XIX. COGNITIVE INFORMATION PROCESSING*

Prof. S. J. Mason
Prof. W. F. Schreiber
Prof. M. Eden
Prof. T. S. Huang
Prof. D. E. Troxel
Prof. O. J. Tretiak
Dr. P. A. Kolers
R. Baecker
W. L. Black
J. K. Clemens
J. R. Disbrow
R. W. Donaldson
J. K. Dupress
D. N. Graham
N. Gramenopoulos
G. S. Harlem
K. R. Ingham
E. Landsman
P. Mermelstein
N. D. Strahm
D. Thornhill

A. PERCEIVED ORDER IN DIFFERENT SENSE MODALITIES

Use of more than one sense has been suggested as a means of increasing the rate at which information can be assimilated by persons having a sensory handicap. One of the first necessary steps is to determine how far apart in time stimuli applied to different sense modalities must be in order for a subject to ascertain which of the two stimuli occurred first. Experiments by Hirsch and Sherrick have shown that the probability of correctly determining the order of two stimuli versus the time interval between their onset is Gaussian, with zero mean and standard deviation of 20 msec.[1,2] Thus, 20 msec is required for judgments having 70 per cent accuracy, 30 msec for 90 per cent accuracy, and 40 msec for 95 per cent accuracy. These results are independent of whether the stimuli consist of lights at two different locations, tones of two different frequencies, identical tones applied to different ears, buzzers that excite similar fingers on different hands, or two stimuli chosen from these pairs and applied to different sense modalities. When tones applied to different ears also differ in frequency, however, only 30 msec is required for 99 per cent correct judgments. Increasing the dimensionality of auditory stimuli reduces the time required to correctly detect stimulus order.

[When stimuli occurring at different places excite the same sense, fusion effects can occur if the stimuli are sufficiently close temporally.[3] For example, if tones of the same frequency are separated by approximately 1 msec and applied to different ears, the tones are perceived as simultaneous. However, the image in the head is displaced from the center toward the ear that received the earlier tone. Since this discussion pertains to tones that are perceived as nonsimultaneous, fusion effects are ignored.]

A pilot experiment in which a poke probe was applied to the index finger and a 1-kc tone was applied to the ears has yielded results very similar to those obtained by Hirsch and Sherrick. The durations of both stimuli were equal. Durations of 50 msec and 500 msec were used, and intensites were well above threshold. The stimulus duration had little effect on the perceived stimulus order.

Research on the use of more than one sense modality to convey information is continuing.

R. W. Donaldson

*This work was supported in part by the National Science Foundation (Grant G-16526), the National Institutes of Health (Grant MH-04737-03), and the National Aeronautics and Space Administration (Grant NsG-496).

References

1. I. J. Hirsch, Auditory perception of temporal order, J. Acoust. Soc. Am. 31, 759-767 (1959).

2. I. J. Hirsch and C. E. Sherrick, Jr., Perceived order in different sense modalities, J. Exptl. Psychol. 62, 423-432 (1961).

3. G. von Békésy, Similarities between hearing and skin sensations, Psychol. Rev. 66, 1-22 (1959).

B. ELECTRONIC COMBINATION LOCK

Figure XIX-1 shows an elementary electronic combination lock, a well-known circuit. In order to connect the output to the voltage source the switches must all be set left, then switched to the right sequentially in the order S_1, S_2, S_3, S_4. If the switches are moved in any other order, no voltage occurs at the output. Thus, although only one relay is employed, the system must pass through five states before an output voltage is obtained. Clearly, with no relay it is impossible to build a sequential lock.

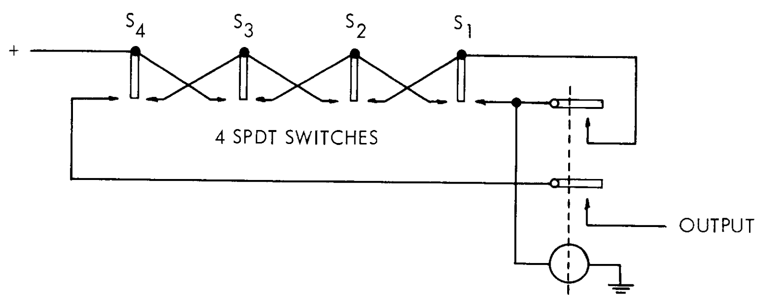

Fig. XIX-1. An elementary electronic combination lock.

With four switches 16 states exist. Is it possible to build a lock with one relay that requires all 16 states to be traversed before an output voltage is generated?

An affirmative example is provided by the circuit of Fig. XIX-2. The relay must be fast enough to drop out in the time it takes a toggle switch to transfer from one side to the other. (I constructed a model using a transistor flip-flop instead of a relay; thus speed was no problem.)

To produce an output voltage:

(i) Set all switches down.

(ii) Flip switches according to the following two rules:

(a) Never flip the switch just flipped.

(b) Flip a switch only if its neighbor to the right is up (or if it has no neighbor to the right) and all other switches to the right are down.

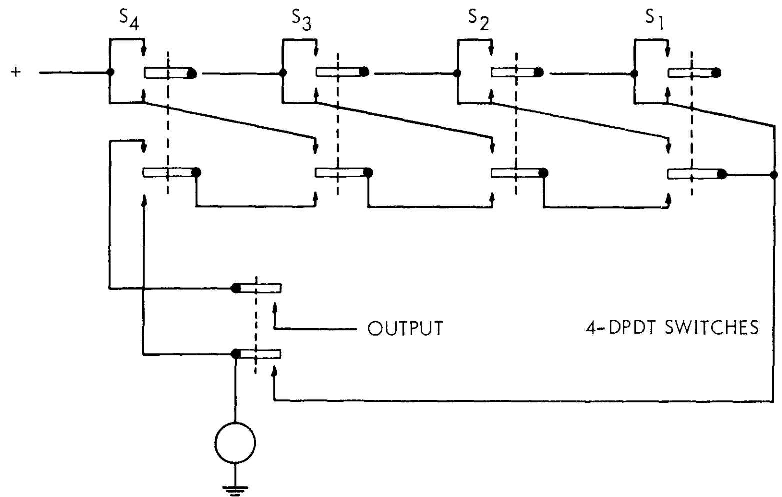

Fig. XIX-2. A complex combination lock.

These rules guarantee that all sixteen states are traversed. In fact, if an up switch is represented by a 1, and a down switch by a 0, the sequence of states is the common reflection gray code. This combination lock is isomorphic to an ancient Chinese ring puzzle. Note that if at any time in the sequence an error is made, the sequence must be resumed at the beginning.

W. L. Black

C. MOBILITY AID SIMULATOR

A survey of research on mobility aids for the blind shows that a great deal of work has been done but progress has been extremely slow. The difficulty is that the development of a specific system requires several years of work before the true usefulness of the scheme can be evaluated. Any nontrivial changes then introduce further delay before a new evaluation of usefulness can be undertaken.

This report describes the development of a Mobility Aid Simulator facility that will be capable of simulating the performance of a wide class of hypothetical sensors and navigation schemes and in which even major changes in the mode of operation can be made simply and quickly. The major components of the simulator facility are a room that serves as an obstructed environment, a box of reasonable size that is the dummy mobility aid carried by the test subject, position-monitoring instrumentation, and a general-purpose computer, probably the TX-0 computer.

The system will simulate the operation of the proposed navigation scheme by using present box location and orientation, and previously stored information about object locations within the test room according to a computer program describing the performance of the navigation scheme.

For example, to simulate an ideal narrow-beam range finder, the computer will determine the distance from the dummy sensor to the nearest object along the "line of sight" of the dummy sensor, and then transmit this information to the subject by means of an audible or tactile display.

As the subject sweeps the dummy sensor through a succession of orientations, the computer will continue to update the range information, so that the subject will experience in real time the performance to be expected from the system being simulated.

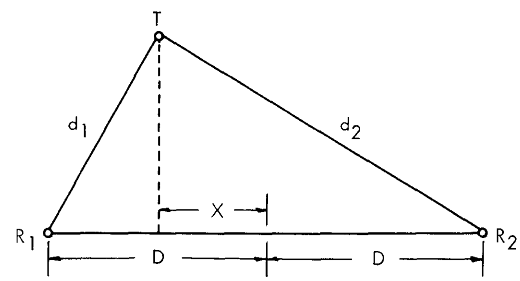

Fig. XIX-3. Coordinate measurement. (T = transmitter; R = receiver; $x = \left(d_2^2-d_1^2\right)/4D$.)

Position-monitoring instrumentation is being constructed and evaluated. The rectangular position coordinates of an ultrasonic pulse sound source is continuously monitored. Figure XIX-3 shows the position coordinate that can be calculated as a difference of two squares.

A block diagram of the analog circuit used for calculation of one position coordinate is shown in Fig. XIX-4.

The ramp generator supplies positive and negative ramps starting when the ultrasonic pulse is transmitted. If d_1 occurs before d_2, a gated positive ramp as shown in Fig. XIX-5a is integrated. If d_2 occurs before d_1, a gated negative ramp is integrated as shown in Fig. XIX-5b.

The output is sampled and held after the integration. The output is then proportional to $d_1^2 - d_2^2$, including the sign. Three coordinate measurements will require four receivers placed in corners of the test room. A second transmitter separated from the first by approximately 1 foot, the position of which is also continuously monitored, will enable calculation of the orientation of the dummy mobility aid.

The circuit that measures distances d_1 and d_2 responds to the first received pulse after the transmitter is triggered. If a reflection occurs within the room such that it can trigger the receiver falsely from a previous transmitted pulse, the position measurement is rendered useless. The test room is therefore being treated with special

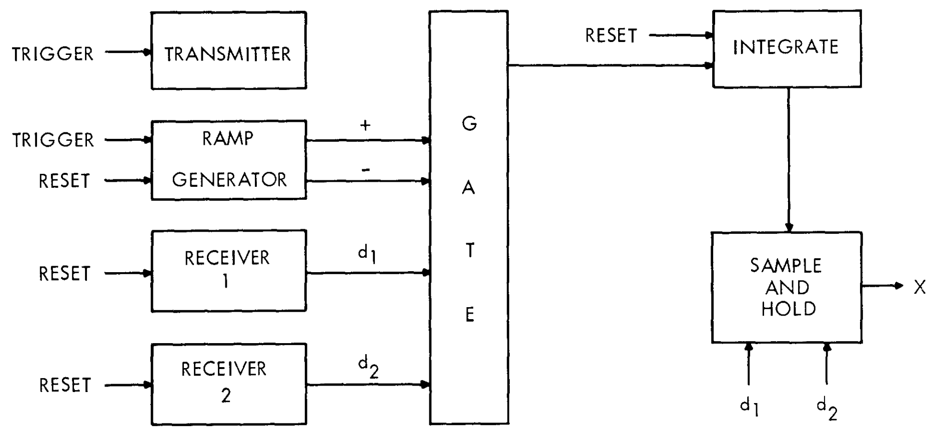

Fig. XIX-4. Calculation of x.

sound-absorbent material. The accuracy of the position measurement depends on how quickly the receivers indicate reception of a pulse. The signal is 40 kc, and the circuits trigger within approximately 2 cycles or 50 msec. This corresponds to approximately $\frac{1}{2}$ inch. Errors in the analog calulation circuit reduce to zero when the transmitter is located on the perpendicular bisector of the line between the two receivers.

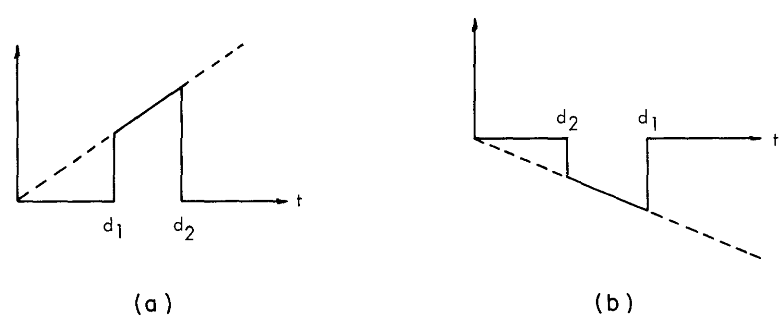

Fig. XIX-5. Integration of a gated ramp. (a) Positive ramp, $d_1 < d_2$. (b) Negative ramp, $d_2 < d_1$.

The ultrasonic transducers being utilized here are of the type widely used for remote control of home television receivers.

E. Landsman

D. EXPERIMENTS ON MACHINE RECOGNITION OF CONNECTED HANDWRITTEN WORDS

The present report, together with our previous reports,[1,2] summarizes work that was submitted to the Department of Electrical Engineering, M. I. T., January 14, 1964, as a thesis in partial fulfillment of the requirements for the degree of Doctor of Science.

The handwriting-recognition experiment reported in Quarterly Progress Report No. 71 (pages 257-265) was repeated on a set of 32 different words generated from the complete lower-case alphabet written by four different subjects and totalling 254 words. The set of words was chosen randomly and therefore reflected roughly the letter distribution in the language, but contained a minimum of two occurrences of each letter. Diacritical marks were omitted and subjects were asked not to cross the letters t and x where it necessitated a break in the continuity of the writing. A dictionary of the 10,000 most frequently used English words[3] was used to limit the recognition results to words of the language.

A total of 249 words was successfully segmented into stroke sequences. In the other five cases, as a result of the smoothing of the normal direction changes in certain contexts under conditions of rapid writing, the last upstroke-downstroke pair of a letter and the following ligature were found to be inseparable by the segmentation algorithm used, as illustrated by writing  instead of . In two other cases the words, as segmented by the program, were not recognizable with the aid of downstrokes alone because the last downstroke of the last letter of the word was not explicitly executed, e.g., .

Recognition was attempted on the remaining 247 words by using a stoke representation of the letters based on downstrokes only and compiled from the strokes constituting those words. The stroke partition utilized consisted of 22 downstroke categories, one of which corresponded to downstrokes found at the beginning and end of words and did not form part of the first or last letter, and was therefore assigned to the null letter. The stroke classification was carried out by using all 12 computed parameters and treating them as if they were independent. Eighty per cent of the words on which recognition was attempted was correctly identified. Of the 49 samples incorrectly identified, 14 had the correct word selected as the second choice, 7 were selected as choices lower than the second, 26 did not give the correct word as one of the possible 20 choices in the threshold range, and recognition of 2 samples had to be terminated when no result was obtained after 20 minutes of processing.

The experimental strategy for word recognition, namely, repeated attempts with successively lower stroke-likelihood thresholds, does not eliminate the possibility that if some lower initial threshold setting were used, a previously incorrect decision might be performed correctly or a correct decision might be upset. In order to observe the frequency of this phenomenon, 12 word samples for which the correct word was initially not selected were reprocessed by using a doubled value for the initial threshold setting. Six of these words were now correctly selected as first choices in the recognition output. The reason for this improvement in performance is that in certain cases the correct category may lie just beyond the likelihood threshold value and therefore be missed, while all of the other strokes are correctly recognized. We may, of course, process all samples with the higher threshold value initially, but this frequently results in an unwarranted sizable increase in the required processing time for word recognition.

Next, independent stroke statistics were computed for each subject, and recognition was attempted on the previously incorrectly recognized samples by using representations derived in each case from the test subject's handwriting. Forty-three of the previous 49 errors were now correctly recognized. Since 40 of the 198 samples previously correctly recognized were correctly recognized now as well, and none was misrecognized, we may assume that all of the 198 samples would have been correctly recognized, thereby resulting in a subject-dependent recognition rate of 98 per cent. Time requirements for recognition in this experiment fell to an average of 9 sec on the IBM 7094 computer.

Our results show that recognition reliability approaching that of humans is attainable for carefully written script if the machine's representation is based on the writing of the same subject who also produced the samples to be recognized. When the machine is not given a representation originating from the same subject, but one based on an ensemble written by a number of individuals, samples from that group of individuals are recognizable, but with lower reliability. Recognition performance deteriorates further when we attempt to recognize writing samples from subjects not included in the ensemble from which the machine's representation is derived.

Our simulation experiments demonstrate that good recognition is obtainable by means of the downstrokes of the writing only. There exist a number of ambiguities that are in general difficult to resolve by means of downstrokes alone, such as differentiation between the letters p and k or the pair cl and the letter d. It is suggested that use should be made of upstrokes as well only in environments giving rise to such ambiguities, and not otherwise, thereby reducing the complexity of and time requirements for the recognition task. The most important deficiency of the program at the present time appears to be the lack of facilities for modifying stoke likelihood decisions on the basis of the stroke environment. Stroke recognition based on information pertaining only to individual strokes is found frequently to be inadequate in low contextual information situations, e.g., differentiation between the words fall and full. It is proposed that in such cases the difficulties, which may in most cases be recognized by a likelihood ratio between the two choices which is nearly one, be resolved by further information to be supplied by specific demand from the recognizer. The retrieval of this information may be implemented by detecting the specific letter-letter difference between the top choices and abstracting the information applicable to the resolution of that particular ambiguity.

P. Mermelstein

References

1. P. Mermelstein, Study of the handwriting movement, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M. I. T., April 15, 1963, pp. 229-232.

2. P. Mermelstein, Experiments on machine recognition of connected handwritten words, Quarterly Progress Report No. 71, Research Laboratory of Electronics, M. I. T., October 15, 1963, pp. 257-265.

3. E. L. Thorndike and I. Lorge, The Teacher's Word Book of 30, 000 Words (Columbia University Press, New York, 1944).