A Plan for the Field Evaluation of an Automated Reading System for the Blind

P. W. NYE, J. D. HANKINS, T. RAND, I. G. MATTINGLY,
and F. S. COOPER

Reprinted by permission from IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS
Vol. Au-21, No. 3, June 1973, pp. 265-268

A Plan for the Field Evaluation of an Automated Reading System for the Blind

P. W. NYE, J. D. HANKINS, T. RAND, I. G. MATTINGLY, and F. S. COOPER

Abstract—After more than two decades of research it is now possible to construct a high-performance reading system for the blind that will produce synthetic speech from printed text. The entire process can be carried out automatically by computer and associated special-purpose devices. As a first step toward the eventual deployment of a reading system, we have begun an evaluation study in collaboration with faculty and students at the University of Connecticut and with trainees at the Veterans Administration Eastern Blindness Rehabilitation Center. Questions to be answered concern the comprehensibility and educational uses of the output and the technical and economic resources required to make automated reading services accessible to progressively larger groups of blind people.

Introduction

Our objective is to obtain an improvement in the variety, the quantity, and the speed of delivery of spoken material for blind people. We believe that this objective can be achieved through the use of an automated reader. However, with so many literate human readers presumably available, it may appear somewhat incongruous that effort should be spent on developing *a machine* to read books to blind people. A few words of explanation, therefore, are in order.

The justification for the work we are undertaking stems from the fact that the direct supply of reading matter to the blind is currently extremely slow and inadequate. This situation imposes severe educational, vocational, and recreational handicaps on blind people of all ages, but to blind college students, in particular, the limitation is critical. At the present time, blind students rely chiefly on Braille for note taking and on Braille, Talking Books, taped voice recordings, or face-to-face readers for more extensive material. However, text books are often unavailable in Talking Book or Braille form or are delivered after long delays, sometimes up to several months. Ultimately, we believe, the solution lies in the establishment of a national network of Reading Service Centers utilizing reading machines capable of generating synthetic speech at a rate many times faster than natural speech (a rate of twenty to thirty times faster would be practical) and of recording the output on tapes moving at a proportionally fast rate. The speech could then be replayed by the listener at a normal speed. Braille could be provided as well when desired. These centers could be based on large regional libraries and would rapidly provide taped synthetic speech in response to requests made by mail, telephone, or in person. The service they would provide could not fail to have significant economic and social value by permitting a far larger segment of the blind population to contribute their skills to society.

The goal of a national network is, of course, far reaching and, in comparison with current expenditure on reading services for the blind, is likely to be considered expensive. But the problems to be faced in establishing such a network are not only economic. As stated elsewhere [1] there are often many other difficulties to be met in our society in establishing an effective interface between a technical capacity and its potential field of application, and these are exemplified in the field of sensory aids for the blind. In fact, most of the difficulties are acutely visible in the whole bioengineering field [2].

The research work on the development and improvement of synthetic speech has been in progress for a number of years. Further progress can be expected. Nevertheless, we believe that the point has now been reached when it is necessary to evaluate our progress and to determine whether the speech is good enough to apply in its present form. Our reasons are the following. First, synthetic speech, although not yet perfectly natural, has been developed to the point where it is intelligible to people who have received no prior exposure to synthetic speech or training in its use. Moreover, this is true of synthetic speech delivered at rates in excess of 150 words per minute (wpm). No other reading machine output intended for use by the blind can make such a claim. It can be argued, therefore, that the value of synthetic speech has already been established and the question of how it may be deployed to provide a useful service can now be given serious consideration. Second, there is an immediate need in the blind community (particularly among students) for an increase in the supply and speed of delivery of spoken text. A reading machine is ideally suited to the task of producing

Manuscript received November 1, 1972. This work was supported in part by the Research and Development Division, Prosthetic and Sensory Aids Service of the Veterans Administration and in part by the University of Connecticut Research Foundation.
P. W. Nye and F. S. Cooper are with Haskins Laboratories, New Haven, Conn.
J. D. Hankins is with the University of Connecticut, Storrs, Conn. 06268.
T. Rand and I. G. Mattingly are with Haskins Laboratories, New Haven, Conn. and the University of Connecticut, Storrs, Conn. 06268.

large volumes of material quickly and can already start to fill the gap in present services by supplementing the material produced by human readers. Third, although synthetic speech appears at present to be at an economic disadvantage when compared with naturally produced speech, the costs of operating reading machines can be expected to fall in the future, whereas human labor costs will certainly increase. The eventual widespread use of reading machines is therefore inevitable. This conclusion leads to our fourth point, which is that the initial entry of automated techniques into any new arena can always be expected to be met by new and often unforeseen problems. Such problems are usually amenable to solution. However, they first need to be identified and then time must be allocated to find ways of circumventing each difficulty. We believe this to be true for reading machines and that the importance of the application and the extent of the need indicate that we should not delay any longer the task of evaluating synthetic speech with blind people under field trial conditions.

Thus the University of Connecticut and Haskins Laboratories are collaborating in the development of an evaluation program leading to the construction of a pilot Reading Service Center on the University campus—initially to serve the blind students enrolled there, but with the eventual goal of extending the service to other blind people statewide. The pilot center will be a first step toward setting up similar centers elsewhere. The evaluation and development work that we are undertaking builds upon the research carried out at Haskins Laboratories under Veterans Administration support. Initial reading tests with synthetic speech texts have already been conducted on blinded veterans, as well as with blind students, with encouraging results [3].

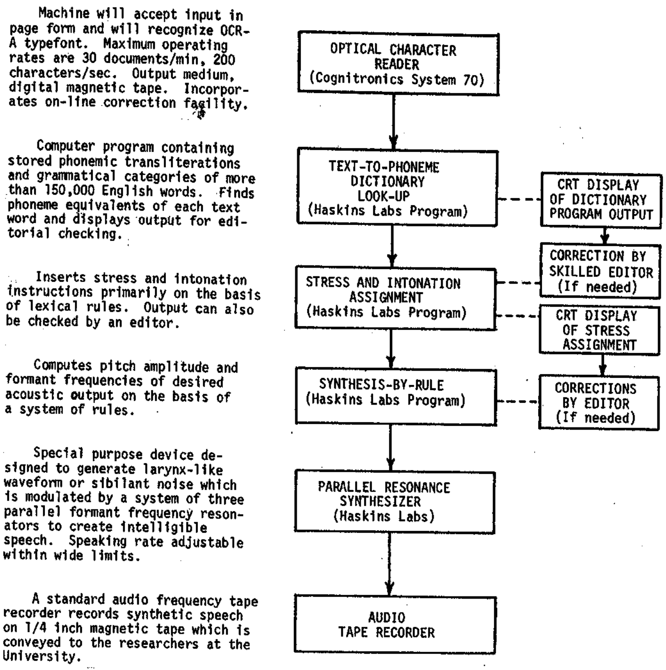

Fig. 1. Text-to-speech processor.

The Text-To-Synthetic-Speech Process

The speech synthesis system that we will use was constructed at Haskins Laboratories, both as a research tool for studies on the perception of speech and as a step toward the development of a reading machine for the blind. Fig. 1 shows the sequence of steps involved in text-to-speech conversion. The characters, which are recognized by the optical reader, are grouped into words and recoded into phonemic form by means of an automatic dictionary. The phonemic text is "punctuated" with stress and intonation assignments and then transformed by another program into instructions for the control of a terminal analog speech synthesizer. Synthetic speech output from the synthesizer is then recorded on tape for use by the blind reader. A substantial part of this system, the speech synthesis procedure that embraces the last three steps of Fig. 1, is already fully operational. Input to this completed portion of the system (by way of a phonetic keyboard) at present requires considerable hand labor. This work will be avoided when the first three stages, which are currently under development, are made operational.

Synthetic speech is currently being produced at Haskins Laboratories by a Honeywell DDP-224 computer, which controls a hardware synthesizer designed by Cooper. To make the machine speak, a phonetically trained typist must transliterate the printed text into a phonemic text and type it on a keyboard attached to the computer. Stress and intonation markers are assigned by programmed rules to "punctuate" the phonemic text, as described by Gaitenby, Sholes, and Kuhn [4]. The typed phonemic symbols and punctuation are then displayed on a storage oscilloscope that allows the operator to examine the input to the computer and to correct typographical errors if necessary. Using this phonemic input, the computer calculates values for the dynamically controlled parameters of the synthesizer on the basis of programmed rules devised by Mattingly [5]. These values are then fed to the synthesizer at a rate set by the operator. In practice, speech can be generated at rates from 60 wpm to over 300 wpm. However, a passage of speech lasting for 10 min at a normal presentation rate may well take the phonetic typist at least an hour to prepare. The way in which we propose to avoid the excessive labor and delay involves the addition of three major component steps that will automate to a large degree the tasks now performed by the phonetic typist and should greatly speed the process of transliteration. These steps will enable us to generate the relatively large volumes of reading material required to provide a reading service.

The first step employs an optical character recognition (OCR) machine. Primarily because the size of our evaluation study does not merit the use of a multifont OCR device capable of reading proportionally spaced ink print, we plan to have text material retyped in an OCR-A upper- and lower-case type face and read by one of the smaller limited-font machines. The output will be recorded on digital magnetic tape for subsequent use by the computer. There are several other reasons why we prefer to use even a limited-capacity optical reader for input rather than an on-line typewriter, punched paper tape, or punched cards. The first is that much larger volumes of reading matter than we have used before must be fed into the computer as rapidly as possible. A good typist can typically work on straightforward text more rapidly and accurately than can a keypunch operator. Moreover, if the work is performed off-line, it need not occupy the computer during the text production process. Once a large volume of typescript has been prepared, an OCR reader can convert it into an alphanumeric code expeditiously and cheaply. A second reason for our interest in OCR input lies in the opportunity it provides to obtain some introductory experience of current OCR technology so that we can better judge which machines and techniques best meet the needs of blind people and what problems still require solution. It is already apparent that the specifications of OCR devices designed for commercial applications do not fully satisfy the requirements of reading machines for the blind. For example, almost all of the commercial multifont OCR development is geared toward high-speed and high-accuracy operation on an input medium for which some or all of the following features are closely specified: size and shape of the page, color, type styles, print quality, and the position of the printed text within the page. In contrast, a reading machine for the blind must be flexible with respect to each of these input specifications. It is possible that by deliberate design this flexibility could be gained at the expense of recognition accuracy which, for a reading machine application, may be a little less stringent than that required for business purposes. However, because of the limited market potential of the blind community it is unlikely that the commercial sector will show an interest in solving its special problems in the near future. The solutions must be sought by those who have a direct interest in the eventual development of automated reading services.

The second step required to speed our production of reading materials is the addition of an automated dictionary for converting the alphabetic representation of each word into its corresponding phonemic representation. In the compilation of our now-completed dictionary we are greatly indebted to the work of Dr. June Shoup of the Speech Communications Research Laboratory. This dictionary will initially include approximately 150 000 words but may have to be expanded. Optimum dictionary size can best be determined through actual use in a practical production system. Throughout this process the system's performance will have to be carefully watched to ensure a proper balance between the size of the dictionary and search time or production rate.

The third step entails automated stress and intonation marking. Following its assembly by the dictionary search routine, the phonemic string will be punctuated with stress and intonation symbols by a program based on the rules of Gaitenby, Sholes, and Kuhn [4]. In the final system the output from the dictionary and the stress assignment routines will be displayed on a storage oscilloscope and monitored by an editor–phonetician. Corrections of errors will be made by the editor, who will also note the circumstances in which errors occur. By this procedure we expect to produce useful text and at the same time to detect defects and omissions in the system and readily correct them. Thus the frequency of error can be expected to gradually fall and the need for an editor to be substantially reduced in the future.

When the combination of phoneme string and stress markings has been formed, it will be processed by the speech synthesis program that is already in operation, converted to synthetic speech, and recorded. It is anticipated that the time taken by the machine to read and speak a page of text will be comparable with the time taken by a human speaker provided that editorial intervention is not required. However, the machine will require a few additional seconds for computation.

Evaluation Studies

The completed production system will generate synthetic speech from text in sufficient quantity to meet the needs of an evaluation study. The purpose of the study is to determine whether the operation of a Reading Service Center is economically and technically feasible and user acceptable. The question of economic feasibility is not easy to answer because of the intangible human values involved, but it is obvious that we need to identify costs and benefits as accurately as possible and to assess them in relation to available resources. To find the answers we need, we propose to operate the production system we have just described, to perform some analytical tests, and to transcribe into synthetic speech some actual reading assignments required by blind students at the University. These assignments will be supplied in a manner similar to that in which existing services operate at the University to provide recordings of natural speech. By examining the way in which these materials are used, we expect to find the answers to two broad classes of questions. The first relates to human factors, the second to technical and economic factors. Fig. 2 illustrates the areas to be explored.

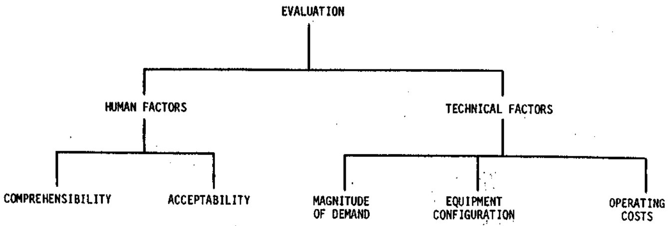

Fig. 2. Evaluation studies.

In the area of human factors, we are concerned with the relative comprehensibility of synthetic speech and natural speech over a range of speaking rates. The key question here is whether any differences in comprehensibility that may emerge are significant enough to affect the educational utility of synthetic speech. Analytical testing procedures will be used. The basic strategy for assessing comprehensibility involves the presentation of a lively passage of general interest followed by a series of questions which seek measures of the number of facts retained (i.e., names, places, distances, colors, etc.) and also the ability of the reader to derive logical inferences from the information. We propose to apply such tests using synthetic speech and natural speech controls (with appropriate counterbalancing) and then to compare the performances of the students. In a series of interviews designed to assess acceptability, we plan to gather data on such subjective factors as the relative preference for synthetic speech versus natural speech, the comparative comfort in use of the media, judgments regarding the aptness of different media for various fields of study, and the influence of delivery rate on all of these factors.

In the area of technical factors, we are concerned with establishing an accurate assessment of the overall demand that a Service Center will be required to meet, the technical quality of the synthetic speech medium required to produce acceptable performance at reasonable cost, the turn-around time that is both acceptable and economic, and the range of speaking rate required of the output. From these data an optimum equipment configuration can be determined and labor and operating costs can be estimated.

Conclusion

In this paper, we have argued that in order to provide better educational, vocational, and recreational opportunities for the blind population of this country, faster and more flexible reading services are required. Moreover, the technical resources are now available to supplement existing services through the use of the reading machines located in Reading Service Centers. We believe that the time is now ripe to make a determined effort to move this technical capability out of the laboratory and into the community it could serve. The preliminary evaluative work we have described here is extensive and time consuming. Nevertheless, it is essential that an exploration of the extent to which the results of our research meet the needs of blind people be carried on in parallel with continued research. This strategy must be followed if we are to apply effectively a laboratory-developed technology to a socioeconomic problem as complex as blindness.

References

[1] P. W. Nye and J. C. Bliss, "Sensory aids for the blind: A challenging problem with lessons for the future," *Proc. IEEE*, vol. 58, pp. 1878–1898, Dec. 1970.
[2] *An Assessment of Industrial Activity in the Field of Biomedical Engineering.* Washington, D.C.: Nat. Acad. Eng., 1971.
[3] F. S. Cooper, J. Gaitenby, and I. G. Mattingly, "Research on audible outputs of reading machines for the blind," *Bull. Prosth. Res.*, vol. 10-15, pp. 241–246, 1971.
[4] J. H. Gaitenby, G. N. Sholes, and G. M. Kuhn, "Word and phrase stress by rule for a reading machine," in *IEEE Conf. Rec., Joint IEEE and AFCRL Conf. Speech Communication and Processing*, 1972, pp. 27–29.
[5] I. G. Mattingly, "Experimental methods for speech synthesis by rule," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, pp. 198–202, June 1968.

Reprinted by permission from **IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS**
Vol. Au-21, No. 3, June 1973, pp. 265-268