Type Reading Machines for the Blind

Glendon C. Smith

The most advanced personal type reading machine described provides the blind user with an aural "spelled speech" equivalent for each upper- and lower-case letter or ligature scanned by a hand-held optical probe. This character recognition machine recognizes the most popular type fonts with moderate accuracy and speed (80–90 words-per-minute). The development of the hand-held probe for this machine has resulted in a family of aural and tactile "direct translation" reading aids which are pocket-sized and battery-operated and may be used independently for low-speed reading.

Interest in providing a useful reading machine for the blind has been gradually increasing since the end of World War I. Most of the emphasis has been placed on simple direct translation machines (Optophone types) which commonly use a vertical row of narrow photocells to scan letters sequentially and generate a different audio frequency for each photocell while it is "seeing" black. While such machines are simple and thus relatively inexpensive, they have the shortcomings of requiring extensive training and yielding low maximum reading speeds (10 to 25 words-per-minute).

The work reported here is part of a continuous effort carried out by Mauch Laboratories, Inc., since 1957, toward the development of a personal reading machine for the blind. The most advanced system now being evaluated and improved will permit reading speeds up to 80-90 words-per-minute. It will be portable

The research and development reported in this paper is sponsored by the Prosthetic and Sensory Aids Service, Department of Medicine and Surgery of the Veterans Administration.

and will be reasonably priced when manufactured in moderate quantities.

In early studies we found that a speech-like aural output is necessary in order to achieve satisfying reading rates. Consequently, an electromechanical Word Synthesizer was designed and constructed which would play back stored tape recordings of individual phonemes in accordance with photoelectric scanning signals derived from the printed text. Full recognition of the entire alphabet was not intended at that time.

Somewhat later, a relatively economical scanning method was developed whereby a letter is seen simultaneously—as in a "snapshot"—by a number of photocells in a two-dimensional array (rather than by a single line of photocells which must progressively scan over a letter to establish its identity). This single snapshot recognition technique was incorporated in a prototype scanner which recognized all the lower-case letters of the IBM Executive typewriter alphabet. In order to recognize upper- and lower-case letters of a wide variety of type styles, a more versatile recognition scheme which utilizes multiple snapshots to collect information on letter features was devised. A Recognition Prototype I which uses the multiple snapshot technique was designed and tested with 498 letters from nine widely-used type styles (Figs. 1–3).

When it was realized that full recognition would be possible, the Word Synthesizer was equipped with tape recordings of "spelled speech" letters, developed by Prof. Milton Metfessel, University of Southern California, instead of phonemes. In the spelled-speech alphabet each letter has a standardized audible pronunciation somewhat like its usual pronounced name, but chosen to give maximum coalescence and smoothness when reproduced rapidly in combination with the sounds for other letters. Thus, the word "type" in spelled speech would sound something like [tiwaɪpii] where the letters in the brackets are International Phonetic Association symbols. The spelled speech output can be understood at 80-90 words-per-minute with little training and is much more suitable for an inexpensive reading machine for the blind than attempting to synthesize English words.

The character-recognition reading machine being developed by

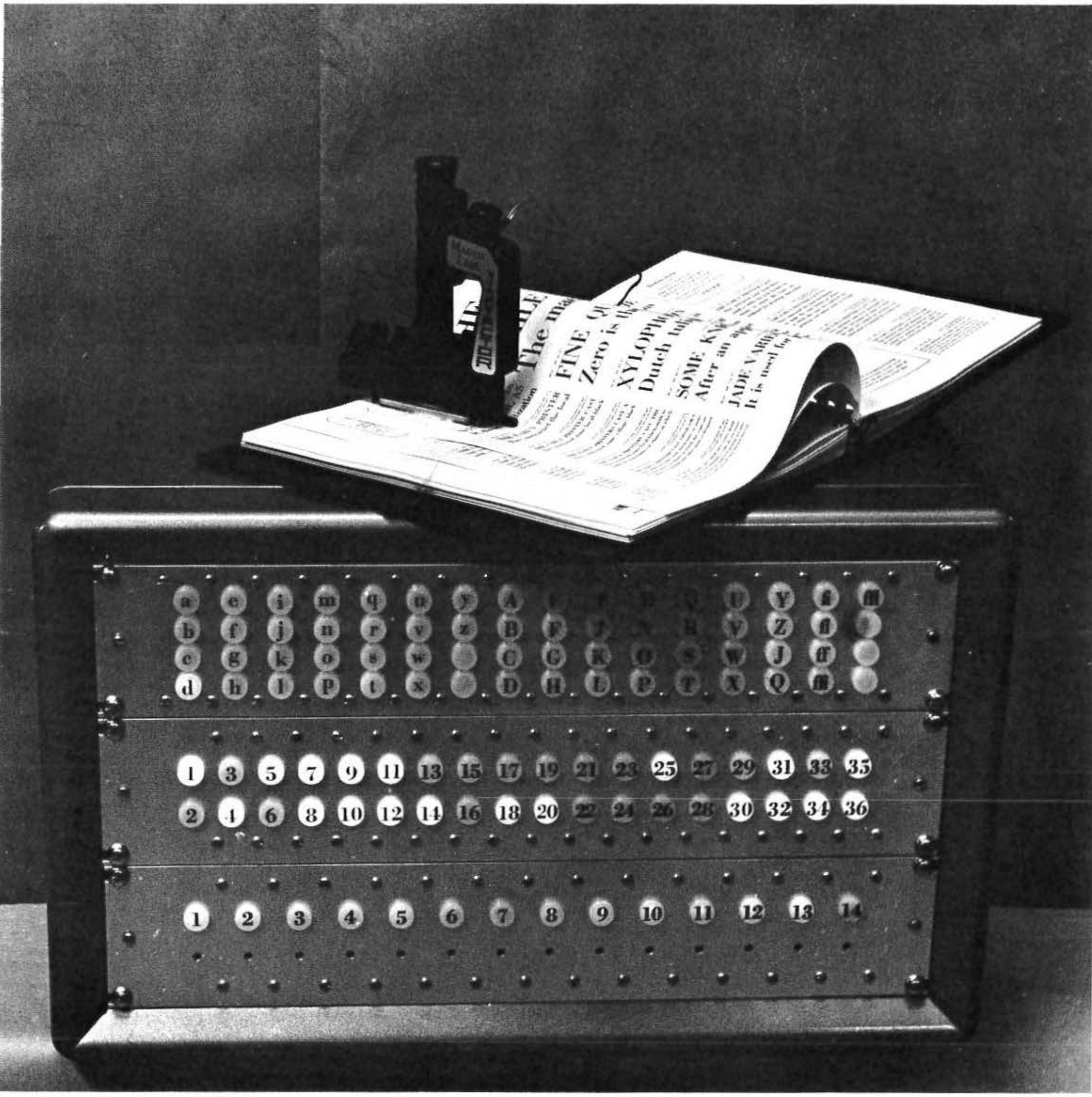


Figure 1. The Recognition Prototype and Visotactor show the indicator lamps after a lower-case d has been recognized but before the memory is cleared (see also Figure 2). The lower-most panel contains 14 Schmitt trigger circuits. Twelve of them connect one-for-one to photocells in the recognition array; the others are spares. The middle panel contains 36 memory units of which nine units are spares. Only memory units numbered 4, 14, 16, 23, and 28 are important in identifying a d. The top panel consists of output indicators.

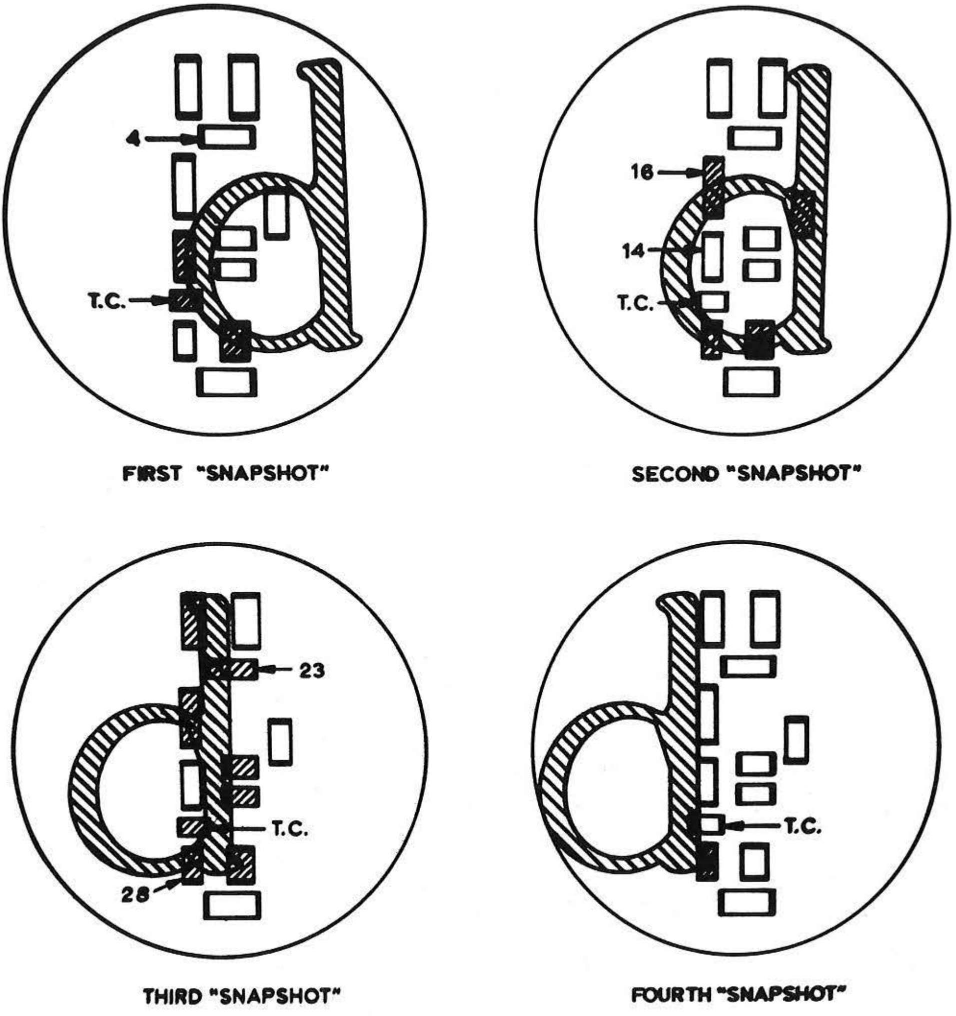


Figure 2. The recognition array is shown scanning a lower-case d. (The photocells used for the tactile output are not shown.) Each "snapshot" is initiated by the trigger cell (T.C.) changing from "seeing" light to "seeing dark." Photocell information is quantised into light and dark and stored in different groups of memory units until the letter is past and the memory read out and reset. The numbers indicate which memory units are used to identify the d. The direction of current flow in each photoconductor is disclosed by the two bold sides which represent electrodes.

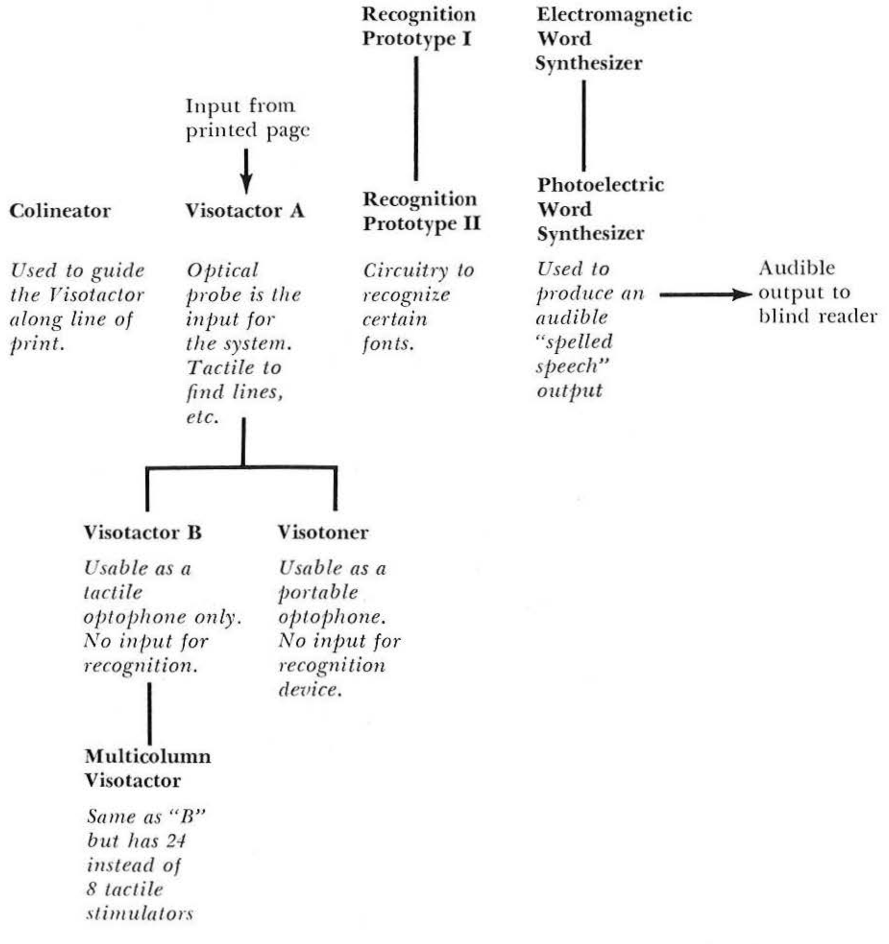
aaaaaaaaa AAAAAAAAA ﬁﬁﬁﬁﬁﬁ
bbbbbbbbb BBBBBBBBB ﬂﬂﬂﬂﬂﬂ
ccccccccc CCCCCCCCC ﬀﬀﬀﬀﬀﬀ
ddddddddd DDDDDDDDD ﬃﬃﬃﬃﬃﬃ
eeeeeeeee EEEEEEEEE ﬄﬄﬄﬄﬄﬄ
fffffffff FFFFFFFFF
ggggggggg GGGGGGGGG
hhhhhhhhh HHHHHHHHH
iiiiiiiii IIIIIIIII
jjjjjjjjj JJJJJJJJJ
kkkkkkkkk KKKKKKKKK
lllllllll LLLLLLLLL
mmmmmmmmm MMMMMMMMM
nnnnnnnnn NNNNNNNNN
ooooooooo OOOOOOOOO
ppppppppp PPPPPPPPP
qqqqqqqqq QQQQQQQQQ
rrrrrrrrr RRRRRRRRR
sssssssss SSSSSSSSS
ttttttttt TTTTTTTTT
uuuuuuuuu UUUUUUUUU
vvvvvvvvv VVVVVVVVV
wwwwwwwww WWWWWWWWW
xxxxxxxxx XXXXXXXXX
yyyyyyyyy YYYYYYYYY
zzzzzzzzz ZZZZZZZZZ

Figure 3. Almost 500 characters were photographed and arranged on transparencies to aid in the design of the photocell array and recognition matrix. The nine type styles illustrated here (Caledonia, Baskerville, Janson, Corona, Imperial, and Regal; plus elite, pica, and IBM Executive typewriter faces) are estimated to make up about 75% of the printed and typewritten material used in the United States.

Mauch Laboratories consists of several components (Fig. 4). The Visotactor A is the hand-held and guided optical scanner which at present contains eight tactile stimulators (two per finger) to operate as a type of tactile optophone enabling the user to adjust probe magnification for size of type (7- to 36-point), to locate the line of print, to scan along the line of print, and to decipher numerals and characters outside the spelled speech "vocabulary." In addition to the eight photoconductive cells in the optophone type array, the Visotactor A contains a two-line array of photoconductive cells which provide sequential signals necessary for our multiple snapshot character recognition technique. Logic circuitry located in the Recognition Prototype II operates from the photoelectric signals to determine the presence or absence of letter features independent, in large measure, of the letter width and type font. A bench model of this recognition machine recognizes the most popular type fonts with excellent accuracy (90-95 per cent). As soon as the letter (upper- and lower-case or ligature) has been determined, the logic circuitry translates its information into the five-bit Baudot code which causes the photoelectric Word Synthesizer to reproduce one of 31 prerecorded outputs through a loudspeaker or earphones. A mechanical tracking aid, the Colineator, is an optional accessory which may be used for easier tracking over extended periods of reading.

The Visotoner is a compact, battery-operated optophone. The Visotoner frame contains nine transistorized audio oscillators which produce the optophone code. The Visotoner and its battery are also pocket-sized. Seven Visotoners (including Colineators) have been built and two are being used by blind individuals previously trained to use the Battelle optophone.

The Multicolumn Visotactor contains three adjacent columns of photocells and associated tactile stimulators which are sensed by the "Braille area" of the finger tips. Potential reading speed places it between the single-column Visotoners or Visotactors and the character-recognition machine. This device may evolve into a Digitactor, a simple direct-translation reading aid which will stimulate the underside of only one finger, probably the finger doing the scanning.



The prinicpal VA-Mauch reading system for the blind as of mid-1967 is as shown in the four horizontal units above. The units above these are earlier variants, those below are separate device developments, offshoots of the principal system.

Figure 4. Evolution chart showing how the several units in the VA-Mauch Laboratories reading machine systems developed.

Development of the Recognition Prototype

The first miniature recognition array encapsulated during June 1965 was mounted in a Visotactor probe and used to scan letters attached to a rotating drum. The drum could be driven at variable speeds from 10 to 90 words-per-minute. Photographic reproductions of the nine type fonts (Caledonia, Baskerville, Janson, Corona, Imperial, and Regal; plus elite, pica, and IBM Executive typewriter) were attached to the drum. The Visotactor probe (containing the first miniature recognition array) was adjusted to the proper magnification and used to scan the upper- and lower-case letters of several type styles on the drum. For two of the type styles, Imperial and Regal, all the lower-case letters were recognized at 20-30 words-per-minute, and approximately 80-90 per cent of the letters in the remaining type styles were correctly recognized at this speed. As the speed was increased, the accuracy of recognition decreased, due mainly to the poor response of the photoconductor material to rapidly changing light levels.

Tests were undertaken to determine how far above or below the line of print the input probe could be placed and still have the Recognition Prototype I not miss more than 20-25 per cent of the letters in a typical test. A 12-point alphabet of lower-case Imperial was attached to the drum apparatus. The Visotactor scanned the alphabet at slow speeds. Beginning at a point below the letter base line, successive scans were made spaced .0016 inch apart. Four scans near the optimum height resulted in recognition accuracy of 90 per cent or better. Two additional scans, one on each side of this group of four, resulted in recognition accuracy of about 80 per cent. The six scans which produced a useful recognition accuracy enclose five spaces each .0016 inch, for a total of .008 inch. Since the x-height of this type style and size is .080 inch, the useful tracking band is 10 per cent of this figure.

These accuracy figures are based on the frequency of letters in common English rather than giving each letter equal weight. For example, at the upper tracking limit (above the base line) the Recognition Prototype I fails to recognize only three letters (a, e, b) out of the 26. Considering only the alphabet, this could be expressed as an 11.5 per cent error; but considering that a and e are very frequently used, the error rate in reading text would be about 22 per cent.

The ideas and plans for an improved Recognition Prototype II were incorporated in a new design during September 1966. Photographic transparencies of the nine type fonts were used to design a photocell array which would tolerate mistracking the line of print up to one stimulator photocell height. This new design also contains additional logic to permit recording in certain memory units whether or not a particular photocell "sees black" (or white) at any time between two snapshots. This new feature is used in several situations, e.g., by letting the photocell which straddles the baseline record "seeing white" between snapshots 2 and 3, one can distinguish more reliably between letters h and b; by letting the photocell which is designed to sense descenders record "seeing black" in a memory unit anytime during the letter scan independent of snapshots, the identification of g and Q becomes more reliable. Other memory units store whether or not certain photocells "see black" at least twice during the letter scan. For three memory units the appearance of selected combinations of photocells "seeing black" or "seeing white" between certain snapshots will flip the memory from one condition to the other. The logic to accomplish these operations was built with discrete components and installed in the chassis containing the 36 memory units.

In July 1966 an order was placed with the Cognitronics Corporation for a Model 632 Speechmaker, a 31-channel photographic film audio memory drum used to provide the spelled-speech output. Experiments indicated that the spelled-speech characters can be standardized to a duration of 140 milliseconds with little degradation of quality. However, the standard message length available with the Model 632 is 600 milliseconds, with an additional 25 milliseconds of suppression time used for switching. Engineers established that slight modifications in circuitry would permit four 140 millisecond messages to be recorded on the same track, allowing 16 millisecond suppression periods between the messages.

Because the memory drum has 31 channels, five channels are available for extra messages in addition to the channels needed for the spelled-speech alphabet. Sinusoidal signals of 440, 880, and 1760 cycles-per-second, as well as spelled speech characters 1 and u played in reverse, were recorded on magnetic tape with the spelled-speech alphabet and the tape was recorded on the 31 tracks of the photographic film drum.

The sinusoidal signals can be used to check the system response or possibly to indicate tracking errors. The reversed l character was included for possible use as an error signal. If, due to misalignment or poor printing, the recognition circuitry fails to make a positive identification, the Word Synthesizer will not respond at all. The resulting silent period may be more disturbing than the soft, rather indistinct, sound of the reversed l. Conversely, the spelled speech u has a very distinctive sound when played backwards and could be used as an additional signal.

Attempts to produce fi and fl ligature sounds short enough to fit the 140 millisecond-letter length requirement were not successful. The spelled-speech f would have been used for ff, the fi sound for both fi and ffi, and the fl for fl and ffl. Since the Word Storage Unit which precedes the Word Synthesizer can accept letters very rapidly, the recognition circuitry will be designed to generate the proper binary codes in sequence for each letter in the ligatures. This method has the advantage that words containing these combinations will sound the same whether or not ligatures are being read.

At present, the film drum contains the spelled-speech alphabet C developed by Prof. Metfessel six or seven years ago and simply cut to 140 milliseconds maximum duration. Although the resulting spelled-speech sounds quite good, a better alphabet might result if one were to develop the alphabet with the constant length requirement as a prime consideration from the beginning. A complete evaluation of the various alphabets perhaps should wait until a number of reading machines and blind users are available to assist in making the value judgments. At that time, several of the more promising alphabets could be recorded on the photographic film drums used in the Word Synthesizer.

Preliminary tests of the Recognition Prototype II were encouraging. A priority circuit which may be useful for a number of letters was installed and tested on the lower-case g. Because this letter has a somewhat irregular shape which varies among the type styles, the memory units which define it most reliably are those two which indicate the absence of an ascender and the presence of a descender anytime during the letter scan. These two factors are common to the g, p, q, and y. The p, q, and y are reasonably uniform in shape and size from style to style and they can be defined more strictly than the g by using a number of memory units. The priority circuit allows the g to appear at the recognition matrix output only if the descender was present, an ascender was not present, and none of the separate sets of conditions for the p,q, and y are met. Additional priority circuits may be useful in other situations.

The Recognition Prototype II was tested for tracking tolerance and recognition accuracy at high speeds. The Recognition Prototype I recognized the lower-case letters of the IBM Prestige Elite with 90 per cent accuracy at low speed; this figure dropped to 79 per cent when the speed was increased to 7.5 characters-per-second. Recognition Prototype II was tested following the same procedure but using IBM Delegate (pica). The measured recognition accuracy was better than 99 per cent from very slow speeds to beyond 15 characters-per-second. During the test, tracking was set only once for best performance over a 10-inch line. Part of this considerable improvement is attributed to the "faster" photocell array and part to the fact that a large percentage of the diodes in the recognition matrix connect to memory units which record data between snapshots and are thus less dependent on photocell response time.

For the lower-case alphabet of IBM Delegate the tracking band of the Recognition Prototype II is .013 inch for a recognition accuracy of 90 per cent or better and .016 inch for 80 per cent or better. Since the x-height of these lower-case letters is .075 inch, these figures correspond to 17.5 per cent and 21 per cent of the x-height, respectively. By comparison, the Recognition Prototype I had a tracking band for 80 per cent accuracy which was only 10 per cent of the x-height.

Evaluation of the Visotactor B

Miss Bonnie Reinicke, a young blind woman, used a Visotactor B prototype and Colineator to read material contained in the 200-lesson Battelle Training Program (Fig. 5). Her reading speed was measured using test material provided at the end of each of the ten lessons (Fig. 6). About 130 hours were required to complete the program.

During January 1966 Miss Reinicke began reading a book of mystery stories, *Alfred Hitchcock's Ghostly Gallery,* with the Visotactor B and Colineator. After several hours practice, she became accustomed to the new type style and the techniques involved in reading a book. In the first test Miss Reinicke read for twenty minutes at an average speed (including line-change time) of 7.0 words-per-minute.

During February Miss Reinicke read in the Hitchcock book at an average rate of 8.0 words-per-minute. From January 1966, when she began reading test and practice material from her book, to June 1966, the average line-change time decreased to about three seconds. Maximum reading speeds of 15 words-per-minute for a two-minute period and 10 words-per-minute for 10 minutes were observed.

During January 1967 Miss Reinicke used the Visotactor A and Recognition Prototype II to read typewritten single sheets. Four reading sessions of about two hours each yielded a number of results. Miss Reinicke learned to identify correctly all the spelled-speech letters and acquired a small vocabulary of spelled-speech words such as: and, the, these, they, etc. Proper tracking was achieved early in the series of reading sessions although line change initially required 10-15 seconds instead of the 3-5 seconds required with the Visotactor B. With such a limited vocabulary, reading speeds were generally 20 words-per-minute or below.

The lower-case s presented some difficulty because it could be missed by the machine even if most of the other letters were correctly identified. To improve the recognition of the s a priority circuit like that used for the g was installed in the Recognition Prototype II. The upper- and lower-case z were connected so they are always pronounced s. (Since z occurs roughly only once in 500 letters, this simplification seems justified.) These changes

Figure 5a. Miss Bonnie Reinicke is shown using the Visotactor B and Colineator to read a book. She has used this equipment in her home since September 1964.

Figure 5b. Miss Bonnie Reinicke began using the Recognition Prototype II at Mauch Laboratories in January 1967. This bench model of the recognition machine contains the photoelectric Word Synthesizer, the Word Storage Unit, and electronic logic similar to that planned for the pre-production prototype.

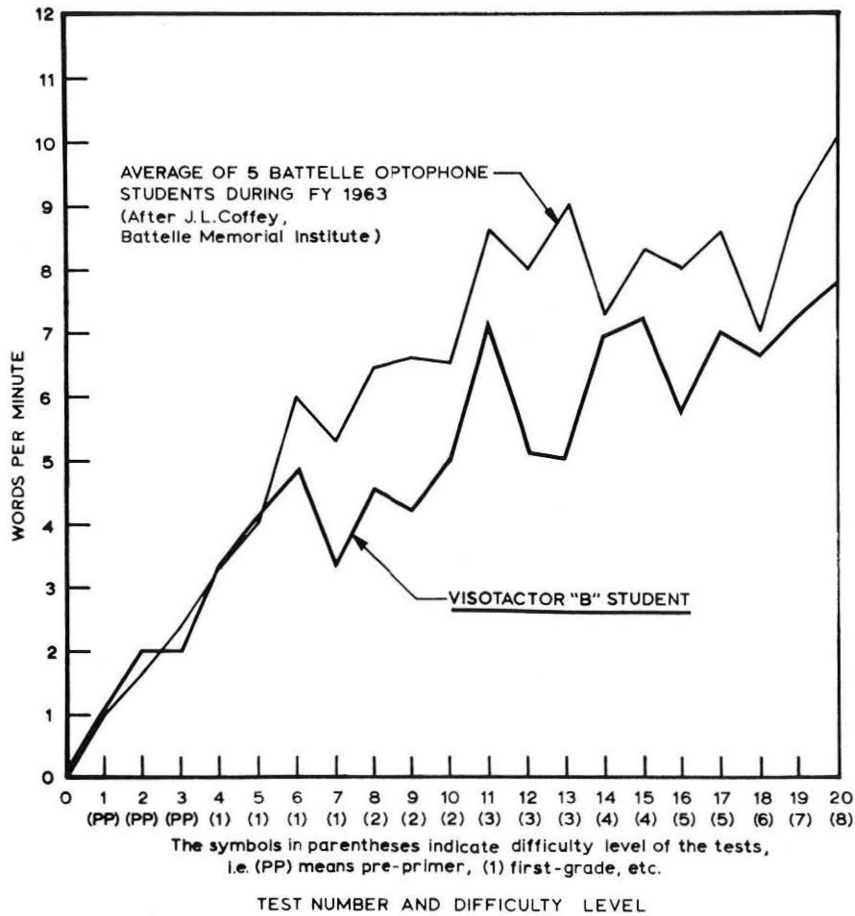


Figure 6. The learning curve for the first Visotactor B student closely follows a curve well established for students of an optophone. Training procedures for the Visotactor B student were very informal, since she read most of the Battelle lesson material at home, without guidance, and lacked the tape-recorded practice material used by optophone students. Both curves include line-change time. Each test score is for ten minutes of continuous reading. Word identification errors, mainly substitutions, are only about 1%.

permitted removing several diodes from the s letter line in the recognition matrix which, in effect, provides a less stringent definition of the letter s.

Sometimes a missed letter resulted in silence rather than the substitution of another letter. This was very misleading when the silence replaced the first letter in a word. The logic was modified so that such silent periods are replaced by the 440 Hz "beep." The silent channel (00000) is still available for use between words, as planned.

Miss Reinicke read additional typewritten material with the Visotactor A and Recognition Prototype II during three reading sessions in February 1967. The recognition errors which seemed to cause the most difficulty for Miss Reinicke were investigated after each session. Frequently, minor modifications in the recognition logic eliminated one or more sources of errors.

During March the Recognition Prototype II was used to read (at a speed of 6 letters-per-second) a list of the 500 most frequent English printed words. The error-free spelled-speech output was recorded on magnetic tape. The words were not arranged in any particular sequence. About one second was added between words to permit the listener to respond if he could identify the spelled-speech version. Replay of the tape recording takes 18 minutes and 40 seconds. Two sighted listeners without previous spelled-speech experience recognized 58 per cent and 63 per cent of the 500 words. Miss Reinicke correctly identified 70 per cent the first time she heard this recording; after playing the tape once per day for five days she scored 94 per cent (or 470 words) correct. Assuming that she would have a spelled-speech vocabulary of 500 words, the probability of encountering a word that happens to be outside the 500 most frequent ones is about 0.3. This means that in a line of ten words approximately three words are new to Miss Reinicke in the sense that the rapid spelled-speech versions are not recognized on the first scan and have to be re-scanned. This speed handicap will decrease gradually as Miss Reinicke's spelled-speech vocabulary increases. At the end of December 1967 the Recognition Prototype II was loaned to Miss Reinicke for daily use in her home to help increase her spelled speech vocabulary.

Several improvements planned for the Recognition Prototype

II were installed and tested during November 1967. The two main changes are a new photocell array to improve recognition of v, w, and y for certain type fonts and to improve word space detection, and a switch to detect reverse scanning.

Development of Multicolumn Visotactors

Exactly what modifications are made in the Multicolumn Visotactor design depends on the evaluation of the present prototype. The possibility exists that a future version of the Multicolumn Visotactor may stimulate the underside of only one finger with a large number of vibrators.

This direct translation reading aid which will fit on the finger doing the scanning has been given the name Digitactor. The Digitactor will not incorporate the lens and the variable magnification mechanism used in the Visotactor and the Visotoner. Instead, the Digitactor will have a number of light pipes which conduct light to the printed page or object being examined and other light pipes to guide reflected light to the array of photocells. By having 16 stimulators per column, the Digitactor can be used to read letters which may vary over a size range of slightly more than 2 to 1 (7- to 16-point size, for example). The resolution at the smallest type size will be as good as that in the Visotactors-Visotoners (eight photocells per column) and much better at the larger, more frequent type sizes. Many factors must be studied before the first Digitactor prototype can be built.

Miss Reinicke was given the Multicolumn Visotactor Prototype during June 1966. Only the center column of eight stimulators was turned on. One of the stimulators was adjusted to make better contact with her finger. Miss Reinicke's reading speed after several hours was less than 1 word-per-minute. After about ten hours practice she had become accustomed to the weak tactile stimuli produced by the piezoelectric stimulators in the Multicolumn Visotactor and she read at an average rate of 4.0 words-per-minute using only one column of the three available in the Multicolumn Visotactor (Fig. 7).

On June 19, 1966 Miss Reinicke read from *Alfred Hitchcock's Ghostly Gallery* at an average rate of 6.6 words-per-minute, using only the center column of stimulators. On July 25 she read at an

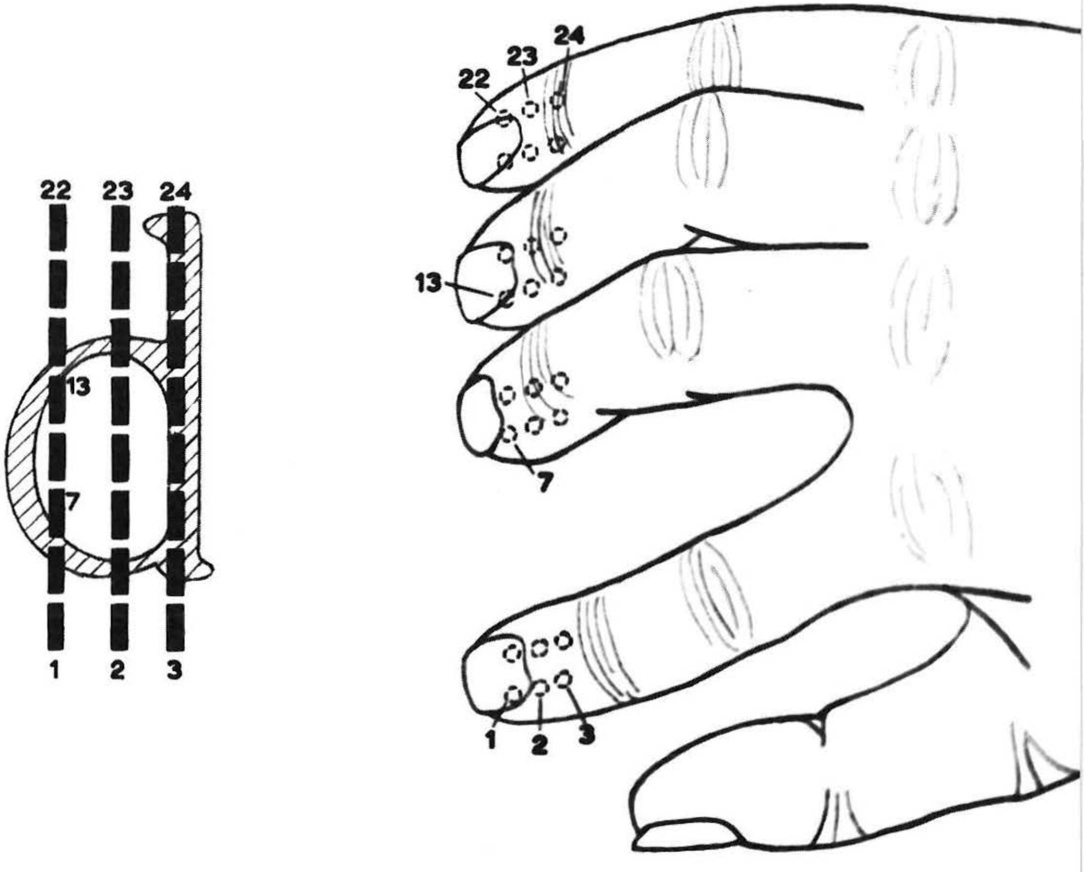


Figure 7. This schematic shows the image of the letter d projected on the photocells of the Multicolumn Visotactor prototype and the location of the stimulator controlled by each photocell.

average rate of 3.8 words-per-minute. An examination of the device disclosed that one stimulator was weak. Miss Reinicke was not aware of this, however, and it is not clear whether or not it affected her reading rate.

During October 1966 Miss Reinicke used the Multicolumn Visotactor with two columns (center and proximal) turned on. Her reading speed was less than 1 word-per-minute at the beginning of the month and increased to 4.0 words-per-minute by October 20, when she read from a second book, *Hitchcock's Solve Them Yourself Mysteries*.

Miss Reinicke continued to read at about 4 words-per-minute during November. Reading speeds of 2.5, 4.3, 3.9, and 3.4 words-per-minute were recorded. On November 29 Miss Reinicke began using the Multicolumn Visotactor with all three columns operating. On December 20 Miss Reinicke was timed for the first time using the Multicolumn Visotactor with all three columns of stimulators turned on. Her reading speed was 3.1 words-per-minute. During January 1967 reading speeds of 1.4, 1.7, 3.2, and 2.3 words-per-minute were measured.

With approximately 30 hours practice under each condition (one, two, or three columns on) her reading speeds averaged 4.1, 3.5, and 2.3 words-per-minute, respectively. Since Miss Reinicke had about 300 hours experience with the Visotactor B (single column of electromechanical stimulators) it may be that the time she spent with the Multicolumn Visotactor was not enough to outweigh the habits acquired earlier. To avoid this complication, it is intended for the future that naive subjects will use the present Multicolumn Visotactor or a modified version with all columns operating beginning with their first exposure.

It may develop that a Visotactor B simulator which presents "perfect" tactile stimulation to a student would speed learning. In its simplest form, the simulator would take the upper eight tones from a tape recorder which replays the tape versions (7-½ ips) of the 200-lesson Battelle optophone training course and use the presence/absence of each tone to turn on/off the corresponding tactile stimulator of a Visotactor. If a tape recorder with three operating speeds (7-½, 3-¾, 1-⅞) is used with additional tone selecting circuits, reading speeds one-half and one-fourth that originally recorded could be simulated. Electronic logic simular to a shift register would permit an approximate simulation of a multi-column tactile reading aid. If a two-channel (stereophonic) tape recorder is used, the teacher's comments (or spelled-speech letters) could be recorded on the second track. In the future, this approach might facilitate learning the tactile output code of the Visotactors A and B and the spelled-speech output of the Recognition Reading Machine.

Summary

Several type reading machines for the blind have been developed which provide the user with immediate access to ordinary printed materials without the need for transcribing into Braille or recording by voice. Preliminary testing of prototypes of these machines by blind users has begun and the results are encouraging.

REFERENCES

Clark, L. L. (ed.). *Proceedings of the International Congress on Technology and Blindness.* New York: American Foundation for the Blind, 1963. 4 vols.

Lende, Helga. *Books about the blind: a bibliographical guide to literature relating to the blind.* New York: American Foundation for the Blind, 1953.

Metfessel, M. F. Experimental studies of human factors in perception and learning of spelled speech. *Proceedings of the International Congress on Technology and Blindness. . . . 1,* 305–308.

St. Dunstan's International Conference on Sensory Devices for the Blind. *Proceedings.* London: St. Dunstan's, 1966.

Zahl, P. A. (ed.). *Blindness: modern approaches to the unseen environment.* Princeton: Princeton University Press, 1950.