

XXVI. COGNITIVE INFORMATION PROCESSING*

Academic and Research Staff

| | | |
|-----------------------|---------------------|-------------------|
| Prof. M. Eden | Prof. D. E. Troxel | F. X. Carroll |
| Prof. J. Allen | Prof. I. T. Young | C. L. Fontaine |
| Prof. B. A. Blesser | Dr. R. R. Archer | E. R. Jensen |
| Prof. T. S. Huang | Dr. J. E. Green | A. J. Laurino |
| Prof. F. F. Lee | Dr. K. R. Ingham | D. M. Ozonoff |
| Prof. S. J. Mason | Dr. D. A. Sankowsky | E. Peper |
| Prof. W. F. Schreiber | Dr. O. J. Tretiak | Sandra A. Sommers |

Graduate Students

| | | |
|--------------------|-------------------|-----------------|
| T. P. Barnwell III | E. G. Guttman | D. S. Prerau |
| W. L. Bass | D. W. Hartman | N. Rashid |
| T. R. Bourk | L. Hatfield | G. M. Robbins |
| J. E. Bowie | L. P. A. Henckels | C. L. Seitz |
| R. L. Brenner | P. D. Henshaw | R. D. Shapiro |
| S. A. Ellias | M. Hubelbank | W. W. Stallings |
| J. R. Ellis, Jr. | J. I. Makhoul | K. P. Wacks |
| A. E. Filip | G. P. Marston | G. A. Walpert |
| R. E. Greenwood | G. G. Matison | T. R. Willemain |
| W. B. Grossman | W. McCandless | Y. D. Willems |
| | G. F. Pfister | |

RESEARCH OBJECTIVES AND SUMMARY OF RESEARCH

The primary interests of the Cognitive Information Processing Group are related to gaining an understanding of visual processes; in particular, to the ways by which humans can process pictorial information. Thus we do research on electronic methods for simplifying pictures (still or motion), without altering significantly human subjective judgments of picture quality, on methods for altering pictures so as to enhance particular features and remove others, on the way humans perform sophisticated visual tasks, such as reading and writing, and on the way humans learn to recognize complicated patterns and categorize them without conscious effort.

These studies have led to a variety of applications. The studies of language and picture processing have suggested ways to substitute other senses for sight, so that a blind person can "read" printed material in ways that are not too different from the ways sighted persons use such information sources. Image processing and pattern recognition studies are being applied to the classification of white cells in blood smears, to the detection of malignant cells in Papanicolau smears, to the diagnosis of blood dyscrasias by measurements on erythrocytes in smears, to the enhancement of x-ray images before radiological diagnosis.

During the past year substantial progress has been made in the development of a reading machine for the blind. This system will be able to provide a blind user with a "translation" from a printed page of text into synthetic English speech. The integrated system was first demonstrated in November 1968. Since then, there have been improvements both in the accuracy of the recognition of printed material and in the quality of

* This work was supported principally by the National Institutes of Health (Grants 5 PO1 GM14940-02 and 5 PO1 GM15006-03), and in part by the Joint Services Electronics Programs (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DA 28-043-AMC-02536(E) and the National Institutes of Health (Grant 5 TO1 GM01555-03).

the speech produced by the machine.

A survey of American publishers revealed that only three type fonts account for 62% of the total hard-cover book production in 1968; six type fonts and minor variants thereof are used in more than 90% of published books. Accordingly, an extensive program has been carried out to train the system on the most frequently used type fonts. A new procedure for recognizing the characters on the page has been devised and tested. It has significant advantage over earlier techniques, in that the number of letter confusions have been reduced to a minimum. The over-all accuracy of recognition after adequate training is better than 99.9%. Since readers have no difficulty in understanding texts with considerably greater (3.5%) error rates, the reading machine's accuracy is more than adequate.

Previously, translation of text into Grade II Braille had been accomplished under computer control. Now, spelled speech (letter-by-letter) output is being generated from PCM stored data at rates up to 120 words per minute. Techniques for mapping scientific notation (e. g., mathematical equations) into Braille have been developed, which adds an important class of texts as acceptable machine input.

The addition of a drum memory has enabled us to expand the size of the working dictionary from 400 to 16,000 root words, suffixes and prefixes. It is estimated that a basic vocabulary of this size will be sufficient for reading college-level textbooks whose vocabulary is not overly specialized.

Several projects have dealt with ways to improve the quality of the synthetic speech. A program has been implemented that performs a grammatical parsing of an English sentence to determine beginnings and ends of phrases. This provides the needed information to improve the selection of the correct phonemes, as well as the information for supplying stress, intonation, and pauses in the synthetic production of an English sentence. These features of a language can compensate in some measure for the inadequacies of the synthetic production of the phonetic sounds themselves. Steps have been taken, however, to improve the quality of the individual phonemes. The synthetic speech depends upon a correct evaluation of certain acoustic parameters that have not been well understood. Until now, these parameters have been specified by ad hoc methods, and the output has not been fully satisfactory. We have now computed all of the necessary parameters for one class of speech sounds, the vowels and fricatives, according to physical acoustic theory. The synthetic sounds generated as a consequence show a definite improvement in quality. Lastly, the acoustic correlates of stress and pauses have been studied in specific syntactic environments. For example, a relation has been found between the length of a vowel in a monosyllabic word and the place of the word in a sentence.

At the same time, work has proceeded on the development of more economical systems. We have demonstrated earlier that the printed page can be translated by machine into Grade II Braille. This year, we have also implemented an output that spells out the word. Clearly, both of these modes require much less computer power and hence a less costly system than does synthetic speech. The Grade II Braille program has been modified so that a sighted typist with no computer experience can type at a regular keyboard and produce instantaneously an embossed paper Braille text. A second development is that of a reading-machine control program whose outputs are Braille or spelled speech and whose inputs include a typewriter, light pen, a joy stick, and other position indicating devices. These inputs will be used to investigate the means by which blind users may control the point at which reading is to begin; that is, to permit the user to be able to "browse," "scan" a page rapidly, and perform the other search procedures that a sighted reader is likely to use before he settles down to read a text continuously.

One related system is intended to aid blind professionals in the performance of their work. A number of independent consoles (4, at present) have been connected to a central computer with programs that provide an auditory response to each of the console users as he interacts with the computer in the preparation of programs

or the acquisition of output results. Thus the usual modes by which computers provide information to the programmer or user – typing or other visual displays – are replaced by a vocal presentation of spoken letters, numbers, mathematical symbols, computer commands, and so forth. In this way, the full power of a computer is placed at the disposal of the handicapped professional and should enhance his potential for employment.

Substantial progress has also been made in understanding the problems of picture compression and image enhancement. While these research projects are important for understanding the broad problems involved in extracting useful information in pictures, they have particular relevance to medicine and biology. This is because the processes of diagnosis and analysis in the health sciences depend heavily on the qualitative description of visual objects, such as microscope preparations, x-ray images, and so forth. Thus image-processing basic research is very closely tied to the Biological Image Processing projects of our group. Recent work has included the development of systems for the automatic typing of white cells in blood smears. The simulation of a system to perform automated leukocyte differential has been achieved through the measurement of certain features or parameters on the cell image. Tests on unknown cell populations show classification errors of the order of 7%. Work has continued on automatic methods for analysis of red blood cells in blood smears by differentiating the smear into normal or one of several abnormal types. The algorithms that perform the analysis have been extensively tested on a variety of disease types, and appear to perform satisfactorily.

In the neural anatomy project, a display processor that produces rapid frame-rate cathode-ray displays under indirect computer control, has been completed, and application to the representation of anatomic structures has begun.

For the Papanicolau smear project, computer algorithms have been developed to locate the nucleus and the cytoplasm, and make several objective measurements on them to differentiate normal from malignant cells.

A new project is under way to do some processing and recognition procedures on histological material. To begin with, stained sections of animal liver are being studied with a view to obtaining computer algorithms for the location of nuclei, cell boundaries, and sinusoids.

A project to obtain a three-dimensional reconstruction from a collection of x-ray photographs taken at various orientations has been initiated. Using techniques of discrete Fourier analysis, we have obtained cross-sectional roentgenographic views through intact long-bone specimens.

A technique has been perfected for the computer generation of optical filters suitable for the preparation of holograms on digital computers. Holography is potentially a very powerful tool for visualizing properties of surfaces, but is still not sufficiently well controlled to be practicable.

The technique of computer generation of optical spatial filters has been extended to the manufacture of holograms on digital computers. Holograms of two-dimensional objects and several two-dimensional objects lying in different planes have been generated on the computer and successfully reconstructed optically.

In most work on image restoration (i. e., compensation for image degradation) it has been assumed that the degrading system is linear and shift-invariant. Many real-life degrading systems are, however, linear but shift-variant. A technique has been found for compensating for image degradation caused by a class of linear shift-variant degrading systems which includes, in particular, the coma aberration of lenses. This technique involves Mellin transforms.

A project in efficiently coding the bit-planes of a quantized continuous-tone picture has been completed. This project was partially motivated by the desire to reduce the memory capacity requirement of parallel-processing computers when pictures are

decomposed into their bit-planes in the memory. By using run-length coding and its extensions, researchers have been able to reduce the average number of bits per sample by a factor of two for 6-bit pictures, and a factor of three for 3-bit pictures.

A new scanner has now been interfaced with the PDP-9 computer, and programs have been written that permit pictures to be scanned or displayed under computer control. The new scanner has been used to process radiographs for the purpose of improving the efficiency of diagnosis. While the processing is not novel, the pictures are of somewhat higher quality than those previously attainable. The technique, which includes real-time analog/digital manipulation on the scanner itself, is simpler than ordinary computer picture processing and suggests that it may be feasible to develop a system that would permit the radiologist who is unversed in computer operation to try a variety of enhancement techniques directly and in real time on the plates that he is examining.

Work continues on the contour coding of images. Preliminary results show that quadratic curves may be fitted to contour data and that good quality images may be reconstructed in this manner.

Two recently completed theses deal with picture coding by Fourier and Hadamard transform methods. Both techniques show promise of substantial reduction in channel capacity using a degree of computational complexity that is rather high but probably acceptable in certain applications. Also, work on the spectral representation of texture information (which may be defined as pictorial data not representable by contours) indicates that a relatively small increase in channel capacity may add a great deal of realism to contour-coded pictures.

Several different methods of coding pictures have been studied. The methods are intended to decrease the amount of computer storage space necessary to hold the pictorial information, and thereby to simplify the very difficult processing problems that are a consequence of the very large storage requirements for high-quality, detailed pictures. We have now been able to decrease the required storage space for photographs of relatively simple natural scenes to approximately 1/6 of the space required by the usual methods of picture scanning and encoding.

Our research objectives for the coming year are extensions of the work described above.

1. The development of a reading machine for the blind is one of our major projects.
 - a. Work continues on a new Vidicon scanner, which will be able to scan intact pages of a book or magazine. This project is also aimed at developing more efficient means of in-putting data to the computer, so that the cost of the scanner part of the system can be reduced.
 - b. Following a survey of type-font usage, we are continuing our study of character-recognition algorithms. In this area, we are trying to develop inexpensive techniques to recognize several fonts, and to deal with the problems of poor quality printing, such as missing segments and touching letters.
 - c. We are attacking the problem of speech quality from a fundamental point of view. The classes of speech sounds (such as stop consonants) are being described from an acoustic-phonetic point of view, and the synthesis algorithms are being designed in accordance with these data. We expect to be able to generate phonemes accurately, and also to provide a better approximation to the suprasequential cues, including stress, intonation, and juncture.
 - d. To be useful, a reading machine for the blind must have convenient controls to allow for page positioning, backup, and other functions that would otherwise require a sighted operator. Experiments are under way to determine what controls are needed and how they should be implemented.

e. We have been expanding our dictionary with new entries in order to be able to handle a wide variety of texts. Present efforts are centered on selection of new words, provision for correct sound representations, and examination of all new words for possible decomposition into prefixes, root words, and suffixes.

2. Image-processing techniques continue to be studied in several areas.

a. The reduction of the number of bits required to transmit a visual image is a continuing goal. Coupled with such reduction is the attendant increase in subjective degradation of these images. Study of visual noise perception in such pictures is being done with controlled spatio-temporal spectra.

b. In the field of coherent optics, we have generated holograms of two-dimensional objects by computer. We are now turning to computer-generated holograms of three-dimensional objects.

c. In the field of image enhancement, we have been studying the effect of linear systems that are not shift-invariant. Initial work has shown that the distortion induced by such shifts can be reduced by use of a form of the Mellin transform on the data.

d. Of great importance is the enhancement of x-ray images. Work continues on the reconstruction of three-dimensional images from x-ray images taken at several different orientations of the object, and algorithms are also being developed to provide improved contrast in x-ray images.

e. We are beginning to investigate the problems of efficient representation of color images. Initially, statistics are being collected for the color components in several images. From these data, it is expected that efficient coding schemes can be developed.

3. A major objective of our research is the automatic analysis of biological objects.

a. The leukocyte differential analysis project will be directed toward obtaining a larger data basis for the classification procedures and toward analysis of some of the pathological cell types found in circulating blood.

b. The erythrocyte classification program has enabled us to measure a number of properties on a fairly large sample of normal erythrocytes. The work is being extended in order to collect a data base of erythrocytes found in disease states so that efficient and accurate diagnostic algorithms can be written.

c. A new project is in progress to process and recognize important features in histological material. To begin with, stained sections of mammalian liver are being examined with a view to creating appropriate computer algorithms that will locate nuclei, cell walls, and sinusoids.

d. The techniques developed in the blood-cell studies appear to be applicable to a wide variety of problems in which discrete and separated objects are identified. Objects that are being studied include malignant and normal cells in smears, malarial parasites, and chromosomes.

4. A variety of projects in pattern recognition continue.

a. Several years ago, we were able to provide a formal description of the dynamics of hand motion that is used in producing English script. This was used as the basis for a highly successful series of recognition programs, but these programs required information about the time course of the writing. Since readers of cursive script do not have such information, it is highly desirable to do the recognition from spatial information alone. We intend to re-examine these algorithms to see if they can be extended to reading script based on spatial information alone.

b. Work continues on a program to read musical scores. A number of insights concerning the use of contextual information have been obtained, and the costly procedure of template matching has been avoided completely.

c. We have been trying to improve the performance of speech recognition systems

by instructing the speaker to modify the manner in which he produces speech sounds. The system has been tested on a restricted set of vowels and consonants, and testing will continue for a larger class of English speech sounds.

d. Research on the invariant properties of three-dimensional visual illusions is being carried out. Results thus far on the geometry of the trapezoidal window and on three-dimensional analogues of the Necker cube have revealed interesting inconsistencies in interpretation. Attempts are being made to interpret these results within a more general theory of recognition of surfaces and textures in depth perception.

5. Our psychophysical and psychological studies of cognitive tasks continue.

a. A comprehensive auditory test facility will be built. Equipment for precision sound generation and recording, as well as analysis, will be included. In our work in speech and auditory psychophysics we shall use this facility heavily.

b. Work will proceed on visual perception with the goal of applying this knowledge to improving character pattern recognition. In particular, we are trying to find those features that are invariant to type font.

c. The role of context in perception is being studied in both auditory and visual modalities. We are most interested in understanding how a complex stimulus can be perceived without specific perception of its constituent individual elements. Such results will be of direct value in understanding reading and speech perception.

M. Eden, J. Allen, B. A. Blesser, T. S. Huang, K. R. Ingham, F. F. Lee,
S. J. Mason, W. F. Schreiber, O. J. Tretiak, D. E. Troxel, I. T. Young

A. COMPUTER MODELS OF SHIFT-VARIANT IMAGE-PROCESSING SYSTEMS

A class of shift-variant image-processing systems represented in the block diagram of Fig. XXVI-1 may be readily modeled in the computer. The first distortion (distortion is the same as a change of coordinates) may be easily performed by nonuniform scanning of the input image. The subsequent filtering operation and second distortion are then performed to produce the output image.

The advantage of this representation is the simplicity either in the representation of the inverse system or in stating conditions under which the inverse

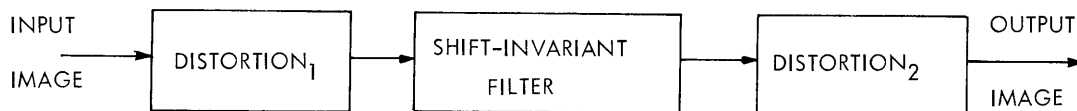


Fig. XXVI-1. Model for shift-variant image-processing system.

system does not exist. For example, when the distortion is not one-to-one over a certain region, then the system is not invertible. Also, this representation

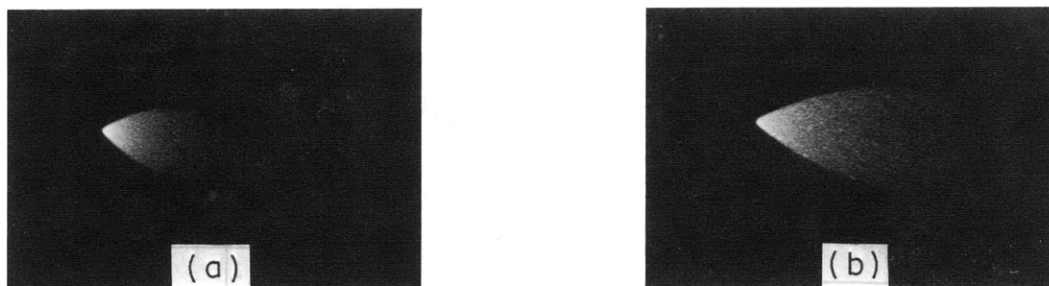


Fig. XXVI-2. (a) Coma $\sim 10^\circ$ off axis.
(b) Coma $\sim 15^\circ$ off axis.

is valuable, since it represents real systems. For a spherical lens, for example, Fig. XXVI-2 shows the coma pattern outputs attributable to point-source inputs at approximately 10° and 15° off axis, respectively. This impulse response is characterized by the fact that it spreads out in proportion to the off-axis distance of the point source. For coma the first distortion may be performed by scanning in polar coordinates with equal angular increments and geometrically related samples along a radius. The impulse response of the shift-invariant system is the distorted coma pattern at a standard position. The second distortion is performed by retaining only those samples along a radius which are logarithmically related, then converting this uniform- r , uniform- θ record to a uniformly sampled Cartesian record.

1. Model of a System with Cylindrical Symmetry

The first system to be modeled is a cylindrically symmetric lens, which essentially reduces the model to one dimension. Let the impulse response be

$$H(t; \tau) = \frac{1}{\tau} h(t/\tau)$$

= response at t to an impulse at τ .

Then with $u(\tau)$ as the input along a scan line and $v(t)$ as the corresponding output

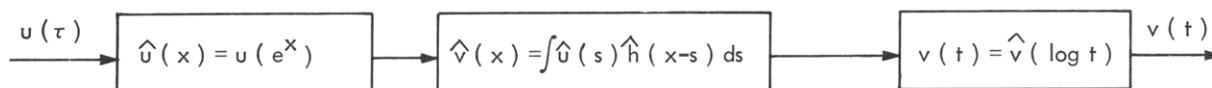


Fig. XXVI-3. Model for system with cylindrical symmetry.

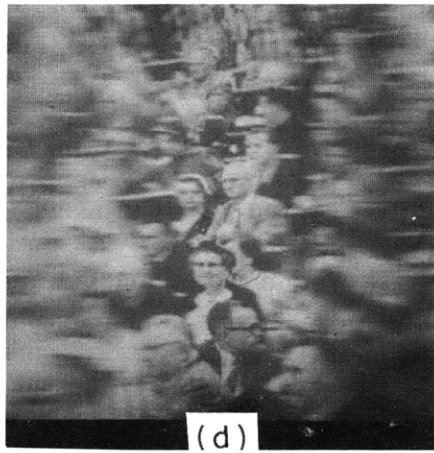
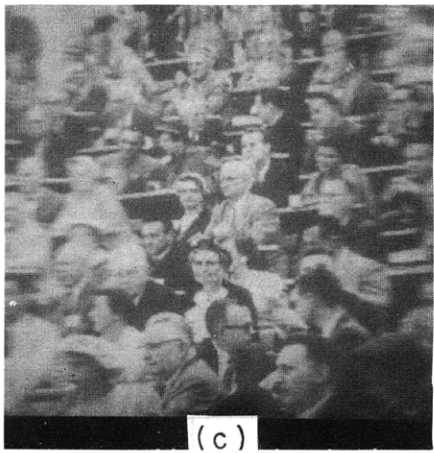
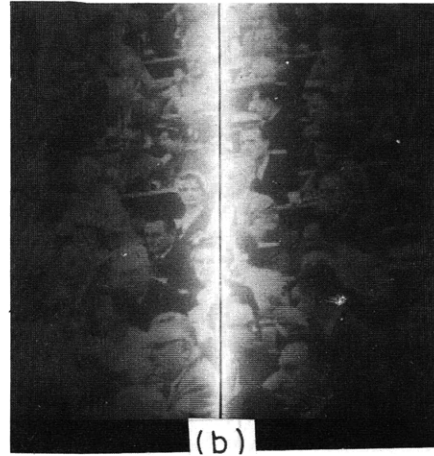
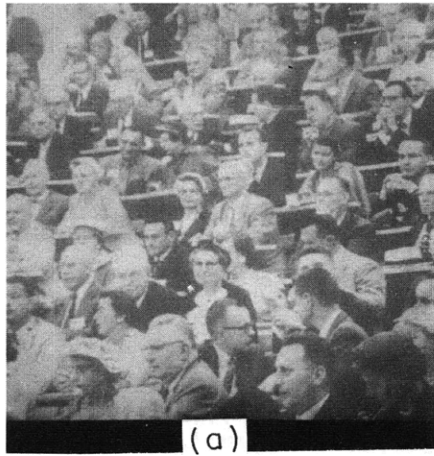


Fig. XXVI-4. (a) Input.
(b) Distorted input.
(c) System output.
(d) System output.

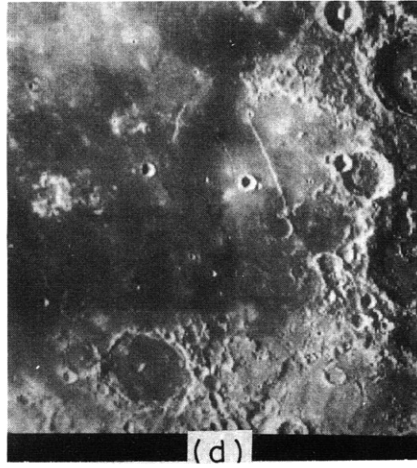
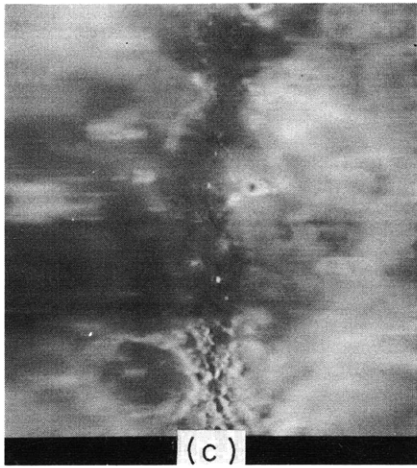
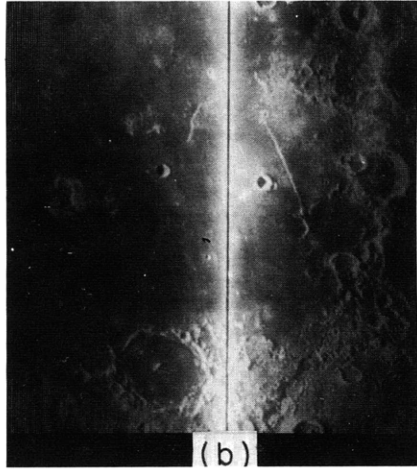
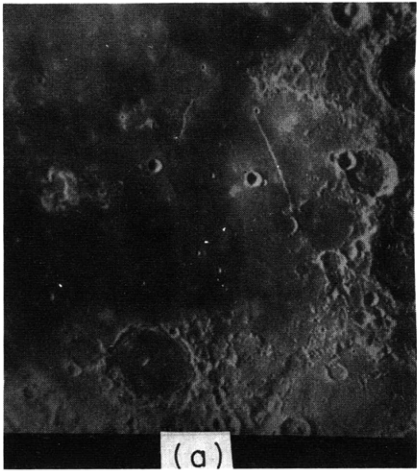


Fig. XXVI-5. (a) Input.
(b) Distorted input.
(c) System output.
(d) Reconstructed input.

Fig. XXVI-1 takes on the specific form of Fig. XXVI-3. With a specific choice of $h(x)$ as

$$h(x) = \begin{cases} 1 & 1 \leq x \leq L \\ 0 & \end{cases}$$

the following examples were generated on the New Scanner under control of the PDP-9 computer. Filtering is performed only along a scan line which is consistent with the cylindrical symmetry assumption.

The input, $u(\tau)$, the distorted input, $\hat{u}(x)$, and outputs, $v(t)$, from two different systems (L was chosen to be different) are shown in Fig. XXVI-4. The axis of symmetry is the vertical line in the center, which is apparent in Fig. XXVI-4b. The distortion $\tau_n = e^{x_n}$ would require an infinite number of samples in the region around $\tau = 0$, so the origin was treated as an exception in the modeling. The outputs shown here would actually extend out farther, but they have been truncated for display purposes. Corresponding directly with Fig. XXVI-4 is Fig. XXVI-5. Figure XXVI-5d shows, however, the results of an inversion performed by means of the Fourier transform. This inversion is essentially a numerical exercise, in the sense that once the picture is scanned everything is performed inside the computer. In the program, the system output samples contain ~ 15 bits/sample. Even so, the reconstructed input is not perfect. This brings up the question of performance in case the system output were truncated to fewer bits. This question is still under investigation.

2. Preliminary Results on Modeling Systems with Axial Symmetry

More common image-processing systems have axial symmetry. Modeling them requires scanning the input in polar coordinates, and such a scanning raster, uniform

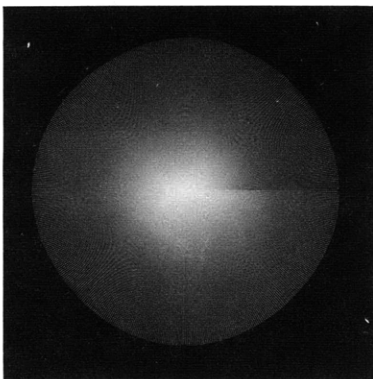


Fig. XXVI-6. Polar coordinate scanning raster.

in r and uniform in θ , is shown in Fig. XXVI-6. Programs to perform shift-variant filtering of the coma type are now being debugged.

G. M. Robbins

B. SPEAKER-MACHINE INTERACTION IN A LIMITED SPEECH RECOGNITION SYSTEM

1. Introduction

In order to recognize the speech of a new speaker, any ambitious automatic speaker-independent speech recognition system will, in some measure, have to adapt its parameters and procedures to the new speaker. This adaptation is often referred to as "tuning in" or "speaker normalization." The extent to which a recognition system must adapt to the speaker depends on the complexity of the task. Keeping this in mind, I set out to answer two questions:

1. Is it feasible to construct a limited speaker-independent automatic speech recognition system without speaker normalization?

2. In cases of incorrect recognition, can speakers change their articulations to effect correct recognition without significantly affecting human perception of the utterances?

In order to answer these questions, a vocabulary of 55 CV syllables was chosen, employing 5 tense vowels and 11 consonants. The vowels are /i/, /e/, /a/, /o/, and /u/. The consonants comprise 6 stops: /p/, /t/, /k/, /b/, /d/, /g/, and 5 fricatives: /f/, /s/, /s̃/, /v/, /z/. Each syllable is preceded by a schwa vowel, and followed by /d/; for example, /əgad/, /ədod/, /əbid/, /əsud/, /əzed/. Many of the utterances are ordinary English words.

The system that was used in the analysis and recognition is shown in Fig. XXVI-7

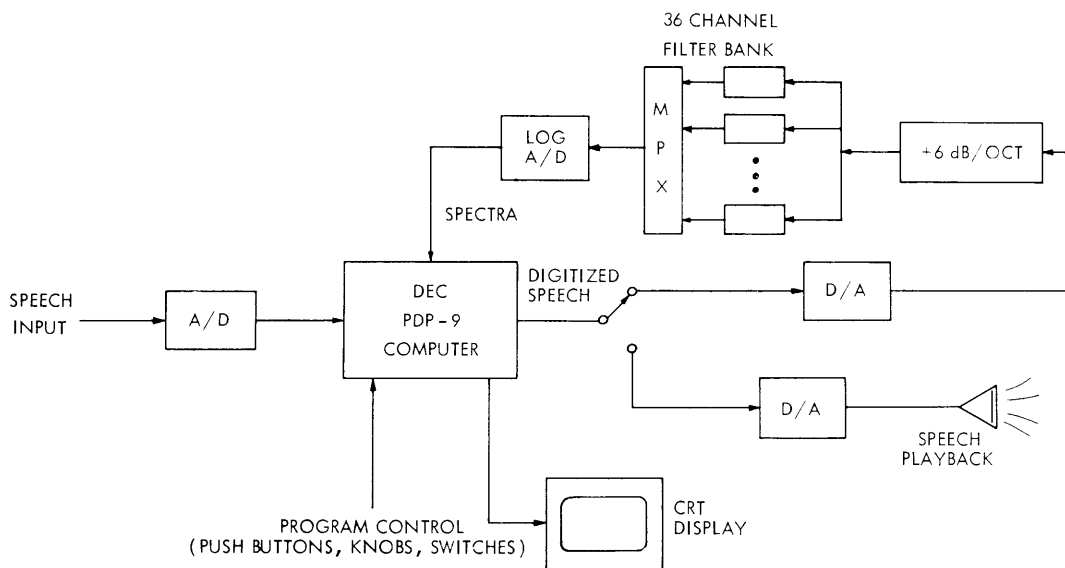


Fig. XXVI-7. Speech-recognition system.

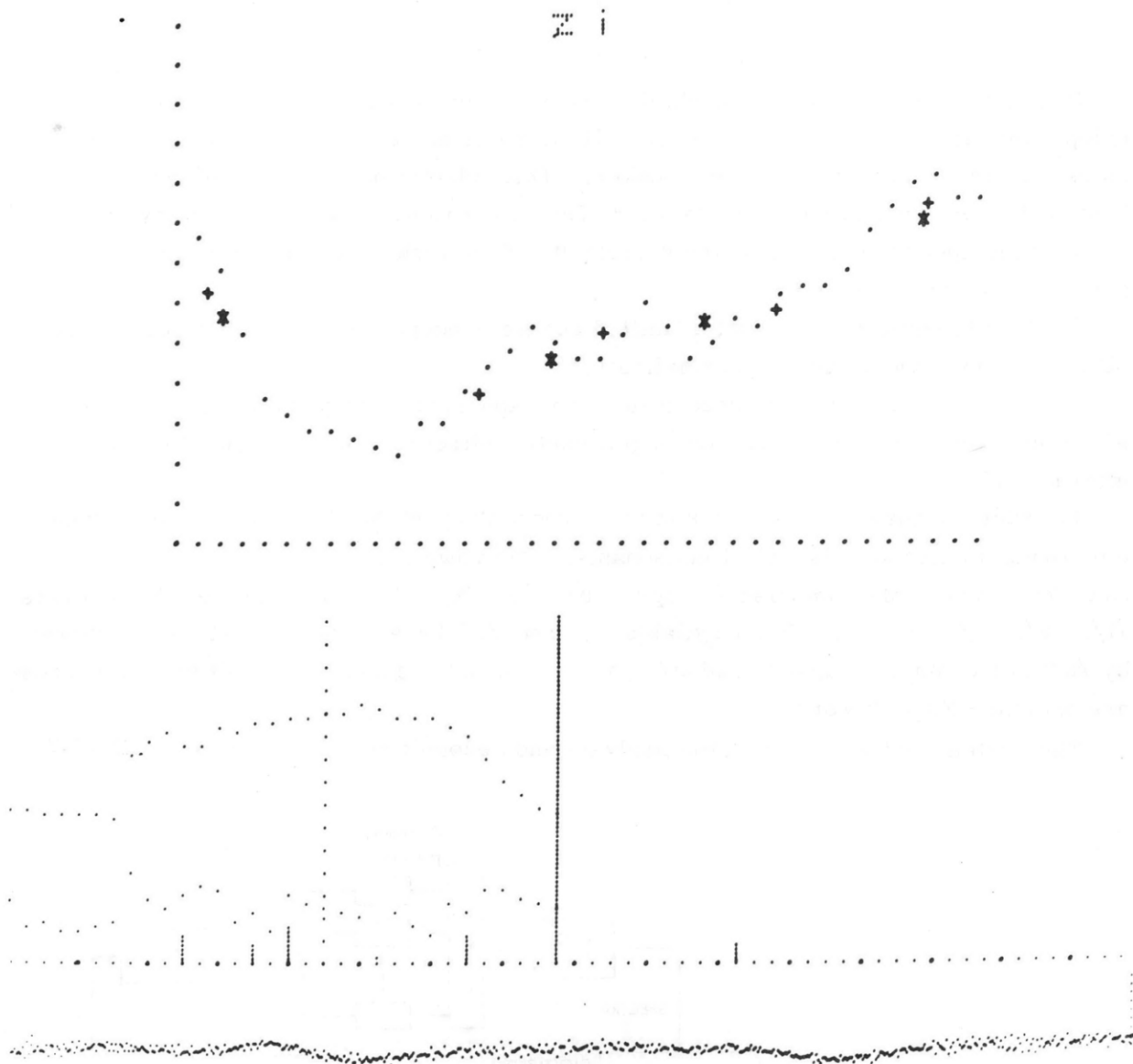


Fig. XXVI-8. CRT display of recognized syllable.

The speech waveform is sampled at 20 kHz and stored in the computer memory. All subsequent operations are performed on the digitized waveform. The frequency spectrum can be obtained at any point in time either by using the 36-channel filter bank¹ shown in Fig. XXVI-7 or by computing the Fast Fourier Transform (FFT). The FFT gives better frequency resolution and indication of voicing, while the filter bank is much faster in obtaining the spectrum. Auditory feedback is supplied to the speaker by the speech-playback push button, to which he has access. Visual feedback about the recognized syllable is also supplied to the speaker through the CRT display.

Figure XXVI-8 shows what the speaker sees as he looks at the CRT display. Note the recognized syllable /zi/ at the top. The rest of the display shows the spectrum of the time waveform that is displayed below, plus other derived graphs, all of which are of concern only to the experimenter and perhaps to the inquisitive subject.

2. Recognition Scheme

Very briefly, the recognition depends on the extraction of several acoustic features. This is accomplished by determining the positions of spectral energy concentrations and tracking their relative movements in time.

After locating the consonant and the vowel, the recognition of the vowel proceeds, followed by the consonant recognition. First, a decision about whether the vowel is front or back is made. This decision is very important because the recognition of stop consonants depends partially on whether the following vowel is front or back. After the front-back decision, the vowel is recognized by deciding whether it is high, mid, or low.

The consonant recognition starts by determining the manner of articulation: stop vs fricative, and then voiced vs unvoiced.

The place of articulation of a fricative is determined by examining the spectrum, as computed by using the FFT at one point in the fricative portion. The place of articulation recognition of a stop consonant is initiated by accurately locating the stop burst. The burst spectrum is used as the major source of information for place recognition. Also, for stop consonants followed by back vowels, the consonant-vowel transition is also used in the recognition.

The different parameters used in the whole recognition scheme were extracted from the recorded utterances of 6 speakers, 3 male and 3 female.

3. Experimentation

The resulting recognition scheme was tested on 12 speakers, 6 males and 6 females ranging in age from 17 to 30. Three of the speakers had been used in the original analysis. The other 9 speakers were those who had responded to an advertisement requesting subjects. The speakers had different accents. They came from

(XXVI. COGNITIVE INFORMATION PROCESSING)

Massachusetts, Ohio, Maryland, Chicago, California, Montreal (Canada), and one had French and Arabic as his first and second languages.

The experimentation had two parts. The first experiment started by having each subject read the randomized list of 55 utterances through once, and the errors were recorded. This was followed by an informal learning session, during which the subject was instructed to change his or her articulation in such a way as to effect correct recognition for those words that had been incorrectly recognized. Different methods of changes in articulation were suggested by myself, since I had a knowledge of which part of the algorithm had caused the incorrect recognition.

Following is a brief list of some kinds of changes in articulation that were employed: rounding and protruding of the lips and diphthongization were used with vowel errors; deliberate efforts at voicing and/or frication were used with consonant errors; place of articulation errors with stop consonants were corrected by a proper production of the stop burst.

The second experiment again started by reading the list through once and recording the incorrectly recognized utterances in an error list. The utterances in the error list were then each repeated twice. Those that were correctly recognized twice in a row were eliminated from the error list, and with the rest of the error list the speaker underwent a more formal learning session. Figure XXVI-9 shows typical learning curves.

The trial number n is plotted on the abscissa, and the total number of correct responses after trial n is plotted on the ordinate. The black dots, representing correct recognition, are joined by straight lines. The ideal perfect recognition is represented by the 45° diagonal line. The learning process is considered successful if the learning curve becomes parallel to the ideal diagonal line, as it does in the two upper learning curves (a) and (b). Note that both curves are typified by a sudden change in slope (at trial 7 in Fig. XXVI-9a and at trial 13 in Fig. XXVI-9b). This indicates that the learning is not a gradual process, but rather comes about as a result of a sudden conscious realization on the part of the speaker as to the articulation needed to effect the correct recognition. This is attested to by the verbal remarks of the speakers. For example, in Fig. XXVI-9b: the /u/ in /ədud/ was being incorrectly recognized as a front vowel because it was pronounced as a diphthong /ədjud/ instead of /ədud/. This fact was immediately explained to the speaker, but it was not until the 12th trial that she suddenly realized what she was doing, and her remark was "Oh, I see! It's because of the /d/," and she immediately proceeded to change her articulation accordingly.

In each of the two lower curves, Fig. XXVI-9c and XXVI-9d, the learning process was considered to be unsuccessful. The learning process in Fig. XXVI-9c looks more hopeful, however, than that in Fig. XXVI-9d.

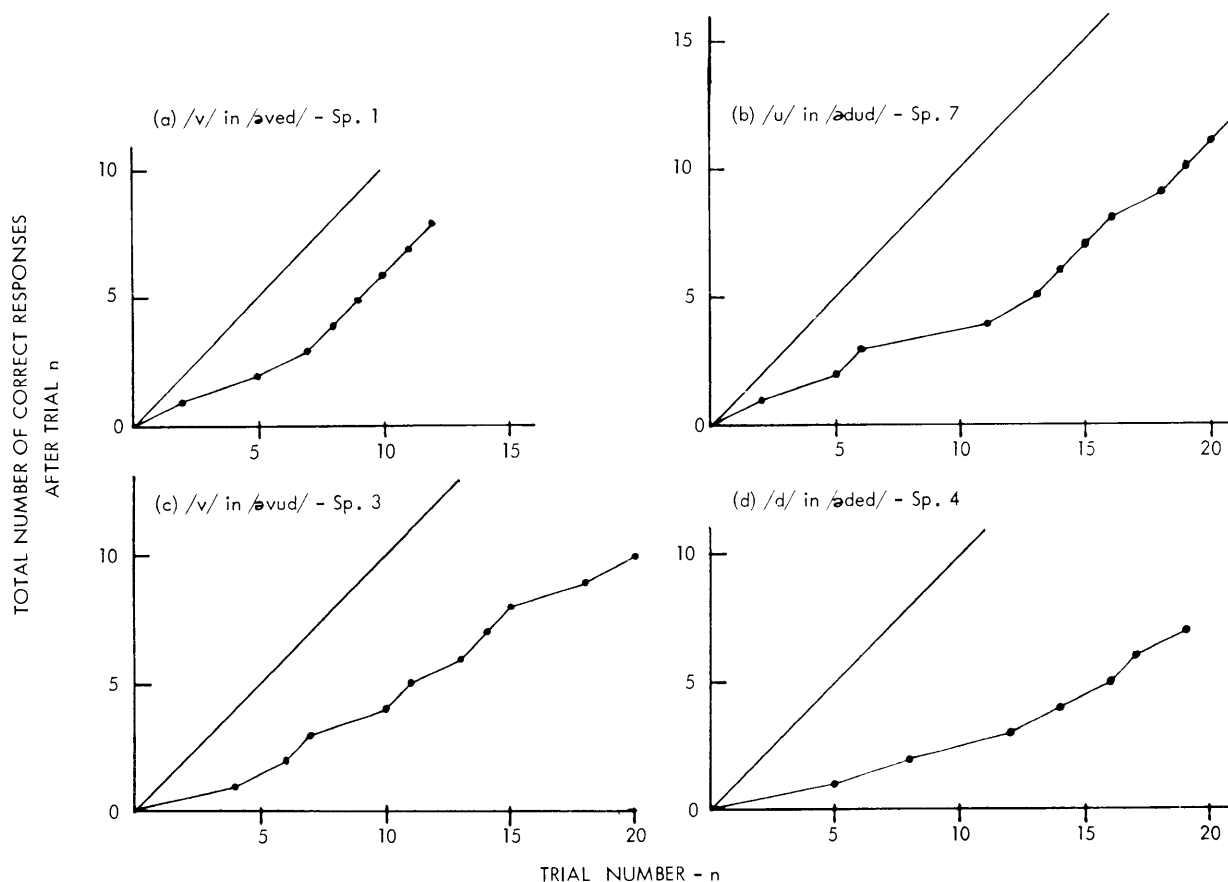


Fig. XXVI-9. Learning curves.

4. Recognition Results

Figure XXVI-10 shows the recognition results of the two experiments. Bar graphs of the error rates are displayed vs the speaker number. The average error rates for all 12 speakers are shown at the far right, above the word "Average." The upper graph shows the consonant error rates, the middle graph shows the vowel error rates, and the total word error rates are shown in the lower graph. The white bars show the results of the first experiment, before any learning had occurred. The solid bars show the results of the second experiment, which took place about a week after the first learning session. Speakers 1-6 are male; speakers 7-12 are female. Speakers 6, 8, and 11 were used in the original analysis, and of these, speakers 8 and 11 were phonetically aware. Note that their performance is not very different from that of the other speakers.

In general, the vowel error rates were much lower than the consonant error rates. Speaker number 10 had a very high vowel error rate in the first experiment because of

(XVII. COGNITIVE INFORMATION PROCESSING)

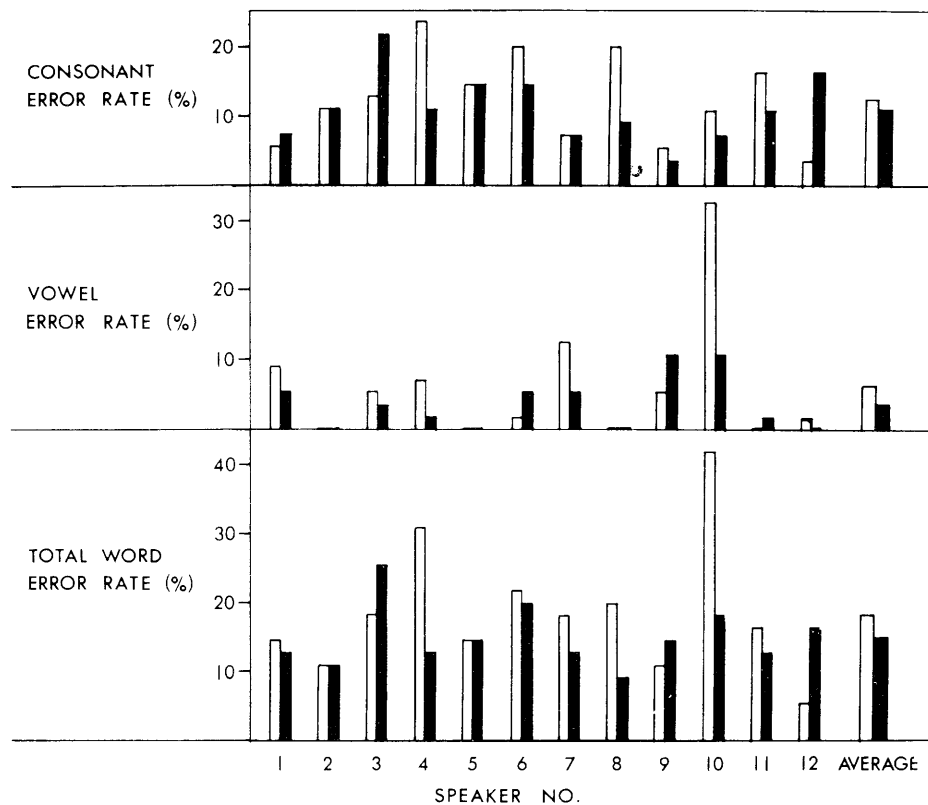


Fig. XXVI-10. Recognition results.
Experiment 1 .
Experiment 2 .

exceptionally high first formants for high vowels, for which there was no hope of a cure through learning. So, instead of quietly eliminating this subject from the experiment, the decision algorithm was slightly changed, and as a result, the recognition improved considerably. That change also resulted in an increased vowel error rate for subject 9, but these errors were successfully eliminated in the second learning session. On the average, the recognition results were better in the second experiment than in the first, but not much.

In order to examine the types of errors more closely, let us take a look at the distribution of errors in the second experiment. The table in Fig. XXVI-11a shows the number of consonant confusions that occurred out of a total of 660 utterances by 12 speakers. The numbers along the diagonal represent place-of-articulation errors, and the rest represent manner-of-articulation errors. Note that with stop consonants most of the errors were place-of-articulation errors, while with fricatives the majority were manner-of-articulation errors, mainly attributable to problems with voiced fricatives /v/ and /z/.

| RESPONSE STIMULUS | | STOP | | FRICATIVE | |
|----------------------|----------|----------|--------|-----------|--------|
| | | unvoiced | voiced | unvoiced | voiced |
| STOP | unvoiced | 12 | 2 | | |
| | voiced | 1 | 23 | | 2 |
| FRICATIVE | unvoiced | 1 | | 9 | |
| | voiced | | 8 | 12 | 4 |

(a)

Place of Articulation Errors for STOP Consonants

with Front Vowels 29 errors
with Back Vowels 6 errors

| Experiment | Consonant Error Rate % | Vowel Error Rate % | Total Word Error Rate % |
|------------|------------------------|--------------------|-------------------------|
| 1 | 12.6 | 6.4 | 18.3 |
| 2A | 11.2 | 3.8 | 15.0 |
| 2B | 5.0 | 0.8 | 5.8 |
| 2C | 2.3 | 0.0 | 2.3 |

(b)

Fig. XXVI-11. (a) Recognition errors – Experiment 2A.
Total number of stimuli 660.
(b) Over-all recognition results.

The middle part of Fig. XXVI-11 shows the distribution of place-of-articulation errors for stop consonants according to the following vowel. There were 29 errors with front vowels compared with only 6 errors with back vowels. Recall that the algorithm for stop recognition was different, depending on whether the stop was followed by a front or a back vowel.

The table in Fig. XXVI-11b shows over-all recognition results in terms of consonant, vowel, and total-word error rates. The total-word error rate was 18.3% in the first experiment and 15.0% in the second experiment. Recall that after the second experiment, the words in the error list were each repeated twice (after a brief reminder to the subject of what he supposedly had learned from the first learning session). Those words that were correctly recognized twice in a row were eliminated from the error list.

(XXVI. COGNITIVE INFORMATION PROCESSING)

The resulting error rates are shown in the table under 2B. The total word error rate thus dropped from 15% to 5.8%. This shows that the majority of errors were easily corrected by simple repetition. Therefore, either those errors were random, or the speaker was immediately able to change his articulation to effect correct recognition.

Now, with the words still remaining in the error list, each speaker underwent a learning process. Those words for which the learning process was successful (determined from the learning curves as in Fig. XXVI-9a and XXVI-9b) were eliminated from the error list. The resulting error rates are shown under 2C. The vowel error rate dropped to zero, and the consonant error rate dropped to 2.3%. These figures represent the optimum results that could be achieved with these 12 subjects, given the present recognition system. The remaining errors all occurred either with the voiced stop /d/ followed by front vowels, or with the voiced fricatives /v/ and /z/. Judging from the learning curves for these errors, and from my knowledge of the recognition algorithm, it seems that the 2.3% error rate can be effectively reduced only by appropriate changes in the recognition algorithm.

5. Conclusion

In answer to the two questions posed at the beginning of this report, this investigation shows that a limited speaker-independent automatic speech recognition system without speaker normalization is feasible to construct, and that speakers can learn to change their articulations, within limitations, to improve recognition scores. Further research should investigate the feasibility of a similar system with a more expanded and functional vocabulary.

A more detailed description of this research may be found in the writer's thesis.²

I wish to thank the Speech Communication Group of the Research Laboratory of Electronics for the use of their computer facility throughout this investigation.

J. I. Makhoul

References

1. W. L. Henke, "Speech Computer Facility," Quarterly Progress Report No. 90, Research Laboratory of Electronics, M.I.T., July 15, 1968, pp. 217-219.
2. J. I. Makhoul, "Speaker-Machine Interaction in Automatic Speech Recognition," Ph.D. Thesis, M.I.T. (forthcoming).

C. MIT-RLE PHONE-INPUT, VOICE-OUTPUT SYSTEM

1. Introduction

For several decades, researchers have been developing specialized sensory-aid devices for the blind and physically handicapped. With the exception of a very few

sensory aids such as the MIT-RLE Reading Machine system few involve the application of system design techniques. The resulting specialized aids have tended to be expensive and of limited usefulness. The MIT-RLE phone-input, voice-output system attempts to provide a flexible, computer-based sensory aid whose usefulness is largely determined by the handicapped themselves. One of the project tasks is to monitor the way in which the blind execute simple tasks such as note taking and typing, and attempt to develop software packages that would perform these functions. Another object is the elaboration of software, together with appropriate spoken vocabularies, which would enable the blind to carry out many employment activities that otherwise require sighted help. Many kinds of employment that are at present off-limits to the blind, as well as to other handicapped people, could then be made available. A third and more long-range aspect of this project is the human factors study which is made possible with a flexible system involving a high degree of man-machine interaction. Custom-designed remote terminals are supplied which maximize the man-machine communications. These terminals, which may include alpha-numeric, Braille, or stenotype keyboards, are acoustically coupled

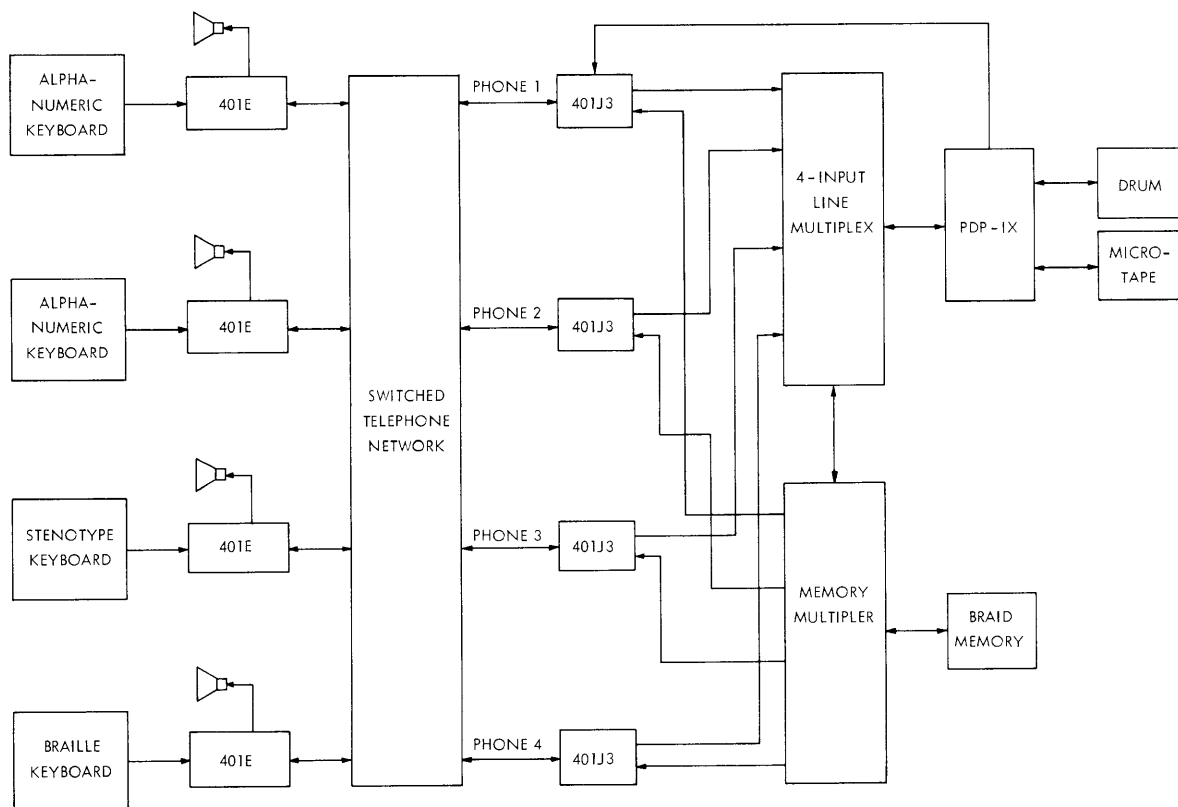


Fig. XXVI-12. Information flow on the MIT-RLE phone-input, voice-output system.

(XXVI. COGNITIVE INFORMATION PROCESSING)

to a home or office telephone handset. The blind operator inputs data or issues commands via the keyboard and receives voice or Braille responses. Figure XXVI-12 is a diagram of data flow. The requirement of eliminating a carrier tone on the phone line necessitates the use of 2 out of 8 and 3 out of 15 touch-tone character sets. This fact, however, is a mixed blessing, since the rotary dial will eventually be replaced by touch-tone instruments that will permit a more universal data entry system.

The system is applied in two ways. The first use is as an experimental console for data manipulation and retrieval experiments as part of the MIT-RLE Reading Machine project. The second application, which will involve the simultaneous operation of the three remaining lines, is intended for the exploration of such systems as sensory aids in employment enhancement. Taking a realistic view, we must concede that the blind, while capable of competitively performing many tasks, lack information in a form that they can use. The Reading Machine system is intended to attack this problem, insofar as printed reading matter is concerned. Unfortunately, much "reading matter" must be read on the spot or be generated by the blind person himself. For example, a blind programmer generates code and reference information that at some time must be put in ink print for the benefit of sighted colleagues. Thus both retrieval and transcribing problems, which are time-consuming for anybody, become far more serious for the blind. The above-mentioned phone system provides means by which low-cost access to the data manipulation power of a computer may be achieved.

In November 1969, the system's implementation reached a milestone, in that a 4-input line multiplexer was successfully operated. A simulation of a basic 93-word speech output, including sound representation for the standard alpha-numeric keyboard, was made by utilizing as input a Peripheral Data Systems keyboard interfaced to the telephone Direct Dial network. The depression of any of its keys causes 3 out of 14 possible touch-tone tones to be generated and transmitted to a 401J3 receiving set, located at the PDP-IX computer. The data output from this receiver is transmitted under program control to user programs. Voice responses to interrogations or calculations can be transmitted back to the user from two data base stores located on the computer. The voice parameters stored in the memory were temporarily stored on the PDP-IX computer's drum, and the simulation involved their reconstruction into analog signals. A second output multiplexer incorporating the RLE-Instrumentation Laboratory wire-braid memory is expected to be completed in December 1969. The same sounds will be output by the braid memory under computer control and, although only 4 lines will be implemented, the memory controller will permit up to 32 simultaneous "conversations." Thus a full 4-simultaneous input/output communication system will be in operation, providing the full power of the

PDP-IX computer to a large group of handicapped individuals in the Boston area.

2. Delta Modulation

The braid-memory data base, which includes such sounds as "integral, differential, sigma," is intended for the presentation of keyboard characters and spelled speech. The experiments, particularly those for employment enhancement, require whole-word presentations of finite but in general different vocabularies. Therefore bit-compression techniques were investigated in order to make more efficient use of the available mass storage. Delta-modulation experiments carried out in both software and hardware have resulted in the elaboration of techniques to enable 10 kbit per second data rates for stored speech. A linear delta encoder and demodulator have been built which gives excellent speech at a 20 kbit per second rate. This unit, while simple to construct, has a limited dynamic range and a bit rate saving at best 2-1 over standard PCM sampling. Therefore adaptive delta modulation was simulated on the TX-0 computer. Here the step size, K , is allowed to vary in time. A signal whose amplitude is increasing would be represented by a series of 1's, where each one caused an increment of voltage of height K to be added to the decoder's leaky integrator circuit. In the adaptive delta-modulation system, K , for a second 1, is allowed to increase, for a third, to increase again, etc., thereby resulting in a faster response to a rising input signal (see Fig. XXVI-13). This approach provides a companding effect that increases the system's dynamic range. Of course the way in which K changes as a function of time, which we call the "algorithm," is variable. Simulations were made with K going as N , $2N$, 2^N , etc. Initial tests with K varying exponentially, as suggested by Marion R. Winkler for high-information delta modulation, indicate that this algorithm is unacceptable. The specification of the best algorithm is still under study. In our case, the $2N$ algorithm provides highly intelligible, and from a quality standpoint acceptable, speech near 10 kbits per second. Therefore if we assume 100-180 ms for average syllable or compressed-speech word duration, up to 60 English words could be stored on 1 drum track. If 10 tracks are used along with the braid-memory store, a comparatively large vocabulary (700 words) would be available for realistic voice response experiments.

3. Software

The present hardware and software has been tested for 4 input lines, although only 1 line was operated at a time. The software incorporates the forking procedure available on the PDP-IX computer to enable 4 simultaneous and essentially

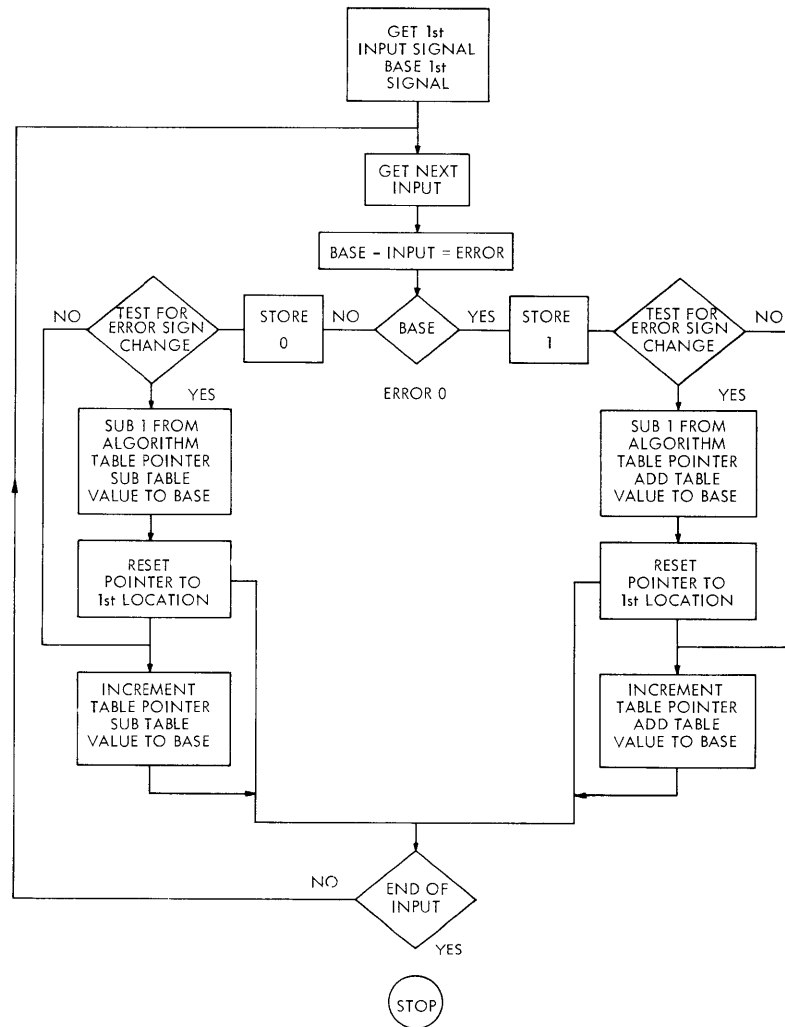


Fig. XXVI-13. Adaptive delta-modulation program. The discrete algorithm values are sequentially stored in the algorithm table, one set for positive increments, and a second for negative increments. The choice of sequence, as well as that for encoding and decoding, is left to the operator.

independent processes, one for each remote station. The current system enables the user to input a program segment or spelled literary segment, and provides output for a simulated braid vocabulary. Software for much more intelligent system operation has been designed and will soon be available along with the memory controller. The functions provided will enable the user to input a comparatively large segment of text (approximately 4000 words), and manipulate these in a variety of ways. Simple editing functions such as SEARCH A, B, where A and B are the beginning and end of a text segment, enable the operator to do string retrieval. This should be a particularly useful function for blind stenographers. Similarly, FIND A, B attempts to locate a portion of text that may or may not exist in the core image. For both SEARCH and FIND, failure to specify B results in the retrieval of A alone. The FIND mode is used with a thesaurus, T, whose terms or descriptors may be changed by the user. This mode of interrogation will help the experimenter to discover how a subject would explore text, since the use of descriptors requires thesaurus-to-text and user-to-machine idea linking. Thus procedures for searching unknown or "new" material can be investigated and strategies developed for information retrieval. The commands to change or append to the text or thesaurus are also provided, and all of these commands are given obvious mnemonics. All input/output specifications will be invisible to the user and this, coupled with the operator's ability to link commands logically, will provide a very flexible research tool. It should be emphasized that nowhere near the full flexibility of the system will be made available to the user. Many of the commands are aids for the researcher and could be supplied to the handicapped only after appropriate training.

The voice-output program utilizes a computational dictionary approach to transform alpha-numeric strings to sound storage parameters. Words to be translated are matched with dictionary entries. If the input word is not found, the spelled combination is decomposed into "syllables." Such syllables in the current program must themselves be whole words having meaning, such as the words "back" and "space," which may be components of the word "backspace." The failure to find a word or a correct set of part-words results in a spelled-speech output. For most experiments the vocabulary along with its linguistic structure should be known in advance and a natural, realistic output will be simulated. The requirement for a natural output will be easily met because many employment situations, such as programming, inventory control, or stenography, involve the use of finite vocabularies. On the other hand, a full-speech output is desirable, and should such a voice become available, we would not hesitate to use it.

K. R. Ingham

(XXVI. COGNITIVE INFORMATION PROCESSING)

D. AUTOMATIC CHROMOSOME KARYOTYPING

1. Introduction

The karyotyping, or classifying, of human chromosomes has become a problem of increasing interest and importance in recent years. The clinical application of karyotypes has increased as medical knowledge in the area has expanded and research interests have produced a need for karyotyping a vastly increased number of chromosome complements. Manual karyotyping is too costly to meet the increased demands; semi-automatic or automatic procedures are necessary.

Briefly, the 46 chromosomes in a normal human cell form 23 pairs. For a female, these pairs are numbered 1-22 and X, with X being reserved for the sex chromosomes and the remaining pairs being numbered roughly in the order of decreasing length. In this work, these 23 pairs have been partitioned into 10 groups (generally agreed as individually recognizable) with the following correspondence:

| | |
|-------------|------------|
| Pairs 1-3 | Groups 1-3 |
| Pairs 4-5 | Group 4 |
| Pairs 6-12 | Group 5 |
| Pairs 13-15 | Group 6 |
| Pairs 16-18 | Group 7 |
| Pairs 19-20 | Group 8 |
| Pairs 21-22 | Group 9 |
| Pair X | Group 10 |

In a few places only 9 groups are mentioned; in those cases, group 10 has been consolidated with group 5.

The automatic karyotyping problem is to find a suitable set of parameters that characterize the individual chromosomes and use them to karyotype any human chromosome complement, normal or abnormal.

2. Karyotyping Normal Cells

The first step, which is at the base of almost any work in this area, is to develop an algorithm utilizing the simplest parameters to karyotype normal chromosomes. All work thus far has used the data provided by Dr. Michael Bender of Oak Ridge National Laboratory. These data comprise the manually measured arm lengths of 50 sets of chromosomes and the independent manual karyotypes of the 50 sets. Two parameters are derived from the data: (i) over-all chromosome length, which is the sum of the average of the two long arms and the average of the two short arms (or one short arm if the chromosome is acrocentric); and (ii) centromere index, which is the ratio of the average of the short arms to over-all length.

(XXVI. COGNITIVE INFORMATION PROCESSING)

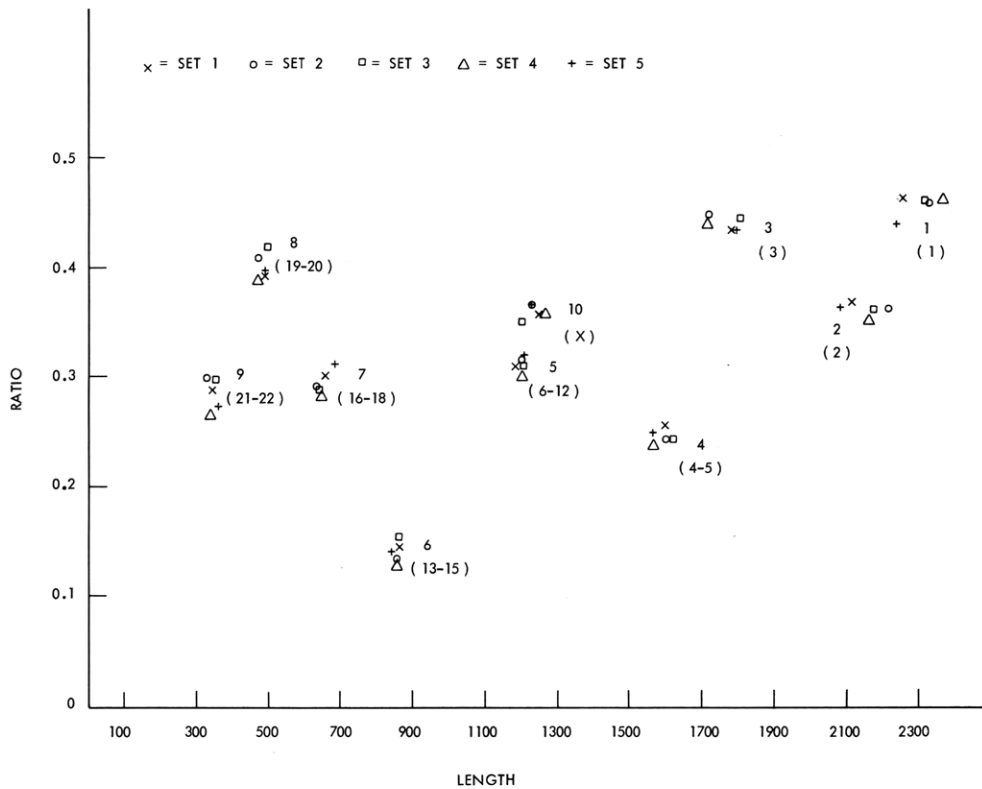


Fig. XXVI-14. Means for sets 1-5 (length normalized - 50,000 X chromosome length/sample length).

This task is more difficult than it might first appear. While the means of sample parameters are fairly stable (i.e., the standard deviation of the mean of the means is reasonably small), the standard deviation of the means themselves is quite large, even for those chromosomes that can be identified with high reliability.

The Bender data were arbitrarily separated into 5 sets of 10 cells each, labeled sets 1-5 in the following discussion. Within each set, the mean lengths and ratios for chromosomes in 10 groups established by manual karyotyping were then calculated. These mean values for all five sets are plotted in Fig. XXVI-14. The group number associated with each cluster is shown beside it and the associated chromosome numbers are shown in parentheses.

More complete data for set 1 are presented in Tables XXVI-1 and XXVI-2 and in Figs. XXVI-15 and XXVI-16. Tables XXVI-1 and XXVI-2 show the numerical values for the means of the lengths and ratios of each group in the set, together with the variance, the standard deviation (labeled SIGMA), the standard

Table XXVI-1. SET 1--LENGTH

| GROUP | MEAN | VARIANCE | SIGMA | % of mean | 1 SIGMA RANGE |
|-------|---------|----------|--------|-----------|-------------------|
| 1 | 2240.18 | 31593. | 177.74 | 7.9% | 2062.44-- 2417.93 |
| 2 | 2102.98 | 20747. | 144.04 | 6.8% | 1958.94-- 2247.02 |
| 3 | 1764.45 | 17630. | 132.78 | 7.5% | 1631.67-- 1897.23 |
| 4 | 1595.73 | 21941. | 148.12 | 9.3% | 1447.61-- 1743.85 |
| 5 | 1185.95 | 30933. | 175.83 | 14.8% | 1010.07-- 1361.83 |
| 6 | 859.93 | 8739. | 93.43 | 10.9% | 766.45-- 953.41 |
| 7 | 648.19 | 8981. | 94.77 | 14.6% | 553.42-- 742.95 |
| 8 | 481.35 | 5100. | 71.42 | 14.8% | 409.94-- 552.77 |
| 9 | 337.40 | 3386. | 58.19 | 17.2% | 279.21-- 395.59 |
| 10 | 1237.42 | 7605. | 87.21 | 7.0% | 1150.21-- 1324.63 |

DATA NORMALIZED--50000 * CHROM. LENGTH/SAMPLE LENGTH

Table XXVI-2. SET 1--RATIO

| GROUP | MEAN | VARIANCE | SIGMA | % of mean | 1 SIGMA RANGE |
|-------|-------|----------|--------|-----------|---------------|
| 1 | 0.468 | 0.00089 | 0.0298 | 6.4% | 0.439--0.498 |
| 2 | 0.374 | 0.00074 | 0.0271 | 7.2% | 0.347--0.402 |
| 3 | 0.439 | 0.00125 | 0.0353 | 8.0% | 0.404--0.474 |
| 4 | 0.261 | 0.00111 | 0.0334 | 12.8% | 0.227--0.294 |
| 5 | 0.316 | 0.00319 | 0.0565 | 17.9% | 0.260--0.373 |
| 6 | 0.147 | 0.00230 | 0.0529 | 36.0% | 0.094--0.200 |
| 7 | 0.304 | 0.00584 | 0.0764 | 25.1% | 0.228--0.381 |
| 8 | 0.396 | 0.00389 | 0.0624 | 15.7% | 0.334--0.459 |
| 9 | 0.291 | 0.00316 | 0.0562 | 19.3% | 0.235--0.347 |
| 10 | 0.360 | 0.00131 | 0.0362 | 10.1% | 0.324--0.396 |

(XXVI. COGNITIVE INFORMATION PROCESSING)



Fig. XXVI-15. Means, 1 sigma range and 2 sigma range for set 1 of Bender data (normalized - 50,000 X chromosome length/sample length).

deviation as a per cent of the mean, and the one standard deviation range of the mean. Figure XXVI-15 displays the means and the one and two standard deviation ranges about the means. Note that the groups containing the shorter chromosomes are almost overlapping at one standard deviation, and that there is considerable overlap at the two-standard deviation mark in virtually all of the groups.

An enlargement of the upper right-hand quadrant of Fig. XXVI-15 is presented in Fig. XXVI-16 with the actual data used to calculate these means and deviations also plotted. The points are identified symbolically (see legend) as to group and numerically according to which cell in set 1 they came from. The karyotyping problem is thus to separate the different symbols. Even here, in dealing with the easiest groups to karyotype, numerous exceptions to any simple rule are evident.

Tables XXVI-3 and XXVI-4 summarize similar data for all 5 sets and verify that set 1 is typical with respect to standard deviation about the mean length and mean centromere index or ratio.

Thus the typical data derived from human chromosomes can be seen to be highly variable. Before describing results of attempts to karyotype despite this

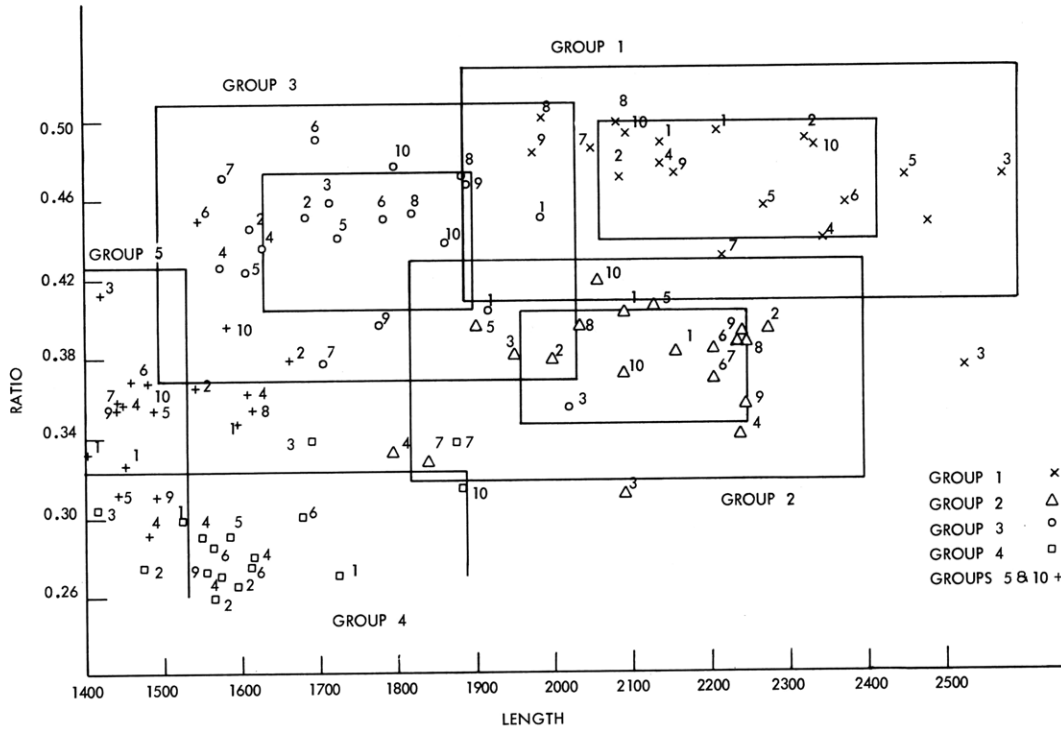


Fig. XXVI-16. Enlargement of upper right-hand quadrant of Fig. XXVI-15 giving data for calculating means and deviations.

Table XXVI-3. Length – One sigma range as per cent of mean.

| Group | Set 1 | Set 2 | Set 3 | Set 4 | Set 5 |
|-------|-------|-------|-------|-------|-------|
| 1 | 7.9% | 9.4% | 7.4% | 10.2% | 11.6% |
| 2 | 6.8 | 7.0 | 8.5 | 6.2 | 8.9 |
| 3 | 7.5 | 7.2 | 5.8 | 8.7 | 7.3 |
| 4 | 9.3 | 8.8 | 9.7 | 7.6 | 8.2 |
| 5 | 14.8 | 13.4 | 15.9 | 14.7 | 13.1 |
| 6 | 10.9 | 13.7 | 12.1 | 8.6 | 12.0 |
| 7 | 14.6 | 13.0 | 15.8 | 14.4 | 13.5 |
| 8 | 14.8 | 14.1 | 15.8 | 14.4 | 13.8 |
| 9 | 17.2 | 25.1 | 27.3 | 17.6 | 21.1 |
| 10 | 7.0 | 9.1 | 9.1 | 6.6 | 12.8 |

(XXVI. COGNITIVE INFORMATION PROCESSING)

variability, some mention should be made of attempts to reduce the magnitude of the standard deviations. The length data presented here has been normalized by the sum of the lengths of the chromosomes of each cell. This worked better than other normalization schemes that were attempted, which included normalization based on the sum of the lengths of chromosomes in subsets of the cell. Another scheme, less fully explored, was to attempt to find some significance in the compactness or state of contraction as measured by the total length of the chromosome complement.

An algorithm has been implemented which imitates the procedure used by laboratory technicians in the karyotyping process. This procedure, as implemented at present, assumes that it has the data for a normal chromosome complement and attempts to duplicate the karyotyping done by a human. It does not attempt to make any judgment on the normalcy of the cell.

This algorithm, which is extremely straightforward, has proved to work quite well. The algorithm itself was developed after spending 3 days working in the Human Genetics Laboratory at the Children's Hospital Medical Center (Boston) under the guidance of Dr. Park Jerrold. It was then tested on the data provided by Dr. Bender.

Data from 20 of the 50 complete sets of the Bender samples were used in the initial evaluation of the algorithm, and the results were used to modify the algorithm.

The algorithm was recently evaluated by running it on 10 samples previously unused in any way. With 46 chromosomes per sample, a total of 460 chromosomes was classified. Thirty-seven chromosomes were classified incorrectly – approximately 8%.

Table XXVI-4. Ratio – One sigma range as per cent of mean.

| Group | Set 1 | Set 2 | Set 3 | Set 4 | Set 5 |
|-------|-------|-------|-------|-------|-------|
| 1 | 6.4% | 4.7% | 4.2% | 4.8% | 18.0% |
| 2 | 7.2 | 8.9 | 8.5 | 6.9 | 8.9 |
| 3 | 8.0 | 7.7 | 6.2 | 6.0 | 4.4 |
| 4 | 12.8 | 13.0 | 13.3 | 11.2 | 11.7 |
| 5 | 17.9 | 17.6 | 18.5 | 17.4 | 16.4 |
| 6 | 36.0 | 30.3 | 36.6 | 36.9 | 31.9 |
| 7 | 25.1 | 25.1 | 26.2 | 27.6 | 25.1 |
| 8 | 15.7 | 13.6 | 12.8 | 18.1 | 16.1 |
| 9 | 19.3 | 23.3 | 19.0 | 22.0 | 24.2 |
| 10 | 10.1 | 7.3 | 10.8 | 10.5 | 10.7 |

Table XXVI-5. Misclassifications.

| <u>Number of Mistakes</u> | <u>Number of Sets</u> |
|---------------------------|-----------------------|
| 0 | 3 |
| 1 | 0 |
| 2 | 1 |
| 3 | 3 |
| 4 | 1 |
| 5 | 0 |
| 6 | 0 |
| 7 | 0 |
| 8 | 0 |
| 9 | 2 |

Table XXVI-5 shows the number of sets in which a given number of misclassifications were made.

Manual inspection of the data reveals essentially 3 types of errors.

1. Misclassifications resulting from the lack of a more sophisticated algorithm (which could be corrected without causing new errors).

2. Misclassifications resulting from conflicting classifications based on length or ratio. Changes in the algorithm to correctly classify these chromosomes will cause new errors.

3. Misclassification because all data (lengths and ratios) of 2 chromosomes conflict with manual classification.

The first type of error is not a serious problem. Relatively simple changes might reduce the number of incorrectly classified chromosomes by 5 to 32.

The second type presents more serious problems. While there are some obvious changes that might be made to classify many more of these chromosomes correctly, they would certainly cause errors in other samples. Whether or not there would really be an over-all gain in accuracy is hard to predict without going through a more laborious "fine-tuning" process.

This type of error is more interesting than the first because it is based on the available numerical data, and no algorithmic procedure appears to be able to duplicate the manual karyotype. But, of course, it is not known which is correct, the manual or the machine karyotype. In at least some of the cases, the human karyotyper undoubtedly used other visual information to make his decision and would defend its validity. In other cases, perhaps the karyotyper would agree with the machine, when presented with the numerical data.

The third type of error cannot be improved on at all; however, the comments about

Table XXVI-6. Misclassification matrix.

| Correct Group \ Assigned Group | Assigned Group | | | | | | | | |
|--------------------------------|----------------|----|----|----|-----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 18 | 1 | | 1 | | | | | |
| 2 | | 18 | 1 | 1 | | | | | |
| 3 | 2 | 1 | 16 | | 1 | | | | |
| 4 | | | 2 | 34 | 4 | | | | |
| 5 | | | 1 | 4 | 153 | 2 | | | |
| 6 | | | | | 2 | 58 | | | |
| 7 | | | | | | | 55 | 4 | 1 |
| 8 | | | | | | | 3 | 34 | 3 |
| 9 | | | | | | | 2 | 2 | 36 |

the second type of error also apply here.

Elimination of 5 more errors through a combination of "fine tuning" and manual reclassification is probably not too optimistic. This would reduce the over-all error rate to approximately 6% for arbitrary sets of normal cells.

Some appreciation can be gained of why the error rate is no lower by considering one example. In one sample used to evaluate this algorithm, one of the chromosomes named as being in pair 1 has an unnormalized length of 2439 units (by far the largest) and a ratio of .440. These values are quite reasonable and present no difficulty in karyotyping. The other chromosome identified as being in pair 1, however, has a length of 1417 units (the tenth largest) and a ratio of .124. These values are totally atypical, and clearly give no way of identifying it as being in pair 1.

In attempting to find an acceptable karyotype for this cell, the algorithm makes 3 errors. Clearly, a single exceptional chromosome can cause many misclassifications in a scheme such as this which relies on the cell being normal or even more strongly on its being "typical."

The misclassification matrix for this test of the algorithm, Table XXVI-6, shows that the bulk of the errors occur either in differentiating group 4-5 from group 6-12, X, or with the smaller chromosomes for which accurate measurements become more difficult and the standard deviations of parameters tend to be large.

3. Summary

An algorithm that closely imitates the approach of a human technician in performing karyotypes of normal human chromosome complements has been implemented

(XXVI. COGNITIVE INFORMATION PROCESSING)

successfully. An error rate of 8% has been achieved and an error rate of under 6% is clearly achievable. Some areas for further work are still apparent, and the extension to handling abnormal chromosome complements must be undertaken.

L. Hatfield, O. J. Tretiak

