

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

**Evaluation of centrally
produced and distributed
synthetic speech**

Hampshire, B. and Rüden, J. and Carlson, R.
and Granström, B.

journal: STL-QPSR
volume: 23
number: 2-3
year: 1982
pages: 018-023



**KTH Computer Science
and Communication**

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

II. SPEECH SYNTHESIS

A. EVALUATION OF CENTRALLY PRODUCED AND DISTRIBUTED SYNTHETIC SPEECH

B. Hampshire, J. Riden*, R. Carlson, and B. Granström

Abstract

The purpose of this project was to illuminate new information technology and its application to visually handicapped. Production techniques and user acceptance of synthetic speech was studied.

During one month, a weekly "small-add" paper was distributed with the help of synthetic speech. 75 visually handicapped subjects received a synthetic speech edition of the newspaper the same morning that the inkprint version became available in the shops.

Listener age had little influence on the rated speech quality. It was as an average found better than fair despite the difficult text material.

Several suggestions were made for future productions with help of the present technique.

Introduction

SRE Tal & Punkt, the production unit within the Swedish Federation of the Visually Handicapped, has, during several years, in cooperation with the Dept. of Speech Communication and Music Acoustics, carried out some tests with synthetic speech. An example of the material that has been converted to synthetic speech is the weekly magazine "Från Riksdag och Departement" (From Parliament and Ministries). This was distributed to some visually handicapped social workers. This test formed a part in a longer project entitled "Provision of Information to Visually Handicapped in School and Work". The purpose of this project was to illuminate new information technology and its possible applications for the visually handicapped. For the conversion to synthetic speech, as with the current project, the speech computer "Texttalk" was used. This was developed at the Department of Speech Communication and Music Acoustics of the Royal Institute of Technology (KTH) in Stockholm (Carlson et al, 1982).

Previous speech intelligibility studies, with an earlier version of this text-to-speech system, have shown that the synthetic speech was readily understandable when listening to paragraph length material after some initial tuning in to the somewhat peculiar "dialect" (Carlson et al, 1976). Some subjects reported a certain amount of fatigue after a

*SRE Tal och Punkt Development Unit, S-122 88 Enskede, Sweden

long listening session. This test was run on a simulated version of the text-to-speech system. It is of considerable interest to investigate how the present further developed microprocessor based module works in practical situations, reading different and more difficult texts.

The increasing developments of information technology, and its effect within the home and at work, makes it important to investigate how different user groups among the visually handicapped are situated with regard to these new possibilities. Furthermore, research in the behavioral science aspect of this problem is not very extensive. The potential applications and influence of synthetic speech are numerous, it is important, therefore, to emphasize the area of use which this project focussed upon. This was the use of a central facility where digitally encoded text can be converted to synthetic speech and distributed to readers.

This should be contrasted to distributing material in digital form and providing readers with their own reading machine with which they can search efficiently in the stored digital text and convert to synthetic speech when they choose to read something. This application would have the advantage of allowing more effective search, retrieval, and storage of information by the reader. However, relatively sophisticated and expensive equipment would be required by each reader. This would not be appropriate to the majority of the visually handicapped population today, before personal text-to-speech systems are commonplace and efficient distribution facilities for digital texts or inexpensive OCR are created.

A centralized facility has the advantage of making the material cheaper to produce and easier to distribute. Furthermore, the reader would not require any special equipment other than an ordinary cassette player. Longer material is, however more difficult to handle if it is distributed in this form because of the relatively limited indexing possibilities.

Tal & Punkt is naturally interested to know more about the potential interest for centrally produced synthetic speech and the acceptability of such a medium for relatively passive listening.

In addition, more experience was needed of the actual technical production of such material, mainly in the editing of digitally encoded material to a form where it was appropriate to convert to synthetic speech.

Tal & Punkt, therefore, initiated a project to evaluate the production and the acceptability of synthetic speech. The evaluation would be

based on the two major components of the project:

1. Production techniques and problems illuminated from a four-week long distribution of a weekly "small-ad" newspaper in synthetic speech.
2. User acceptance of this material based on a questionnaire follow-up to the distribution.

Project description

After considerable discussion and some preliminary tests with different kinds of material converted to synthetic speech, a weekly newspaper consisting entirely of "small-ads" was chosen for conversion and distribution in synthetic speech form over a period of four weeks. The main reasons for choosing this type of material were:

- the text was available in digital form,
- such material is not otherwise directly available at all to visually handicapped people (in Sweden there is furthermore no broadcasted advertisements)
- it is material which is of potential interest to all members of the visually handicapped population, i.e., independent of age, sex, educational level, interests, etc.
- a considerable interest for such material was shown by the visually handicapped people who listened to the test material; so much so that it was considered not to be too great a risk of people gaining a negative attitude because of the relatively difficult nature of the text
- the highly structured nature of the text (i.e., the paper) was divided into about a dozen subject areas, each of which was further sub-divided into quite specific areas of interest
- no realistic alternative method for making this material available to the visually handicapped exists bearing in mind that it must be available at the same time as the ink-print version. Fast access is relatively more important than esthetic aspects on the speech in this situation. In reading, e.g., fiction, the priorities are possibly reversed.
- this type of text is difficult to read both for man and machine since it uses a lot of non-standard abbreviations, lack of punctuation and frequent spelling errors. Our impression is that sighted persons are often in doubt of the intended interpretation of the advertisement.

All subjects took part voluntarily by ringing a telephone-answering machine and saying which sections of the newspaper that they would like to receive (they had earlier received a copy of the content headings in the newspaper). A condition for their taking part in the experiment was that they should answer a questionnaire after the completion of the experiment. This questionnaire was conducted over the telephone and a 100% response was obtained. The age of the subjects ranged from 18 to 81, with a mean of 42.1 years. Totally 75 persons participated (60 male

and 15 female).

The project was carried out successfully - the synthetic speech edition of the newspaper arriving by post at the readers' home the same morning that the newspaper became available in the shops. Table 1 illustrates the production procedure for this experimental distribution.

Results

Numbers 34-37 of Gula Tidningen were distributed in edited form, according to the participants' interests. 95% of the requested synthetic speech editions reached their readers on the same morning as the ink-print edition became available. The final experimental edition included a total of 10 C-90 cassettes, which corresponded to approximately half the complete ink-print edition of the newspaper.

Table I. Production scheme for experimental distribution of synthetic speech newspaper

SRF Tal & Punkt (Development Unit)	Tuesday	Telephone-answering machine checked to see which sections of the newspaper had been ordered
Siffer-Service/ Liber (Printers)	Wednesday morning	Compositor's tape fetched from printers
Tele-ekonomi (Equipment manufacturer where special computer services are bought)	Wednesday morning/ afternoon	Text "cleaned" of compositor's codes and sorted according to sections requested. Text divided into digital equivalent of 45 min of synthetic speech; digital cassettes transported to Dept. of Speech communication, KTH
Dept. of Speech Communication	Wednesday pm/Thursday am	Text converted to synthetic speech with help of TEXT-TALK. Audio master tapes transported to SRF
SRF Tal & Punkt (Talking Book Dept.)	Thursday	Tapes duplicated onto cassettes; cassettes distributed by post
Ink-print edition available on Friday; synthetic speech cassettes delivered by post Friday morning		

The time subjects spent in listening to the recordings changed during the experiment. 37% listened one hour or more to the first edition compared to 60% for the last edition. Several subjects reported that they actually bought things advertized, one accuired a second hand juke-box!

Reactions of nearly all the readers were positive, perhaps to a large extent due to the high motivation of the readers, which seems to be a very significant factor in acceptability. At the same time, many useful negative, although constructive, comments on the quality of the speech were received.

The subjects' estimate of the speech quality is shown in Fig 1, where the scores are divided according to age. The youngest age group gives a somewhat higer rating of the speech quality. Interesting is that the post-retirement age group seems to understand and accept the synthetic speech fairly well especially considering the low penetration of Braille in this group.

Furthermore, what was particularly interesting from the point of view of the project was that there were many suggestions for what kind of material readers would like and would accept in synthetic speech, if this was the only method of making it available within the time, economical, and practical constraints which existed. Examples of mentioned material are: continued distribution of "small-ads", encyclopedia, telephone directories, zip-code listings, radio and TV programs, time tables for public transportation, recipie books, municipality information, mail-order catalogues etc.

Conclusions

In the first hand the highly miniaturized synthetic speech equipment - TEXT-TALK - has its greatest application as an output alternative in personal, sensory aids. However, there is a role for this equipment in the centralized production of synthetic speech where the user needs topical information quickly and cheaply and is not too bothered about the quality as long as it is not too difficult to understand. What this simple introductory experiment has shown is that the tolerance level for synthetic speech is high when motivation is also high; that the quality of the output from the presented text-to-speech system is "understandable" for the majority of people is beyond all doubt.

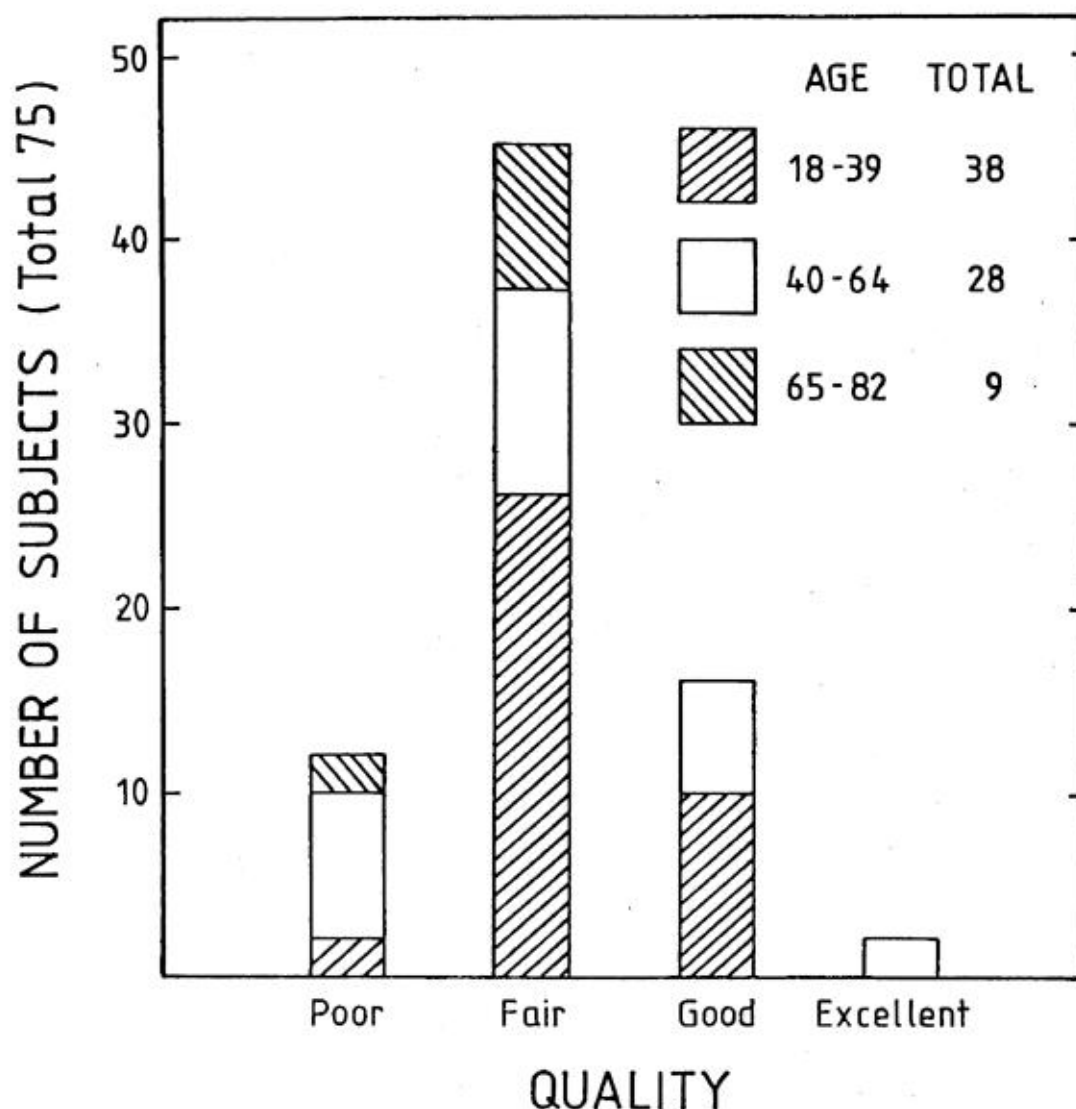


Fig. 1. Estimated quality of the synthetic speech.

How synthetic speech stands in relation to human speech and Braille is more difficult to comment on. This will be strongly dependent on each individual and the situation they are in.

In general, it is possible to maintain that synthetic speech has a place to fill. But even if development is going fast it is still a very long way to human, dramatized speech produced by machine. To obtain this, the complete "thinking computer" must be available and controversy still reigns as to whether such a machine is possible.

References

CARLSON, R., GRANSTRÖM, B., and LARSSON, K. (1976): "Evaluation of a text-to-speech system as a reading machine for the blind", STL-QPSR 2-3/1976, pp. 9-13.

CARLSON, R., GRANSTRÖM, B., and HUNNICUTT, S. (1982): "A multi-language text-to-speech module", Proc. ICASSP 82, Vol. 3, pp. 1604-1607.