende test, met zowel dove als horende proefpersonen. Het bleek dat alle proefpersonen beter konden spraakafzien, als hen ook tactiele informatie over het spraaksignaal werd aangeboden op een vingertop. De training die nodig bleek om de herkennings score significant te laten stijgen was relatief kort: afhankelijk van de proefpersoon lag deze tussen ca. 5 tot 10 uur. Voor de dove proefpersonen werd een gemiddelde relatieve stijging van 50% waargenomen in de herkenningsscores, wanneer tactiele informatie wordt aangeboden.

- 178 -
Curriculum vitae

26 augustus 1962 Geboren te Roosendaal


Bedrijfsstage: Philips Medical Systems
onderwerp: een snelle interface voor in een CT-scanner.
Afstudeerrichting: Medische Elektrotechniek
Afstudeeronderwerp: Ontwerp van een aansturing voor een lineair array t.b.v. een ultrasone tomograaf.

sep. 1986 - dec. 1990 Assistent in Opleiding (AIO-4)/Promovendus
TU Eindhoven, werkzaam bij de vakgroep Medische Elektrotechniek, bij de interfacultaire werkgroep “Communicatie hulpmiddelen t.b.v. gehandicapten” voor het verrichten van onderzoek naar de mogelijkheden van tactiele informatie overdracht t.b.v. totaal doven.
STELLINGEN

behorende bij het proefschrift

An Experimental DSP-Based Tactile Hearing Aid
a feasibility study

door

R.W.M. Mathijssen

Eindhoven, 20 september 1991
Het blijkt mogelijk om na een relatief korte trainingstijd informatie voor ondersteuning van spraakafzien te onttrekken aan correct gekozen tactiel aangeboden informatie.

(B dit proefschrift)

Bij tactiele displays waar vrijwel constant (spraak)data wordt aangeboden, vormen de aangeboden data een referentie voor zichzelf.

(B dit proefschrift)

Indien spreker-onafhankelijke spraakherkenning van lopende spraak op dit moment beschikbaar zou komen, zal dit voorlopig nog niet gerealiseerd kunnen worden in draagbare, batterij-gevoede communicatie hulpmiddelen voor gehandicapten.

De uitdrukking ‘Speed Power Product’ zou voor een beter begrip veranderd moeten worden in ‘Speed Power Quotient’.

(B dit proefschrift)

Het feit dat iemand tactiel aangeboden karakters uitstekend kan onderscheiden, is geen waarborg dat die persoon tactiele ondersteuning voor spraakafzien snel correct kan waarnemen.

(B dit proefschrift)

Datgene wat een ‘pitch-extractor’ uit een signaal ‘extraheert’ hoeft niet als zodanig aanwezig te zijn in het beschouwde signaal.


Voor de ontwikkeling van een kind is het van wezenlijk belang dat het in de eerste levensjaren wordt blootgesteld aan een taal die het zich —op min of meer natuurlijke wijze— eigen kan maken; mocht deze eerste taal slechts door een klein aantal mensen in de leefomgeving gebezigd worden, dan kan het kind altijd nog een meer gebruikelijke taal als tweede taal aanleren.

De bioscoop in Nederland en Vlaanderen is beter toegankelijk voor auditief gehandicapten dan in de omringende landen dankzij de traditie om buitenlandstalige films slechts in bijzondere gevallen na te synchroniseren.

Ondanks de dwang om te publiceren ten einde de financiële middelen voor wetenschappelijk onderzoek te waarborgen (publish or perish), dient men bij het bekend maken van gedane of te verwachte ontwikkelingen —in het bijzonder voor gehandicapten of ernstig zieken— te allen tijde de nodige omzichtigheid te betrachten, om te hoog gestelde of valse verwachtingen te voorkomen.

Zolang wettelijk niet kan worden opgetreden tegen overtredingen van het rookverbod in openbare gebouwen, kan de gezondheid van niet-rokers extra worden geschaad door spanningen en het gevoel van onmacht —naast het verplicht moeten meeroiken— als een roker de vaak overduidelijk zichtbare niet-roken bordjes weer eens “niet heeft gezien”.

Men mag stellen, dat de V-stelling de belangrijkste stelling is voor vliegeraars.

In het land der blinden zal éénoog eerder opgesloten in een inrichting zitten dan op de koningstroon.