SAN JOSE RESEARCH LABORATORY

MEMO TO: Mr R L. Haug

April 28, 1958

SUBJECT: A preliminary survey of the application of communication theory and related theoretical techniques to data transmission and other machine-component areas

ABSTRACT

The consideration of recent voice-recognition proposals in the light of information theory has lead to the preparation of charts. figures, and tables comparing the application of information theory and associated theoretical techniques to the general machine and component areas of data transmission. character recognition, voice recognition, and storage mediums A table of the bandwidth of the logic and/or transmission circuits required for the different types of input has been prepared Preliminary proposals on the theoretical analysis of the voice input problems are included

F. B. Wood

F B. Wood

FBW:mjm

cc: Mr W C. Dersch
Mr W A Goddard
Mr. M. L. Lesser
Mr J A. McLaughlin

A Preliminary Survey of the Application of Communication Theory and Related Theoretical Techniques to Data Transmission and Other Machine-Component Areas

Table of Contents

List of Illustrations

Tables

1. Introduction

The consideration of possible ways to use theoretical techniques to implement various voice-recognition proposals[1-5] as elements of the proposed voice-input secretarial machine project has led to the tabulation of a chart (Fig. 1) This chart is not considered complete but is prepared as a basis for further discussion

The philosophy of this chart is to explore what can be gained from an overall view of related projects, which might lead to the utilization of communication theory techniques in machine and component areas outside of the field of original development

2. Machine Component Areas (Tabulated in Fig. 1)

The chart is divided vertically into machine and component areas as follows:

a. Data Transmission
   1. Digital
   2. Voice
   3. Analog

b. Character Recognition
   1. Special style
   2. General range of styles, printed
   3. Handwritten

c. Voice Recognition
   1. Phoneme
   2. Letter
   3. Syllable
   4. Word

d. Storage Medium
   1. Optical
   2. Magnetic

The objective of this chart is to explore the potential overlapping of theoretical techniques used in the study of digital data transmission problems with the techniques useful in the other areas listed above.

3. Theoretical Techniques

The vertical division of the chart by rows identified by theoretical techniques such as Cybernetics, Communication Theory, Information Theory, Statistics, etc., is a rough approximation. Many of the fields are overlapping. The interpretation of the terminology as used here can be deduced from the volume of the detailed notes in the following sub-sections:

a. Cybernetics:

In this discussion the field of "cybernetics" is considered as restricted to systems where there are feedback circuits for control, stabilization, or error-correction

A comparison of error checking by feedback requesting a repeat without using a higher order redundant code for error-correction by logic at the receiver has been included in a report on optimum block length.[6]

The primary use of a general feedback circuit in the porposed secretarial machine appears to be in a checking system whereby the sylable or word which the machine selects as having the closest correlation with the voice input would be displayed for visual checking by the operator. This is illustrated in Fig. 3. If the error rate is high for a particular voice, the operator could vary some control which would select a different decision rule more suitable for minimizing the loss for the kind of noise the particular operator introduces. The overall system of the secretarial machine could be illustrated by Fig. 3, which is a modification of Fig. 2.

b. Communication Theory:

The basic problem of communication theory is to transmit information through a channel that is disturbed by noise. The influence of this disturbing noise can usually be compared for different systems by defining a signal-to-noise-ratio, usually a ratio of the root-mean-square values for a particular noise distribution. The basic relationship between transmitted signal, noise, and received signal for a data transmission system is illustrated by Fig. 4. The difference between a "detection" problem of deciding between two states as in digital transmission and the "estimation" problem of deciding between an infinite number of states is explained by Abramson in Ref. 9, p. 37

One way to approach the voice recognition problem is to resolve the class variation and statistical variation of different voices into a "signal" representing a weighted norm and a "noise" representing a statistical variation within a class as illustrated by Fig. 5 and the notes of Ref. 10, pp. 6-1/4. In this comparison a major class distinction such as the difference between the male and female voices could be represented as the frequency shift due to an asynchronous carrier system followed by a set of filters. The filter could be represented by artifical transmission coefficient $(\alpha + j\beta)\ell$. An example of the choice of a signal waveform has been given by Abramson[11] for a digital data transmission system.

c Information Theory:

In this discussion Information Theory is considered as a division of Communication Theory dealing with channel capacity (maximum rate of transmission of information) and methods of coding to reduce the error rate. The basic theorems on channel capacity has been outlined by Shannon[12] and extensive bibliographies have been maintained by Stumpers.[13]

For digital data transmission the channel capacity is given by Shannon. For multilevel transmission, S. Watanabe[14] (IBM, Yorktown) has developed an analysis. For intra-plant cables the available theoretical and experimental data on channel capacity is summarized on pp. 11-17/18 of Ref. 10.

The channel capacity for a binary channel above a certain signal to noise ratio is twice the bandwidth of the circuit. The channel capacity can also be stated as a time-bandwidth product. That is, if one is willing, have a longer time between sampling times a narrower bandwidth circuit can be used.

Consider the information capacity of different voice recognition systems.

10 words $p = 0.1$	$H = -\sum_i p_i \log_2 p_i = 3.32$
1000 words $p = 0.001$	$H = -\sum_i p_i \log_2 p_i = 9.97$
100,000 words $p = 0.00001$	$H = -\sum_i p_i \log_2 p_i = 16.61$
400,000 words $p = 0.000\,0025$	$H = -\sum_i p_i \log_2 p_i = 18.60$

In applying information theory to voice recognition work, we see that the negentropy of the vocabulary gives us a measure of the complexity of the circuits required to identify words in the vocabulary.

In considering first phoneme identification and then other systems to the right we see that the phoneme system is a generating space out of which a large vocabulary in any language can be generated. This corresponds to a multi level system of data transmission. Until a more detailed study is made we can use the hypothesis that this most general system (phoneme identification) is limited in realizability in a way similar to multilevel transmission. If the signal-to-noise-ratio is low we are limited to sylable or word identification, which correspond to lower levels of multilevel transmission. This could also be approached as a level-time product. This approach to understanding the quivalent information content of different voice recognition systems would be dependent upon the validity of the equivalence of the variation of different voices to "noise". See Fig. 6 for channel capacity curves for digital data transmission used as an anology to explore the prospects of voice recognition.

In the reduction and amplification of optical storage images there is a potential application of information theory. A listing of the optical systems corresponding to double- sideband amplitude modulation, single side-band, and phase modulation is included in Ref. 10, p. 9-36. These correspondences indicate a correspondence between certain optical systems and specific carrier communication channels.

Several problems in magnetic recording have been approached from the point of view of information theory. A. B. Fontaine[15] and J. E. MacDonald[16] of IBM have calculated the channel capacity of certain magnetic tape systems. A. S. Hoagland[17] has shown for a particular case that redundant coding has no advantage in magnetic recording. A potential similarity between the Ferranti system of magnetic recording and the dipulse data system of E. Hopner is pointed out in Ref. 10, p. 9-30.

The part of information theory dealing with error detection and correction codes which is so important in digital data transmission would be of less significance in voice recognition by reason of the proposed visual checking display.

d. Statistics:

The statistical distribution of noise is needed to utilize the theorems of communication theory. For thermal noise the probability of the noise voltage exceeding a given value is a gaussian probability function. For impulse noise on telephone lines the probability is different in that the probability of the logarithm of the noise voltage approaches a gaussian probability function. For a comparison see Ref. 10, p. 3-12.

Similarly, statistics of inking variation would introduce noise in character recognition. For reading handwriting the variation of style may have multiple peaks about several "norms." For comments on statistics in voice recognition see notes under Communication Theory.

e. Probability Theory:

From the noise statistics and the decision rules used, the basic error probability functions are derived. The probability of multiple errors is then predicted from probability theory The construction of different codes from Information Theory determines the probability of undetected errors slipping through a digital data transmission system.

In voice recognition, detection of phonemes or letters can permit some errors due to the redundancy in the language In some way similar to data transmission the redundancy of language can be used to correct single errors, so that in a fixed vocabulary system, it is the probability of multiple errors which sets the criterion of successful operation of a voice-input system

f Decision Theory:

An introduction to decision theory has been prepared by N. Abramson [9] In data transmission, the use of decision theory permits us to systematically compare different modulation-demodulation systems or different signal waveforms. The demodulation system constitutes a decision rule. Similarily in character recognition the choice of the method of scanning has similarities to choosing different modulation systems In the area of voice recognition the selection of the "norms" for phonemes, letters, syllable, or words could be handled as a decision theory problem, provided sufficient statistical data is available.

g. Fourier Synthesis:

The use of Fourier Synthesis for generating quasi-transient response curves for transmission lines where there are difficulties in using the more elegant Laplace transforms may have application in other areas. The possibility of synthesizing the waveforms for letters, syllable, or words by computer programs using clipped sonograms is a possibility

h. Correlation Functions:

Correlation functions are used to determine the ambiguity between different signal waveforms, interference of the transient decay tail of earlier pulses with present pulse, and interference of noise with signal in a correlator type detector. These correlation functions also have application in measuring the intersymbol interference and establishing "norms" in both character recognition and voice recognition.

i. Tensor Transforms and Matrix Algebra:

In decision theory matrix algebra is used to calculate the risk by matrix multiplication of the symbol error probability matrix by the symbol distribution vector by the loss matrix by the decision rule vector. There is a possibility of starting with a collection of sonograms for the phonemes which would correspond to a three-dimensional space with multilevels in each dimension from which the simplified waveforms for letter, syllable, or word identification could be developed by a computer program. This proposed theoretical analysis would be of advantage if the vocabulary of the machine is to be greater than some critical value such as forty words. In this way a group of n voices could be recorded for 28 or 40 phonemes instead of 1000 words each. Then the computer could process the sonograms to generate the word waveforms for use in the machine.

j. Logic (Boolean):

Logical operations described by boolean algebra are needed in realizing error-correcting codes in digital data transmission. Similarly if voice recognition is based on distinguishing phonemes or letters a logical decoding system is needed to translate the group of phonemes into one of the allowed words or code points in the system

In data transmission another use of logical operations is also used, namely the transformation of the signal in a way which valid signals add and invalid signals (noise) partially cancel out. In a similar manner combinations of logic and filters could be used in voice recognition.

k. Computer Programming:

All of the theoretical techniques considered above involve numerical calculations which generally can be programmed on a 650 computer However, the more extensive calculations of high order error-correcting codes in digital data transmission and much of the voice recognition work can be more easily done on the 704 computer.

l. Electromagnetic Theory:

In data transmission the channel capacity calculated from information theory requires either experimental or theoretical data from which an equivalent bandwidth can be calculated. For the channel capacity of cables the primary cable constants $R(\omega)$, $L(\omega)$, $G(\omega)$, and C are calculated from the physical dimensions and properties of the conductors and dielectric by use of electromagnetic theory

In a similar way we can expect that any clarification of the limiting resolution in optical storage systems that can be obtained with help of information theory will be dependent upon an adequate description of the physical phenomena of diffraction through electromagnetic theory. The analysis of writing and reading in magnetic storage systems uses electromagnetic theory in a limited way.

m. Network Theory:

In digital data transmission use can be made of Wiener's filtering theory if noise and signal have different waveforms or spectra, so that the filter can be designed to pass the signal and alternate the noise. In distinguishing between different analog signals or signals generated by scanning printed type, the signal can be passed through a set of parallel networks, each adjusted for maximum output from a particular waveform.

In voice recognition a combination of logic and filters may be required. In the simple approach to word recognition a filter network may be used to reduce the waveform to a "simplified" form and this simplified form can be compared with "normal" waveform for the different words.

The use of network analog in determining magnetic field plots in the analysis of magnetic recording has been described by A. S. Hoagland.[18]

n. Systems Theory:

For lack of a more specific definition of measures of complexity of different systems, the preliminary data from an information retrieval proposal (Ref. 2) are used and extended here. For a rate of sixty words per minute compare the different transmission and recognition systems:

(1) Data Transmission

(a) Digital: $60 \frac{\text{words}}{\text{min}} \times 6 \frac{\text{letter}}{\text{word}} \times 7 \frac{\text{bits}}{\text{letter}} \times \frac{\text{min}}{60 \text{ sec}} = 42$ bits/sec.

The theoretical maximum channel capacity for a binary data transmission system is

$$C = 2 W$$

where W is the bandwidth of the circuits. For practical cases the noise and distortion on practical telegraph lines generally requires that the transmission rate be limited to one-sixth of the theoretical limit. This gives a revised equation:

$$W = 3 C$$

For this case:

$$W = 3 \times 42 = 126 \text{ cycles/sec.}$$

(b) Analog Transmission:

For a first approximation use facsimile transmission of a 7 x 9 matrix, then

$$C = 60 \times 6 \times (7 + 2)\ 9/60 = 486 \text{ bits/sec.}$$

(c) Voice Transmission:

Here we use the typical voice channel bandwidth.

$$W = 2500 \text{ cycles/sec.}$$

(2) <u>Character Recognition</u>

Since no accurate estimates are available, the following estimates are proposed:

(a) Special Character Style: (5 x 7 matrix)
i.e., 7 amplitude levels, and 5 sampling times.

$$C = 60 \times 6 \times \frac{35}{60} = 210 \text{ bits/sec}$$

$$W \simeq 3\,C = 630 \text{ cycles/sec.}$$

(b) General Style:

If four comparisons must be made using correlation functions for four classes of styles, multiply bandwidth by four for same word rate

$$W = 4 \times 630 = 2520 \text{ cycles/sec.}$$

(c) Handwritten:

Assume a factor of three for having to make corrections for variations in the way a single individual writes a character at different times.

$$W = 3 \times 2520 = 7560 \text{ cycles/sec.}$$

(3) <u>Voice Recognition</u>

(a) Phoneme Recognition:

Assume a sonogram with 200 bits/character as in Ref. 2, p. 10. Also assume four different sets of correlation functions are stored to cover the range of different voices.

$$C = 60\,\frac{\text{words}}{\text{min}} \times 6\,\frac{\text{letter}}{\text{word}} \times 200\,\frac{\text{bits}}{\text{letter}} \times \frac{4 \text{ (passes)}}{60 \text{ sec/min}} = 4800 \text{ bits/sec.}$$

$$3\,C = 14{,}400 \text{ cycles/sec.}$$

This estimate is approximate. It is hoped that someone will be inspired by this rough estimate to make a better approximation.

(b) Word Recognition:

To obtain a rough approximation to the equivalent bandwidth for word recognition of a vocabulary such as the Basic English vocabulary of 800 words plus a special list of 200 for the particular business, consider a reduction in vocabulary from 400,000 words to 1000 words. In paragraph 3C the information capacties were calculated as follows:

1,000 words	H = 9.97
400,000 words	H = 18.60
Difference:	8.63

Dividing the bandwidth in the above paragraph by 8.63 gives a lower bound on the equivalent bandwidth for a 1000 word vocabulary:

$$W = 14,400/8.63 = 1670 \text{ cycles/sec.}$$

This procedure is valid only if signal-to-noise ratio is above the turning levels of these cases as is illustrated in Fig. 6. The turning levels of S/N for 2-levels and 4-levels are marked S_2 and S_4 respectively in Fig. 6.

4. Proposed Analyses of Voice-Recognition

I propose to assist the development of the voice actuated secretarial machine in the following ways:

a. Establish of scale of complexity of the levels of recognition system such as: Special format printed and typed letter; handwriting; spoken words; spoken letters, etc. A preliminary set of such data is calculated in Section 3n and is summarized in Table I.

b. Analyse the relationship between the basic sonogram representation of spoken letter (phonemes) with different vector spaces such as "Basic English" word space, a special 1000 word space, speech compression systems, and logical systems such as is being done at Yorktown.

c. Plan and execute computer correlation programs on experimental word-shapes and also derive word-shapes from simplified sonograms by facsimile input methods.

d. Analyse the variation of word-shapes between different persons on the basis of the data transmission analogies.

5. Conclusions

I propose to analyse new problems such as the voice-recognition problem for the secretarial machine as an extension of theoretical techniques utilized in related problems. Similarly to the way probability functions calculated for an old project on coherers has been used for plotting noise in a data transmission system and in calculating the usefulness of redundant coding in magnetic recording, I expect that more total work can be done from a viewpoint where we look for potential overlapping of areas or of the critical techniques that will enable us to develop insights into new problems more quickly.

6. References

1. R. W. Porter "Voice Actuated Adding Machine," January 7, 1958.

2. F. B. Wood, "Information-Retrieval-Problems of the Distant Future," January 27, 1958

3. W. C. Dersch, "Character Recognition Proposal," February 4, 1958.

4. W. C. Dersch, "Proposal For a Speech Input Translative Device," February 11, 1958

5. W. C. Dersch, "A Proposal For Coordinating the Character and Speech Recognition Proposals Recently Submitted and Expanding the Scope to Include Word Recognition," February 24, 1958

6. F. B. Wood, "Optimum Block Length for Data Transmission With Error Checking," RJ-MR-11, February 28, 1958.

7. F. B. Wood, "Preliminary Note on the Application of Statistical Decision Theory to Business Data Systems," SJC-12, June 6, 1957.

8. David Middleton and David Van Meter, "Detection and Extraction of Signals in Noise from the Point of View of Statistical Decision Theory," J. Soc. Indus. Appl. Math, Vol. 3, No. 4, December 1955; Vol. 4, No. 2, June 1956.

9. N. M. Abramson, "Seminar on Decision Theory and Its Application to Communication Theory," RJ-MR-8, January 1958.

10. F. B. Wood, "Notes and Questions on Communication Theory." Unpublished loose-leaf notes.

11. N. M. Abransom, "Inter-Pulse Interference and Noise in a Synchronous Pulse Detection System," RJ-MR-13, April 1958.

12. C. E. Shannon, "A Mathematical Theory of Communication," B.S.T.J., Vol. 27, pp. 379 - 423, 623 - 656, July, October, 1948.

13. F. Louis H. M. Stumpers, "A Bibliography of Information Theory (Communication Theory - Cybernetics)," Trans. IRE: PGIT - 2 November 1953; First Supplement, IT-1, No. 2, 31-44, September, 1955; Second Supplement, IT-3, No. 2, 150-165, June 1957.

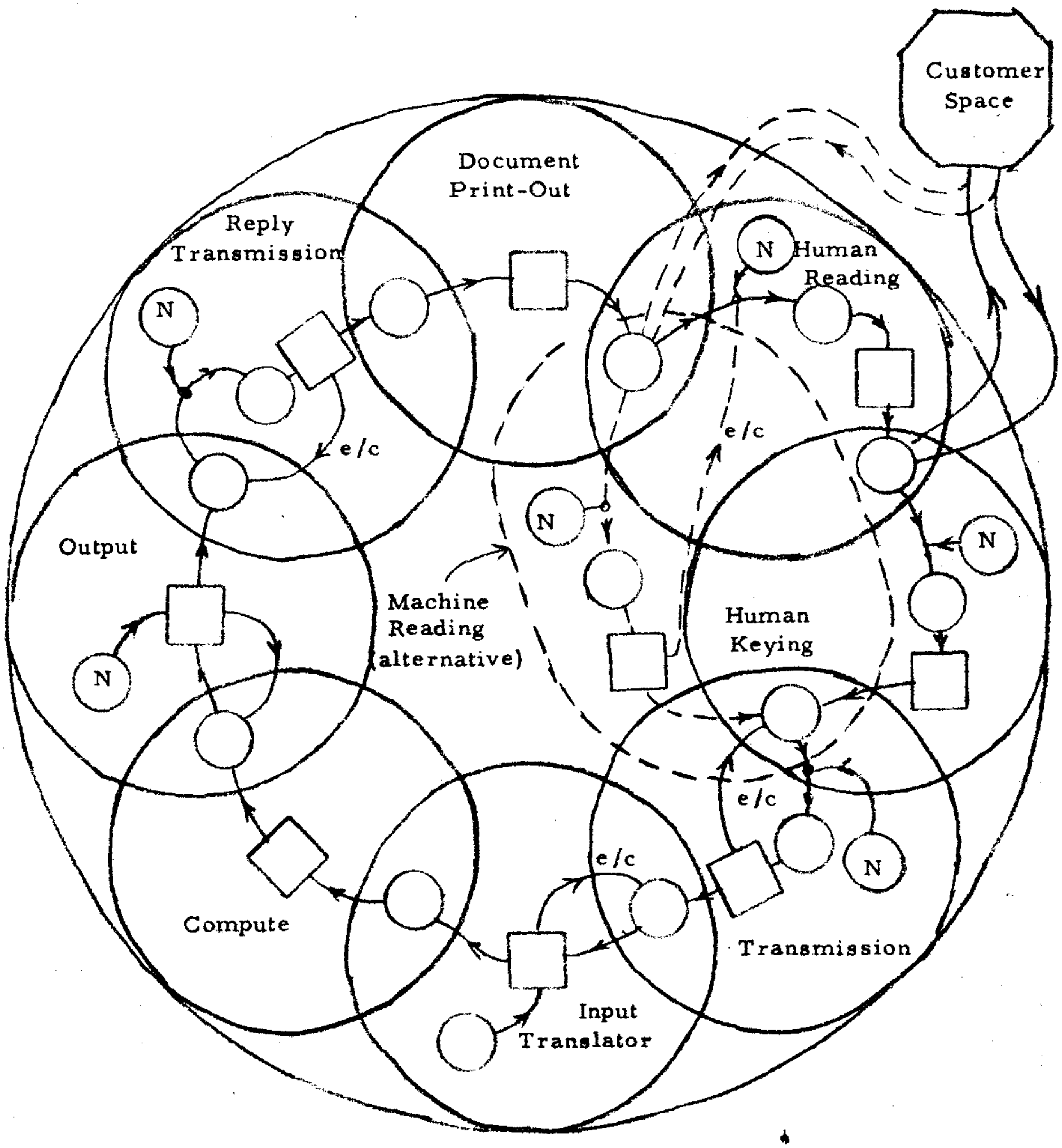
14. S. Watanabe, "A Theory of Multilevel Information Channel With Gaussian Noise," IRE Trans., Vol., IT-3, pp. 214 - 219, December, 1957

15. A. B. Fontaine, "Bounds on Probability of Error for Error Correction Codes," Unpublished notes December 27 1956, pp. 13 - 17 Application to Tape Channel.

16. J. E. MacDonald, "A Calculation of Tape Channel Capacity for the Binary Case," Unpublished notes, October 11, 1956.

17. A. S. Hoagland, "Analytical Study of Redundant Coding of Magnetically Recorded Data When Reading in Presence of Random Noise," RJ-RR-120, February 10, 1958.

18. A. S. Hoagland, "High-Resolution Magnetic Recording Structures," IBM Journal, Vol., 2, pp 91 - 104, April 1958.



Legend:

□ = Decision Rule

○ = Information Space

e/c = Error correcting or detecting link

Fig. 2 -- Organization of a Business Data System into Sub-Spaces

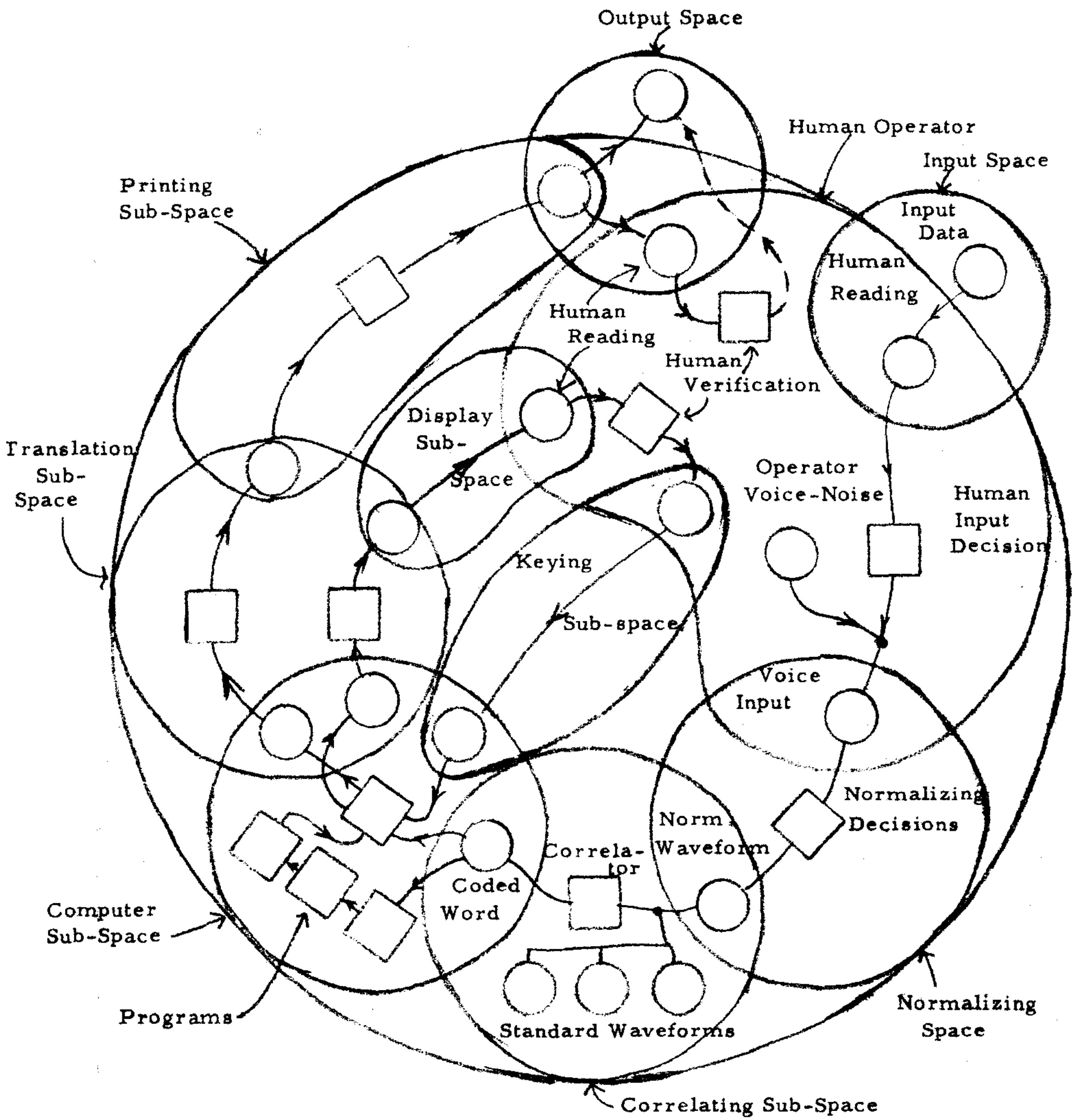


Fig. 3 -- Organization of Secretarial Machine into Sub-Spaces

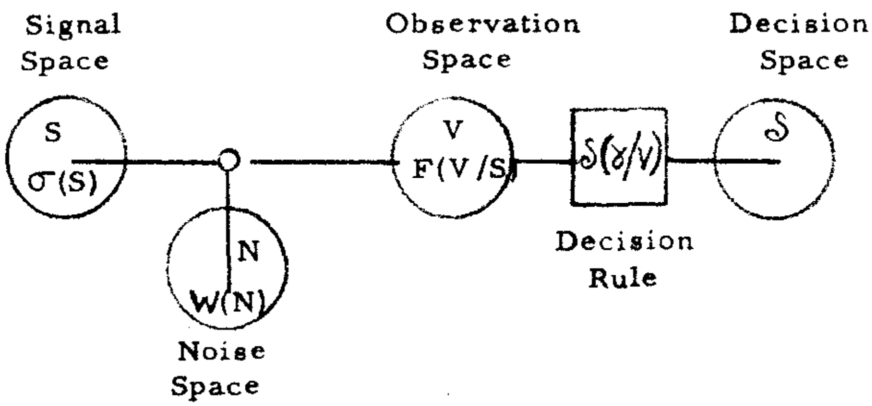


Fig. 4 -- Data Transmission Sub-Space

4

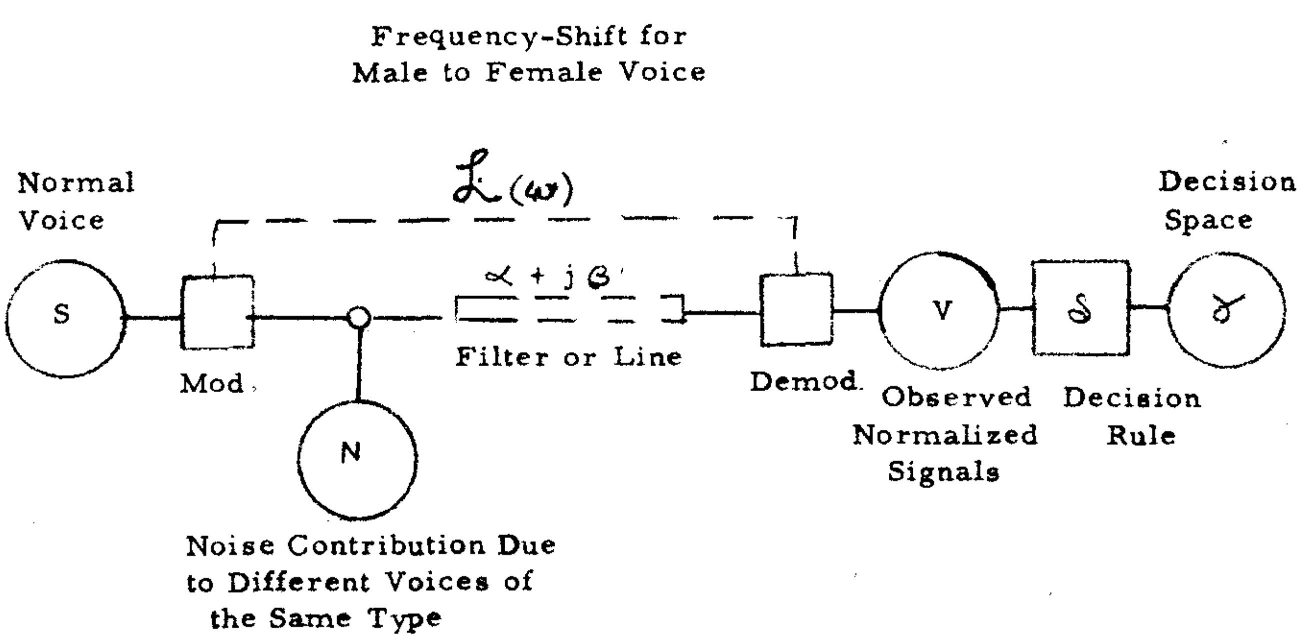


Fig. 5 -- Voice Communication System Resolved into Signal and Noise Components

5

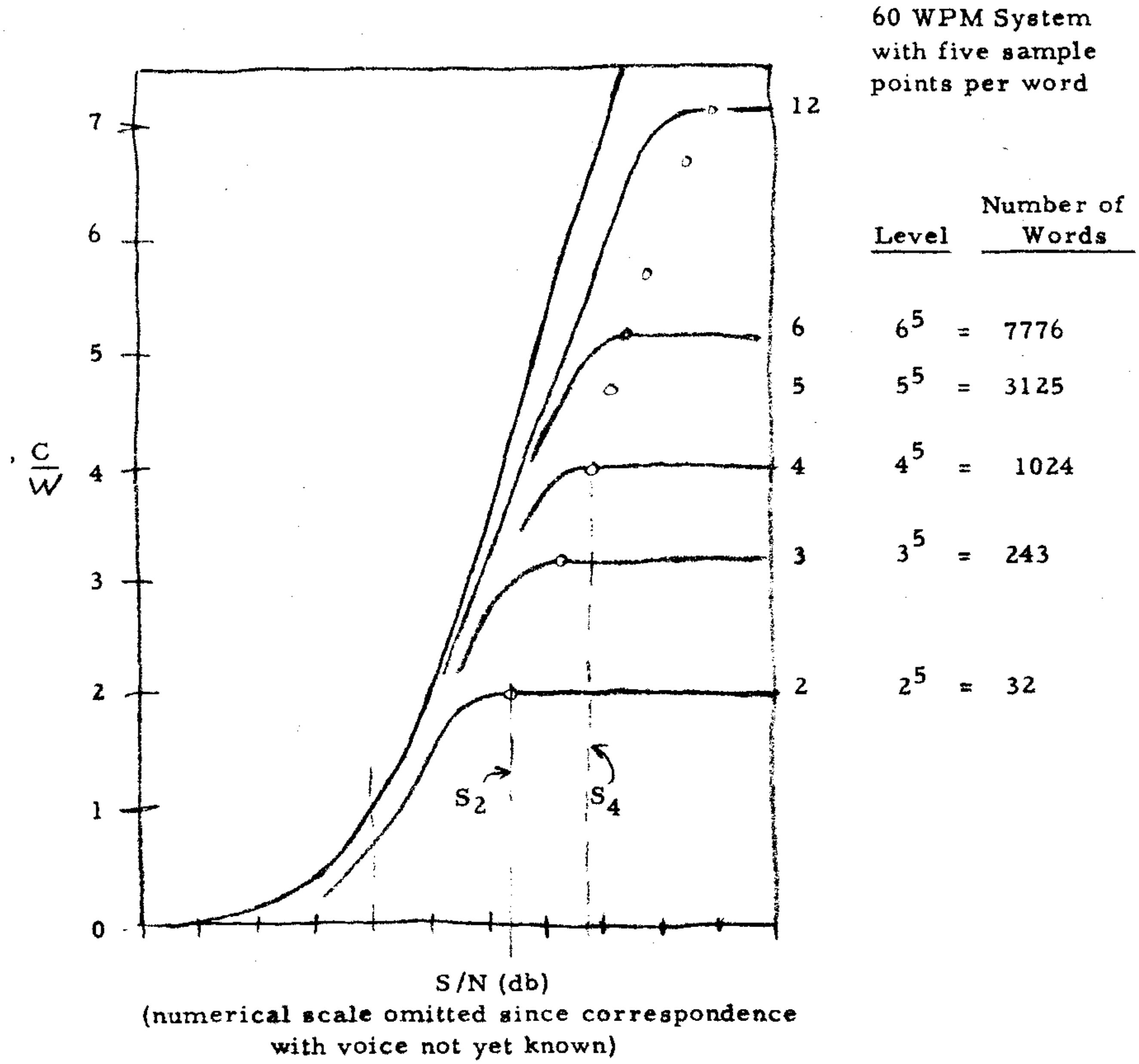


Fig. 6 -- Potential Anology of Channel Capacity vs. Signal-to-noise Ratio Curves in Determining the Limiting Vocabulary of Types of Voice Recognition

TABLE I

Equivalent Bandwidth for Different Data Transmission and Recognition Systems

Machine-Component Area		Equivalent Bandwidth (cycles/second) for 60 words/minute
General Area	Sub-Section	
Data Transmission	Digital (Binary)	126
	Analog (Facsimile)	1458
	Voice (Telephone)	2500
Character Recognition	Special Style (Printed)	630
	General Style	2520
	Handwriting	7560 (l.b.)
Voice Recognition	Word (1000 basic)	1670 (l.b.)
	Syllable	--
	Letter	--
	Phoneme	14,400 (l.b.)

Note: "l.b." means lower bound, where insufficient data is available to make an accurate calculation.