XXVI. COGNITIVE INFORMATION PROCESSING*

Academic and Research Staff

Prof. M. Eden
Prof. J. Allen
Prof. F. F. Lee
Prof. S. J. Mason
Prof. W. F. Schreiber
Prof. D. E. Troxel
Prof. T. S. Huang
Prof. K. Ramalingasarma
Dr. R. R. Archer
Dr. W. L. Black
Dr. J. E. Green
Dr. K. R. Ingham
Dr. P. A. Kolers
Dr. O. J. Tretiak
F. X. Carroll
C. L. Fontaine
E. R. Jensen
A. J. Laurino
D. M. Ozonoff
E. Peper
Sandra A. Sommers
Y. D. Willems

Graduate Students

G. B. Anderson
T. P. Barnwell III
W. L. Bass
A. K. Bhushan
B. A. Blesser
J. E. Bowie
A. E. Filip
A. Gabrielian
A. M. Gilkes
R. E. Greenwood
E. G. Guttmann
R. V. Harris III
D. W. Hartman
P. D. Henshaw
A. B. Hayes
M. Hubelbank
A. N. Kramer
W-H. Lee
J. I. Makhoul
G. P. Marston III
O. R. Mitchell, Jr.
D. R. Pepperberg
G. F. Pfister
R. S. Pindyck
D. S. Prerau
J. E. Richards, Jr.
G. M. Robbins
S. M. Rose
C. L. Seitz
D. Sheena
W. W. Stallings
R. M. Strong
J. W. Woods
I. T. Young

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, of 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. A grant from a donor who wishes to remain anonymous made it possible for us to buy a PDP-9 computer and use it almost exclusively for this project. In November 1968, we were able to demonstrate for the first time, an integrated system in which a text page taken from a Fourth Grade reader was inserted into the

*This work was supported principally by the National Institutes of Health (Grants 5 PO1 GM-14940-02 and 5 PO1 GM-15006-02), 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 GM-01555-02).

video scanner, the text page characters were located and scanned by the computer, recognized, and identified as English words. The system then used a Grade II (or advanced) Braille translation program to actuate a modified electric typewriter capable of Braille embossing, and in this way translated the original page of printed text into Grade II Braille. The system also translated the English spelling into phonetic spelling, which in turn provided the commands to an electronic artificial speech generator and produced synthetic "reading" of the text page. Although the quality of the synthetic speech is somewhat mechanical at the present stage of development, it is basically intelligible and can be understood with a little practice. In any case, the demonstration establishes that it is technologically feasible to "read aloud" printed text at speeds entirely comparable with that of human readers.

Several improvements have been made to component parts of the system, including procedures for recognizing the printed characters and a translation and speech synthesis process, which have not yet been incorporated into the complete system. We have demonstrated, however, in a series of experiments with the speech generator that the lengthening of vowels at certain places in words and the addition of pauses at suitable places determined by analysis of the sentence structure can profoundly improve intelligibility, even for the current monotone speech of the generator.

Other preliminary experiments with sighted and with blind subjects have showed us that comprehension of synthetic speech can be learned relatively quickly and, as expected, the comprehension of sentences is much higher than that of simple words.

The development of spelled speech displays has continued toward providing special vocabularies for the use of blind persons employed in a variety of professions. A "read-only" memory containing 90 different word sounds has been completed. It has been shown that this memory can produce most of the words needed by a blind programmer or mathematician in order to communicate with a computer.

Research into human cognitive tasks has continued. The ability of humans to learn to understand and use transformed text has been studied here for several years. We have recently studied the task of learning to understand speech inverted in frequency, that is, high-frequency speech components are transformed to low-frequency and vice versa. Inverted speech can be learned in a relatively short time; this suggests that there is no immediate organic connection between the sounds that can be produced by the vocal tract and the cognitive ability to recognize sounds as language. In another experiment subjects were tested for their ability to learn to write transformed block letters. The order of difficulty of the transformations is significantly different from that of reading transformed text.

Several procedures for digital coding of images have been studied. Efficient and rapid algorithms have been found for coding the margin or contour of figures, as well as for arbitrarily connected figures. The significance of this research lies in the fact that almost all computer pattern recognition studies utilize similar algorithms, but existing algorithms are probably too slow for use in practical large-volume analysis.

Image enhancement studies have led to the design of a high-quality scanner capable of doing certain image-processing tasks independently of the computer. In one such process an image is digitized and recorded on tape. During playback the same image is somewhat defocussed and combined with the original sharp signal in such a way as to give the effect of an arbitrary optical spread function. While this process is less accurate and less flexible than computer filtering, it is very much cheaper and faster and holds promise for use in large-volume radiographic diagnosis.

Progress has been made in the analysis of biological images. A very recent achievement was a demonstration that classification of white cells can be made on the basis of only 7 measurements on each cell and that the machine classification is as accurate as that of a careful laboratory technician. Much work still remains to be done, however, in order to translate these laboratory results into a practical working system.

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

1. Reading Machine Project

The principal objectives of this project are as follows.

a. The design and construction of a machine that is capable of reading printed English aloud in connected speech form, and also of producing Braille output.

b. The study of problems associated with character recognition, the transformation into symbolic speech representation, and the generation of speech sounds.

c. The study, adaptation, and comprehension of such synthetically generated material and its utility as sensory aid for the blind, as well as a general machine-to-man communication vehicle.

d. The research makes very heavy usage of general-purpose electronic computers. No computer, at present, is sufficiently powerful (regardless of the cost) to perform the over-all task at a normal reading rate; hence, it is necessary to design and construct specialized digital apparatus for an economical real-time operation. Consequently, a systematic approach to the design of digital systems is of substantial interest. A description language, a digital simulator, a system realizer, and miscellaneous topics associated with design automation are being studied.

Specific projects that we are undertaking include the following.

e. A new vidicon scanner has been designed and construction will begin soon. This will enable us to scan book pages without removing the pages from the book as we now have to do. It is also a step toward the design of a system of lower cost than the research system now in operation.

f. Our current synthetic speech generator produces intelligible speech, but the quality is still too monotonous and "lifeless." Research will continue on the analysis of speech sounds and on better procedures for computing the individual speech sounds and – most important – the transitions between phonemes.

g. Experiments have begun to find out what design problems need to be solved before blind subjects can operate the system without the continuous presence of a sighted operator.

h. Braille produced by the reading machine and synthetic speech will always contain certain errors in transcription. Consequently, we shall continue our studies of learning and error-correction rates, optimal presentation modes, and subjective acceptability.

i. A new drum is being installed in the system. This, together with the new scanner and an additional 8,000 word core memory, will permit us to raise the level of sophistication of the text material from Fourth Grade reader to at least High School or general reading material. We shall also be able to effect substantial efficiency in computation time.

j. Work continues on more adequate recognition algorithms so that relatively poorly printed material such as newspapers can be read with no catastrophic increase in error rate.

2. Cognitive Processes

Our psychophysical and psychological studies on cognitive tasks will continue.

a. The ability of normal adults to learn to write transformed text, that is, "upside-down," "mirror-image," etc., is being studied. Preliminary results suggest that the psychological mechanism of writing is different from that of reading. Further studies will be made in which the feedback loop is opened (the subject does not see what he is writing), and in which the feedback phase is reversed (the subject sees normal writing only when he is writing upside down).

b. Studies of depth perception of real and illusory objects to continue, again with and without the feedback loop being prepared.

3. Pattern Recognition Studies

Our pattern recognition research program objectives include the following.

a. The development of a computer program that will be able to "read" printed sheet music.

b. Preliminary studies on speech recognition have been performed in which the producer of the speech will interact with the recognition system so that he can modify in a natural way the production of his speech sounds so as to enhance the computer recognition rate. This work will be extended to a larger set of speech sounds.

c. We have shown that cursive writing can be read by machine if the writing is done on-line. This research will be extended to reading handwriting without using the time information.

4. Biological Image Processing

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

a. The white blood cell differential counting project now classifies the 5 types of cells found in normal peripheral blood. This program will be extended to identify abnormal cell types, and we shall investigate the feasibility of constructing a system based on this principle that would be applicable in clinical laboratories.

b. The red blood cell classification project has progressed to the point where it will soon become a useful quantitative tool for clinical research. To gain experience in using this type of tool, we hope to begin a limited number of joint medical research projects in conjunction with hematologists from local hospitals. As experience is gained and deficiencies are unveiled, continuous improvement will be made in the system. We shall also search for ways to reduce the cost and complexity of the system, with the ultimate goal of developing a marketable clinical test instrument.

c. During the coming year, we shall begin investigation of ways to automatically handle microscopic slices from whole tissues other than blood. Initially we shall concentrate on normal tissue slices with relatively regular structure, with the goal of ultimately handling pathological tissues.

d. Our chromosome karyotyping project is proceeding along two lines. The more powerful system will abstract data from a photograph with a scanner, and construct a caryogram from this information. This project had been suspended during the past year, while our new facility was assembled. Our second approach is to trace chromosome outlines with a graphic input device, and to perform the analysis from the outline tracing information.

e. In the coming year, we expect to complete the initial phase of the Papanicolau smear analysis project. We shall accumulate data on nucleus and cytoplasm sizes of normal and neoplastic isolated cells in a Papanicolau preparation. The primary problem is one of overcoming the difficulties that arise, because of the low contrast between the background and the cytoplasm.

f. We shall continue our work on applying computer graphics to the study of neural anatomy. We plan to complete two projects: (i) construction of three-dimensional drawings from serial sections; and (ii) three-dimensional recording and display of Golgi-stained neurons.

5. Picture Processing

A significant part of our objectives relate to image-processing techniques.

a. Studies of subjective effects of visual "noise" on picture quality continue. Also, a study of the visual system's time and frequency response characteristics will be performed. Noise images of known spatio-temporal spectrum will be produced and psychophysical response parameters measured. This method appears to be a more general one than methods used by previous investigators.

b. Recently developed algorithms for contouring and coding pictures will be programmed and tested for efficiency and subjective picture quality.

c. From theoretical considerations, we have recently come to realize that, in principle, a three-dimensional reconstruction of an object can be made from x-ray images taken at several different orientations of the object. We intend to perform some experiments, making the image reconstruction by digital computation.

d. Work on coherent optical processing continues. We have been able to reconstruct two-dimensional spatial functions, that is, pictures, following space and frequency domain filtering. But there is a need to eliminate sources of error that distort the reconstruction.

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

A. VOLUNTARY ATTENTION SWITCHING BETWEEN FORESIGHT AND HINDSIGHT

As you move your eyes about a room, the room and its contents remain approximately stationary, despite the relative displacement between rays of light and the retina on which they impinge. With very careful observation, some displacement of objects can be detected in a direction opposite to eye motion; displacement can be detected even during a slow blink of the eyes.[1] In ordinary viewing of a scene, however, these motions go undetected. A continuing problem in the study of visual perception is how to account for the stability of perceived objects in space during movements of the eyes, the head, and the body. The observations reported here raise the possibility that the stability is achieved by means of learned compensations for directed image motions across the retina.

The experiments were made with a helmet-mounted system of mirrors that simultaneously provides a view of the straight-ahead and the straight-behind (foresight and hindsight) (Figs. XXVI-1 and XXVI-2). Mirror A before the eyes is only partially coated, so that it transmits and reflects equal amounts of light. Mirrors B and C above and behind the head are fully reflecting. The path of light rays from an object behind the observer is from the object to Mirrors C, B, A in that order, and then into the eyes; objects in front of the observer are seen directly through partial-mirror A.

In the tests cardboard frames and black cloth are used to make the fields of view equal in size and to prevent light from the sides of the apparatus from reaching the eyes. I have performed two experiments with this apparatus, one with the

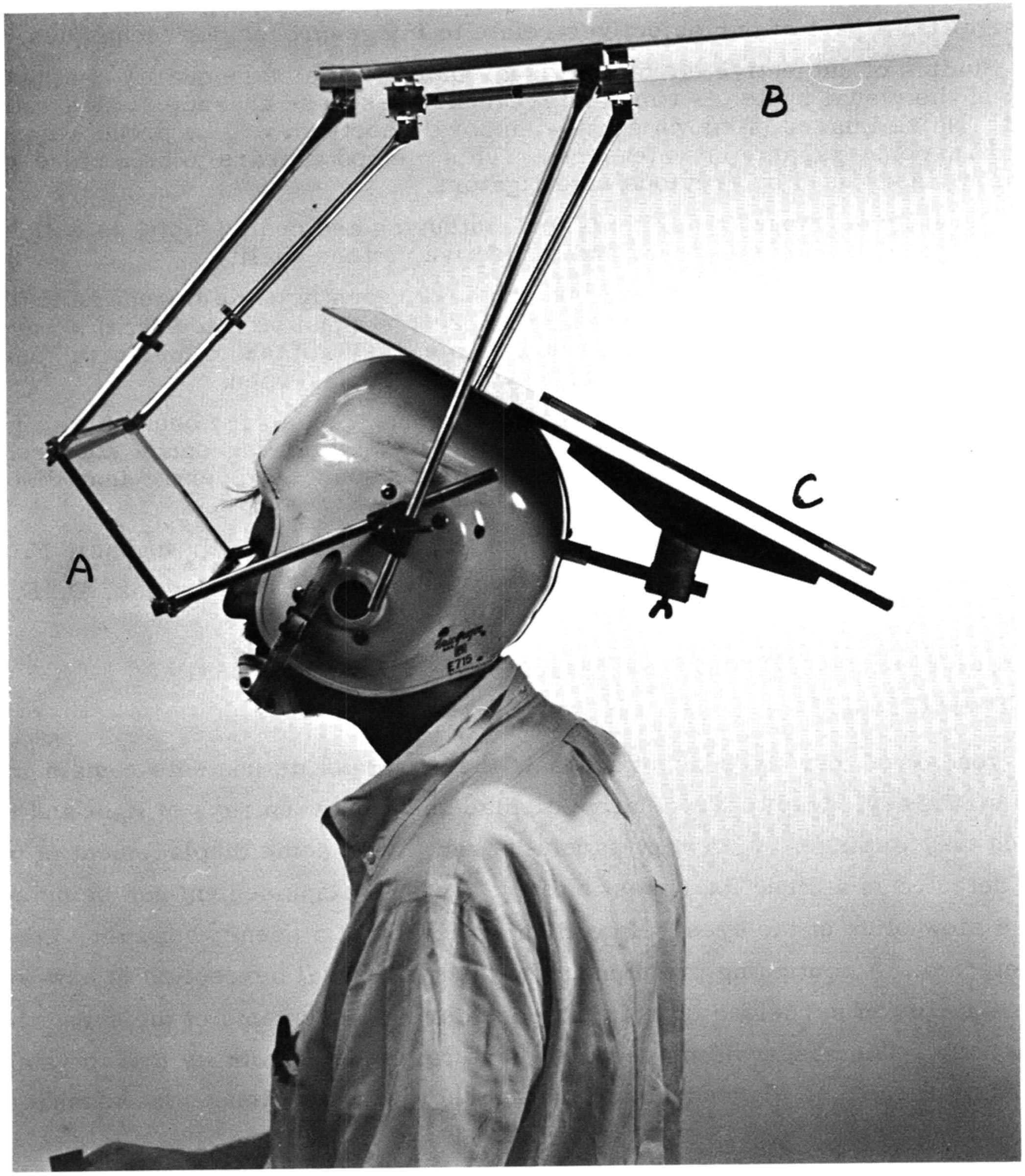

Fig. XXVI-1. Helmet-mounted system of mirrors provides simultaneous views of straight-ahead and straight-behind. In a normally lighted environment users can voluntarily select one scene or another for attention. (Light baffles normally present for experiments are not shown.)

Fig. XXVI-2. Front view of the device. The half-silvered mirror is in place before the observer's eyes.

observer stationary and the second with the observer ambulating.

The device that is illustrated creates an optical melange on the eyes, a combination of the scenes ahead and behind. With only a moment's hesitation the stationary observer can select one or another scene for attention and experiences little if any difficulty in alternating his choices. When he directs attention to one scene, it dominates the field of view. When he switches to the other, the first scene disappears and the second comes into view. The switching is voluntary and in this respect is markedly different from the uncontrollable switching of retinal or binocular rivalry. Viewers report that they switch by concentrating upon features of a scene, which then comes into view; that is, the construction of a scene, while rapid, is piecewise.

Switching and selective attending are facilitated by context. Objects that normally belong together are seen together. Two people known to an observer and stationed at optically equal but opposite distances from him are convenient targets. The head of each is always seen on its own body, and only one of the people is clearly perceived at a time, the other person becoming invisible or, at most, being seen only dimly. As nothing in the mirror system conveys information to the stationary observer about location of the targets, his knowledge or inference about what he is looking at and its structural unity[2] must be the aids he uses in switching attention.

When we manipulated intensity, contrast, or other parameters that enhance the distinctiveness of objects, the enhanced objects, regardless of their position, intruded into the visual scene. The seated observer could not successfully choose, for example, to see a dimly lit scene of low contrast (either in front or behind) when the other contained well-lit objects of high contrast. Otherwise, the switching was trouble-free and easy to accomplish. The significant feature of this result is that the optical melange on the retinas was resolved into two scenes only one of which was attended to at a time while the other was more or less suppressed or ignored.

This selective switching of visual attention and its concomitant suppression of simultaneously present but unwanted information has a possible analogue in audition: the well-known cocktail party effect.[3] In neither case can the information processing be identified with peripheral coding of inputs.

When the viewer ambulates, directional information is provided in a dramatic way. The scene in front of the eyes remains stationary, as in normal viewing, but the scene behind the head moves. The direction of movement that can be detected in normal viewing of the straight-ahead is opposite to the direction of eye motion (moving the eyes rightward induces a slight sense of leftward object motion); the direction of motion of the scene from behind the head that is reflected by the mirror system is the same as the head motion. Turning to the right causes the reflected scene to turn rightward and tilting the head produces a like tilt of the reflected scene. Thus a single willed movement of the body produces two perceptual consequences, one that maintains "world constancy" and the other that violates it.

Why should hindsight be so sensitive to disruption and foresight be so stable? A potential clue to the answer is provided by considering the direction of image motion across the retina. Image motion in foresight is opposite to direction of eye motion, while image motion in hindsight, as pointed out above, is in the same direction as eye motion. It is implausible that direction of image motion across the retina by itself is the cue for compensatory computations. Instead, it seems plausible that the skilled observer has had to learn to correlate not merely motions, but directions of motion with respect to his own body. It is not the intention to move or the movement itself that provides the information needed for compensation; rather, the information must be interpreted within a directional frame of reference or orientation set.[4]

Some years ago Stratton,[5] wearing lenses that reversed right and left, experienced a "swinging field" when he first put them on. With continued wearing of the lenses, object motion induced by motions of the eyes and head diminished. While his experiments contain many ambiguities,[6] this observation suggests that compensation for object motion is indeed an acquired visual skill that may be brought to operate within different orientation sets.[7]

P. A. Kolers

References

1. H. Carr, Psychol. Rev. Monogr. Suppl. 7, No. 3, 1906.
2. P. A. Kolers, Recognizing Patterns: Studies in Living and Automatic Systems, P. A. Kolers and M. Eden (eds.) (The M.I.T. Press, Cambridge, Mass., 1968), pp. 50-56.
3. C. Cherry, On Human Communication (The M.I.T. Press, Cambridge, Mass., 2d edition, 1966), p. 279ff.
4. P. A. Kolers and D. N. Perkins (to appear in Perception and Psychophysics). We note too that K. Duncker [Psychologische Forschung 12, 180 (1929)] analyzed a number of induced motion effects in terms of the "frame of reference" within which objects were perceived. It is possible that Duncker's frame of reference for foresight is the fixed scene itself, but that the frame of reference for hindsight is established by the moving mirror system.
5. G. M. Stratton, Psychol. Rev. 4, 341; 463 (1897).
6. G. L. Walls, Am. J. Optom. 28, 55; 115; 173 (1951); C. S. Harris, Psychol. Rev. 72, 419 (1965).
7. See also J. G. Taylor, The Behavioral Basis of Perception (Yale University Press, New Haven, 1962), p. 198ff.

B. DERIVATION OF PULSE-INVARIANT DIFFERENCE EQUATIONS

There is a simple way of finding the difference equation with an impulse response equal to the sampled impulse response of a given continuous filter.[1] But in modeling a continuous filter by this difference equation one must approximate the waveform input to the continuous filter by a sequence of impulses. Consequently, the analogy between the continuous system and the discrete system is not obvious.

A difference equation will be derived here whose response is equal to the sampled pulse (with width equal to the sample period) response of the continuous filter. The analogy between the continuous and discrete systems should be clearer as shown in Fig. XXVI-3. The analogy is clearer because the sampled

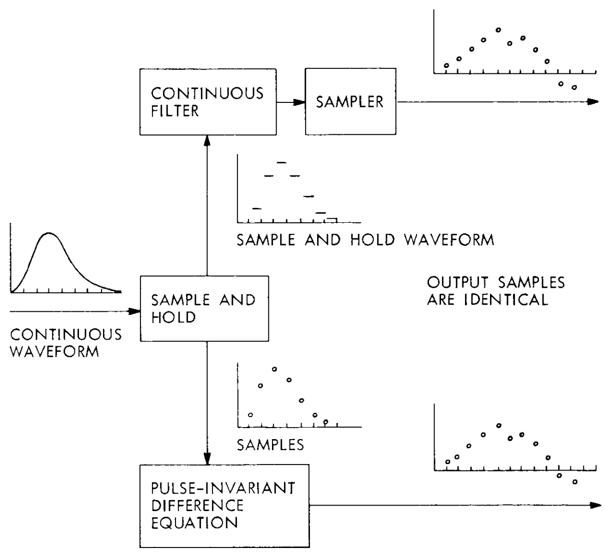

Fig. XXVI-3. Relation between the pulse-invariant difference equation and the differential equation filter.

output of the continuous system for a sequence of pulses (more tangible than a sequence of impulses) will be exactly equal to the output of the difference equation.

Let the continuous system be represented in state variable form:

$$\dot{\underline{x}} = \underline{A}\,\underline{x} + \underline{b}\,u$$

whose solution at time t is written[2]

$$\underline{x}(t) = e^{\underline{A}t}\,\underline{x}(0) + \int_0^t e^{\underline{A}(t-\sigma)}\,\underline{b}u(\sigma)\,d\sigma.$$

Let the sampling period be T and write

$$\underline{x}_n = \underline{x}(nT).$$

Thus

$$\underline{x}_n = e^{\underline{A}nT}\,\underline{x}_o + \int_0^{nT} e^{\underline{A}(nT-\sigma)}\,\underline{b}u(\sigma)\,d\sigma$$

$$\underline{x}_{n+1} = e^{\underline{A}T}\,e^{\underline{A}nT}\,\underline{x}_o + e^{\underline{A}T}\int_0^{nT} e^{\underline{A}(nT-\sigma)}\,\underline{b}u(\sigma)\,d\sigma$$

$$+ e^{\underline{A}T}\int_{nT}^{(n+1)T} e^{\underline{A}(nT-\sigma)}\,\underline{b}u(\sigma)\,d\sigma.$$

Consequently, the state-variable difference equation is

$$\underline{x}_{n+1} = e^{AT}\,\underline{x}_n + e^{AT}\int_0^{T} e^{-AS}\,\underline{b}u(s+nT)\,ds.$$

Up to this point no approximation has been introduced, and the difference equation above gives the output of the continuous system exactly at the sample times. If one now introduces the approximation $u(t) = \Sigma\, u_i\delta(t-iT)$, then impulse-invariant difference equations may be found. Of course, for the impulse approximation it is much easier to use a different method.[2]

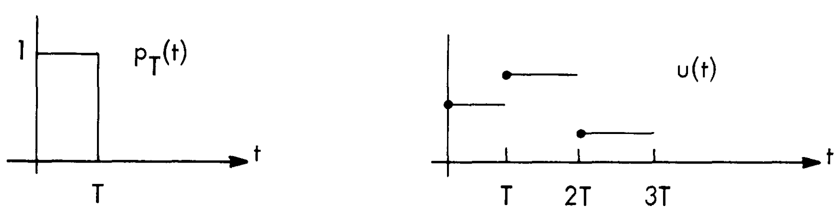

Fig. XXVI-4. Definition of $p_T(t)$ and an illustration of u(t).

To obtain the pulse-invariant filter let

$$u(t) = \sum u_i p_T(t-iT),$$

where $p_T(t)$ and u(t) are as shown in Fig. XXVI-4. The difference equation becomes

$$\underline{x}_{n+1} = e^{\underline{A}T}\,\underline{x}_n + e^{\underline{A}T}\int_0^{T} e^{-\underline{A}S}\,ds\,\underline{b}u_n.$$

Evaluating the integral, we obtain

$$\underline{x}_{n+1} = e^{\underline{A}T}\,\underline{x}_n + \underline{A}^{-1}(e^{\underline{A}T}-\underline{I})\,\underline{b}u_n.$$

At this point we must evaluate $e^{\underline{A}T}$ and $\underline{A}^{-1}$. Then if a difference equation for a particular one of the state variables in $\underline{x}_n$ is desired, the others must be eliminated from a system of algebraic equations.

As an example of this procedure a difference equation for a single pole-pair resonant filter will be found. The differential equation governing such a system is

$$\ddot{y} + 2\sigma\dot{y} + (\sigma^2+\omega^2)y = (\sigma^2+\omega^2)u.$$

Notice that the DC response is unity.

Let $x = y$, $w = \dot{y}$ be the state variables, then

$$\begin{bmatrix} \dot{x} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -(\sigma^2+\omega^2) & -2\sigma \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma^2+\omega^2 \end{bmatrix} u$$

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -(\sigma^2+\omega^2) & -2\sigma \end{bmatrix}$$

$$e^{\underline{A}T} = \begin{bmatrix} \frac{\sigma}{\omega} e^{-\sigma T} \sin \omega T + e^{-\sigma T} \cos \omega T & \frac{1}{\omega} e^{-\sigma T} \sin \omega T \\ -\frac{\sigma^2+\omega^2}{\omega} e^{-\sigma T} \sin \omega T & e^{-\sigma T} \cos \omega T - \frac{\sigma}{\omega} e^{-\sigma T} \sin \omega T \end{bmatrix}$$

$$\underline{A}^{-1} = \frac{1}{\sigma^2+\omega^2} \begin{bmatrix} -2\sigma & -1 \\ \sigma^2+\omega^2 & 0 \end{bmatrix}.$$

Now, for ease of writing, let

$$e^{\underline{A}T} = [a_{ij}], \qquad \underline{A}^{-1}(e^{\underline{A}T} - I)\,\underline{b} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}.$$

Thus

$$x_{n+1} = a_{11}x_n + a_{12}w_n + c_1u_n$$

$$w_{n+1} = a_{21}x_n + a_{22}w_n + c_2u_n.$$

Using the third equation

$$x_{n+2} = a_{11}x_{n+1} + a_{12}w_{n+1} + c_1u_{n+1},$$

one may eliminate w_n and w_{n+1} to obtain

$$x_{n+2} = 2e^{-\sigma T} \cos \omega T\, x_{n+1} - e^{-2\sigma T} x_n + \gamma_1 u_{n+1} + \gamma_o u_n,$$

where

$$\gamma_1 = 1 - e^{-\sigma T} \left[\cos \omega T + \frac{\sigma}{\omega} \sin \omega T\right]$$

and

$$\gamma_o = e^{-2\sigma T} + \frac{\sigma}{\omega} e^{-\sigma T} \sin \omega T - e^{-\sigma T} \cos \omega T.$$

As a partial check this equation does have unity gain at DC.

It is possible to generalize this technique to better approximations to the input u(t), such as the trapezoidal or parabolic approximations. It might be noted that these will leave the homogeneous difference equation unchanged, but will make the excitation terms more complicated.

G. M. Robbins

References

1. F. Radar and B. Gold, "Digital Filter Design Techniques in the Frequency Domain," Proc. IEEE 55, 149-171 (1967).
2. R. W. Brockett, "Course 6.60 Notes," M.I.T., 1968 (unpublished).