A DESCRIPTION OF THE KURZWEIL READING MACHINE AND A STATUS REPORT ON ITS TESTING AND DISSEMINATION

Aaron Kleiner
Executive Vice-President

Raymond C. Kurzweil
President

Kurzweil Computer Products, Inc.
68 Rogers Street
Cambridge, Massachusetts 02142

ABSTRACT

This paper presents a technical overview of the different subsystems of the Kurzweil Reading Machine—the nature of the information processed by each subsystem and the type of transformations performed. A description of how the user directs the machine is provided. The Veterans Administration's testing program, as well as those of the Bureau of Education for the Handicapped, the Rehabilitation Services Administration, and the National Federation of the Blind, are outlined.

Preliminary tests of the machine with blind secondary school students have focused on the intelligibility of the synthetic speech output, and the results of these tests are presented.

INTRODUCTION

The Kurzweil Reading Machine converts print to speech, and is designed as a reading prosthetic aid for the blind and visually handicapped. The system handles ordinary printed material—books, letters, reports, memoranda, etc., in most common styles and sizes of type. The output produced is a synthetic voice using full-word English speech.

The user operates the device by placing printed material face down on the glass plate which forms the top surface of the scanning unit; he then presses the "Page" button on the control panel, and listens to the synthetic speech produced as an electronic camera scans the page and transmits its image to a minicomputer housed within the device (Figs. 1, 2, and 3). The computer separates the image into discrete character forms, recognizes the letters, groups the letters into words, computes the pronunciation of each word, and then produces the speech sounds associated with each phoneme.

The machine operates at normal speech rates, about 150 words per minute.

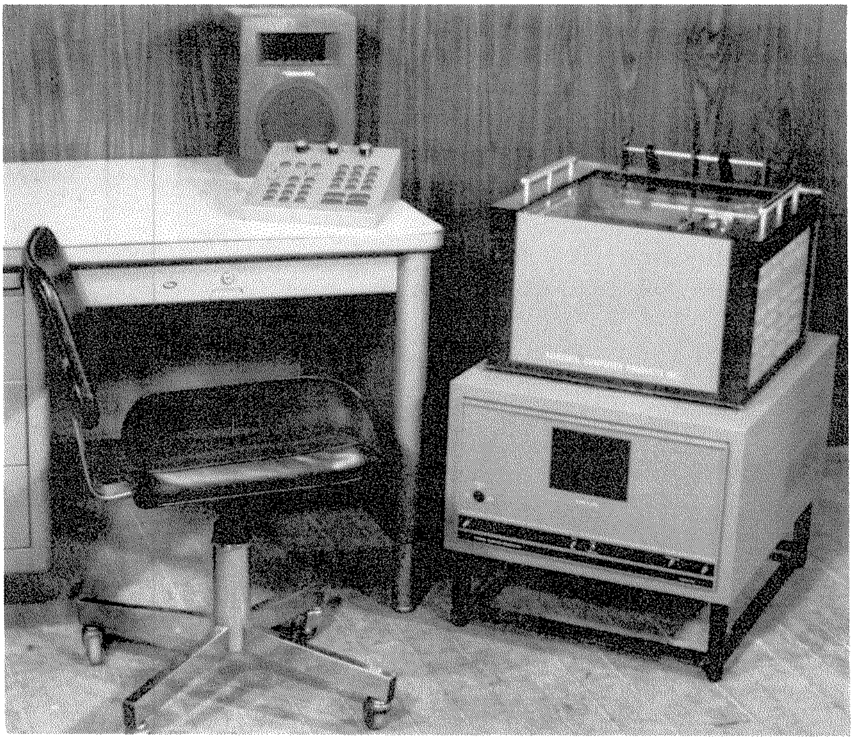

FIGURE 1.—Kurzweil Reading Machine for the Blind. On the desk are the loudspeaker and the user's control panel, and to the right the glass-topped scanner which has been placed on top of the electronic unit.

THE SUBSYSTEMS

The Scanner

The first subsystem of interest is the scanner. A high-contrast electronic image of the page to be read is generated by the scanner, which includes a linear photosensor array at the focal plane of a

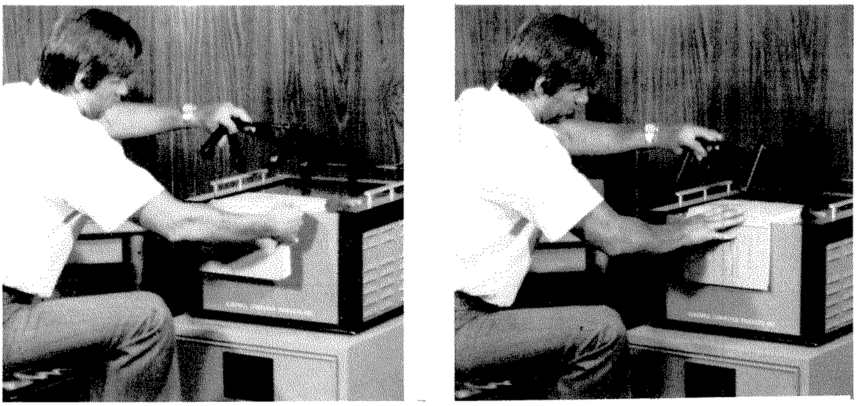

FIGURE 2.—Adjusting a book for reading on the Kurzweil Reading Machine.

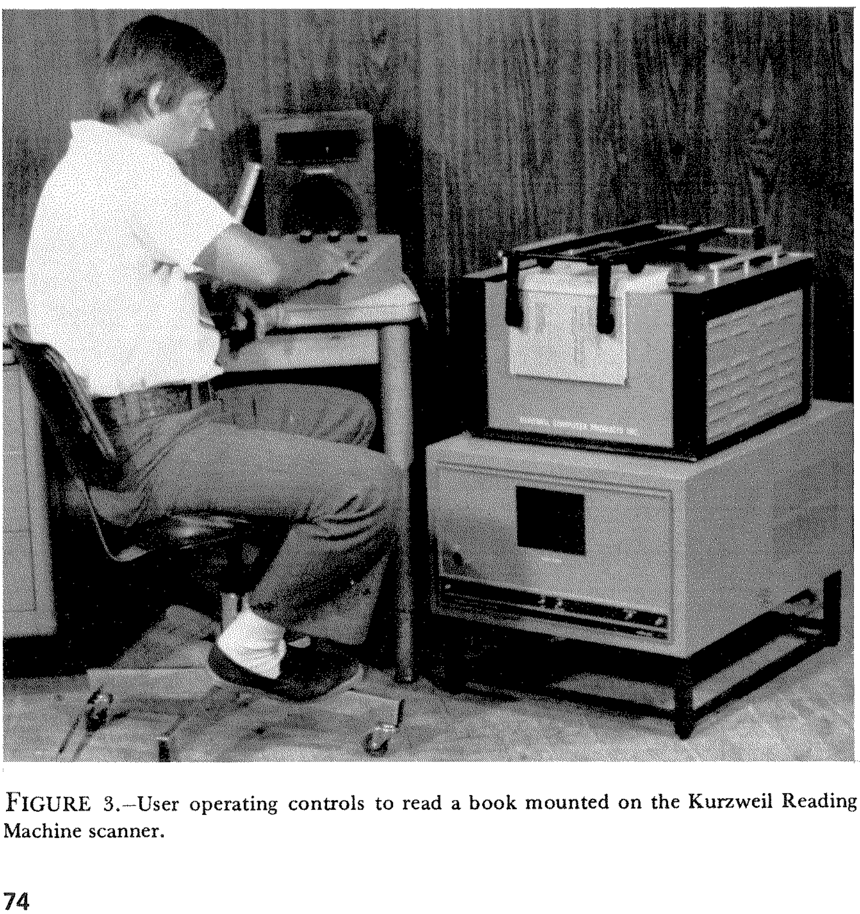

FIGURE 3.—User operating controls to read a book mounted on the Kurzweil Reading Machine scanner.

camera mounted on an X-Y mover. The text page is placed face-down on a glass plate above the X-Y mover, which moves the camera in a plane parallel to the plane of the text. The photosensor array is perpendicular to the direction of text lines; thus, a single scan across a page results in an electronic image covering a strip of the page which includes several text lines.

Immediately after the user presses the "Page" button, the scanner moves to the limit of its travel in the direction of the top of the page, and begins to search for the first text line. When the first line is found, the camera tracks that line and moves automatically down to the following lines, tracking each in turn. Scanner efficiency is maximized by scanning left-to-right and also right-to-left; material scanned right-to-left is reversed by the computer before the output stage.

Character Recognition

Scanner output, which can be imagined as a simple two-dimensional matrix of black and white points, is processed by the character recognition system, beginning with the isolation of each contiguous area of print in the page image. Each contiguous print area is presumed to be a single character, with the exception of certain special cases requiring additional processing; e.g., touching or fragmented characters, and dots associated with particular characters. Each individual character is analyzed by a set of feature extraction routines. The features, or properties, extracted are those that have been found to be relatively invariant for the same character with respect to the kinds of changes that occur across different typestyles. These properties are basically geometric—line segments, concavities, loops, loop extensions, and the positional relationships among these elements. For example, the properties of a standard capital "A" include a single loop and a single south-facing concavity. Once the properties have been extracted, they are compared to stored lists describing each character in the identification set. (Examples of characters outside the identification set, that is, characters not presently identified, include various special mathematical symbols, and non-Latin alphabets.)

The output of this process is usually a final identification, although it is occasionally an ambiguity code indicating one of several possible final identifications based on contextual cues. In this case, the final identification is made by a special module (called the "disambiguator") which analyzes positional and other contextual cues to "disambiguate" the output of the shape recognizer. An example is a single vertical bar. The shape recognizer output will indicate that this shape is the lower-case letter "i" if there exists a dot above the character, or the lower-case letter "1" if there exists no dot and the character is in a letter context, or the number "one" if there exists no dot and the character is in a number context. (A character is in "number context" generally if it is adjacent to a number.)

The final step in character recognition is text line formation—that is, the collation of individual characters into words and punctuation to be analyzed by the speech generation system.

Speech Generation

The primary function of the speech generation system is to determine the proper pronunciation of each word; i.e., grapheme strings are converted to phoneme strings. This is accomplished by a set of over 1,000 phonetic rules supplemented by a dictionary of exceptional cases. First, prefixes and suffixes are stripped off by algorithms which reconstruct the root—for example, "Rating" is converted to "rate"-"ing." The exception dictionary is checked for the resulting root, and if the root has been entered into the dictionary, the dictionary phonemes are used. Otherwise, the root is passed to the phonetic rules module. Whether the rules module or the exception dictionary is used, the prefix and suffix, if any, are looked up in a special affix dictionary and properly appended to the root phoneme string.

Stress marks, which are modifiers tied to each phoneme in the form of relative pitch and volume parameters, are also generated by the exception dictionary and the phonetic rules for intra-word stress-assignment. A stress contour over each sentence, for improved prosody, is computed by a set of primitive syntactical rules that look primarily for certain lexical cues indicating cutoff phrase boundaries. These rules range from the very simple:

> If a word ends in a question mark or exclamation point, stress it. (This generally gives the appropriate stress pattern to questions and exclamatory sentences.)

to the more complex:

> Relative pronouns are stressed if the *preceding* word was not a relative pronoun ("that which is . . .").

Rules of the second type utilize an internal dictionary of syntactic types, entirely distinct from the phonetic exception dictionary.

The final step in speech generation is a conversion of the phoneme string into the speech waveform. This operation is performed by a hardware synthesizer, which is essentially a set of variable electronic filters designed to model the human vocal tract. Synthesizer output is amplified and transmitted to a speaker—and the machine "talks."

User Controls

For everyone, sighted or blind, reading is properly an active rather than passive process; one re-reads interesting or difficult passages, pauses at unfamiliar words, skims the page to find material of particular interest, and so on. The Kurzweil Reading Machine allows its user to interact actively with the text he is reading via the user controls, which are a set of keyboard-mounted buttons. When the "Page" button is pressed, the scanner automatically finds the first line of text on the page and begins reading, continuing down the page until the "Pause" button is pressed, at which point the scanner (and the speech output) is halted. When "Resume Scan" is pressed, output is resumed at the current scanning position.

After "Pause" is pressed, and until "Page" or "Resume Scan" is pressed, the machine is in "pause mode." While in pause mode, the user may utilize buttons to manipulate the reading machine in several ways. He may read a single line or a single word, repeating as often as he wishes, and he may move forward through the text in line or word increments; he may have words spelled out rather than normally pronounced.

If the user presses "Set Mark," the computer will note the current scanner position; if at a later time he presses "Go To Mark" the scanner will return to the "Set Mark" position. Using a keyed numeric input in conjunction with the mark facility allows the user to mark, and return to, several different positions.

This is only an overview of the control options available to the user; our keyboard presently includes 30 buttons. The physical configuration, however, is designed to allow the novice user easy access to the five or ten buttons most important to his reading. Still, the favorite button of many inexperienced users will probably be the "Nominator" key; if that key is pressed immediately preceding the pressing of any other key, the machine will not perform the command associated with the second key, but rather will tell the user its function (by speaking the name of the key).

ANTI-OBSOLESCENCE STRATEGY

An important design goal was to produce a high-performance (rapid, easy to use) reading machine for the blind which would not become obsolete. With continuing rapid advances in digital component fabrication techniques, the problem of a design becoming obsolete by the time it is completed is a serious one. Therefore, we have implemented any information-handling process likely to change, in software. The software can and will be modified over time as the system continues to be shaken down, and as further improvements are implemented. Software changes, as they are developed, can be distributed to units in the field in the form of software update tapes, which can be loaded into each reading machine using the digital cassette tape drive provided with each unit. In this way, even the first set of reading machines will be able to take advantage of improvements such as the development of other languages. Special character sets (mathematical symbols, for example) can also be introduced to units in the field using special cassette tapes, again without requiring hardware modification.

A final advantage is that for future production runs we can take advantage of more efficient components—new microCPU's, new memory chips—as they are introduced, without having to redesign the heart of our technology, which is primarily in software.

HISTORY AND CURRENT STATUS

The project began approximately 10 years ago. After approximately 5 years of feasibility study, algorithm simulation, component review, and other preliminary steps, Kurzwell Computer Products was organized to implement a fully working model. About 1½ years ago (Fall 1975), the first system capable of scanning a printed page in multiple type styles and converting it to comprehensible full-word speech was completed. The current system is now the fifth version of the Print-to-Speech system.

The machine was demonstrated last summer (1976) at the national convention of the National Federation of the Blind in Los Angeles, at the national convention of the Blinded Veterans Association in Philadelphia, and more recently in Washington at the Bureau of Education for the Handicapped and before the Senate Subcommittee for the Handicapped.

While the initial development of the reading machine has been completed, further improvements are possible. Reading is a complex activity. Print specifications differ significantly among printed documents, varying in the style and size of type, the grade of paper and ink, the quality of printing, and the page format. The purpose of reading may also vary, ranging from skimming to intensive study. Extensive user experiences in the field will enable us to optimize the machine's ability to handle the largest possible diversity of reading situations.

To guide this continuing refinement, feedback is currently being collected under several programs from those who will be affected by the machine; blind children and adults, teachers, and rehabilitation professionals.

Veterans Administration Program

The Veterans Administration Program will test the reading machine with blinded veterans, and will seek to improve the device based on the reading needs of this population. An initial placement will be made at the Central Rehabilitation Section for Visually Impaired and Blinded Veterans, at VAH, Hines, Illinois.

The National Blinded Veterans Association has endorsed the VA's purchase of the machine and its clinical evaluation. Jerry Monroe, the BVA national president, has been particularly helpful in providing guidance, enthusiasm, and insightful user input for the project.

The short-range goal of the work at Hines is to evaluate the usefulness of the machine in its present state for purchase by the VA for the various blind centers as a teaching and experimental tool. The long-range goal is to establish the ultimate usefulness of the machine at its eventual market price, for blind and visually-impaired persons. The first step in this project will be an examination of the performance and functioning of the machine by Harvey Lauer and others. Then the following areas will be explored:

1. Interaction with the world of print—the percentage of character-recognition accuracy as a function of type style, page format, and printing quality.
2. The control interface—the machine's provision for interaction with the user will be examined, and the aptitude and training requirements will be described.
3. The output interface—the intelligibility rate of the synthesized speech output as a function of character-recognition accuracy rates, pronunciation error rates, pace of output, intelligence of user, length of user experience with the machine, etc., will be described. The ultimate effect of these variables on reading and comprehension rates will be described.
4. Finally, the Kurzweil machine will be compared with other modes of reading—braille, recordings, sighted readers, other reading devices. The machine's applicability to the needs of blinded veterans will be tested, and the VA's 30 years of pioneering experience in the development and evaluation of reading machines for the blind will be utilized in this effort to apply new technology to the needs of blind people.

Other Federal Programs

The machine is also being evaluated under a program with the

Department of Health, Education, and Welfare (HEW) which includes tests sponsored by the Bureau of Education for the Handicapped (BEH) and the Rehabilitation Services Administration (RSA), both in that department.

The BEH program is focusing on the use of the reading machine in educational situations, particularly with school-age blind children. Data are being collected concerning the use of the machine, and the machine will be improved on the basis of this feedback from blind children, their parents, and their teachers. The first placement under the BEH program was made in the fall of 1976, at the Perkins School for the Blind in Watertown, Massachusetts.

The RSA program will focus on the use of the reading machine by blind adults in vocational situations and will provide feedback from blind consumers and rehabilitation specialists. The first RSA placement will be made shortly at the West Virginia Rehabilitation Center. Additional placements will then be made at Arkansas Enterprises for the Blind, and the Iowa Commission for the Blind.

NFB Participation in Development

In addition to the Federal evaluation programs, a comprehensive human engineering consumer-based study of the reading machine has been organized in collaboration with the National Federation of the Blind.

The National Federation of the Blind is the largest organization of blind people and has for the last 3 years advised the company on the development of the machine. As part of the current study, the Federation is placing six reading machines in a variety of user situations—in the office of a blind executive, a lawyer, an engineer or scientist, in a rehabilitation center, and so on. On the basis of this experience and suggestions that result from use of the machine, the Federation will provide guidance on ways in which the machine can be made more responsive to the needs of blind readers through such things as additional user controls and improved physical configuration.

The program has been designed, and is being implemented and directed, by the NFB's own scientists and engineers.

The funds for this program, incorporating a new concept of consumer involvement in the design process, have been contributed by five major foundations across the country. Initial testing of the reading machine by Michael B. Hingson of the NFB's staff began several months ago in Boston.

We believe it is important for designers and consumers to involve each other in this kind of intensive way, and thus increase the relevance and quality of new technology for the disabled.

Preliminary Tests on Secondary School Children

In the preliminary tests of the machine conducted under the BEH and RSA programs, seven secondary school students whose only disability was visual impairment and who were of normal intelligence, were subjects for this initial study. The study was structured as a comparison of the subjects' performances in decoding (understanding single words in isolation) and comprehending (understanding paragraph-length material) the reading machine's synthetic speech.

The Durrell Listening-Reading Series, Advanced Level (for grades 7-9) was selected as the test instrument. Four sets of word-lists and paragraph-length passages, each of equal level of difficulty, interest value, and representativeness of vocabulary sample were presented to each student as follows: (i) test using human speech; (ii) test using synthesized speech—subjects had no previous exposure to synthetic speech; (iii) test using synthesized speech after moderate practice; and (iv) test using synthesized speech after more extensive practice.

Due to limitations, no student received more than 3 hours total exposure to the synthetic speech before testing ended. Briefly, the results of the tests were as follows (using the mean scores for the subjects):

The highest scores were obtained using natural human speech—9.83 for word repetition and 12.00 on paragraph comprehension;

The lowest scores were obtained upon first exposure to the synthetic speech—4.83 for word repetition and 9.50 on paragraph comprehension;

The scores for final tests using synthesized speech were most encouraging—9.37 for word repetition and 11.17 on paragraph meaning.

Given the limited sample size, there is no statistically significant difference between the human speech scores and the final synthetic speech scores. Later studies will utilize larger and more diverse population samples, as well as more extensive test material, but the results obtained from this preliminary study agree with the experience of nearly everyone who has practiced with the reading machine—it is not difficult to comprehend the synthetic speech after only a few hours of exposure.