

XV. COGNITIVE INFORMATION PROCESSING*

Prof. S. J. Mason	J. K. Clemens	M. B. Lazarus
Prof. M. Eden	R. W. Donaldson	F. F. Lee
Prof. T. S. Huang	J. K. Dupress	R. M. Mason
Prof. O. J. Tretiak	C. L. Fontaine	Kathryn F. Rosenthal
Prof. D. E. Troxel	K. R. Ingham	T. L. Saxton
Dr. P. A. Kolers	E. E. Landsman	F. W. Scoville
A. L. Citron		A. Spiridan

A. COGNITIVE PROCESSES

1. VARIATIONS OF PERCEIVED DISTANCE WITH APPARENT MOTION

Apparent visual movement occurs when two visual targets are flashed briefly in sequence. If the space separating the two flashes and the time relations between them are set properly, one sees a single target appear briefly in the region of the first flash and then move smoothly and continuously across the intervening space to the region of the second one, where it disappears. Apparent visual movement under optimal conditions is indistinguishable from "real", or veridical, movement, and this similarity in appearance of the two perceptions has been taken as evidence that the two have an identical physiological basis. This argument of identity has recently been shown to be wrong, however, and several criteria distinguishing the two perceptions have been discovered.¹

Our finding of a basic difference in the coding of the two perceptions with respect to their rates, formation times, and masking effects prompted us to examine other properties attributed to the illusion, particularly the variations in distance that are often reported.

If one observes two lights flashing at a rate that produces apparent movement, the distance seemingly travelled by the illusory object varies; however, the published accounts are ambiguous about the nature of this variation. Some writers report that the apparent distance is less than the physical distance; others report that it is greater. Irrespective of whether the apparent distance is less or greater than the physical distance, the fact that there is a difference, and one which has been related more or less systematically to the stimulus conditions producing the basic phenomenon, creates a condition requiring further investigation. The question under study has to do with the formation of the visual space in which we ambulate, and the conditions affecting its psychological representation.

We have studied the variations in apparent distance of the apparently moving object in three ways. In one experiment two fiducial marks, dim points of light (M_1 and M_2), were set some distance apart, just below two lines of light (A and B) whose flashing gives the appearance of motion (see Fig. XV-1). The distance separating the two fiducial

*This work was supported through the Joint Services Electronics Program by the U. S. Army Research Office, Durham, under Contract DA36-039-AMC-03200(E); and in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-05), and the National Aeronautics and Space Administration (Grant NsG-496).

(XV. COGNITIVE INFORMATION PROCESSING)

marks was varied and made either greater or less than the physical distance separating the flashing lines. The subject's task was to say, after each flash of the lines,

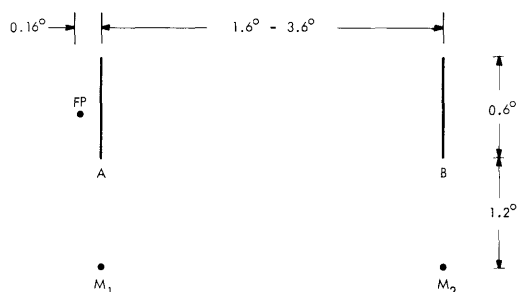


Fig. XV-1.

Physical arrangement of flashing lights (A, B), marker lights (M_1 , M_2), and point of fixation (FP).

the distance traveled by the object in apparent motion appeared greater or less than the distance separating the fiducial marks. The flashes were varied in duration, but were typically 40 msec each; the interval between the offset of the first and the onset of the second (the interstimulus interval (ISI)) was varied 10-250 msec; and the interval between the offset of the second and the recurrence of the first, the intercycle interval, was 3250 msec. The experiment was performed with 3 subjects who made between 20 and 30 observations each at each temporal separation of the interstimulus interval.

When the flashing lights were 12.5 cm apart at a viewing distance of 282 cm ($\sim 2.4^\circ$) separations of the marker lights greater than 14 cm and less than 12 cm were almost always perceived correctly; that is to say, the subjects reported that the distance traveled by the apparently moving object was either less or greater than the distance set by the fiducial marks, which is in conformity with physical reality. When the fiducial marks were 13 cm apart, the subjects tended to report the apparent distance as greater, the probability of these reports increasing slightly with increases in ISI. Thus there was an apparent overshoot of perceived distance, a physical separation of 12.5 cm being perceived as greater than 13 cm, but not as great as 14 cm. This, at most, is an overshoot of 4-8%, far less than the 25-40% reported by others.² Furthermore, we find the overshoot is not a U-shaped function of that temporal interval (see Fig. XV-2.)

In an effort to reproduce the phenomena reported by others, another effort was made to measure the overshoot in apparent distance. In this case the two lights were flashed repeatedly with the dark intervals equal, so that the subject saw a line of light in a continuous oscillation whose period was varied between trials. During each trial, which lasted 3-5 sec, the fiducial lights were moved physically until they appeared to match the distance traveled by the oscillating object. In this case again, there was no systematic deviation from physical reality as a function of the temporal interval separating the two lights: the subjects' judgments describing a match between the marker lights and the apparently moving line were in good agreement with the physical conditions.

Despite these two negative results we persisted with a third experiment, for it will be recalled that the apparent distance does appear to vary with stimulus conditions. The results of this third experiment provide a clue to the basis of the phenomenon.

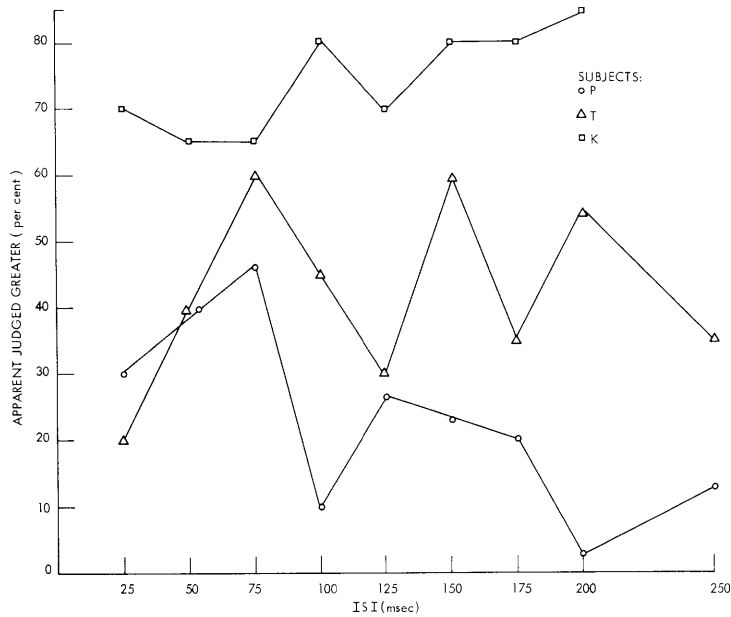


Fig. XV-2.

The percentage of judgments showing apparent distance perceived as greater than physical distance, as a function of the interval between the flashes.

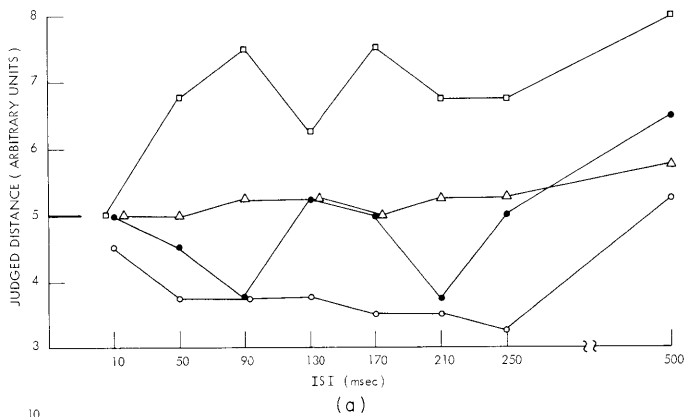
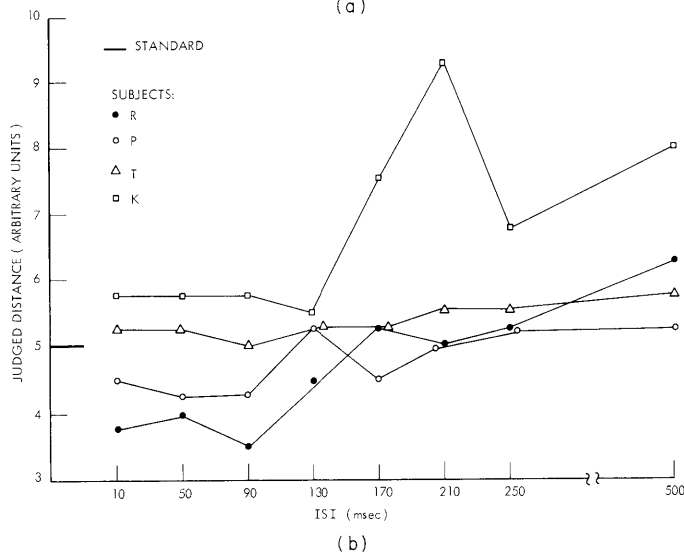


Fig. XV-3.

Judged distance travelled by illusory line. (a) Two lines flashed at 90 msec each; (b) Two lines flashed at 310 msec each. The ordinate value 5 indicates the identity of illusory and physical distance. Values less than 5 indicate that the apparent distance is perceived to be less than the physical distance.



(COGNITIVE INFORMATION PROCESSING)

First, subjects estimated numerically the magnitude of the spatial separation of the two fiducial lights alone, using a subjective scale of distance. Four subjects tested individually generated essentially linear relations between judged distance and physical distance. The subjects were then shown the line in apparent motion, with the fiducial lights no longer present, and again used a subjective scale to estimate the distance traveled by the illusory object. The duration of each flash was 90 msec in one set of trials and 310 msec in another. The judgments were made with an intercycle interval (ICI) of 3250 msec, one judgment per presentation; and with an ICI of 2000 msec, a judgment made on the basis of three flashes. The results in both cases were similar but highly variable. Moreover, the subjects found it difficult to make a single estimation based on three flashes because, as they reported it, the apparent distance varied from flash to flash. Figure XV-3 shows the results for the last estimation; it can be seen that two subjects characteristically produce an undershoot for the shorter ISI values, two characteristically produce an overshoot, and all subjects produce an overshoot at the longest ISI values. Notice also that the variations in apparent distance are not correlated with the ISI values that produce good apparent motion. If they were so correlated, the curves would begin at 5 (the value for the two lamps that are on continuously), depart smoothly from that value to some peak, and then return to 5.

In these experiments, the subject always looked with one eye only through a 4-mm artificial pupil at a small dot of red light (the fixation point) placed just to the left of the first of the flashing lines. It is known, however, that a fixated eye is not entirely immobile; small involuntary motions characterize even the best fixated eye.³ These involuntary motions, we now believe, are the basis of the variations in apparent distance of an object in illusory motion. The effects do not seem to be due to any particular excitation pattern established by the temporal relations between the flashes per se, nor to the occurrence of the illusion of motion, as others have claimed. They appear to be due to involuntary motions of the eye that occur between the offset of the first flash and the onset of the second. We have not, of course, proved this assertion; our apparatus does not lend itself to such exactitude. Significant proof would come from studies of apparent motion with the lines of light established as stabilized retinal images, either optically or by means of a suitably programmed computer that moved a stimulus display to compensate for movements of the eye. Our data, however, are consistent with this belief.

That it is involuntary eye movements that seem to play the predominant role is suggested also by the fact that voluntary movements of the eye undertaken during a stimulus sequence do not affect the apparent distance between the lights when the ISI is short. Thus, if ISI and ICI are both approximately 25 msec with the on-times of the lamps approximately 40 msec, sweeping the eyes across the display does not change the apparent distance between the lights. Involuntary eye movements on the other hand, can directly affect the spatial representation of an object, as demonstrated by Matin.⁴ The

(XV. COGNITIVE INFORMATION PROCESSING)

variations in perceived distance in apparent movement seem to be due to just such movements and not to any figural or temporal interaction. What is more intriguing now is the manner in which such eye movements contribute to the representation of visual space. The implication of these results is that the coding of space occurs in terms of some form of monitored retinal local sign.

P. A. Kolers, G. E. Touchstone

References

1. P. A. Kolers, "The Illusion of Movement," *Sci. American*, Vol. 211, No. 4, pp. 98-108, 1964.
2. W. Scholz, "Experimentelle Untersuchungen über die phänomenale Grösse von Raumstrecken, die durch Sukzessiv-Darbietung zweier Reize begrenzt werden," *Psychol. Forsch.* 5, 219-272 (1925); W. Neuhaus, "Experimentelle Untersuchung der Scheinbewegung," *Arch gesam. Psychol.* 75, 315-458 (1930).
3. T. N. Cornsweet, "Determination of the Stimuli for Involuntary Drifts and Saccadic Eye Movements," *J. Opt. Soc. Am.* 46, 987-993 (1956); R. M. Steinman, "Effect of Target Size, Luminance, and Color on Monocular Fixation," *J. Opt. Soc. Am.* 55, 1158-1165 (1965).
4. L. Marin and G. E. MacKinnon, "Autokinetic Movement; Selective Manipulation of Directional Components by Image Stabilization," *Science* 143, 147-148 (1964).

B. PICTURE PROCESSING

1. EFFICIENT FACSIMILE TRANSMISSION BY SUPERPOSITION OF PSEUDORANDOMLY SCANNED PICTURES

Introduction

Pseudorandom scanning has been used by Deutsch¹ to reduce television frame rate and hence bandwidth. In pseudorandom scanning, the scanning beam does not travel sequentially as in ordinary television, but hops from point to point in a seemingly random fashion. In order to recover the picture, the scanner at the receiving end has to be completely synchronous with the transmitter scanner. If the receiver scanner is out of step, or if it is controlled by a different pseudorandom sequence than that of the transmitter scanner, the reconstructed picture will appear as random noise. In this report, we describe an efficient facsimile transmission system utilizing the property of pseudorandom scanning which has just been mentioned.

The System

Our efficient facsimile transmission system is shown in Fig. XV-4. It is to be used for transmitting two-level (black and white) digital pictures in which the white areas constitute only a small fraction of the total frame. The M pseudorandom scanners are

(XV. COGNITIVE INFORMATION PROCESSING)

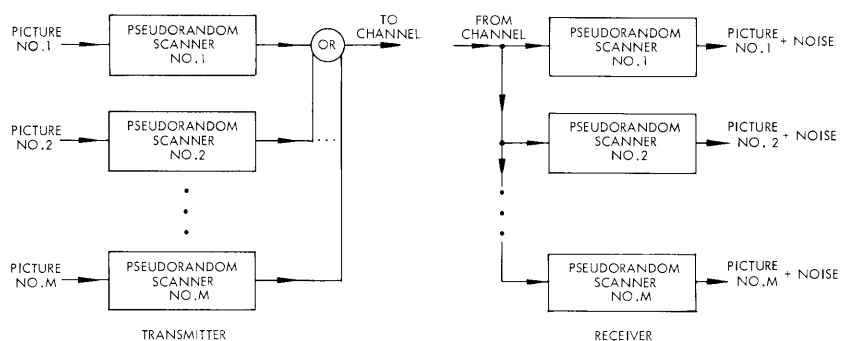


Fig. XV-4. An efficient facsimile transmission system.

controlled by different pseudorandom sequences. At the transmitting end, the output binary sequences (black = 0, white = 1) of the scanners are "ored" and the resulting binary sequence is sent to the channel. At the receiving end, the received binary sequence is fed to M scanners that are identical to those at the transmitter. The output of each receiver scanner will be the picture put into the corresponding transmitter scanner plus random noise caused by the other pictures and the channel. The noise will consist of white points scattered more or less randomly over the picture. Under the assumption that the channel is noiseless, the probability of error (black points received as white) is

$$P \leq (M-1) p, \quad (1)$$

where M is the total number of pictures transmitted at the same time, and p is the fraction of white in each picture. For example, if a probability of error of 0.3 can be tolerated and if the amount of white is 1/20 of the total picture, then 5 pictures can be transmitted at the same time.

Experimental Results

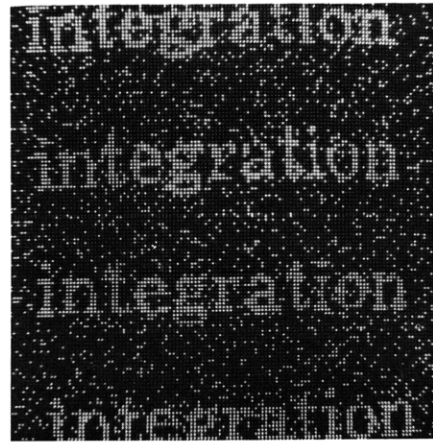
The system described here was simulated on the IBM 7094 computer. Some of the results are shown in Figs. XV-5 and XV-6. The original pictures used were white characters on black, the amount of white being approximately 1/10 of the total picture. Figure XV-5 shows the received pictures for M = 2 and Fig. XV-6 for M = 4. Since the noise is randomly scattered over the picture, it can probably be reduced substantially by some noise-reduction scheme.

Discussion

The system described here is quite suitable for use as a party-line system. As the number of users increases, the noise in the received pictures will increase; however, in our system, as contrasted with ordinary party-line systems, any user can transmit

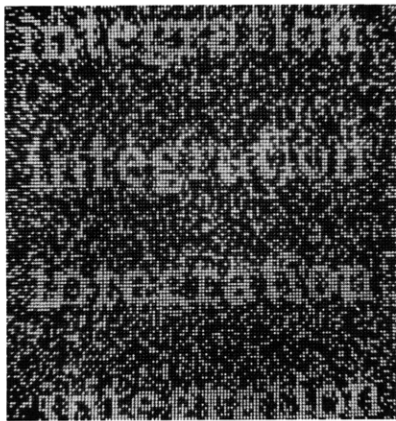


(a)

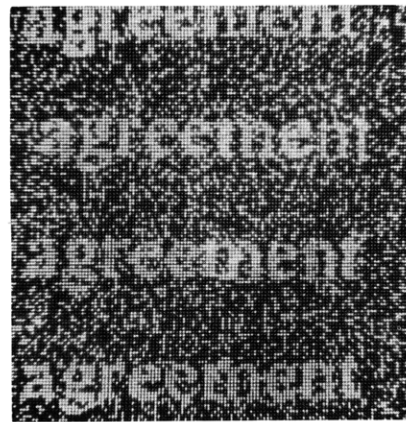


(b)

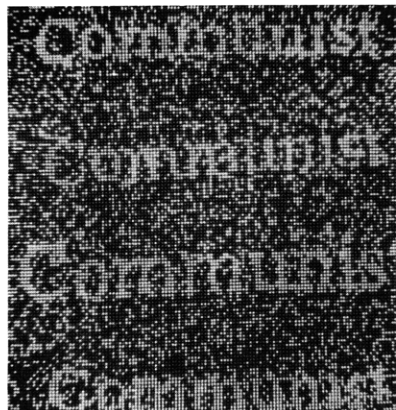
Fig. XV-5. Pictures from computer simulation of the system of Fig. XV-4 ($M = 2$).



(a)



(b)



(c)



(d)

Fig. XV-6. Pictures from computer simulation of the system of Fig. XV-4 ($M = 4$).

(XV. COGNITIVE INFORMATION PROCESSING)

pictures at any time: the line never refuses service to the use, it only degrades gracefully.

The computer programming for this experiment was skillfully done by John Currano.
T. S. Huang, O. J. Tretiak

References

1. S. Deutsch, "Narrow-Band TV Uses Pseudorandom Scan," *Electronics*, Vol. 35, pp. 49-51, April 27, 1962.

2. OPTIMUM BINARY CODE

Introduction

In Quarterly Progress Report No. 78 (pages 231-233) we considered the problem of transmitting digital data over a binary symmetric channel (BSC) using fixed-length binary codes. In particular, we were interested in finding the optimum code that minimizes the mean-square error or average noise power. We reported that if the input data are uniformly distributed over the integers 0 to $2^n - 1$, then the natural code minimizes the single-bit average noise power, e_1^2 , which was defined as the part of the average noise power resulting from single-bit errors in the received code words. If the error probability p of the BSC is small, then the total average noise power is essentially the same as the single-bit average noise power, and therefore the natural code minimizes the total noise power. We believed that the natural code, in fact, minimizes the total average noise for any value of $p \leq \frac{1}{2}$; however, we have no proof thus far. We remark that the results of the last report depend only on the first-order statistics of the channel, and hence are valid for all binary channels, if p is interpreted as the average bit-error rate.

It was shown in the previous report that both natural and Gray codes are good as far as the average noise power is concerned. The purpose of the present report is to compare these two codes in more detail. First, we make a conjecture about the total average noise power of the Gray code. Then, we discuss the effect of the input data on noise power. Finally, natural and Gray codes are compared in connection with picture transmission.

Average Noise Power of the Gray Code

We first write formally our previous conjecture.

Conjecture 1. If the input data are uniformly distributed over the integers 0 to $2^n - 1$, and if the average bit-error rate is $p \leq \frac{1}{2}$, then the natural code minimizes the (total) average noise power.

A particular case of this conjecture maybe expressed as follows.

Conjecture 2. Under the same assumptions as in Conjecture 1, the natural code yields a smaller (total) average noise power than the gray code.

An expression for the total average noise power of a natural code was given in the previous report. One approach to proving Conjecture 2 would be to find an expression for the total average noise power of the gray code; then we could compare the expressions for the average noise powers of the two codes. In carrying out this approach, we formulated the following conjecture.

Conjecture 3. Under the same assumptions as in Conjecture 1, we have

$$(G_n - N_n) = (1-2p) G_{n-1}, \quad (1)$$

where G_i is the total average noise power of an i -bit gray code, and N_i is the total average noise power of an i -bit natural code.

Equation 1 has been verified up to $n = 6$, but we have not been able to prove it in general. It is clear that Conjecture 2 follows readily from Eq. 1.

Noise as a Function of Signal Amplitude

In order to calculate average noise powers for input data with various probability distributions, it is helpful to have curves of noise power (or rms amplitude) as a function of the signal amplitude. The average noise power is equal to the area under the curve which is the product of the input probability distribution curve and the curve of noise power versus signal amplitude.

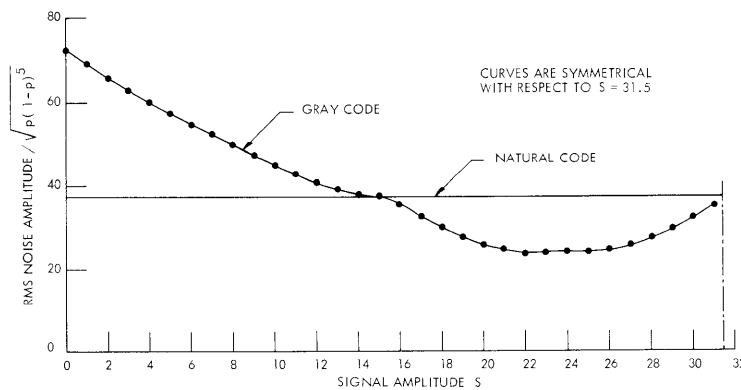


Fig. XV-7. Curves of rms noise amplitude vs signal amplitude, for natural and gray codes.

For natural and gray codes, the curves of single-bit rms noise amplitude versus signal amplitude are shown in Fig. XV-7 for $n = 6$. It is seen that if the input probability distribution has a peak near the center and is small at the extremes, then the gray code might yield less single-error average noise power than the natural code.

(XV. COGNITIVE INFORMATION PROCESSING)

For a natural code, it is clear that the curves of noise power versus signal amplitude are flat for all n . For the Gray code, we observed empirically that the curves for different values of n are very close to each other if they are properly normalized by powers of 2 in both the horizontal and the vertical directions. An interesting problem would be to find the asymptotic ($n \rightarrow \infty$) form of the normalized curve and estimate the deviation of the normalized curves for finite values of n from this asymptotic curve.

Picture Transmission

In our discussion we took the mean-square error as our fidelity criterion. If the input data are samples of a digitalized picture, then the question in which we are interested is, Which received picture has better quality? This "best-quality" criterion depends, of course, on high-order statistics of the channel.

Picture transmission through the BSC has been simulated on the IBM 7094 computer, with both natural and Gray codes used. The received pictures are shown in Fig. XV-8. The picture consisted of 256×256 samples, and the bright was quantized to 64 levels (6 bits). The probability of error of the BSC was $p = 0.003$. The distribution of the brightness levels in the input picture is shown in Fig. XV-9. The peak-signal-to-rms-noise ratio is 29.9 db for the natural-code received picture, and 27.5 db for the Gray-code received picture. The noise in the Gray-code picture appears more visually objectionable than that in the natural-code picture. The poorer performance of the Gray code is evidently due to the fact that the input brightness distribution had peaks at the two extremes. It is quite possible that the subjective quality of a picture received through a BSC could be measured objectively by a weighted mean-square error, the weighting function being chosen to account for the noise visibility over the Gray scale.

Y. Yamaguchi, T. S. Huang

C. SENSORY AIDS

1. MOBILITY AID SIMULATOR

This report describes the development of the mobility aid simulation facility which was described in Quarterly Progress Report No. 73 (pages 233-235).

More accurate position-monitoring equipment has been constructed. The original system used receivers located in the corners of a room. The signal processing involved taking the difference of squares of voltages. The path between the "dummy"



(a)



(b)

Fig. XV-8. Pictures received through a binary symmetrical channel ($p = 0.003$). (a) Natural code. (b) Gray code.

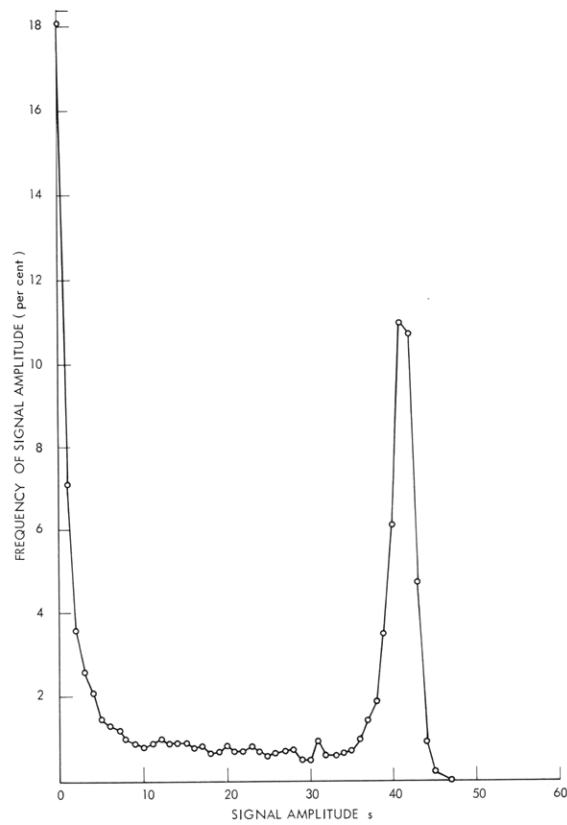


Fig. 9. Frequency distribution of signal amplitudes (brightness levels) of the picture used in Fig. XV-8.

(XV. COGNITIVE INFORMATION PROCESSING)

mobility aid and a receiver could easily be interfered with by the subject who was holding the aid, and the difference of squares operation was implemented by an analog scheme that was subject to drift.

The new system uses a set of three mutually perpendicular polystyrene rods. Each rod provides one dimension of a Cartesian coordinate system. Two microphones are attached to each rod, one at each end. The microphones are Barium Titanate units and

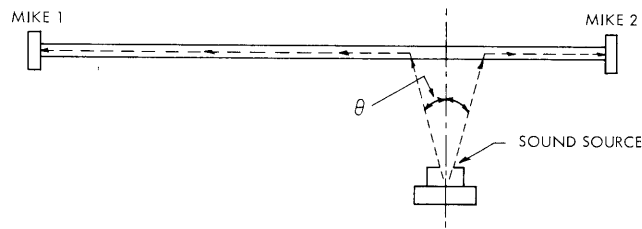


Fig. XV-10. Sound-path geometry.

are fastened to the rods with beeswax. The "x" rod, 20 ft long, is composed of four 5-ft rods held together by epoxy. The first signals to be received at the ends of the rod after each transmitted pulse take the paths shown in Fig. XV-10. The angle θ between the perpendicular and the sound path in the air is just the critical angle for the passage of sound from air to polystyrene and is given by

$$\theta = \sin^{-1} \left(\frac{v_A}{v_R} \right),$$

where v_A is the velocity of sound in air, and v_R is the velocity of sound in the rod. Since v_R is approximately 5 times that of air, $\theta = 11^\circ$. Since the rod is mounted overhead and θ is a small angle, it is much more difficult to interfere with the sound path. Polystyrene was chosen as the rod material because it has the lowest transmission loss of any plastic tested and v_R is not high enough to produce timing problems. It can be seen from Fig. XV-10 that the difference between the arrival times of the first pulse received at one end of the rod and the first pulse received at the other end is directly proportional to distance along the rod, is independent of the radial distance from the rod, and depends only on the velocity of sound in the rod. The velocity of sound in air does not matter, as long as it is the same for both air paths, nor does the time of emission of the sound matter as it did in the old system. A zero crossing of the signal arriving at one end is used to turn on a counter, and the counter is turned off by the corresponding zero crossing of the signal arriving at the other end. A sign bit is derived by detecting which

end received a pulse first (see Fig. XV-11). The linearity of the system is now being

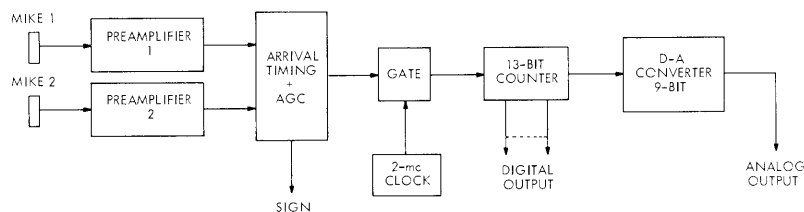


Fig. XV-11. System diagram.

determined. The peak-to-jitter is less than 0.1 inch over at least 15 ft of the "x" rod, and the radial distance from the rod can be as much as 10 ft without degrading performance.

E. Landsman

2. OPTICAL CHARACTER RECOGNITION FOR READING MACHINE APPLICATIONS

Introduction

Optical character recognition (OCR) was applied to a very special problem, that of a reading machine for the blind. Since the proposed machine was, in fact, a personal one, interplay between the machine and the reader could be used to advantage. Figure XV-12 shows a possible reading machine system.

In normal operation the output of the data-reducing equipment will be connected by the switch to the recognition equipment. Whether or not the recognition equipment recognizes a letter depends on the distortion of the printed letter. If it is recognized it will be presented. If we assume, for example, that the output is spoken letters, the letter will be spoken to the operator. If it is not recognized an appropriate signal will be given to the operator. It is now up to the operator to decide whether he wants the distorted letter to be presented by a recoding device such as a two dimensional pattern, or wants the machine to go on.

As an example let us assume that the "e" in the word "the" is distorted beyond machine recognition. The machine could say t - h - beep. The operator would either understand the word or not. If he did not he would throw the switch and the pattern of the distorted "e" would be presented as a tactile two-dimensional pattern (or some other recoded output), independently of how many correct letters had been read after this bad "e". The

(XV. COGNITIVE INFORMATION PROCESSING)

operator could now feel the letter and tell the machine to go on.

The requirements of OCR in this application were found to be quite different

than the requirements in electronic data processing. The most significant changes are the reduction of speed and error-rate requirements and the increase in the requirements of low cost and the ability to recognize type fonts of various styles. No existing OCR techniques were found to be suitable for this purpose and a new technique was developed.

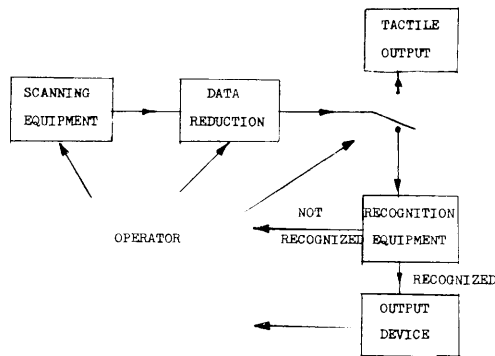


Fig. XV-12.

A personal reading machine system.

Contour Tracing Technique

The technique used here is based upon tracing the outside black-white boundary of the printed character. It was found that in the Roman alphabet the outside contour contained enough information to describe the characters. The outside contour of the character is traced by means of a digitally controlled flying-spot scanner. A column scan is generated which proceeds from left to right until a character is contacted. The scanner then traces the contour of this letter by means of simply generated local operations until the trace is complete. The basis of description of this trace is the local maxima and minima of the horizontal and vertical waveforms of the trace as it proceeds around the character and the positions at which these extrema occur. The details of the trace generation are given in the author's thesis.¹

A method for smoothing had to be developed which would eliminate extraneous extreme caused by printing noise, and yet preserve the extrema arising from significant features of the character. The smoothing method used is a nonlinear technique employing hysteresis. This smoothing is analogous to "gear-backlash." Figure XV-13a shows the electrical model of a "gear-backlash" system. Figure XV-13b shows an input curve and an output curve that is the smoothed curve. Notice that the smoothed curve still has sharp turns and inflection points along it. If a maxima or minima is only detected when the slope of the smoothed curve changes sign, then the extrema pairs that are separated by less than 2E units are eliminated, where 2E is the smoothing threshold.

We found that this smoothing threshold could be one-fourth the letter extent and still retain enough information to recognize the characters of the Roman alphabet. With this large noise threshold most of the variability of the characters caused by style variations

(XV. COGNITIVE INFORMATION PROCESSING)

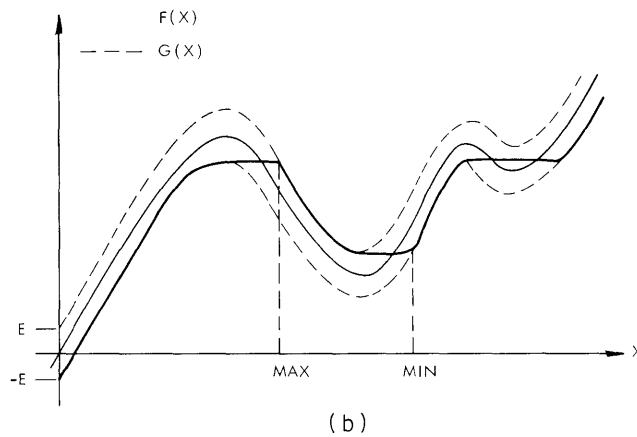
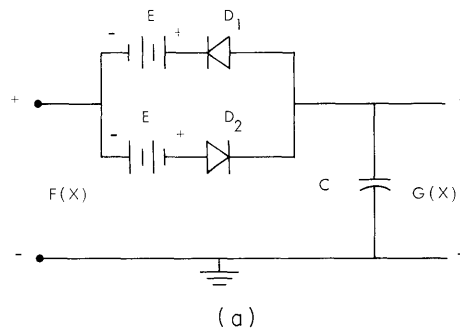


Fig. XV-13. Hysteresis smoothing.

and printing noise is eliminated in the scanner itself.

Generation of Binary Descriptions of Characters

There are three separate binary words used in the recognition of the letters, the height-to-width ratio of the letter, the sequence of the maxima and minima, and the coordinates of the maxima and minima.

The generation of these binary words will be explained by an example. Figure XV-14 shows a letter to be scanned and recognized. The search is started by a column scan, bottom to top, which moves from left to right until the letter is contacted. This will occur at the point labelled "start" in Fig. XV-14. The captive scan is then initiated and the outside contour of the letter is traced. During this trace only two pieces of information are recorded, the letter extent in the vertical, or y direction and the letter extent in the horizontal, or x , direction. From this information, the height-to-width ratio is calculated. This information is also used to set the actual values of the smoothing thresholds. The waveforms occurring in the horizontal and vertical directions during

(XV. COGNITIVE INFORMATION PROCESSING)

with one bit of information if it is assumed that the starting position is at a y minimum, as well as at an x minimum. The first extremum encountered on the trace must then be either a y maximum or an x maximum. This is a binary choice and can be designated with 1 bit of information. Each extremum can be designated by 1 bit, since it is always known what the past extremum was in each coordinate direction. That is, although it may appear that there is a four-way choice of the next extremum along a trace, there is in fact only a two-way choice, because of the history of extrema along the trace.

The code word is generated by recording a 1 for the start position, a 1 for an x extremum and a 0 for a y extremum. This is done in sequence with new extrema added on the right of the word. The code word for the letter in Fig. XV-14 would be 10110010. The start position is labelled as a 1 so that the number of extrema is recorded, otherwise leading zeros would make the code word appear short.

In order to give information about where the extrema were detected, another binary word, called the coord word (coordinate word) is generated. The letter is divided in half in both the vertical and horizontal direction with the information obtained in the first trace. The four areas are labelled as shown in Fig. XV-14. Whenever an extremum is detected it is appropriately recorded in the code word. Also, the label of the area in which the extremum is detected is recorded in the coord word. The coord word is formed in the same manner as the code word. That is, the two bits describing the coordinates of each extremum are added sequentially to the right side of the coord word. The coord word for the letter in Fig. XV-14 is 0011110010101000.

To summarize, the binary description of a character is very simply generated. It is based on a contour trace of the letter which is performed twice. The first trace determines the height-to-width ratio and the thresholds used in the smoothing operation. During the second trace the smoothing and the code-word and coord-word generation are performed simultaneously. When the second trace is complete all pertinent information has been taken and the scan moves on to the next letter.

For purposes of efficient data handling, the maximum length of the 3 binary words is fixed. A maximum length of 12 bits for the code words was found experimentally to be sufficient. The maximum length of the coord word was fixed as 18 bits instead of 24. We found that if the letter was so complicated that it had more than 9 extrema the code word itself carried enough information to determine the letter uniquely. The height-to-width ratio is divided into 4 classes and therefore requires a two-bit word.

The total information about the unknown letter is thereby contained in 32 bits of information. Most of the multifont machines in use today use a factor of 10 more information to describe the character in binary form immediately after scanning. Since, by means of this technique, a large part of the variability of the unknown is eliminated by the scanning technique, a great deal of data reduction has taken place in the scanner itself.

SpACE

CloSl*ng the Gop
(See Cover)

He stood on top of his spaceship*s
white titanium hull*. He touched it with
his bulky thermal gloves*. He burned
around like Duck Rogers propelling
himself with his hand*held jet*. He
floated lazily on his back*. He loked and
laughed*. He gazed down at the earth
id3 miles below*. spotted the Houston*.
Galveston Day area where he lives and
tried to take a picture of it*. Like a gas
station Dttendant* he echecked the space*
craft*s thrusters* wiped its wlnshield*.
Ordered to get back into the capsule* he
protested like a scolded kid*. **I*m doing
great*** he said*. **It*s fun*. I*m not eoming
in*** When* after 2d minutes of space
gymnasties* U.S.* Astponaut Edward
Higgins White II* 34* finally did agree
to squeeze himself baek into his Gemini
q ship* he still had not had enough of
spaae walking*. Said he to Command
Pilot James Alton McDivitt* **It*s the
saddest day of my life***

White*s exhilarating space stroll pro*
vided the moments of highest drama
during Gemini 4* scheduled 62* orbit*
98*hour* i*7dd*ddd mile flight*. White
spent twice the time outside the space*
eraft that Soviet Cosmonaut Aleksei
Leonov did last March 18* and he had
much more maneuverablilty* all Leonov
did was somersault around at the end of
a tethep* getting dizzy* while White
moved around pretty much at will*.

Seecnd Genepotl*on*. Still* Gemini*s
planners would have serubbed White*s
EVA (for Extra*Vehicular Actlvityl
expedltion in a second if they had
thought it might detract from the
flight*s basic missions*.

Fig. XV-15. Example of reading "Time Magazine."

Recognition of the Unknown Character

The classification of the unknown character is made in three steps. The code word of the unknown is compared with the list of previously found code words. When a match is found, the coord word of the unknown is compared with all coord words that previously

occurred with that code word. Finally, the height-to-width ratio is compared and the appropriate character is given as output. The search can be terminated at any step if the unknown is uniquely classified at that step. If no match is found, an appropriate signal is given.

Experimental Data

In order to make a realistic test of the variable font capability of this character-recognition technique, a set of 10 popular scripts was compiled. These type fonts were all of the Roman style but are quite different in design. Transparencies of these type fonts were taken from actual printed material so printing noise was included.

These 10 fonts were used also to experiment with various threshold settings. For each setting of the thresholds the letters were scanned and the results compiled. On the basis of these data the letters were then classified in terms of their code words, coord words, and height-to-width ratio. The number of code words encountered was recorded, as well as the number of coord words needed to distinguish letters that had the same code word. This is a measure of the amount of storage required to recognize the letters. The number of confusions represents the number of letters which could not be distinguished. If 10 B's fell into the classification used for D's, this was called 10 confusions.

In order to make the results meaningful, it was assumed that ascenders and descenders of the letters were detected. This would certainly be done in any reading machine using this technique. The method for obtaining these is straightforward (see Earnest²). It can be done even when there are only several letters in a line of type. Although almost all letters are distinguished by their contours alone, the i, j, and l of the mixed type fonts are more readily distinguished with ascender and descender detection. Rarely are other letter pairs distinguished too.

Details of the various runs are given in the author's thesis.¹ Possibly the best run has a large smoothing threshold with a high recognition rate. This was obtained with the smoothing threshold for the vertical waveform set to 1/4 the letter height and the smoothing threshold for the horizontal waveform set to 1/4 the letter width. The vertical division of the character was set at 9/32 rather than 1/2. The reason for this was that in some cases, notably in the letter "P", the bulge of the letter, although shorter than the letter height, was larger than 1/2 the letter and was not detected as being short. In this case 61 code words and 67 coord words were needed.

In the run of the 10 fonts there were no confusions between upper-case and lower-case letters. There were 30 confusions among the 260 upper-case letters and 1 confusion among the 260 lower-case letters.

(XV. COGNITIVE INFORMATION PROCESSING)

If the letters are assumed to be equally likely, the error rate for the lower-case letters, which make up the majority of letters in common usage, is 0.4 per cent. In this run the following confusions were made: B for D, Q for O, R for A, Y for V, c for e, H for N and C for G. It is interesting to note that the machine confuses letters that are readily confused by the eye. Possibly the two most surprising confusions were H for N because noise made the crossbar appear slanted, and C for G. The crossbar on the G is always more than 1/4 the horizontal extent of the letter; however, the line leading to the crossbar is in general very thick and hides the fact from the contour trace. All of these letters can be distinguished by further operations based on the technique described here. They are discussed in the author's thesis.¹

In order to test the validity of this character-recognition technique for reading-machine applications a full page of "Time Magazine" was read. To be as impartial as possible the first page of text was taken from the issue of June 11, 1965. Over 3300 characters were encountered. There were no touching letters and only one broken letter. The page was scanned once and the data from all the characters were compiled. From these data the appropriate division for the 4 classes of height-to-width ratio was set and the characters were classified according to these code words, coord words, and height-to-width ratio. Only 68 code words and 86 coord words were needed. The number of points around the contour of a letter was counted. If it was more than 16 but less than 64 points, it was called a general punctuation mark and no further data were recorded. The page was then read from the data collected. This is equivalent to reading with a completely trained machine.

Unfortunately, the property of detecting ascenders and descenders was not incorporated in this reading program. Therefore the reading error rate is higher than it need be. The error rate, including letters and punctuation from titles and text, is approximately 3 per cent. If ascenders and descenders were detected, the reading error rate would drop to 1 per cent. This error rate is quite sufficient for reading-machine applications, because of the redundancy of the English Language. English text can be reconstructed from passages for which the error rate approaches 10 per cent.

The major confusions in reading this printed page were the letters l for i and j, p for r, and e for c. Numbers were also confused with letters, the major confusion being d for 0. The titles were printed in sans-serif type so some additional confusions exist there, mainly a for 0. Other confusions are listed in the author's thesis.

Figure XV-15 shows the results from the passage in "Time Magazine," including the confusions. The page is relatively easy to read if the possible confusions are kept in mind.

Conclusions

It is the contour tracing with the incorporation of the large-threshold hysteresis smoothing which permitted an approach to the OCR for the reading-machine applications

problem. The variability of the data after scanning causes expensive data reduction when other techniques are used. Almost all of the variability of the character is eliminated in the scanning process itself, while the information describing the character is maintained. This is evidenced by the fact that the unknown characters can be described by 32 bits of information immediately after scanning and by the small amount of storage needed for the classification of characters. The simple operations used to generate the trace, the hysteresis smoothing technique used to reduce variability, and the small storage requirements suggest the real possibility of an inexpensive reading machine. We estimate that a reading machine operating as described here, including all optical scanning equipment, could now be built for a few thousand dollars, which makes a personal reading machine feasible.

J. K. Clemens

References

1. J. K. Clemens, "Optical Character Recognition for Reading Machine Applications," Ph.D. Thesis, Department of Electrical Engineering, M.I.T., August 1965.
2. L. D. Earnest, "Machine Recognition of Cursive Writing," Information Processing 1962, Proc. IFIP Congress, 1962, edited by C. M. Popplewell (North-Holland Publishing Company, Amsterdam, 1962), pp. 462-466.

3. PERCEPTUAL MODEL

When humans are required to identify a stimulus selected from a given set, the probability of a correct identification usually depends on several stimulus parameters. There is a class of stimuli, however, for the probability of a correct identification depending on only three parameters: the stimulus probability distribution, the stimulus rate, and the error costs.¹ The purpose of this report is to present a model that allows the channel matrix between stimulus and response to be expressed in terms of these three stimulus parameters.

The model is valid under the following conditions. (In this report a stimulus is characterized by its coordinate values along a set of stimulus dimension, analogous to the way in which a vector is characterized by its coordinates along a set of spatial axes.)

1. Each stimulus can assume one of at most two values along any one stimulus dimension.
2. The distance between coordinates along each dimension is well above threshold.
3. Only one of a finite number of stimuli occurs at any time.
4. The probability of occurrence of each stimulus is known by the observer.
5. The cost associated with each type of error is known by the observer.

(XV. COGNITIVE INFORMATION PROCESSING)

Postulates of the Model

The model is based on the following postulates.

(i) The channel between the sensory stimuli and that portion of the nervous system where stimulus identification occurs can be modelled by a large number of identical parallel binary channels (not necessarily symmetric), followed by a decision-making network.

(ii) Each binary channel can transmit information about the coordinates of certain stimulus dimensions. Each channel can transmit information about only one stimulus dimension during any short space of time.

(iii) In general, the probability that a binary channel will correctly transmit the coordinate value of any one stimulus dimension depends on the coordinates along the other dimensions.

(iv) Each individual channel is unreliable, and operates with a probability of error which is slightly less than 1/2.

(v) Decisions as to the identity of the original messages are made in such a way that the observers' risk is minimized.

The following discussion should help to illustrate these postulates.

Assume that a human is faced with the task of distinguishing 4 pure tones that vary along the dimensions of frequency and duration. The frequency is low or high; the duration long or short. In accordance with the second assumption, N_1 of the channels can be used for discrimination of frequency and N_2 for discrimination of duration, while N_{12} of the channels can be used for the discrimination of frequency or duration (see Fig. XV-16).

The fraction of the N_{12} channels used for the discrimination of frequency will depend on the message probabilities. If, for example, the two messages that are represented by tones of high frequency occur with probability zero, then all N_{12} channels will be used

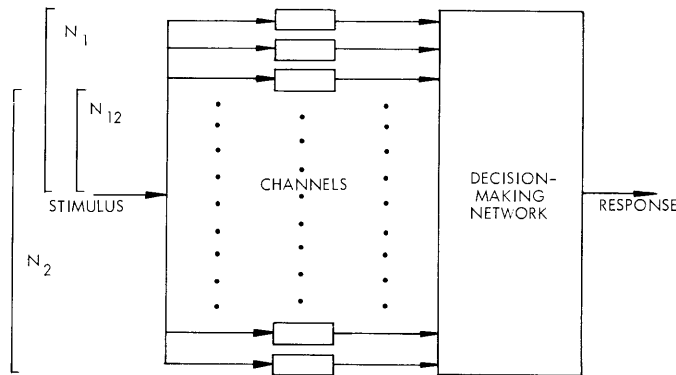


Fig. XV-16. Illustrating the model.

for discrimination of duration, in accordance with the fifth assumption.

In general, the probability of correctly identifying a high-frequency tone will depend on its duration, in accordance with the third assumption. Similarly, the probability of correctly identifying a short-duration tone will depend on its frequency.

Some Special Cases

Consider the problem of discriminating between two stimuli that vary along only one dimension. Stimulus s_0 is represented by the digit zero, stimulus s_1 by the digit one. Each digit is transmitted through N identical channels as a zero or a one. These transmitted digits constitute an output sequence, on which a decision as to the identity of the original stimulus is based.

Define

- v , any one of the 2^N possible output sequences;
- c_{ij} , cost of receiving the i^{th} stimulus when the j^{th} stimulus was sent;
- $p(s_i)$, probability of occurrence of the i^{th} stimulus;
- $p(v|s_i)$, probability of occurrence of the v^{th} output sequence, given that stimulus s_i was sent;
- w_i , the set of all channel output giving rise to a decision that the i^{th} stimulus occurred; and
- p , the probability of error per channel. (Without loss of generality, the channels be considered symmetric.)

The expected cost, or risk c , can be minimized by choosing w_0 to contain all sequences v for which

$$\frac{p(v|s_1)}{p(v|s_0)} \leq L_{01}$$

$$L_{01} = \frac{(C_{10} - C_{00}) p(s_0)}{(C_{01} - C_{11}) p(s_1)}$$

If v_r denotes any output sequence containing r ones (and therefore $N-r$ zeros), then the set w_0 contains all sequences for which

$$\left(\frac{1-p}{p}\right)^r \left(\frac{p}{1-p}\right)^{N-r} \leq L_{01}$$

Substitution of $\sigma = 1/2 - p$ in this inequality yields

$$\left(\frac{1}{2} + \sigma\right)^r \left(\frac{1}{2} - \sigma\right)^{N-r} \leq L_{01}$$

(XV. COGNITIVE INFORMATION PROCESSING)

If this inequality is solved for r , and if second and higher order terms in σ are neglected, the set w_o is found to contain all output sequences v for which $r \leq Nc$.

$$N_c = \begin{cases} N & N'_c > N \\ \frac{N}{2} + \frac{\log L_{01}}{8\sigma} = N'_c & 0 \leq N'_c \leq N \\ 0 & N'_c < 0 \end{cases} \quad (1)$$

From Eq. 1, the channel matrix for the over-all system can be evaluated. If $P(w_i | s_j)$ is the probability that the output sequence v falls in the set w_i when the stimulus s_j occurs, then

$$P(w_o | s_o) = \sum_{r=0}^{N_c} \frac{N!}{r! (N-r)!} p^r (1-p)^{N-r} \quad (2)$$

$$P(w_o | s_1) = \sum_{r=0}^{N_c} \frac{N!}{r! (N-r)!} (1-p)^r p^{N-r}. \quad (3)$$

Approximation of these sums by integrals (2) yields the channel matrix for the over-all system.

$$\begin{bmatrix} P(w_o | s_o) & P(w_1 | s_o) \\ P(w_o | s_1) & P(w_1 | s_1) \end{bmatrix} = \begin{bmatrix} \operatorname{erf} \left(2\sqrt{N}\sigma + \frac{\log L_{01}}{4\sqrt{N}\sigma} \right) & \operatorname{erf} \left(-2\sqrt{N}\sigma - \frac{\log L_{01}}{4\sqrt{N}\sigma} \right) \\ \operatorname{erf} \left(-2\sqrt{N}\sigma + \frac{\log L_{01}}{4\sqrt{N}\sigma} \right) & \operatorname{erf} \left(2\sqrt{N}\sigma - \frac{\log L_{01}}{4\sqrt{N}\sigma} \right) \end{bmatrix}$$

$$\operatorname{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy.$$

Next, consider the problem of discriminating among 4 stimuli that vary along two dimensions. In this case, n_1 channels are used for discrimination along dimension 1, and n_2 channels are used for discrimination along dimension 2. Thus,

$$n_1 = N_1 - a_{12}^1 N_{12}$$

$$n_2 = N_2 - a_{12}^2 N_{12}$$

$$a_{12}^1 + a_{12}^2 = 1.$$

The constants a_{12}^1 and a_{12}^2 are chosen for minimum average risk.

If one assumes a unity cost matrix, minimizing the average risk is equivalent to maximizing the probability of a correct response, $P(c)$.

$$P(c) = \sum_{i=0}^3 p(w_i | s_i) p(s_i).$$

This sum is maximized by choosing w_i to include all output sequences v for which

$$\frac{p(v | s_j)}{p(v | s_i)} \leq \frac{p(s_i)}{p(s_j)} \quad (j=0, 1, 2, 3; j \neq i).$$

Thus, the set w_0 corresponding to stimulus s_0 contains all sequences v for which

$$\frac{p(v | s_j)}{p(v | s_0)} \leq \frac{p(s_0)}{p(s_j)} \quad (j=1, 2, 3).$$

If one assumes that the probability of correctly identifying the coordinate value along one dimension is independent of the coordinates along the other dimension, then substitution of the appropriate likelihood ratios in this inequality gives the following equations. (r_1 is the number of one's transmitted by the n_1 channels; r_2 is the number of one's transmitted by the n_2 channels. It is assumed that n_1 , and n_2 lie between 0 and $n_1+n_2 = N_1+N_2-N_{12}$.)

$$r_1 \leq \frac{n_1}{2} + \frac{\log \frac{p(s_0)}{p(s_1)}}{8\sigma}$$

$$r_2 \leq \frac{n_2}{2} + \frac{\log \frac{p(s_0)}{p(s_2)}}{8\sigma}$$

$$r_1 + r_2 \leq \frac{n_1 + n_2}{2} + \frac{\log \frac{p(s_0)}{p(s_3)}}{8\sigma}.$$

The members of w_1 , w_2 , and w_3 are determined in a similar way.

Once the numbers of w_0 , w_1 , and w_3 have been determined, the over-all channel matrix can be evaluated. For example, if all four messages are equiprobable, then

(XV. COGNITIVE INFORMATION PROCESSING)

$$\begin{aligned}
 P(w_o | s_o) &= \sum_{r_1=0}^{n_1/2} \frac{n_1!}{r_1! (n_1-r_1)!} p^{r_1} (1-p)^{n_1-r_1} \sum_{r_2=0}^{n_2/2} \frac{n_2!}{r_2! (n_2-r_2)!} p^{r_2} (1-p)^{n_2-r_2} \\
 &= \text{erf} (2\sqrt{n_1} \sigma) \text{erf} (2\sqrt{n_2} \sigma).
 \end{aligned}$$

If one lets

$$p_1 = \text{erf} (2\sqrt{n_1} \sigma)$$

$$p_2 = \text{erf} (2\sqrt{n_2} \sigma),$$

then the channel matrix for the over-all system is

$$\begin{bmatrix}
 P(w_o | s_o) & P(w_1 | s_o) & P(w_2 | s_o) & P(w_3 | s_o) \\
 P(w_o | s_1) & P(w_1 | s_1) & P(w_2 | s_1) & P(w_3 | s_1) \\
 P(w_o | s_2) & P(w_1 | s_2) & P(w_2 | s_2) & P(w_3 | s_2) \\
 P(w_o | s_3) & P(w_1 | s_3) & P(w_2 | s_3) & P(w_3 | s_3)
 \end{bmatrix}
 =
 \begin{bmatrix}
 p_2 p_1 & p_2 (1-p_1) & (1-p_2) p_1 & (1-p_2) (1-p_1) \\
 p_2 (1-p_1) & p_2 p_1 & (1-p_2) (1-p_1) & (1-p_2) p_1 \\
 (1-p_2) p_1 & (1-p_2) (1-p_1) & p_2 p_1 & p_2 (1-p_1) \\
 (1-p_2) (1-p_1) & (1-p_2) p_1 & p_2 (1-p_1) & p_2 p_1
 \end{bmatrix}
 \tag{4}$$

Once the parameters $\sqrt{N_1} \sigma$, $\sqrt{N_2} \sigma$, and $\sqrt{N_{12}} \sigma$ have been evaluated, the quantities p_1 and p_2 can be determined by choosing a_{12}^1 and a_{12}^2 to minimize the probability of error.

The parameters can be determined by evaluating the channel matrix that results when an observer is required to discriminate between judiciously chosen pairs of stimuli. By measuring the channel matrix that results when an observer discriminates between two equiprobable stimuli that vary along dimension 1, $\sqrt{N_1} \sigma$ can be determined from

$$P(w_o | s_o) = \text{erf} (2\sqrt{N_1} \sigma).$$

By measuring the matrix resulting when an observer discriminates between two equiprobable stimuli that vary along dimension two, $\sqrt{N_2} \sigma$ can be evaluated from

$$P(w_o | s_o) = \text{erf} (2\sqrt{N_2} \sigma).$$

Finally, determination of the matrix resulting when an observer discriminates between two equiprobable stimuli that vary along both dimensions allows $\sqrt{N_{12}} \sigma$ to be determined from the two previous equations and from

$$P(w_o | s_o) = \text{erf} [2(\sqrt{N_1} \sigma + \sqrt{N_2} \sigma - \sqrt{N_{12}} \sigma)].$$

General Case – Discrimination among 2^m m-dimensional Stimuli

A set of stimuli which can assume one of two coordinate values along m stimulus dimensions can contain 2^m different stimuli. In order to evaluate the confusion matrix for the over-all system, considerable algebraic manipulation is sometimes required. In principle, the computation is always possible. It can be shown that when the N 's and σ 's appear in various terms in the over-all matrix, they always occur as products of the form $\sqrt{N} \sigma$. It can also be shown that all such products can be determined from the channel matrices that result when an observer is required to discriminate between judiciously chosen pairs of stimuli.³

a. Qualitative Comparisons of the Model with Experimental Data

The model is in qualitative agreement with several experimental facts. For example, when 2^m messages are mapped into sensory stimuli that vary along m dimensions, the model predicts that the probability of error will increase with m , since the number of channels available for discrimination decreases as m increases. When two messages are mapped into stimuli that vary along dimension 1, N_1 channels are available for discrimination. When the messages are mapped in stimuli that vary along dimension 2, N_2 channels are available. When 4 messages are mapped into stimuli that vary along dimensions 1 and 2, $N_1 - a_{12}^1 N_{12}$ channels are available for discrimination along dimension 1, and $N_2 - a_{12}^2 N_{12}$ channels for dimension 2. If N_{12} is non-zero, then the probability of a correct identification along each dimension decreases. Such an effect was observed by Pollace,⁴ and is consistent with earlier results.¹

The model also predicts that a redundant use of stimulus dimensions will reduce the probability of a transmission error. When two messages are mapped into stimuli that vary along two dimensions, $N_1 + N_2 - N_{12}$ channels become available for discrimination. Since N_1 and N_2 are always larger than N_{12} , the number of channels increases, and $P(e)$ decreases. Some earlier experiments,¹ as well as experiments performed by Ericson,⁵ support this prediction.

The fact that many of the channels can be used for discrimination along one of several dimensions suggests that people are able to focus attention on a few stimulus features, while paying little attention to others. We as humans seem to possess this flexibility.

b. Qualitative Comparisons of the Model with Experiments

When humans are required to discriminate between 2 m -dimensional stimuli, it can be shown that

(XV. COGNITIVE INFORMATION PROCESSING)

$$P(w_0 | s_0) = \operatorname{erf} \left[\frac{8(\sqrt{N}\sigma)^2 + \log L_{01}}{4\sqrt{N}\sigma} \right] \quad (5)$$

$$P(w_0 | s_1) = \operatorname{erf} \left[\frac{-8(\sqrt{N}\sigma)^2 + \log L_{01}}{4\sqrt{N}\sigma} \right], \quad (6)$$

where N is the total number of channels available for discrimination.

$$\left(N = \sum_{i=1}^m N_i - \sum_{i=1}^m \sum_{j=1}^m N_{ij} + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m N_{ijk} \dots (-1)^{m+1} N_{ij\dots k} \right)_{i \neq j \dots k}$$

These equations support the hypothesis that when one of two stimuli can occur, sensory data can be described as varying continuously along a single axis, partitioned to minimize the observer's expected risk. In fact, (5) and (6) are almost identical to those used by Swets, Tanner, and Birdsall,⁶ Egan,⁷ and others to derive their receiver operating characteristics. The only difference is that the variances in (5) and (6) are equal. Although some workers do not impose this restriction, the experimentally determined variance ratio is often close to unity. Even when the data are best fitted with unequal variances, reasonable agreement between theory and experiment is obtained by assuming equal variances.

When an observer was required to discriminate between 4 equiprobable pure tones which varied in frequency and duration, the channel matrix shown in Fig. XV-17 resulted. It was found that the probability of correctly identifying the coordinate along one dimension was independent of the coordinate value along the other dimension. Thus, the theoretical matrix is given by (4). If dimension 1 is the frequency dimension, it was found that $\sqrt{N_1} \sigma = .46$; $\sqrt{N_2} \sigma = .19$; and $\sqrt{N_{12}} \sigma = .08$.

	1	2	3	4					
1	.69	.11	.17	.03	.58	.10	.27	.05	
2	.11	.69	.03	.17	.17	.55	.04	.24	
3	.17	.03	.69	.11	.20	.05	.64	.11	
4	.03	.17	.11	.69	.07	.22	.09	.62	
	<u>Theoretical</u>				<u>Experimental</u>				

Tone 1: 1000 cps, 50-msec duration

Tone 3: 1000 cps, 150-msec duration

Tone 2: 3000 cps, 50-msec duration

Tone 4: 3000 cps, 150-msec duration

Stimulus Rate: 1.6 stimuli/sec

Fig. XV-17. Auditory channel matrices.

When an observer was required to decide whether or not each of two adjacent fingers was stimulated tactually by a poke probe, the probability of correctly identifying a poke on one finger was larger when the other finger was poked than when it was not poked (in accordance with Weber's law). An equal number of channels were available to each finger. If the subscript of σ is used to designate a poke (1) or no poke (0) on the other finger, then, at a stimulus rate of 2.6 stimuli/sec, $\sqrt{N_1} \sigma_0 = .41$; $\sqrt{N_1} \sigma_1 = .33$; $\sqrt{N_{12}} \sigma_0 = .15$; and $\sqrt{N_{12}} \sigma_1 = .13$. It was found that the number of channels common to pairs of fingers decreased as the distance between fingers increased, as did the ratio of σ_0/σ_1 . The same was true when the fingers were stimulated kinesthetically, rather than tactually.

When an observer was required to simultaneously identify pure tones differing in frequency, and decide whether or not a finger was poked, approximately one-quarter of the total number of channels was common to the auditory and tactile dimensions.

Variation of the Parameters with Stimulus Rate

When the distance between the stimulus coordinates is well above threshold, the $\sqrt{N} \sigma$'s are independent of all stimulus parameters except rate.¹ When subjects were required to discriminate between several pairs of stimuli, it was found that the $\sqrt{N} \sigma$'s were linear functions of stimulus period.

$$\sqrt{N} \sigma = c(T-0.25).$$

The constant c depends on the stimulus pair. The time T is in seconds.

Thus, parameters measured at T_1 can be scaled to T_2 as follows:

$$\sqrt{N} \sigma (T_2) = \sqrt{N} \sigma (T_1) \frac{(T_2 - 0.25)}{(T_1 - 0.25)}.$$

Dimensional Goodness and Dimensional Independence

The model suggests a quantitative comparison of the goodness of a set of sensory dimensions, and a measure of the degree of dependence of two or more dimensions. The numbers $N_1 \sigma^2$, $N_2 \sigma^2$, $N_3 \sigma^2$, etc., all evaluated for the same coordinate values and stimulus rate, can be used as a numerical measure of dimensional goodness. Evaluation of the ratio N_1/N_2 allows one to say whether dimension 1 is twice as good, three times as good, or half as good as dimension 2.

The ratio $\frac{N_{12}}{N_1 + N_2 - N_{12}}$ provides a measure of the degree of interdependence of two dimensions. A unity ratio indicates strong dependence, as all channels can be used for discrimination along other dimensions. A zero ratio indicates dimensional

(XV. COGNITIVE INFORMATION PROCESSING)

independence. Similarly, the ratio

$$\frac{N_{123}}{N_1 + N_2 + N_3 - N_{12} - N_{13} - N_{23} + N_{123}}$$

is the measure of the degree of interdependence of three dimensions.

When the coordinates along a set of stimulus dimensions are statistically independent, and when all the N_{ij} 's, N_{ijk} 's, etc. are zero, the information transmitted is equal to the sum of the information transmitted by each dimension. The probability of a correct identification is the product of the probabilities of a correct identification of each coordinate value. These statements follow directly from the discussion in Fano.⁸ When the dimensions are not mutually independent, the information transmitted and probability of error is less than if they were independent.

Although the numerical goodness of each dimension varies with coordinate values and stimulus rate, ratios of these goodness values do not. Thus, the relative goodness and interdependence of a set of dimensions are virtually independent of all stimulus parameters.

It is interesting that no independent dimensions were found experimentally. One might expect that dimensions belonging to different senses would be independent, especially when signals from the receptors of different senses to different regions of the cortex. Apparently such is not the case.

R. W. Donaldson

References

1. R. W. Donaldson, "Information Rates for Multidimensional, Multimodality Sensory Stimuli," Quarterly Progress Report No. 78, Research Laboratory of Electronics, M. I. T., July 15, 1965, pp. 241-248.
2. E. Parzen, Modern Probability Theory and its Applications (John Wiley and Sons, Inc., New York, 1963), pp. 239-245.
3. T. W. Donaldson, "Multimodality Sensory Communication," Ph. D. Thesis, Department of Electrical Engineering, M. I. T., June 1965.
4. I. Pollack and L. Fitts, "Information of Multidimensional Auditory Displays," J. Acoust. Soc. Am. 26, 1954, pp. 155-158.
5. C. W. Ericson, "Multidimensional Stimulus Differences and the Accuracy of Discrimination," U. S. A. F. WADC Technical Report (1954), pp. 54-165.
6. J. A. Swets, W. P. Tanner, and T. G. Birdsall, "Decision Processes in Perception," Psychol. Rev 68, 301-340 (1961).
7. J. P. Egan, A. I. Schulman, and G. Z. Greenberg, "Operating Characteristics Determined by Binary Decisions and by Ratings," J. Acoust. Soc. Am. 31, 768-773 (1959).
8. R. M. Fano, Transmission of Information (the M. I. T. Press, New York, 1961), p. 31.

4. EXPERIMENTS IN READING NONVISUAL TEXT

It has been shown experimentally¹ that the speed and accuracy with which humans are able to identify sensory stimuli depends on the dimensions along which the stimuli vary. These experiments showed that, in general, identification is easiest when stimuli vary along many dimensions and assume one of two coordinate values along each dimension. In this case, a few simple measurements allow the channel matrix to be evaluated once the stimulus probabilities, stimulus rates, and error costs are known (see Sec. XV-C. 3).

These results were applied to the problem of transmitting English text by means of the nonvisual senses. Each letter was coded into a unique sensory stimulus. In the initial tests, a restricted alphabet of eight letters was used. Finally, the entire alphabet was used.

Description of the Experimental System

In order to enable stimuli to be presented sequentially to a human, the system shown in Fig. XV-18 was designed and constructed.

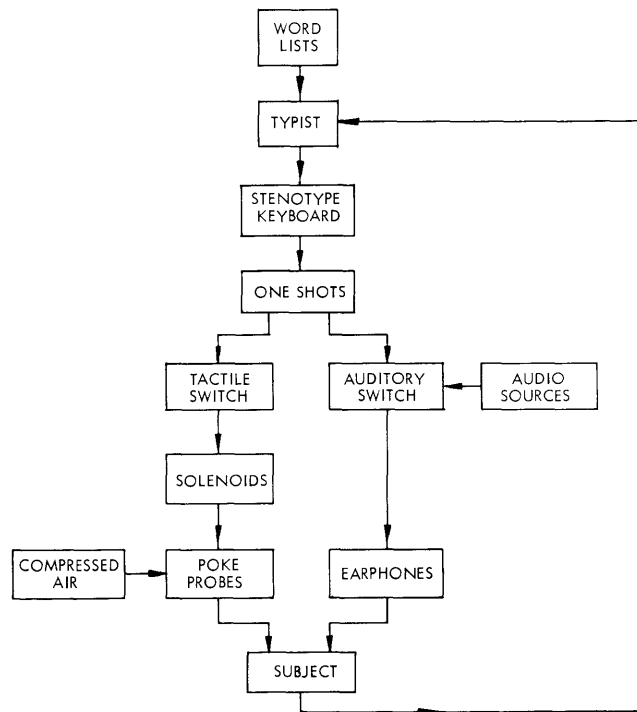


Fig. XV-18. Experimental system.

(XV. COGNITIVE INFORMATION PROCESSING

By depressing certain keys on a stenotype keyboard, a typist was able to convert English letters into a unique set of voltage pulses. These pulses triggered appropriate one shots. The triggering of a one shot resulted in a change in its state. This change in state caused an auditory, tactile, or kinesthetic switch to open for the duration of the state change.

Each auditory stimulus was delivered by separately wired PDR-8 earphones. The tactile stimulators were in air-operated poke probes, designed by Troxel.²

The time between successive stimuli was controlled by the typist. Other stimulus parameters, such as stimulus intensity and duration, were controlled by dial settings on the one shots, on the audio-energy sources, and on the compressed-air source.

Eight-Letter Multidimensional Auditory Alphabet

The eight most common letters of the alphabet were coded into sensory stimuli which varied along the following dimensions.

1. frequency of a pure tone (700 cps or 3000 cps)
2. duration of a tone (30 msec or 150 msec)
3. intensity of a tone (zero or well above threshold)
4. ear stimulated by tone
5. frequency range of a noise burst (0-400 cps or 10 kc- ∞ ; 12 db per decade fall-off rate).
6. duration of a noise burst (30 msec or 150 msec)
7. intensity of a noise burst (zero or well above threshold)
8. ear stimulated by noise burst.

Thus, there are 256 points in an eight-dimensional stimulus space. Those points with coordinate values of zero along the third and seventh dimensions could not be used. Also, a pilot experiment showed that identification of a tone or noise is easier when either one occurs in the absence of the other. To facilitate learning, all vowels were pure tones, and all consonants were noise bursts. With these restrictions, the code that minimizes the probability of error is the following one. (The stimulated ear is designated by a capital letter.)

a: 700 cps, 30 msec, L	n: 0-400 cps, 30 msec, L
e: 3 kc, 30 msec, L	r: 10 kc- ∞ , 30 msec, L
i: 700 cps, 150 msec, R	s: 0-400 cps, 150 msec, R
o: 3 kc, 150 msec, R	t: 10 kc- ∞ , 150 msec, R

Four subjects were given practice in identifying letters and words containing only these letters. The subjects worked in pairs, each subject transmitting a word to his partner. After the typist had transmitted the word, the reader identified it orally. If his response was correct, the typist sent the next work. the response was

(XV. COGNITIVE INFORMATION PROCESSING)

incorrect, the typist transmitted the word again until the correct response was made. The roles of typist and reader were reversed several times each day.

For half-hour periods were spent sending and receiving words (or letters) selected at random from a given list. The lists appear below.

List 1: a, e, i, o, n, r, s, t.

List 2: an, as, at, in, is, it, on, or, no, so, to.

List 3: are, its, one, not, ran, sit, sat, set, see, too.

List 4: into, near, none, rain, rose, seat, seen, soon, star, tree

List 5: enter, noise, start, stone, store, entire, insist, reason, series, street

List 6: a, I, and all the words in Lists 2, 3, 4, and 5.

A contest was held every 5 days. On the contest days, the subjects' pay depended on their performance relative to the performance of the other team. The error-free transmission rate in words per minute on the contest days appears in Table 1. The rates opposite each subject were computed on those trials for which the subject was the reader. Team 1 did not participate in the last contest.

Table XV-1. Reading rates.

<u>Team</u>	<u>Subject</u>	<u>List Number</u>					
		1	2	3	4	5	6
1	B. P.	39	22	23	29	23	
1	D. M.	38	27	23	30	18	
2	W. O.	54	49	37	35	39	33
2	J. B.	59	41	40	43	39	29

During the contests, most of the messages were transmitted correctly the first time. Approximately one message in ten had to be sent more than once; one in 30 had to be sent more than twice; and 30% of the errors were typing errors. The tendency toward error, especially by the reader, increased with the length of the words. This increase was most pronounced for the words from List 5. Approximately 20% of these words had to be sent more than once. In any given list, the words that were confused most often were those that differed from other words in only one letter. In List 3, "sit" was often confused with "sat" or "set", but rarely with "are" or "one".

All four typists tended to type in such a way that the time between successive letters in a word was the same for all letters. When a new list was being learned, the time between successive letters was approximately 1 second. This time was gradually decreased as the readers became more and more familiar with the words.

All 4 subjects felt that the reading rate would increase if oral responses and sending errors were eliminated. In order to test this hypothesis, computer prepared punched

(XV. COGNITIVE INFORMATION PROCESSING)

paper tape and a tape reader replaced the word lists, the typist, and the stenotype machine. The time between successive letters was controlled by the subject. The time between successive words was four times the time between successive letters. After 2 days of practice (1/2 hour per day), W.O. was able to read 42 wpm, while J. B. was able to read 35 wpm. At these rates, 90% of the words were correctly identified.

Both subjects were given an additional day of practice in making silent responses. The speed control on the tape reader was set in such a way that each subject felt that he was identifying 90% of the words correctly. After one day of practice, W.O.'s rate was 71 wpm, while J.B.'s was 59 wpm. Both subjects felt that a proportional reduction in the duration of all tones, a relative increase in the time between words, and more practice would permit even higher rates.

Eight-Letter Multimodality Alphabet

The eight most common letters were coded into auditory or tactile stimuli. The vowels were pure tones which varied along the same dimensions as in the previous section. A tactile stimulus represented each consonant. The finger stimulated was the one used to produce the stimulation by typing. With these constraints, the optimal coding scheme is the following one. The fingers are numbered consecutively from left to right, thumbs excluded.

a, e, i, o: Same as above

n, r, s, t: A poke on fingers 3, 4, 5, and 6, respectively

The experimental procedure was identical to that described above. The general remarks concerning errors also apply here. The transmission rate in words per minute on the contest days appear in Table XV-2. As before, the senders typed in such a way that the time between successive letters was constant.

Table XV-2. Reading rates.

<u>Team</u>	<u>Subject</u>	<u>List Number</u>					
		1	2	3	4	5	6
1	A.G.	57	40	41	33	23	27
1	G.T.	53	43	38	35	20	29
2	D.C.	54	48	53			
2	F.K.	45	39	52			

Both subjects were given two days of practice in receiving the words in List List 6, presented by the tape reader. At the end of the practice period, A.G. was able to read 38 wpm, while G.T. read 44 wpm. At these rates, 90% of the

(XV. COGNITIVE INFORMATION PROCESSING)

words were correctly identified.

After an additional day of practice during which oral responses were replaced by silent responses, A. G. felt that he was correctly identifying 90% of the words at a rate of 61 wpm. G. T.'s rate was 55 wpm. Both subjects stated that an increase in the time between words relative to the time between letters, as well as more practice, would result in an increase in reading rate.

Eight-Letter Tactile Alphabet

The 8 available tactile dimensions gave the intensity with which a finger could be poked. The finger poked was the one used to produce the stimulation when typing. The coding scheme is the following.

<u>Letter</u>	<u>Finger Stimulated</u>	<u>Letter</u>	<u>Finger Stimulated</u>
a	1	n	3
e	2	r	4
i	7	s	5
o	8	t	6

The experimental procedure was identical to that above, with one exception. Only one team was used. The earnings on the contest days were computed by averaging the transmission times of all other subjects who were tested on the same word lists. The transmission rates on the contest days appear in Table XV-3. The general remarks concerning errors made above apply here, except that approximately 25% fewer errors were made by the reader.

Table XV-3. Reading rates.

<u>Subject</u>	<u>List Number</u>					
	1	2	3	4	5	6
P. S.	61	48	42	38	32	37
T. B.	60	56	43	42	31	36

After two days of practice with the tape reader, P. S. was able to read orally 49 wpm. T. B. read 46 wpm. At these rates, 90% of the words were correctly identified.

After an additional day of practice with silent responses, P. B. read at a speed of 68 wpm, and T. B. read at a rate of 63 wpm. Both subjects thought that they

(XV. COGNITIVE INFORMATION PROCESSING)

were correctly identifying 90% of the words.

Twenty-six Letter Alphabet

The dimensions along which stimuli representing the 26 letters of the alphabet could vary were the 8 auditory dimensions and the 8 tactile dimensions described above. The vowels were either pure tones or pokes on the left or right little finger. (The subjects were told to regard y as a vowel.) No more than two fingers were poked at any one time.

The coding scheme appears in Table XV-4. (The stimulated ear is designated by a capital letter. The fingers are numbered consecutively from left to right.)

Again, subjects worked in pairs; one subject always acted as the typist, the other as the reader. The teams practiced half an hour each day, three days a week. During the first two days, only letters were sent. Thereafter, the few minutes of each day were spent sending letters. The remainder of the time was spent sending

Table XV-4. Twenty-six letter multimodality code.

a: 700 cps, 30 msec, L	n: 0-400 cps, 30 msec, L
b: 1,8	o: 700 cps, 150 msec, R
c: 6	p: 2,7
d: 4	q: 5,8
e: 3 kc, 30 msec, L	r: 10 kc-∞, 30 msec, L
f: 4,5	s: 0-400 cps, 150 msec, R
g: 3,6	t: 10 kc-∞, 150 msec, R
h: 3	u: 1
i: 3 kc, 150 msec, R	v: 2,3
j: 1,5	w: 2
k: 6,7	x: 1,4
l: 5	y: 8
m: 7	z: 4,8

English text. The passages were chosen from Aesop's "Fables," "The Little Prince," "The Reader's Digest," and a daily newspaper. The learning curves appear in Figs. XV-19, XV-20, and XV-21. Common syllables, and short common words were quickly learned by the receivers. These were sent faster than the longer less familiar words. Words whose length exceeded 6 letters sometimes had to be sent two or three times. The reason given by all subjects was that they tended to forget the first letters of the word while receiving the last letters. This difficulty was

Fig. XV-19.
Reading rates—Team 1.

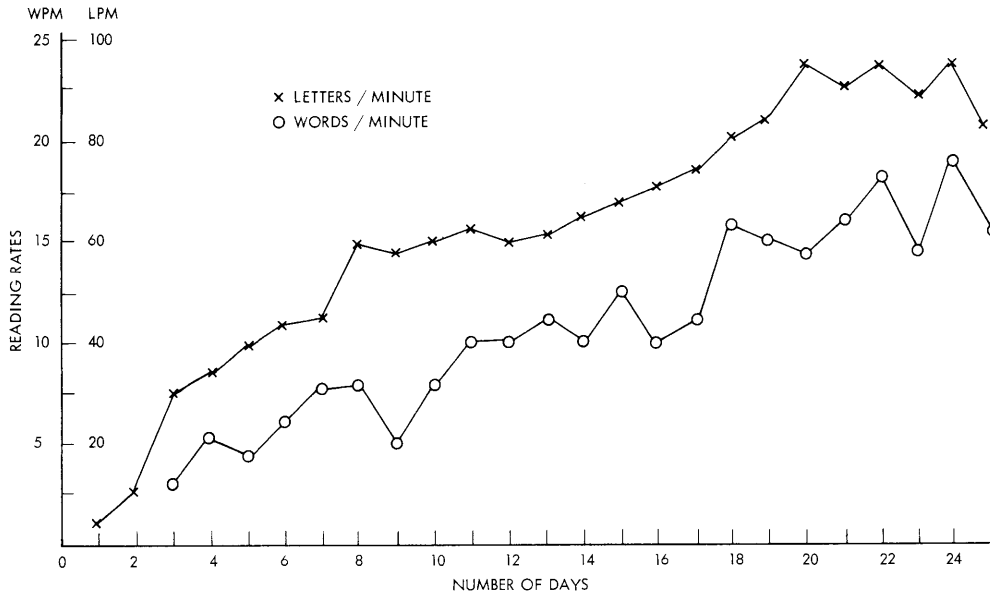
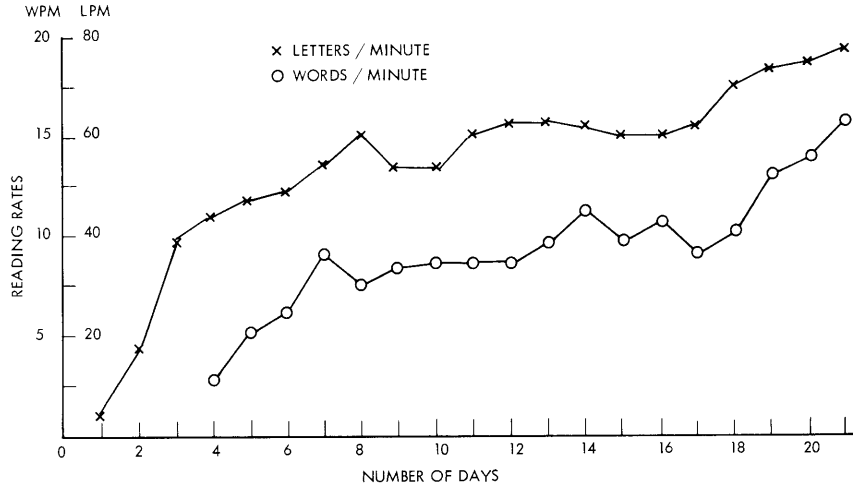


Fig. XV-20. Reading rates — Team 2.

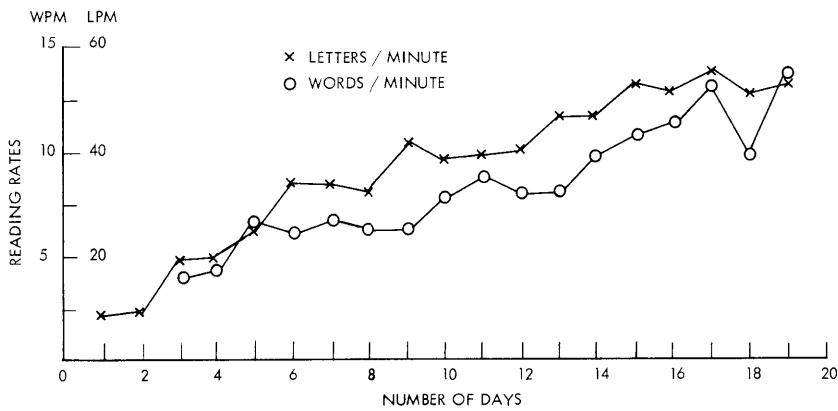


Fig. XV-21.
Reading rates—Team 3.

(XV. COGNITIVE INFORMATION PROCESSING)

partly overcome by having the sender pause between successive syllables of long words.

Discussion

All subjects who learned the twenty-six letter multimodality alphabet stated that words and syllables consisting of both auditory and tactile stimuli felt as natural as words and syllables consisting of single modality stimuli. This result is in accordance with earlier experiments¹ which indicated that stimuli that simultaneously excite more than one sense modality can be identified as easily as single-modality stimuli.

In order to read text at reasonably high rates, it is necessary that the letters be easily distinguished one from another, and that the reader by given practice in identifying letters, words, and groups of words. The relative ease with which subjects learned to identify words made up from letters coded into easily distinguishable stimuli suggests that these two conditions (letter distinguishability and practice) may also be sufficient for high reading rates.

R. W. Donaldson

References

1. R. W. Donaldson, "Information Rates for Multidimensional, Multimodality Sensory Stimuli," Quarterly Progress Report No. 78, Research Laboratory of Electronics, M.I.T., July 15, 1965, 241-248.
2. D. E. Troxel, "Tactile Communication," Ph.D. Thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, August 1962.

