Using speech technology in the field of aids for the handicapped

USING SPEECH TECHNOLOGY IN THE FIELD OF AIDS FOR THE HANDICAPPED.

Authors[1]: W.H. Leliveld and R.P. Waterham.

Address:
Dr. ir. R.P. Waterham, Institute for Perception Research (IPO),
IPO 2.27, P.O. box 513, 5600 MB Eindhoven, the Netherlands.
Ir. W.H. Leliveld, Eindhoven University of Technology, (TUE-EME),
Division of Medical Electrical Engineering, EH 3.09, P.O. box 513,
5600 MB Eindhoven, the Netherlands.

ABSTRACT

Developments in the field of speech technology and microelectronics create new possibilities in aids for the handicapped. In this paper we will present several examples of applications meant for different categories of disabled persons.

The categories of disabled we generally focuss on are the visually impaired, the speech impaired, the motorically handicapped and in some cases the hearing impaired.

The research done at IPO and TUE-EME, on aids for the handicapped, generally follows a strategy in which the process of defining requirements, realizing and subsequently evaluating an aid, is carried out as many times as it requires to come up with a useful aid. During the last phase of the project industrial contacts are looked for in order to transfer the results and to make the aid commercially available.

Speech technology can be divided in speech recognition and speech production. Speech recognition can be used to create a voice operated environmental control system based on the Bush-Timac system. Speech production can be used in two ways. First it can be used as an alternative communication channel meant for the visually impaired (e.g., Typophone, Reflotalk, speaking elevator and speaking household balance). Second it can be used to replace natural speech when it is lost (e.g., Tiepstem, Pocketstem and speaking Possum). In the lastmentioned case social acceptance of the synthetic speech is a vital aspect.

Not all projects that have resulted in a useful aid have been transferred to industry because of the complexity of this transfer and the poor market expectations.

[1]Authors in alphabetical order.

Key words: Aids for the handicapped, Synthetic speech, Speech recognition.

PREFACE

In 1984 the Institute for Perception Research (IPO) and the division of Medical Electrical Engineering of the Eindhoven University of Technology joined forces in the research on communication aids for the handicapped. Knowledge in the field of aids for the handicapped together with developments of both speech technology and microelectronics were used to investigate the possibilities of speech production and recognition techniques in communication aids for the handicapped. Speech recognition techniques can be used to control devices in case normal operation is impossible (eg.; environmental control), where as speech production techniques can be used to replace lost speech or to serve as an alternative communication method (eg.; speech feedback instead of visual feedback). In this paper we will give an overview of the activities of the IPO and EME on the field of communication aids for the handicapped, in particular communication aids using speech technology.

RESEARCH STRATEGY

The development of communication aids for the disabled is hampered by the fact that researchers and technicians are not fully aware of the real communication problems of the potential users because they do not have these problems themselves. Apart from that it is very hard to extract communication requirements when communication is a problem. Furthermore it is very difficult to estimate the remaining abilities of the handicapped person (eg.; a visually handicapped person can generally interprete sounds better than non-handicapped). Because of forementioned complexity of the situation it is essential that potential users are included in the development. The contribution of potential users to the development can best be carried out by including them in an extensive evaluation. One way to be sure that assistance of potential users is included is the use of a research strategy (Soede, 1980; Waterham, 1989; Deliege, 1989a) as given in fig. 1.

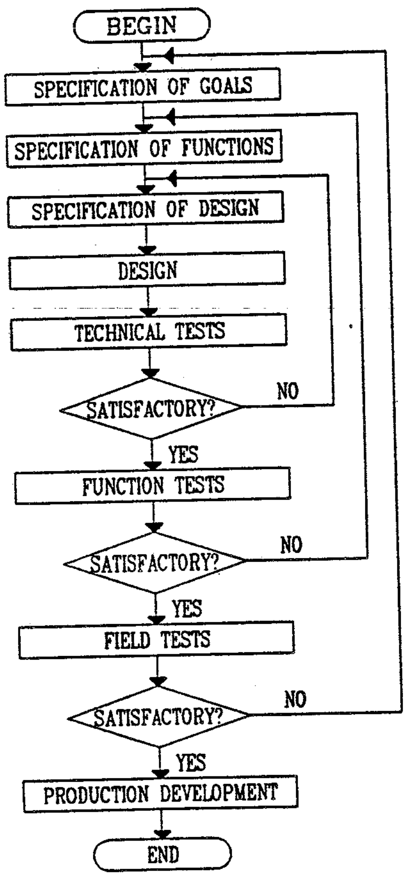

Fig. 1: Research strategy (Soede, 1980).

The project starts with a definition (as good as possible) of the requirements for the intended aid. In this process the assistance of potential users or therapists should be included by means of interviews and inquieries. After the realization and technical tests, the aid should undergo an extensive field test in which potential users can fully use the aid. In general the first field test is used to redefine the requirements. According to these adjusted requirements a new design has to be made. This process should continu untill the field test has a positive result. If this is the case the aid should be transferred to industry in order to make the aid commercially available.

There are some general requirements that can be made beforehand. Speech output devices for communication purposes should be (1) compact and robust, (2) have fast access, (3) incorporate an acceptable speech quality and (4) should be affordable. These demands stem from various aspects of the application situation. The first demand stems from the fact that the devices are normally used in ADL situations. The second demand garantees immediate response, while the third relates to the fact that the speech should at least be intelligible. Of course this all should be realized at a price that is not out of proportion with the goal. General requirements for speech recognition systems are: (1) fail-safe or not used in critical situations, (2) reasonably effective and (3) affordable. The first demand stems from the fact that speech recognition is not (yet) capable of interpreting different voices and/or emotions correctly, so misunderstanding can occur in critical situations (eg.; wheelchair control). The second demand (meaning a reasonably correct and fast recognition) prevents frustration in use. The third demand is the same as the last demand of speech output devices.

Ideas for applications are generally the result of specific questions, put forward by individuals, to solve a specific problem. The result of developments, however, may be of general interest. Occasionally industry comes up with a question or technological developments lead to new application areas.

TECHNOLOGY

The developments in microelectronics resulted in the availabilty of low-power high capacity microprocessors suitable for application in aids. In our case generally the MCS51 series of microcontrollers is used. The availability of large memory circuits (EPROM) and integrated speech synthesizers (which drastically reduce the data rate) made it possible to create speech output devices that are relatively cheap and very robust. In our case a Philips PCF8200 formant synthesizer is used which offers intelligible speech at a data rate of 2.5 kbits per second (or even less). This implies that a 32 kbyte EPROM can store up to 100 seconds of speech.

Speech can be stored in different ways, varying from hole sentences to sound fragments (eg.; phonemes). The reproduction of stored messages (sentences, words) is called speech resynthesis, where the creation of an utterance from sound fragments is called speech synthesis. The form in which speech has to be stored depends on the application. In cases where a limited number of previously known messages is required, speech resynthesis can be applied. In this case all messages are recorded, analyzed and stored and will be reproduced when necessary.

In case the number of messages is too big, but the number of words that can form these messages is still relatively small, the technique of word concatenation can be applied (this is still speech resynthesis). However, in this case sentence melody is not in order if no action is taken. If the message consists of a carrier part supplemented with a specific part (eg.; the result is: *9.4* millimol per liter), this technique is quite appropriate.

In case the speech required is not known beforehand, the technique of speech synthesis is required. In this technique only speech (sound) fragments (eg.; phonemes) are stored and when necessary concatenated to form the message desired. In this case both word and sentence melody are not at hand and if monotonous speech is not applicable serious action has to be taken to cover this aspect. Even when measures are taken to cover lastmentioned aspects, the speech will still sound less natural because of the temporal aspects. The intelligebility of the speech, however, is very adequate.

For the application of speech recognition a SP1000 signal processor is used to perform the signal analysis and forementioned hardware (a MCS51 series microcontroller) is used to perform pattern recognition. The system can perform word recognition, which is speaker dependant, on isolated speech fragments. This implies that the system has to be trained by the user to learn all the commands.

APPLICATION AREAS

Fig. 2: The Typophone.

We will now discuss several applications for the handicapped realized with the speech techniques peviously mentioned. The first example is the Typophone (Kroon, 1986; Stovers, 1990), a device that gives auditory feedback of the typing process (see Fig. 2). In this case all messages are stored and reproduced when required. The Typophone gives feedback of the keystrokes and gives additional information on the typing process. The Typophone can be used in the process of learning to type and for checking what has been typed. In the same category we can mention the Pocketstem (Waterham, 1989), a device that can produce 28 prestored messages and is ment for the speech impaired with additional motor and/or cognitive handicaps (see Fig. 3), where messages are reproduced simply by pressing a labled key. The lables are specially designed pictograms to indicate the message stored. A third example of this category is the speaking version of the Possum communication aid. This is a single key operated device to select communication symbols. In this version each symbol represents a word or sentence, spoken when the symbol is selected.

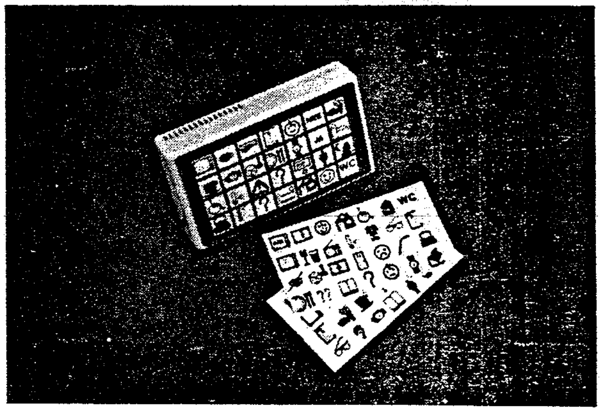

Fig. 3: The Pocketstem.

Several applications we realized use the word concatenation technique with carrier sentences. A speaking clock will say the time when requested in the way people normally say the time. For instance 13.46 will be spoken as: "It is fourteen minutes to two o'clock". The clock can be put in a mode in which it states the time every fifteen minutes automatically.

A speaking elevator (Nijtmans, 1985; Stok, 1985) indicates where it is going to when a button is pressed, at what floor the elevator is at the moment the door opens and gives additional messages (see Fig. 4).

Two examples in this category will be discussed a bit more in detail. The first example is a speaking digital household balance (v. Uitert, 1990), specially designed for the visually disabled, which speaks the information shown on the display (see Fig. 5). To announce stable values, however,

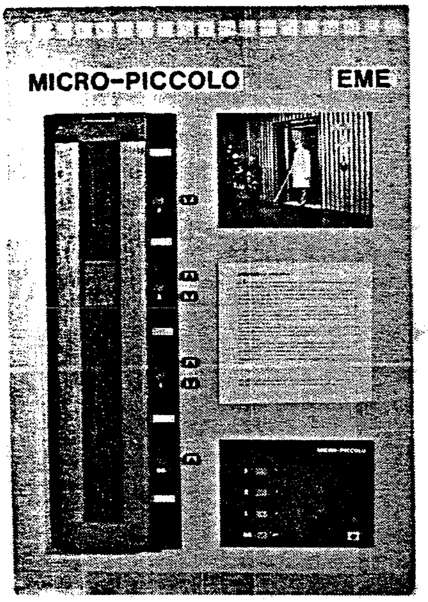

Fig. 4: The speaking elevator.

Fig. 5: The speaking household balance.

is sufficient for stationary measurements but is not usefull for standard household situations, where the measurement most of the time is used to get a certain amount of some matter (eg.; poor sugar in a bowl untill you have 350 gram).

The balance guides this process by estimating the weigth of the load at the end of an utterance, given the inflow is more or less stable. As a result of an evaluation we found that this works even better than was expected beforehand. A second ex-

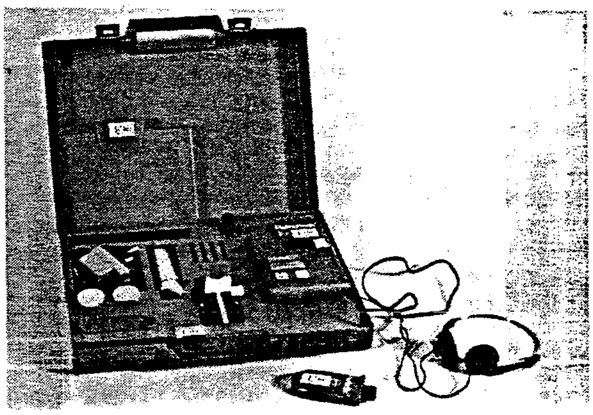

Fig. 6: The Reflotalk.

ample is a blood glucose measurement system (Reflotalk) (Waterham, 1983; Steeksma, 1988) which is specially designed for visually handicapped diabetics (see Fig. 6). As a visual disorder is often the consequence of diabetes, a blood glucose measurement system was equipped with speech feedback to enable visually handicapped diabetics to monitor their blood glucose level individually.

The most flexible speech production method (speech synthesis) is used in the Tiepstem (Deliege, 1989a). This is a keyboard-to-speech system (see

Fig. 7: The Tiepstem.

Fig. 7) ment for the speech impaired who still are able to control a keyboard and who have sufficient language skills. This communication aid uses a limited text-to-speech conversion program (some words need to be typed in more or less phonetic) and has facilities to save and edit previously typed messages. At the moment at IPO a full text-to-speech system has been developped which will be the hart of a new version of the device.

Speech recognition is used in combination with the Bush-Timac X10 system (see Fig. 8) to realize a voice-operated environmental control system (Bosch, 1985; v.d. Krol, 1990). The Bush-

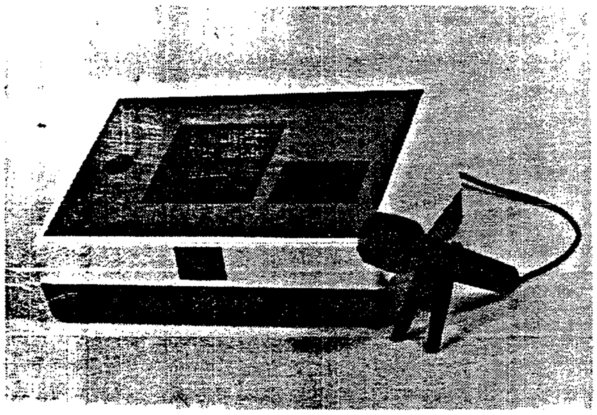

Fig. 8: A voice-operated environmental control system.

Timac X10 is a wireless remote control system that controls sixteen receiver units via the wires of the mains. These units can switch on and off devices or lamps. Special units are able to dimm and brighten lamps. The speech recognition system that is able to identify 40-60 words is used to operate the functions (16 channels and 6 function commands) of the environmental control system.

RESULTS

Although for each application effort is made to make the device become commercially available, this is not always achieved. The Typophone for instance was more than once the subject of negotiations, but never made it to production. The talking lift indicator and the speaking version of the Possum also were never transferred to industry. The speaking clock will, in a slightly adjusted form, be produced in the near future, while the speaking household balance is already in a final state for production preparation. The Reflotalk (English) has been transferred to industry in Germany and the United States. The Pocketstem has been transferred to a Dutch industry and a prototype series of 50 copies will be available this year. The Tiepstem and the voice-operated environmental control system are still in a research phase. The Tiepstem will undergo a second development loop, while the voice-operated environmental control system has to undergo an extensive field test, although functional tests had promissing results. Attempts will be made to tranfer the lastmentioned projects to industry as soon as evaluation shows a satisfactory result.

CONCLUSIONS

One of the main results of our research is that speech resynthesis is a very usefull technique for application in aids for the handicapped. The application of speech synthesis is fairly new but already there are indications that this technique is very useful and the speech is socially acceptable. Field tests will have to prove that this statement holds. Anyway synthetic speech is an adequat alternative communication channel for the visually handicapped as well as a powerful speech replacing tool. The speech recognition techniques has not yet learned us much about the possibilities of its application. Research and developments in the near future will probably lead to other applications specially for Text-to-speech systems (Deliege, 1989b).

REFERENCES

Bosch H. (1985), Een goedkope spraakherkenner voor toepassing in hulpmiddelen voor motorisch gehandicapten, *Afstudeerverslag Technische Universiteit Eindhoven, vakgroep EME.*

Deliege R.J.H. (1989a), The "Tiepstem": an experimental Dutch keyboard-to-speech system for the speech impaired, *Doctoral thesis, Eindhoven University of Technology.*

Deliege R.J.H. (1989b), A stand-alone text-to-speech system, *IPO Annual Progress Report*, 24, pp. 43-46.

Krol R.C.P. v.d. (1990), Een demonstratiemodel van een spraakherkenner ten behoeve van omgevingsbesturing door gehandicapten, *Afstudeerverslag Technische Universiteit Eindhoven, vakgroep EME.*

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

Nijtmans J.J. (1985), Een sprekende liftindicator, *Stageverslag Technische Universiteit Eindhoven, vakgroep EME.*

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

31-32.

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

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

Stok L. (1985), Een liftsimulator voor de micropiccolo; een liftbedieningsaanpassing voor visueel gehandicapten, *Stageverslag Technische Universiteit Eindhoven, vakgroep EME.*

Stovers M.Q.C. (1990), De Typofoon: modernisering van de sprekende schrijfmachine, *Afstudeerverslag Technische Universiteit Eindhoven, vakgroep EME.*

Uitert A.G. and Ossevoort H.J.M. (1990), Een sprekende keukenweegschaal voor visueel gehandicapten, *technische beschrijving, EME rapport 90-EME-03.*

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

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