THE LEXIPHONE: AN EXPERIMENTAL READING MACHINE FOR THE BLIND

by

CHARLES GARRY AKERMAN CAPLE

B.A.Sc., University of British Columbia, 1963

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in the Department of
Electrical Engineering

We accept this thesis as conforming to the

required standard

Members of the Department

of Electrical Engineering

THE UNIVERSITY OF BRITISH COLUMBIA

February, 1966

ABSTRACT

An experimental reading machine for the blind has been built to test a proposed multidimensional audible code. This device, patterned after the popular Optophone reader, can generate either the multidimensional code or a simulated version of the Optophone code. The results of tests carried out with two blind subjects show that multidimensionally-encoded letters and words can be learned and "read" with reasonable accuracy, even when entirely different dimensions of the code are utilized. A comparative evaluation of the multidimensional and Optophone codes, based on the performance of 52 sighted persons, suggests that the multidimensional code provides a better basis for letter discrimination.

A detailed study of the discrete print signals produced by this machine is presented. The results of this study suggest that this particular print scanning system does not lend itself to automatic letter recognition, but that, with some pre-processing of the print information, some optimization of the audible code can be achieved. It is also demonstrated that the information produced by this machine is highly redundant, and that the discrete nature of the print translation process may psychologically limit the maximum reading speed, regardless of the audible code employed.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGEMENT

The author would like to thank Dr. M.P. Beddoes, the supervisor of this project, for his encouragement and assistance throughout the course of the work. The author is indebted to Professor E.S.W. Belyea for his considerable assistance with the code comparison experiments.

To Mrs. Eva Williams and Miss Cynthia Moffatt, who so willingly worked with the reading machine code, and to the volunteers of the Canadian Institute of the Blind who provided transportation, sincere thanks are offered.

Special thanks are given to my wife, Marcia, for her help and encouragement.

Acknowledgement is gratefully given to the National Research Council for financial assistance, and for their co-sponsoring, with the Mr. and Mrs. P.A. Woodward Foundation, the work described here.

THE LEXIPHONE: AN EXPERIMENTAL READING MACHINE FOR THE BLIND

1. INTRODUCTION

1.1 The Problem

For more than a half century the workers of many varied disciplines have laboured, but with limited success, to provide a simple, personal machine with which the blind might "read" ordinary print. A multitude of simple reading machines, known as direct-translation machines, have been proposed and built, but all have failed in one important respect: no direct-translation machine tested to date has enabled a blind person to read more quickly than 30 words per minute. In fact, average machine reading rates have been closer to 10 words per minute. Researchers in this field consider that in order for a reading machine to be truly useful and widely accepted, it should provide a reading speed of at least 60 to 100 words per minute.

Consequently, the reading machine problem still remains to be solved: to construct a simple, cheap ($500 - $1000), and portable machine capable of translating print into a coded form easily assimilated by a blind person at speeds greater than 60 words per minute.

1.2 Previous Attempts to Solve the Problem

Most of the earlier reading machines presented the print information in terms of some audible code. Each new machine was based on a different letter scanning principle, and each different principle inherently gave rise to a new type of audible code. While the character of the resulting codes differed considerably, each code was founded on the assignment of various audio frequencies to certain parameters of the print detected by its scanner. The machine user had to associate with each letter a characteristic pattern of frequencies. While none of these codes provided the reading performance expected, it still seemed reasonable that the bottle-neck to higher reading speeds lay in the type of audible code employed; it was simply a matter of discovering a code well-suited to the human channel.

1.3 The Proposed Code

An audible code easily assimilated by the human processes is that of speech. While many of the components of speech arise from pure frequencies or tones (vowels and voiced consonants), a great number of the components (fricatives, sibilants, plosives, clicks, etc.) have a completely non-tonal basis. This observation suggests that an audible code more likely to yield the higher reading rates desired would, unlike the earlier codes which utilized only a single auditory variable (frequency), be one that incorporates the use of several simultaneous variables. Hence we propose just such a new code, called a multidimensional code, whose components are derived from several auditory variables, or dimensions. An auditory dimension is defined as a stimulus variable whose value can be manipulated independently of any other stimulus variable. For example, the frequency and amplitude of a tone are considered to be independent dimensions because the value of one variable may be altered without affecting the value of the other. Of course, certain dimensions may to some extent be subjectively inter-dependent. This consideration does not affect the definition of a dimension, but it does control the choice of dimensions. Only experiments can show which dimensions are most subjectively independent and therefore most effective.

The decision to investigate a multidimensional code is also supported by the experimental results of many workers. Pollack and Ficks[1] have demonstrated that human subjects are able to assimilate a greater amount of information per audible stimulus when the stimuli vary along several dimensions, than when they vary along only a single dimension. This means that, for a given number of alternative stimuli, a multidimensionally-encoded stimulus is more easily discriminated than one encoded in a single dimension. If it can be demonstrated that a person working with a multidimensional machine code is able to "read" more accurately than a person working with a unidimensional code produced by the same machine, then there is a reasonable possibility that the multidimensional code can also provide higher reading rates. We intend to make just such a comparison between our multidimensional code and a popular unidimensional code known as the Optophone code.

The multidimensional machine code we propose is based on the audible dimensions successfully employed by Beddoes, et al.,[2] in a letter recognition experiment. These investigators used the dimensions of tone frequency, tone timbre, hiss bandwidth, and click, to encode each letter with a single multidimensional stimulus. These same dimensions are appropriate for our purposes where we wish to encode each of many elements of a letter with a single multidimensional stimulus. These elements are those normally encoded by the direct-translation machine into the frequency dimension alone.

1.4 Scope of the Project

A versatile direct-translation reading machine was constructed for this project. Patterned after one of the more popular reading machines known as the Optophone (discussed in detail in Chapter 2), our machine was designed to produce either of two audible codes: the original Optophone code or the multidimensional code. The valuable feature of this machine is that the components of each code are controlled by exactly the same print signals. Consequently, if the print noise is the same for both codes, any difference exhibited between the two codes, in terms of reading performance, would be due to the codes alone, independent of the print scanning technique.

Both absolute and comparative evaluations of the multidimensional code were carried out using coded material produced by this machine. The reading performance of two blind subjects provided the absolute evaluation, while the relative performance of two groups of sighted students working with the two codes offered a comparative evaluation of the multidimensional and Optophone codes.

The print signals produced by our experimental machine were subjected to a statistical study for the purposes of suggesting audible code simplification, and of determining the amount of informational redundancy and psychological entropy produced by a machine of this type.

1.5 Thesis Outline

A review of past and continuing research in the reading machine field is presented in Chapter 2. Chapter 3 presents a detailed description of our experimental machine and of the two audible codes. Chapter 4 deals with the entire multidimensional code evaluation. In Chapters 5 and 6 the results of the statistical study are presented and discussed. In addition, psychological limitations associated with reading machines of this type are considered in Chapter 6. The final chapter, Chapter 7, summarizes the results of Chapters 4, 5, and 6, and presents appropriate conclusions and recommendations.

2. A REVIEW OF READING MACHINES FOR THE BLIND

This chapter is devoted to a discussion of the considerable effort that has been made to produce a practical reading machine for the blind. A short historical review of previous and continuing research in the reading machine field is given. It is seen that all machines fall broadly into two classes according to their over-all sophistication of print translation, and that these classes can be subdivided into groups corresponding to the human sense modality employed.

The practicality and code problems of the simple direct-translation machine are discussed as an introduction to the Lexiphone.

2.1 Historical Review

Research into reading machines for the blind was launched by Fournier d'Albe of England when, in 1914, he invented the Type-Reading Optophone—a machine designed to transform print into sound.[3, 4] In this first machine, called the White-sounding Optophone, a narrow vertical bar of pulsed light, one letter in height, falls on the print. Eight regions constituting the bar are pulsed at eight audio frequencies by means of a spinning disc. Light reflected from the illuminated area falls on a selenium cell and causes corresponding frequencies to be heard in an earphone. Letters passing horizontally under the illuminated area delete certain frequencies according to the regions occupied by segments of a letter. Consequently, each letter creates a characteristic sound pattern as it is scanned.

However, with this machine, the presence of all eight tones in the absence of print was considered to be a serious drawback. The problem was solved in a modified version of the Optophone produced in 1920 by the English firm of Barr and Stroud. This model is called the Black-sounding Optophone.(4) It utilizes two selenium cells in a bridge which is balanced when no ink is detected. Tones are generated, rather than deleted, by letter segments intersecting the illuminated area. Silence is therefore present during the spaces between letters. This model employs only five tones (<u>sol</u>, doh, ray, me, sol) in place of the original eight, and its sound code consists of a pattern of chords for each letter. The term "chord" here means the simultaneous sounding of one or more tones.

From its creation, to the present, the optophone* has been subjected to continuous evaluation and modernization—perhaps more so than any other reading machine invented since. Reading speeds attained by blind subjects using the optophone have not improved over the years, even with the use of up-dated models. Speeds of 30 words per minute have been recorded for exceptional persons, while speeds of 2 to 10 words per minute are more typical of average performance. Such limited success with the optophone led many research workers to search for other reading machine schemes. These other schemes are now reviewed.

---

* The generic term, "optophone", is now used in the literature to describe a reading machine whose reading mechanism and acoustic code are similar in principle to the modified (Black-sounding) Optophone.

The first tactile reading machine, the Visagraph, was invented in 1928 by R.E. Naumburg.(5) This machine produces an enlarged embossed facsimile of printed letters on aluminum foil. It embosses the foil in six rows corresponding to the print information sensed by a vertical line of six photocells. This particular type of print sensing device is similar in function to that of the Optophone, and identical to the photocell sensing device used in later Optophone simulations. The principle difference between the Optophone and Visagraph, then, is the human sense stimulated. The Optophone is designed to communicate to the aural sense; the Visagraph, to the tactile sense. However, it turned out that embossed letters were even more difficult to "read" than Optophone-coded letters.

During World War II, the Committee on Sensory Devices was formed by the U.S. Veterans Administration in order to stimulate and sponsor research leading to the development of sensory aids, particularly reading machines for the blind. Under the auspices of the Committee, during the period 1944-49, several machines were built and tested. This work was caried out, primarily, at the Haskins Laboratories and at RCA.

The RCA Labs first developed the "Type A-2 Reading Machine."(6) It employs a small point of light rapidly oscillating up and down to scan the vertical extent of a letter. A variable frequency oscillator is coupled to the oscillating light beam in such a way that a high audible frequency is generated when the spot is at the top of its scan and a low frequency at the bottom. This audio frequency is keyed ON whenever the spot of light intersects a letter segment. An optophone-like sound code results when the scanning beam is drawn slowly over a line of print, but superfluous noise is added to the normal code by the keying process.

A later development, the RCA Recognition Machine,(9) utilizes a flying-spot scanner as the input to a 26-letter recognition circuit. A letter is scanned horizontally in eight bands, and the number of intersections are counted and logically processed to trigger an audible output. The spelled pronunciation of each letter, prerecorded on magnetic tape, is heard as the corresponding output. This type of audible code, later termed "spelled speech," has since undergone much serious investigation.

Research at the Haskins Labs included the simulation and evaluation of previous reading machines, a thorough investigation of various audible codes for use with optophone-type and recognition machines, and the development of the FM-Scan reading machine.(8) This latter machine utilizes a frequency-modulated audio tone to indicate the letter scanned. An illuminated narrow vertical slit, one letter in height, is passed over a line of print. As various parts of a letter pass beneath the slit, the total reflected light, which varies according to the proportion of slit area darkened by print, is sensed by a photocell. The resulting analogue signal is used to frequency-modulate an audio oscillator whose output frequency varies between 100 cps (no print "seen" in the slit) and 4000 cps (slit entirely black). A "Zero-suppressor" detects the absence of print between letters or words and squelches the

100 cps output during these periods. After a 90 hour training period with this machine, the average reading speed attained was 4.2 words per minute.

In 1952, P.E. Argyle of Royal Oak, B.C., invented the Argyle Reader. This machine employs a rapidly moving point of light which scans the vertical extent of one letter from top to bottom at a uniform rate and returns instantaneously to repeat the operation. This operation is repeated at 200 cycles per second. The total instantaneous light reflected is collected by a photocell whose signal current is amplified and fed to a loudspeaker. As letters pass horizontally under the flying spot of light, each letter intercepts the spot and generates a pattern of audible transients whose spectral components are multiples of 200 cps. The result of this process is a directly generated audible code in which each scanned letter is represented by a characteristic "growl." This machine was later evaluated at the National Physical Laboratory by Clowes et al.,[9] who found that the reading facility afforded by the Argyle Reader was equivalent to that given by the optophone.

Since 1957, the Veterans Administration has sponsored reading machine research on a broad front. Many new transistorized models of the optophone, now employing nine photocells, have been developed and thoroughly tested for the Administration by the Battelle Memorial Institute, Ohio.[10, 11] An extensive training programme has also been developed by Battelle for users of their device. The Mauch Laboratories, Ohio, have pursued the development of a "multiple snapshot" character recognition machine with spelled-speech output.[12] The Haskins Laboratories have concentrated on the synthesis of speech, with the idea of providing a speechlike code for use with recognition-type reading machines.[13] Also under VA auspices, Professor Metfessel at the University of Southern California has evolved techniques by which synthetic spelled speech may be produced in smooth coalescent form, intelligible at 90 words per minute.[14, 15] His continuing work offers a reasonably simple and familiar output code for character recognition machines.

Nye[16] has carried out a thorough investigation of audible outputs for reading machines, including multidimensional codes, an artificial larynx code, and a synthesized speech code known as "Wuhzi," originally developed by the Haskins Labs. The results of Nye's code tests, in which the same vocabulary of eight four-letter words was used for each, are shown in Figure 2-1. These were not speed tests, but tests for comparative code performance.

Beddoes[2] has studied the multidimensional auditory coding of single letters in connection with a Braille reader. With four-dimensional coding of letters, he found that 80% of the letters could be identified correctly. He has also demonstrated that artificially compressed spelled speech can be assimilated comfortably at 110 words per minute with 82% accuracy.[17]

Other workers have investigated the value of output codes employing the tactile and kinesthetic senses, and of codes utilizing several sensory modalities simultaneously.[18, 19]

J.G. Linvill recently devised a tactile reading aid which displays an enlarged facsimile of the print on an 8 x 5 array of vibrating piezoelectric reeds.[20, 21] The entire device, similar in principle to earlier designs (described by Freiberger and Murphy[22]), can be contained in a hand-sized package. Reading speeds of 15 words per minute have been recorded for one user of this device.

Many varied character recognition schemes have been proposed.[23, 25] Because of size or cost limitations normally imposed by the average blind consumer, most of these schemes are of value only in high-speed business applications.

The characteristics of the reading machines reviewed here are discussed in groups in Section 2.3.

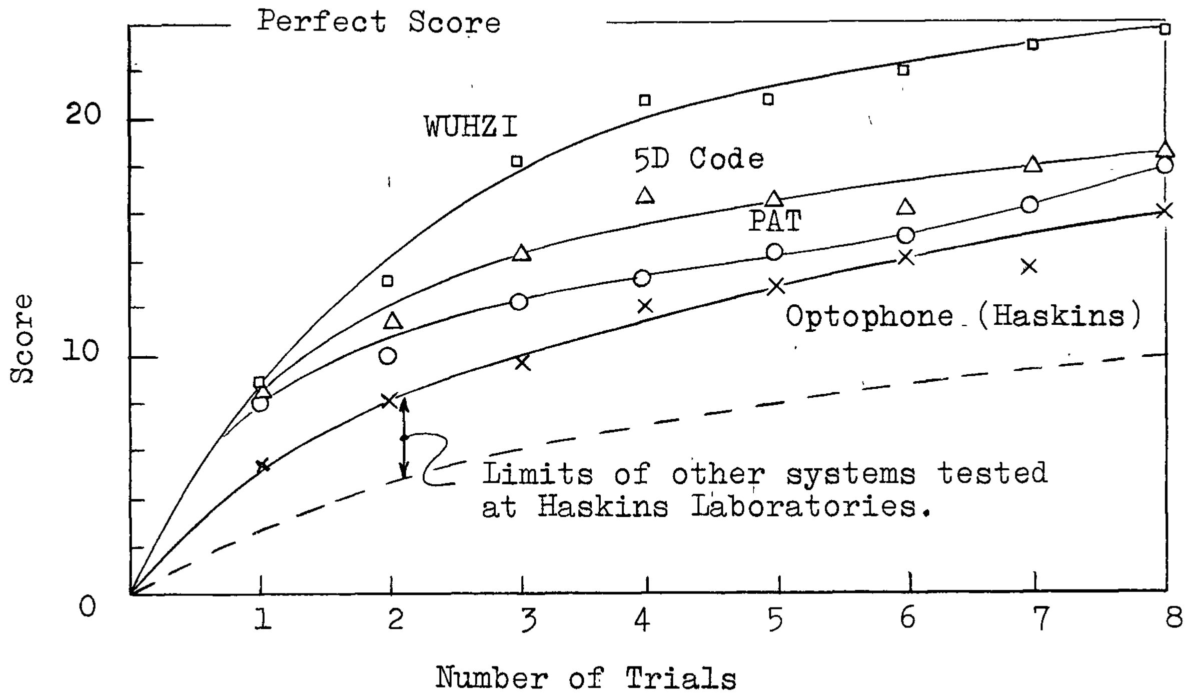

Figure 2-1. Nye's Comparison of Several Audible Codes

2.2 Classification of Reading Machines for the Blind

Reading machines can be categorized according to overall function and sensory output. Functionally, the machines tend to divide into two distinct classes: direct-translation machines and character recognition machines. These classes can also be subdivided into the two output code groups, audible and tactile-kinesthetic, corresponding in name to the senses most often used in supplanting loss of vision. Table 2-1 summarizes this classification scheme.

	DIRECT-TRANSLATION	(Ref.)	RECOGNITION	(Ref.)
Audible	d'Albe Optophone RCA A-2 Reader FM-Scan Machine Argyle Reader Other uni and multi-dimensional optophones	(3) (6) (8) (9)	Output: Haskin's "Wuhzi" Metfessel's spelled speech Other speechlike codes Multidimensional codes	 (8) (14) (16) (19)
Tactile	Naumburg Visagraph Mauch Tactile optophone Linvill's device Other tactile devices	(5) (12) (20, 21) (21, 22)	Tactile systems	(18, 21)

Table 2-1. Classification of Reading Machines for the Blind

The direct-translation machine, as its name suggests, generates output signals directly in accordance with the changing contours of print. No internal processing of the print information takes place. The output signals and the print scanning technique are so closely related that little choice exists concerning the types of sounds that may be generated. In the design of a direct-translation machine, the scanning method and the output code are, unfortunately, effectively inseparable factors.

On the other hand, a character recognition reading machine permits the separate optimization of its two principal functions which are:

1. To scan the print and identify each letter discretely and
2. To trigger, on the basis of this identification, a circuit which generates an output sensory stimulus corresponding to the letter identified.

This basic letter recognition machine generates a distinct output for each letter identified—for example, the spoken sound of each letter. More sophisticated recognition schemes utilize sequences of identified letters to trigger the generation of pre-stored voiced phonemes, syllables, or even entire words. Of course, such sophistication causes machine complexity and associated expense to mushroom.

A third class of reading machines has been proposed,(5) although no complete model has been demonstrated. This machine, termed the "intermediate" reading machine, lies in terms of complexity and cost somewhere between the simple direct-translation machine and the recognition machine. Such a machine theoretically retains the instrumental simplicity of the direct-translation device, but operates upon the derived print information in such a way that speechlike or simple sound units are generated directly from letter features. Mauch(26) initiated work on a machine of this type but soon found that the desirability of a speechlike sound output necessitated a recognition-type input.[12]

2.3 Class Characteristics

To each functional class of reading machines considered in the previous section can be ascribed a set of characteristics typical of all machines in that class.

2.3.1 Direct-translation Machine

The direct-translation machine is generally simple to instrument, inexpensive, and can usually be made portable. However, its simple translation process generates an abundance of code units for each letter, ignoring the fact that the translated graphical information is highly redundant. (See Chapter 6.) An audible code produced by direct-translation cannot be made similar to the efficient natural code, speech, owing to the fact that no unique relationship exists between the shape of a letter and its phonetic equivalent. As a consequence, an audible or tactile code produced by such a machine is not easily learned, and in fact, as experience has shown, limits reading speed to about 30 words per minute. In addition, with the direct-translation machine, an appreciably different code must be learned for every new font style encountered.

2.3.2 Recognition Machine

The character recognition machine allows much higher reading rates to be attained by assigning no more than one output stimulus to each letter. For example, speeds of 90 wpm are possible using spelled speech.[14, 17] At these higher reading rates the fatigue associated with reconstructing contextual material from coded letters is also reduced. With a letter-reading recognition machine only 26 stimuli need be employed (compared with the 52 stimulus patterns generated by the direct-translation machine). In addition, the same 26 stimuli can be associated with a number of font styles, given a multifont recognition machine. The code training period may be made almost negligible with a recognition machine by employing a sensory code consisting of familiar code units such as those of spelled speech.

Of course, the drawback with such an otherwise ideal machine is that it requires relatively complex instrumentation; it is therefore costly to implement and difficult to make portable.

2.3.3 Cost

A functional classification of reading machines for the blind is illustrated in Figure 2-2, showing the relationship of a class to an approximate level of data processing and output code sophistication. For each level is indicated the estimated cost of a complete machine produced in quantity. The first three estimates are quoted from the literature, while estimates 4 and 5 are pure guesses. In fact, the possiblity of even instrumenting the components of level 4 is in question. Also level 4 may not be a desirable or necessary step to the realization of level 5.

Cost can be used as a rough index of instrumental complexity and of machine dimensions.

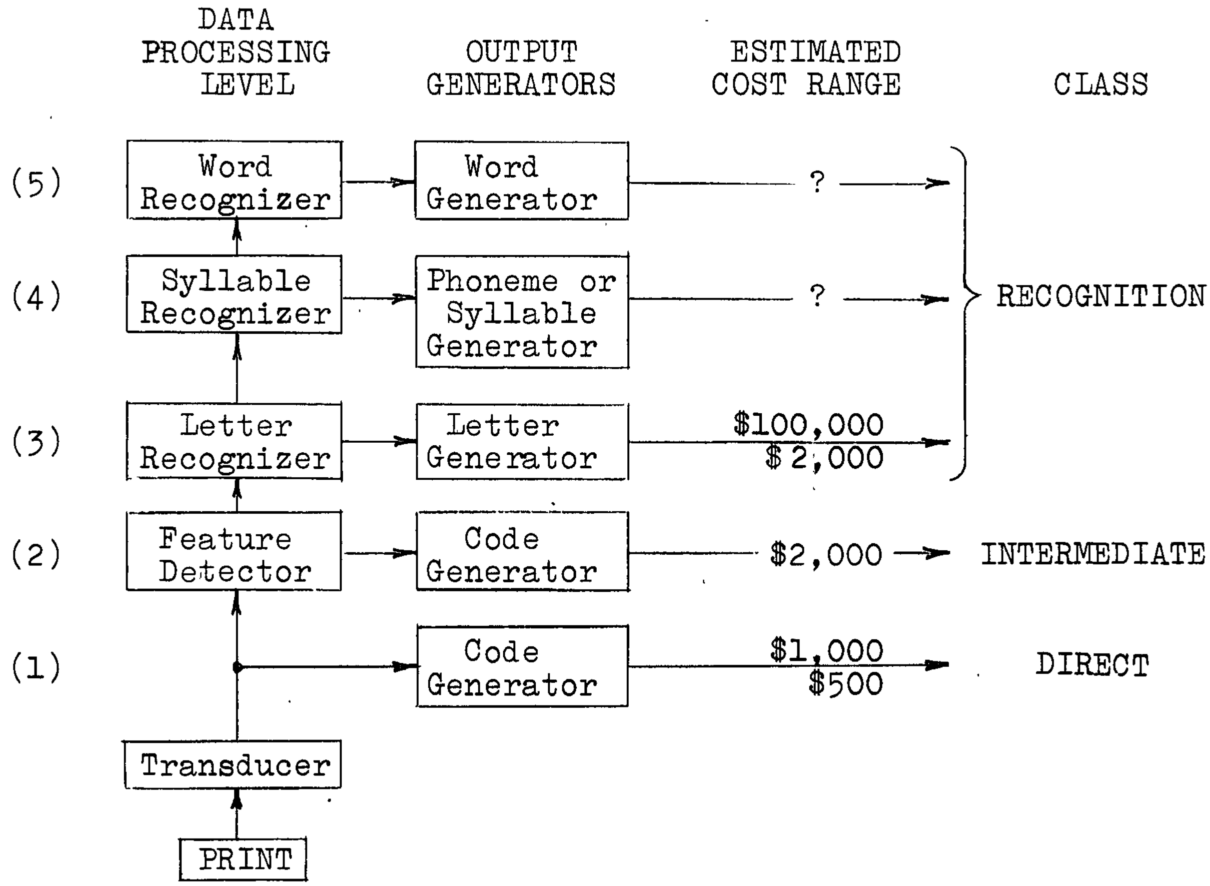

Figure 2-2. The Reading Machine Hierarchy

2.4 Introduction to the Lexiphone

Figure 2-2 immediately indicates why so many workers have pursued the development of a simple direct-translation machine, rather than a more sophisticated machine. It is by far the least expensive machine to produce. Also, as pointed out earlier, it is easily made portable. Hence, such a machine is well suited for and well within the means of the average blind person, and consequently deserves much attention.

For these reasons it was decided to investigate further the direct-translation machine, in an attempt to resolve the major problem experienced with machines of this type; namely, the low reading speeds normally attained.

With the optophone, it was felt the problem lay in the type of audible code employed. If, in place of the normal optophone "chord" code, a multidimensional auditory code were employed, then perhaps an increase in reading speed might be possible. This would be due to the fact that auditory stimuli varying along several dimensions are more easily identified than those stimuli arising from a single dimension.(1)

In order to test reading performance with a multidimensional sound code, a direct-translation machine was built. While its print pickup device is similar to that of the optophone, its output code employs several dimensions in sound. For the sake of brevity, this machine, a multidimensional optophone, is called the Lexiphone.* Its description and performance characteristics form the subject of subsequent chapters of this thesis.

---

* This name, meaning "words in sound," was suggested by Professor E.S.W. Belyea.

3. DESCRIPTION OF THE LEXIPHONE

The Lexiphone (Fig. 3-1), although designed principally to generate the multidimensional Lexiphone code, was constructed to generate as well a simulated version of the optophone code. The latter code provides a suitable standard against which the hopefully superior Lexiphone code can be evaluated. The two code generators, while controlled by the same print information

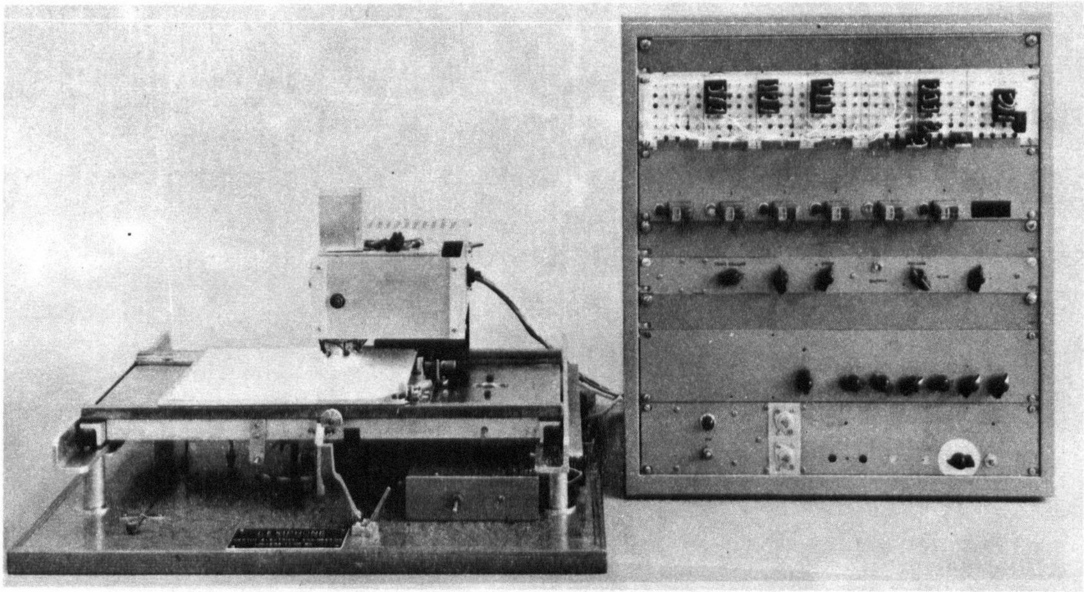

Figure 3-1. Photograph of the Lexiphone

signals and machine components, develop distinct codes by assigning different acoustic variables to the same signals.

Discussed in this chapter are the system components used to produce these signals, and the audible codes themselves.

3.1 Principle of Operation

The print information signals are generated by a linear array of six photocells, situated at the image plane of a lens which is focused on the printed page. Oriented in a direction perpendicular to the horizontal line of print, the six cells together span the vertical extent of letters in a line. As the page is transported horizontally from right to left, the stationary cells scan a line of print and convert the print into six electrical signals. When a photocell detects the presence of ink it activates a corresponding relay. The instantaneous state of all six relays determines the code sound generated.

Illustrated in the block diagram of Figure 3-2 are the system components which constitute of Lexiphone. They are discussed individually in the sections that follow.

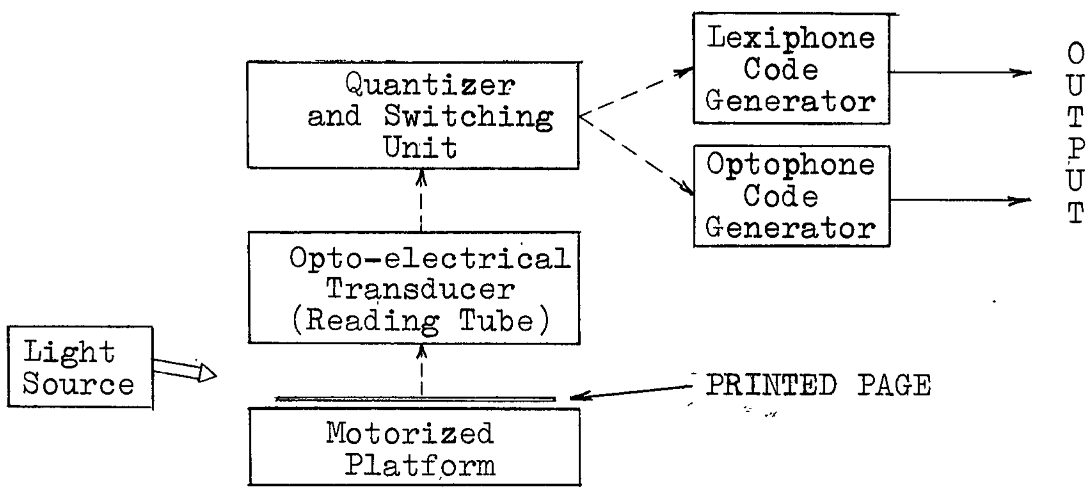

Figure 3-2. Block Diagram of the Lexiphone

3.2 The Light Source

Initially, a stroboscopic light source was used to illuminate the print. Flashing illumination was employed so that the photocell signals could be ac-amplified, thus circumventing the temperature drift problem associated with dc-connected transistor amplifiers. However, this type of light source was discarded in favour of a constant intensity incandescent source because the xenon flashtube in this unit generated a distracting amount of acoustic noise, and was subject to short life and variable light output. As well, the stroboscopic unit was undesirably bulky.

It was found that if the high-impedance photocells used were properly matched, the temperature drift effect would be negligible with respect to the total photocell signal. Thus, simple, reliable dc-connected circuits could safely be employed between photocells and relays.

The present print illumination source proved to be the simplest, most compact incandescent source of those investigated. It utilizes a ring of six pre-focused flashlight bulbs positioned just above the page so as to illuminate an area the size of one letter. The bulbs, rated at 2.2 volts, are connected in parallel and driven by a regulated dc voltage supply whose output is continuously variable between 1.4 and 2.2 volts, is constant to within 5 millivolts, once set, and has less than 1% peak ripple.

These last two supply voltage characteristics are important for reliable operation of the machine. First, the illumination level, once set, must be kept within certain bounds in order that the range of photocell voltages generated is constant. These bounds dictate that bulb voltage must drift no more than 30 millivolts. The supply drift is well below this figure. Second, the thermal time-constant of the bulb filament is so short that any 120-cycle ripple voltage impressed across the filament is reproduced as illumination ripple. This ripple is detected by the photocells and causes the relays to "chatter" when switching. The 1% supply ripple is well below that necessary to cause "chattering."

3.3 The Motorized Platform

To ensure the reliable translation of print into code, the page of print is transported at a uniform rate on a flat, motor-driven platform. A small induction motor moves the platform from right to left until a complete line of print is scanned. A microswitch then reverses the motor, mutes the audible output, and returns the platform to the opposite end where a ratchet assembly moves the page up one line. This cycle is repeated until the entire page is scanned. There are vernier adjustments to make the lines of print parallel to the motion (horizontal alignment) and to position the print directly under the photocells (vertical alignment).

The scanning speed is continuously variable in two ranges: 1 to 2½ and 2 to 5 characters per second (corresponding to the total equivalent speed range of 12 to 60 words per minute). A variable frequency phase-shift oscillator supplies the ac voltage which is amplified by a 100-watt amplifier and which in turn drives the induction motor. The two speed ranges are supplied by a rack-and-double-pinion gear assembly.

In Figure 3-1, pictured on the left are the platform and the cantilevered unit containing the light source and reading tube. On the mounting board at the front are situated three controls for the user: a platform reversing switch, a drive-release lever, and a motor switch. The speed control knob is located on the cabinet front.

For reasons of portability, a commercial model of the Lexiphone would probably substitute a hand-held reading probe for the bulky motorized platform and cantilevered unit. However, in our experimental model, code uniformity, not machine portability, was the factor of greater concern.

3.4 The Reading Tube

Figure 3-3 shows the Lexiphone lens-photocell assembly that transforms print into electrical signals. This assembly, called the reading tube, contains six cylindrical photocells each 0.082 inches in diameter, located at the image plane of a two-element magnifying lens. The lens provides print magnification sufficient to ensure reasonable photocell resolving power. For letters of the standard Pica typewriter font, the font that this reading tube was designed to scan, resolution sufficient to distinguish all significant letter features is obtained when the standard letter width (0.1") is magnified to 9.5 times the photocell diameter. This leads to print magnification of 7.8.

A top view of the reading tube in Figure 3-4 shows

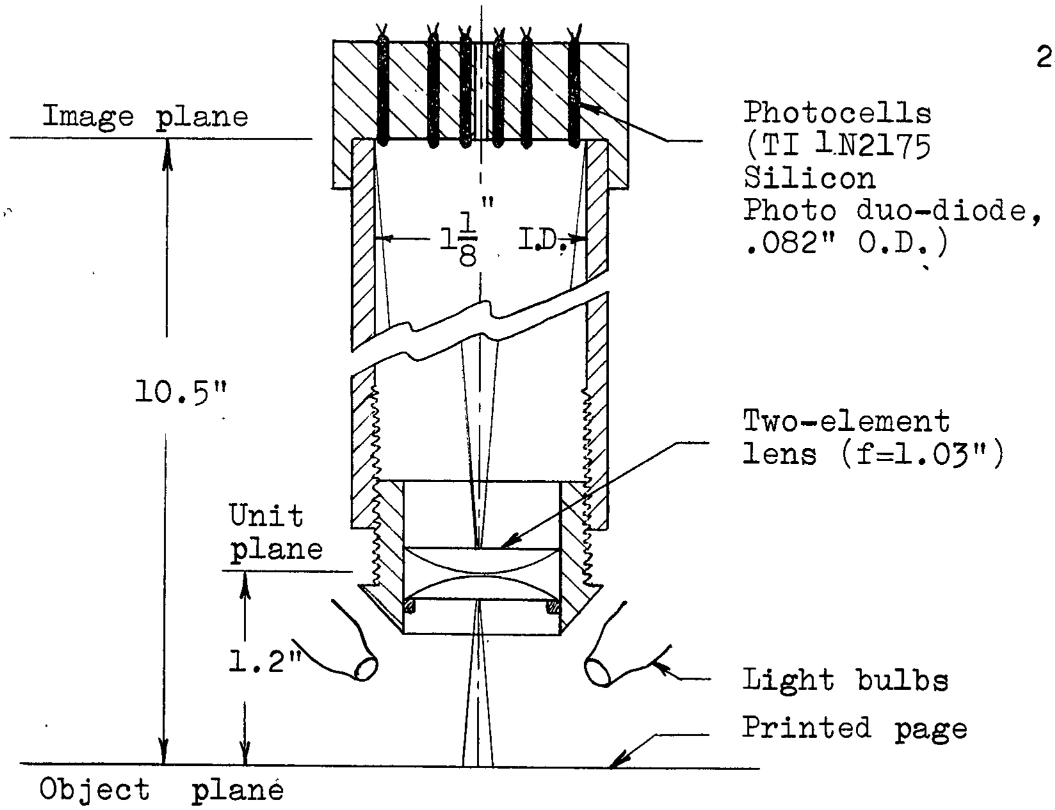

Figure 3-3. Lexiphone Reading Tube

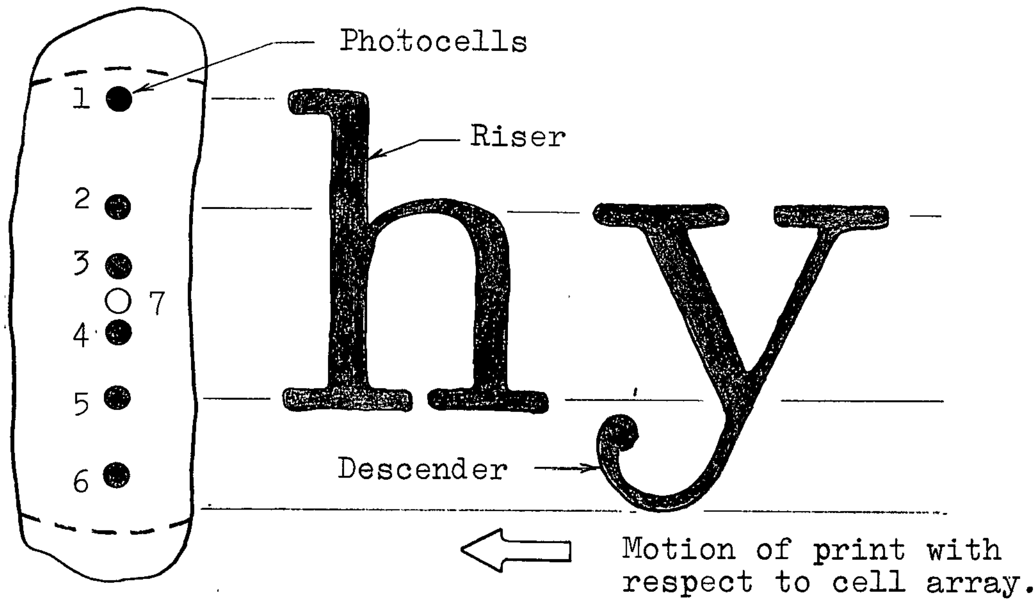

Figure 3-4. Relation of Photocells to Print

the position of the photocells with respect to correctly aligned, magnified print. The print is drawn horizontally past the cell array as shown. Cells 2 to 5 detect print information concentrated in the area defined by the lower-case main body height. All lower- and upper-case letters cause these cells to be activated. The uppermost cell (#1) detects risers, dots, and the top of all upper-case letters, while the lower cell (#6) detects descenders. This particular cell configuration is called the even-cell configuration.

The seventh cell location shown provides for an odd-cell configuration in which cell 7 substitutes for cells 3 and 4. This 5-cell configuration is necessary to distinguish between the letters e and c. When cell 7 is not substituted, the horizontal line of the e falls between cells 3 and 4 and fails to activate either of them, and therefore e is translated as c.

3.5 Quantizer and Switching Unit

When a photocell detects segments of print on the illuminated page its resistance varies between the approximate values of 100 kilohms and 1 megohm, corresponding to no-print and print. These variations produce an analogue voltage signal which is quantized by a subminiature relay into a succession of binary states: "white" or OFF, and "black" or ON.

One of the six identical quantizing circuits is shown in Figure 3-5. In this circuit the photocell voltage appears as the output voltage v of a Darlington circuit. Resistor R_1

sets the "white" output voltage v_W. With the photocell seeing white, R_1 is increased until the Darlington buffer is just turned off ($v_W \approx 0.5$ volts). Resistor R_2 sets the switching threshold

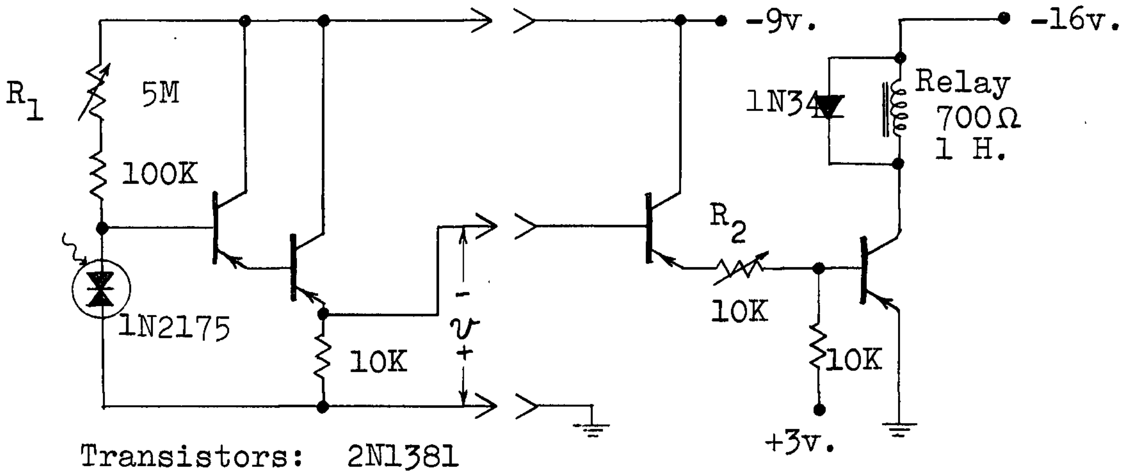

Figure 3-5. Photocell Signal Quantizing Circuit

voltage v_T ($v_T \approx 2.0$ volts). The "black" output voltage v_B, which depends upon the ratio of R_1 to the photocell "black" resistance and upon the Darlington leakage current, generally exceeds 5 volts. These voltage values are summarized in Figure 3-6.

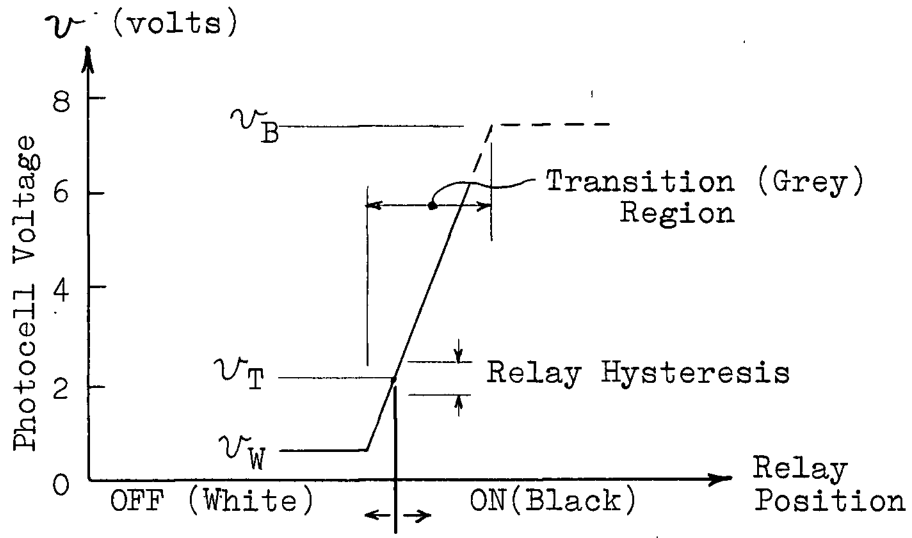

Figure 3-6. Photocell-Relay Quantizing Characteristic

Because the transition region (Fig. 3-6) is passed through so quickly in practice, and the relay hysteresis region is so small a fraction of the entire output voltage swing, the use of a Schmitt trigger in this circuit is unnecessary. Voltage drift of $\sim$(0.2 volts) due to ambient temperature variation is also insignificant compared with the total voltage swing.

Each relay is fitted with four type-C (SPDT) contact sets, two of which are allotted to general code switching functions and appear on a patch board. The third set of each relay is wired in to the Lexiphone code circuit to switch on the audible output if any relay is ON and to mute the output if all relays are OFF. An individual pilot light for each relay is activated by the fourth set.

The functional simplicity and sufficient switching speed (< 10 ms.) of subminiature relays adequately suited Lexiphone switching requirements. Faster semiconductor switches were considered unnecessary; in fact, the need for versatile switching facilities in this experimental machine precluded their use.

3.6 The Lexiphone Coder

Four acoustic dimensions are employed in the production of the Lexiphone code: tone frequency, tone timbre, hiss bandwidth, and click. The value of each dimension is controlled by the state of one or two photocells (relays), according to the scheme of Figure 3-7. Referring to Figures 3-4 and 3-7, one can see that particular letters or letter features control two

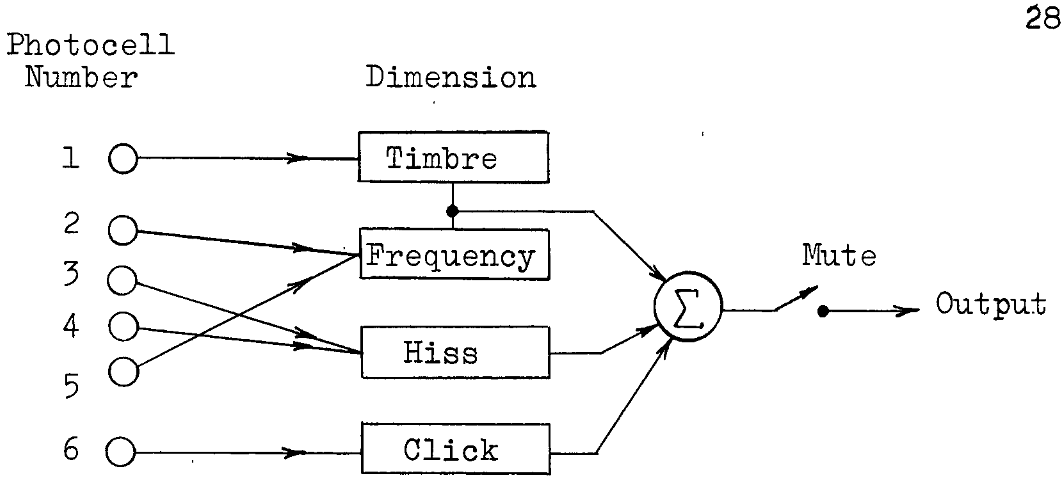

Figure 3-7. Control of Lexiphone Code Dimensions

of the dimensions. Specifically, risers and capital letters cause a change in timbre; descenders produce a click. Of course, all letters cause variation in the tone frequency and hiss bandwidth because cells 2 to 5 are activated by every letter.

Listed in Table 3-1 are the dimensional values employed in the Lexiphone code. The four frequencies are generated by an astable multivibrator whose feedback capacitor is switched to four different values by relays 2 and 5. The output of the astable is a spike waveform whose timbre is "harsh". A smoothing filter is switched in by relay 1 to produce "soft" timbre. A noisy transistor generates wideband hiss which is amplified and passed through one of four filter connections, determined by the state of relays 3 and 4, to yield four hiss bandwidths. When the 5-cell configuration is employed, only two values of hiss can be used: wideband hiss and no hiss, corresponding to the two states of photocell 7. The click is produced upon energization of relay 6, which applies a step voltage to an RC differentiator connected to the output. The various components of the code are summed in a mixer, amplified, and fed to a loudspeaker.

Dimension	Dimension Values			
Tone Frequency (cps)	250	298	334	420
Tone Timbre	Soft (Sinewave)		Harsh (Spiked wave)	
Hiss Bandwidth (6 cells)	Low-pass (1000 ~)	Band-pass (500-1000 ~)	High-pass (500 ~)	No hiss
(5 cells)	All-pass		No hiss	
Click	Click		No Click	

Table 3-1. Values of the Lexiphone Code Dimensions

3.7 The Optophone Coder

The optophone code employs only one acoustic dimension: frequency. A distinct audio frequency is assigned to each photocell, and that frequency is heard whenever its corresponding cell senses print. With six photocells, up to six frequencies may be heard simultaneously. This control scheme and the code frequencies used in the optophone simulation are indicated in Figure 3-8.

In our machine, the six frequencies are spaced at equal logarithmic intervals, after the Battelle Optophone. Six phase-shift oscillators provide these individual frequencies, and the sinusoidal output of each oscillator is clipped to produce a square waveform similar to that of the original Fournier d'Albe Optophone. Each oscillator signal is connected to the output mixer only when its corresponding relay is energized;

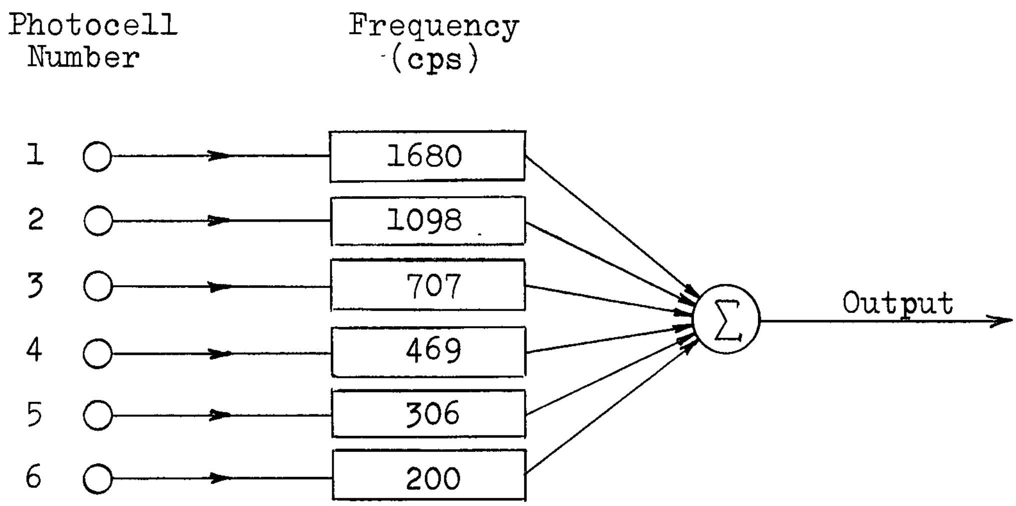

Figure 3-8. Optophone Coding Scheme

if no print is sensed, all relays are OFF and the optophone output is silent. The necessary relay contacts are connected in to the optophone circuits by means of the patchboard evident in Figure 3-1.

4. EXPERIMENTAL EVALUATION OF THE LEXIPHONE CODE

4.1 Introduction

In order to establish the value of the Lexiphone multidimensional code, it was necessary to determine the answers to two questions:

1. Is it possible for a blind person to "read" letters and words presented in the Lexiphone code? and if so,
2. Is the Lexiphone code more effective than the optophone code in conveying print information?

Two series of experiments were performed to answer these questions. In the first series, two blind subjects worked with the Lexiphone code over a period of two months (about 36 hours training time). They learned to read a large number of Lexiphone-coded words constructed from a restricted alphabet of nine letters.

The second series of experiments was designed to demonstrate any difference in coding efficiency that might exist between the Lexiphone and optophone codes. Based on the same nine-letter alphabet, training and test tape-recordings were prepared for each of the two codes. These tapes were used to test the code learning ability of two groups of sighted students. It was intended that the quantitative results of these tests be used to assess the relative value of the two audible codes.

In these experiments the nine-letter alphabet consisted of the letters a e s c r g y h t. These letters were chosen to represent those found most often in printed English (e t a o n r i s), those with risers and descenders (g y h t), and those normally found most difficult to decode because of their apparent similarity in code (a e s c). The need to conserve code training time necessitated our resorting to such an abbreviated alphabet. However, difficult letters were included to make the decoding task more realistic.

It became apparent during tests with the blind subjects that the six-cell configuration (Sec. 3.4) provided insufficient information for them to distinguish between letters e and c. With a 52-letter alphabet this difficulty might have been overlooked, but with a 9-letter alphabet, the decoding problem was severely aggravated. For this reason, throughout the audible-code experiments the five-cell configuration was employed.

4.2 Blind Reading Experiments with the Lexiphone Code

For a period of two months, two blind persons, both female, volunteered an hour a day to work with the Lexiphone code. During the first ten sessions of code training they learned to recognize the nine coded letters presented individually. For the remaining twenty-two hours, they concentrated on decoding words composed of these letters. Some additional time was spent in testing their ability to perceive certain parameters of the Lexiphone code.

Each training session comprised fifty minutes of code instruction and ten minutes of testing. This training method made possible a continuous measure of their progress.

4.2.1 Letter-reading Tests With the Blind Subjects

For the letter-reading experiments, code training proceeded as follows. The experimenter announced the letter to be heard in code. Then the Lexiphone presented the coded letter three times in succession allowing a space of one second between repetions. This training cycle was repeated throughout the instruction period, with each of the nine letters appearing randomly. At the end of the instruction period a short test was given in which letters were presented randomly and without the voice clue. During the 10 seconds of silence following each presentation the subjects were asked to record in Braille their response to each coded letter. Their test results are illustrated in Figure 4-1.

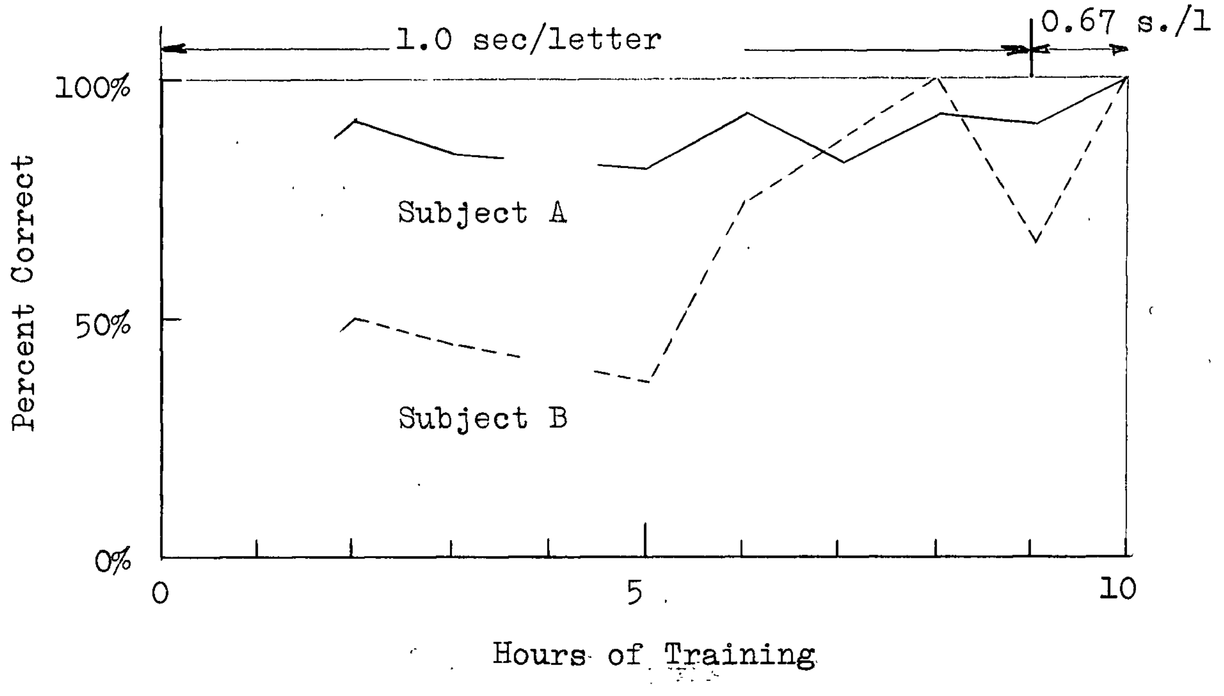

Figure 4-1. Letter-Reading Test Results

The Lexiphone scanning speed was initially set at one letter per second; that is, each single letter occupied one second in time. The rate at which successive letters were heard, on the other hand, was determined by the pause between letters. After eight hours training at a speed of one second per letter, both subjects could identify coded letters with almost no error (Fig. 4-1). The scanning speed was then increased to yield a duration of 0.67 seconds per letter. Within two hours both subjects managed to identify perfectly letters presented at this higher speed. Because of these encouraging results, the more realistic task of word-decoding was begun (Subsection 4.2.2).

Shown in Figure 4-2 are the letter confusion matrices

——— Response ———>

	a	e	s	c	r	t	h	g	y	-
a	88				6					6
e	4	85	7							4
s	7	21	72							
c			21	79						
r		4		7	85		4			
t						95				
h							100			
g		5	5				5	85		
y									95	5

Subject A

	a	e	s	c	r	t	h	g	y	-
a	74	6		14						6
e	4	85	11							
s	14	36	43	7						
c	5		21	63	11					
r	5	12		9	69					5
t						82	18			
h						19	81			
g								90	10	
y								12	88	

Subject B

<u>Figure 4-2</u>. Letter Confusion Matrices

for each blind subject, compiled from 100 responses made during the last four letter tests. The matrices indicate different confusion patterns for the two subjects. For example, while subject A sometimes gave the responses e, s and h for the stimulus letter g, subject B confused stimulus g only with y. (The reason for these differences is discussed shortly.) Matrix asymmetry indicates that confusions were often non-reciprocal. For instance, letter a was sometimes given as a response to stimulus letter s, but the reverse confusion never occurred. The main diagonals of the matrices summarize the relative difficulty of decoding each letter. Those letters with risers or descenders (t h g y) were most easily decoded, while s was the letter most readily confused.

4.2.2 Word-Reading Tests With the Blind Subjects

A vocabulary of 46 three-letter words (Appendix I), constructed from the nine-letter alphabet, provided the material used for word training and testing. Words of equal length were employed in these tests to exclude the possibility of identifying a word by its length.

During the training period of each session a code-voice-code technique was employed. The coded version of a word was presented followed by a five second pause; the experimenter then announced the word and, finally, the coded word was repeated. During instruction successive words were often chosen to illustrate certain important code differences or similarities. A twenty-word test was given at the end of each training session, and the results of these tests are illustrated in Figure 4-3.

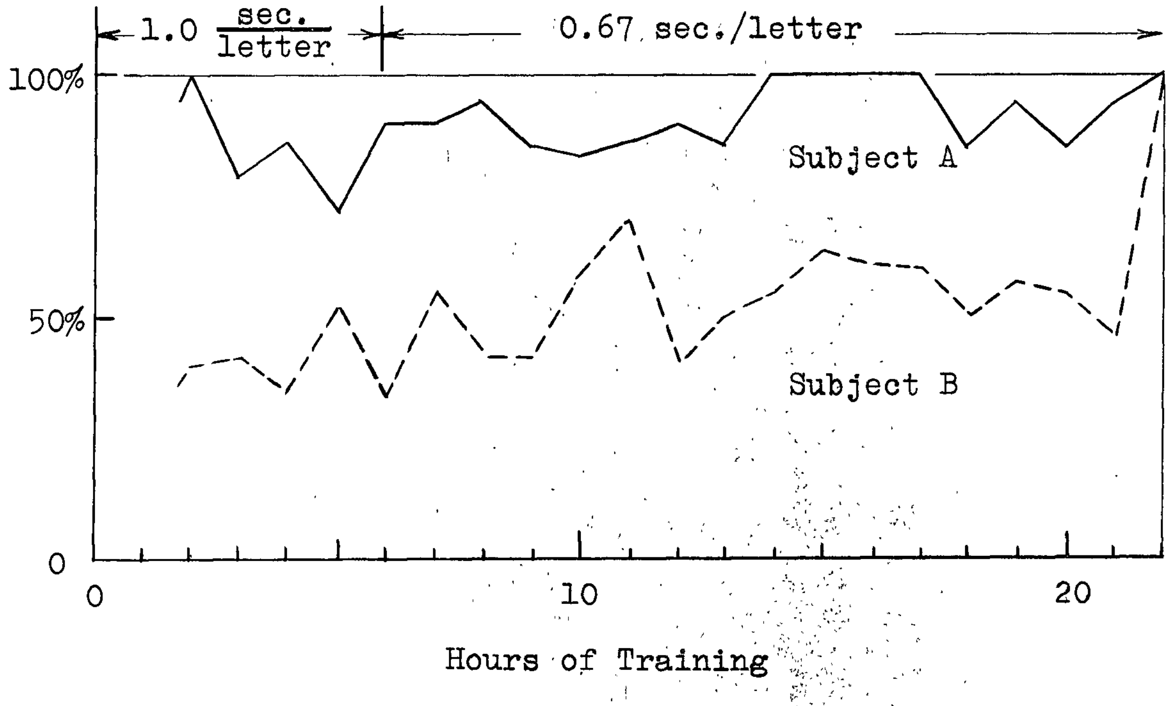

Figure 4-3. Word-Reading Test Results (Three-letter words)

Similar to the procedure employed in the letter-reading experiments, the subjects learned to decode individual words presented at two scanning speeds. For the first five sessions the presentation speed was one letter per second, corresponding to a duration of three seconds per word. For the remaining seventeen sessions the higher speed of 0.67 seconds per letter was used.

For the first seven word-training sessions, as for all letter-training sessions, the Lexiphone was utilized to produce the coded material directly from typewritten sheets. However, the remaining fifteen word-training sessions were conducted with tape-recorded material in order to ensure code uniformity, to exclude operating noises of the Lexiphone, and generally to expedite the sessions.

4.2.3 Comments on the Blind Reading Experiments and Further Tests

The consistently superior performance of Subject A in these Lexiphone code experiments is obvious from the curves of Figures 4-1 and 4-3. Her facility with the code is attributed to her exceptional musical ability. She experienced no difficulty in recalling the exact pattern of pitch variation corresponding to each letter. This particular ability is important where two coded letters (for example, a and s) are characterized principally in the frequency dimension, and to distinguish between them requires detecting a subtle difference in the order of frequencies. Subject A stated that she relied heavily on the dimension of frequency and rarely upon that of hiss or click. This fact explains why she sometimes confused the letter g, whose descender triggers the click, with letters not having descenders but whose musical patterns are somewhat similar to the pattern generated by g.

Subject B, on the contrary, depended almost entirely on the dimensions of hiss and click. The variations of pitch not reinforced by a change in timbre seemed to escape her entirely. A short test to determine the extent of her pitch perception problem was given, and it showed that she was unable to distinguish between various patterns of two notes unless the notes differed in frequency by nearly an octave. Apart from this partial tone-deafness, her hearing seemed normal. In fact, in another similar test with both subjects, this time using two bandwidths of hiss, Subject B made no errors in identifying different hiss patterns while A did. Subject B's clear perception of clicks and hiss accounts for the fact that, unlike subject A, she confused letter g only with y. Both letters have descenders and therefore both cause a click to be heard. She (B) always heard the click and hence knew the letter to be either g or y.

Another short and informal experiment was performed during the last few minutes of each of the final four sessions. Coded sentences consisting of two to eight words were presented to the subjects at an equivalent rate of 12 words per minute (1 letter per second). The component words, varying in length from one up to six letters each, were constructed from the nine-letter alphabet. The majority of these words were words never heard before in coded from by the subjects (Appendix II). Both subjects demonstrated an amazing ability to decode nearly every word upon its first presentation. Sometimes a word, usually a long word, had to be repeated. Also surprising is the fact the Subject B, whose word-reading ability averaged about 50% throughout the formal word-training sessions compared with A's 90%, usually responded more quickly to these new words than did A.

Although no quantitative results were obtained from the sentence-reading tests, the qualitative results are apparently encouraging, but may indicate no more than that new words of varying lengths, presented at one letter per second, can be easily decoded on a letter-by-letter basis. It was not possible, of course, for the subjects to decode each word as a single sound pattern or word-unit because for most of these words no previous training had been given. The subjects themselves could offer no specific reasons for their unusual performance during the sentence-reading tests.

4.2.4 Summary of Lexiphone Tests with the Blind Subjects

The results of our machine reading tests show that these two blind persons can indeed read Lexiphone-coded words from the nine-letter alphabet. Although one subject demonstrated superior ability with the code, both subjects seemed equally well motivated and interested throughout the period of training. The fact that Subject B, not having a musical ear, performed as well as she did (50%) helps justify the choice of a multidimensional code—its non-musical dimensions can provide useful information.

While the Lexiphone experiments were not designed to test reading speed as such, nevertheless we were interested in determining the effect of an increased scanning speed upon performance. At the two speeds employed, 1.0 and 0.67 seconds per letter, performance did not alter perceptibly.

This phase of the Lexiphone code experiments was concluded because the blind subjects were no longer available. Also it was desired at this point to procede with the comparative evaluation of the Lexiphone and optophone codes. Unless the Lexiphone code could be shown to be significantly superior to the optophone code, further Lexiphone training of the blind subjects would be of token interest only.

4.3 Lexiphone-Optophone Code Comparison with Sighted Subjects

Test performances of two comparable groups of sighted university students provided the data for the Lexiphone-optophone code comparison. Each group consisting of 26 students was trained and tested with each of the codes. Group 1 spent one hour learning the Lexiphone code, and Group 2, the optophone code; the groups then switched and worked with the second code during an additional hour one week later. This technique of reverse code presentation was intended to control the effect that the learning of the first code might have upon the learning of the second. Single letters of the restricted nine-letter alphabet were used in these tests, and were presented at a speed corresponding to 1.0 seconds per letter.

The following outlines the training and testing procedure employed in each of the hourly sessions. Each session was divided into two parts. In that part given to training, the code-voice-code technique was used (Figure 4-4(a)). Each letter was sounded three times in code, then a seven-second pause was followed by the letter spoken once, then the coded letter was sounded twice again. During the pause, the subjects recorded their identification of the coded letter, as a means of reinforcing their learning of the code and of checking their own progress. Ten groups of nine letters each were presented in this way with each letter occurring randomly, yet appearing ten times in all.

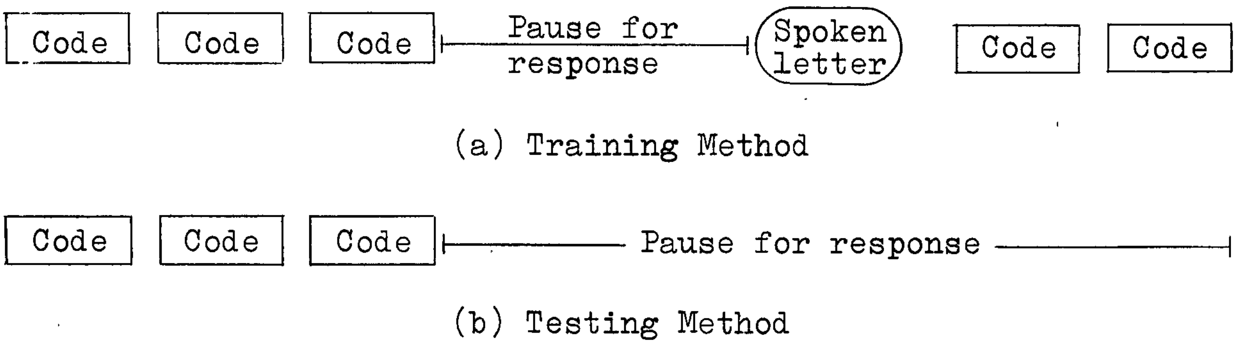

Figure 4-4. Training and Testing Presentation Used for Each Letter

In the second part of the session, the students' code learning abilities were tested with 27 coded letters presented in random order with each of the nine different letters appearing equiprobably. Each letter was sounded three times only, without a voice clue, and during the pauses between letters the students recorded their responses (Figure 4-4(b)).

4.3.1 Test Results

Table 4-1 summarizes the results of the Lexiphone-optophone code tests. The tabulated scores shown are the average number of letters correctly identified by each group out of the total of 27 letters presented during each code test and the standard deviations serve to indicate how individual scores in a group varied.

Comparing the results of the first code tests, it is seen that group performances differed somewhat: Group 1 working with the Lexiphone code achieved slightly better results (mean = 10.51), compared with Group 2 working with the optophone

	Lexiphone		Optophone	
	Avg. Score	S.D.	Avg. Score	S.D.
Group 1 (Lex.-Opt.)	10.51	4.40	11.69	3.04
	(N = 26)		(N = 26)	
Group 2 (Opt.-Lex.)	11.50	4.47	8.57	3.55
	(N = 26)		(N = 26)	
Code Average	11.01	(N=52)	10.12	(N=52)

Maximum possible score = 27.

Table 4-1. Group Mean Scores and Standard Deviations of Code Tests

code (mean=8.57). The statistical significance of the difference between these means (D=1.94) was examined by applying a single-tail t-test. The value of t computed (2.01) indicates the difference in performance is just significant (at the α=.05 level, df=25).

To include the results of the second code tests and to evaluate other variables, further examination of the results was carried out with a special t-test[27] that utilizes the mean difference in performance experienced by individual subjects between their first and second code tests. This t-test offers a measure of the three important variables operating in this series of code experiments: the code effect, practice effect, and the effect of interaction between code and test-order. Table 4-2 summarizes the results derived from this t-test (single-tail).

Effect	Mean Difference	t	P (df=50)	Significance (α=.05)
1. Code effect	Lex.-Opt.= 1.019	1.66	>.05	Not Signif.
2. Test-order effect (Practice effect)	1st-2nd= -1.904	-3.11	<.005	Very Signif.
3. Interaction (Code x test-order effect)	(L-O)-(O-L)= -3.810	-3.58	<.001	Very Signif.

Table 4-2. Significance of Lexiphone-Optophone Code Test Variables

The positive value of the mean difference in line 1 of the Table indicates that performance with the Lexiphone code, averaged over both groups, was slightly superior to that with the optophone code. Unfortunately, the corresponding value of t reveals that this superiority is so slight as to be statistically insignificant ($P > .05$). Lines 2 and 3 of the Table show two secondary effects which are quite significant. The t-value calculated for the test-order effect indicates that, regardless of which code was learned first, experience acquired with the first code significantly improved performance with the second. The codes were similar enough in character that practice with the first code could successfully be applied to the second. The third calculation, which measures the interaction between code performance and code order, shows that greater improvement in performance occurred between the first and second tests when the optophone code preceded the Lexiphone code.

Confusion matrices, constructed from the 1,404 pooled test responses for each code, are shown in Figure 4-5. Confusions are widely dispersed in both matrices, demonstrating that the students had not received enough training to enable them to narrow their confusion field. Note, however, that the optophone matrix may be partitioned to yield two high-confusion submatrices, a e s c and r t h g y. This observation (supported by observations of the students' progress during their training period) may indicate that optophone code learning had reached

——— R e s p o n s e ———→

Optophone Code

	a	e	s	c	r	t	h	g	y	-
a	30	14	18	13	4	3	8	4		4
e	17	26	17	26	3	1	6	3		2
s	17	19	24	29	3	1	3	1	1	1
c	13	21	20	34	4	3	1	2		2
r	6	5	5	5	30	18	8	7	10	4
t	1	1	2	1	10	49	12	3	20	1
h	6	3	4	1	10	17	31	15	7	4
g	4	4	3	3	6	7	13	51	4	3
y		1			8	15	12	4	60	1

Lexiphone Code

	a	e	s	c	r	t	h	g	y	-
a	28	8	13	1	11	8	13	8	7	4
e	8	29	23	12	5	11	3	8	1	
s	8	9	40	31	5		1	3	1	1
c	13	8	20	36	7	3	3	4	2	4
r	18	5	6	8	38	6	7	6	3	3
t	3	4	3		6	46	17	3	18	1
h	8	7	5	2	6	10	42	9	7	4
g	10	7	6	3	10	6	10	42	2	3
y	2	1			3	13	2	3	74	1

Figure 4-5. Letter Confusion Matrices (Groups pooled)

some kind of plateau even at the end of one hour of training. The optophone submatrices suggest that the students had learned to group coded letters into two classes, but they then found, perhaps, that the optophone code offered no further clue with which they could accurately identify individual letters. Or, perhaps more likely, after listening to the optophone code continuously for one hour, the students became bored with the "monotonous" code (according to their comments) and therefore made no effort to identify letters more accurately. After a period of rest, it is possible that performance with the optophone code would have seen some improvement.

The partitioning phenomenon is certainly not evident in the Lexiphone matrix; confusions are more uniformly distributed than in the optophone matrix. This fact suggests that the students had learned to classify each Lexiphone-coded letter into its own category and not just into one of two or three multi-letter categories. The main diagonal values show that the average number of Lexiphone confusions were less than the number of optophone confusions. No evidence of Lexiphone code saturation is exhibited in these results, and hence, we conclude that a second Lexiphone code training session would have improved performance quite markedly. The remarks of the following paragraph also bear out the conjecture that the students could have experienced more substantial gains in performance with the Lexiphone code than with the optophone code given further training time.

At the conclusion of the code comparison experiment the students were asked to record their opinion of the two codes. The majority preferred the Lexiphone code, saying that the optophone code quickly became monotonous, or that all optophone-coded letters came to have an annoying "sameness." One student stated specifically that the hiss and click dimensions of the Lexiphone code provided the only distinguishing clues for him. He admitted not having a musical ear. Although the remainder of students preferring the Lexiphone code were not so specific, they explained their preference by saying that the Lexiphone code components were more "variable." At the same time, some of them complained that the "noise" (hiss) present in the Lexiphone code tended to obscure the "real variation," thus revealing their inability to appreciate the hiss dimension. With further training this unpleasant dimension likely could have become a subjectively worthwhile variable to those disturbed by it in the first training session.

4.3.2 Summary of the Lexiphone-Optophone Code Comparison

All of the results obtained from this Lexiphone-optophone comparison experiment suggest that the Lexiphone code is the superior code. Unfortunately, we have not been able to show that the difference in performance between the two codes is statistically significant, but this failing is attributed to the short period of code training involved. The difficulty of correctly interpreting the short-term training results is outlined in the following paragraph.

Both groups performed better on their second code test, despite the fact the second code differed from the first. Had the same code been presented in both training sessions, the improvement in performance certainly would have been at least as great. Hence, the significant improvement that was noticed demonstrates that at the end of the first hour of training the code learning process was not yet complete; the students' ability to improve their decoding performance had not yet reached a saturation point. (Average score was 40%.) It is therefore not valid to extrapolate the results of these short one-hour training sessions as truly representing the results that would be obtained after extended code training. As an illustration of the danger of such extrapolation, the initial learning rate with code A might be greater than that with code B, but after sufficient training the learning curves could level off with code B in the superior position. Although desirable, it was not practical to carry out a more decisive code comparison experiment involving several hours training with each code.

5. STATISTICAL STUDY OF LEXIPHONE-CODED LETTERS

5.1 Introduction

This chapter describes a study carried out to determine the exact print information signals produced by the Lexiphone, and to ascertain whether letter characteristics derived from these signals lead to a simple scheme for letter or letter-feature recognition. If simple feature-detection circuits based on these characteristics could economically be added to the Lexiphone, then the sound code generated by the machine could be made less complicated and consequently more efficient.

It was implicit in this study that the signals be obtained from the existing six-cell print scanner. The idea was merely to establish the possibility of incorporating simple improvements in the present machine, not to suggest the design of a new print scanning method. Also it was intended to use the statistics compiled from the signals of the present scanner to study the quantity of information generated by a simple direct-translating machine of the Lexiphone type. This informational study is the subject of Chapter 6.

5.2 Quantizing and Recording the Print Signals

As print passes beneath the Lexiphone scanner (reading tube), six binary signals are generated by six relays connected to the photocells. At any instant, each of the six relays is in one of two conditions: ON (binary state "1"), or OFF (binary state "0"). We call the instantaneous condition of all six relays the relay "state." Physically, each state registers the instantaneous distribution of ink detected in a narrow vertical sample of that letter, and the entire letter is then described by the succession of states generated during its scanning.

There are 64 possible states, each of which can be characterized by a six-digit binary number denoting the binary position of the relays. The zero state, all relays OFF, occurs when no print is detected. Any non-zero state signifies the presence of print. If the six binary digits of a state are considered to form a column vector, and the successive column vectors describing the letter are placed next to each other, a binary array or matrix results, and it constitutes a quantized facsimile of the letter (Figure 5-1).

```
Relay
 No.

  1     ....................     ......111111111.....
  2     ...1111111..........     111111111111........
  3     .1111...1111........     ....1111............
  4     11111...1111........     ....111.............
  5     .111111111111111....     111111111111........
  6     ....................     ....................
```

Binary zeros are represented by dots.

<u>Figure 5-1.</u> Quantized Facsimiles of the Letters "a" and "f"

In this study, the 52 letters of the standard Pica typewriter font were used. Ten samples of each were typed on a sheet with an electric typewriter in preparation for Lexiphone scanning. Because the horizontal letter-spacing in this font is constant (0.1 inches per letter) independent of the letter typed, we define this length (0.1 inches) to be the "letter-space." The contours of a letter, however, actually span just a proportion of the letter-space, and this proportion defines the "length" of a letter (Fig. 5-3).

An Alwac III-E digital computer performed the state sampling and storing functions. Because the input unit of the Alwac conveniently reads six-bit characters, the six Lexiphone relay contacts were connected directly to it, thus allowing the computer to sample the instantaneous relay state with a single READ instruction.

It was determined experimentally that twenty equally-spaced samples per letter-space were sufficient to resolve the shortest state durations occuring during the scanning of a letter. The sampling process was therfore carried out as follows. A convenient Lexiphone scanning speed of one letter-space per second (0.1 in./sec.) was chosen, and the computer was programmed to sample the relay state every 50 milliseconds once the presence of print was detected. The white space before each letter generated the zero state. When this state was encountered by the computer, it cycled back to the READ instruction every 4 milliseconds. The first non-zero state read-in constituted the first sample of a letter, and it initiated the 50 millisecond sampling cycle for the remaining nineteen equi-spaced samples of that letter. Each of the 520 letters was sampled in this way and stored.

This sampling method, in which the samples refer to print information detected at specific locations of a letter left-registered in the letter-space (Fig. 5-1), seemed to be the simplest method that might be instrumented in connection with the Lexiphone. The samples are simply and reliably synchronized to each letter because the sampling process is triggered by the left-hand edge of a letter. In the actual machine sampler, the sampling cycle rate would be tied directly to the scanning speed and this would ensure the correct spacing of samples.

For the purposes of displaying the quantized letters in an abbreviated form, the 64 six-bit states are most easily represented by means of hexadecimal numbers (0,1,...,9,a,b,...,f). If position 6 in Figure 5-1 is considered the least significant bit, and position 1 the most significant, then each state can be described by a two-digit hexadecimal number between 00 and 3f (corresponding to the decimal equivalents 0 and 63). The binary matrix shown in Figure 5-1 for a quantized letter can then be represented as a row matrix of twenty elements, as indicated in Figure 5-2. The binary matrix is spread out in Figure 5-2 to allow for the clear designation of each state.

The row matrix of Figure 5-2 can be considered a one-dimensional representation of the two dimensional spatial information contained in the printed letter "a". The row matrix is therefore a one-dimensional code for that letter, and the decimal values of the matrix can be plotted on a scale, as in Figure 5-3, to yield a graphical display. Such a display illustrates the signal variation which would be encountered with a unidimensional sound code where there exists a one-to-one correspondence between the 64 state and code values.

-- STATES --

B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
I	.	.	.	1	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	.
N	.	1	1	1	1	.	.	.	.	1	1	1	1	.	.	.	.	.	.	.
A	1	1	1	1	1	.	.	.	.	1	1	1	1	.	.	.	.	.	.	.
R	.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.
Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HEX.	04	0e	0e	1e	1e	12	12	12	12	1e	0e	0e	0e	02	02	02	00	00	00	00

Figure 5-2. Binary Matrix and Hexadecimal Row Matrix Representing the Quantized letter "a"

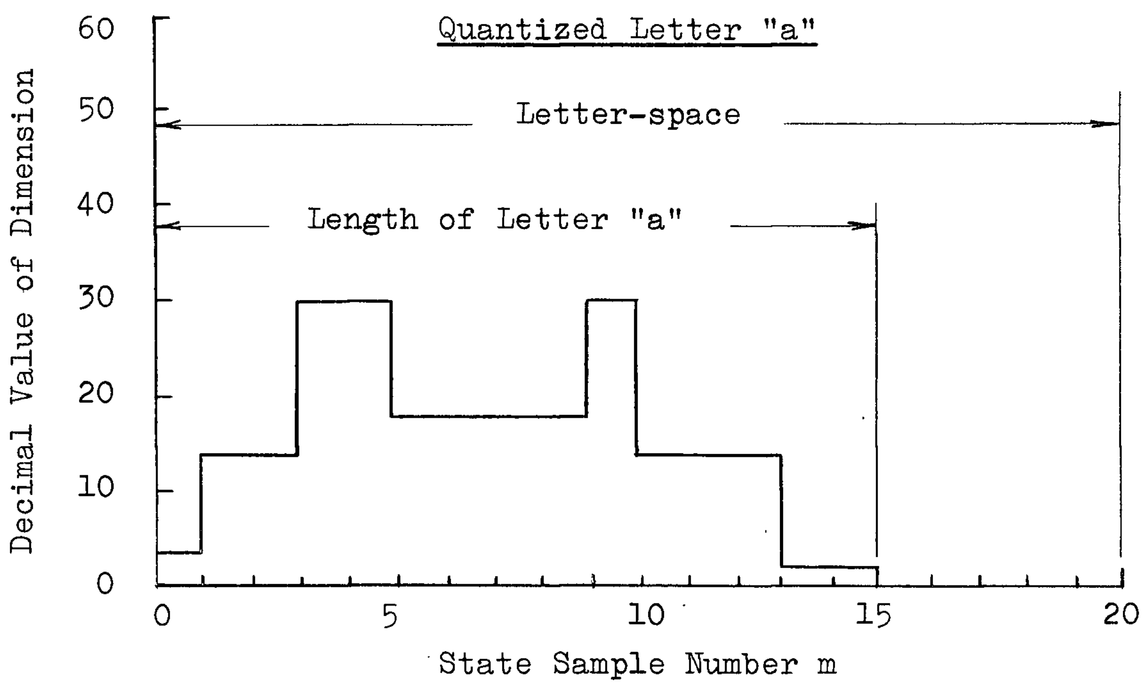

Figure 5-3. One-dimensional Representation of Quantized Letter On a Time or Space Scale

5.3 Study of Correlation Between Pairs of Cells

A preliminary study was carried out on the quantized print data to establish the possibility of optimizing the Lexiphone code with respect to photocells 2,3,4, and 5. These cells are activated by all letters and therefore generate the bulk (two-thirds) of the information produced by the print scanner. Because these cells are located symmetrically with respect to the lower-case body (the vertical extent of lower-case letter x), it seemed possible that some of the symmetry evident in letters such as d,b,p,q,g,e,c,o, and x might cause cells 2 to 5 to duplicate each other to some extent. If the coding action of two of the cells were highly correlated, then one of them could be ignored and the sound code made consequently simpler, or the less well correlated cells could be made to control the more effective dimensions of the sound code.

Pairs of cells expected to exhibit a high degree of correlation were selected, and from their binary signals were computed correlation coefficients r.

$$r = \frac{\overline{(x_1 - \overline{x_1})(x_2 - \overline{x_2})}}{s_1 \, s_2}$$

where x_i is the binary signal of the i-th cell considered, and s_i is the standard deviation of the cell signal. The correlation was also measured between two groups of two cells each, in which case x_i refers to the combined signal of the i-th pair of cells.

In Table 5-1 are summarized the correlation results.

Between Cells No.	2&5	2&3	3&4	4&5	2&3 and 4&5	3&4 and 2&5
r	-0.41	-0.36	-0.28	-0.46	+0.30	+0.24

Table 5-1. Correlation Coefficients Computed From the Quantized Photocell Signals

The results indicate that all pairs of cells are correlated to about the same extent and that the degree of correlation is not very great. Each cell appears to contribute an approximately equal proportion of the coded information generated by cells 2 to 5. No special assignment of audible code dimensions is therefore indicated.

5.4 Characteristics of the Quantized Letters

A statistical study was carried out on the quantized letter data to test for the presence of certain characteristics. The characteristics evaluated were those that

1. might lead to a letter recognition scheme, and
2. could be measured by a simply-instrumented device connected to the Lexiphone scanner. Some additional statistics were calculated for the whole field of data, independent of particular letters, which indicate the average distribution of print information in the letter space.

The statistical calculations tabulated in these sections were performed by an IBM 7040 digital computer. Punched cards containing the quantized Lexiphone data were produced for the 7040 by first converting the normal six-bit punched paper tape output of the Alwac to corresponding tapes coded in the five-bit telegraphic code. These latter tapes were then converted to cards by an IBM Type 47 Tape-to-Card Converter.

We first introduce terminology to be used in this and the following sections. The instantaneous relay state, denoted by s_i, may be any one of the 64 possible states ($i = 0,1,2,\ldots,63$). The j-th character of the 52-letter alphabet is denoted by Y_j ($j = 1,2,\ldots,52$). Every Y_j is described in quantized form by twenty particular states s_i occuring in the twenty successive locations m ($m = 1,2,\ldots,20$) of the letter-space. (See Figure 5-3.) We refer to location m as column number m; for if the 520 row matrices of the ensemble of letters (10 examples of each letter) are placed one beneath the other to form an ensemble array (520 x 20), all locations numbered m fall in a column. The rows of the binary matrix for each letter are numbered n ($n = 1, 2,\ldots,6$), corresponding to the photocell and relay number. $C_i(k_i)$ refers to the i-th characteristic of the data, and depends on the parameter k_i ($k_i = 1,2,\ldots,K_i$). When the probability of Y_j is written it is assumed the probability is evaluated over all ten examples of that letter.

5.4.1. Letter-Independent Characteristics

Some trends of the whole field of data were determined first without regard to the letters generating the data. The trends evaluated were: the average state density, $P(s_i)$; the preference of state s_i for column m, $P(m|s_i)$, and vice versa, $P(s_i|m)$; the probability $P(C_1(k_1)|s_i)$ that state s_i occurs exactly k_1 times per letter; the probability $P(C_2(k_2))$ that the average letter length is k_2 columns long; the probability $P(C_3(k_3)|n)$ that k_3 "1" states occur in row n of the binary matrix (in other words, that the integrated "black" information detected by photocell n per letter amounts to k_3); and finally, the probability $P(C_4(k_4)|n)$ that there are k_4 intersections per letter in row n of the binary matrix (that is, photocell n on the average, intersects k_4 distinct parts of a letter).

The following trends are evident from the results. Forty-three of the non-zero states occur, of which three in particular occur most frequently by quite a margin: state 22_{hex} (cells 1&5), state 12_{hex}(cells 2&5), and state $3e_{hex}$(cells 1,2,3,4,&5). Some states show a preference for certain groups of columns m, but for the most part are widely distributed across the letter-space. Averaging characteristic $C_1(k_1)$ over all states shows that states occur once, twice, three times, and four times per letter with about the same frequency, although a small peak at $k_1 = 3$ is evident. Characteristic $C_2(k_2)$ shows that the average length of a letter lies in the range $k_2 = 11$ to 20 columns, with a definite average peak at $k_2 = 18$. No particular patterning of the integrated information given by $C_3(k_3)$ is evident, except for the obvious fact that cells 1, 2 and 5 encounter a greater amount of information than do the other cells. The results obtained from $C_4(k_4)$ show that a cell encounters no more than three intersections per letter, and that all cells encounter single and double intersections with equal frequency.

5.4.2 Letter-Dependent Characteristics

The power of the above four characteristics $C_i(k_i)$ to discriminate between letters was determined by computing the conditional probabilities $P(Y_j|C_1(k_i), s_i)$, $P(Y_j|C_2(k_2))$, $P(Y_j|C_3(k_3), n)$, and $P(Y_j|C_4(k_4), n)$. Based on these statistics, the grouping of letters according to their characteristics is certainly possible if small deviations between samples of the same letter are ignored.

Characteristic C_2, the letter length, offers classification of the alphabet into two groups only because of the inconsistency and similarity of letter lengths: lengths 11 to 16 - ijsclgIeftozarpSCZgnOTHGQ; lengths 17 to 20 - nOTHGQdbuvxLBhmERFJPYDkywAKMNUWX.

Characteristic C_3, the integrated information per letter per cell, leads to nineteen exclusive groups of letters. The groups are determined by first representing the integrated information per cell with the digits 0, 1, or 2 (no units of information, 1 to 10 units, or 11 to 20 units), and second, collecting together all letters characterized by the same set of six digits. The nineteen groups (some consisting of just single letters) are: ESTKLIPYZDX, srzmnux, RFHGBN, CJUV, gpy, fli, khd, eco, tb, a, j, q, v, w, A, M, O, Q, W.

Characteristic C_4, the number of intersections per cell, when summarized over all cells as for C_3, leads to twenty-five exclusive groups of letters: escrv, fliIY, EODZB, tCLJ, hdbG, aoz, kKR, gy, pq, FT, PY, MN, u, j, m, n, w, x, A, H, Q, U, V, W, X.

The classification of letters according to C_1 is unreliable; most of the groups resulting from such classification overlap.

A discrete partial letter recognition scheme based on characteristics C_2, C_3, and C_4 would be simple to instrument: an integrator and comparator for C_2, six sets of integrators and comparators for C_3, six sets of two-bit counters for C_4, a sampling clock pulse, and necessary AND and OR gates to make the discrete letter-grouping decisions. Those letters not classified uniquely could be coded conventionally (modified according to the classification group into which they fall), while letters identified individually could be coded with a single characteristic sound.

The drawback with such a discrete GO, NO-GO recognition process is that imperfect letters are readily rejected, or classified in the wrong group. A more sophisticated identification scheme based on "maximum-likelihood decoding" circumvents this imperfection problem to some extent. In this scheme, J summing junctions are established corresponding to the J letters Y_j to be identified. The K_i outputs of each characteristic calculator $C_i(k_i)$ ($k_i = 1,2,\ldots,K_i$) are individually connected to each Y_j junction through a weighting attenuator whose attenuation value depends on the statistical importance of $C_i(k_i)$ to the identification of Y_j. After an unknown letter is scanned, the Y_j junction with the highest signal is selected. If its value is above a certain threshold and is sufficiently greater than the value of its nearest competitor, that Y_j is the letter identified; if not, the unknown letter is rejected as being unidentifiable. The advantage of this scheme is that the effect of small imperfections in a letter do not appreciably reduce the possibility of it being correctly identified.

In the following section, we examine the value of the characteristics C_2, C_3, and C_4 in such a maximum-likelihood recognition process. A "goodness" measure calculated for each characteristic establishes its relative merit and allows the effectiveness of all characteristics operating together to be established.

5.5 Recognition Effectiveness of Characteristics

For each characteristic C_i is calculated a goodness measure G_i, a single non-negative number measuring the correlation between C_i and the Y_j's.

$$G_i = \sum_j \sum_{k_i} P(C_i(k_i), Y_j) \log \left[\frac{P(Y_j | C_i(k_i))}{P(Y_j)} \right] \quad \ldots (5\text{-}1)$$

In information theory terms, G_i is precisely the average information about the Y_j's given by C_i. Consequently, individual characteristics or groups of characteristics may be evaluated, for letter recognition purposes, by comparing individual G_i's or sums of G_i's, respectively. Lewis[28] points out that if the characteristics chosen are statistically independent, then a linear relationship exists between the percentage recognition attainable by a system employing the C_i's ($i = 1,2,\ldots,\ell$) and the corresponding sum of the G_i's. This means that a quick evaluation of proposed characteristics (and groups of characteristics) is possible without the problem of computing ℓ-th order conditional joint probabilities $P(Y_j|C_1,C_2,\ldots,C_\ell)$. He also suggests that if the argument of the log term of equation (5-1) is very near unity (that is, $C_i(k_i)$ by itself is not very effective in identifying Y_j), then the log term can accurately be represented by the first-term of its power series expansion, so that equation (5-1) becomes

$$G_i = \sum_j \sum_{k_i} P(C_i(k_i),Y_j)\left[\frac{P(Y_j|C_i(k_i))}{P(Y_j)} - 1\right] \ldots (5\text{-}2)$$

Utilizing the IBM 7040 computer, G_i values were computed by means of equation (5-2) for $C_2(k_2)$, $C_{3(n)}(k_3)$, and $C_{4(n)}(k_4)$, ($n = 1,2,\ldots,6$), where n is the row number of the binary matrices. The values are listed in Table 5-2. It is seen from the G_i values that characteristic C_3 (integrated "black" information) is about twice as effective as characteristic C_4 (number of intersections). This difference is actually overrated. The value k_3 of C_3 computed for a letter depends on the sensitivity of the Lexiphone scanner which can change slightly from one day to the next. C_3 is therefore not a reliable characteristic. Characteristic C_2 (letter length) is unreliable for the same reason that C_3 is unreliable, although not to the same extent. C_4, however, is quite reliable; the number of intersections counted for a letter is reasonably independent of the machine setting.

							$G_i = \sum_n G_{i(n)}$
G_2	1.63						1.63
n	1	2	3	4	5	6	7.67
$G_{3(n)}$	1.84	1.60	1.15	1.02	1.56	0.50	
n	1	2	3	4	5	6	4.35
$G_{4(n)}$	0.94	0.78	0.82	0.73	0.64	0.44	
Sum of all G_i							13.65

Table 5-2 Goodness Measures G_i Computed for the Characteristic C_i

To relate G_i to the percentage recognition P_i possible employing a single characteristic C_i, equation (5-3) was evaluated for each $C_{i(n)}$.

$$P_{i(n)} = \sum_{k_i} P\left[Y_j^* \middle| C_{i(n)}(k_i)\right] \quad P\left[C_{i(n)}(k_i)\right] \qquad \ldots(5\text{-}3)$$

where Y_j^* is that letter for which $P\left[Y_j \middle| C_{i(n)}(k_i)\right]$ is a maximum. Plotting the $P_{i(n)}$ against the corresponding $G_{i(n)}$ indicates that 10% recognition per unit of G_i is obtained for characteristics C_2 and C_3, while C_4 offers about 5% recognition per unit G_i. If we take the 5% value to be the more realistic, then the total percentage recognition possible with a system using characteristics C_2, C_3, and C_4 together, assuming the characteristics to be independent and identification errors to be equally weighted, is (13.65) x (0.05) = 68%.

Although time did not allow further investigation of the characteristics, it is possible to determine whether or not characteristics are independent. To do this, one simulates a maximum-likelihood recognition system and measures the percentage recognition obtained for a given set of characteristics. In each of several successive trials another characteristic is added to the previous set. For each trial, the sum of the G_i's corresponding to the C_i's employed is plotted against the percentage recognition. If the resulting curve formed by the series of trial points is linear, the characteristics chosen are statistically independent. If a dependent characteristic is added, the curve exhibits saturation. Of course, eventually, the curve must saturate at 100% recognition.

The cursory examination we have given the above characteristics suggests that they by themselves cannot yield a perfect letter-recognition system. (Lewis required thirteen independent characteristics to achieve 82% recognition of fifteen alphabets.) Perhaps the introduction of registration-dependent characteristics (dependent on m) would improve recognition performance. Nevertheless, as demonstrated in Section 5.4, these simple characteristics can be used to identify some letters uniquely or, at the least, to classify them into groups. By delaying the audible coding of a letter, this classification information could be used to make the output coding more efficient. The conventional sound-coded version of that letter could be modified (or simplified) according to the group into which it is classified.

6. MACHINE READING RATES - INFORMATIONAL AND PSYCHOLOGICAL CONSIDERATIONS

6.1 Introduction

The assimilation of printed language presents no problem to normal sighted readers. Typically, they can manage to read 150 to 400 words per minute. Yet when the same print is translated by a simple reading machine and presented to blind persons in a different sensory form, only 30 words per minute or less can be handled. This comparison indicates that the print recoding process performed by the machine must introduce certain speed limiting factors not normally present or operative in the direct visual process. While the existence of these factors is self-evident, it is not obvious how many there are nor upon what they depend. The choice of a particular audible code, the factor with which we have primarily been concerned up to this point, may be only one of several, more significant factors. Naturally, the reading machine problem would be greatly clarified if the identity of these limiting factors were established.

In this chapter we speculate upon the effect of several informational and psychological variables, some or all of which may be limiting factors responsible for low machine reading rates. Selective and psychological order source entropies are calculated for Lexiphone-coded print, taking Lexiphone characteristics as typical of simple translating devices. When the total psychological entropy calculated is related to the human channel capacity, maximum machine reading rates similar to recorded experimental results are obtained.

6.2 Lexiphone Source Entropy and Coding Redundancy

Consider the Lexiphone to represent an information source capable of transmitting any one of a set of n source events x_i (i = 1,2,...,n). If event x_i occurs with probability p_i, then the average entropy H associated with the x_i is given by[29]

$$H(x) = -\sum_{i=1}^{n} p_i \log_2 p_i \quad \text{(bits/event)} \qquad \ldots(6\text{-}1)$$

where H(x) denotes specifically source entropy. If the x_i are equiprobable, H(x) attains a maximum value and equation (6-1) simplifies to

$$H(x) = \log_2 n$$

Let us first calculate the maximum amount of coded information per letter that could be transmitted by the Lexiphone in the optimal coding situation. Maximum H occurs only when the 63 states generated by the machine are used equiprobably in the coding of letters. If this is done, then $H(x)_{max} = \log_2 63 = 5.98$ bits per state. From the statistical data collected for Chapter 5, it has been calculated that a letter scanned by the Lexiphone generates, on the average, 8.4

distinct states. This leads to a maximum possible source entropy of

$$H(x)_{max} = 5.98 \times 8.4 = 50.3 \text{ bits/letter}$$

Now let us compare this maximum with the source entropy actually transmitted by the Lexiphone (or for that matter, by any reading machine that presents no more than one letter per stimulus). The source material from which the reading machine entropy arises is simply the set of letters making up the 52- (or 26-) letter alphabet. In the case of a simple reading machine there are generated 52 different stimulus patterns corresponding to the 52 input letters. If the 52 letters are assumed to occur equiprobably and independently, the source entropy generated by the machine must be just the zeroth-order approximation to the entropy of printed English, H_0.

$$H(x) = H_0 = \log_2 52 = 5.7 \text{ bits/letter}$$

Of course print "noise" can cause the actual machine source entropy to rise above H_0 by causing the machine to generate more than one stimulus pattern per letter, but under normal circumstances this rise is not serious.

The values of actual and maximum Lexiphone source entropy are now compared by calculating the coding redundancy R

$$R = 1 - \frac{H(x)}{H(x)_{max}}$$

Hence, the Lexiphone coding redundancy is

$$R_{Lex} = 1 - \frac{5.7}{50.3} = 89\%$$

Shannon[29] points out that printed English is about 50% redundant, and that much of this redundancy is useful in reducing error. (In a non-redundant language, every misprinted letter gives rise to a new word that is perfectly meaningful in the sentence, but a sentence whose meaning is different from that intended.) With the Lexiphone code it is a question of whether or not all of the redundancy is useful. If the Lexiphone user is able to appreciate and use to his advantage the patterning and interdependence of code sounds, then the code redundancy is useful. If he cannot appreciate (or ignore) the redundant parts of the code, the redundancy is not useful, and he experiences a consequent increase in effective source entropy. The result of such an increase is to reduce his maximum reading rate, as is demonstrated in Section 6.3.

We have not included alphabetic or contextual redundancy in the above calculation of Lexiphone source entropy because this type of redundancy is effective regardless of what reading machine code is used; the machine merely translates the alphabetic information it receives. Unequal letter probabilities evident in printed English reduce the average entropy per letter; for a 26-letter alphabet, the entropy is reduced from $H_0 = \log_2 26 = 4.7$ bits per letter to $H_1 = 4.1$ bits per letter.[30] Further reduction occurs when higher-order redundancies (digram, trigram,.., and word frequencies) are taken into account.[31, 32]

The redundancy we have calculated for the Lexiphone code is over and above that language redundancy present in the source material.

6.3 Psychological Source Entropy

If the discrete sounds of the Lexiphone code comprising a letter cannot be perceived by the human subject as a single image or Gestalt, then a psychological limitation suggested by Crossman(33) may tend to increase the effective source entropy. The limitation deals with the cost of confusing, on recall, the exact serial order in which the code sounds were perceived. It is not unlikely that this limitation is effective here because many Lexiphone-coded letters are differentiated only by a subtle difference in the order of certain code sounds.

Consider a coded letter to be a message consisting of n distinct, sequentially ordered message-units (which are merely the n distinct states generated by the letter). After a message has been received by the human subject, he recalls the set of message-units from immediate memory, and to make a correct identification, he must recall exactly their original serial order. The cost of his preserving the serial order of n units in a message can be considered the psychological order source entropy, $H_o(x)$, given by

$$H_o(x) = \log_2 \left[\frac{n!}{n_1! \ldots n_i! \ldots n_r!} \right] \qquad \ldots(6\text{-}2)$$

where n! is the number of permutations of n message units, n_i is the number of times message-unit i appears in the message if i appears more than once, and r is the number of different units appearing more than once per message. The denominator of equation (6-2) is present because we consider that two identical units n_i may be transposed without any loss of order information. As we would expect, the cost of preserving serial order increases with the length of the message.(34)

Equation (6-2) was evaluated for each of our 520 letter-samples, and the average order source entropy calculated was

$$H_o(x) = 13.7 \text{ bits/letter}$$

(When the individual letter entropies are weighted according to the frequency of occurrence of letters in English, $H_o(x)$ is reduced to 12.0 bits per letter.) Crossman shows that the total psychological source entropy $H_t(x)$ is simply the sum of the original selective source entropy $H(x)$ and the order source entropy $H_o(x)$:

$$H_t(x) = H(x) + H_o(x) \qquad \ldots(6\text{-}3)$$

If we assume that all of the order source entropy is effective as far as a Lexiphone user is concerned, then, by inserting in equation (6-3) this value of $H_o(x)$ and that value of $H(x)$ determined in Section 6.1, we arrive at the following amount of total psychological source entropy:

$$H_t(x) = 5.7 + 13.7 = 19.4 \text{ bits/letter}$$

We are now in a position to calculate the maximum possible reading rate based on this estimate of source entropy.

6.4 Human Channel Capacity and Maximum Reading Rates

Peak information transfer rates $\dot{I}_{max}$ attained by human subjects performing well-practised sequential tasks are reported to be in the range 15 to 44 bits per second.(32, 35) The lowest figure is typical of tasks requiring motor-control, such as piano-playing and typing; intermediate values correspond to tasks such as casual conversation, oral reading, and expert mental arithmetic; while the upper limit is an estimate of the peak rate attained for the motorless task of silent reading.

If we suppose a person reading by means of the Lexiphone is so highly experienced with the machine that his errorless reading rate has reached a saturation speed of $\dot{L}_{max}$ letters per second, then it is reasonable to assume he is processing information at a rate $\dot{I}$ approaching his channel capacity C (bits/sec.) for this particular task. Let the channel capacity of our subject assume the generous value of $C = 40$ bits per second. We will assume that his equivocation rate $\dot{H}_y(x)$ is negligible; that is, that

$$C \triangleq \max\left[\dot{I}(x;y)\right] \triangleq \max\left[\dot{H}(x)-\dot{H}_y(x)\right] = \max\left[\dot{H}(x)\right] = \dot{H}(x)_{max} \quad \ldots(6\text{-}4)$$

where the dot denotes a time-rate quantity. His maximum reading rate is then given by

$$\dot{L}_{max} \text{ (letters/sec.)} = \frac{C}{H(x)} \frac{\text{(bits/sec.)}}{\text{(bits/letter)}} \qquad \ldots(6\text{-}5)$$

Substituting into equation (6-5) the value of total psychological source entropy calculated in the previous section in place of H(x), and the value of C = 40 bits per second, we obtain the following estimate for the maximum Lexiphone reading rate (assuming 4.5 letters per word):

<u>Lexiphone</u> $\dot{L}_{max} = \frac{40}{19.4} = 2.06 \frac{\text{letters}}{\text{sec.}} = 27.5 \frac{\text{words}}{\text{min.}}$

Another similar calculation is appropriate at this point. Consider a person who reads by means of spelled speech (say, the audible output from a recognition-type reading machine). The source entropy he encounters is just the selective entropy of one of 26 equiprobable letters, $\log_2 26 = 4.7$ bits per letter, considering that no psychological order entropy is created by a single letter sound. If we assume his channel capacity for this task is also C = 40 bits per second, then his maximum spelled-speech reading rate is, according to equation (6-5),

<u>Spelled Speech</u> $\dot{L}_{max} = \frac{40}{4.7} = 8.5 \frac{\text{letters}}{\text{sec.}} = 110 \frac{\text{words}}{\text{min.}}$

Although these maximum reading rates have been arrived at using rough informational approximations, the rates correspond quite closely to the experimental rates recorded for human performance with Lexiphone-type reading machines and spelled speech. How accurate are the approximations, or how valid is this informational approach is difficult to say, for the human information channel does not easily lend itself to informational analysis in absolute terms. We can, however, offer comment on the informational (and some of the psychological) parameters involved.

The maximum reading rate $\dot{L}_{max}$ given by equation (6-5) employs the ratio of channel capacity C to source entropy H(x). The value of C = 40 bits per second chosen for our hypothetical subject is certainly an upper limit under any condition. Decoding errors and lack of familiarity with the task will reduce C to a value well below 40 bits per second. The upper limit of $\dot{L}$ is therefore dependent only on the lower bound of H(x), or, in the case of the human channel, $H_t(x)$. So one asks, how can the components of $H_t(x) = H(x) + H_o(x)$ be reduced in magnitude psychologically?

The true statistical value of H(x) cannot of course be reduced below the average amount of information necessary to differentiate 52 possible letters, but a good proportion of the information in excess of this minimum (i.e. the excessive coding redundancy) can certainly be excluded with psychological advantage. To achieve minimum H(x), a character recognition machine is obviously indicated; but near-minimum H(x) can be achieved by transmitting only those states or patterns of states which are almost non-redundant. In other words, near-optimum coding (minimum psychological H(x)) can be realized only with some pre-processing of the print information.

The psychological order source entropy $H_o(x)$ depends on the number of distinct units per message (sounds per letter) that must be recalled serially by the human subject. $H_o(x)$ may be reduced in two ways. First, we can minimize the actual number of discrete sound units generated per letter—even to one sound per letter, which again necessitates a character recognition device. Or, second, we can associate with the signal units generated by a letter a coalescent audible code whose components combine in time to produce a single psychological sound unit. Of course it is just such a coalescent code for which reading machine workers have unconsciously, or otherwise, been searching—a code which gives rise to easily remembered audible patterns. Most of the simple reading machine codes previously studied have generated sensory code units direct from discrete print signals. So it would be worthwhile to investigate a multidimensional audible code that is controlled by analogue electrical signals derived from the print. Such a code, whose components would be smoothly varying and necessarily coalescent, should encourage the perception of a letter as no more than a single psychological unit.

7. SUMMARY AND CONCLUSIONS

An experimental direct-translation reading machine, based on the print scanning technique of the Optophone, has been built and its operation studied. This machine has made possible the evaluation of a four-dimensional audible code (the Lexiphone code), and the determination of print signals produced by a machine of this type.

Two blind subjects were trained for 36 hours to "read" both Lexiphone-coded letters and words selected from a limited alphabet of nine lower-case letters. Their reading performance was measured in terms of the percentage of letters or words correctly identified. While one subject performed consistently better, both managed to attain 100% at some point during the training period. It was demonstrated that blind persons can learn to "read" multidimensionally-encoded words, and more significantly, that they can do so by utilizing entirely different dimensions of the code. This latter fact alone is sufficient reason, in terms of wide appeal, to employ a multidimensional code in preference to a unidimensional code.

Also carried out was a comparative evaluation of the Lexiphone and optophone codes. Two comparable groups of sighted persons were trained with single coded letters of the two codes for one hour. The results of tests given at the end of the training hour showed performance with the Lexiphone code to be slightly superior. While this superiority was not statistically significant, other factors indicated that further code training would widen the performance gap in favour of the Lexiphone code. First, the results indicated that optophone performance had reached a partial saturation point even during the first training hour. Second, the majority of subjects concurred in the opinion that while the Lexiphone code was not as pleasant to listen to, it provided a better basis for letter discrimination.

The results of these audible code experiments lead to the conclusion that with a machine of the Lexiphone type a multidimensional code is to be preferred to a unidimensional code. The further question of whether or not multidimensionally-encoded stimuli can yield reading rates higher than those experienced with unidimensional stimuli of the optophone type can be settled only by carrying out an exhaustive code training program. The danger and inadequacy of interpreting short-term training results has been pointed out.

A statistical study of the quantized print signals was carried out to determine possible sound code simplifications, and to establish the redundancy and amount of information presented by the machine. It was shown that, with a small amount of logic circuitry, characteristics derived from the print signals could be used to classify letters into groups. If the encoding of letters into sound were delayed, then this classification information could be used to optimize the encoding of particular letters or groups of letters. The use of these characteristics in a maximum-likelihood letter recognition scheme was also investigated. Although it was demonstrated that 68% recognition could be achieved if the characteristics acted independently, this low figure and the fact that some of the characteristics are rather unreliable leads to the conclusion that this particular print scanning system is not well-suited to automatic letter recognition.

To facilitate the hypothesis of reasons why simple reading devices can offer only low reading speeds, the Lexiphone-human combination was treated as an information processing channel. Depending upon the validity of certain assumptions it was shown that higher reading rates could be obtained by reducing the psychological value of the two components of source entropy. The first component, selective source entropy, may be reduced by the decreasing coding redundancy, which in the case of the Lexiphone is about 90%. That proportion of the redundant information the subject is unable to use to his advantage or to ignore becomes useless information he must digest to the detriment of reading speed. The reduction of this component can be realized by first determining the redundant print signals and then processing the print information in such a way that these signals are deleted. The second component, psychological order source entropy, can be minimized by employing a sound code without discrete components—one which gives rise to just a single psychological sound pattern per letter. It was pointed out that both of these information components could together be reduced by utilizing a character recognition device.

APPENDIX I

THE THREE LETTER WORD VOCABULARY

sag	sat	say	set	she	shy	sty	car	cat	cry
act	art	ash	ass	ate	age	ace	aye	are	
eat	ear	erg	eye	egg	etc	rat	ray	rag	rye
her	hat	hay	hag	has	tee	tar	tea	the	tag
gas	gee	gat	get	gay	gag	yet	yea	yes	

APPENDIX II

SENTENCE LIST

1. her tears are rare
2. the races that he sees are easy
3. her saga rates cash
4. he has a scar that each eye sees
5. she says that he acts gay yet cagy
6. she hates her rash
7. she acts her age at teas
8. there are the three tarts he eats
9. yes she teases
10. the crate sags at the stress
11. teach her that cats eat grey rats

REFERENCES

1. Pollack, I. and Ficks, L. "Information of Elementary Multidimensional Auditory Displays," J. Acoust. Soc. Amer., 26, 155-158; 1954.

2. Beddoes, M.P., Belyea, E.S.W., and Gibson, W.C., "A Reading Machine for the Blind," Nature, 190, 874-875; 1961.

3. Fournier d'Albe, E.E., "The Type-Reading Optophone," Nature, 94, p. 4; 1914.

4. Fournier d'Albe, E.E., "The Optophone: An Instrument for Reading by Ear," Nature, 105, 295-296; 1920. Fournier d'Albe, E.E., The Moon Element, An Introduction to the Wonders of Selenium, D. Appleton and Co., New York; 1924.

5. Cooper, F.S., "Research on Reading Machines for the Blind," in P.A. Zahl (ed.), Blindness: Modern Approaches to the Unseen Environment, Princeton Univ. Press, Princeton, N.J., 512-543; 1950.

6. Zworykin, V.K., and Flory, L.E., "Reading Aid for the Blind," Electronics, 19, 84-87; August, 1946.

7. Zworykin, V.K., Flory, L.E., and Pike, W.S., "Letter Reading Machine," Electronics, 22, 80-86; June, 1949.

8. Cooper, F.S., and Zahl, P.A., "Research on Reading Machines for the Blind," Progress Report to the Committee on Sensory Devices of work done at the Haskins Laboratories, New York; June 30, 1947.

9. Clowes, M.B., Ellis, K., Parks, J.R., Rengger, R., Communication from the National Physical Laboratory, U.K.; February 22, 1961.

10. Abma, J.S., Laymon, R.S., Mahler, D.S., Mason, L.J., Rice, D.R., and others, "The Development and Evaluation of Aural Reading Devices for the Blind," Battelle Memorial Institute Research reports to the Veterans Administration; 1958, 1959, 1960, 1961, 1962, 1963.

11. Coffey, J.L., "The Development and Evaluation of the Battelle Aural Reading Device," Proc. Int. Cong. on Technology and Blindness, Vol. 1, A.F.B., New York, 343-360; 1963.

12. Mauch, H.A., "The Development of a Reading Machine for the Blind," Summary Report to the Veterans Admin. from Mauch Laboratories, Inc., Dayton, Ohio; June 30, 1964.

13. Freiberger, H., "Summary Report of Developments in the Veterans Administration Research Program on Aids for the Blind," Veterans Administration, New York; October, 1962.

14. Metfessel, M.F., "Experimental Studies of Human Factors in Perception and Learning of Spelled Speech," Proc. Int. Cong. on Technology and Blindness, Vol. 1, A.F.B., 305-308; 1963.

15. Metfessel, M.F., and Lovell, C., "Controlled Association in Learning the Auditory Code of Spelled Speech," paper read at the 72nd Annual Convention of the American Psychological Association, Los Angeles; September 7, 1964.

16. Nye, P.W., "An Investigation of Audio Outputs for a Reading Machine," Report from the National Physical Laboratory, U.K.; February, 1965.

17. Beddoes, M.P., "Possible Uses of a Printed Braille Reader with Spelled Speech Output," Proc. Int. Cong. on Technology and Blindness, Vol. 1, A.F.B., New York, 325-341; 1963.

18. Bliss, J.C., "Kinesthetic - Tactile Communications," IRE Trans. on Inf. Theory, IT-8, 92-98; 1962.

19. Donaldson, R.W., "Multimodality Sensory Communications," Ph.D. Thesis, M.I.T.; June, 1965.

20. ______, "Medical electronics for blind readers," Electronics, 38, 35-36; January 25, 1965.

21. Shrager, P.G., and Susskind, C., "Electronics and the Blind," Adv. in Electronics and Electron Physics, 20, 281-292; 1964.

22. Freiberger, H., and Murphy, E.F., "Reading Machines for the Blind," IRE Trans. on Human Factors in Electronics, HFE-2, 8-19; March, 1961.

23. Fischer, G.L., Pollack, D.K., Radack, and Stevens, (ed.), Optical Character Recognition, Spartan Books, Washington, D.C.; 1962.

24. Scharff, S.A., "Letter Scanning for Character Recognition," Proc. Int. Cong. on Technology and Blindness, Vol. I, A.F.B., New York, 227-244; 1963.

25. Clemens, J.K., "Optical Character Recognition for Reading Machine Applications," Ph.D. Thesis, M.I.T.; September, 1965.

26. Emmanuel, A.F., and Mauch, H.A., "The Development and Evaluation of a Personal Type Reading Machine for the Blind," Mauch Laboratories report to the Veterans Administration; September, 1958.

27. Scott, W.A., and Wertheimer, M., Introduction to Psychological Research, John Wiley and Sons, Inc., New York, 258-262; 1962.

28. Lewis II, P.M., "The Characteristic Selection Problem in Recognition Systems," IRE Trans. on Info. Theory, IT-8, 171-178; 1962.

29. Shannon, C.E., "A Mathematical Theory of Communication," Bell Syst. Tech. J., 27, 279-423; 1948.

30. Quastler, H., "A Primer on Information Theory," in H.P. Yockey (ed.), Symposium of Information Theory in Biology, Pergamon Press, New York, 3-49; 1956.

31. Shannon, C.E., "Prediction and Entropy of Printed English," Bell Syst. Tech. J., 30, 50-64; 1951.

32. Attneave, F., Applications of Information Theory to Psychology, Holt, Rinehart, and Wilson, New York; 1959.

33. Crossman, E.R.F.W., "Information and Serial Order in Human Immediate Memory," in C. Cherry (ed), Information Theory, Fourth London Symposium, Butterworths, London, 147-161; 1960.

34. Murdock, B.B., "The Immediate Retention of Unrelated Words," J. Exp. Psychol., 60, 222-234; 1960.

35. Pierce, J.R., and Karlin, H.E., "Reading Rates and Information Rate of a Human Channel," Bell Syst. Tech. J., 36, 497-516; 1957.