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

Speech synthesis for the non-vocal in training and communication

Carlson, R. and Galyas, K. and Granström, B. and Pettersson, M. and Zachrisson, G.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 21
number: 1
year: 1980
pages: 013-027

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

I. SPEECH SYNTHESIS

A. SPEECH SYNTHESIS FOR THE NON-VOCAL IN TRAINING AND COMMUNICATION

R. Carlson, K. Galyas, B. Granström, M. Pettersson*, and G. Zachrisson*

Abstract

A prototype system has been built for evaluation of synthetic speech as a communication and training aid for the disabled. Based on a minicomputer and the OVE IIId synthesizer, the computer program translates any Swedish text into speech. The system is not portable but it can be moved around on wheels. Experiments have been running since 1978 in the Bräcke school in Gothenburg to test the practical use of synthetic speech in different situations as an aid for communication and education. One teenager has used the system for communication, and three primary school children have studied writing and spelling with audio reinforcement. The results are positive and valuable information has been received regarding psychological and social factors and the functional design of aids with synthetic speech.

Requirements and future development of aids with speech output is discussed.

Introduction

Ever since synthetic speech was produced with acceptable quality, hopes have run high that it would become a useful aid for the non-vocal. Although there are alternative non-oral communication methods, the spoken word, being the primary mode of communication between human beings, is superior in most situations. That is why aids with spoken output are of great importance. A close analysis of the advantages of artificial speech compared to other means of non-oral communication is made by A. Warrick et al (1977) both from the user's and the message receiver's point of view. Encouraged by their experiments with audio reinforcement, they incorporated a speech synthesizer in a classroom communication system. In their paper they also discuss some basic requirements which must be considered. These are:

- portability
- natural sounding voice and correct grammar
- large vocabulary
- individual voice for each user
- the ability to express emphasis and attitude in various situations
- an adequate speed of communication

---

* Bräcke Östergård, Special School for Motorically Handicapped Children, Gothenburg, Sweden

No system known to us can fulfill all these requirements. Portability has been achieved thanks to microprocessor technology and circuit integration. One example is the Handivoice device which has a fixed vocabulary and an additional phoneme compiling facility. Apart from portability there is a deficiency concerning the other basic requirements. Another portable speech synthesizer, developed in Finland (Karjalainen & Laine, 1976), offers for a Finnish user a general text-to-speech conversion (the Finnish orthography is very close to phonetic spelling). This device, still in the prototype stage, is being evaluated as a communication aid.

When saying that no presently existing synthesizer fulfills all the basic requirements, we do not want to minimize their value. Within their limitations they can contribute to a higher degree of independence and improved social interaction for a non-vocal individual. These devices certainly deserve a careful evaluation. Apart from the above basic requirements there are a lot of other factors which may influence the practical use of synthetic speech. Being mainly of social and psychological nature, these factors cannot be predicted. Only field tests and evaluation can give reliable answers to questions, like what users experience when an artificial voice expresses their messages and how different people react to this kind of communication.

Based on these kinds of considerations and discussions with therapists we decided to investigate the usefulness of synthetic speech for the non-vocal. Work with speech synthesis has been conducted at the Department of Speech Communication since the fifties which led to the different versions of the OVE synthesizer and a complex program for text-to-speech conversion. We felt that the speech quality was acceptable and we started to look for cooperation in evaluating its usefulness. We received positive interest from Bräcke Östergård, a special school for motorically handicapped children in Gothenburg, and we opened our cooperation with a short trial at our computer laboratory.

One student, Mikael (then 15 years old) visited us for one week and tried out the use of the synthesizer. Mikael is suffering

from CP and cannot produce any speech. He knew how to spell out words and sentences, he could type on an electric typewriter with a mouth-stick and he always carried a small letter board on his wheel-chair. This and the mouth-stick was his only communication channel. He learned easily to type on the CRT terminal board and listened to the speech he produced. We could communicate with each other and discuss how he liked the speech quality and the way certain words were pronounced. After that week it was decided to continue the training once a week via the public telephone line. Mikael used a text telephone to type on and could listen to the resulting synthetic speech on a loudspeaker (see Fig. I-A-1). In the meantime the experimental system was built and programmed.

The speech synthesis system

In the experiment to be described we make use of a prototype synthesis system first used in spring 1978 (Carlson & Granström, 1978). Let us first mention some general features of this system. It consists of a modified OVE IIId speech synthesizer (Liljencrants, 1968), an Alpha LSI 4/90, mini-computer and some general input/output facilities. What makes this system unique compared to other synthesis systems are the synthesizer and the software (the computer program). The ease of changing the software is of great value for the experiments and will be discussed later in more detail.

The synthesizer is a serial synthesizer with a special kind of glottal source (Rothenberg, Carlson, Granström, & Lindqvist-Gauffin, 1975) that makes the speech quality more natural than normally is the case. We could control physiologically related parameters like pressure drop over the vocal folds, fast movements of the lips compared to slow movements of the tongue.

Why use such a synthesizer instead of a synthesizer with built-in "phonemes" or sounds? If we use a synthesizer of the latter kind we have limited ourselves in quality and could only produce one kind of voice. The speech act is not a concatenation of discrete sounds, but rather a continuous flow of movements. These movements could be described by linguistic rules.

Fig. I-A-1. Mikael "talking" through the text-to-speech system.

The linguistic rules are the second important feature of the system. Some years ago we defined a new computer language that was closely related to a notation used in linguistics (Carlson & Granström, 1976). We use this language to define each part of the system. These rules are developed and tested on a big computer and when we are satisfied for the moment, they are moved to the prototype system.

In that way we do not have to re-program the prototype system. Each new version is described by the rules, presently around 400, for Swedish. Other languages have also been implemented in this way.

Even though speech synthesis, phonetics and linguistics have a long tradition at the Dept. of Speech Communication we still have a lot to learn about speech. As a result of the basic research efforts, the quality of speech synthesis could be continuously increased and the rule system will have to be modified over and over again. The possibility of updating the system in an easy way is hence of great importance.

If we look in more detail into the prototype system, we will find some features of importance for the experiments. First we

have rules that convert a normally spelt text into a phonetic transcription. A special set of rules takes care of numbers and mathematical expressions.

A lexicon makes it possible to specify the pronunciation of a certain word. In the same way parts of a sentence or even complete sentences can be stored and called by abbreviations or "codes" by the user. This whole process is programmable by the user. New entries could be inserted, old deleted or examined. This facility is one way of speeding up the message entry. In entering a dialogue or discussion, especially via telephone, this is very important.

Phonetic rules convert the phonetic text into control parameters for the synthesizer. These rules include rules for intonation, direction, and coarticulations. These three areas are of great importance for the final quality.

Some new features have been added to the system during the experiments. It is now possible to get each word pronounced when entered before the full sentence is read out. This is helpful in at least two ways. First the user could listen, accept, or change an entered word and second, the listener will maintain his interest during the formation of a new sentence. When the whole sentence has been entered, it is read out and new prosodic rules give the sentence a better quality.

The speech rate could be controlled by a separate knob. A listener is very sensitive to this parameter and speech synthesis is often regarded as either too fast or too slow ("boring").

We regard the present prototype system as more or less completed. We will try to include as much knowledge from the experiments as possible in the next generation of synthesis systems. The future will be touched on in the final remarks in this paper.

The system has been in use at the Bräcke School for two years. In this part of the paper we will present experiences from two different studies. One of the studies concerns the use of synthetic speech as an educational aid for the first and second grades in primary school. The other is a continuation of the experiment with Mikael, now 18 years old.

Mikael

Mikael, diagnosed as suffering from CP with dystonic syndrome with anartria, was engaged to work on this project three years ago. His most useful motor function is located around the head and he can write on a keyboard with a mouth-stick. His left hand function is sufficient to make it possible for him to control an electric wheel chair. At present he communicates by pointing with a mouth-stick at a letter board or writing on a Cannon Communicator.

Since the spring of 1978 when the first prototype system was installed at the Bräcke School, Mikael has had several short and intensive periods of working with the synthesizer, before he left school in June 1979. Now he receives job training. Since October 1979 he and Gerd Zachrisson meet once a week for extensive training sessions.

While in the school, Mikael received special training in Swedish. His teacher was very positive about using the synthesizer for this purpose. She could establish a better contact with her pupil than before and the increase in his motivation was apparent.

At the beginning he had difficulties in spelling. During his training he improved remarkably though he is still uncertain when facing difficult words or others unknown to him. According to his teacher the positive effect on motivation was carried over to other schoolwork not assisted by synthetic speech. Mikeal himself says that the synthesizer made training interesting and fascinating for him. He found it very useful that he could immediately hear when he made an error.

Gerd Zachrisson had the most extensive training with Mikael. Short and intensive periods were followed by longer breaks. He explains the reason being demands put on him from everywhere, which he felt so pressing that he couldn't mobilize any energy for extra activities. However, to work with the system was fascinating for Mikael and he was interested to find out all the possibilities of the computer programs. We received valuable points of view and a lot of suggestions for improvements.

The use of synthetic speech for communication purposes was for us the most interesting task. In their sessions Gerd and Mikael spent more and more time for communication with each other. Gerd explains: "In the beginning Mikael had trouble thinking of something to write. This has become prograssively better and now when he uses OVE III we both experience it as a natural conversation. He can now use OVE to explain, ask questions, or ask for help as well as answering questions and express opinions. Training sessions become more and more like an ordinary conversation (between two people)."

Mikael has started to talk about his situation, the drawbacks of his handicap and how technical aids in the future could improve his life. He is interested in taking part in evaluation programs and other work which can lead to better aids. Quite naturally, their discussions are often about communication aids. Mikael considers speech output to be very important and he has mentioned several situations where he would need such an aid. When meeting small children, for instance. Children, who do not know him, sometimes talk to him. When they do not get any answer they become aggressive, start teasing. In a situation like that he would like to be able to scream loudly. There are small children Mikael knows and he would like to say something to them sometimes. He has a baby cousin and he would like the baby to learn from the beginning to talk to him. Perhaps he could ask single questions like: "Can you say papa?"

Another example of such a situation is telephoning. Being able to call and chat with a fried might be more important for a handicapped person than for others. Call for help, for a taxi, or leave a message would mean a much more independent life. Synthetic speech will also make it possible to participate in discussions and other group activities.

Mikael has also contributed many ideas as to how an aid with synthetic speech should function in practice. The most important demand is to increase the speed of communication, for example through programmed abbreviations which is a feature we have tested. Mikael has found that it was best to use letter codes for words and

phrases. For example JV = Jag vill... (I want ...); MNP = Mitt namn är Mikael Pettersson (My name is Mikael Pettersson) etc.

He has also demands as to portability, a little keyboard which can be used with or without write-out on paper tape or light display. He also wants a keyboard he can see in the dark: "I have never been able to talk in the dark". He would also like to have adjustable sound volume and a memory for longer messages.

He assumes that synthetic speech would have been useful to him at a much earlier stage in his life, partly because it would have made it easier for him to learn to read and write and partly because it would have made independent communication possible while reducing his general dependence on others. When one always communicates through others, for example an assistant in class, one becomes more passive. People are not always willing to transmit the entire message, especially emotional expressions.

In response to a discussion about possible psychological problems in using a non-standard aid and talking with an artificial voice he said that he does not experience any difference between the use of Cannon Communicator and synthetic speech. The only difference is that a speaking aid is so incredibly more practical and has a wider range of functions.

He is positive to the suggestion that small children use synthetic speech in connection with Bliss symbols or pictures. It is important that children are introduced to the use of technical aids (even complicated ones) at an early age.

In the beginning stages of this test we suggested to place OVE III in the classroom. At that time Mikael was against this, but now he says that that was due to his unfamiliarity with the system. Today, however, he would gladly use OVE III at his jobtraining.

Synthetic speech in primary school

Three small CP-children with anartria aged 8-10 years have also participated in our experiments. Individual training has taken place twice a week for one year. It was difficult to train more

often because of the children's schoolwork and other activities and therapies. It was not possible to train these beginners with OVE III in the classroom, because it would interfere with the other children's activities. The training sessions were 10-60 minutes long, but sometimes, especially in the beginning, only part of this time was effective. One of the children could use his hand to write on the keyboard. The other two with more serious motor handicaps used the head stick and for them the keyboard operation was time-consuming and laborious. At the beginning of the experiment the children could read and write a few simple, memorized words and names, but did not seem interested in using this knowledge. They communicated with Bliss symbols; one by pointing, the other by means of a head lamp. All the children have enjoyed the use of OVE III and they always were anxious to be first in turn.

All three had different qualifications and they have worked differently, but their development has been common in many ways. They all started with playing with letters, writing their names and simple words, such as "mamma" and "pappa". In the beginning the sound games consisted of isolated letters without any systematic choices and this developed into a complex series of random sounds, a kind of "meaningless babble". Later they tried to sound out the words, especially one of the boys, and they also wrote isolated words which corresponded to printed test or Bliss symbols. After a few months they showed interest for syntactic structures in the form of simple, incomplete sentences of the type: "Åka mamma" (Go mama); "Hej pappa" (Hello papa). They also curiously and constructively examined the knobs and keys on the system, tested the adjustments of the loudness, speech rate and the delete function. The possibility to make the system whisper was discovered by the children themselves, and was frequently used. They wrote a word and then tried the same in a whispered voice. Half a year later one child left the school but the two children using head-sticks have continued the experiments. They have made great improvements the last months. They have mainly been interested in writing full sentences, and very often they start with Bliss symbols to tell what message they want to express. Playing with sounds is a more unusual thing now. Very often the message is

directed to a certain person, e.g. the assistant: "Hey. Could you help me to the bathroom?" The two children very often plan messages to each other, e.g. "What are you going to wear to the party?". Often they need help with the spelling, but mostly they have problems with the grammatical construction of sentences. They feel disturbed when incorrect sentences have been formed. The help with spelling has changed from just telling the letters in a word to carefully pronouncing the word.

Some differences between the children could be noted. The boy, who could write with one hand, was improving faster than the other. He worked fast and concentrated, but only for short periods. He started early to form words by combining sounds. He also used known words and wanted to write what he thought without any help. He could immediately write longer sentences and was amused by that. After two months he started to make two-word sentences that rather soon were expanded to three words. He left the school after half a year of experiments. During this short time his ability to read and write was improved and he also became more interested to spontaneously use an ordinary typewriter.

The other boy has a more severe motor handicap and could only with great difficulty control the keyboard by using a head-stick. He showed a similar but slower development compared to the first boy. He was often caught in a meaningless and stereotype playing with sounds. As a start he only managed to copy short words but that has improved and presently he writes longer and more complicated words. He also managed to make new words with help of hearing them pronounced carefully. Lately his interest has increased to use the system to write messages to someone else. Then he tries to spell what he wants to say, but he still needs help with some spelling and with the forming of sentences.

The third child, a girl, also uses the head stick, but then she can manage it rather well. She was different from the other children, since she had a "language" and she used the Bliss symbols very well to communicate. She wanted to write complete sentences at once and started with Bliss symbols. It took a long time for her to find

out what to write. She had more problems to copy words than the boys. She also had problems to discriminate different letters, e.g. "d" and "b". Getting the word pronounced or spelt out did not help. Instead she preferred to get it written down with big letters on a paper. Even simple words, like "mamma" and "pappa", were difficult. After several months' practice she became interested in playing with sounds and she still continues, but not as actively as the boys. Presently, she is able to copy words and is also doing that by listening. She does mix the order of letters as frequently as before. Even if she still plays with sounds, she has started to use the system to make small messages and is eager to make grammatically correct sentences.

Generally speaking, we think that the result has been very positive. We also think that synthetic speech increases the children's interest in reading and writing. First, the working situation is more motivating for the children. Second, they get an immediate visual and auditory feed-back, thus reinforcing their learning. We also noted that for at least one of the children playing with letters and sounds resulted in spontaneous attempts to form words.

A lot of children show a tendency to reverse the order of letters in a word without noticing the error. Using synthetic speech the errors were immediately noticed and corrected.

Future use of speech synthesis as a technical aid

We hope to develop better routines and schedules that permit an integration of speech synthesis in the education, something that is encouraged by the staff of the Bräcke School. OVE III will be equipped with a printer and computer-aided training program. This will make it possible for students to work independently with the system without constant supervision by the teacher. We also hope that more students will be given the opportunity to use the device in language training and communication. To this end more individually adjusted keyboards or other input facilities have to be connected.

There is also a great interest in using synthetic speech in connection with Bliss symbols in the classrooms. The effort to develop

written language communication is often very great and not always successful. That is why it is so important that the non-verbal children at an early stage get alternative means of communication, to help them to work independently, interact in class or other groups, use the telephone etc.

Synthetic speech is also efficient in teaching Bliss and encourage the Bliss-using children to produce complete sentences. We believe that it is important to start using synthetic speech very early.

At this point in time we could only speculate what effects such an early start could have for the future development.

Discussion on the future development of speech prosthesis

Let us return to the basic requirements listed in the introduction. Based on our experiences we want to focus on areas where future development and research is needed. We will also indicate in what problem areas we plan to contribute in the context of the Swedish speech prosthesis project.

Portability is the requirement which was especially emphasized by Mikael. In the educational setting, portability is not important but as soon as the device is to be used as a communication aid, it has to be both portable and battery operated. The advances in integrated circuit technology will make that possible and we plan to have a production prototype of this kind ready within one year.

A natural sounding voice and correct grammar. Even though the intelligibility of the present voice is satisfactory the naturalness still has a lot to gain. For fixed vocabulary systems, coded human speech is still superior but for unrestricted text-to-speech synthesis that method is out of the question. The improvement of quality is a long-range research topic. Since continuous improvements could be expected for a long time to come, it is important that speech prosthesis could easily be updated with better programs.

Large vocabulary. The vocabulary should provide the user with full capability to express any thought. In most cases this implies

a need for unrestricted vocabulary. This goal can be achieved only by a system for text-to-speech conversion which, however, presumes reading and writing ability of the user. In very special cases with primitive communication needs, a fixed vocabulary system could be used; at the present stage preferably with coded human speech.

An individual voice for each user can in most synthesizers only be achieved by changing the pitch within a rather limited range. Outside this range the voice quality deteriorates. Most synthesizers are not capable of producing a child or a female voice with the same quality as a male voice. It is not sufficient to just raise the glottal pitch or formants. There is a need for more basic research. This is going on the whole time and improvements are to be expected.

Voice type differences and dialectal variations ask for a rather complex re-programming of the synthesizer on the parameter level. The details of this reprogramming are not at present fully understood. Again, this makes the programmability of the speech prosthesis necessary. In the experiments at the Bräcke School the voice quality per se was not much discussed. It is conceivable that the more regular use of the speech synthesis as a communication aid makes the demand of an appropriate individual voice more apparent. Apart from the psychological advantage of a well-fitted voice prosthesis it is of course convenient in a group situation with several non-vocal members to be able to identify the speaker by voice.

The ability to express emphasis and attitude is very important in normal voice communication. From the research point of view there exists a quite good understanding of emphasis, whereas the means for conveying attitudinal information is much less studied. Two ways of implementing these factors in a speech prosthesis exist. First, it is possible to add extra information to the text input that controls rules for the realization of emphasis and attitude. This presupposes a good understanding of the phenomena and also a very conscious use of the extra symbols by the user. The other possibility is that the user interacts directly with the synthesizer, e.g.,

by controlling the voice inflexion. We have carried out some preliminary tests with this, but it is still unclear how useful such a facility would be, especially since many of the potential users of the speech prosthesis also have severe motor handicaps.

The speed of communication is of essential importance when it comes to practical use of the communication aid. In conversational situations the user can be by-passed if the listeners have to wait. The worst problem in using a speech prosthesis is the often very time-consuming input of the message. There are two ways of dealing with this problem. One is to adjust the input facility to the user in an optimal way thus maximizing the speed of input. The other is to reduce the needed input by special symbols, abbreviations, or pre-stored utterances. Through the user-programmable dictionary we have one solution of the latter kind, which has been very useful and appreciated in the experiment at the Bräcke School. The harder problem with the individual optimization of the input facility still needs a lot of work. There is no single solution to this problem, rather the optimal solution appears different for most individuals.

Among the things we plan to try are enlarged keyboards, communication boards and a Canon Communicator. In a longer perspective we want to implement a Bliss symbol input to the speech prosthesis and also a linguistically organized word memory system.

The development of communication aids with speech output is only partly an engineering problem. There is a need for more humanistic research to study the psychologic, linguistic, and human engineering factors involved in communication with an artificial voice. To fully make use of this kind of communication aids, there also has to be a development of educational programs.

There are several groups working on problems related to speech prosthesis around the world. It is important that information about projects and results are spread to other groups. An intense international cooperation is desirable, in the interest of the disabled population.

Acknowledgments

Our thanks are due to Professor Gunnar Fant, who together with Professor Olle Höök, head of the Institute of Rehabilitation Medicine, University of Gothenburg initiated this project. We are especially grateful to Lena Samuelsson, who participated in the early experiments with Mikael, and Ingemar Olov, M.D., head physician at Bräcke, who has constantly supported and inspired the project from its start. Without the technical ingenuity of Björn Larsson of the Department of Speech Communication, the adoption of the synthesis to the need of the students had been impossible. We also want to thank the staff and pupils both at the Bräcke School and the Department of Speech Communication for support and valuable discussions.

The support from the Swedish Board for Technical Development and the State Inheritance Fund is also thankfully acknowledged.

References

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

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

Karjalainen, M.A. & Laine, U.K. (1976): "Development of communication aids for the handicapped based on speech synthesis techniques", Digest of the 11th Int.Conf. on Medical and Biological Engineering, Ottawa, 304-305.

Liljencrants, J. (1968): "The OVE III speech synthesizer", IEEE Trans. AU-16, No. 1, March.

Rothenberg, M., Carlson, R., Granström, B., & Lindqvist-Gauffin, J. (1975): "A three-parameter voice source for speech synthesis", pp. 235-243 in Speech Communication, Vol. 2 (ed. G. Fant), Almqvist & Wiksell, Stockholm.

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.