APPLICATIONS OF SPEECH TECHNOLOGY IN AIDS FOR THE DISABLED

Rolf Carlson, Björn Granstrom and Sheri Hunnicutt*
Dept. of Speech Communication and Music Acoustics

Royal Institute of Technology
Stockholm, Sweden

*names in alphabetical order

ABSTRACT

A number of technical aids which include speech synthesis or speech recognition have been developed at the Department of Speech Communication and Music Acoustics and are now being used by disabled individuals. Applications of synthetic speech include a communication aid, a symbol-to-speech system, word predictors, talking terminals and a daily newspaper. Speech recognition is also being used in a communication aid.

INTRODUCTION

Because speech is the most natural way for people to communicate with each other, there has been a great deal of interest in the use of synthetic speech and stored or concatenated speech in technical aids for blind and nonvocal individuals. In the aids that have been developed at the Department of Speech Communication and Music Acoustics, the speech synthesis system which was developed there has been used. This system is being presented in another contribution to this conference by the same authors (Carlson, Granström and Hunnicutt, 1988). It is a text-to-speech system, accepting any text input. Presently it exists in nine languages (or dialects): Swedish, American English, British English, Spanish, German, Norwegian, Danish, French and Italian.

This text-to-speech system has several components. A first set of components produces a phonetic text from the orthographic text. The phonetic text is produced by a set of grapheme-to-phoneme rules for regular pronunciations and a lexicon for irregular pronunciations. A user lexicon is also available for special user-specified pronunciations. Special number rules convert digit sequences to phonetic text. A second component is the set of phonetic rules which create appropriate allophones by adjusting formant frequencies, aspiration and transition timing, and determine phoneme duration and utterance fundamental frequency depending on context. This information is stored in parameters which are sent to the digital speech synthesizer each 10 msec. This synthesizer has a combined parallel/cascade filter structure and can use either voiced (pulses) and/or unvoiced (noise) excitation. The synthesis hardware is based on a Motorola 68000 and a NEC7720 speech processing chip.

The system contains programs that allow it to be adapted to different applications. It operates in several reading modes: spelling, word, line and sentence. Reading speed is variable, and this speed is adjustable at any time. Voice parameters are also adjustable at any time. These include pitch level, dynamic range and breathiness of voice. A voice type that has been created using these parameters may also be stored and retrieved for a particular user. It is also possible to generate phonetic text if the pronunciation given is inappropriate. This pronunciation may be permanently stored in a user lexicon. One may save and retrieve sentences and may delay or continue output from the synthesizer. The language may also be easily changed. It is possible to connect to either a terminal or a host computer. This is necessary in the talking terminal application, but is also helpful in other contexts.

TEXT-TO-SPEECH IN VOICE PROSTHESIS

The first use of the Swedish text-to-speech system as a communication aid was in 1978. This early system was based on a minicomputer and could be moved around on wheels. A teenager, diagnosed as suffering from cerebral palsy, used the system for several years before beginning to use one of the newer variety. Writing on the keyboard using a mouthstick, he used the system in school for about a year, and then in job training. It was apparent that the system increased his ability to communicate and his linguistic competence. At the same time, it increased his general motivation which was marked in other areas of his schoolwork as well. The results were quite encouraging, and valuable information was gained regarding psychological and social factors as well as regarding the functional design of aids with synthetic speech (Carlson et al., 1980).

In the following years, when the new systems containing microcomputers with signal processing chips became available, a number of other persons with various needs began to use them for verbal communication. Some examples from a study by the Swedish Institute for the Handicapped follow (Schildt and Sterner, 1986):

One man, who has had both a laryngectomy and a glossectomy uses the system quite flexibly. He types quickly, and, particularly with the abbreviations possible in the user lexicon, can carry on a conversation in near normal time. He finds speech a much more natural way to communicate than writing, and uses it in conversations with his family. Telephone conversations have also become a possibility once again. This seems to be a very important use for others, as well, since it gives a great deal more independence and privacy. Another man, left without speech after a stroke, began to use the system while still in a hospital. He later moved to a service apartment where he lives independently, having the help of an electric wheelchair, a personal computer and his speech synthesis device.

The device has also been used rather much for training reading and writing, particularly for practicing spelling. One non-speaking boy who seemed not to be able to learn to write more than a few short words, learned easily with the speech synthesis device. Using a "sounding out" mode, he listened to the sound of each letter until he could begin to write words. Another child, who has a brain injury and autism, was able to immediately transfer his experience with a communicator, and began writing right away. He was further motivted by the device and greatly enjoyed hearing words with particular meaning for him. A group of non-speaking students has used it for reading a theatre play together The play was typed in a file in the computer. When a particular actor's turn came, only his/her keyboard would trigger the reading of that part.

MULTI-TALK

The speech synthesis device has been packaged in an attache case as a special-purpose communication aid called Multi-Talk. It is being produced by another Swedish firm, Fonema AB. To use it, one simply lifts the attache case lid, turns on the device, and begins to type on the Epson keyboard. It runs on either rechargeable batteries or mains power. Multi-Talk comes equipped with up to four of the available languages; a language can be chosen by simply pushing a function key.

In addition to the usual text-to-speech features such as abbreviation and control of voice quality, Multi-Talk includes several features specially designed to aid communication. There is a contrast-adjustable screen to see what is being written, and a function key to hear what has already been written in the current sentence. The last sentence can be repeated, even in the middle of the following sentence, and can be repeated word by word if desired. Any word can even be spelled out if it has not been understood. The printer included in the Epson can always be used to print out the text which is visible on the screen.

The speed of communication can be increased very much by the availability of two "higher levels." One higher level allows the user to access stored messages with any single key. This level can be accessed for saying a message without disturbing work in progress at the base level. The other higher level also allows the user single-key access to messages. These messages, however, can be completed with further typed text, or can be copied into the sentence in progress on the base level.

In early trials with Multi-Talk, it was observed that young school children became more interested in practicing reading and writing when speech output was provided to them. This led to the design of a series of computer programs for developing phonological awareness. These programs include practice in sound discrimination, phoneme identification, letter and phoneme relationships and spelling. While constructing a sentence, the pupil may listen to phonemes, syllables, words and sentences. Listening augments the ability to detect spelling errors such as omitted letters, reversed words, and missing words. The student's activities are logged continuously, providing the teacher with information about his/her progress. The most apparent effect of working with these programs has been the high level of motivation. Controlled experiments and case studies have shown unexpectedly large improvements. An increase in the pupil's self esteem has also been observed in many cases.

BLISSTALK

The use of Blissymbols in Sweden began in 1976 at two regional habilitation centers, Since then, their use has spread widely in Scandinavia. Because of this wide interest, another device containing speech synthesis and taking Blissymbol input was developed. It is an electronic communication board called Blisstalk (Hunnicutt,1986) which was first built and tested in 1981. As a symbol board with a "voice," it makes its users heard. Communication is not so highly dependent upon the willingness of a "listener" to watch, and to interpret the symbols for himself.

Blisstalk is now produced by a Swedish company, Rehabmodul AB. On it are up to 500 Blissymbols which are selected by a magnet or by scanning. The board can be reprogrammed with any of 1400 available symbols: a few large symbols may be chosen for a beginner, and more can be added as the user progresses. Each symbol is represented in a lexicon by one or two corresponding words, their pronunciation and grammatical category. Some symbols have grammatical functions themselves, e.g., plural, verb tenses, possessive. Blisstalk also contains a special phrase structure grammar which takes account of word order, phrase order and grammatical information in the lexicon to produce well-formed sentences.

Besides this sentence mode are also a word-by-word mode, which does not access the grammar, and a character mode for pronouncing numeral and letter names. A sentence or other completed expression may be repeated, and up to 10 sentences may also be temporarily stored and quickly retrieved. The letters may be used to supplement the symbols by spelling out words, just as in the usual text-to-speech system.

Systems for Swedish, English and French have been distributed. A system for Spanish is also ready for use.

WORD PREDICTION

An adaptive word prediction scheme was developed several years ago to be used in conjunction with our speech synthesizer (Hunnicutt, 1987). It was developed in response to the need of a non-vocal user of the synthesizer who found that her listeners were guessing the word she was typing before she finished. In order for the word to be synthesized and pronounced correctly, however, it was necessary for her to finish typing it. It seemed that a word prediction scheme, based on a frequency-sorted lexicon, could be a decided help.

Given an initial letter by the user, a word is predicted based on word frequency and recency of use, and optionally, on a simple phrase structure grammar. Typing the first letter of a word results in accessing the most frequent word beginning with that letter. It is also possible to input letters by using speech recognition. Successive predictions are made from a large lexicon and from a lexicon that stores new typed words. A lexicon containing word pairs is consulted when a word is terminated. If found, the word following it is automatically predicted without its initial letter being typed. The program is currently being modified to serve as part of a word processing system.

Another word prediction program has been developed for use with aphasics and children learning to spell. It also provides word prediction to aid in easier message construction. However, it does not presuppose the ability to spell a word. The goal is to use any information that the person has about a word, such as any letter it contains, the number of syllables it has or the word class to which it belongs, to access it. The word, or a list of word predictions, can then be synthesized for the user to hear and to choose among.

Emphasis has been placed on the use of substantial linguistic support to make the programs give reasonably expected choices.

TALKING TERMINALS

The most widespread use of the speech synthesis system as a technical aid at this time is as a "talking terminal." In this application, information which is printed on a computer screen is read by the device. About 300 systems have already been installed as talking terminals. This technique is most used by blind persons and persons with low vision, and has also been implemented in a number of work stations in Sweden. These work stations are typically built around a personal computer, and include a Braille display and printer as well as a speech synthesis device. The results have been quite promising, particularly in several office applications such as word processing, register handling and local switchboard operation.

There are also several distributors of screen-reading programs, for ABC80, ABC800, IBM-PC and VT100 terminals. These programs detect certain commands from normal text input such as "Read current line," "Give cursor position" or "Read word by word" which are interpreted in special routines. These routines access the appropriate text and send it to the synthesis device to be read.

DAILY NEWSPAPERS

Another application of speech synthesis for the visually handicapped is in the area of reading text which has been typeset by computer -- a common practice in printing offices nowadays. A project which has continued for several years in Sweden is to make daily newspapers available to persons with visual disabilities. (Carlson and Granström, 1986) At present, newspaper text is broadcast digitally to the homes of about 30 blind subscribers where it is stored on a magnetic disk during the night. The user can then, at his leisure, search the material for sections, headlines, or particular words with the help of a small microcomputer. The text can then be presented as synthetic speech.

SPEECH RECOGNITION

Although it has not been the major emphasis of this discussion, an area which will be increasingly useful in technical aids in the future is speech recognition. The system which was developed at the Royal Institute of Technology in Stockholm is a pattern-matching system (Elenius and Blomberg,1986). The system digitally implements a 16-channel filter bank. This filter bank covers frequencies from 200 to 5000 Hertz in bands spaced according to the critical band scale which represents the frequency characteristics of the human auditory system. Thirty-two sample points derived from this information are matched with the stored reference patterns by dynamic programming time alignment. Speech analysis and dynamic programming are accomplished using a NEC7720 signal-processing chip. Control of the recognition process and storage of the reference vocabulary are handled in the microprocessor and memory of the PC.

There are several ways in which the recognition device could be used as a technical aid. It may be used as a stand-alone unit to voice-control another device connected to an I/O-port of the PC. Such a use would be voice control of an environmental control device. It may also be used to add speech control to any already existing program. After speech input is initialized, the response strings of recognized utterances look like keyboard entries to the program to be run. A motorically disabled person capable of producing different (but consistent) utterances for each key on the keyboard would thereby have full control of any user program. One particularly useful application is word processing in which utterances would access both keys and editing commands. Another use would be to fully integrate speech control into an applications program with calls to the recognizer's special functions. This could be an especially useful tool for a disabled programmer.

A system using speech recognition has been developed as an alternative for environmental control for a motorically handicapped person who has full use of his head and neck. With his previous sipp-and-puff device he could switch on the radio, TV and lights. He could also open the door to his appartment and make telephone calls. These functions were controlled through infrared light. The system has now been augmented by a personal computer with speech recognition connected to a a programmable remote control. With the new system he can control, by speech, all the old devices but also get the full use of all the functions on standard consumer electronic devices such as TV, VCR and radio. Software on the PC gives the appropriate feedback such as selections from the telephone list. Now he can also use the PC for tasks such as creating and editing messages. It is planned to include the previously described word predictor to allow faster input of text while giving oral typing commands.

CONCLUSIONS

It is now possible for both speech synthesis and speech recognition to be used in technical aids for disabled persons. Aids which have been developed at the Royal Institute of Technology were described and ongoing work discussed. Included were word predictors, the use of text-to-speech as a voice prosthesis, a specialized communication aid called Multitalk, a Blissymbol-to-speech system called Blisstalk, talking terminals, a daily newspaper source for the blind and a communication system including speech recognition.

REFERENCES

Carlson, R. Galyas, K., Granström, B., Pettersson, M. & Zachrisson, G. (1980) "Speech Synthesis for the Non-Vocal in Training and Communication," STL-QPSR, Vol 4.

Carlson, R. & Granström, B. (1986) "Applications of a Multi-Lingual Text-to-Speech System for the Visually Impaired," in *Development of Electronic Aids for the Visually Impaired,* P.L. Emiliani, ed., Martinus Nijhoff/Dr W. Junk Publishers, Dordrecht.

Carlson,R., Granström, B. & Hunnicutt,S. (1988) "Rulsys - The Swedish Multilingual Text-to-Speech Approach," SST-88.

Elenius, Kj. & Blomberg, M. (1986) "Voice Input for Personal Computers," in *Electronic Speech Recognition,* Geoff Bristow, ed., Collins Professional and Technical Books, London.

Granström, B. (1987): "Speech technology for the visually impaired - the Swedish perspective", STL-QPSR 1/1987, pp. 29-38.

Hunnicutt, S. (1986) "Bliss Symbol-to-Speech Conversion: 'Blisstalk'," Journal of the American Voice I/O Society, Vol. 3.

Hunnicutt, S. (1987) "Input and Output Alternatives in Word Prediction," STL/QPRS 2-3.