THE OPERATION OF A VERSATILE PHONEME-BASED SPEECH AID USING A SPECIAL HIGH BANDWIDTH PERSON-AID INTERFACE

M. Walsh, L. Westphal, M. Darveniza

Electrical Engineering Department
University of Queensland

ABSTRACT - A novel method of man-machine communication is described. This is used to control a multi-purpose communication aid designed for persons with an acquired speech disability. The hardware and software required to realize the complete system are briefly described. The result of recent field trials are also given, indicating the ease of use of the system.

INTRODUCTION

Speech aids for the vocally handicapped are chosen so that the disabled person's ability to communicate is optimized for that person's physical, mental and vocal ability. To overcome an apparent lack of suitable aids for a class of disabled, a new type of person-aid interface has been designed.

Special purpose hardware and software are required to allow the interface to be fully utilized to control a phoneme-based aid. These details are described by Walsh et al. (1985). The resulting system has the potential to allow communication via synthetic speech at rates approaching that of normal speech.

INPUT DEVICE

Before an input device can be selected for an aid, careful consideration must be given to the desired performance of the aid and the group of handicapped persons who are expected to benefit most from the aid. From the vast array of aids available, it is seen that the communication needs of persons who are vocally handicapped but not significantly physically or mentally handicapped are not adequately catered for. The alphanumeric keyboard-based aids used by these persons (e.g. the Canon 'Communicator' (R)) may allow simple entry and compilation of messages, but are slow to operate. These aids do not allow the user to communicate at rates approaching normal human speech, the inadequacy being with the person-aid interface (i.e. the keyboard).

To overcome the limitations of the alphanumeric keyboard, a new kind of interface needs to be designed. Vanderheiden (1976) indicates that the correct selection of the input device is the first basic step in the development of any communication system. To allow the user to communicate at 'normal' rates, the required interface must be capable of transducing information (from the user) at reasonably high rates. Several different approaches were initially investigated, including an analogue joystick and EMG switches. However, the operation of the joystick was cumbersome, and the EMG switches required extensive preprocessing and careful placement of electrodes. A more direct interfacing system was required, one which did not require elaborate signal processing and one which was easy to operate. The obvious body site for such an interface is the hand.

Some of the features needed for such an interface are:

(1) The use of a limited (small) number of switches, typically one per finger, each finger positioned close to a unique switch. In this way, visual and hand-eye coordination skills are not required to operate the interface.

(2) The use of an encoding approach, as opposed to a scanning or a direct selection approach.

(3) The use of parallel encoding (i.e. the pressing of a combination of switches at one time instant) rather than serial encoding (e.g. Morse Code).

These features indicate that the operation of the interface could be analagous to playing a musical instrument.

INPUT STRATEGY

A commercially available device which fits the above description should now be considered. The 'Microwriter' (R) uses 5 switches, one for each of the digits of one hand, and performs the function of a one-hand keyboard. However, the method of control, the number of switches and the portability makes this device unsuitable for use in a phoneme-based aid. To allow a user to select phonemes quickly from a phoneme set with more than 32 elements (which is the limit of a device with 5 switches), a different approach is needed.

Firstly, at least 6 switches are required, so that the user may access phonemes from a 'large' phoneme set (i.e. up to 64 phonemes). Secondly, to improve the speed of operation, a 'high bandwidth' interface is required as described in Goldenberg (1981). The design procedure for such an interface is given below:

(1) Use high bandwidth input devices that transduce unintentional as well as intentional components of the user's motion (i.e. don't throw away information before input).

(2) Process the signal obtained from the input devices to improve the signal-to-noise ratio.

(3) Provide feedback to the user, so that the nature of the enhancements provided by this scheme can be learnt, allowing the user to develop strategies to effectively use the device.

The switches in such a device must be easy to operate, requiring little pressure and only a small motion. They must also operate quietly and have a minimum number of moving parts (which may enhance reliability).

The input scheme used by the Microwriter (R), although adequate for entering text, is not suitable for the real-time selection of phonemes. A more appropriate scheme would be accepting data after a combination of switches has been held constant for a predefined period. Another predefined period can be used to give the user sufficient time to change the combination after the aid has verified that the previous one has been accepted (by saying the requested phoneme). This approach is, in effect, the filtering of the input signal using timing restraints, and is the approach taken in the design of the interface.

PROTOTYPE DESCRIPTION

Based on the features listed above, touch sensitive switches were used in the interface. Each finger was assigned a switch, each switch being activated by a small movement of the finger causing the finger to touch a small conductive contact. The thumb was assigned 3 switches, arranged in a triangle formation. A switch was also provided for the wrist, giving a total of 8 switches. A diagram of the basic interface is shown below in figure 1.

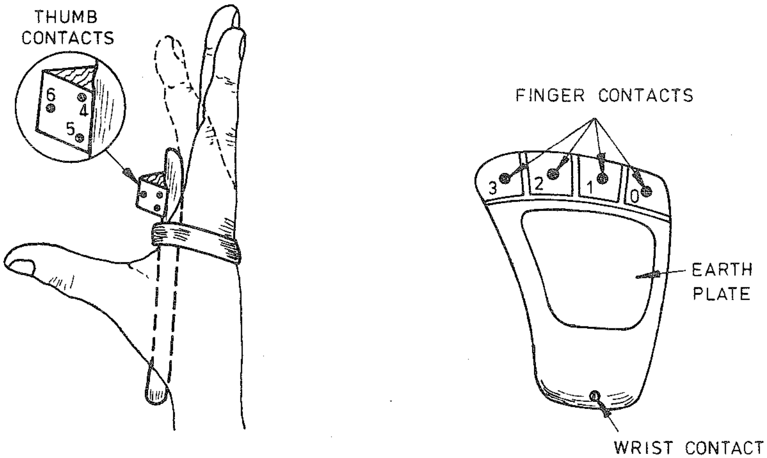

Figure 1. Hand Interface.

The device is strapped on to one of the user's hands, so as to maintain a constant positional relationship with the hand. This also allows the interface to be used while the user is standing. This is important as it does not require the user to be seated to operate the device, nor does it require the user to look at the device to use it.

Prototypes have been made out of wood and P.V.C., each of which is made to suit a particular user's hand.

The hand interface is required to control the speech generated by the aid. This is done by selecting phonemes with the interface, each phoneme being selected by touching a unique combination on 6 of the contacts on the hand interface. By selecting an appropriate sequence of phonemes, words can be constructed. Given that the aid has an adequate phoneme set, the user can construct any word in the English language. The user can also select from a variety of preprogrammed phrases using the hand interface, enabling commonly used words to be said with ease. This function can be conveniently selected, since only 6 of the 8 contacts are used for phoneme generation leaving the other 2 for preprogrammed phrase selection.

HARDWARE FUNCTIONS

To allow the touch switches to serve as input devices, special purpose hardware was designed. This circuitry extracts a digital signal from each of the touch switches, with the digital signal then being processed by software to reject unwanted signals. In this way, the filtering of the input signal can be readily altered by changing constants in the software.

The main control system of the aid is based around a CMOS microprocessor and associated CMOS LSI devices. The speech synthesizer used is a 'third generation' device and is capable of high quality speech. The whole system is powered by rechargeable batteries, allowing the aid to generate at least 6 hours of continuous speech between recharges.

SOFTWARE FUNCTIONS

Special purpose software has been written to allow the user to alter the input filtering. This provides the user with a mechanism for matching the aid to his/her abilities. The default parameters of the synthesized voice can also be changed to suit the user, all of these functions controlled by the hand interface. Inflection and emphasis functions are available, allowing the user to change a small number of these parameters in real-time. The facility is available to allow the user to program a number of personalized phrases into the aid, thus customizing the aid's pre-programmed vocabulary.

TRIAL RESULTS

Trials have been carried out over the past 6 months with vocally handicapped persons using the aid. A total of six persons were evaluated in this period. Two basic modes of testing were used, the mode dependent on the home location of the users.

(1) Initial coaching followed by independent learning.

(2) Periodic coaching (1 hour/week).

It was found with all of the users that it was essential to ensure that the hand interface 'fitted' the hand correctly. Once this was done, evaluation of the aid could commence.

With the users who learnt the aid as in (1), it was found that after a period of approximately 2 weeks of use (an average of 1½ hours per day), most of the aid's phoneme set could be remembered and accessed without any prompting. Many different words could be said, with a few words said fluently. These users indicated that, given sufficient time, the skill required to control the aid to its full potential could be acquired. Some problems were experienced, mainly in the selection of phonemes which have a very short duration (e.g. stops).

The users who were coached for one hour every week (and with no contact with the aid for the rest of the week) found that only 4-5 hours of use was required before most of the phoneme set could be accessed. The extent of the learnt phoneme set was demonstrated by the users being able to say all of the numbers from 1 to 20, and all of the letters in the alphabet without need to constantly refer to the Users Guide. These users generally felt that the aid could be 'mastered' in a period of 2 weeks, at which stage any word could be generated. Once this stage was reached, the skill required to control the aid would improve, yielding a speed of operation which is only limited by the physical and mental dexterity of the user. However, concern was expressed about the quality of some of the aid's phonemes, as well as the ease of selection of some of the phonemes.

CONCLUSIONS

A type of person-machine interface has been described which can allow for high bit-rate communication between a person and a machine. This type of interface has been specifically designed to allow a vocally handicapped person to control a phoneme-based speech aid. Using this interface and specialized hardware and software, a communication system suited to a class of vocally handicapped persons has been realized.

The results of early trials with such a device are given and indicate that users can accept the concept of the aid and learn how to operate it in a relatively short time.

REFERENCES

GOLDENBERG, E.P. (1981) "Flexible High Bandwidth Communication for Motorically Impaired Persons". Proceedings of the Johns Hopkins First National Search for Applications of Personal Computing to Aid the Handicapped, October 31 1981, 190-192.

VANDERHEIDEN, G.C. (1976) "Non-Vocal Communication Techniques and Aids for the Severely Physically Handicapped", p. 19 (University Park Press).

WALSH, M.J., WESTPHAL, L.C. "Communication Aid for the Vocally Handicapped", presented at IREECON International, October 4 1985.