Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

A multi-language, portable text-to-speech system for the disabled

Carlson, R. and Galyas, K. and Granström, B. and Hunnicutt, S. and Larsson, B. and Neovius, L.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 22
number: 2-3
year: 1981
pages: 008-016

http://www.speech.kth.se/qpsr

II. SPEECH SYNTHESIS

A. A MULTI-LANGUAGE, PORTABLE TEXT-TO-SPEECH SYSTEM FOR THE DISABLED*

R. Carlson, K. Galyas, B. Granström, S. Hunnicutt, B. Larsson, and L. Neovius

Abstract

Previous experience with speech output aids for blind and non-vocal persons has shown great promise. The need for individual adjustments and the relatively small market makes flexible, programmable aids necessary. We have developed a modular microprocessor text-to-speech system that is portable and battery operated. This prototype has been adjusted to several different users. Programs for different languages have been developed. Connection of text sources is made simpler by standardized interfaces. One special attachment is a 500-symbol Bliss board and a related Bliss-to-speech program that transforms the symbol string to well-formed sentences.

Introduction

Advances in speech research and electronics have made available a new class of aids based on different kinds of artificial speech for the handicapped. The aids range from devices with a small fixed vocabulary of digitized natural speech to unrestricted text-to-speech systems. We are here concerned exclusively with aids of the latter type that transform any text to synthetic speech. Since many of the aids, such as speech prostheses and reading machines, offer new possibilities, the functional specification of the aids are in many cases still to be made. At our laboratory a minicomputer-based text-to-speech system was developed three years ago and has since been used in different experimental applications for blind and non-vocal persons(Carlson and Granström, 1978; Carlson et al., 1980). This equipment was generally very well received, but in this context we want to focus on some of the negative reactions.

---

* An earlier version of this paper was presented at the Annual Conference of Rehabilitation Engineering, Washington, D.C.

- Too space consuming, not portable.
- It has to be mains-connected.
- Too noisy (fan noise)
- Spelling mode not available.
- Speech quality sometimes too low.
- Need for different text input devices.
- Production cost too high.
- Not in regular production.

The last remark is, of course, related to the preceding. It proves to be impossible to engage any commercial company in production of this kind of aid. The general objections are that the market is too small and too uncertain, especially in a small language community such as Sweden, and that the cost of developing a production prototype is too high. This is a situation that we face in supplying different kinds of technical aids for the handicapped. Part of the answer to this dilemma lies in developing aids at research institutes that are closer to producible equipment. Modular construction also makes it possible to use the same units in different kinds of aids. Standardization and sharing developing efforts internationally also allow a potentially greater market.

Our present project is focussed on such a development, and we have tried to accommodate as many of the suggestions for improvement as possible based on the experience from the previous prototype mentioned earlier. A substantial part of the work is also devoted to evaluations and application-oriented hardware and software adjustments.

Text-to-speech-hardware

The present text-to-speech system consists of a conventional formant speech synthesizer, a powerful microcomputer based on the MC68000 16-bit microprocessor and a variety of text input equipment. In the basic configuration, the microcomputer consists of four euro-cards; the processor card, one RAM memory card (32k byte), and two PROM memory cards, each with 32k bytes of memory and one serial I/O port. The microcomputer box also contains rechargeable batteries, a loudspeaker, and different kinds of controls (Fig. II-A-1).

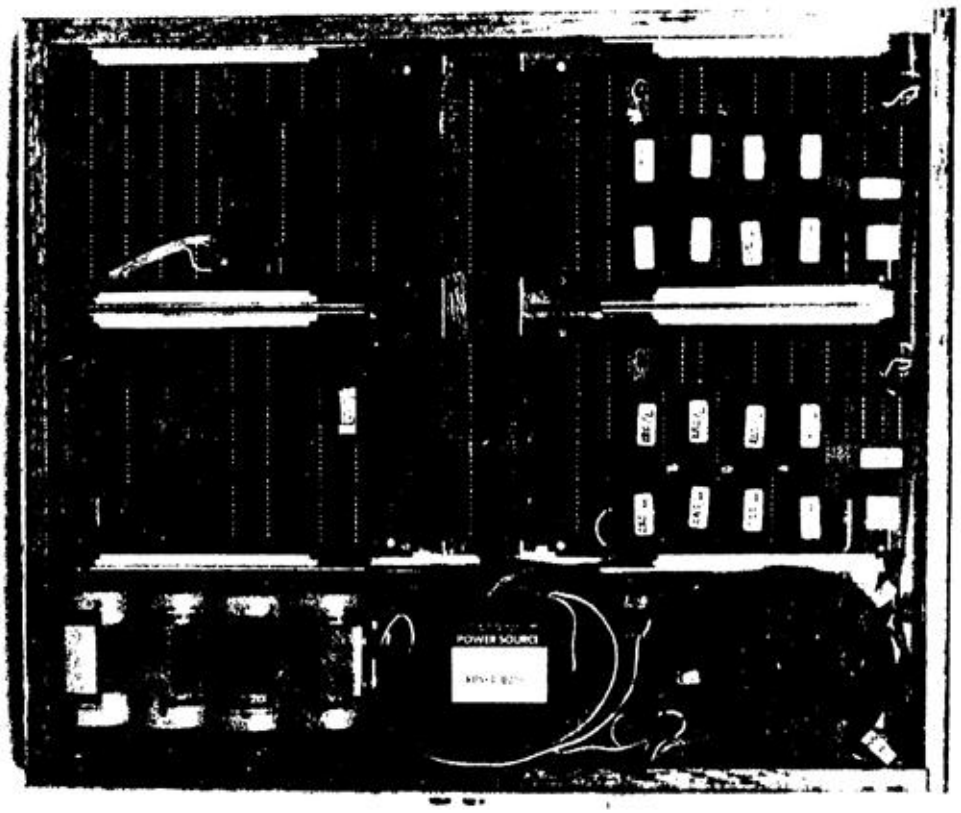

Fig. II-A-1. Top view of the text-to-speech system.

The synthesizer box fits under the microcomputer and is of the same size. As powerful signal processors become more available this component of the system will, of course, shrink drastically both in price and physical dimensions. These boxes, connected to any kind of text source such as a conventional computer terminal, operate as a general text-to-speech system. To make the system fully portable, we have constructed a special "lid" to the computer box that contains a low profile keyboard and a 16-character liquid crystal display (Fig. II-A-2). Other attachments include enlarged keyboards and a 500-symbol Bliss board which will be described later in this paper.

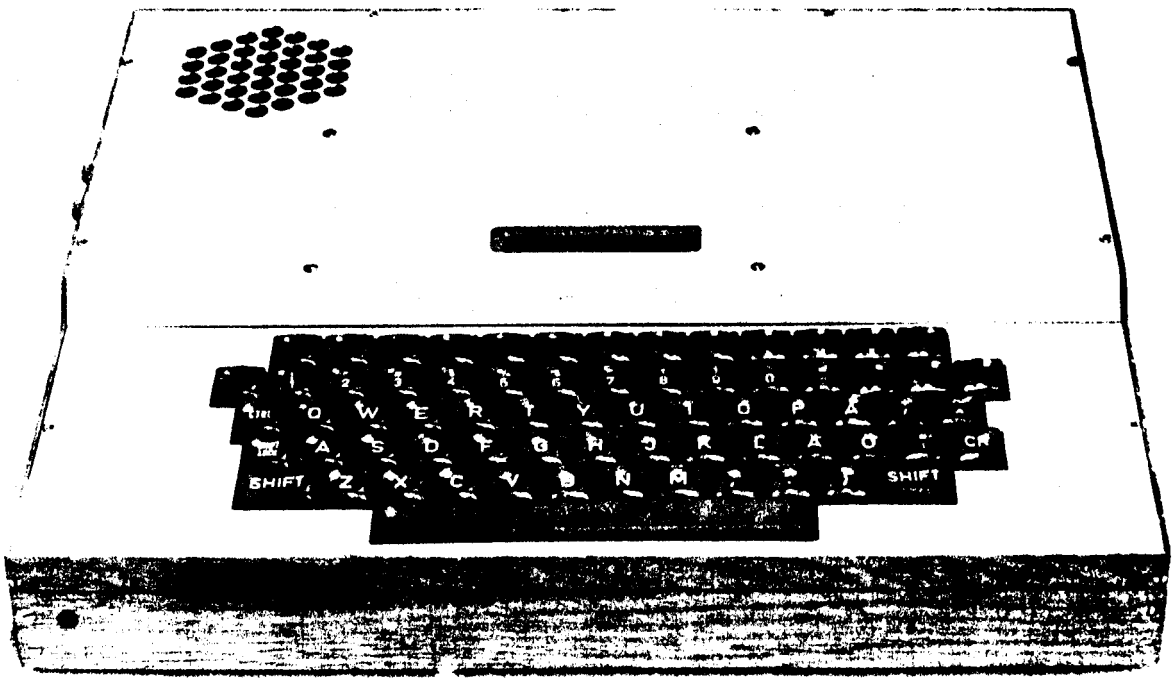

Fig. II-A-2. The text-to-speech system with a key-board lid.

General software features

The text-to-speech program is initiated on "power on." It operates in three basic modes: spelling, word by word, and sentence mode. An additional mode allows a combination of Bliss symbol and ordinary alphanumeric input. The principles of text-to-speech conversion are described elsewhere (Carlson and Granström, 1976). Unlike commercial synthesizers which operate with a fixed set of sounds (allophones), this system has a sound inventory which is fully under program control. The realization of the sounds is language-dependent and adjusted according to the linguistic context.

It should be emphasized that the system accepts any kind of text, including numbers.

The talking speed is adjustable both by a knob and by keyboard commands. In sentence mode, continuous reading at above 250 wpm (words per minute) is possible.

Besides the ordinary keyboard connection facility there is the possibility of connecting to a host computer. This is necessary in the talking terminal application, but is also useful in other contexts.

Keyboard commands to the system consist of a command prefix plus a command character. Some of the implemented commands are:

- Change to spell mode
- Change to word mode
- Change to sentence mode
- Stop output from synthesizer
- Continue output from synthesizer
- Slow down speech
- Speed up speech
- Save last sentence
- Retrieve a saved sentence
- Reinitiate the program
- Write to a host computer

The text-to-speech system has been programmed for several languages. Versions now exist for Swedish, English, Spanish, Finnish, German, and Chinese.

Applications for the blind

The generality of the text-to-speech system module makes it possible to use it in different kinds of aids for the blind. Some applications have already been explored.

Automatic "talking book" production

In cooperation with the Swedish Federation of the Visually Handicapped (SRF), we use the system to record different kinds of text material on audio tapes. The text is supplied from ordinary casette tapes which are used for text editing and printing of Braille books at the SRF printing office. In the future the blind person will be able to choose which kind of media he wants the text to be presented in: Braille or synthetic speech.

Reading machine

So far we are missing suitable OCR (optical character recognition) equipment to realize a reading machine for the blind. Special characters in the Swedish language make it necessary to develop special recognition schemes. Such work is, however, in progress in several places and we hope to be able to realize a reading machine for the blind within the next few years.

Talking terminal

In cooperation with the Swedish Institute for the Handicapped, a special working place for a blind user is being developed during 1981. It will be equipped with a small computer and a talking terminal. The goal is to gain experience with how such an equipment should look and what facilities are needed. Special emphasis will be placed on studying the ergonomics of a normal working situation. The computer will be used for both programming and editing of text.

Talking Braille recorders

The recent development of Braille recorders with editing and search capabilities has created a possibility for blind persons to interactively work with great volumes of text. Speech is in many cases a more convenient and fast mode of presentation than Braille. Giving this kind of equipment a speech output option will greatly enhance its usefulness. This kind of interface will be implemented.

New kinds of information systems, such as TV-text, Viewdata and electronic mail, will be directly accessible by blind users through an appropriate connection to a text-to-speech system.

Applications for the non-vocal

Speech prosthesis with "typed" input

The previous prototype has been used extensively for non-vocal in both communication and language training. The smaller size and the possibility to run on rechargeable batteries will make the use as a personal speech prosthesis feasible.

The main problem seems now to be with the typing of text, since most non-vocal persons have additional motor disabilities. Several different keyboards, such as a Canon Communicator and enlarged keyboards, have already been interfaced to the system. Interfacing other kinds of control devices is planned.

Communication speed can be increased by storing whole phrases as 2-3 letter codes.

Bliss communication with speech output

The value of synthetic speech as a tool in learning has been demonstrated (Warrick et al., 1977). In their experiment, each symbol was pronounced separately and the use was restricted to the school setting. The utilization of synthetic speech as a means of communication for Blissymbol users has been discussed. Speech allows communication with people at some distance and also with people who have no prior knowledge of Blissymbols.

McNaughton has expressed some caution in applying speech synthesis to Blissymbol communication (McNaughton, 1977). She points out that it must be ensured that "the spoken utterance triggered by the Blissymbol statement provides the user and the listener with an acceptable English translation". Research on these lines has been reported from France (Emerard et al., 1980) and the USA (Eulenberg and Rahimi, 1978). Such considerations have also been taken into account in developing the special attachment to the text-to-speech system for Blissymbol users described in this paper.

A standardized chart containing approximately 500 symbols has been used. Symbols on the board are actuated by touching the appropriate square with a magnet. After an adjustable delay, the code for that symbol is transmitted to the text-to-speech system via a standard RS-232C interface. The symbol set contains numbers and digits. The addition of a few special symbols has made the whole capability of the general text-to-speech system, such as temporarily storing and quickly retrieving sentences, accessible from the Bliss board.

The Bliss-to-speech program is an expansion of the standard text-to-speech program. Therefore, Bliss symbols and spelled text can be intermixed to produce a spoken message. As each symbol is chosen, the synthesis system pronounces it as if it were the only word in a sentence. When the user terminates the sentence with a punctuation mark, the complete grammatically well-formed sentence is spoken.

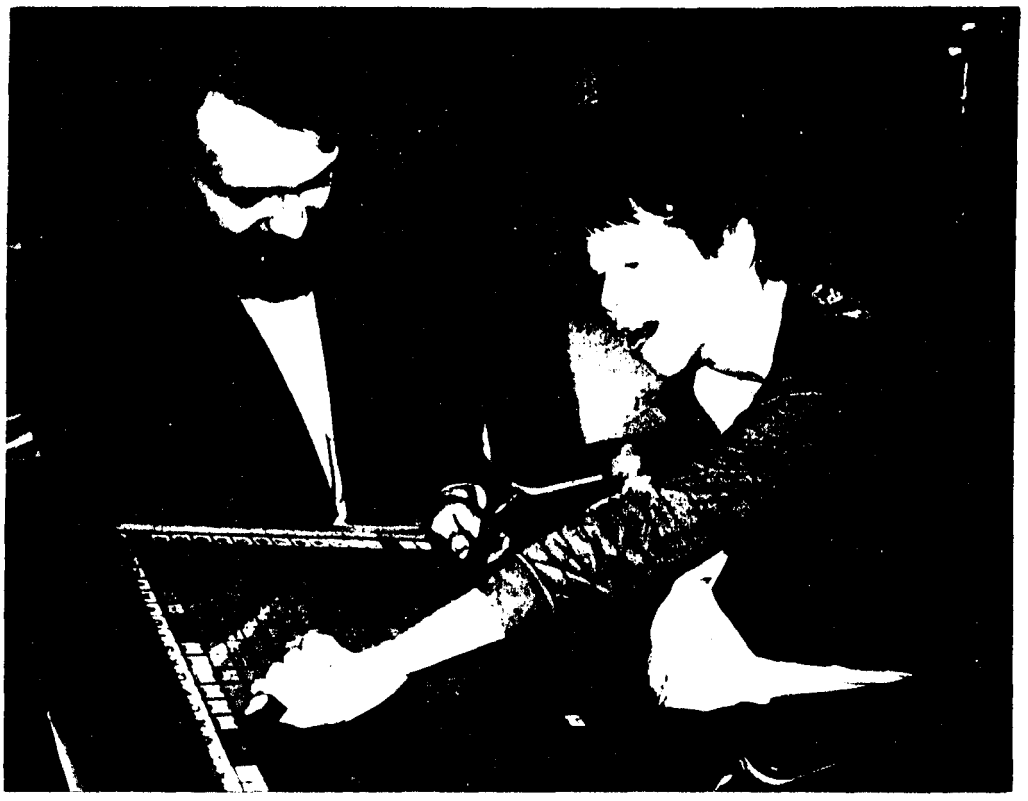

Fig. II-A-3. A Bliss board connected to the text-to-speech system. When a square is touched the corresponding word will be pronounced.

Each symbol on the Bliss board corresponds to a lexical entry giving its pronunciation and a code for its part-of-speech and other features such as declension category. Using these codes and a grammar, the phrase structure is established; then rules for well-formed sentences are applied. The system has been successfully tested on several types of declarative and interrogative sentences including some with more than one clause. These sentence types can be composed of around 25 types of noun phrases and 25 types of verb phrases. This facility allows a user to compose many well formed sentences with flexibility of expression.

The possibility to spell words has already proved to be valuable in that it adds to the expressional capability. Many Bliss users have good language comprehension and this kind of device is judged to be of great help in language production training.

Final remarks

The text-to-speech system presented here is still not in regular production. We feel, however, that practice with the system in many different applications is of primary importance. The experiences gained are

necessary for a good specification of the final product. Some of the features of the present system are worth mentioning. The modularity has made it easy to adjust to different needs. Using state-of-the-art technology makes the system small and relatively inexpensive. The choice of a powerful processor makes future expansions and improvements of the text-to-speech program easy to accommodate without sacrificing speech rate.

Experience from applications tells us that often the greatest problem in utilizing this kind of aid is the individual adjustments. This makes it even more important that the basic text-to-speech module is general enough to interface to a great variety of text sources and control devices.

Text-to-speech technology will, with improved quality, spread to more applications in everyday life. This will also promote greater production volumes and lower costs, and increase the availability for the disabled. It is important that the needs of the handicapped are well understood so that the demands could be considered in a commercial product.

References

Carlson, R. & Granström, B. (1976): "A text-to-speech system based entirely on rules", Conf.Record 1976 IEEE-ICASSP, Philadelphia, PA, USA.

Carlson, R. & Granström, B. (1978): "Experimental text- to-speech system for the handicapped", J.Acoust.Soc.Am. 1978 vol. 64, S163.

Carlson, R., Galyas, K., Granström, B., Pettersson, M., Zachrisson, G. (1980): "Speech synthesis for the non-vocal in training and communication", STL-QPSR 1/1980, Royal Inst. of Technology, Stockholm, Sweden.

Emerard, F., Graillot, P., Cyne, G., Lucas, JJ. (1980): "Protheses de parole destinees a la communication des handicapes moteurs deficients de la parole", pp 133-144 in Recherces / Acoustique vol. VI-1979/80, Centre national d'etudes des telecommunications centre de recherches de Lannion.

Eulenberg, J.B. & Rahimi, M.A. (1978): "Toward a Semantically Accessible Communication Aid", Proc. of the National Electronics Conference, Vol. XXXII, 1978.

McNaughton, S. (1977): "Technical Aids in Blissymbol Communications", pp 96-100 in Proc. of the Workshop on Communication Aids for the Handicapped, Ottawa 1977.

Warrick, A., Nelson, P.J., Cossaltor, J.G., Cote, C., McGillis, J., Charbonneau, J.R. (1977): "Synthesized Speech as an Aid to Communication and Learning for the Non-Verbal", pp 120-135 in Proc. of the Workshop on Communication Aids for the Handicapped, Ottawa 1977.