XXVI. COGNITIVE INFORMATION PROCESSING*

Academic and Research Staff

Prof. S. J. Mason
Prof. W. L. Black
Prof. M. Eden
Prof. T. S. Huang
Prof. B. Prasada
Prof. O. J. Tretiak
Prof. D. E. Troxel
Prof. D. Cohen
Dr. M. P. Beddoes
Dr. P. A. Kolers
Dr. N. Sezaki
K. R. Ingham
G. L. Wickelgren

Graduate Students

A. K. Bhushan
J. D. Bigham, Jr.
D. Caldwell
A. L. Citron
R. W. Cornew
A. Gabrielian
B. P. Golden
G. S. Harlem
P. H. Hartmann
M. Kawanami
E. E. Landsman
M. B. Lazarus
F. F. Lee
J-H. Liu
A. Neuberger
J. A. Newell
L. C. Ng
D. H. Pruslin
C. L. Seitz
S. D. Shoap
A. Spiridon
J. A. Williams

RESEARCH OBJECTIVES AND SUMMARY OF RESEARCH

The primary research interest of this group is in the real-time acquisition and processing of visual information for display to the visual and nonvisual senses, and in the psychology of human utilization of such information, for both communication and control. The motivation is an interest in human capabilities for information processing and in human information requirements. Applications include sensory-aids systems for the blind and the blind-deaf, picture-transmission systems, and special information-display systems for enhancement of human performance under conditions of stress.

Major projects now in progress include studies on reading machines, mobility-aids, picture processing, and pattern recognition.

1. Reading-Machine Studies

Results during the past two years from a number of research investigations, each concerned with some facet of the reading-machine problem, now point the way toward the development of an integrated real-time reading-machine system that will serve as an invaluable facility for further research on character recognition, special-purpose picture processing, tactile pattern perception, and auditory displays, including artificial speech. The system will also constitute a prototype for possible practical reading machines for the blind.

The system will include an opaque scanner for the sensing of printed matter, real-time data-processing facilities for real-time scan control, for operating upon the data, for controlling tactile and auditory displays, and for monitoring and analyzing the performance of human subjects. In its most sophisticated mode of operation, the system will recognize printed characters and generate artificial speech. A less sophisticated output mode will utilize Braille for tactile sensing and spelled speech for auditory reception. In its simplest mode of operation, the system will simply reproduce, as tactile

*This work is supported through the Joint Services Electronics Program by the U. S. Army Research Office, Durham, under Contract DA 36-039-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).

patterns, the black-white shapes sensed by the scanner, without automatic character recognition. Between the simplest and most sophisticated modes, there lies a spectrum of intermediate modes. In particular, information may be extracted and displayed at any intermediate step of a character-recognition procedure, leaving the final steps to the human user.

Perhaps the most interesting and important aspect of this research will be the studies of human learning ability and performance when subjects are provided with a flexible real-time information acquisition and display system and taught how to use it effectively.

Character Recognition. A character-recognition procedure for the identification of letters and numbers in newsprint has been developed.[1,2] The approach is based upon automatic tracing of the outer black-white edge of each letter. After nonlinear smoothing, of the "gear backlash" type, the horizontal and vertical position-coordinate waveforms are processed to detect the sequence of horizontal and vertical changes of direction, together with low-resolution information on the spatial locations where such changes occur. The result is a 30-bit code word from which the letter is identified by look-up in a stored list, in which there may be several different code words associated with the same letter. The system design was not aimed at ultrahigh data rates and microscopic error rate, but rather at human reading rates, a humanly comfortable error rate, and economical simplicity of logic and storage. With only 3000 bits of look-up table storage, produced by "training" on one page of news-magazine text, the system "read" the next page with an error rate of only a few per cent, so that the output was intelligible. Most of the errors were either previously unencountered characters or one of a small set of idiosyncratic letter-pair confusions that are easily resolved by the human reader. Reduction of the error rate to 1 per cent, or less, is fully expected on the basis of obvious improvements.

Opaque Scanners. The character-recognition work thus far has been carried out by photographing the printed page and then transmission-scanning a photographic transparency. An opaque (reflected-light) scanner, suitable for direct real-time acquisition of data, is under development. A digital-deflection, cathode-ray tube, flying-spot scanner will be utilized.

Tactile Displays. Two tactile displays are under development, both of which are two-dimensional arrays of Braille-like tactile "dot" elements, in which each dot is either present or absent, according to external control. The first display, made up of solenoid-controlled poke probes, has a size of 7×7 and a frame rate (tactile-pattern replacement rate) of several per second. The second display is of the "moving-belt" variety with size 64×128, and a frame rate of three or more per minute. These displays will be used as tactile outputs in the reading-machine system and also for general studies of tactile pattern perception, both dynamic and static.

Phonetic Translation of English Text. The problem of automatic translation from letter spelling to phonetic spelling in English text has been studied.[3] A procedure has been developed, based upon a "dictionary" of approximately 30,000 root words and a set of combination and generation laws, such that the correct pronunciation (phonetic spelling) of 250,000 English words can be automated. The results of this study provide new and useful information on the detailed properties of the English language and also specify the design of a practical system, utilizing read-only optical storage, for phonetic translation.

Artificial-Speech Generation. The general research problem of artificial speech generation is under continuing investigation in Professor K. N. Stevens' Speech Communications Group and particular application to the reading-machine system is being considered in cooperation with his group. Recent work in this country and abroad[4-7] indicates that a real-time phoneme-sequence-controlled speech generator can be made to perform satisfactorily, provided speech-formant control signal transitions are properly governed by the phoneme sequence, in accordance with certain programmable rules.

Errors in phonetic spelling, arising from erroneous character recognition and resulting in phonetic nonwords, can be accommodated by spelled-speech display of the erroneous English word. Spelled speech is also a reasonable possibility for phonetically intractable proper names.

Spelled-Speech Storage, Display, and Compression. Spelled speech offers a slower but simpler auditory display than artificial speech. Under consideration are magnetic and optical storage and read-out of the spoken-letter sounds, and under investigation is the compression of spelled speech by sampling in sections, temporal truncation, and temporal-overlap superposition.

Effects of Spelling Errors upon Reading Performance. In the early stage of investigation is the problem of reading performance in the presence of spelling errors, for visual reading, grade-one Braille tactile reading, and spelled-speech auditory reading. Our present plan is to utilize imbedded-question text for a running measure of comprehension, and to allow this performance measure to control the presentation rate, the error signal being the deviation of performance from some reference level. The experimental system will also serve as a teaching machine for the training of subjects in tactile and auditory codes.

Picture Simplification. The nonlinear processing of pictures, for elimination of all features other than principal edges, is being investigated. The aim is to simplify pictures to the extent that tactile display and perception of the principal edge patterns becomes practical. Edge extraction, which involves processes related to spatial differentiation of the picture intensity function, is a nontrivial operation, especially for noisy pictures. Edge extrapolation and smoothing, which are desirable operations from the standpoint of noise elimination, place the problem in the realm where general pattern-recognition considerations become important.

Real-time Data-Processing Facilities. Work is proceeding on the development of real-time remote input-output subsystems that will allow two small general-purpose digital computers to be utilized simultaneously and cooperatively in the reading-machine system, one computer being mainly occupied with phonetic translation and artificial-speech control-signal generation, and the other with scan control, character recognition, and the operation of nonspeech displays.

Although general-purpose computers will be employed in the experimental system for desirable flexibility, all of the data processing is of such a character that realization by means of much more economical hardware is a future possibility. Such a system may lead to practical reading machines for the blind and the blind-deaf.

2. Mobility-Aid Studies

The problem of human mobility in an obstructed environment, under conditions of limited information availability, falls into the realm of automatic control system dynamics. With respect to the interest of this group in human communication and control, the reading-machine studies may be classified as communications, and the mobility-aid studies as control. There are, of course, close relationships between these areas.

A simulation system is being developed for the study of human information requirements for human mobility in an obstructed environment under conditions of limited availability of visual information. A recent result is the successful real-time operation of a monitoring system for continuous tracking of two coordinates of a point (a small ultrasonic sound source) in three-dimensional space. The monitoring system is necessary for simulation of the operation of radarlike guidance devices because the position of the dummy device must be known in order to predict the interaction of that device with obstructions in the environment. A novel characteristic of the monitoring system is that it utilizes acoustical receiving antennas designed for a linear relationship between transmitter-to-receiver transmit times and rectangular coordinates of the moving transmitter.

The simulation system will make major use of the same real-time data-processing facilities that are being developed for the reading-machine studies.

A principal use of the system will be the simulation of mobility-aid systems, enabling a sightless subject to experience, in real time and real space, the operation of hypothetical mobility aids. The monitoring and data-processing capabilities of such a facility would provide flexibility and control in basic experimentation on human information requirements, and would enable optimizing the performance of specific systems before major investment of effort in realization of the complete system.

A mobility aid is a device or system to help a sightless person detect obstacles, locate "step-downs" and "step-ups," identify landmarks, or otherwise facilitate his travel in either familiar or unfamiliar environments. Like every sensory aid system, a mobility aid can be visualized as consisting of a sensor or pickup, a processor, and a display or stimulator. The sensor may be either active (ultrasonic transmitter and receiver, infrared projector and photosensitive pickup) or passive (a lens backed by masks and photosensors, a television camera, a light meter).

Several mobility aids have been devised (e.g., by Kay and Pye in England, and Witcher, Benham, Russel, Melkin, Bliss, and Kallman in this country) but none has come into common use by the blind population.[8-12] The difficulty is that the development of a specific system usually requires several years of work, only after which the true usefulness of the scheme can be evaluated. Moreover, any nontrivial changes in the system design then introduce further delay before a new evaluation of usefulness can be undertaken.

We propose to develop a facility that will enable simulation of the performance of a wide class of sensory-aid systems. Operation of the simulator is to be as follows: The sightless subject walks about in an obstructed environment, carrying a dummy mobility aid. The dummy sensor can be moved freely but its position and orientation are continuously monitored by a tracking system.

The dummy-sensor orientation and position data are sampled at frequent time intervals and the current data are stored in a digital computer. Information about the positions and shapes of obstacles in the environment is also stored in a rapid-access memory as fixed data. The subject is outfitted with a stimulator or display (tactile or auditory) which is controlled by the computer. Finally, the computer is provided with a mathematical model, in the form of a program, describing the performance characteristics of a hypothetical sensor-processor combination.

The ability to monitor the position and orientation of a hand-held dummy sensor will, of course, permit conducting many other investigations of interest, including detailed studies of unaided mobility, as well as studies of more sophisticated processing between existing sensors and displays.

3. Picture-Processing Studies

Picture-processing research is concerned with the processing and encoding of picture information for efficient or economical transmission, subject to the requirements of satisfactory subjective quality in the reconstructed received picture.

Objective Measures of Subjective Picture Quality. Our long-range goal is to develop a mathematical measure of picture quality that approximates the quality judgment made by human observers. The general problem is difficult, but special cases help provide entries to the general case. Special studies of interest include subjective effect of noise in pictures that have been transmitted through noisy digital channels, subjective effect of additive white noise as dependent upon the probability density function of the noise, quality of lowpass pictures as a function of the shape of the lowpass filter, optimum amplitude quantization, and relation of noise perception to the two-dimensional sine-wave response of the eye.

Processing of Motion Pictures. Interframe constraints in motion pictures offer the possibility of bandwidth compression superior to that obtainable from frame-by-frame processing. The technical problems of simulating various motion-picture processing schemes are being considered. The associated psychological studies would be concerned with apparent motion, fusion, flicker, and dynamic resolution requirements of the human observer.

Optical-Image Processing. Our principal interest is in using optics as a tool for the processing of pictures. Certain operations, such as two-dimensional spatial filtering, can be readily accomplished, in principle, with coherent-light optics. In order to use the tool effectively, however, its capabilities and limitations must be known. Problems under consideration include: the effect of film-grain noise on the performance of a coherent optical system; the relation of film thickness and exposure; techniques for the making of spatial filters; and the effect on the reconstructed picture of various operations (such as sampling, quantization, noise addition) upon the hologram.

4. Pattern-Recognition Studies

"Pattern Recognition" is a common theme underlying the projects described above, whether the task is accomplished by human cognition, by computer data processing or, most commonly, by a combination of the two.

Complementary to the task-oriented projects, the following research studies in the general area of pattern recognition are being carried out.

Psychophysics of Depth Perception. This project is designed to provide information on a subject's ability to make accurate distance and orientation judgments with respect to real and illusory three-dimensional objects. The variation of such judgments will be studied without feedback, as well as with a continuous auditory or visual error signal.

Effects of Spatial Transformation of Text upon Reading and Writing Performance. It has been shown that systematic transformations or printed text (such as inversions or rotations of letters and of words) produce regular changes in the learning rate, as well as in the asymptotic reading rate. These studies are being extended to a wider class of transformations and also to reading and writing performance with cursive script used as the text.

Studies of Apparent Motion. It appears that one of the differences between the perception of apparent motion and real motion is an alteration in the psychological space metric. Experiments are being conducted to verify this observation.

In addition to these psychological studies specific pattern-recognition problems are being investigated to determine the extent to which the recognition tasks can be automated.

Automatic Extraction of Information from Printed Sheet Music. With the aid of the scanner and ancillary equipment developed for the projects described earlier, it is possible to program a computer to determine note value, duration, phrasing, and certain other music indicators from ordinary printed sheet music. This study is being continued to see what additional parameters can be extracted by automatic recognition.

Recognition of Cloud Types in Satellite Photographs. This will be a new project.

Recognition of Plant Species from Leaf-Vein Patterns. We begin this study under the assumption that an algorithm can be developed which will generate a venation pattern in which certain stochastic features perturb an essentially deterministic process. The parameters of the algorithm could then be regarded as the features serving to classify the leaf types according to species.

S. J. Mason

References

1. J. K. Clemens, "Optical Character Recognition for Reading Machine Applications," Ph. D. Thesis, Department of Electrical Engineering, M. I. T., September 1965.

2. H. Teager, "Multidimensional Visual Information Processing," Conference on Advances in Biomedical Computer Applications, New York Academy of Science, New York, June 3-5, 1965.

3. F. Lee, "A Study of Grapheme to Phoneme Translation of English," Ph. D. Thesis, Department of Electrical Engineering, M. I. T., October 1965.

4. J. N. Holmes, I. G. Mattingly, and J. N. Shearme, "Speech Synthesis by Rule," Language and Speech, Vol. 7, Part 3, pp. 127-143, July-September 1964.

5. L. J. Gerstman and J. L. Kelly, "An Artificial Talker Driven from a Phonetic Input," J. Acoust. Soc. Am. 33, 835(A) (1961).

6. S. E. Estes, H. R. Kerby, H. D. Maxey, and R. M. Walker, "Speech Synthesis from Stored Data," IBM J. Res. Develop., Vol. 8, pp. 2-12, January 1964.

7. R. S. Tomlinson, "SPASS – An Improved Terminal-Analog Speech Synthesizer," a paper presented at the Seventieth Meeting of the Acoustical Society of America, November 3-6, 1965.

8. J. A. Leonard and A. Carpenter, "Trial of an Acoustic Blind Aid," NBS Research Bulletin No. 4, January 1964, p. 70.

9. C. W. Shilling, "Identification and Teaching of Auditory Cues for Travelling in the Blind," Final Progress Report, Auditory Research Center, Inc., December 1, 1959-March 1, 1963.

10. P. Worchel, D. Byrne, and R. Young, "The Evaluation of the Haverford Bionic Instruments Obstacle Detector," Summary Report to Verterans Administration by Tracor, Inc., November 27, 1963.

11. Proceedings of the Conference on Mobility for the Blind, held in Rotterdam, The Netherlands, August 3-7, 1964.

12. "Evaluation Report on Work in Progress on Sensory Aids and Prosthetics," Report No. 9211-2, Engineering Projects Laboratory, M. I. T., April 1964.

A. COGNITIVE PROCESSES

1. PHONOLOGICAL CODE-SWITCHING TIME IN BILINGUALISM

Bilingualism as a phenomenon is particularly interesting to students of human information processing because it provides an opportunity to ask a number of questions about information storage, access, and retrieval. It is sometimes the case, for example, that particular kinds of knowledge or skill are expressible only in one of a bilingual's two languages; while in other cases the information that he has can be expressed in either of them.[1] That is, certain skills and some knowledge are coded in linguistically tied forms, and other kinds of information may be in some linguistically free or metalinguistic form. Thus it appears that the nervous system codes different kinds of information and skill in different ways, some of which limit the individual's access to the information.

Consistent with these differences is the naturalistic observation that when a bilingual

speaks, his use of one of his two languages reveals certain system properties. While intrusions from the second language will occur during the use of the first, the intrusions rarely are random; instead, they tend to exhibit a grammatical or thematic regularity. The intrusions are reported to occur at major syntactic boundaries, or to introduce concepts that are more readily expressed in one language than the other.[2] Thus, individual words from each of a bilingual's two dictionaries are not substituted for each other in haphazard fashion, but are made to conform in the way they are used to the major organizing or system properties of the languages of which they are members.

It would be interesting to learn something detailed about the way the bilingual typically switches between his languages, for this would shed some light on the way information is actually processed for output. One cannot learn very much from naturalistic observation, however, partly for the reasons having to do with the systemlike properties of the use of language; and partly because in free speech the bilingual actually is processing his information along syntactic, semantic, and phonological lines in very close temporal sequence. No means are now known for measuring the time taken to go from conceptualization to production that would enable us to study separately the various contributions of syntax, semantics, and phonology. One can, however, drive a wedge into these processes that enables us to get some approximations about phonology. This report describes an experiment in which we succeeded in measuring a phonological code-switching time by having bilinguals read aloud. The text that they read sampled haphazardly from their two dictionaries; in order for it to be read with some fidelity to the accent of the language, the subjects had to switch back and forth between an English and a French-generating phonological system.

An example of the kind of text that they read is the following selection which preserves French word order.

Son cheval, suivi by two hounds, en marchant d'un pas égal, made resound the earth. Drops of ice se collaient à son cloak. A wind strong soufflait. Un côté of the horizon s'éclaircit; et, in the whiteness de crépuscule, he saw des lapins sautillant au edge of their burrows. Les deux bassets suddenly se précipitèrent on them; and ça et là, quickly of them broke the backs. Soon he entra dans un bois. At the end d'une branch a coq de bruyère, engourdi by the cold, slept, la tête sous its wing. With a stroke d'épée, il lui faucha the two feet and without le remasser went on his way. Three hours later, il se trouva sur la pointe d'une mountain.

Twelve subjects were tested, six Americans who knew French and six Europeans whose native language was French. All were students at either the Massachusetts Institute of Technology or Harvard University. Their task was to read aloud as rapidly as they could. The passages that they read were all approximately 110 words long; some passages were French or English; some were alternating, alternate sentences appearing in each of the two languages; and some were mixed, half preserving French word order (as in the example) and the other half English word order. Thus two pages of alternating

or mixed text have approximately the same number of words in French and English as one page of unilingual French and one of unilingual English. The texts and subjects were run according to a Latin square so that any given passage (in the sense of the information it contained) was read in a different form by different subjects.

The results of the reading tests are shown in Table XXVI-1. Material in the native language is read most rapidly and mixed text most slowly. Passages of alternating sentences are read at a speed midway between the speeds for unilingual French and English.

Table XXVI-1. Time required to read aloud passages in various linguistic forms.

	Unilingual		Alternating		Mixed	
	English	French	English[a]	French[a]	English[b]	French[b]
	(seconds)		(seconds)		(seconds)	
Americans	30	47.4	38.4	39.6	49.2	51.6
Europeans	41.4	30	36.6	36	45	45
Average, Americans	38.5		39		50.4	
Average, Europeans	35.7		36		45	

[a]Language of the first sentence.
[b]Favored word order.

Since the same number of words in each language appears in two pages of text of each form, averages for unilingual, alternating, and mixed forms reflect time taken to read identical amounts of material. The averages are also shown in Table XXVI-1. Clearly, the need to switch a phonological system slows the reader down considerably when he is reading mixed text. If we take the difference in time for mixed and unilingual passages and divide it by the number of transitions in mixed text between French and English words, separately for each subject (that is, $(t_m - t_u)/X$), we can estimate a phonological code-switching time. For subjects reading aloud as rapidly as they can, these times average out to approximately 0.33 second per switch, with ranges of 0.24-0.66 sec among the Americans, and 0.12-0.48 sec among the Europeans. While there are 29.5 transitions in mixed text, there are only 4.5 in alternating text, so that we cannot compute a code-switching time for alternating text confidently with this small number.

We can also compute the cost of a switch, the amount of unilingual material that could have been read if no switches were made. This is given by the average reading rate in words per minute for unilingual French and English multiplied by the switching time,

$\delta(r_e+r_f)/2$, and is equal to 1 + words.

P. A. Kolers, P. Perrolle

References

1. P. A. Kolers, "Specifity of a Cognitive Operation," J. Verb. Learn. Verb. Behav. 3, 244-248 (1964); "Interlingual Word Associations," ibid. 2, 291-300 (1963).

2. A. R. Diebold, Jr., "Code-Switching in Greek-English Bilingual Speech, Georgetown University Monograph Series on Language and Linguistics Number 15, 1963, pp. 53-62.

2. GRAPHEME-TO-PHONEME TRANSLATION OF ENGLISH

An efficient scheme has been developed for the automatic translation of English text from letters to phonemes. This process of translation is an indispensable part of a reading machine that produces synthetic-speech output.

An investigation of the possibility of generalization of phonic rules for the translation process was made. It was found that the interaction of word final mute 'e' and vocalic suffixes requires the use of both complex rules and large exception lists. With the phonic approach one is unable to handle mute 'e' in medial positions in compound words. The handling of stress would require additional rules that depend on the determination of parts-of-speech through the process of suffix identification.

The lexicon look-up approach was next looked into. By a judicious choice of elements for the lexicon and the use of a set of very simple algorithms it is possible to decompose a word in an orderly fashion into the stored elements from which the phonemic translation can be found. A lexicon containing approximately 32,000 elements is sufficient for the decomposition of all entries in the Seventh Edition of Webster's "New Collegiate Dictionary." Additional markings in the lexicon provide the basis for the resolution of syntactical ambiguities which cannot be handled by the phonic approach.

F. F. Lee

B. PATTERN RECOGNITION

1. AUTOMATIC RECOGNITION OF SHEET MUSIC

The recognition of sheet music is being studied as an example of pattern recognition in partially structured pictures. "Partially structured" refers to a picture composed of well-defined forms or "characters" and of forms that are only partially defined. The need to handle both types of forms and to deal with the relations between a number of forms simultaneously makes this problem more pictorial in nature than character recognition alone. The scheme that is being utilized includes nonlinear background elimination, contour tracing, and feature extraction. The following progress has been made.

A simple flying-spot scanner for binary (2-level) pictures has been developed in our laboratory. The desired scan is programmed on the cathode-ray tube of the TX-0 computer. The scanning point of light is focused on a transparency and the transmitted light is detected by a photomultiplier tube, and applied to a threshold circuit. The output of the circuit triggers the light gun flip-flop of the TX-0 if the photomultiplier output is on the proper side of the threshold. A core image of the transparency is obtained by checking the flip-flop and packing away ones and zeros.

By utilizing this scanner, pictures of music were scanned into the TX-0 and written on magnetic tape in a format compatible with FORTRAN. Nonlinear vertical and horizontal defocusing (background elimination) were effected on these pictures by using the IBM 7094 computer. Tapes were then generated containing pictures obtained by various logical combinations of picture arrays, such as complementation and logical "anding." The results of using these processes were various types of modified pictures, such as those with horizontal and vertical lines erased, those with nothing present except horizontal lines, and so forth.

Contour tracings were derived from the modified pictures on the TX-0, and the coordinates of these tracings were placed on magnetic tape in the FORTRAN-Compatible format.

A large FORTRAN program was written to accept the contour tracings and perform recognition of general classes of patterns found in sheet music. The scheme involves both intratrace and intertrace operations. The intertrace operations are carried out without requiring that the coordinates of all tracings involved be stored in the computer simultaneously. Thus only storage enough for the longest expected tracing need be reserved.

At present, the program successfully processes and recognizes one general class of patterns, namely the timing-bar complexes attached to notes containing the time information for the notes. Progress continues along these lines with work now being directed toward recognition of general note clusters. A program has also been developed to extract the spacing between the lines of the musical stave, with the aim of using this number to set parameters in the general recognition program.

D. H. Pruslin