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

Evaluation of a text-to-speech system as a reading machine for the blind

Carlson, R. and Granström, B. and Larsson, K.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 17
number: 2-3
year: 1976
pages: 009-013

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

II. SPEECH SYNTHESIS

A. EVALUATION OF A TEXT-TO-SPEECH SYSTEM AS A READING MACHINE FOR THE BLIND

R. Carlson, B. Granström, and K. Larsson *

Abstract

A preliminary study on the feasibility and usefulness of speech synthesis in automatized text reading for the blind has been carried out. Data on intelligibility and learning effects of the synthetic speech has been gathered together with subjective reactions to the speech quality. A group of visually handicapped subjects appeared to gain useful knowledge of the synthetic speech in a very short time. This knowledge is kept for a long time. Subjects are in general positive to the prospect of getting textbooks and newspaper material in speech synthesis.

Introduction

In the spring 1976 a thesis work was carried out with the goal to study speech synthesis in relation to its use by visually handicapped, Larsson (unpublished). This was done in cooperation with The Swedish Association of the Blind (DBF). Different aspects were studied covering both intelligibility, acceptability, and learning effects as well as possible applications and the design of the ultimate tool, a reading machine for the blind. The latter part included a study of the OCR market and the handling of text in printing offices. Only the first part concerning speech quality and learning effects should be dealt with in this paper. It should be noted that the result is dependent on the status of the system in January.

The synthesis system

The synthesis system has been described elsewhere, see Carlson and Granström (1976); Liljencrants (1968); and Rothenberg et al (1974), but some summarizing remarks might be necessary. The system is built around a minicomputer and a terminal analog speech synthesizer. In the computer a synthesis program formulated as a sequence of rules is working. The first rules transform a digitally coded ordinary spelt text into a phonetic transcription. This transcription is by the following rules used to derive control parameters to the synthesizer. The user could either listen to the synthesized speech immediately or record the result on a tape recorder. The latter is the case in our present experiment.

---

* K. Larsson was a thesis student at the Dept. of Speech Communication, 1975/76.

Test material I

This material consisted of eight different lists of 25 unrelated synthesized sentences with a mean length of six words, Korsan-Bengtsen (1973). The speed was about 140 words/min with a pause of 10 sec between each sentence.

Test material II

Two test lists were added to study increased speed (200 words/min). This was done either with a higher tape speed (list S), normally used by visually handicapped or by a higher clock-frequency in the system, keeping the fundamental frequency, and formants, on an appropriate level (list C).

Test material III

This material consisted of a synthesized short story of about 7 minutes' duration.

Subjects

Eight visually handicapped persons acted as subjects. Data on these subjects are presented in Table II-A-I. They had never listened to synthetic speech earlier and no training was offered before the actual test. In the table their previous use of recorded books has been noted to show how accustomed they were to listen to this type of material.

The experiment

Two test sessions were arranged with a one-week interval. Each session contained four test lists of type I. The subjects' responses were recorded and analyzed. No feedback concerning the correct interpretation was given. The test lists were balanced so that each list was presented once on each position in the list sequence, i. e. each list acted once as the first, once as the second, and so on. The difference between the lists could therefore be disregarded.

In the end of session I the short story (test material III) was presented and spontaneous responses were collected. In the end of session II the speeded material (test material II) was used as stimuli and responses were recorded and analyzed in the same way as the other lists.

Finally, the subjects were asked questions about the test as well as the synthetic speech in general as a tool.

Subject number	Year of birth	Visual handicap	Experience in reading of recorded books	Normal reading activity book/month	Is synthesis a good tool?	Would you like to use complete synthesized books?	Do you think the test session has been tiring?	Mean word response score for all lists
1	1912	.20	5 years	4	yes	no	no	65
2	1951	blind	10 years	2	"	yes	yes	81
3	1952	blind	10 years	4	"	yes	yes	79
4	1951	blind	20 years	1	"	yes (only text books)	yes	91
5	1951	blind	15 years	4-5	"	yes (peri-pheral)	no	75
6	1949	blind	13 years	4	"	- " -	yes	91
7	1957	.03	no	no	"	- " -	no	72
8	1953	.10	16 years	1	"	- " -	no	63

TABLE II-A-I. Data and results from questionnaire and lists.

Result and discussion

The responses from the test lists were analyzed in two ways: (a) correct words, and (b) correct sentences. One incorrect word can by this method spoil a whole sentence. That is why the sentence curve is well below the word curve in Fig. II-A-1 presenting the result. As can be seen in the figure the learning effect is well marked, i.e. the intelligibility increases rapidly. One week interval does not cause the subjects to forget their gained knowledge of the synthesis "dialect". The result of speeded synthesis (200 words/min) is indicated in the figure. It is obvious that a higher synthesis rate is superior to a higher tape speed.

In Table II-A-I the mean result for each individual is given. The figures indicate that those who think the test is easy have a lower score than those who think it is harder or more tiresome. Since the two groups are so obvious we have calculated the word score corresponding to each group. The result is presented in Fig. II-A-2. Group I contains subjects 2, 3, 4, and 6, while group II contains subjects 1, 5, 7, and 8. The fact that these curves do not increase monotonously is probably due to the different difficulties of the lists. The difference between the groups, is, however, more emphasized in the first lists than in the later indicating that the invested effort will not determine the ultimate result.

Conclusion

Summarizing the answers on the questionnaires, given to the subjects in the end of the experiment, all subjects thought that the speech synthesizer will become a useful tool in the near future. Synthesis should in the first hand be used for textbooks, peripheral literature, and newspaper material and to shorten or eliminate the time between printing publications and recordings as talking books.

In the result we find a rapid learning effect that is present even during a one-week interval. This effect is present despite the degree of concentration.

The result of this experiment is encouraging in the sense that it has demonstrated the feasibility and usefulness of a reading machine for the blind developed on our concept of speech synthesis.

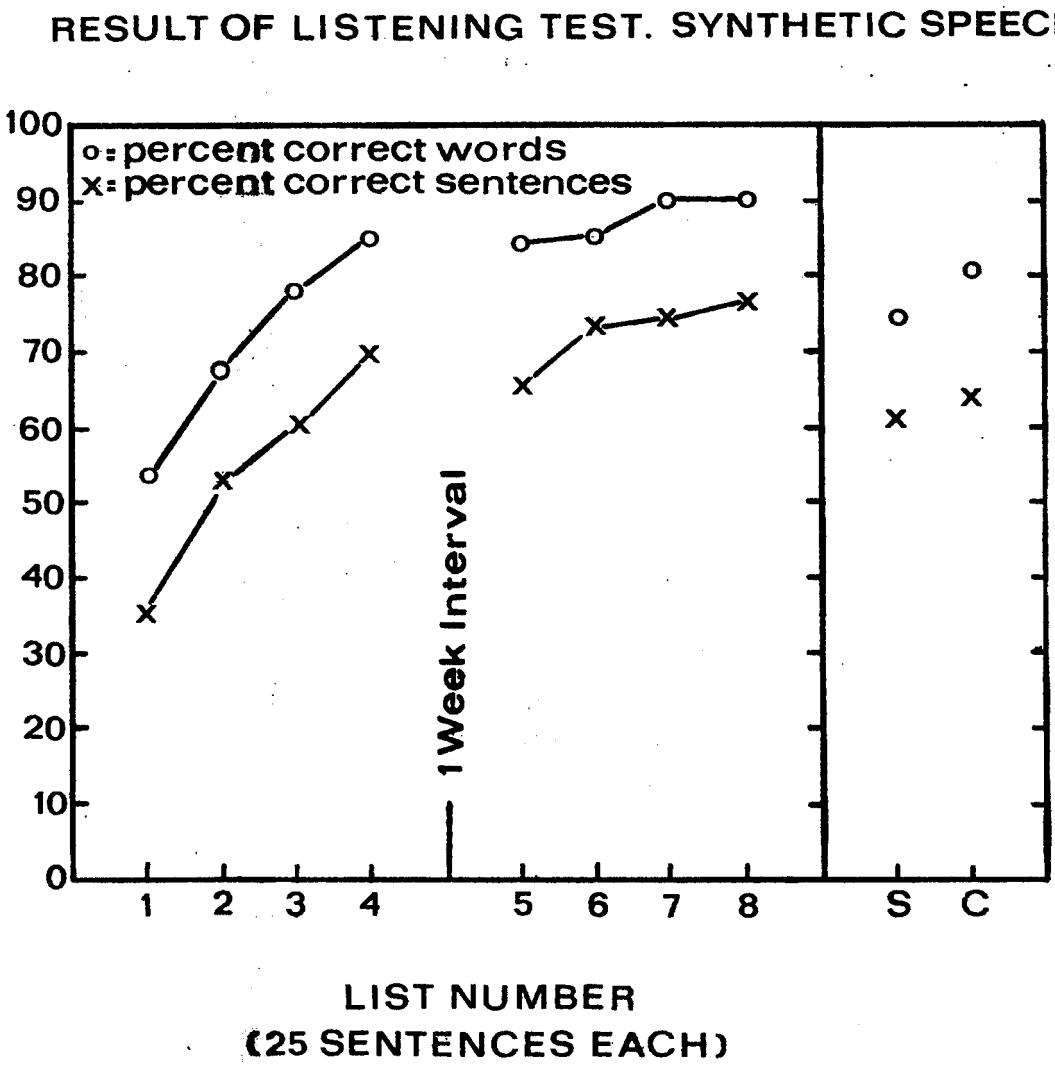

Fig. II-A-1. Result of listening test using testmaterial I and II. See the text.

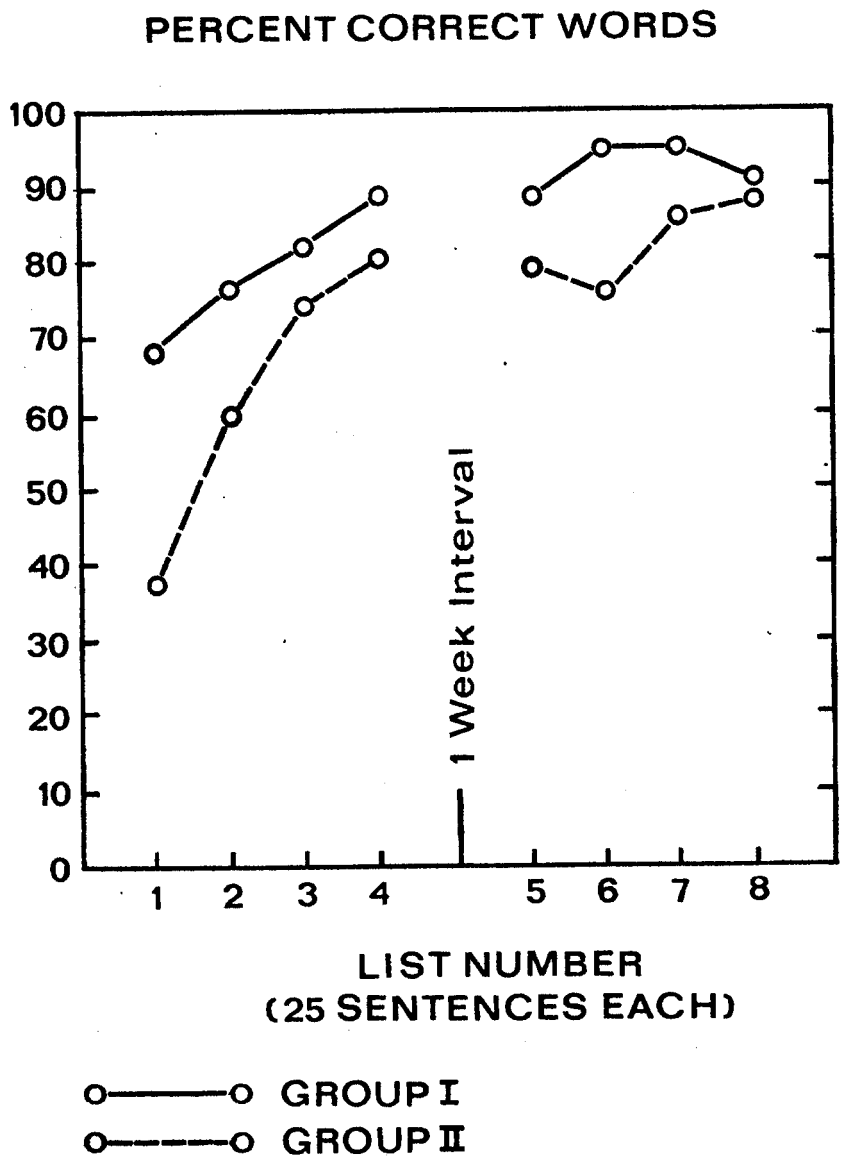

Fig. II-A-2. Result of listening test using testmaterial I.
Group I contains subjects 2, 3, 4, 6.
Group II contains subjects 1, 5, 7. 8.
See also Table II-A-I.

Acknowledgments

This work was supported by the Swedish Board for Technical Development under Contract STU 74-3547.

References:

CARLSON, R. and 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, pp. 686-689.

KORSAN-BENGTSEN, M. (1973): "Distorted Speech Audiometry", Acta Oto-Laryngologica, Suppl. 310.

LILJENCRANTS, J. (1968): "The OVE III Speech Synthesizer", IEEE Transac. on Audio and Electroacoustics, Vol. AU-16, pp. 137-140.

LARSSON, K. (unpublished), master thesis, KTH, March 1976.

ROTHENBERG, M., CARLSON, R., GRANSTRÖM, B., and GAUFFIN, J. (1974): "A Three-Parameter Voice Source for Speech Synthesis", pp. 235-243 in Speech Communication, Vol. 2, ed. by G. Fant (Proc. of the Speech Communication Seminar, 1974), Almqvist & Wiksell Int., Stockholm 1975.