VOLUME 11 □ NUMBER 7 IEEE spectrum JULY 1974

IEEE spectrum

articles

departments

Not long ago, you couldn't get a Nova 2 for love or money.

Now all you need is money.

If you're willing to buy five Nova 2's at a time with 16K or 32K memory in each, you can get them in any quantity, within 60 days.

And for those who only want one, we've come up with five different Nova 2 systems you can buy one at a time.

The smallest is built around our RTOS operating system. Which makes it a great little real-time data acquisition satellite.

Next up the line is our SOS system. A small single-user computation system for people who don't need a disc.

Then there's the medium-sized RDOS system you can use to develop software and hardware interfaces, or to build your own medium-sized computation system.

Which is followed by a big RDOS system with more storage and faster output for higher throughput.

And finally, for the people who need a system with a built-in backup computer, there's the Dual Nova 2.

Now, everybody who wants Nova 2's can get them.

And 60 days after our factory gets your order, you'll get your system.

DataGeneral

The computer company you can understand.

Data General Corporation Southboro, Massachusetts 01772, (617) 485-9100.
Datagen of Canada Ltd., Hull, Quebec (819) 770-2030 Data General Europe, Paris, France 504-23-44.

Circle No. 1 on Reader Service Card

the cover

Skylab 3 astronaut Owen Garriott deploys an experiment to collect material from interplanetary dust particles as the space station travels in Earth orbit. An exclusive interview with Garriott begins on page 50.

IEEE Publications Operations

IEEE SPECTRUM is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: 345 East 47 Street, New York, N.Y. 10017. Cable address: ITRIPLEE. Telephone: 212—752-6800. Published at 20th and Northampton Sts., Easton, Pa. 18042. **Change of address** must be received by the first of a month to be effective for the following month's issue. Please use the change of address form below. **Annual subscription:** IEEE members, first subscription $3.00 included in dues. Single copies $3.00. Nonmember subscriptions and additional member subscriptions available in either microfiche or printed form. Prices obtainable on request. **Editorial correspondence** should be addressed to IEEE SPECTRUM at IEEE Headquarters. **Advertising correspondence** should be addressed to IEEE Advertising Department, at IEEE Headquarters. Telephone: 212—752-6800.

OTHER IEEE PUBLICATIONS: IEEE also publishes the PROCEEDINGS OF THE IEEE and more than 30 Transactions for IEEE Societies and Groups with specialized interests within the electrical and electronics field. Manuscripts for any IEEE publication should be sent to the editor of that publication whose name and address are shown on pp. 101–102 of the January 1974 issue of *Spectrum*. When in doubt, send the manuscript to E. K. Gannett, Publishing Services, at IEEE Headquarters, for forwarding to the correct party.

IEEE also copublishes, with the Institution of Electrical Engineers of London, England, *Electrical & Electronics Abstracts* and *Computer & Control Abstracts*, as well as their companion titles journals, *Current Papers in Electrical & Electronics Engineering* and *Current Papers on Computers and Control*.

MOVING? PLEASE NOTIFY US 4 WEEKS IN ADVANCE

Name (Please Print)

Address

City State/Country Zip

MAIL TO: IEEE CODING DEPARTMENT
345 East 47 Street, New York, N.Y. 10017

ATTACH LABEL HERE

- This notice of address change will apply to all publications to which you subscribe.
- List new address above.
- If you have a question about your subscription place label here and clip this form to your letter.

spectral lines

Who gets published?

You may recall a reader's recent letter to *Spectrum* in which he raised the question of what the criteria are for having an article accepted for publication in the technical press, and, specifically, whether an author's previous "track record" has any bearing on the matter. Undoubtedly the answer varies depending upon the association or publishing house, and even upon the inclination of an individual editor or reviewer. Nevertheless, a summary of *Spectrum*'s policy and methodology in this regard is in order. In subsequent discussions, we shall hope to describe how the matter is handled by *Proceedings* and *Transactions*, too.

As a general Institute policy, each Editor is given the authority and responsibility for the content of the publication of which he is in charge. The import of this will become clearer later. It is further the intent of the Institute that all material published by IEEE undergo review procedures, and that such procedures be designed to insure that reviewers are chosen from among recognized experts in the appropriate portions of the field.

In the case of *Spectrum*, here is how it works. First of all, many of *Spectrum*'s articles are invited as opposed to "over the transom." These are intended to fulfill a planned matrix of topics extending a year or more into the future. The types of articles *Spectrum* seeks differ from those sought by *Proceedings* and the various *Transactions*. Specifically, we want review and tutorial articles (not as research oriented as *Proceedings* articles), articles of application and economic significance, and, occasionally, theoretical articles of broad and fundamental import. Furthermore, material limited in interest to a small or specifically defined portion of the membership is inappropriate for *Spectrum*, though not necessarily for *Proceedings* or *Transactions*.

Taken together, these criteria suggest that *Spectrum* editors are seeking the "best" author to handle a broad, significant topic. More often than not, this means an author with a track record—one who has had published many articles in his field.

But whether a manuscript is solicited or unsolicited, it must pass the muster of peer review. At *Spectrum*, an initial review is conducted by staff members. Having survived that, at least two expert, nonstaff reviewers are assigned the manuscript. Occasionally, a generalist is added to help judge an article's breadth of interest. On complex or controversial topics, as many as ten reviewers have been enlisted. In all cases, reviewers are asked to critique not only the manuscript's technical accuracy, but its timeliness, significance, and suitability to *Spectrum*'s broad audience of EEs working in a variety of disciplines. *Spectrum* encourages reviewers to provide detailed comments and suggestions for additions to manuscripts. We'd rather not have the one-line review: "Excellent article. Recommended for publication," or "This article is not for *Spectrum*." We strongly discount such reviews.

Likewise comments such as "This would be a fine article; unfortunately, this is not the man to write it" are not very useful. We'd want to know what is specifically wrong with the manuscript in question, who the reviewer recommends as author, and why.

In addition to solicited manuscripts, about one half of the articles published in *Spectrum* are staff-written, and these, too, are subjected to expert-review procedure. An incidental comment on when and why we elect to staff-write an article is appropriate. Basically, if we cannot identify an expert author (or group of coauthors) for a given article, we assign it as a staff project. This usually happens when a subject is too unwieldy (requiring many inputs from many, sometimes competitive, sources), too controversial for a single expert to report on objectively, or the like.

Readers raise the question of reviewer bias. Is it a factor? Does not-invented-here creep in? We think so. This is why we sometimes invest in a second and even third round of reviews. But in such cases we are not conducting an election (5 yes's, 4 no's—so publish!). We are really seeking sufficient reasoned inputs to help us reach a decision. In the end, we may even override what seems to be a consensus. Taken together, a set of reviews may lean toward "don't publish," but we may elect to, anyway. The judgment is often based on editorial intuition, bolstered by the editorial independence cited earlier. (We remember a technique used by Edward Bursk, now retired from his post as editor of the *Harvard Business Review*. In each issue of *HBR*, the staff would select one article that failed to meet the standard criteria for publication, but nevertheless had intuitive appeal. Often that article turned out to be the most popular of the issue.)

Finally, on the question of reviewer bias as it relates to the author's status or prior accomplishments, we think it does exist. For example, even the most conscientious and careful reviewer is likely to assume accuracy of certain portions of a manuscript written by say, Dennis Gabor, while he might be more skeptical in the case of an author unknown to him. But is this all bad? An alternative would be to protect the identity of the author from reviewers. Readers' opinions on the advantages and disadvantages of this proposal are welcome.

Donald Christiansen, Editor

Putting numbers where your hunches used to be

Decision analysis techniques help define 'best bets.' But finding a green path through gray statistics is part art, part science

Mixing midnight oil and raw nerves, Astable Industries is doggedly pursuing another important corporate decision. The manager of engineering has presented the division vice president with a new design for two assemblies of a new computer. This design replaces older bipolar discrete circuits with metal-oxide semiconductor, large-scale integration (MOS/LSI), which the engineering manager believes is superior. The vice president has a very uneasy feeling about MOS/LSI, and the two argue for an hour without reaching agreement.

Tired, but perseverant, the vice president then asks the company's newly formed decision analysis group for assistance. Since the main advantage of MOS/LSI, low cost, is partially offset by new technology risks, the problem seems basic: do the potential savings warrant taking chances?

This particular case is typical of the problems facing design engineers trying to determine the "best" design for an assembly or a product. Solutions involve performing three tasks:

- Identifying the alternatives.
- Defining the criteria (i.e., what is meant by "best").
- Developing a method for evaluating the alternatives in terms of the criteria.

Curt F. Fey Xerox Corp.

Key issues and alternatives are not always what they first appear to be, and defining "best" is very subjective. (The manager of engineering defines "best" as "least cost"; the vice president as "minimal risk." Others think of "best" in terms of technical performance.) For this discussion, "best" will be that alternative which produces the most desirable consequences in terms of the preferences of the decision-makers.

Should the first two tasks be accomplished, evaluating alternatives can still pose formidable obstacles since the data required may not be available. But valid analysis may still proceed even when hard facts are absent.

Decisions grow on trees

A quick calculation indicates that MOS/LSI could, indeed, produce substantial savings in parts cost. These savings have to be weighed against the risks which include:

- Subsystem specification changes.

Example A: The confidence game

Buoyed by an unexpected Government contract, Astable Industries developed a computer which is undergoing flight tests in each of three spacecraft. Its vice president is aboard spacecraft number three. Before the first launch, he asks his engineering manager, "What is the probability of a computer malfunction in spacecraft three?"—"Less than one in a million," answers the engineer confidently. During launch, computer error causes the first spacecraft to explode. "What is the probability of a computer malfunction in spacecraft three?" repeats the vice president anxiously.

Fortunately, the engineering manager is familiar with Bayes' conditional probability theorem and updates his initial "one-in-a-million" estimate based on new information provided by the explosion. Bayes' Boolean expression relates two events, A and B, and the current state of knowledge E in terms of conditional probabilities:

$$P(A|BE) = P(A|E)\frac{P(B|AE)}{P(B|E)}$$

Applied to the computer failure:

$P(B|E)$ = Probability of computer one failure *prior* to launch of spacecraft one = 10^{-6}.

$P(A|E)$ = Probability of computer three failure *prior* to launch of spacecraft one = 10^{-6}.

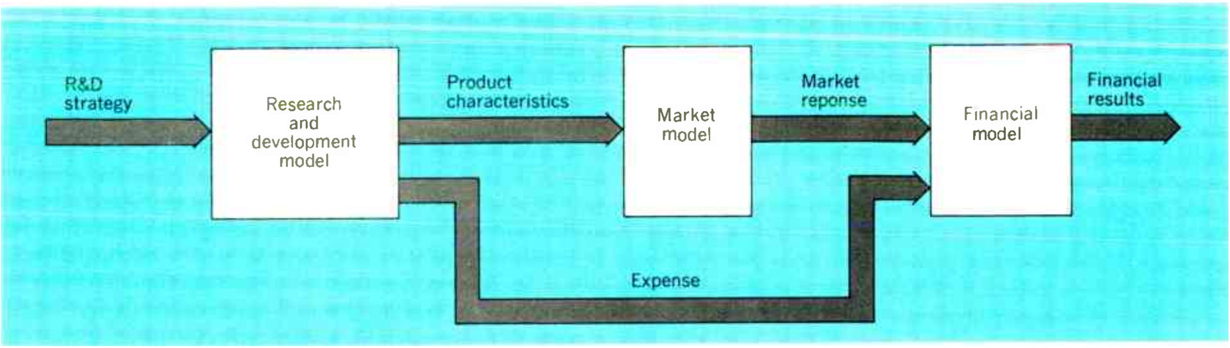
$P(B|AE)$ = Probability of computer one failure *given* computer one failed = 1.

$P(A|BE)$ = Probability of computer three failure *given* computer one failed = ?

Note that *two* states of knowledge are considered, one involving prelaunch and the other postlaunch data for the first spacecraft. Substituting the known probabilities into Bayes' relationship and solving for $P(A|BE)$, we find:

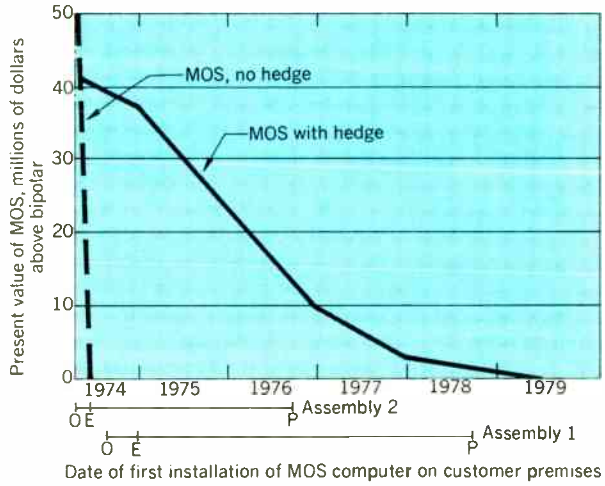
$$P(A|BE) = (10^{-6})\cdot\frac{(1)}{(10^{-6})} = 1$$

Confronting further disaster as a statistical certainty, the vice president promptly resigns his role as astronaut. Identification of design problems (plus rework) should precede additional flight tests.



[1] Working through three basic model structures, decision analysis predicts how various research and development strategies affect the financial return from a new product while determining when key events occur in time.

[2] Delivery problems drastically affect the advantage MOS alone can provide over bipolar design, but hedging MOS development with a bipolar backup softens the impact of delayed shipments considerably. Optimistic (O), expected (E), and pessimistic (P) installation dates are shown for two computer assemblies.



• Delays of prototype circuits if subsystems specifications change before machine introduction.
• Delays in production of circuits in quantity.
• Difficulties if a computer must be modified after introduction.

Some of these risks could be reduced through management actions that amount to buying insurance:
• Developing both designs sidestepping late completion of MOS.
• Ordering production quantities of both circuits, circumventing MOS manufacturing difficulties.
• Delaying release of MOS circuits until production lots have been tested, further negating manufacturing problems.
• Establishing inventories to buffer short supplies.

Risk related information is frequently processed in terms of probability. Probability is not an independant, measurable property of materials (like size, shape, color, etc.), but rather represents a numerical encoding of information by an observer. As new information becomes available, an observer will change his probability assignment even though the underlying situation has not changed. The box on page 34 illustrates this point in the development of a formal method for updating probability assignments on the basis of new information.

Per the boxed example, an analysis of risks and risk reduction leads to a number of options. These options or alternatives can be presented by a graph called a *decision tree* (see boxed example on page 39). Each square node of the tree indicates a decision point; each branch an alternative. The preferences for alternatives depend upon information not presently known. For instance, the computer design eventually issued will be influenced by prototype test results that are not now available. These chance events will be brought into the analysis later. For the moment we want to obtain quickly a rough estimate of the key issues and the most attractive alternatives.

Best quest

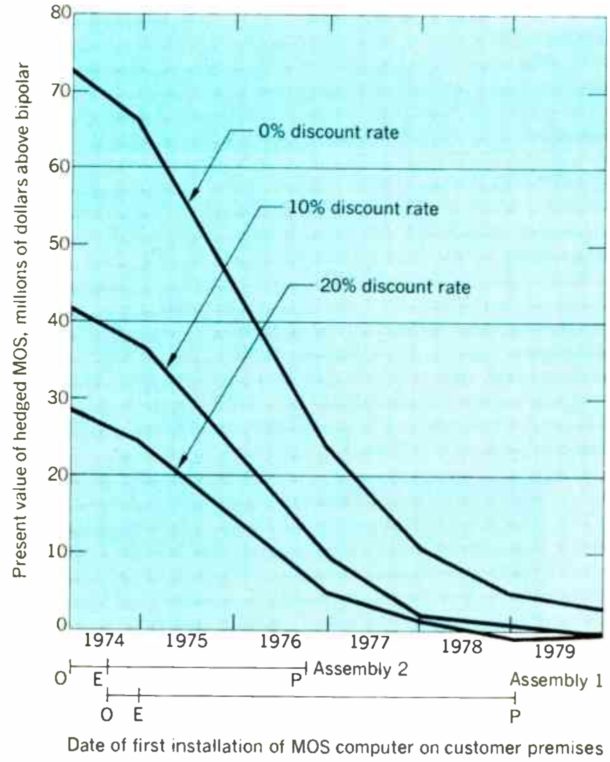
The second task in the analysis is to define "best." In practice, the "best" machine is the one the decision-maker prefers. Ideally, the "best" machine should be the cheapest, the most reliable, and the first available, but in reality we seldom find one machine "best" in every respect. Which characteristic is more desirable? Some single-choice criterion is needed that incorporates all features of the machine (reliability, cost, availability, etc.). In this analysis, a financial measure is used as the single criterion.

Profit provides a good first approximation. Using this yardstick, one soon discovers that two machines can be equally profitable but differ in attractiveness. For example, one machine could start making profits this year; the other, ten years from now. So a financial measure of "best" must include the *time value of money*. Money available today is worth more than money accruing later (the earlier the profit the better). To be precise, we must determine how much higher one values a dollar today than a dollar next year.

One measure of this differential is the *discount rate*. A discount rate of 20 percent signifies that two business ventures are considered equally attractive if one venture yields one dollar today and the other yields one dollar and 20 cents next year. This implies a borrowing cost (after taxes) of 20 percent or the availability of other investments that will yield at least 20 percent (after taxes). More details are given in the box on the top of page 39.

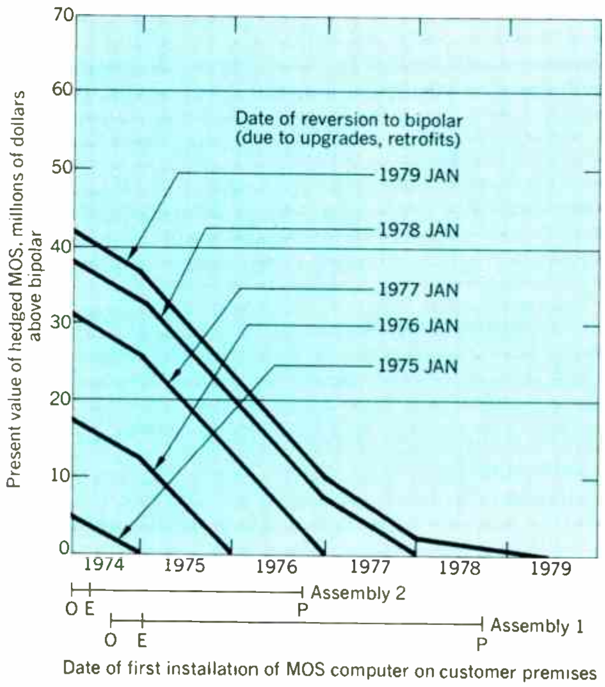
With the financial measure as our criterion, we now determine the criterion value. For this, a model is developed (Fig. 1), which consists of three parts: research and development, marketing, and finance.

The research and development model takes research and development strategy as its input and yields product characteristics (timing, costs, and performance).



[3] High discount rates on future dollars earned can significantly reduce the computer's present value, but the decision to develop MOS with a bipolar hedge is *not* affected by such money market conditions.

[4] Present value estimates (for a MOS with bipolar hedge program) increase the longer MOS computers can be sold without reverting to bipolar designs.



The market model takes a product of given characteristics and determines the market response in terms of units sold, lease cancellations, and revenues. The market model is *not* concerned with how the product was developed (in fact, market response to hypothetical machines can be determined).

The financial model takes the revenue flow and expense generated by research, development, manufacturing, and market conditions to determine financial results. One of this model's key functions is to determine when certain events occur and then identify costs. The important events are the start of the last redesign, the duration of development, parts lead time, parts acceptance, cut-in of MOS, duration of assembly and shipping, machine introduction, identification of major upgrade, lead time for upgrade, cut-in of upgrade. The method of determining times and costs will be further explored after examining some basic costing sensitivities.

First, two cases are evaluated to determine the primary nature of financial estimates (Fig. 2). Below the x-axis are two lines indicating when computers in the field will first have MOS in assemblies 1 and 2. Three estimates are shown for each assembly: optimistic, expected, and pessimistic. The y-axis denotes any increase in present value from using MOS instead of bipolar in the computer. The solid curve describes this incremental present value as a function of when MOS assemblies are available (if MOS strategy is hedged through a backup bipolar design). The dashed line describes incremental present value as a function of when MOS assemblies are available if no hedge is maintained. If all goes well, the no-hedge strategy is clearly superior; but should a small slip occur, the results are disastrous.

How sensitive is the analysis to various assumptions? Figure 3 shows the effect of varying discount rates. Note that the value of the project changes drastically with discount rate, but hedged MOS remains

I. Spotting the sensitive among the expensive

Variable	Present Value, Millions of Dollars: Optimistic	Present Value, Millions of Dollars: Pessimistic	Sensitivity, Millions of Dollars
Time of MOS cut-in	40	6	34
Discount rate	67	35	32
Date of major bipolar upgrade cut-in	36	9	27
Duration of development	39	23	16
Date of assembly performance spec	44	29	15
MOS chip cost	45	30	15
Computer sales	42	19	13
Rate of board or replacement	44	33	11
Duration of procurement	39	30	9
Computer introduction date	39	32	7
Reversion to bipolar	39	33	6
Probability of inability to obtain MOS	41	35	6

preferable. The *decision* is insensitive to the discount rate, but the *outcome* is very sensitive to that rate. Since decision analysis concerns factors that affect decisions rather than outcomes, we need not become further involved with the discount rate in this particular case. Figure 4 shows the effect on present value of potential reversion to bipolar. An early reversion could indeed make MOS unattractive.

The finances of any large technical undertaking hinge on many variables (discount rate is just one example) and each can be considered from both optimistic and pessimistic viewpoints. The differences in present value between two such projections (sensitivity) helps identify those variables that bear further investigation (Table I).

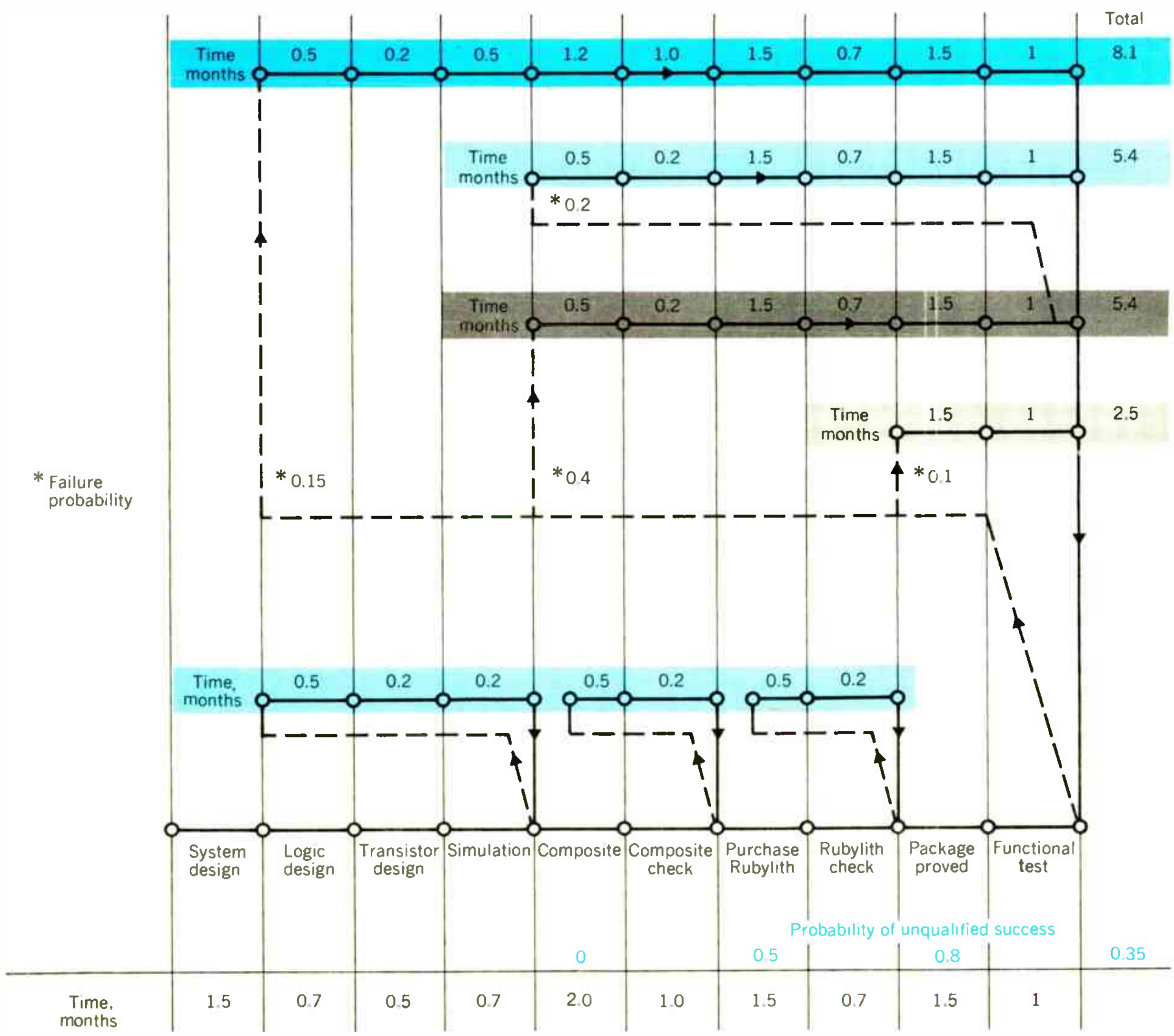
A path around the pitfalls

While our preliminary discussion concerned only the known alternatives open to a decision-maker, responsible, detailed analysis must also handle certain chance events. Incorporated in a decision tree, chance events are represented by circles; decision points, by squares (again, see boxed example on page 39). Selection of strategy (bipolar, MOS, MOS and hedge) by definition, is strictly up to the decision-maker, while other events can only be described by probability distributions (i.e., the date for issuing final specifications). One design exists when the project starts, but it is subject to many updates before the computer is introduced.

Determining probability distributions for chance events is illustrated in Fig. 5. In predicting assembly one design and redesign times, an engineering manager first lists the steps that must be completed to design the assembly (each column represents a particular activity). Next, he states the probability of unqualified success for the activity. When complete success is not achieved, repeating a number of steps (dashed lines) is required. For example, the chance of unqualified success after the first simulation is nil, requiring repetition of logic design, transistor design, and simulation.

A more complex situation occurs after functional testing where four probabilities exist: unqualified success (0.35), mask error (0.1), composite error (0.4),

[5] Contingency planning accounts for probable errors, rework, and the time needed to correct such deficiences. In this example (at final functional testing), there are four significant outcome probabilities: unqualified success, 0.35; mask error, 0.1; composite error, 0.4; and logic design error, 0.15. Any slip initiates a "make good" cycle.



and logic design error (0.15). The graph provides a simple, quick way of identifying design steps, outcome probabilities, and the effect of outcomes on project duration.

Good engineering management considers all possible outcomes, not only the most desirable, and identifies remedial actions required. Such an analysis frequently leads to creative innovation and always provides more realistic estimates.

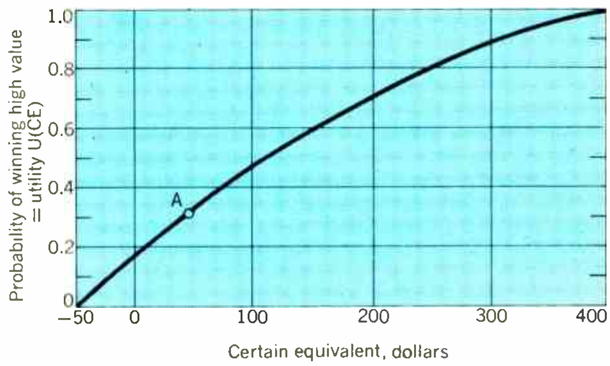
Information: what kind? what cost?

Often, before product design is started, potential applications are explored to determine whether the engineering effort (if successful) is worthwhile. This activity is supported by conventional wisdom, since future market conditions will drastically affect financial achievement. The trap here is confusing *outcomes* (end item sales) with *decisions* (use bipolar or MOS). Assume, for example, that the distribution of present values for a MOS computer is always higher than for an equivalent bipolar design—regardless of favorable or unfavorable market reception. Since MOS has this advantage, further marketing information will not affect our choice of MOS over bipolar. *Information that does not influence decisions is worth nothing to decision analysis.*

If, by contrast, older bipolar products definitely meet military specs, while similar success for new MOS assemblies is undetermined (although estimates exist), further data could be most significant. A MOS testing program will cost $1 million, but is this money well spent? The tests should have some merit, since with mil spec approval, MOS has the more attractive present value; without approval, bipolar dominates (Table II).

Suppose mil spec approval and nonapproval are equally likely for MOS, and no other information about the testing is available. Financial impact of the outcomes can still be estimated. Bipolar has an expected value of 1 × 160 = $160 million and MOS has

[6] **An indifference function for money is derived from a reference lottery. Here, the prizes are always minus $50 million and $400 million. Point A indicates that the particular decision-maker being studied is indifferent to a 30-percent chance of winning $400 million coupled with a 70-percent chance of losing $50 million, or having $50 million for certain (lottery selling price).**



II. Calculating the value of perfect information

Mil Spec Status	Probability	Present Value (PV), Millions of Dollars Bipolar	MOS	Δ
MOS passes	0.5	160	189	29
MOS fails	0.5	160	150	−10
		Expected value (EV), Millions of Dollars		
		160	169.5	

Decision Conditions	Choice	Reasoning
Case #1: No test information on MOS Bipolar will always meet mil spec	Use MOS only	Expected value of MOS ($169.5 million) is greater than expected value for bipolar ($160 million) based on above PV data and probability assignments.
Case #2: MOS has a 0.5 chance of passing mil spec MOS has a 0.5 chance of failing mil spec Bipolar will always meet mil spec	Use accepted MOS (EV = 0.5 × 189 = 94.5) Use bipolar (EV = 0.5 × 160 = 80)	Total expected value increases if MOS is used provided it meets spec, while bipolar is used should MOS fail ($94.5 million + $80 million = $174.5 million).

Note The value of perfect (MOS test) information is $174.5 million − $169.5 million = $5 million

Example B: Time is money

No one knows better than today's prospective homeowner about the time value of money. Mortgages—when and if available—are being written with interest rates approaching usury.

Business ventures are sensitive to such "money market" conditions too, but figuring projected expenses and returns is done via a percentage discount, whereby all future dollars are discounted against the value of a dollar today (1974). The graph here illustrates this process, and the grid is already set up for a discount rate of 20 percent ($1.20 earned a year from now is worth only $1.00 today after discounting). Such a chart can be used to project cash flow for a prospective project over future time periods on an equitable basis.

Cash flow (heavy solid line) is a record of income generated minus operating expenses and investment outlay plotted on a year-to-year basis. This curve yields two cash values at each point in time. In 1978, for example, cash flow will be minus $140 million (in undiscounted 1978 dollars), but only minus $80 million (in discounted 1974 dollars).

Present value (dashed line) is the algebraic sum of all monies paid in and out on a cumulative, discounted basis. Returns exceed the discount rate (profits begin) when this curve first crosses the line where present value equals zero (≈ 1981).

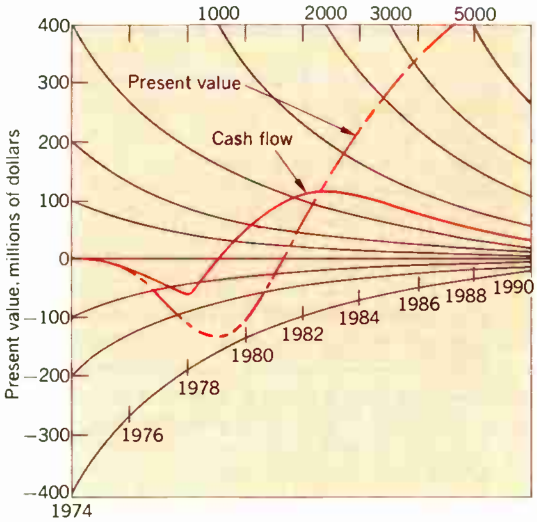
This plot allows the present value of a project to be easily found over any future time interval (i.e., present value of cash flows from 1974–79 equals minus $150 million). Present value (PV) is defined mathematically as:

$$PV = \sum_{N} C_N(1 + R)^{-N}$$

Where C_N equals cash flow in year N, and R equals the discount (interest) rate.

Present value alone is not a foolproof indicator of success since two long-term projects with similar present values can produce very different effects on cumulative cash flow. The project with the longest delay before profits begin is usually the most risky.

Other common indicators include *payback period* (time for present value at a zero discount rate to become positive) and *internal rate of return* (discount rate where present value equals zero).



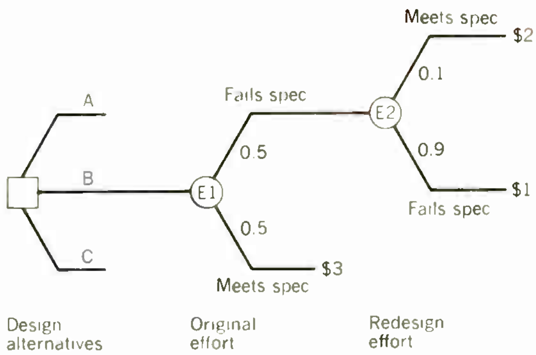
Example C: A decision tree bears fruit

Making engineering management decisions is like buying a ticket on a complex lottery where the payoff is determined by a chain of chance events (time to complete design, test results, etc.). Suppose a manager chooses design approach B among three alternatives (A, B, C). Uncertain if design B will meet spec (chance event *E*1), he estimates the probability of initial success as 0.5. Market information says the payoff for an early product is $3.

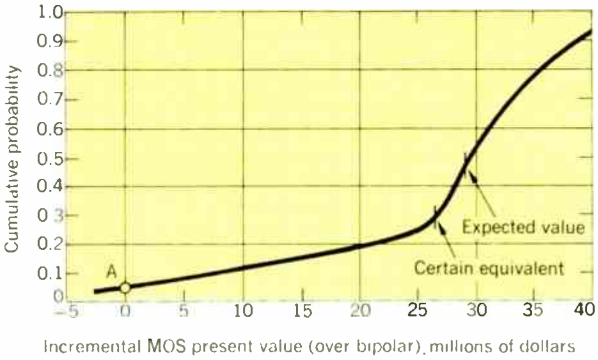
If trouble develops, redesign can still smooth out the kinks. Should rework solve the remaining problems, the delayed product is worth $2, but probability for success here is only 0.1 (chance event *E*2). Product value drops to $1 if design goals remain unfulfilled.

This decision tree defines three different outcomes, the probabilities affecting each, and their potential values. Such estimates provide the groundwork for establishing a probability distribution (see the figure at the top of page 40).

Design approach B has a 0.45 probability of paying $1, a 0.05 probability of paying $2, and a 0.5 probability of paying $3. Although not shown here, design approaches A and C can also be examined in terms of probability, payoff, and expected value from available estimates. Complete analysis of an actual project involves many more decision points (square nodes) and chance events (circles).



Outcome	Probability	Payoff	Expected value
Original meets spec	0.5	$3	0.5 x $3 = $1.50
Redesign meets spec	0.5 x 0.1 = 0.05	$2	0.5 x $2 = $0.10
Redesign fails spec	0.5 x 0.9 = 0.45	$1	0.45 x $1 = $0.45
	1.00		$2.05



A distribution of payoffs for each design choice can be determined by the method illustrated at the bottom of page 39. In the graph here for example, the cumulative probability distribution for MOS + hedge indicates that the chance for a negative present value is less than 0.05 (point A).

an expected value of 0.5 × 189 + 0.5 × 150 = $169.5 million; hence MOS is the preferred choice.

Now suppose a commitment is made to test MOS, and we predict that MOS has a 0.5 probability of failure. Test information would improve the present value of MOS, if MOS fails, from $150 to $160 million ($\Delta = -10$) since our first estimates gave MOS an overall edge even without test data. Present value remains at $189 million where MOS passes muster.

Refiguring the expected value for MOS, we get $174.5 million compared to $169.5 million with no test information. The difference, $5 million, is the *value of perfect information*. Since the testing costs $1 million, this investment should yield $4 million in benefits.

Coefficients of conservatism

Not only does decision analysis handle external factors surrounding the options, but it even probes the subjective nature of the decision-maker himself. Has he a gambler's spirit or a banker's reserve? These attitudes are determined by encoding subjective probabilities. Here the decision-maker is presented with a number of lotteries. One such wager offers a 90-percent chance of winning $100 and a 10-percent chance of winning nothing [expected value = (0.9) (100) + (0.1) (0) = 90]. At what price should such a lottery ticket sell? Called the *certain equivalent*, this price indicates where the decision-maker is indifferent; he has no preference between having the selling price or having a lottery ticket.

A certain equivalent determines a point on an empirical indifference function for money[1] (Fig. 6). The curve gives the certain equivalent of a number of "reference" lotteries for a particular decision-maker. Data on persons with different instincts or expectations will yield a unique indifference function for each.

A mathematical expression frequently used in such analysis is the exponential utility function:

$$U(R) = 1 - e^{-\gamma R}$$

where U = utility, R = reward and γ = risk aversion coefficient (or coefficient of conservatism[2]). The function is independent of personal wealth: if all lottery prizes are increased by C, the certain equivalent is increased by C. The advantages and disadvantages of various utility functions are detailed by Arrow.[3]

Given a decision-maker's risk profile, expected present value can be adjusted via a utility function to determine the certain equivalent of a lottery. The reasoning involves four basic steps: construct a decision tree; replace the rewards by their utilities; determine the expected value of the utilities; determine the certain equivalent of the expected utility from the indifference function. Other views on incorporating utility in decision analysis are expressed by Fama and Miller.[4]

Results of our bipolar-MOS case study indicate that MOS (with proper hedges) provides the best engineering alternative. We progressed from shouting "MOS is best" vs. "I have an uneasy feeling" to identifying issues, risks, and risk-reducing measures plus remedial actions and, finally, flagging those situations where bipolar becomes more profitable.

Decision analysis is a powerful tool, but successful application depends not so much upon the method as on the analyst's experience and the organization in which he works. Skill in handling interpersonal relations is required to gain cooperation and sell results.

REFERENCES

1. Raiffa, H., *Decision Analysis*. Reading, Mass.: Addison-Wesley, 1968.
2. Tribus, M., *Rational Descriptions, Decisions and Designs*. New York: Pergamon, 1969.
3. Arrow, K. J., *Essays in the Theory of Risk Bearing*. Chicago: Markham, 1971.
4. Fama, E. F., and Miller, M. H., *The Theory of Finance*. New York: Holt, Rinehart and Winston, 1972.
5. Fellner, W., *Probability and Profit: A Study of Economic Behavior Along Baysian Lines*. New York: R. D. Irwin, 1965.
6. Howard, R., *Dynamic Probabilistic Systems* (two vols.). New York: McGraw-Hill, 1971.
7. "Special issue on decision analysis," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-4, Sept. 1968.
8. Pratt, J. W., Raiffa, H., and Schlaiffer, R., *Introduction to Statistical Decision Theory*. New York: McGraw-Hill, 1965.
9. Aoki, M., *Optimization of Stochastic Systems*. New York: Academic, 1967.
10. Raiffa, H., and Schlaiffer, R., *Applied Statistical Decision Theory*. Boston, Mass.: Graduate School of Business Administration, Harvard University, 1961.
11. Schlaiffer, R., *Computer Programs for Elementary Decision Analysis*. Boston, Mass.: Graduate School of Business Administration, Harvard University, 1971.
12. DeGroot, M. H., *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.

Curt F. Fey, manager of systems analysis at Xerox Corporation, Rochester, N.Y., is currently developing methods for evaluating technology resource allocation alternatives and new product development strategy. He has been active in the IEEE serving currently on the board of the Systems, Man, and Cybernetics Society. He has been a member of the board of the Computer Society. His career includes positions with IBM and Texas Instruments. He received a B.S. with honors from Haverford College and a Ph.D. from the University of Pennsylvania.

Reprints of this article (No. X74-071) are available at $1.50 for the first copy and $0.50 for each additional copy. Please send remittance and request, stating article number, to IEEE, 345 E. 47 St., New York, N. Y. 10017, Att: SPSU. (Reprints are available up to 12 months from date of publication.)

Intelligent machines are on the way

Are they just around the corner or 'blue sky'? Some robots are already at work, while automatic thinkers may 'control' our future

Speculation about robots and other intelligent machines has long been the unchallenged preserve of the science-fiction writer and properly so, as long as the technology needed to realize these machines was a distant dream. With chunks of that technology now at hand, it is high time for engineers to speculate more knowledgeably about the real machines we are likely to see in the near future. We *do* have choices about that future; engineers have more than ordinary power to choose technology innovations, and those choices can help shape the ways we live.

Some probable machines of the near future are depicted in the drawings on these pages. To understand the background for these conceptions, let us consider some of the intelligent machines already at work, and under development in the laboratory.

Oscar Firschein, Martin A. Fischler
Lockheed Research Laboratory
L. Stephen Coles, Jay M. Tenenbaum
Stanford Research Institute

A number of practical machines are currently being used. Industrial robots, like the Versatran (American Machine and Foundry, New Rochelle, N.Y.) and the Unimate (Unimation, Inc., Danbury, Conn.) perform sightless assembly line operations, and the more-advanced HIVIP Mark I (Hitachi Ltd., Japan) uses TV vision to pick objects from a moving conveyor belt. In automated warehouses, parts are loaded into the correct storage bins by computer-controlled positioning (Mobility Systems, Santa Clara, Calif.). Voice-operated systems allow supermarket checkout clerks to speak prices instead of punching them into cash registers, and airport baggage handlers can call out flight numbers and destinations to route luggage automatically (Threshold Technology, Cinaminson, N.J.).

A medical monitoring belt called IBLMS transmits individual bio-activity data to a remote console for analysis (Lockheed, Sunnyvale, Calif.). A device called the Identimat 2000 (Identimation Marketing, Northvale, N.J.) uses hand measurements to provide unique personal identification. Voice verification equipment is to be used to identify credit card users

Automated intelligence system—In concept, this is a general augmenter of human intelligence, capable of automatically monitoring ongoing streams of input data, coordinating facts, and making logical inferences to obtain insights. Prototypes of such systems are expected by the early 1980s.

Some systems have already been built that augment intelligence by clustering data, by identifying critical factors in multivariant analysis, and by searching for inferential chains that relate two arbitrary facts in a symbolic data base. However, such systems cannot acquire their own knowledge. Inputs must be manually filtered for relevance and supplied in an appropriate internal representation.

Applications of fully automated systems in such fields as social and natural sciences, medicine, law, and publishing are expected to play an important role. Government uses of these systems could include monitoring citizens' actions as well as providing new inputs to guide policy changes.

(Bell Telephone Labs), and a voice response system with a 16 000-word vocabulary is capable of complex sentence responses (Quantel Corp., Hayward, Calif.).

Beyond these practical systems, there is a wide-ranging artificial intelligence research effort that opens many additional possibilities for future engineering developments and products. Here are a few selected examples.

Perception research has produced systems that can identify objects in scenes consisting of polyhedral blocks, arbitrarily stacked on a table top,[1,2] or can find human faces in a cluttered room environment and recognize a limited number of those faces.[3,4] Another research effort is striving to interpret real-world office environment scenes so that a mobile robot will be able to identify specified objects.[5]

Experimental language-understanding systems now include one that comprehends fairly complex declarative and imperative sentences about a simulated-robot environment,[6] and another that can follow portions of children's stories in a reading text.[7,8] Question-answering systems respond to an impressive subset of natural English queries about information stored in a data base.[9,10,11] Speech-recognition systems can handle a limited vocabulary and a small number of speakers.[12] A chess program accepts spoken moves, and uses chess knowledge to resolve any phonetic ambiguities in favor of the more probable move.[13]

Laboratory robots now disassemble and reassemble towers of blocks,[14] thread bolts,[15] move through rooms to fetch large polyhedral objects,[16] and assemble parts for toys and other objects.[17]

Game-playing programs now include a class B chess program[18,19] and a master checker player,[20] a master Kalah player,[21] and a trivial Go player.[22]

Predicting future machines

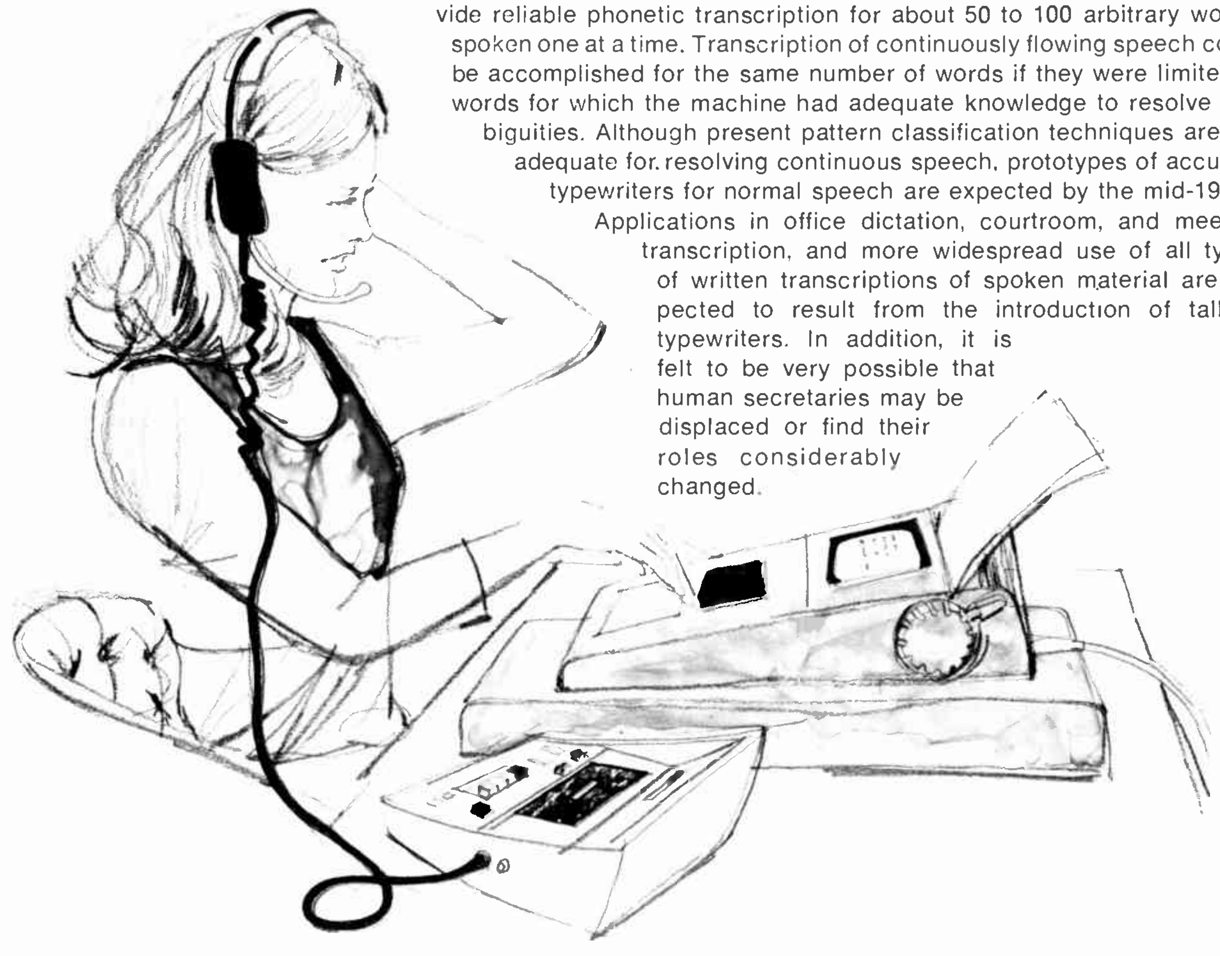
With a survey of relevant artificial intelligence products and research in hand, the authors launched an organized effort to study possible future products that might be based on the developing technology.* Since this was to be a look into the uncertain future, the advice of a number of specialists was sought.

An initial list of 21 such products was placed before a workshop session of the IEEE Systems, Man, and Cybernetics San Francisco Chapter, where participants were asked to respond to the following scenario:

"The Board of Directors of International Artifi-

* Further details of the study described in this article can be found in Tenenbaum, J. M., and Firschein, O., "Forecasting and assessing the impact of artificial intelligence on society," a paper presented at IEEE INTERCON '74, as well as in a forthcoming IEEE Report to the National Science Foundation (Project GJ-37696).

Talking typewriter—This voice-operated typewriter would be capable of converting spoken language into typewritten form at the speed of normal speech, and with an error rate as good as that of a human typist. Today, a voice typewriter could provide reliable phonetic transcription for about 50 to 100 arbitrary words, spoken one at a time. Transcription of continuously flowing speech could be accomplished for the same number of words if they were limited to words for which the machine had adequate knowledge to resolve ambiguities. Although present pattern classification techniques are not adequate for resolving continuous speech, prototypes of accurate typewriters for normal speech are expected by the mid-1980s. Applications in office dictation, courtroom, and meeting transcription, and more widespread use of all types of written transcriptions of spoken material are expected to result from the introduction of talking typewriters. In addition, it is felt to be very possible that human secretaries may be displaced or find their roles considerably changed.



cial Intelligence Products, Inc., after investing considerable capital to attract and hire some of the best workers in the field of artificial intelligence, is now ready to choose their initial product line. A tentative list of 21 products has been prepared by the technical staff, and the board has called together a consulting team to evaluate them.

"The team is to consider whether the products are feasible, what the problem areas are, how they might be marketed, and what the societal effects might be."

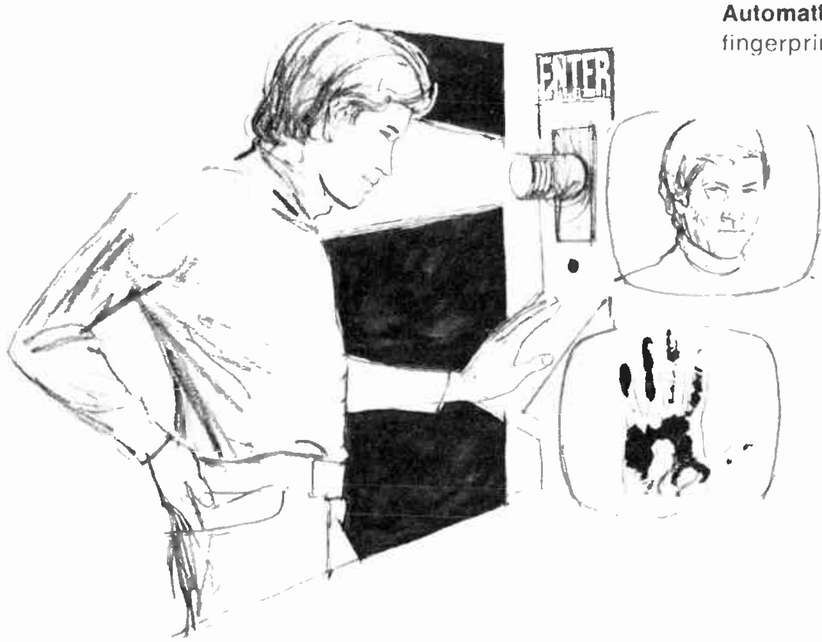
The workshop team consensus was that all the products depicted on these pages were feasible and were likely to be available within the next 15 years. Not pictured in this article are four additional products placed in the same feasible category: an automated inquiry system, a personal biological model to monitor patient's body functions, a machine–animal symbiont with an animal visual system and brain to augment mechanical functions, and an insightful weather analysis system.

Paralleling the workshop session, 41 artificial intelligence experts—including members of the International Joint Artificial Intelligence Council—took part in a Delphi study (see p. 47) designed to shed light

Voice-response order taker—Capable of handling spoken questions about orders for merchandise, this system would make up orders on request, look up catalog information, and inform the user with audible messages.

Such systems are already in existence and an order-taker of their type could be built today to handle about 500 items. If the catalog were more extensive, items would have to be specified by giving code numbers. Limited voice response by such a system, using combinations of prestored words and phrases, is also available now. More capable systems are expected to be in prototype by about 1980.

Order takers promise more efficient and accurate operation of retail and wholesale businesses, as well as remote control of devices by telephone. Use of such systems would probably displace some workers from their jobs, alter shopping habits, and perhaps change the character of retail outlets.



Automatic identification system—Using voice prints, fingerprints, or facial matching, this system would automatically recognize individuals, performing a unique personal identification.

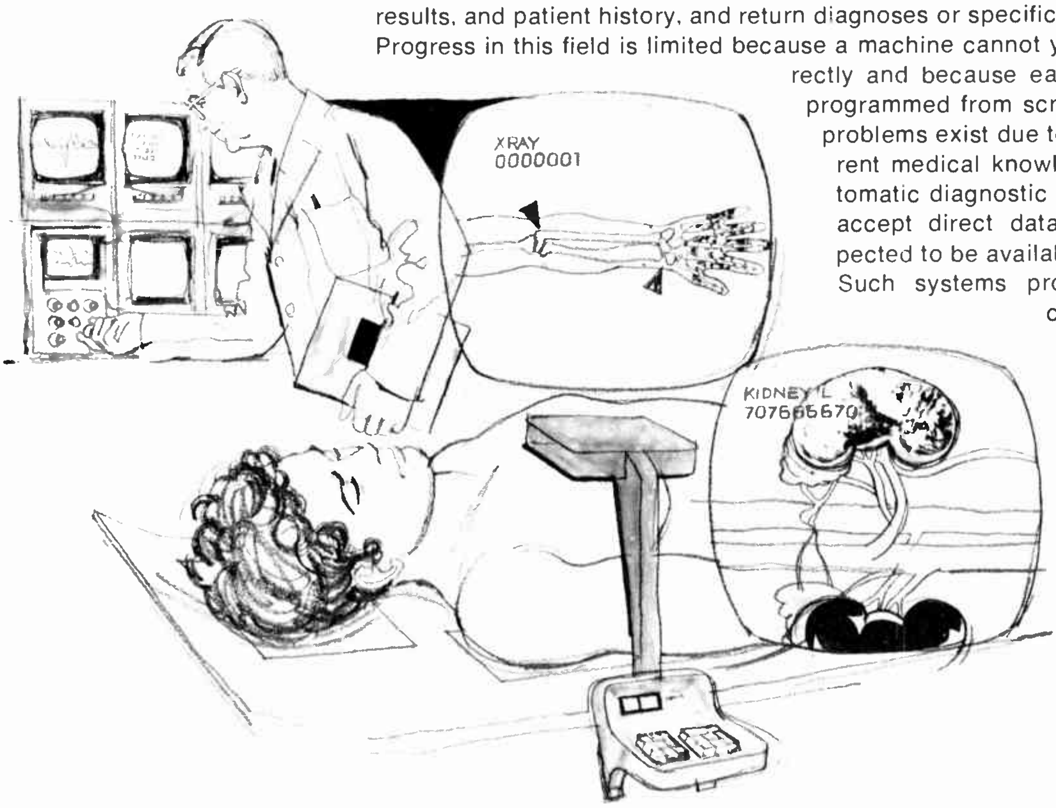
Existing forerunners of this type of system now suffer from pattern classification problems similar to those that limit the voice typewriter. It is difficult to build systems that discriminate more than 50 to 100 categories of faces, fingerprints, or voices. An interactive system in which humans provide descriptions of important features for a mechanized decision-maker might be made practical in the near future. Prototypes of fully automatic systems are expected to be available by the late 1970s.

These systems are expected to be useful for control of secure areas and for security in business transactions. There is apprehension that they may also be used by Government to monitor individuals.

Automatic diagnostician—Based on biological tests and on questions answered by patients, this system would interactively or automatically supply medical diagnoses.

Many diagnostic techniques have already been successfully automated, including chemical laboratory analyses, cell counting, and X-ray and electrocardiogram interpretation. Several interactive systems have been built that accept symptoms, test results, and patient history, and return diagnoses or specific requests for more data. Progress in this field is limited because a machine cannot yet observe a patient directly and because each application must be programmed from scratch. More fundamental problems exist due to the crude state of current medical knowledge. Prototypes of automatic diagnostic systems that are able to accept direct data from patients are expected to be available by about 1980.

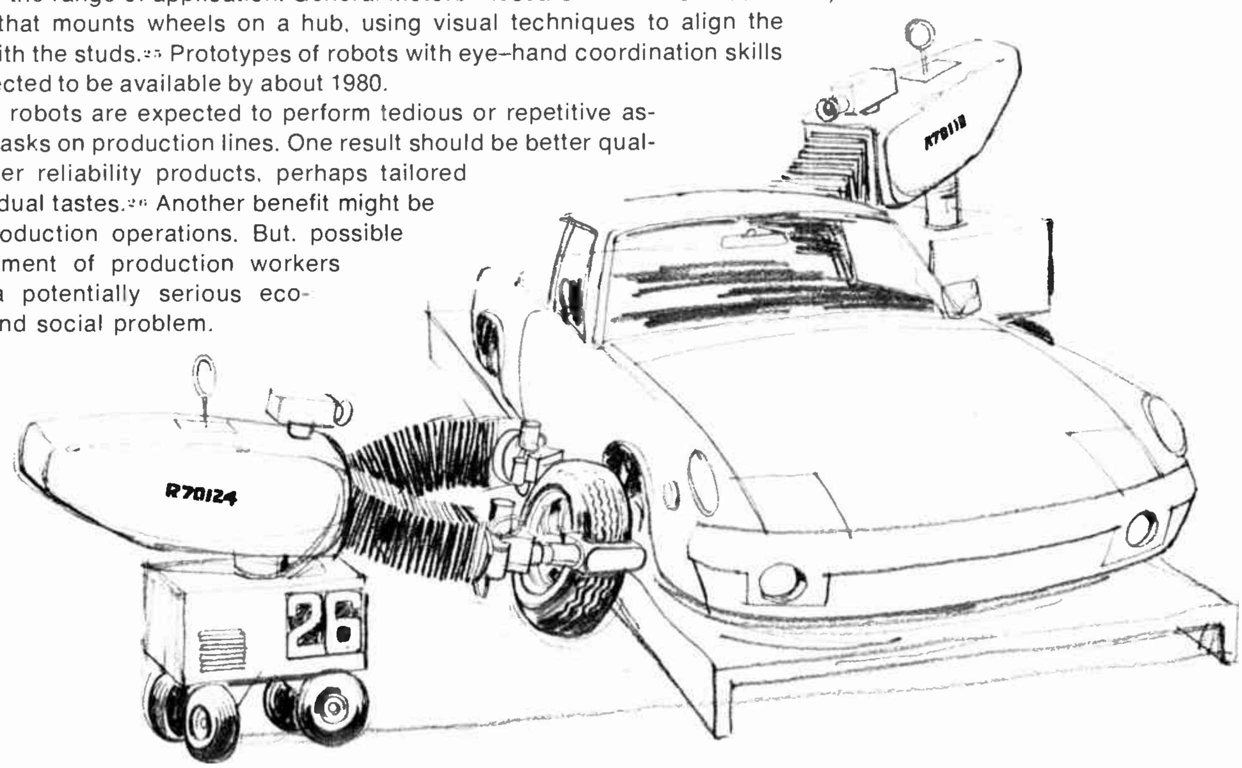
Such systems promise improved health care for patients located far from existing centers of medical care. The role of doctors and their educational needs would likely be affected by the availability of such systems. Extended life spans, with possibly increasing population pressure, is another possible outcome.



Industrial robot—Using both visual and manipulative skills, this type of robot would perform product inspection and assembly in a wide variety of automated factory working environments.

Today, robot manipulators are being used in increasing numbers on automobile assembly lines. They now do repetitive tasks that can be preprogrammed and can operate without feedback—spot welding, for example. The addition of simple visual and tactile sensors would significantly broaden the range of application. General Motors Research Lab has successfully demonstrated a system that mounts wheels on a hub, using visual techniques to align the wheel with the studs.[25] Prototypes of robots with eye-hand coordination skills are expected to be available by about 1980.

These robots are expected to perform tedious or repetitive assembly tasks on production lines. One result should be better quality, higher reliability products, perhaps tailored to individual tastes.[26] Another benefit might be safer production operations. But, possible displacement of production workers raises a potentially serious economic and social problem.

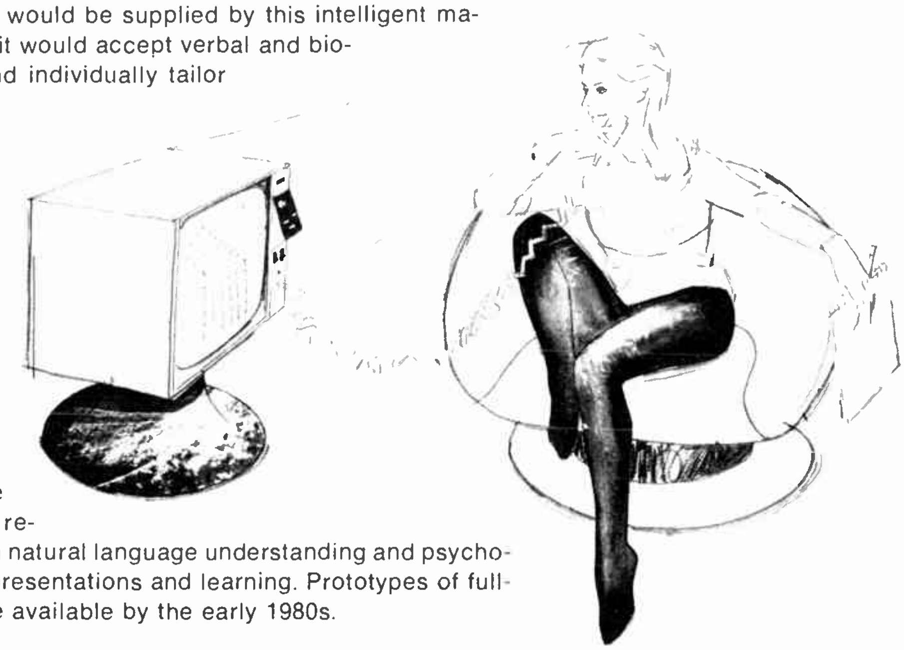


on possible applications, implications, and societal effects of future intelligent machines. In the course of this study, it was necessary to define several of the proposed products more precisely. Some of the results of the study are given here as commentary included in the descriptions of the drawings of intelligent machines. In addition, a number of more general questions were raised by the study.

The future we do not know

During the workshop and the Delphi study, a problem that dogged our efforts was the difficulty of looking into the future. We were especially conscious of the rapidity of technological change, and of tendencies to overestimate near-future possibilities and to underestimate longer term prospects.

By contacting critics of artificial intelligence, we



Robot tutor—Not simple-minded programmed learning but complex interaction near the level provided by a human teacher would be supplied by this intelligent machine. Given an appropriate data base it would accept verbal and bio-sensed responses from the student and individually tailor the course of instruction.

The best existing computer-aided instruction systems now allow mixed-initiative interactions; the student can alter the course of instruction by asking the machine unanticipated questions about the subject area to test his own understanding. Researchers are currently planning systems that will ask probing questions to model a student's comprehension of a subject area and then plan a customized tutorial strategy intended to transform the student's existing conceptual structure into a desired one. The success of such research is very dependent on progress in natural language understanding and psychological investigations of knowledge representations and learning. Prototypes of full-fledged robot tutors are expected to be available by the early 1980s.

Universal game player—Capable of operating at skill levels from novice to master, this machine would play such games as chess, checkers, Kalah, Go, bridge, Scrabble, and Monopoly.

Many existing game-playing computer programs have been written in the course of studying programming and learning problems. Machines can now play champion-level checkers and dominoes, B-level chess, Kalah, Scrabble, and low-level Go. However, each of these games are individually programmed using information deduced from careful human introspection and representations painfully derived to allow effective use of that information. It is expected that by 1980 there will be prototypes of a general game player that can be taught to play a new game as one would teach a human opponent. Interactive home recreation is seen as a probable main use of these machines. It seems possible that such machines might lead to increased public interest in intellectual games, but it remains somewhat doubtful that they could displace or attenuate interest in more passive forms of entertainment like TV.

became aware of opinions that there has been little significant recent progress in the field.[23,24] One critic felt that artificial intelligence researchers were like the man who, having succeeded in climbing a tree, believes he has solved part of the problem of going to the moon.

On the other hand, we were also well aware of the need to make decisions practically every day that must reach beyond a decade, or occasionally even a century! For example, new buildings are normally designed for an occupancy of one hundred years. Because intelligent machines also have the potential for long-term effects on human life and society, it is important that the implications of such systems be understood as well as possible.

A number of general possibilities and potential problems were highlighted by the study.

Some of the machines promise to supply not only intellectual stimulation or instruction, but also domestic and health care, social conversation, entertainment, companionship, and even physical gratification. That could mean a decrease in direct contact and interaction between human beings. We are uncertain about the possible consequences of such a trend, but it should be of intense interest to behavioral scientists and, probably, to mental health practitioners.

The machines also promise to erode or eliminate activities that have, in the past, served to distinguish man from the other animals and from machines. With the coming of intelligent machines, human beings must face the question of their own uniqueness more directly than before. This may well have a profound psychological impact.

Other possible implications of increased machine takeover of human tasks include displacement of human workers, concentration of decision-making in fewer hands, and complete dependence on machines for such functions as weather prediction and economic forecasting.

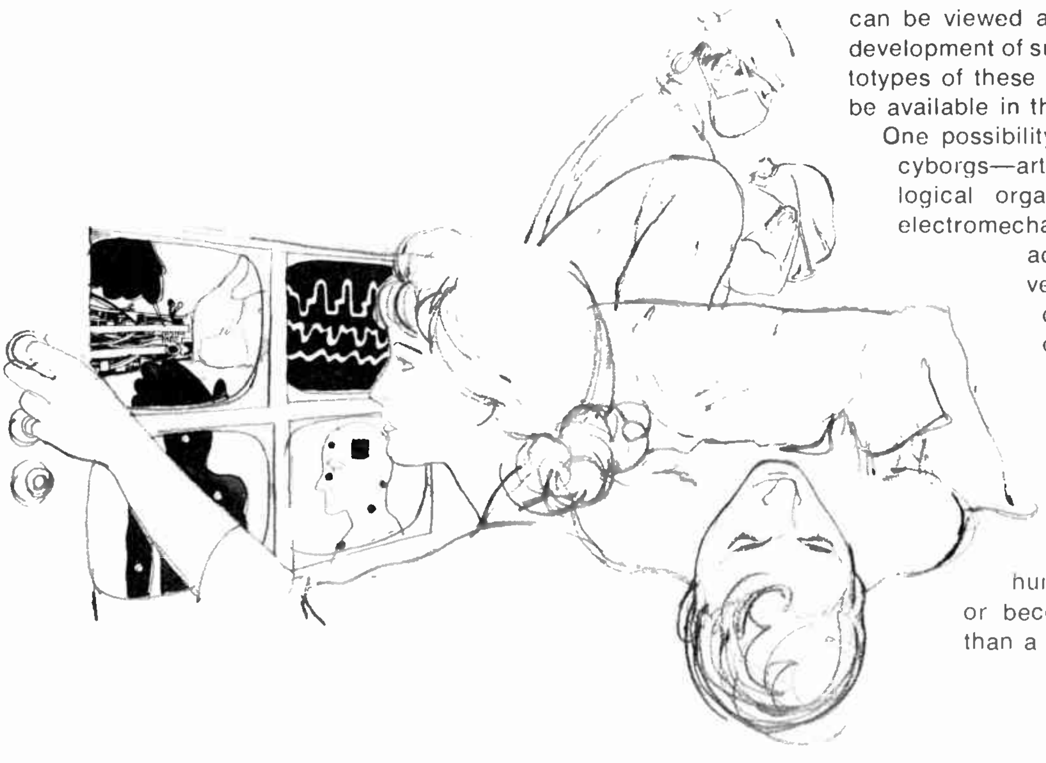
A number of questions come up in connection with safety. Is a designer to be allowed to produce autonomous devices that are free to navigate over city streets and in homes and offices without some review procedure? For devices that require an algorithmic formulation of judgmental factors, who shall decide whether the formulation is antisocial? Furthermore, how can such formulations be validated, and by whom? These are not simply hypothetical questions. Some of the mobile robots already developed have enough speed and power to cause possible injury to humans encountered by the robots in the course of performing their tasks. Any device that includes the characteristics of mobility, speed, strength, and unpredictability can be extremely hazardous, especially in an uncontrolled environment or one where children are present.

Where there are potential dangers to society, these possibilities should be communicated to the appropriate decision-makers. One vehicle for keeping both the public and the decision-makers informed about intelligent machines might be a review board, similar to those in the medical profession that deal with safety and ethical problems. Members of the board could be selected from the artificial intelligence community, but could also include legal, political, and other advisors, as required. The board could recommend legislation, establish standards, and suggest safety guidelines for the use of robots in human-environments.

Computer-controlled artificial organs—These devices would be capable of replacing human hearts, livers, lungs, and other organs, while maintaining overall body-function balance. Present-day experiments with feedback-controlled heart pacemakers, and satellite life-support systems can be viewed as first steps toward the development of such artificial organs. Prototypes of these devices are expected to be available in the mid-1980s.

One possibility is the development of cyborgs—artificially augmented biological organisms that incorporate electromechanical components to adapt the organism to severe environments. A more doubtful possibility is that of significantly extending human lifetimes—barring accident.

Philosophical and legal problems may include deciding just when a mechanically augmented human is considered dead, or becomes a machine rather than a person.



Bridges still to be crossed

We found it exciting to take a peek into the possibilities of the future, but to round out our understanding of what developments can be realistically expected, we felt it was also necessary to take a hard look at some of the fundamental problems in artificial intelligence that still remain unsolved.[27]

One of the most remarkable human problem-solving abilities is knowing how to look at, arrange, or convert a given problem in such a way that the solution is easier to obtain. This is called the ability to *represent* a problem effectively, and we haven't yet been able to formalize this ability in a general manner for the computer. That is, although we can use a particular representation to program the computer to solve a particular problem, we have not been able to provide the computer with the ability to find its own best representation of a problem.

Although human beings are able to separate interesting features from an otherwise confusing torrent of visual and speech signals, difficult problems arise in machine perception of such real-world information. These problems include sensory overload due to the large mass of data to be processed, the related problem of extracting relevant objects from the irrelevant detail, forming internal representations for these objects, and finally, matching these representations against accumulated knowledge in order to identify the words or objects.

The foremost problem appears to be extracting meaningful entities in the context of the problem at hand. Thus, in speech, a continuous waveform must be divided into word segments that are appropriate to the situation, while in vision, objects that relate to the visual task being performed must be isolated from the general background.

Gestalt psychologists have long observed that such human interpretation of parts or features of an object is often guided by an overall interpretation of the whole object. Because general techniques for taking advantage of Gestalt or global information are not practical, machines must usually proceed from parts to the whole, and are easily overwhelmed by the task of examining all the possible interpretations for each part. Therefore, we are still far from any satisfactory solution to the perceptual problems of intelligent machines.

Human communication by means of the spoken or written word is amazingly complex. In order to solve the problems of ambiguity (more than one meaning), pronoun reference, and other difficulties which are handled effortlessly by the human, it is necessary to use an enormous knowledge data base.

For example, consider the amount of extra knowledge needed to identify what "it" refers to in the following statement:

The car ran over the toy in the driveway; *it* was your fault; *it* shouldn't have been there; *it* was nicked and had to be repaved.

Even to *hear* spoken language properly, requires that the listener know the language. Thus, a computer cannot separate (segment) speech into individual words unless it knows the structure, and often the meaning, of the language being used. Language is not

Delphi: a disciplined peek at the future

The Delphi method, used for systematic solicitation and collation of informed judgments on a particular topic, is distinguished from a polling procedure in that *feedback* is used. Judgments of the individual participants are collected, organized, and reformulated as a group response and fed back to each individual. Each participant then reviews the group response and has an opportunity to revise his earlier views or to contribute more if he desires. Responses are anonymous, hence the individual can express ideas freely and without embarrassment.

Delphi can be thought of as a group communication process that includes responses that ordinarily would not be available. Because the questionnaire is the basic information medium used, it is important that a major effort be devoted to properly phrasing the items that appear in it. Particular care must be taken, so the phrasing will not prejudge possible responses. We found that communication is maximized if personal interviews with the experts can be held after a round of the questionnaire has been completed. Comments and analyses that would not appear in written form can be captured by the study team in this way.

We found that the categorizing and organizing required to derive the questionnaire were very useful exercises since they helped us analyze the field of artificial intelligence. Because we were forced to derive a meaningful product list, we had to examine capabilities, postulate products, and iterate the process until a satisfactory list was obtained. After the questionnaire was prepared in draft form, it was tried on experts at Stanford Research Institute, and on Delphi experts at the Institute for the Future, Menlo Park, Calif., and Pacific House Associates, Palo Alto, Calif. Their suggestions led to several modifications in the questionnaire and the proposed product list. We found the preparation of the Delphi study a valuable organizational experience, and one that forced us to confront many of the basic problem areas.

We also found some problems in using this method. About five hours of expert time is required to supply answers for all three rounds of the Delphi process, and most experts do not have much time to spare for such extracurricular studies. Some experts disagreed with part or all of the Delphi approach, but did not feel that the effort required to revise the questionnaire was worthwhile. The result was that they responded to the questionnaire in a half-hearted manner.

Another basic problem in a Delphi study is the individual's rating of his own expertise. Because the true expert may rate himself modestly while the novice inflates his capabilities, we found a tendency for the responses to cluster in a "moderately expert" range. We thus lost the greater importance of the true expert, while inflating the importance of the novice's response. Although there does not seem to be any valid way to overcome this difficulty, we made an attempt to 'calibrate' the responses using a technique developed by Andrew Lipinski of the Institute for the Future.—O.F., M.A.F., L.S.C., J.M.T.

yet formalized to a point where we can program the computer to deal with sophisticated written or spoken communication. Efforts to date have dealt only with limited vocabularies that are used in well-defined specialized environments.

Humans are able to draw general conclusions from a set of specific examples, and to deduce facts from other facts in a very flexible and power manner. While there are now theorem-proving computer programs that can be used for deduction, the more facts these programs have, the longer it takes them to prove something. Like the confused student who cannot separate extraneous from needed facts in an examination problem, these programs are unable to ignore information that has no relation to the problem at hand.

Sometimes people have to deal with contradictory facts. For example, we understand that a whale is a mammal, not a fish. Yet, for certain kinds of reasoning we may find it useful to treat the whale as a fish. Computer programs for theorem-proving don't have that kind of flexibility. In general, formal methods of computer deduction are still in a primitive state, and have so far only been applied successfully in very special ways, such as proving theorems in geometry, or in very simple robot tasks. Artificial intelligence programs that aspire to great generality typically exhibit unacceptably poor performance.

REFERENCES

1. Minsky, M. L., and Papert, S., "Artificial intelligence progress report," Memo AI-252, Artificial Intelligence Lab., M.I.T., Cambridge, Mass., 1972.
2. Falk, G., "Scene analysis based on imperfect edge data," *A. I. Journal*, Summer 1972.
3. Kelly, M., "Visual identification of people by computer," Memo AI-130, Computer Science Dept., Stanford Univ., Stanford, Calif., July 1970.
4. Fischler, M. A., and Elschlager, R. A., "The representation and matching of pictorial structures," *IEEE Trans. on Computers*, Jan. 1973.
5. Tenenbaum, J. M., "On locating objects by their distinguishing features in multisensory images," *Computer Graphics and Image Processing*, Dec. 1973.
6. Winograd, T., "Understanding natural language," *Cognitive Psychology*, vol. 3, Jan. 1972.
7. Badre, N. A., "Computer learning from English text," Memo ERL-M372, Electronics Rsch. Lab., College of Engrg., Univ. of Calif., Berkeley, Calif., Dec. 20, 1972.
8. Charniak, E., "Toward a model of children's story comprehension," Artificial Intelligence Report 266, M.I.T., A. I. Lab., Dec. 1972.
9. Thompson, F. B., *et al.*, "REL: a rapidly extensible language system," *Proc. 24th National ACM Conference*, 1969.
10. Woods, W. A., "Procedural semantics for a question-answering machine," *Proc. AFIPS Conference FJCC*, vol. 33, 1968.
11. Coles, L. S., "The application of theorem proving to information retrieval," *Proc. Fifth Hawaii International Conf. on Systems Sciences*, Honolulu, Hawaii, Jan. 1972.
12. Paxton, W. H., "A best first parser," *Proc. IEEE Symposium on Speech Recognition*, Pittsburgh, Pa., p. 218, Apr. 1974.
13. Reddy, D. R., *et al.*, "The hearway speech understanding system," *Proc. Third International Jt. Conf. on Artificial Intelligence*, Stanford, Calif. 1973.
14. Winston, P. H., "The MIT Robot," in *Machine Intelligence*, vol. 7, Edinburgh University Press, 1973.
15. Paul, R. L., "Modelling, trajectory calculation and servoing of a computer controlled arm," AIM-177, A. I. Lab., Stanford Univ., Stanford, Calif., 1972.
16. Nilsson, N. J., "A mobile automation: an application of artificial intelligence techniques," *Proc. First Intl. Jt. Conf. on Artificial Intelligence*, Washington, D.C., pp. 509-520, 1969.
17. Ambler, A. P., *et al.*, "A versatile computer-controller assembly system," *Proc. Third Intl. Jt. Conf. on Artificial Intelligence*, Stanford, Univ., Stanford, Calif., 1973.
18. Mittman, B., "Can a computer beat Bobby Fischer?" *Datamation*, pp. 84-87, June 1973.
19. Report on Chess, SIGART Newsletter, vol. 44, p. 13, Feb. 1974.
20. Samuel, A. J., "Some studies in machine learning using the game of checkers," *IBM Journal of R and D*, vol. 3, 1959.
21. Slagle, J. R., "Artificial intelligence: the heuristic programming approach," in *Programs that Play Kalah*. New York: McGraw-Hill, 1971, section 2.5, pp. 26-28.
22. Ryder, J. L., "Heuristic analysis of large trees as generated in the game of GO," A. I. Memo 155, Artificial Intelligence Project, Stanford Univ., Stanford, Calif., 1971.
23. Dreyfus, H., *What Computers Can't Do: A Criticism of Artificial Reason*. New York: Harper & Row, 1972.
24. Fein, L., "The artificial intelligencia," *IEEE Spectrum*, vol. 1, Feb. 1964.
25. Olsztyn, J. T., *et al.*, "An application of computer vision to a simulated assembly task," *Proc. First Internl. Jt. Conf. on Pattern Recognition*, Washington, D.C. p. 505, Oct. 1973.
26. Rosen, C. A., "Robots, productivity and quality," *Proc. ACM National Convention*, Boston, Mass., Aug. 1972.
27. Nilsson, N. J., "Artificial intelligence," *Proc. IFIPS-74*, Copenhagen, Denmark, Aug. 1974.

Oscar Firschein (M) has conducted studies of information systems and of long-range forecasting, as well as performing research in pattern recognition and in pictorial description.

Currently a member of the Information Sciences Directorate at the Lockheed Palo Alto Research Laboratory, Palo Alto, Calif., Mr. Firschein also teaches classes in artificial intelligence and its societal impact at San Jose State University, Natural Sciences Department. He received his B.E.E. from the City College of New York in 1948, and his M.S. in applied mathematics from the University of Pittsburgh in 1953. He is a past chairman of the San Francisco Chapter of the IEEE Systems, Man, and Cybernetics Society, and headed the Technology Forecasting and Assessment Project of the Chapter.

Martin A. Fischler (M) has conducted research, and published in the areas of artificial intelligence, picture processing, switching theory, computer organization, and information theory. He is currently employed as a staff scientist at the Palo Alto Research Laboratory of the Lockheed Missiles and Space Co.

Dr. Fischler received his Ph.D. in electrical engineering from Stanford University in 1962. He is a member of Tau Beta Pi, Eta Kappa Nu, ACM, the Pattern Recognition Society, the Mathematical Association of America, and is Past Chairman of the San Francisco Chapter of the IEEE Systems, Man, and Cybernetics Society. He is currently an associate editor of the journal, *Pattern Recognition*.

L. Stephen Coles (SM) is a research mathematician in the Artificial Intelligence Center at the Stanford Research Institute, where he is helping to develop novel applications of computers, including problem-solving programs for a computer-controlled robot. He has served as a lecturer with the Computer Science Department at Stanford University, the Department of Electrical Engineering and Computer Sciences at U.C., Berkeley, and the Systems Program of the Federal University of Rio de Janeiro.

Dr. Coles received his Ph.D. in systems and communication sciences from Carnegie-Mellon University. He has served as a national lecturer for both the ACM and the IEEE and is currently the newsletter editor of the ACM Special Interest Group on Artificial Intelligence.

Jay M. Tenenbaum (M) directs machine perception research at the Artificial Intelligence Center, Stanford Research Institute. He holds B.S. and M.S. degrees in electrical engineering from M.I.T. and received a Ph.D. in 1970 from Stanford University, where he was associated with the Artificial Intelligence Project. Dr. Tenenbaum's interests include human and machine perception, and robotics and practical applications of artificial intelligence research.

A return visit with EE-astronaut Owen Garriott

Seven years of preparation and a blast off like riding the rumble seat of a Model A . . . but the two-month trip was worth it!

The prospect of sitting atop a Saturn missile and being propelled to orbital heights is probably not every engineer's idea of fun, but EE-scientist-astronaut Own Garriott experienced just that—and then circled the earth in Skylab 3 on an orbital mission that lasted nearly two months and brought him both pleasure and satisfaction.

Since *Spectrum's* first interview with Dr. Garriott (February 1966), he has become famous with the public for his experiments with those celebrities in-their-own-right, the "flying spiders," (Arabella and Anita).

In view of his fascinating experiences, *Spectrum* thought it appropriate to revisit Dr. Garriott, and so, at the Johnson Space Center, south of Houston, Texas, and site of "Mission Control," Nilo Lindgren, who conducted the original interview, met with him once again.

At the outset, *Spectrum* asked Dr. Garriott whether he would have preferred the moon mission over the Skylab mission. He replied that he would, in fact, have liked to have had the chance to fly both to the moon *and* in earth orbit, but surprisingly, he said that if he had the choice of only one, he would *not* have passed up the chance of spending two months in orbit for the opportunity to land on the moon.

Dr. Garriott, an EE by training and a member of IEEE, is now the deputy director of Science and Applications at the NASA Space Center, and is busy with many aspects of postflight analysis; he continues to spread the word, through lectures and contacts, on the continuingly important role of scientists and engineers in space.

The Skylab mission was his preference principally because of his background. His previous research experience was not related to studies of the lunar surface, but related much more closely to the sorts of things that can be done in earth orbit. Dr. Garriott had been doing satellite studies of the earth's upper atmosphere and the influence of the sun upon it and, of course, one of the major areas of work on Skylab related to studies of the sun. Other experiments pertaining to earth resources as well as medical experiments, which involved the long-duration missions, were also of great interest to him—both personally and professionally—more so than the relatively short missions in which the experience of a geophysicist or geologist would have been valuable.

While Dr. Garriott had had no previous experience in the medical area, he had always been inquisitive about the body make-up, how it generates energy, and so on—an interest that partly stemmed from years of enthusiastic jogging. And many of the experiments on board—measurements of the cardiovascular performance, monitoring of oxygen consumption, and CO_2 production as the astronauts delivered measured amounts of work—were directed precisely at this question.

Because we were interested in the trials, tribulations, and successes of the mission itself, we launched into this series of questions:

Spectrum: **One gets the impression from reading about Skylab that much of the mission was data gathering. Did this mean that you personally were intervening in the experiments as the mission evolved and as new things happened?**

Garriott: Absolutely! Much of the experimental work was data gathering but the important thing is that a significant fraction, if not the major fraction, did require that the Skylab operator interpret what he was looking at in real time. It was then necessary on occasion for him to intervene in the preplanned activities and to repoint the solar instruments at new targets and to alter their operational modes to take advantage of unforeseen events. As a matter of fact, our flight which lasted from July 28th to September 25th of last year happened to fly during perhaps the most interesting period of solar activity in the past few years. Many of our instruments and some of our scientific objectives had been originally devised to look at solar maximum conditions, but with the slippage in the program timing, we believed that we would be encountering solar minimum. It turned out very fortunately, however, that last August and September the sun increased its activity very markedly over our expectations. The numbers of sun spots and active regions and the X-ray levels were actually running as high as they are often observed to do during sun spot maximum conditions.

Just after we were launched, on about the third day of ATM [Apollo Telescope Mount] operation, we were very pleased to observe an event for the first time that had never before been seen by eye. It was a "solar transient"—a huge bubble of gas that was produced by an eruptive prominence on the solar disk and that was moving out through the solar corona at a speed of about 500 kilometers per second. Transients had been observed on a prior OSO [Orbiting Solar Observatory], but this sort of bubble transient (see p. 53) is really a first. I can remember very clearly after the first Skylab mission came back, just about a week before our

Nilo Lindgren Contributing Editor

NATIONAL TELECOMMUNICATIONS CONFERENCE 1974

NTC 74

December 2, 3, 4
Sheraton-Harbor Island Hotel, San Diego, Calif.

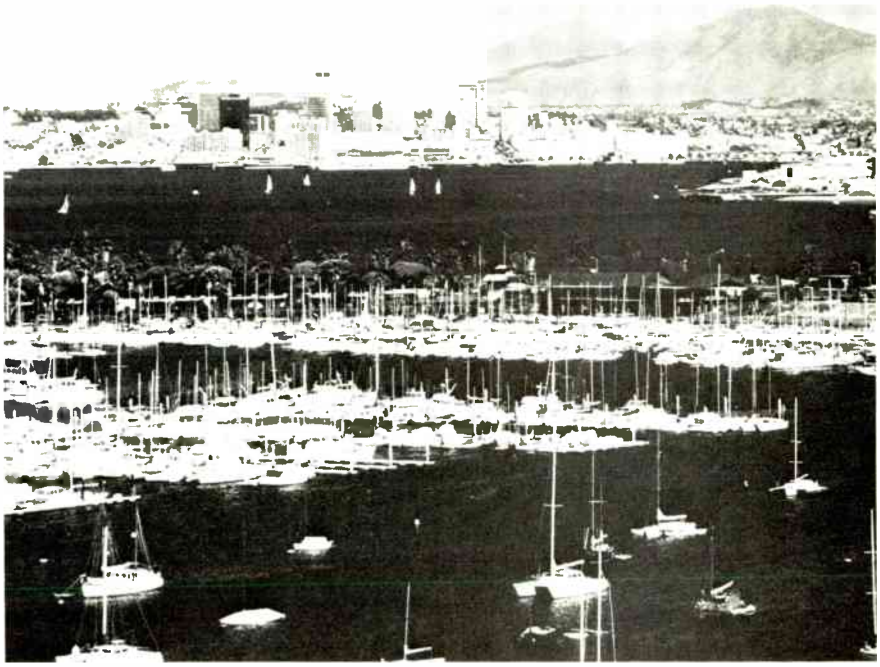
TECHNICAL SESSIONS
covering communications electronics, communication switching, satellite systems, space communications, radio communications, wire communications, data communication systems, communication theory, adaptive communications, channel effects, tactical communications, and more.

TUTORIAL LECTURE
Computer Networks, by Prof. Mischa Schwartz, Columbia University.

WORKSHOP
Data Links for Vehicular Communications, by Prof. Stephen Riter, Texas A. and M. University.

EXHIBITS
with over 50 booths showing the latest in telecommunications equipment.

ACTIVITIES
San Diego provides a unique setting for winter enjoyment of the sun and sea. A full schedule of social activities is planned for the entire family. The reduced conference rates are in effect at the Sheraton for the duration of your visit.



New Trends in Communications

for exhibit, registration, and activity information,

WRITE: NTC '74
10453 Roselle Street
San Diego, CA. 92121

CALL: 714-453-7007

Chairman
Irwin M. Jacobs
Linkabit

Technical Program
Phil Migdal
Micronetics

Exhibits
Bud Hale
General Dynamics Electronics

Registration
Ray Boettger
Pacific Telephone

Sponsors

Institute of Electrical and Electronics Engineers, Incorporated:

- Communications Society
- Aerospace and Electronic Systems Society
- Geoscience Electronics Group
- San Diego Section

Circle No. 26 on Reader Service Card

Astronaut Garriott flies the experimental maneuvering equipment in the forward dome of the Skylab orbital workshop, thus demonstrating man's maneuverability in zero gravity.

launch, we were shown a photograph taken while the previous crew was asleep; it indicated a bubble structure which we interpreted to be a transient since it was not a normal shape for a quiet corona. So when this bubble transient showed up on our third day, we changed our program and went into a new operational mode to observe it. Now, there was no way of predicting this event from the ground at all; instead, we were the first to observe it, as it was happening, from on board the spacecraft.

Before Skylab was launched, we were hopeful that there might be two or three transient events of this nature to be observed. As it turned out, on our flight alone, we brought back over 20 examples of these events. Such transients could conceivably be observed from the ground, but only if they occur in an interval, say, of two or three, maybe ten, minutes out of every year when there is a total solar eclipse. So, you see, it is essentially impossible to catch this kind of event from the ground, since the available time is so short, and since the eclipses during which you can observe the full corona usually occur in some remote corner of the world. Instead, during the Skylab program, we essentially saw the full corona nearly continuously for almost eight months straight. So we were able to see many of these new sorts of events for which there had previously been no opportunity at all.

Spectrum: **You could see them by eye as well as observe them with your instrumentation?**

Arabella, one of the two Skylab 3 common cross spiders, spins its web in the zero-gravity of the orbiting space station. Arabella and companion, Anita, were housed in an enclosure onto which a motion picture and still camera were attached.

ferred to a manned configuration and went through some evolutionary stages before being developed into the package that was flown.)

The Earth Resources Experimental Program (EREP) differs from the Solar Program in that it had involved a large facility—an array of instruments, really—that was placed in orbit without having a single investigator associated with any one instrument. The program had evolved from proposals from all over the world, many from outside the United States and literally hundreds from within. Each had a different objective: some related to the study of crops; others, hydrology, the oceans, and so on.

In view of serious in-flight problems associated with Skylab II, we asked Dr. Garriott the following:

***Spectrum:* Let's talk about the technical problems in your flight, doing repair work and the like.**

Garriott: Yes, we had some repairs, but nothing like the first mission. Our first problem was with our attitude control thrusters. Because one of our quads—a set of four thrusters—went out after liftoff, we had to rendezvous using the three remaining to us. Then, after six days in orbit, a second quad went out, cutting us to half power. If you remember, there was some concern then about the possibility of a rescue mission. Fortunately, it didn't come to that, but we couldn't do anything about repairing the quads. One major repair we did get into, however, was during the extra vehicular activity or EVA. We emplaced a new thermal cover . . . we called it a sail . . . out over the top of the parasol which had been left there by the first Skylab crew. The sail dropped the internal temperatures down another 5°F or so to a more comfortable level.

We also had to replace a complete set of attitude control gyros, the ones that were used to establish the attitude of the spacecraft. These had apparently been damaged by high temperatures during the first part of the mission, and they were erratic in their performance, causing our spacecraft attitude to alter unpredictably on occasion. We had a doubly redundant set of gyros but this was still not adequate to hold a stable attitude. We had brought up a new set of gyros as replacements, which meant EVA work outside the spacecraft. This work didn't require any particular technological background, except for some experience with what is involved and a bit of extra work in the pressure suit.

***Spectrum:* How did you feel outside the spacecraft?**

Garriott: It was a pleasant experience. We looked forward to each of the three EVAs. I was most fortunate in having the chance of going outside all three times. It provides you with a view different than you've ever had before. At the end of the array of solar instruments where we change film, you can lean back in your shoes and it really feels as though you are uncoupled from the spacecraft and floating along through space. When you're at a window looking out, it's more like being in an airplane flying at a very high altitude. But when you are outside, with the full peripheral vision of your helmet, you really can see from horizon to horizon. It's just remarkable. Of course, you are busy most of the time, but on one oc-

The crippled Skylab space station as it looked to astronauts circling it in the Skylab command module, preparatory to docking. The solar array is completely missing on one side, and the parasol sunshield shades the workshop where the micrometeoroid shield is missing.

casion when I was out at the end of the solar telescope, I had to wait while Jack Lousma finished some work, and I managed to lean back and enjoy myself for a minute. We were just coming across the South Pacific. Out to my left, all the way to the horizon there were cloud layers covering half the Pacific. Off to my right was Chile, and I could see all the way across Argentina to the Atlantic Ocean. And straight ahead was the great spine of the South American continent—the Andes. It was just the most phenomenal and beautiful panorama . . . just hard to describe. I'd like to have a painting of it someday.

***Spectrum:* How did it compare with flying in a jet?**

Garriott: There's a vast difference in the way things look from 270 miles! I like to compare it to the difference you perceive between flying in a light plane at 1000 feet and flying in a jet at 30- or 40 000 feet. Every time you increase that dimension by a factor of ten or more, your whole perspective becomes entirely different. You see less detail at high altitudes, but you do see large structures and relationships for which you have no appreciation at the lower altitudes. For instance, at 270 miles you can see all of California at a single glance, which reveals the San Andreas Fault cutting right down through the mountain ranges. Its orientation and delineation are perfectly clear.

The same goes for meteorological phenomena. When you see the whole of a hurricane in one glance, you get a better appreciation of the cloud structure around the eye, the altitudes of the different cloud structures, and so on.

***Spectrum:* And this represents an important scientific outcome of Skylab . . . in having trained**

human observers discerning these unexpected patterns?

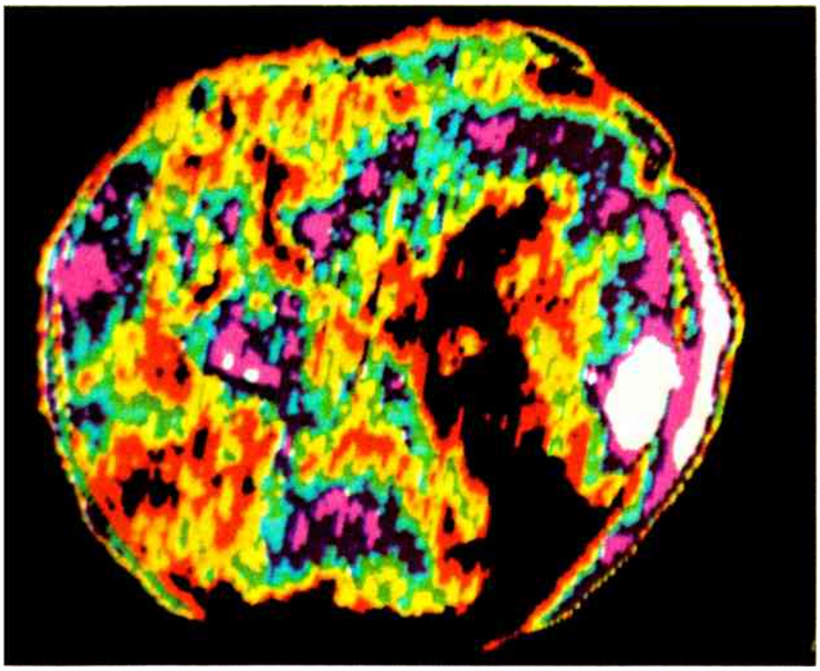
Garriott: Yes, it's very clear now that the more specialized the background a man has, and the greater his store of knowledge, the more he is going to contribute to a flight. In one sense, we were all lay meteorologists by virtue of our piloting experience—knowing something about it, but not knowing it through and through as a science. I'm convinced that when we get a meteorologist up there who really understands the important problems and their significance, he'll do a far better job than we were able to do.

Just after launch, we observed an event for the first time: a solar transient—a huge bubble of gas—that had never been seen before by the naked eye.

***Spectrum:* What you're saying, then, represents some of the wisdom that is coming out of the postflight debriefings and analyses that you're now involved in?**

Garriott: The amount of data to wade through in this postflight phase is so great, it will take years. As I mentioned, it will probably take 50 astronomers five years to get through most of the solar data. In my own case, I'm working with part of it related to atmospheric extinction measurements. I've participated with the principal investigators in devising the operating plans for this segment of the research program, and we plan to do some of the analyses jointly.

False color isophoto reveals day-to-day changes in location of coronal hole, due to solar rotation. The isophoto was processed from a television transmission from Skylab.



***Spectrum:* Are there any other activities you're involved in now that are a consequence of Skylab and your special background?**

Garriott: Yes, as a matter of fact, one of the things I tried to do in flight was to provide useful demonstrations of science that could be converted to 15- or 20-minute movies for distribution to high schools all around the country and perhaps to college undergraduates. I'm pursuing that right now. The first two are almost complete and we have outlines for half a dozen more. The first one, "Magnetic Effects in Space," shows the oscillation of dipoles, and magnetic torque causing the precession of a gyroscope, which was an idea for an experiment I thought of during flight. The second one demonstrates the conservation laws for energy and angular momentum. They don't demonstrate any new physical principles, but they do show how the zero-G environment of space provides a laboratory for illustrating these known principles in a cleaner, more interesting fashion. My interest in these science demonstrations stems, of course, from my teaching background.

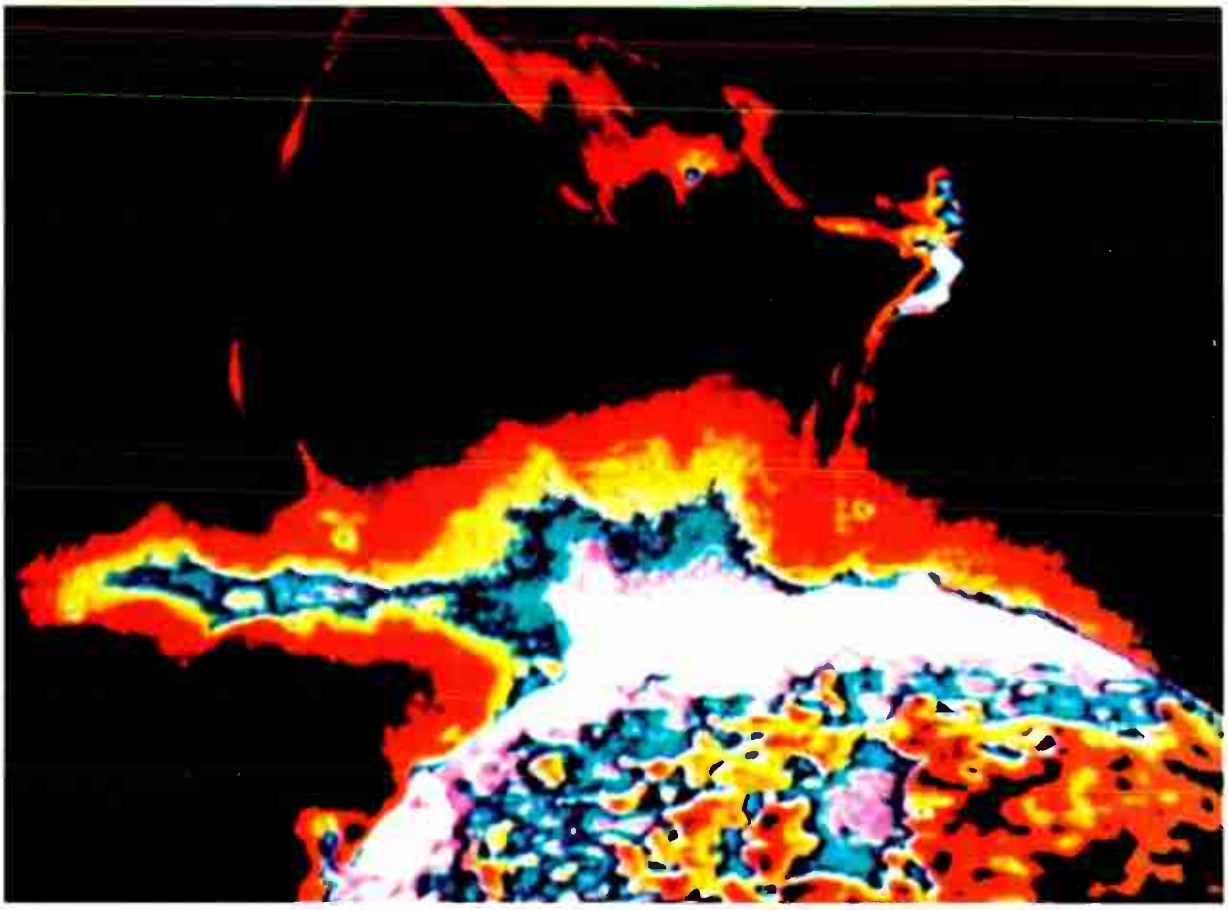
My interest in education has been deep and personal for many years, and Skylab gave me an unparalleled opportunity of carrying teaching to a new stage. Everywhere I've visited and talked around the country, I find the young people are still extremely fascinated by the things we are doing in space science. These films, I believe, will scratch their curiosity a bit more, and perhaps lead them to think up other, more sophisticated experiments that some of them may wind up doing someday as investigators in space.

***Spectrum:* Speaking of that, how many scientists who joined the space program as scientists have stayed with it, and how many have had a chance to fly?**

Garriott: In our initial group, there were six scientists, though one man had to drop out almost immediately and a second resigned to return to university teaching and research a few years ago. All four of us who remained have now flown. The next group selected included 11 men which, for one reason or another, has now shrunk to six. Of these six scientists, none has flown yet, though it's very likely they will in the shuttle program sometime around 1979. The crew for next year's flight—the Russian/U.S. mission—has already been selected but because science is not one of its major objectives, no scientist-astronauts will be on that mission.

***Spectrum:* What about your space training? Did your Skylab experiences match your original expectations?**

Garriott: Reasonably well, I would say, except in terms of time. Early in the program, we had planned to fly both the lunar landing missions and what was to become known as our earth orbital Skylab missions in parallel. However, to reduce annual costs, the lunar landings were stretched out and completed in sequence before we undertook the earth orbital flights, which delayed that program until the 1973



A color density rendition of a solar eruption. The photograph was made using an ATM component covering the spectrum from 150 to 650 angstroms (extreme ultraviolet).

The southern aurora—luminous streamers of light in the Southern Hemisphere—photographed from Skylab as it moved into the sunlight. The permanent aurora over the South Pole is in the background. Garriott used a 35-mm Nikon (4 seconds at f/1.2, high-speed Ektachrome).

time period. Aside from the fact that we all had to wait longer, I must say that our opportunity to fly—in my case for a full two months—was more pleasant and a greater opportunity than I would have expected six or seven years ago.

***Spectrum:* Were the space studies a real continuation of your earlier work—that is, did you remain a researcher in the way you were at Stanford?**

Garriott: No, two things were basically very different. First, time for research was diminished, because most of the work was oriented toward operational activities. The design of Skylab equipment, reviews of design, and preparations for operation, and so on, really took up a good many years of effort. Second, there was not a direct continuation of my earlier relatively narrow interests, rather a very great broadening of those interests. When I was teaching Ph.D. students and pursuing my own research, I didn't have time to look over a broad spectrum of research activities. But in Skylab, with only three men in the spacecraft, and with only one of these having a research background to begin with, you are simply obliged to cover many disciplines. In our case, this included solar astronomy, stellar astronomy, aeronomy, medical experiments . . . not just physiology, but a range of medical activity . . . and it was necessary to be acquainted with this wide spectrum.

All the way to the horizon there were cloud layers, covering half the Pacific... Chile to my right... and I could see across Argentina to the Atlantic.

***Spectrum:* How did that work? Did you have a close association with scientists from all these fields?**

Garriott: Yes, because it's a more efficient learning process to work with people who have these special skills. For instance, in preparation for our solar studies, Frank Orrall of the University of Hawaii presented a set of background lectures to all the Skylab personnel, which provided us with a thorough basic understanding of solar phenomena.

Those of us who were to be scientist-pilots then began working closely on an individual basis with the scientists who were developing the program of experiments that were being flown on the ATM or Apollo Telescope Mount. In fact, we collaborated with them in devising what we called a JOP or Joint Observing Program for ground and space experiments, so that when we were gathering information on particular phenomena—like transient events, chromospheric network structure, or prominences and filaments—both earth and Skylab observations could be coordinated. In the process of devising these operating programs, we naturally became more deeply involved in the leading edge of solar science, in the knowns and unknowns of interesting particular problems in which our observations would be most useful.

***Spectrum:* Have the training programs for the scientist-astronauts changed at all as a result of the Skylab experience? Maybe we should ask this differently: In retrospect, would you personally have wanted your own training to differ in any respect?**

Garriott: Somewhat. The flight training was very useful; I would not have wanted to miss that for anything. But I think we could very well have used scientific training in additional disciplines. To be precise, I'm most annoyed that I did not have adequate training in many of the disciplines for which visual observations would have been important—meteorology, oceanography, even geography. I would certainly press for a more thorough grounding in these aspects if I were to start all over again.

***Spectrum:* What does the future look like, then, for space scientists?**

Garriott: My guess is we'll be taking up more specialists. These first few groups—myself in particular—are really generalists. Any specialization we've done, we brought to the program, and we became generalists thereafter. We'll still need generalists, but specialists will play a larger fraction of the role. For example, in our shuttle we will probably have two crew members—the commander and a pilot—whose training will be largely pilot-oriented. The rest of the crew . . . anywhere from an additional two to five persons . . . probably will be oriented toward payload operations. Fifty percent or more of the crew in the shuttle era will be associated with research activities. So, very definitely, scientists and engineers will play an increasing role in the years ahead.

***Spectrum:* How would you advise potential candidates to prepare themselves?**

Garriott: I hear that question a lot; I don't think there is any special preparation other than the realization that it is still a multidisciplinary job and one that requires a certain innate perspicacity—the ability to look at all sides of a problem, whether it has to do with the generation of concepts for experiments or with the tradeoffs associated with the design and engineering of equipment. You need to be highly motivated, able to work hard, and emotionally be pretty stable in a sustained mission. The actual research activities that need to be done may be obvious—they include the life sciences, solar and stellar astronomy, and even some kinds of manufacturing and technological experiments that rely on hard vacuum and zero G, such as materials processing. As a matter of fact, my crewmate Jack Lousma ran a furnace on Skylab for growing crystals. The results of his work show that certain crystals can be grown up to six times the size single crystals have been grown on earth, and their chemical homogeneity is far greater than for those made here on earth.

***Spectrum:* Can you tell us about your subjective experience in all this. Some people wonder about new awareness or consciousness changes in the as-**

tronauts . . . for instance, the kinds of ESP things that Mitchell is interested in. Have you experienced any such perceptible changes in your own outlook?

Garriott: No, I've not experienced any remarkably different feelings, though I've talked about this a bit with other astronauts. Even in Ed Mitchell's case, I believe most of the views he's expressed . . . that is, his outlook on life . . . were really present before he flew. He has had a long-standing interest in ESP and related activities, but during the time he was preparing for flight he had to concentrate so hard on the mechanics of accomplishing the mission that he didn't have much opportunity to express these other viewpoints. My own view is that we all have many feelings we need to suppress somewhat during the arduous preparation for flight, and once the mission is over, we may reevaluate our situations and may find the opportunity and desire to spend time on ideas and feelings and viewpoints that we've probably been carrying for a good many years but have never found the time to develop.

In space, your thinking processes are just the same as here, your values are just the same. But in space you realize your unique vantage point and the limited amount of time available, and you work hard to take proper advantage of it. Halfway through our mission, on about day 30, I tried asking myself how much time I had accumulated to look out the window and just enjoy myself. And though I may exaggerate a bit, my best evaluation at that point was only three or four hours. I have no complaints; but though we were there for two months, we were always pressed for time to do the many valuable and fascinating things that needed to be done.

***Spectrum:* Now that this particular epoch of your life is over, how do you see your immediate future?**

Garriott: That still is not clear, though I have several options. Right now, as deputy director of Science and Applications, I've got a lot of interesting work to do. I'll just have to give myself more time to decide which alternative appears best. I don't think I'll start off in a radically different direction, and I believe the experience I've gained over the past seven or eight years could be valuable in other programs.

***Spectrum:* In this long epoch you've lived through with the space program, have you ever despaired that things wouldn't work out as well as you've indicated?**

Garriott: No, I can't remember any despair. I suppose there were ups and downs, but my personality is such that I tend to smooth off the tops and bottoms a bit . . . I've really enjoyed most of it except for the extensions of time. I would rather it had been done in four years rather than seven, but looking back on the unique experience I was able to participate in, even thre[illegible]ra years were worth the investment.

***Spectrum:* One last question for those who might be seriously entertaining the idea of a space-scientist career, and especially for those of us who have only watched the Saturn liftoffs. What was the experience like? Was it scary?**

Garriott: Well, scared is not the word. Certainly there is a normal amount of apprehension, regardless of how many flights one may have made . . . call it a healthy regard for the situation. Heart rates are somewhat elevated; they always are. The commander's heart rate is always highest, because he's got the most responsibility at that point . . . which is about what you would expect.

But I really didn't know what to expect from other people's descriptions. For each of us, it's apparently a little different. In my case, I compare the initial blast with riding in a rumble seat of a Model A down an old country road. It really shakes, rattles, and rolls there for a minute or two. At the low altitudes, the acoustic waves and atmospheric turbulence provide vibrations so that you know your machine is very noticeably in motion. Then, after you get above most of the atmosphere the vibration diminishes, the ride becomes more and more smooth, the acceleration builds up, and you get pressed more and more firmly against the back of your couch. At staging, when the rocket cuts off its engine suddenly, you bounce right out of your chair up against the straps. Almost immediately you're thrown back against your seat again as the second stage ignites. Then you have a nice smooth ride until "Zap!"—a very abrupt cutoff—and there you are, floating for two months.

Nilo Lindgren, a contributing editor for *Spectrum*, is developing a program for technical communication at the Xerox Corporation's Palo Alto Research Center. For five years as a staff writer for *IEEE Spectrum*, he wrote broad reviews on machine recognition of human language, human factors, artificial organs, electric cars, speech research, cybernetics, biomedical engineering, and art and technology. Prior to that, he was an editor for *Electronics*, and a McGraw-Hill correspondent in Finland. More recently, he was a senior staff member of Technology Communication, Inc., a member of its creative board, and a senior editor of *Innovation*. During his career, he has been associated with Philco Corp., Hughes Aircraft, and Grumman. A member of the American Society for Cybernetics, Mr. Lingren holds a B.S.E.E. from the Massachusetts Institute of Technology. He has also coauthored a book on art and technology and is the editor-in-chief of a series of forthcoming volumes on cybernetics.

Owen Garriott (M) received his B.S.E.E. from the University of Oklahoma and a master of science degree and a doctorate in electrical engineering from Stanford University. Dr. Garriott was an electronics officer while on active duty with the United States Navy, serving aboard the U.S.S. Cowell and U.S.S. Allen M. Sumner. From 1961 until 1965 he taught electronics, electromagnetic theory, and ionospheric physics in the Department of Electrical Engineering at Stanford. He has logged over 2000 hours flying time, including over 1500 in jet aircraft. In addition to NASA ratings, he maintains FAA commercial pilot and flight instructor certification. Dr. Garriott was selected as an astronaut by NASA in June 1965. He is a member of the American Geophysical Union, the International Scientific Radio Union (URSI), and the American Astronautical Society, as well as Sigma Xi and Tau Beta Pi. He was awarded the NASA Distinguished Service Medal in 1973.

WELCOME TO WESCON!

September 10-13
Los Angeles Convention Center

The nation's largest and most complete electronics show and convention this year opens in Los Angeles on September 10. And of course, you're invited.

This 23rd annual Wescon is a super week in an exceptional electronics year. There are 54,000 net square feet of product demonstrations by 300 electronics producers — that's a full house, literally.

The 540 exhibits are grouped in four product/system categories on one beautiful carpeted and air-conditioned floor of the Los Angeles Convention Center, and will be open for hands-on demonstrations for four days and one evening.

The four-day **professional program** is an escalator ride away — also in very large, carpeted and air-conditioned auditoriums. This 29-session program is both current and choice. Four knowledgeable sessions on microprocessors; a full day on CCD applications; an expert update on medical electronics; a first-rate panel on uninterruptable power problems; sessions on component and board testing, plastic encapsulated circuits, management in periods of crisis, and taking high-technology products from lab to market.

And there's more — a keynote address by **Donn Williams,** president of Rockwell Electronics Operations, a Distributor-Manufacturer-Representative conference, an all-industry reception, and three California-colorful social events for visiting wives.

Your share of **four idea-filled days** awaits you in Los Angeles. Registration for all four days of the program and show is just $5, payable at the door. And if you'll assist us in a simple questionnaire-survey at the show, you'll be eligible for a very special prize drawing. On each of Wescon's four days, a drawing winner will receive an HP-45 hand-held calculator!

Better **plan right now** to join us in Los Angeles. Write directly to the Wescon Housing Bureau, P.O. Box 3669, Los Angeles, CA 90054, for accommodations. We're holding large blocks of rooms at the Regency Hyatt House, Los Angeles Hilton, Biltmore, Sheraton-West and Wilshire-Hyatt.

WESCON74

Circle No. 27 on Reader Service Card

Western Electronic Show and Convention, 3600 Wilshire Blvd., Los Angeles, Ca 90010 (213) 381-2871

Sponsors: IEEE Los Angeles Council and San Francisco Section; Electronic Representatives Association (Southern and Northern California chapters).

Automatic voice response: interfacing man with machine

This developing technique can relieve the human from repetitive instructional chores by piecing together his recorded words

One of the current goals of speech research is to simplify the problem of communication between man and machine by making it possible to exploit the human voice. Toward this end, much work has been devoted, in recent years, to designing voice-response systems. These systems, which simulate the human voice, are automatic and computer-controlled.[1]

For example, a wireman—beset with the tedious problem of assembling and installing telephone equipment—can work much faster and more accurately from spoken instructions than from written directions.[2,3] In this case, voice response techniques can be used to produce the voice instructions necessary for complex wiring programs (see the box on pp. 64–65).

Other applications, made possible through the wide availability of the wire communications network, can utilize the Touch-Tone® telephone much like a remote computer terminal, providing keyboard input and voice output and making computer data bases accessible to virtually anyone with a telephone. For example, credit-card numbers could be automatically verified. After being inputted via telephone, they would be accepted or rejected by the centralized voice response system over the same telephone. Automatic confirmation of reservations and other services requiring interactive communication with a computer (e.g., telephone directory assistance) would work the same way—inputting the information in question via the pushbutton dial and receiving spoken answers from the voice response system.

Although research is accelerating on *speech recognition*[4] and *speaker verification*[5]—two areas closely related to voice response—systems based on these "pattern recognition" technologies are still not developed enough to be successfully applied on a wide scale. However, combined with voice response systems, future voice recognition and verification systems could provide limitless on-line applications needing user authentication. For example, while present voice response systems have the capability of authenticating credit-card numbers, future voice verification/response systems would actually be able to attest to the authenticity of the voice itself. In this way, a credit-card subscriber could immediately be validated while making purchases. Similarly, bank depositors will be able to make transactions directly over the phone—a banking concept that is virtually revolutionary (see box on page 66).

L. H. Rosenthal, L. R. Rabiner, R. W. Schafer, P. Cummiskey, J. L. Flanagan
Bell Laboratories

Voice response systems

Using digital techniques, we can now design a computer voice response system with real-time multichannel capabilities. A voice response system operates as illustrated in Fig. 1. The three basic components of the system are: vocabulary propagation, vocabulary storage, and message composition. In preparing the vocabulary, one may use individual words, phrases, or even entire sentences. The vocabulary must be stored so that the entries can be conveniently accessed to form messages. Voice output can be generated by using the required entries in sequence to form the desired message. This type of message composition, if performed faster than real time, is capable of multiline output in which several message requests are processed in parallel.

One such system, designed and built at Bell Laboratories, is based on a recently developed digital speech-encoding technique, and is capable of providing voice output simultaneously on ten channels. This system uses an all-digital approach, and is made with existing off-the-shelf hardware and software.

The design of a voice response system depends on both the unit of vocabulary storage (e.g., words, phrases, etc.) and the method of storing these units. Efficient vocabulary storage generally requires that the vocabulary entries be individual words, since many different messages can be composed from combinations of the same words. On the other hand, highly contextual messages composed from individual words can sound very unnatural unless a fairly sophisticated algorithm is used to interpolate pitch, amplitude, and formant frequencies across word boundaries. Thus, a tradeoff exists between efficiency of vocabulary storage and natural output speech.

Nevertheless, for many applications of voice response it is adequate to generate speech by a simple concatenation or linkage of the waveforms of individual words and phrases. These applications generally require noncontextual output speech (the elements of which can be stored as individual words) and a small number of contextual messages (which can be stored as entire phrases or sentences). This restriction on the output speech greatly reduces the complexity of the voice response system, and does not severely limit the range of application. It also permits the design of a totally digital multiline system based on a medium-size vocabulary (under 200 words or phrases) that can be corrected with a minimum of effort.

The primary feature of the voice response system of Fig. 1 is that all speech processing is done digitally. Input speech is converted to digital form before being edited, catalogued, and stored, and the speech remains in digital form until a desired output message is formed.

On the input side of the system, the most important feature is the *ADPCM* (adaptive differential pulse-code modulation) method of speech encoding, which permits both efficient storage of speech on a disk and automatic editing of encoded speech.

On the output side, the most significant feature is *real-time multitask programming*, which makes it possible for the CPU (central processing unit) to run as many as 21 independent tasks, interrupt routine services, and perform data channel transfers so that these different operations appear to be going on simultaneously while producing ten real-time outputs.

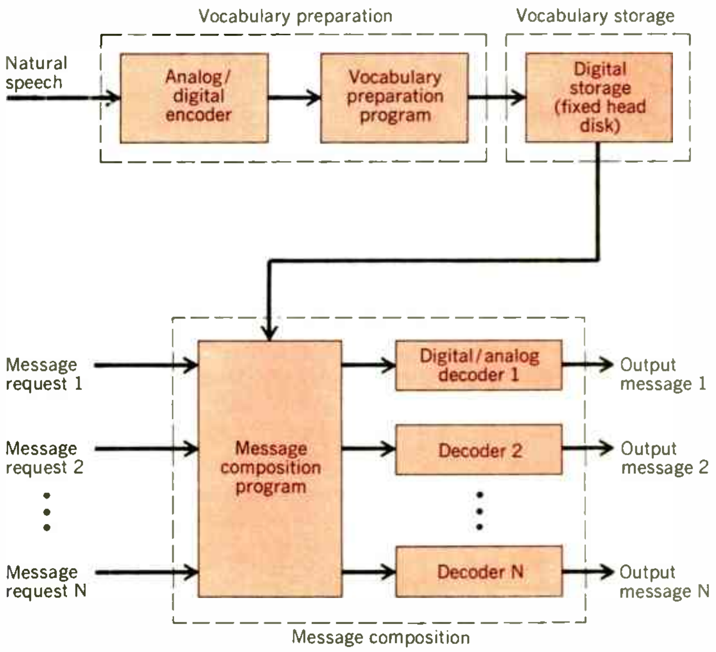
Designing the system . . . making it work!

Design of the voice response system can be logically divided into two parts, the first part dealing with creating and maintaining a *vocabulary*, and the second with the characteristics of the output—or *message composition*—system.

Once a vocabulary has been created and edited, it must not only be efficiently stored, but be able to accept continual updating. Further, entries must be quickly accessible and, when linked to form messages, as natural sounding as possible. Currently existing voice response systems employ a variety of both *analog* and *digital* recording techniques in an attempt to meet these requirements.

One means of storage is to record speech entries in analog form on a tape, film, or magnetic drum. This technique is used mainly in systems with relatively small vocabularies, since the storage is inefficient. Generally, the storage medium is divided into tracks, each of fixed duration, and an individual word or phrase is recorded on each track. Although this is the simplest recording technique, it is also the least flexible. For example, it is often necessary to compress a long word to fit a single track of the storage medium, whereas a short word may not fill an entire track. This, plus the fact that pauses between words must also be of fixed duration, makes the output speech sound more unnatural than is necessary. In addition, retrieval time for a given word is usually fairly long because of the time required to position a reading device at the beginning of the correct track.

[1] An all-digital voice response system.



Finally, the choice of a storage device will determine the ease in changing vocabulary entries. For example, generation of a new vocabulary on photographic film requires a new storage medium. Although analog storage limits system flexibility, such systems are now used and are commercially available.

The alternative to analog storage is some form of digital storage of speech. Digital storage devices for automatic voice response systems include magnetic disks, drums, and solid-state memories. These devices all provide faster access times than analog storage media and allow the possibility of variable length vocabulary entries. Solid-state memories provide the fastest access times, but large versions are still economically impractical. The best compromise between access time and storage capacity is provided by a fixed head, or head-per-track, disk. This device has tracks similar to those of an analog drum, but it rotates at a much higher speed (3600 r/min is typical), thus providing rapid access to any section of a track. In addition, since the disk is a read/write memory, vocabulary entries can be easily changed.

Why ADPCM?

Although it is only one of many alternative digital waveform coding schemes (see box on page 67), ADPCM coding has several advantages for digital voice response applications, one of which is good-quality speech (20-dB signal-to-quantizing-noise ratio) at a rate of 24 000 bits/s (4 bits/sample × 6000 samples/s). ADPCM-coded speech at this bit rate was found to be perceptually comparable in quality to log PCM* at a rate of 42 000 bits/s (7 log-bits/sample × 6000 samples/s).[6,11] Another factor in favor of ADPCM coding is that the entire encoding/decoding process can be performed by inexpensive hardware in real time. Thus, no CPU time is needed for processing the speech data. Finally, by exploiting the unique features of ADPCM coding, it is possible to create and edit a vocabulary automatically from an analog tape recording of the spoken entries with minimal human intervention.[7]

ADPCM coding achieves its advantages over PCM coding by quantizing a *difference signal* derived from the actual speech waveform and by allowing the quantizer to change its step size to match the amplitude of its input.† Figure 2 indicates the operation of an ADPCM coder and decoder. The difference signal $\delta(n)$ is obtained by subtracting from each sample of

* The objective signal-to-noise ratio improvement of ADPCM over log PCM is only about 8–10 dB; the subjective improvement, however, is of the order of 18 dB or about 3 bits. At a rate of 24 kbits/s, ADPCM-coded speech also has about a 6-dB signal-to-noise ratio (SRN) advantage over adaptive delta modulation.

† If the quantizer is fixed, the system is a conventional differential pulse-code-modulation (DPCM) coder.

the input speech signal $x(n)$ an estimate of that sample $y(n)$. The estimate is obtained by multiplying the sum of the previous estimate $y(n-1)$ and the previous quantized difference $\hat{\delta}(n-1)$ by the constant a (which is close to 1.0). By using this technique for deriving the difference signal, it can be shown that quantization errors do not accumulate with successive differences.[8] In fact, for each sample of the input signal $x(n)$, the quantity $x(n) - \hat{x}(n)$, which is the difference between the actual speech sample and the sample reconstructed by the decoder, is equal to $\delta(n) - \hat{\delta}(n)$, the quantization error for the most recent difference. It has already been shown that the use of differential coding in the ADPCM coder yields a 4-6-dB advantage in signal-to-quantizing-noise ratio over PCM coding for speech.[6,11]

Further improvement in the ADPCM coder is obtained by using an *adaptive quantizer*, which decreases the step size upon detection of quantizer underload (i.e., when the quantizer input is small) and increases the step size when quantizer overload is detected (i.e., when the quantizer input is large).

In the input–output relation for the adaptive quantizer, the binary numbers labeling quantizer levels are 4-bit code words $c(n)$.* Quantizer step sizes $\Delta(n)$ are modified on the basis of the previous quantizer output according to the rule $\Delta(n) = M\Delta(n-1)$, where M (the step size multiplier) is a function of the previous quantizer output. The particular values used for M in the present case were selected by computer simulations of the ADPCM coder.[11]

In the hardware implementation of the coder, the quantizer has a finite library of 21 possible step sizes with a range of 100 to 1. Thus, the quantizer step size $\Delta(n)$ obeys the above relation only for a certain range of values, beyond which it saturates either at a maximum or minimum step size. The use of an adaptive quantizer has been shown to give an additional 4-dB advantage in SNR over PCM.[6,11] (It should be noted that only the 4-bit code word representing the quantizer output level must be transmitted since the step-size adaptation algorithm in the decoder is identical to the one in the coder and therefore the step size is completely determined from the sequence of code words.)

In the decoder, the same adaptation logic employed in the coder is used so that, in the absence of transmission errors, $\hat{\delta}'(n)$ (the decoded quantizer output) is identical to $\hat{\delta}(n)$.† A differential feedback loop identical to that used in the coder reconstructs the signal $\hat{x}(n)$, and this signal is then low-pass filtered to produce the analog output.

For the voice response system that was implemented, one ADPCM coder/decoder and nine decoders were constructed in hardware. Each coder was separately interfaced to the computer. The coder interface packs four 4-bit coded speech samples to each 16-bit computer word. Each decoder interface in turn unpacks the coded samples and clocks them to the decoder. By performing this packing/unpacking process in the interface instead of in the computer, the computation time saved can be utilized more efficiently to control the multiple, simultaneous speech lines.

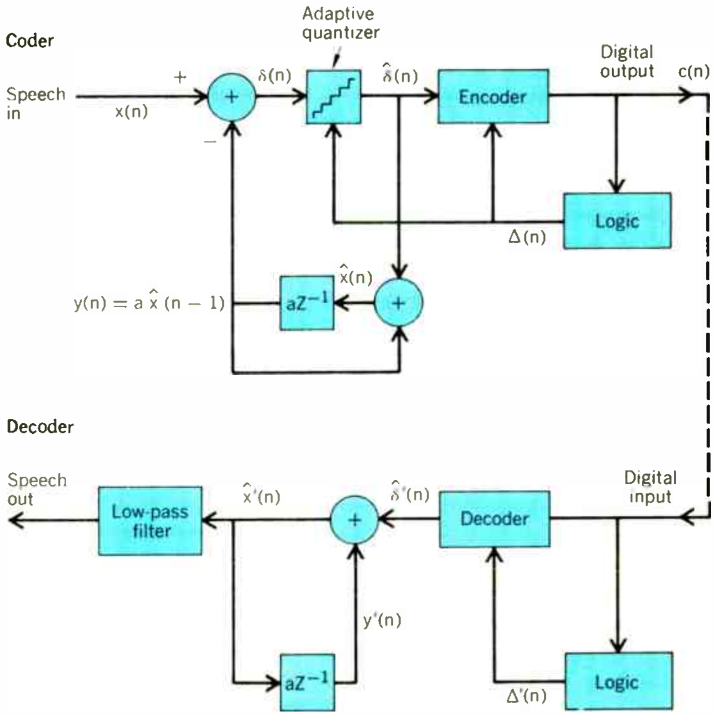
Automatic editing—an ADPCM advantage

After a file of speech has been created using the ADPCM coder and stored in digital memory, the individual words and phrases in the file must be isolated to store speech efficiently and without intervening periods of silence between entries in the word catalog. Conventionally, this editing is done manually by a combination of listening to the speech and studying a visual display of the speech waveform. However, this approach is both time consuming and subject to error, especially for an unvoiced segment of speech at the beginning or end of a word. In such cases, manual editing usually results in the *shortening* of an utterance, both at its beginning and end. Such chopped words are distracting when linked to form a message.

ADPCM coding, however, provides a novel alternative to manual editing. Because of the step-size adaptation used in ADPCM coded speech, an ADPCM coder effectively has an automatic gain control (AGC). This causes the coded ADPCM samples to exhibit considerable variation during both voiced and unvoiced speech, but not during lower-level background silence, which causes the ADPCM quantizer to saturate at its minimum step size. Thus, a very sharp threshold between speech and no speech is obtained due to the effective AGC of the ADPCM coder and can therefore be used to distinguish between silence and unvoiced segments.

The basic principle of operation is as follows: For each ADPCM-coded speech sample $c(n)$, an energy calculation $E(n)$ is made. That is, the energy of the code-word sequence is computed over a 16-ms or 101-

[2] The ADPCM coder/decoder circuit.



* The largest negative quantizer output level is labeled 0000, and the largest positive output level is labeled 1111.
† Note that this is the case in the voice response application since we are only retrieving the code words from digital memory.

point window, centered around the most recent speech sample.* The energy values that are obtained this way are then compared sample by sample with a threshold, which is set midway between the measured code word energy of background silence and the average for speech. When the code word energy exceeds this threshold for 50 ms or 300 consecutive samples, the point at which the energy first exceeded the threshold is recorded as the *beginning* of an utterance. The algorithm then begins at that point and continues to make the energy comparison with the threshold. When the code word energy falls below the threshold for 160 ms or 1000 consecutive samples, the point at which the energy first passed below threshold is recorded as the *end* of the utterance. The 160-ms criterion ensures that a stop consonant within a word or phrase will not be mistaken for the end of the utterance.

The end-point algorithm was tested on 50 typical entries for a voice response system vocabulary and made essentially no errors. Auditory and visual verification indicated no evidence of shortening of any of the words. Two other measures of the coded speech signal—the energy in the difference signal and the energy in the quantizer step size—were also studied as possible signals for use with the automatic editing algorithm. However, the results based on the code words themselves were far more accurate than those obtained from any of these other signals.

* For hardware implementations, the magnitude of $c(n) - 7.5$ can be used in place of the square with little sacrifice in the performance of the algorithm.

Computer-aided voice wiring

One fairly straightforward and currently used application of a voice response system is that of computer-aided wiring by voice.[2,3]

Conventionally, a wireman works from a printed list that contains the information for each wiring operation. However, in many wiring situations it is awkward for the wireman to divert his eyes from his work in order to consult such a list. In these cases, it becomes convenient to record the wire list in spoken form on a cassette tape and allow the wireman to work from a spoken list. Typically, a foot switch is used to start the playback unit, and recorded tones on the tape automatically stop the unit after each wiring instruction.

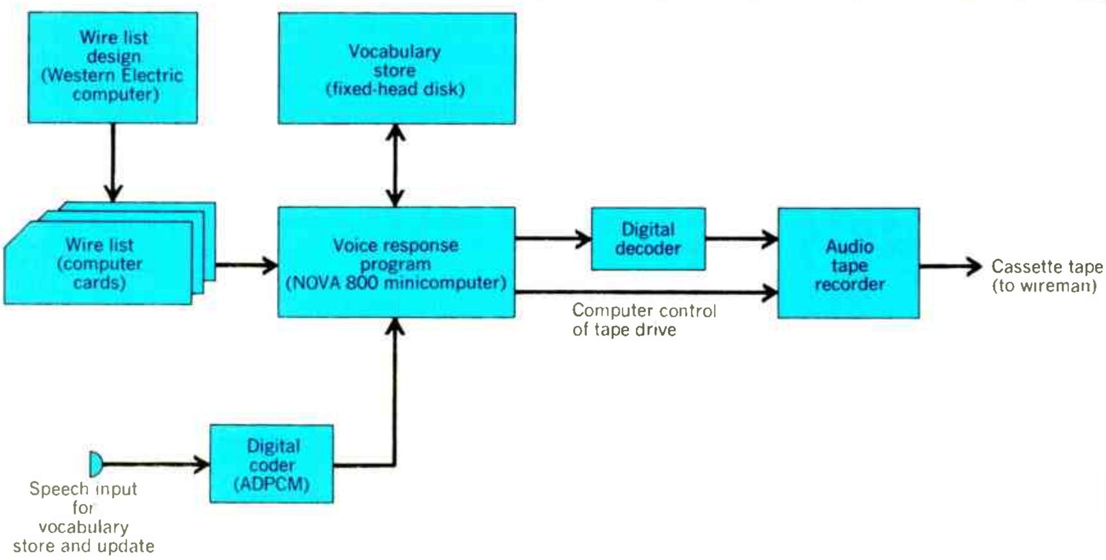
Generating wiring lists can be done either by man or by a computer. However, if the list is made without the aid of a computer, one person must read the list (which may be several hours long), and another must monitor the audio input. Any errors detected must then be edited and corrected. Even after a wire list has been recorded successfully, any future updating of the list requires that this entire process be repeated. This could occur several times in the course of a few days or weeks. The resulting tedium factor would be extremely high and would tend to increase human errors, making the entire procedure impractical.

This problem can be eliminated, however, by generating the spoken lists automatically with a voice response system. Moreover, if such a system operated faster than real-time, it would have multichannel capabilities. As a result, a number of independent lists could be generated simultaneously, or a long single list could be generated in pieces at a faster rate.

Figure A shows a system for generating spoken wiring instructions. A deck of computer cards is created on a Western Electric Company computer for use as input to the wire-list system. On these cards are punched entries that describe the words to be used in the specific wire list. The waveforms of individual utterances are inputted to the system for storage through the hardware digital ADPCM coder and are stored on the fixed-head disk. The computer accesses the words in the message from the disk and sends them out to a cassette tape recorder through a hardware decoder.

This system has been used for generating instructions spoken by a computer for wiring telephone equipment on production lines, and for directing equipment

[A] System for generating voice response wiring instructions.



Creating a vocabulary

The first step in creating a vocabulary for the voice response system is to store coded speech on the disk. Coded 4-bit speech samples are packed four samples to a computer word and transferred directly to memory before being stored on the disk.

Once speech is recorded on the disk, a cataloging program is next required to add, delete, and rename stored entries. Figure 3 shows a flow chart for the cataloging program. Basically, this program must perform bookkeeping on the speech directory and on the speech file for the tasks of adding, deleting, and renaming vocabulary entries. For adding entries, this program makes use of the automatic end-point algorithm already described.

The average storage requirement for an isolated word entry is four disk blocks or 1024 words. Thus, a typical vocabulary of 100 words would require approximately 100 000 words of disk storage. This amount of storage is readily available in most commercial fixed-head disk systems. The fixed-head disk used in our implementation (a NOVADISC®) had 786 432 words of storage. Hence, about 800 words can be stored on such a disk.

This is your voice response speaking!

The retrieval, or message composition, section of the voice response system must access the vocabulary entries necessary to form the desired output message and disburse them with minimal delay and no unwanted pauses. In the case of a multiline voice response system, the retrieval section must supply each output channel with its desired message, again with minimal delay and no unwanted pauses. In order to produce the most natural speech possible, the amount of silence between spoken entries must also be controlled. Finally, the system should be as modular as possible, so that it can be used in a variety of applications. The output system described in this article

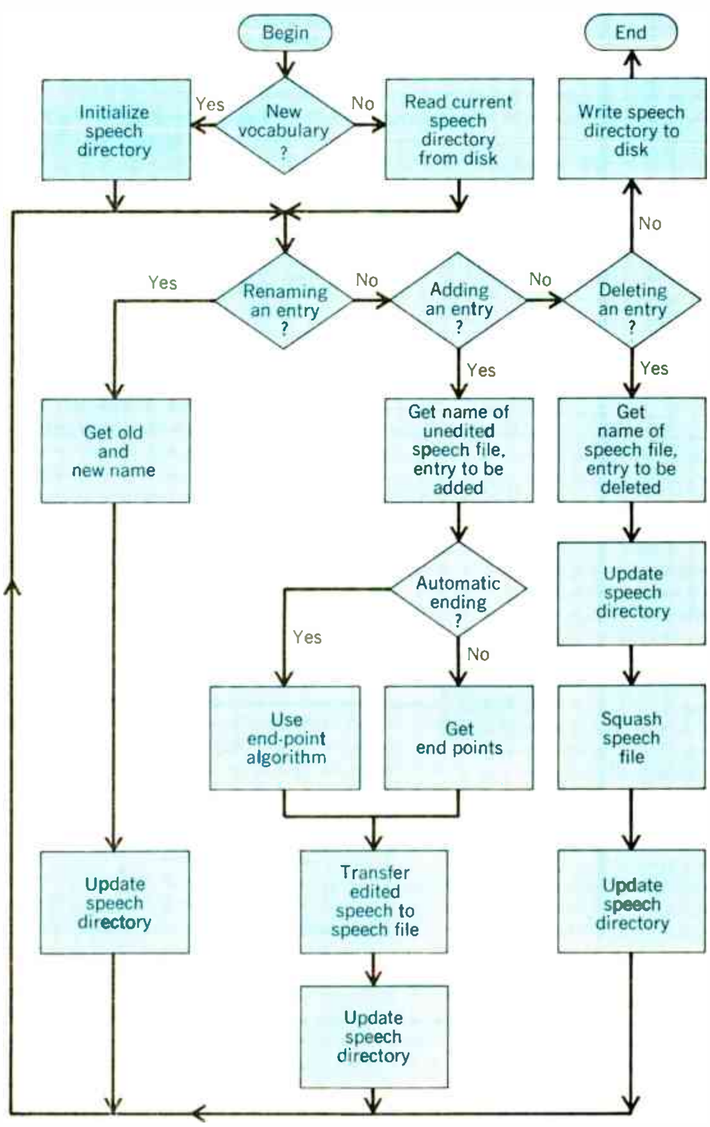
installers, while working on the wiring of an electronic central office.

Figure B shows a wireman using a specially designed cassette-recorder belt attachment to install wires in setting up a central office. Tone control signals are provided to automatically stop the tape recorder and key the installer at various points during the job.

This system has been used as a means of generating a fairly large number of voice-wiring lists for production line assembly, as well as two extremely large instruction lists for use in central office installations. Preliminary timing experiments indicate potential speedups of the order of 25 percent for some types of voice wiring over wiring from printed lists.

[B] Headphone instructions are received by this telephone installer from a belt-secured cassette player.

[3] A word cataloging program.



meets all of these requirements and the design principles discussed here can be applied in the design of a variety of sophisticated voice response systems.

The maximum number of output lines that a voice response system can support depends on many factors. However, a first estimate can be obtained by considering the hardware constraints imposed by the storage disk and coders. The ratio of disk data rate to coder data rate (960 kbits/s:24 kbits/s) sets a theoretical maximum of 40 output lines for our devised voice response system. In practice, however, this estimate cannot be attained, since it assumes perfect disk-read scheduling and ignores the possibility of a conflict occurring when too many coders request data from the same sector of the disk.

On the basis of this argument and various memory considerations, a goal of ten output lines was chosen as adequate for any of the applications intended for the system, while still keeping within the capabilities of the disk. Speech buffers of 256 words each were chosen to minimize core buffer requirements. With these constraints, the average demand on the system turns out to be approximately 12.5 disk reads in ten revolutions of the disk. Depending on the status of the channels at a given time, this ratio can vary from a minimum of ten to a maximum of 20 disk reads.

Real-time multitask programming

To handle up to ten output channels in real time, a voice response system must be capable of minimizing any wasted idle CPU time. For example, after a speech buffer for a given channel has been filled, the system must immediately proceed to service the next channel needing a buffer. To accomplish this, it is necessary to synchronize the emptying and filling of speech buffers for each channel. On the other hand, the ten output channels are completely independent of each other and should be kept isolated in the software. These requirements were met by using the real-time multitasking capabilities of Data General's Fortran.[12] These features include the ability to run several independent program units or tasks simultaneously on a priority basis, to communicate data between tasks through Fortran *common* areas, and to synchronize two tasks or a task and an interrupt service routine by means of special tasking routines. In addition, all codes generated by the compiler are reentrant, thereby permitting different tasks to share

Speaker verification

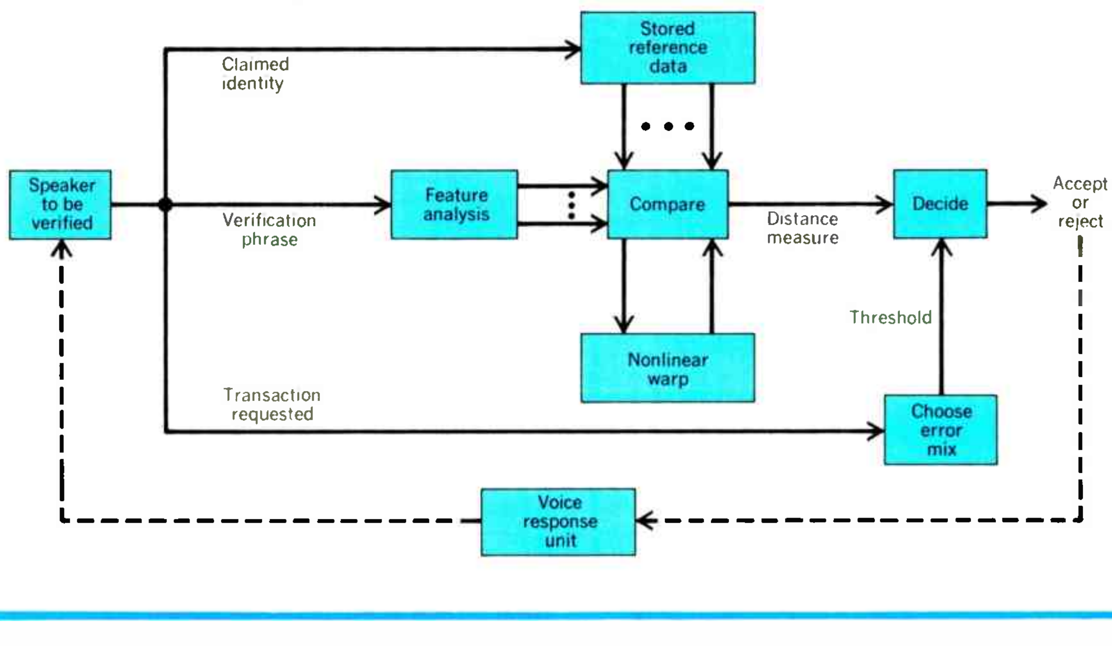
The voice response system has been used as an aid to experiments in speaker verification.

In the illustration, which depicts one such method for speaker verification,[5] the speaker desiring to be verified enters his claimed identity, speaks his verification phrase, and requests a transaction. (For example, in a banking application the customer may request withdrawal of a certain amount of money.) The computer performs a feature analysis on the verification phrase and—based on a comparison between stored reference data and a nonlinear time warped version of the features of the verification phase—computes a distance measure of similarity between the input phrase and reference data of the claimed identity. A decision is then made on the basis of the distance measured and the transaction requested as to whether to accept or reject the speaker. Clearly, a much lower threshold of acceptance is suitable for depositing into an account than for withdrawing from that account.

Since the scenario just depicted must take place over a conventional Touch-Tone® telephone, the job of the voice response system is to guide the user and provide a feedback as to whether he has been accepted or rejected. Of course, such a situation is a natural for voice input (i.e., speech recognition) to instruct the computer, but at the present time the input must be restricted to Touch-Tone telephone signals.

Other foreseen applications of the voice response system include directory assistance and automatic Dataphone® testing experiments.

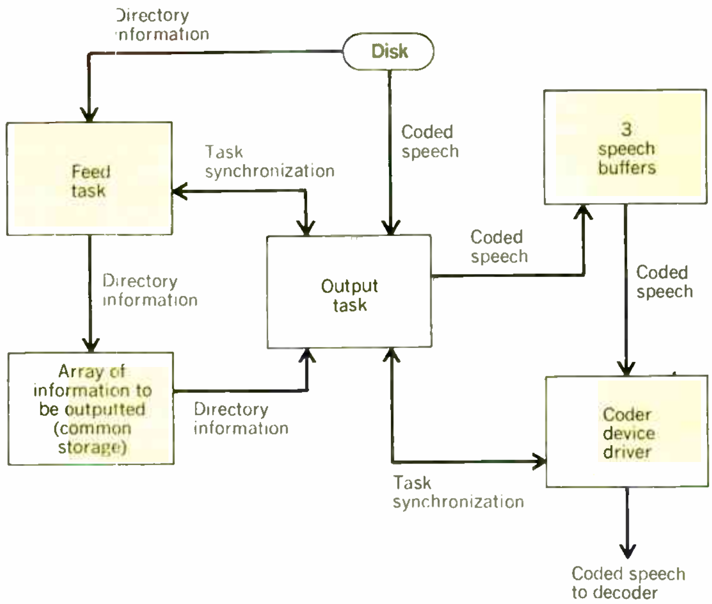
A speaker verification system.



the same code.

To illustrate how multitasking was used for the voice response system, Fig. 4 describes a typical channel in the output system. The software for each channel consists of three distinct program units. The first of these units, the *feed* task, must access the information to be converted to speech and load it into a common buffer area. This module will be different for each application of the voice response system. The second program unit, the *output* task, controls the output speech by making the appropriate disk read to fill the speech buffers. The information obtained by the *feed* task is used to determine which blocks of speech from the disk are to be read. Finally, the third program unit is the *driver* for the *speech coder*. This program must direct the coder to the next full buffer of speech to be outputted and signal the output task to fill any empty buffers. A separate driver program is necessary for each output channel. However, most of the code for the other two program units can be shared by all channels. (See Ref. 13 for details.)

Using the ideas discussed, a ten-line voice response system was built and tested. With up to ten simultaneous output lines of speech, no errors were detected



[4] Typical speech output channel.

Digital speech

A basic consideration in using digital storage is the selection of a method for digitizing speech.[8] These methods include *waveform quantization* schemes, such as linear PCM, log PCM, differential PCM (DPCM),[9] adaptive delta modulation (ADM),[10] and adaptive DPCM (ADPCM),[11] and *analysis/synthesis* schemes such as channel vocoders, linear prediction vocoders, and formant vocoders.[2] Information storage rates (bit rates) for these various techniques range from 60 000 bits/s down to about 16 000 bits/s for waveform coding schemes and from 10 000 bits/s to 1000 bits/s for analysis/synthesis schemes. All of these schemes require processing of the speech signal and they all provide greater flexibility of message composition and vocabulary storage/editing than analog storage.

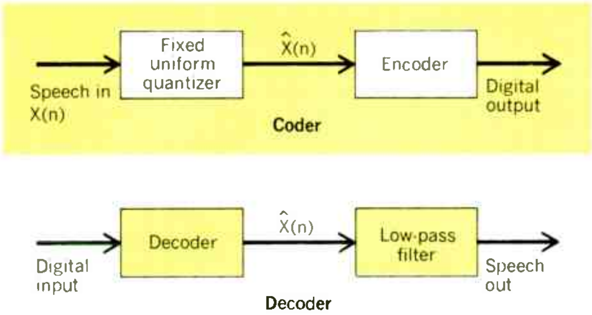
In general, the lower the desired bit rate, the greater the complexity and cost of the encoding and decoding hardware. However, the analysis/synthesis schemes also offer greater flexibility in composing messages, even though the use of these schemes for representing vocabulary elements is not required in situations where the vocabulary is small and when the message outputs can be composed of strings of words and phrases. Thus, in many applications, the waveform quantization schemes are most attractive.

The most direct method for digitizing speech is linear pulse-code modulation (PCM), which simply quantizes a band-limited analog signal both in time and in amplitude. The accompanying illustration simply shows the operation of a linear PCM coder and decoder. The input to the coder is the sampled speech waveform $x(n)$, which is discrete in time but not in amplitude. Digital output is obtained by quantizing $x(n)$ in amplitude with a uniform quantizer to produce $\hat{x}(n)$ and then encoding each quantized sample to produce a binary representation. In the decoder, the quantized signal $\hat{x}(n)$ is reconstructed from the bit stream and then low-pass filtered to produce the analog speech output.

The main disadvantage of linear PCM coding for voice response applications is that a rather high information rate (bit rate) is required to represent the speech signal faithfully. The bit rate is simply the product of sampling rate and the number of bits per sample. To obtain speech quality suitable for most voice response applications involving telephone lines requires a bit rate of about 60 000 bits/s (6-kHz samples × 10 bits/sample). Lower bit rates can be obtained by reducing the sampling rate and the number of bits per sample, but only at the expense of speech quality. If the sampling rate is reduced further, intelligibility is sacrificed. Alternatively, if fewer bits are used to represent each sample, the quantization noise will be greater. The latter problem is due to the limitations of a fixed, uniform quantizer. This quantizer must have both a step size small enough to provide good resolution at low signal levels and a large enough number of quantization levels in order to cover the peak-to-peak range of an input signal. For a speech signal, which has a very wide dynamic range, these requirements necessitate a high bit rate for linear-PCM-coded speech.

One solution to the dynamic range problem is to use a nonlinear quantization scheme such as logarithmic compression and expansion. This reduces the bit rate requirement at the expense of increased coder complexity. (For log PCM, the required bit rate is about 42 000 bits/s; i.e., 7 bits/sample times 6-kHz samples.) Because the bit rate of both linear and log PCM is rather high, they are still somewhat undesirable for voice response systems that require either a medium-size vocabulary or the capability of servicing many output lines simultaneously.

Basic linear pulse-code modulation coder/decoder.



I. Memory requirements for a ten-channel system

	Words
System	5k
Voice response control program	7k
10 speech buffers: 768 words/channel	7.7k
10 input buffers: 128 words/channel	1.3k
21 run time stacks: 500 words/stack	10.5k
	31.5k

II. Timing requirements for a ten-channel system

	Percent
Data channel transfers	3
Interrupt servicing	1
CPU computation	29
Disk read (≈ 18 percent)	
Other (≈ 11 percent)	
Available CPU time	67
	100

on any of the channels. Tables I and II summarize the memory and timing requirements for implementing a ten-channel system using multitask programming. As can be seen, in terms of memory requirements, a ten-line system is about the largest that can be implemented under system control (32k-word memory).

As the Tables show, the timing requirements for a ten-line system are much more flexible than the memory requirements. In fact, the CPU is working only about 33 percent of the time when running all ten lines. Consequently, of the order of 30 lines could theoretically be handled by the CPU before timing became critical.

REFERENCES

1. Flanagan, J. L., *et al.*, "Synthetic voices for computers," *IEEE Spectrum*, vol. 7, pp. 22–45, Oct. 1970.

2. Rabiner, L. R., Schafer, R. W., and Flanagan, J. L., "Synthesis of speech by concatenation of formant-coded words," *Bell System Tech. J.*, vol. 50, pp. 1541–1558, May–June 1971.

3. Flanagan, J. L., *et al.*, "Wiring telephone apparatus from computer-generated speech," *Bell System Tech. J.*, vol. 51, pp. 391–397, Feb. 1972.

4. Newell, A., *et al.*, "Speech understanding systems," Carnegie-Mellon Report, May 1971.

5. Rosenberg, A. E., and Sambur, M. R., "An improved system for automatic speaker verification," submitted for publication.

6. Cummiskey, P., "Adaptive differential pulse-code modulation for speech processing," Ph.D. thesis, Newark College of Engineering, 1973.

7. Rosenthal, L. H., *et al.*, "An algorithm for locating the beginning and end of an utterance using ADPCM coded speech," *Bell System Tech. J.*, vol. 53, pp. 1127–1135, July–Aug. 1974.

8. Jayant, N. S., "Digital coding of speech waveforms," *Proc. IEEE*, vol. 62, pp. 611–632, May 1974.

9. McDonald, R. A., "Signal-to-noise and idle channel performance of differential pulse code modulation systems—particular applications to voice signals," *Bell System Tech. J.*, vol. 45, pp. 1123–1151, Sept. 1966.

10. Jayant, N. S., "Adaptive delta modulation with a ore-bit memory," *Bell System Tech. J.*, pp. 321–342, Mar. 1970.

11. Cummiskey, P., Jayant, N. S., and Flanagan, J. L., "Adaptive quantization in differential PCM coding of speech," *Bell System Tech. J.*, vol. 52, pp. 1105–1118, Sept. 1973.

12. FORTRAN IV User's Manual, Data General Corporation, 1973.

13. Rosenthal, L. H., *et al.*, "A multiline computer voice response system utilizing ADPCM coded speech," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Oct. 1974.

The authors wish to acknowledge the contributions of K. L. Shipley and D. E. Bock of Bell Laboratories and R. H. Rabiner, G. Franzen, and J. Green of Data General Corporation, whose hardware and software assistance helped in implementing the voice response system.

Lewis H. Rosenthal received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology in 1974. Between 1971 and 1973, he participated in a cooperative program in electrical engineering at Bell Laboratories, where he worked on power conversion, digital multiplexers, and, most recently, automatic voice response. His study of the digital voice response system described in this article was the basis for a thesis submitted in partial fulfillment of the requirements for the B.S. and M.S. degrees.

Lawrence R. Rabiner (M) received the S.B. and S.M. degrees in June 1964, and the Ph.D. degree in E.E. in June 1967, all from Massachusetts Institute of Technology. From 1962 through 1964, he participated in the EE cooperative plan at Bell Laboratories, where he investigated digital circuitry, military communications, and binaural hearing. At present, he is engaged in researching speech communications and digital signal processing at Bell's Murray Hill installation.

Ronald W. Schafer (M) received the B.Sc.E.E. and M.Sc.E.E. degrees in 1961 and 1962, respectively, from the University of Nebraska, and the Ph.D. degree in EE in February 1968 from the Massachusetts Institute of Technology. During 1962–1963 he was an instructor in the Department of Electrical Engineering at Nebraska, and from 1964 to 1968 an instructor in the EE Department at M.I.T. During 1965, he was with the M.I.T. Electronic Systems Laboratory, and from 1966 he was a member of the Research Laboratory of Electronics at M.I.T. In 1968, he joined the Acoustics Research Department of Bell Laboratories, where he is now engaged in research on speech analysis and synthesis and digital signal processing techniques.

P. Cummiskey (M) joined Bell Laboratories in 1962, where he has been engaged in a variety of research projects involving formant synthesis and time waveform encoding of speech signals. He received his Dr. Eng. Sci. degree from Newark College of Engineering in 1973. As a member of the Bell Labs staff, he is currently involved in exploratory development of digital telephone systems.

James L. Flanagan (F) joined Bell Laboratories in 1957 after completing the doctor of science degree in engineering at the Massachusetts Institute of Technology. His interests have centered on voice communication and digital techniques for signal analysis and transmission, and he has been concerned with signal processing methods for bandwidth conservation and fundamental acoustical studies of speech and human hearing. In 1961, he became head of the Speech and Auditory Research Department, and in 1967 was made head of the Acoustics Research Department. He holds patents in the fields of speech coding, digital processing, and underwater acoustics, and has published technical papers in these and related fields. He is the author of *Speech Analysis, Synthesis and Perception*, which has been translated into Russian and Spanish.

Electric power's role in the U.S. energy crisis

A rapid increase in electrification is consistent with a necessary transition away from excessive dependence on oil and gas

The escalating imbalance between energy supply and demand may be the most difficult and pervasive problem facing the world today. And most responsible forecasts anticipate a major aggravation of the problem in the future. In the U.S., the root cause of the problem has been stated, almost universally, as the insatiable demand of the American public for energy as reflected in the exponential growth of its demand. The solution, as proffered by some, is the adoption of a no-growth or limited-growth philosophy and, by others, as the conservation of end-use consumption, either voluntarily or by government edict. These tenets will be examined here together with the role of electric power, a form of energy that has been growing at nearly double the rate of total "raw" energy demand (7.2 vs. 4.4 percent annually).

Although the sustained high rate of growth in energy demand in the past has had a direct and major impact on the emerging energy crisis, it is doubtful that it is the root cause. The causes are manifold. They relate to governmental tax, pricing, and regulatory policies; inadequate planning on the part of the energy industries themselves; political considerations; and, more recently, attempts to impose instant solutions to long-standing environmental problems.

The proposition that the problem is simply one of excessive demand created by a "cheap energy" policy is also fallacious. Energy demand has risen steeply in many European nations and in Japan even though these nations have depended heavily on energy imports and have not had cheap energy policies.

Energy is such an intrinsic part of modern life that price variations, even over a wide range, may not have a pronounced near-term effect on demand. Consumer decisions regarding the purchase and utilization of energy-consuming devices are usually based on the need or desire for such devices and on their price, not on the cost of the energy to operate them.

Energy, in terms of all fuels and electricity, accounts for only some 3 percent of the expenditures constituting the Consumer Price Index.[1] Also, the low content of value added by energy to the total cost of a product in industry hardly constitutes—except for a very few industries—a major factor in decision-making. While energy costs are rising rapidly, so are the costs of raw materials and labor. The availability of energy makes possible a highly industrialized society; it does not, through pricing, create one.

T. J. Nagel
American Electric Power Service Corp.

The no-growth theory

In a democratic society, the resolution of major public issues can be accomplished only after enlightened discussion aimed at reaching a solution which the majority of its citizens will support. A solution to the energy crisis cannot be advanced either by arguing the merits of a no-growth society or by projecting unlimited exponential growth. In fact, there is no way to enforce a no-growth or limited-growth policy on the bulk of the American public at this point in our nation's history without creating domestic havoc and further jeopardizing our security in an unstable world. The causative factors already exist in our society which, at the present or an even higher rate, will require continued growth in the demand for energy during the next 10 to 15 years.

The Arab oil embargo escalated the shortage in our energy supply, mandating measures for reducing energy consumption. The public's willingness to share these burdens may have led some to conclude that energy growth can be curtailed with only a mild sacrifice in comfort and convenience. Such thinking is fallacious because it does not recognize that the component of energy that Americans have temporarily relinquished is a small part of total energy use. The remaining energy demand must increase rapidly in the years ahead to insure the well being of our economy.

In industry and agriculture, the need to compete in world markets—particularly in view of our inevitable reliance on imported petroleum—and the need to provide jobs and food for a growing population make the thought of near-term reductions in energy growth unrealistic. Even in the residential and commercial sectors there is every indication that energy must inevitably play an increasing role.

Forecasting a sustained or even increased rate of growth in energy demand during the next 10 to 15 years, however, does not imply exponential growth forever. All physical as well as societal processes have a growth cycle and reach levels of maturity and saturation. Predicting the point of change in slope on an S-curve of energy growth in a society where growth depends on a multitude of interdependent decisions—economic and otherwise—is more in the nature of soothsaying than forecasting.

Questions of economic growth and energy demand must ultimately be resolved in the U.S. form of government through constitutional processes. But it is incumbent upon the power industry to commit generating capacity additions on the basis of today's evidence and trends and not on any self-assigned wisdom regarding the public good. Electric energy—unlike other energy forms—cannot be stored and must be immediately available when needed. The power industry, therefore, must be able to respond quickly to demands that result from the interaction of a myriad of private decisions made within the framework of prevailing public policy.

What about the environment?

We all have become increasingly concerned over the unwanted byproducts that accompany the expansion of energy processing and electrification. While electric power production does contribute to pollution, it is important to recognize that it also has greatly reduced the pollution that would otherwise have existed in its absence. Just as utilization of energy has brought us to our present state of an improved human environment, it can also help to create a better natural environment. It is up to us to provide the next generation with a lifestyle at least as good as our own, including their material well being and a healthy natural environment, both of which will surely require a substantial increase in energy utilization.

Full internalization of costs—environmental or otherwise—is an accepted economic theory, but in our complex, interdependent economy, it cannot be applied solely to energy, but must encompass all resources and commodities. If it is the public will that such a policy be adopted, it needs to be done on the basis of carefully established scientific evidence and not on the basis of a credo adopted in an emotional atmosphere. A sufficient period of transition would be essential both to allow for technological development and to avoid severe economic dislocation. It could be economically disasterous and harmful to the public interest, for example, if the power industry were required to install costly equipment for the removal of sulfur oxides before such equipment proved operationally capable of doing the job.

The concept of eliminating economic subsidies to energy industries, while seemingly sound, does not address the full issue. The intricacies of today's economy make simple, broad-gauged solutions oftentimes inadequate. It is unrealistic to think that we have a totally free economy where the laws of supply and demand work automatically. Rather, governmental policy and regulation have a pervasive effect on all segments of our economy. The elimination of subsidies in one area may, in turn, call for other efforts and incentives to encourage exploration and stimulate research and development.

Energy supply and demand

The most important study period for any discussion of energy supply and demand is the near-term, extending from now into the 1980s. If we cannot arrive at an accommodation during this period, it will do little good to worry about the year 2000 and beyond.

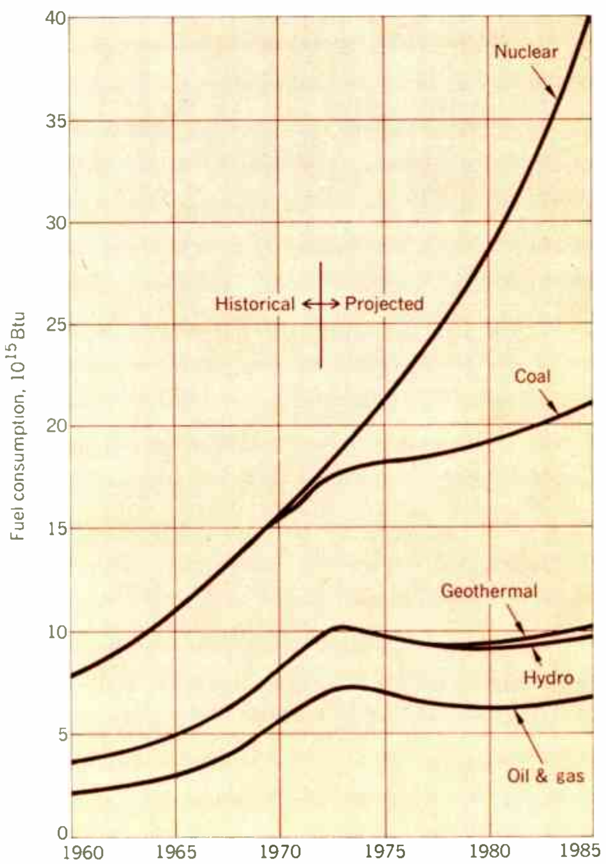
Some perspective may be gained by examining our energy consumption to date as compared with our primary energy supplies. Although geological estimates vary, a conservative accounting of our indigenous domestic energy resources in their original state, including petroleum, natural gas, coal, and uranium, is about 55Q (Q is defined as 10^{18} Btu).[2-3] The effect of breeder technology could increase this energy resource base to 1000Q. Our annual consumption is presently about 0.07Q and could reach 0.2Q by the year 2000 if the present growth rate continues. To date, we have consumed less than 3Q, or less than 6 percent, of these resources—a clear indication that the nation is not running out of primary energy supplies.

It so happens, however, that oil and gas combined supply about 75 percent of our present energy needs, yet these are the least abundant of our fossil fuels. Our original resource base of oil and gas is estimated at less than 5Q[2] and we already have consumed about one fourth of this amount. Unfortunately, government policy—through taxation, regulation, and, more recently, environmental standard-setting—has not only failed to encourage exploration for and recovery of these fuels but, in fact, has encouraged their use while discouraging the use of our more abundant energy resource, coal. In addition, the timely introduction of nuclear energy has suffered temporary setbacks.

It is obvious that we must now hasten the exploration for new oil and gas supplies and encourage the increased use of coal, while in the longer term both nuclear energy and coal must take up the burden of energy supply. Electric power production provides for the most effective use of coal and is the only technological means for the conversion of nuclear fuel to useful energy. Increased electrification is the only logical way to ease the burden on our premium fuels.

Figure 1 shows the recent history of primary fuel

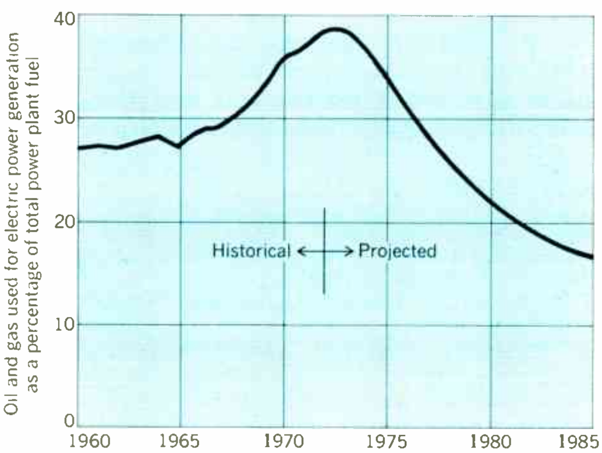
[1] How primary fuels have been, and will be, used for electric power generation to 1985.



consumption for electric power generation as well as a forecast to 1985.[4] In Fig. 2, it is shown that gas and oil have supplied an increasing portion of the fuel for electric power generation, reaching approximately the 38-percent level in 1972. But the historical upward trend is expected to reverse drastically during the years ahead. The increase during the recent past was, in large measure, attributable to the power industry's heavy conversion from coal to oil in an attempt to meet unrealistic air-pollution standards.

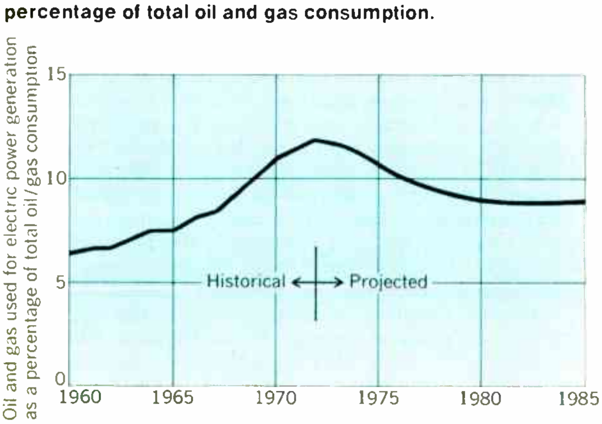
In the context of total oil and gas consumption, electric power generation accounts for a relatively small share (Fig. 3). This portion of the oil and gas market, as shown in Fig. 3, is likely to drop to 9 percent or less by 1985.

The exploitation of coal and nuclear energy is not without problems. Surface mining, which now provides 50 percent of the total coal produced, is under serious attack because of its possible environmental impact on the terrain. Wise and reasonable reclamation measures can alleviate the problem. In the eastern portion of the country, where the greatest demand for electric energy exists, much of the coal is of relatively high sulfur content. With presently available technology, its combustion results in the release to the atmosphere of considerable amounts of sulfur oxides. In spite of the absence of a proven technology for the removal of such gases from the stack or removal of the sulfur from the coal, and in the absence of adequate scientific documentation on the primary (health) and secondary (other) effects of varying ambient levels of sulfur oxides in the atmosphere, severe emission standards have been established that will prohibit the use of the abundant reserves of Eastern high-sulfur coal.



[2] The power industry's heavy consumption of oil and gas is expected to drop off drastically in the next ten years.

[3] Electric power generation accounts for only a small percentage of total oil and gas consumption.



The contribution of nuclear power to the energy crunch has also been handicapped. Construction difficulties, compounded by licensing delays and litigation resulting from environmental-based opposition, have aggravated the problem. If present trends continue, the forecast for nuclear energy shown in Fig. 1 may well be too optimistic.

A solution to the crisis?

Confronted with the energy crisis and seemingly endless problems on all fronts, what is the solution? While no one of us can claim to have all the answers, there are some widely held misunderstandings which need to be examined.

Energy conservation—eliminating waste, sacrificial abstinence, and increased efficiency (doing more with less)—as a means of solving our near-term problems has been greatly overstated. There is no justification for outright waste and it should be discouraged—with or without an energy crisis—but to classify a particular energy use as waste often strikes at an individual's set of values.

The sacrificial reduction of energy consumption, which we have observed during the last several months, is important during times of national crisis. It helps to prevent shortages of energy in areas critical to the national economy. Voluntary sacrificial abstinence can only be temporary, however, since people eventually rebel against discomfort, inconvenience, and the economic effects of loss of production. Forced abstinence, through rationing, encroaches on human liberty and should be used only if absolutely necessary, and then only on a temporary basis.

Conservation through improvements in efficiency of both energy processing and utilization—probably the most desirable and effective approach—is not new. History is replete with examples: improvements in average energy conversion efficiencies in electric generation (5 percent in 1900 to 32 percent in 1970) and electricity for illumination (1 percent for early-day incandescent lamps to 20 percent for modern fluorescent tubes); electricity required for aluminum reduction (10 kWh/lb in the mid-1940s to 4 kWh/lb in the near future)[5] However, such improvements involve technological and economic factors apart from energy efficiency and take significant periods of time for implementation.

Transportation, a sector of the economy that now accounts for 25 percent of our primary energy usage, is much discussed as an area of vast potential for energy savings. The inefficiency of the automobile when measured in terms of energy consumed per passenger-mile has been well publicized. But the nation has an investment of over 100 billion dollars in some 90 million automobiles, most of which will last through the 1970s. And the highly desirable goal of substantially improved intracity rail and local mass transit systems is still more than a decade away, even assuming the unlikely immediate commitment to such goals.

Much has been said about the desirability of energy savings through improved home insulation. It is interesting, in this regard, that homes designed for electric heating typically have superior insulation and may have perhaps half the heat loss of the average fossil-heated home.[6] This factor, together with the unique control features of electric heat, makes it at least as efficient, if not more efficient, in primary energy consumption on an overall basis.[7] Electric heat has the added benefit of being able to utilize the more plentiful fuels—i.e., coal and nuclear. The electric heat pump, which is commercially available but only sporadically used, should be given greater consideration for energy conservation.

Apart from the type of heat source, it is possible to retrofit existing homes with additional weatherstripping, thermally "tight" windows, and roof insulation. These improvements, applied universally, might reduce heat losses in residential and commercial buildings by about 20 percent, with primary energy savings of about 2 percent of total requirements in 1985.[8] Extensive reinsulation in the walls of existing homes is unlikely since it is very expensive and would require disassembling the walls in many cases.[9]

Improved standards in the building industry and in energy-consuming equipment are certainly worthy objectives. However, improved efficiency in energy use should not be the sole goal. In order for true resource conservation to be effective, overall economic efficiency, rather than merely energy utilization efficiency, should be the real objective. This requires a balancing of the impacts of reduced energy use against the increased consumption of other limited resources, including the energy required in their development.

On the basis of the foregoing, it is certainly doubtful that conservation, by itself, can provide a solution to the energy crisis. While efficiency in energy processing and end use are desirable objectives, their implementation is long term in nature and their impact in the near term can only be marginal. While such improvements should be pursued, it is essential to direct such efforts not simply toward efficiency in energy processing and utilization but toward the goal of overall economic efficiency.

Another proposed solution is a heavy commitment to research and development in new energy technologies. This is often based on the mistaken belief that if only enough money is invested, as was done in the Manhattan Project or the space program, quick solutions will be forthcoming. An enlarged program of research is essential, but recognition must be given to the time frame involved. The timetable for the integration of a major new energy technology from the laboratory stage to commercial reality must be measured in decades, not years. Nuclear power, already about 25 years old, is just now making a measurable contribution (approximately 2 percent of total energy needs), despite the vast investment in money and manpower by both government and industry. The breeder reactor cannot have a substantial impact on energy supply before the late 1990s, even if a demonstration plant goes on line successfully by 1985. Near-term solutions to our energy crisis must be based on existing technology not on technologies yet to come.

Realities we must face

Without attempting to define anything as complex as an energy policy (if a definitive policy can be prescribed at all), there are a number of actualities that we must face and some trends that, of necessity, will characterize the energy picture in the near term.

1. We must recognize that substantial growth in U.S. energy consumption during the next two or three decades is both necessary and desirable for the well being of our citizenry.

2. We must hasten the exploration for and recovery of additional indigenous oil and gas supplies. However, off-shore and deep-well drilling can hardly have a major impact before the 1980s, and other forms of energy must play a greater role in the near term.

3. We must encourage alternative fuel consumption, particularly high-sulfur coal and nuclear power, wherever prudent, to reduce our dependence on oil imports.

4. We must accept the fact that a rapid increase in electrification is consistent with the necessary transition away from excessive dependence on oil and gas, the scarcest of our endemic fossil fuels.

5. We must carefully reconsider our current and proposed environmental legislation, giving proper attention to our indigenous energy reserves and presently available technology, and arrive at a suitably balanced accommodation of energy needs and environmental safeguards.

6. We must pursue and encourage expanded energy R&D with the greatest emphasis on those areas that are in advanced stages and that have the greatest promise of meeting our needs in the near term.

REFERENCES

1. Netschert, B. C., "Energy vs. environment," *Harvard Business Review*, p. 134, Jan.–Feb. 1973.
2. Landsberg, H. H., *et al.*, *Resources in America's Future*. Johns Hopkins Press, 1963: pp. 390, 407, 414.
3. Sherman, J. T., "Uranium 1971," *Engr. and Mining Jnl.*, vol. 173, no. 3, 1972.
4. Certain background information for consideration when evaluating the *National Energy Dilemma*, staff report, Joint Committee on Atomic Energy, 1973.
5. *The Wall Street Journal*, Jan. 12, 1973.
6. "Space conditioning passes close scrutiny, gets nod," *Elec. World*, p. 88, Feb. 15, 1972.
7. "Electric heat conserves energy, cuts ground level air pollution," *Contractors Elec. Equip.*, p. 8, Aug. 1973.
8. *National Energy Dilemma*, Joint Committee on Atomic Energy, 1973, pp. 7–8.
9. Rosenberg, R. B., "Energy usage in the home—consumption and conservation," paper presented at the 1974 annual meeting of the AAAS, San Francisco, Calif., Feb. 24–Mar. 1, 1974.

Theodore J. Nagel (F) is senior vice president, system planning for American Electric Power Service Corporation, a position he has held since June 1973. Mr. Nagel is a member of various industry technical and advisory committees and has had a particularly active role in advisory committee work with the Federal Power Commission, participating in its first National Power Survey as a member of the Transmission and Interconnection Special Technical Committee and in its second survey as a member of the East Central Regional Advisory Committee. He was formerly a vice-chairman of the Commission's Advisory Committee on Reliability of Electric Bulk Power Supply, and is currently a member of the National Electric Reliability Council's Technical Advisory Committee.

Germany automates its rails

The key to West German railway efficiency resides in electronic signals and continuous automatic train control

German signaling and train-control technology was not created in a quantum jump from the shambles of 1945 to the present-day level of sophistication. The advances came in stages beginning with automatic magnetic train control, followed by continuous train control, automatic train operations, and finally integrated transport control.

Thus this article bites off a chunk of German Federal Railway (DB), suburban commuter lines (S-Bahnen), and urban mass-transit systems (U-Bahnen) technology that is so complex that it would be difficult for the reader to chew in an article of limited length—unless it is somewhat "digested" in content. Therefore, we shall present a broad-brush overview of the railway signaling techniques, track-to-train communications, and safety systems that have evolved in the Federal Republic since the end of World War II (with some chronological assistance from the box on p. 76).

Automatic magnetic train control

To protect trains from accidents that are caused primarily when stop signals are not obeyed by train operators, wayside equipment is necessary. At first, mechanical stop devices, consisting of trip levers to set a train's brakes were employed. But these devices were, at best, only partially reliable and were unsuitable for trains traveling at high speeds because the force of emergency braking could injure passengers or damage merchandise. Further, these mechanical equipments were expensive to maintain and had to be replaced frequently.

More advanced technology, however, demanded supervision and surveillance of the train driver, a check of the braking systems before application upon a warning signal, and supervision of train speed prior to stop signals. When these criteria were met, it was possible to use equipments that

- Controlled train motion without mechanical components
- Could be used at all operational train speeds
- Differentiated between various control functions and could transmit control signals from the wayside to the multiple-unit or locomotive cab
- Could operate in accordance with a closed-circuit principle—both wayside and on-board—to ensure fail-safe functioning of equipments

Of the various methods used to achieve these objectives, West Germany's Siemens Electric devised the inductive multifrequency resonance and the permanent-magnet transmission systems. The former system, called "Indusi" (the German acronym for "Inductive Train Control for Rail Traffic Safety"), dates back to 1959, and it still fulfills operational fail-safe requirements for high-speed mainline service. The latter permanent-magnet system is employed primarily on S-Bahn and U-Bahn systems. The first such successful installation was made back in 1930. After

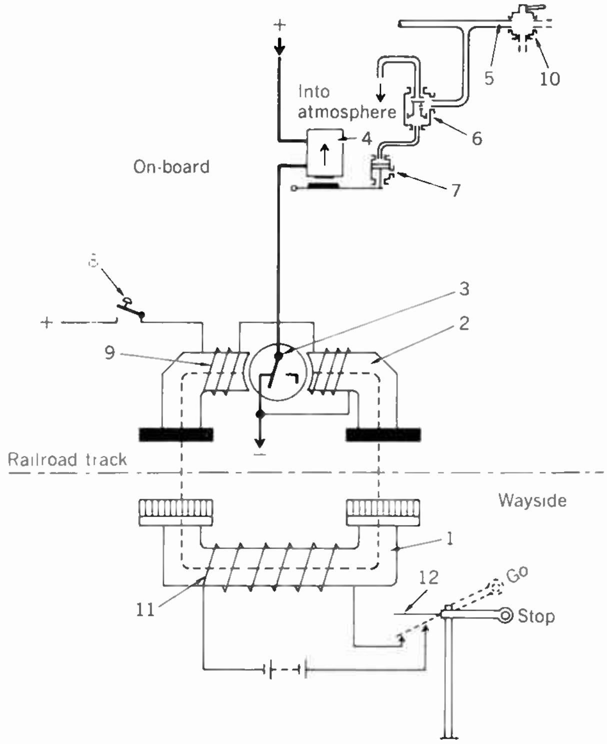
Gordon D. Friedlander Senior Staff Writer

In the April issue of *Spectrum* the author discussed the traction equipment, rolling stock, and other aspects of the amazing post-war revival of the German Federal Railways (DB), the computer lines serving major cities (S-Bahnen), and the urban mass-transit systems (U-Bahnen) within the Federal Republic.

[1] Automatic magnetic train control. Basic layout of permanent-magnet train-control system that has been in effect for many years on German mainline railways, S-Bahn, and U-Bahn systems.

Legend:

1. Track magnet	5. Main-air pipeline	9. Auxiliary winding
2. Train magnet	6. Emergency valve	10. Driver's braking valve
3. Magnetic relay	7. Transmission valve	11. Quenching winding
4. Brake-valve relay	8. Release key	12. Signal contact



World War II, design improvements were made on this basic system—thanks in no small measure to advanced magnetic materials.

Figure 1 is a line diagram of the wayside and on-board magnetic system. The wayside equipment includes a track magnet (1), which is generally placed between the rails near a block signal. This magnet contains a permanent magnetic element whose upward-directed force field is captured by a pole of the train's magnet (2) as the latter passes the wayside equipment and is fed to a relay (3). In normal position, the relay contact is closed and is interposed in the closed circuit of a braking relay valve (4).

An emergency valve (6) is connected to the train's main air-brake pipeline (5). One of the two chambers of the emergency valve is connected with a transmission valve (7), by means of a small-diameter tube; the other is attached to the main air pipeline.

If a train passes over the active track magnet in violation of a "stop" signal aspect, the magnetic force field in the train magnet reverses the relay contact (3). The closed circuit is interrupted and the relay (4) becomes deenergized; its armature opens the transmission valve to produce a vacuum in one chamber of the emergency valve. The membrane of this valve actuates the air-braking system of the entire train until it is either brought to a stop or there has been an adequate speed reduction. The enforced braking can be cancelled by means of a release key (8). When this key is manually actuated, a current circuit for the auxiliary winding (9) is closed. Thus, the contact in the relay reverses position, and the braking valve relay again reverts to its initial position. The use of the key is recorded in a counter mechanism; following which, the main-air pipeline can be refilled by the driver's brake valve (10). The train is then ready to proceed.

In the event the wayside signal aspect is green, then the appropriate track magnet will not interfere with the train's progress. To allow for nonintervention, the quenching winding (11) is incorporated in the track magnet; and, when the signal aspect is green, this information is fed, via the signal contact (12), from a dc source via the "green signal" current.

CTC and CATC

In operating its mainline trains at speeds ranging from 160 to 200 km/h (speeds now being attained on the Intercity System), the DB management believes that existing wayside visual signals must be augmented by continuous track-to-train communication. Starting with an experimental continuous-train-control (CTC) installation in 1970, between Munich and Augsburg, more than 700 route kilometers of mainline right-of-way are now equipped with track conductors as the basis for an eventual LBZ (Linienzugbeeinflussing), or Continuous Automatic Train Control (CATC) system, that will meet the on-board cab signaling and speed-supervision requirements prior to its integration into a cybernetic CATC operation.

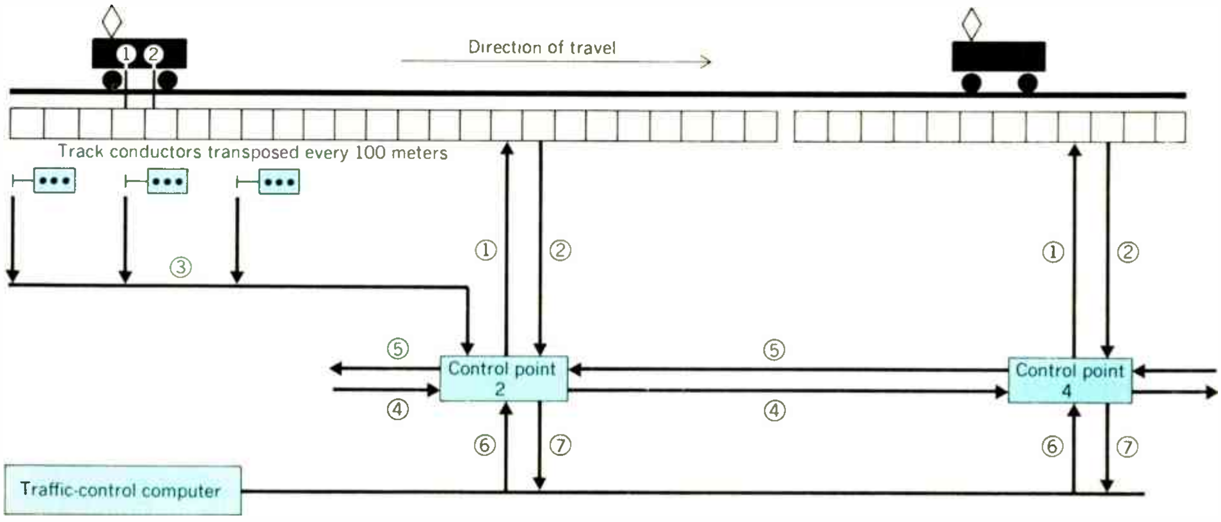
LBZ combines both automatic train-control (ATC) and automatic train-operation (ATO) features. Thus, the functional scope of LBZ encompasses

- Continuous protection of train movements under the direct surveillance of wayside signals and line sections that have permanent as well as temporary speed restrictions.
- Sensitive vehicle speed-checking equipment that permits headway reduction.
- Cab instrument displays to eliminate the dangers of misinterpreting signal aspects due to adverse weather conditions.
- The possibility of fully automatic train operation (at some time in the future).
- The modification of present braking practice to minimize the necessity of emergency-situation braking.

Figure 2 is a block diagram of such an integrated system, in which operational and safety information is transmitted from the control point after being processed as a coded instruction in one message to the train (channel 1).

The reply message (channel 2), from the train to the control point completes the closed-loop control. CATC safety information is a function of the charac-

[2] **A CATC (continuous automatic train control) communications network. The significance of channels (1) through (7) is explained in the text.**



teristics of the right-of-way (gradients, curves, etc.), the type of traction unit (locomotive-hauled train or multiple-unit motor-coach sets), and spacing of wayside signals. Taking these factors into consideration, a speed limitation may be set. This information is transmitted to the train operator as a visual display (Fig. 3) on the cab's instrument panel.

If a speed limitation is determined by operational factors (such as a heavily trafficked CTC junction or interlocking, late-running trains on the line, or other unforeseen events), the control points receive—via channel 6, Fig. 2—instructions from a central computer to reduce the permissible speed below the specified limit. Whether the execution of the speed command is actuated manually or automatically depends upon the equipment in the cab of the motive-power unit. However, if more detailed information is required by the train's operator, it can be transmitted simultaneously over carrier-frequency channels.

Information exchange

The exchange of information between train and wayside control points is accomplished by continuous track conductors (Fig. 2) transposed every 100 meters. In operation, each motive-power unit equipped with CTC (or CATC) measures the distance it covers from the start of the supervisory zone and reports its exact instantaneous position by transmitted message. This allows the control point (a fixed station) to determine what is the most restrictive information relevant to this position; it then transmits a "selective program" to the train.

The trains within the controlled zone are addressed, in turn, following their entry. They report their position, instantaneous speed, and other data, and these data are processed in the control point for possible generation of new command instructions.

Information is transmitted by a frequency-shift method, using time-division multiplex coding. In the track-to-train direction (channel 1, Fig. 2), 36 Hz is employed; in the train-to-track direction, the frequency is 56 Hz (channel 2). A standardized telegraph speed is employed: for the track-to-locomotive direction, it is 1200 baud; in the reply message (locomotive-to-track), it is 600 baud.

However, additional information can be interchanged between track and train via a telephone link over which compatible transmission channels are furnished. Speech communication is imperative, of course, in emergency situations.

Adjacent control points, however, must exchange pertinent information in order to ensure a uniform message flow when passing from one supervisory zone to the next. Thus, in Fig. 2, although the train is in the zone controlled by control point 2, control point 4 must have the details of the braking characteristics and the train's length to ascertain the speed restrictions of entry to the next zone (channel 4). Therefore, control point 4 transmits to control point 2—channel 5—information provided for the first position (in 4). This is used in conjunction with information in control point 2 to present the situation ahead.

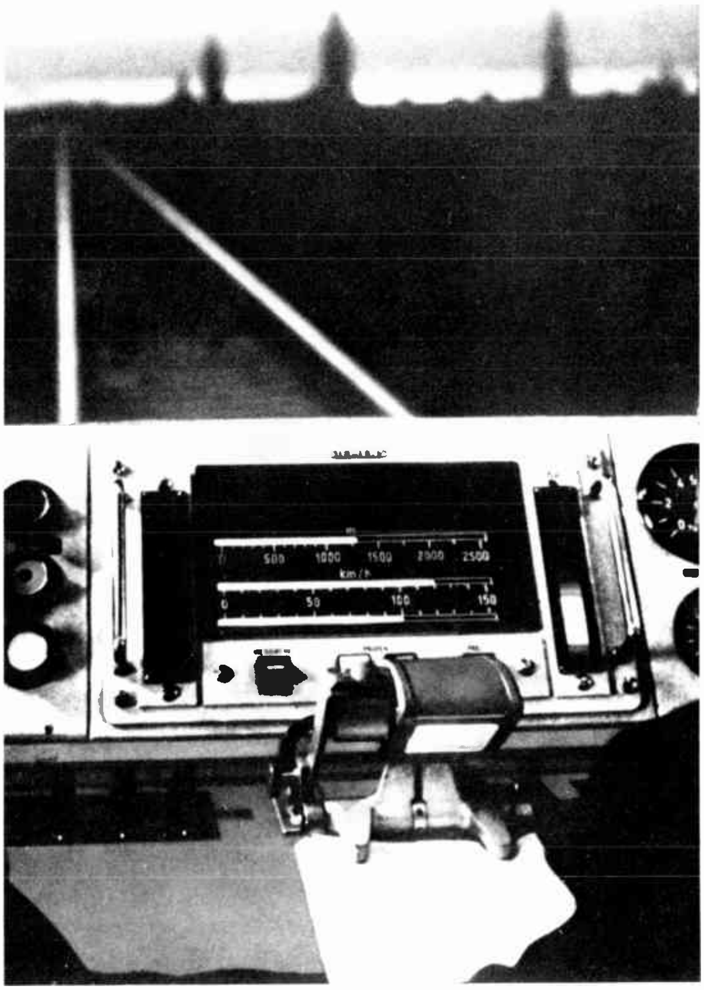
The supervisory operations center is always apprised of the operating situation in the controlled zone, especially the flow of traffic from adjacent lines and the relative importance of trains arriving at junctions, or terminal interlockings, by means of channel 7 (Fig. 2).

In order to optimize passage at junctions, or passenger interconnections, it is often advisable—despite the fact that the particular train may be running late—to have that train approach the junction at reduced speed in the interest of maintaining the schedules of other trains on the same or adjacent systems.

In the present developmental stage of the DB's CTC system, the technically permissible speed is transmitted from the control point to the motive-power unit. But, when CATC is inaugurated, a central supervisory computer will instruct the appropriate control point to transmit to a particular train either the maximum permissible (technical) speed limit or a lower speed command (Fig. 3) for operational reasons.

Each control point serves a zone 10- to 12-km long,

[3] CATC visual-display panel in the drivers' cabs of DB locomotives and S-Bahn train sets. The top analog indicator shows the distance, in meters, of a "target" ahead (either a train ahead on the same track, or a restrictive wayside signal); the middle indicator tells the driver the permissible speed in his zone; the bottom indicator shows the instantaneous speed at which the train is running. All speeds are in kilometers per hour.



Railway signaling technique timetable: centralization, mechanization, and automation

1948: First German push-button-controlled track-diagram interlocking display installed at Dusseldorf-Derendorf station, complete with a control desk and switching equipment that consisted of plug-in standardized relay groups; route control with start–destination buttons in the track diagram; and 380-volt three-phase ac power-point circuit, with extended control distance.

1951: First train-number indication and teletransmission installation for automated train-run supervision at Cologne (Koln) Central Station. Automatic block CTC, with pulse code, and automatic train-number communication introduced between Bebra and Cornberg on DB mainline.

1952: CTC system placed on line from Regensburg to Nürnberg, with pulse-code rotary-selector control. CTC central office established at Nurnberg.

1953: General introduction of standard push-button interlockings for intermediate DB stations. Development of geographic circuitry interlocking technique by Siemens in cooperation with DB.

1957: Central interlocking, with automatic train routing for greater Frankfurt area installed; indications given in advance of type of train, its destination, etc., by alphanumeric train codes.

1959: Inductive train-control development completed (known as "Indusi I 60" system). In use on DB since 1960.

1960: Train destination platform indicator, with electrooptic punch-card program control put in operation at Braunschweig Central Station.

1962: Carrier-frequency manual block developed for DB and placed on one line. Development of a three-phase magnetic-core transistor circuit system, with logic element operating on the fail-safe principle also occurred that year.

1964: Largest geographic circuitry interlocking system installed at Munich main station in conjunction with a pulse-code-relay CTC system, alphanumeric train description, and automatic block installations for an automatic train routing.

1965: Electronic CATC equipment (for train speeds up to 200 km/h) placed in operation on the DB's Augsburg–Munich mainline. Introduction of an automatic block system on the DB, in which all signals of the block sections (up to a distance of 6.5 km) joining a departure route are connected, with their switching elements to the geographic circuitry interlocking system.

1966: Placed in service at Kassel—U-Strassbahn system with inductive transmission of data between trains and fixed track (wayside) devices for control of routes, signals, train destination platform indicators, etc. Installation of an audio-frequency track circuit for an automatic "track clear" and "track occupied" indication without insulated rail joints.

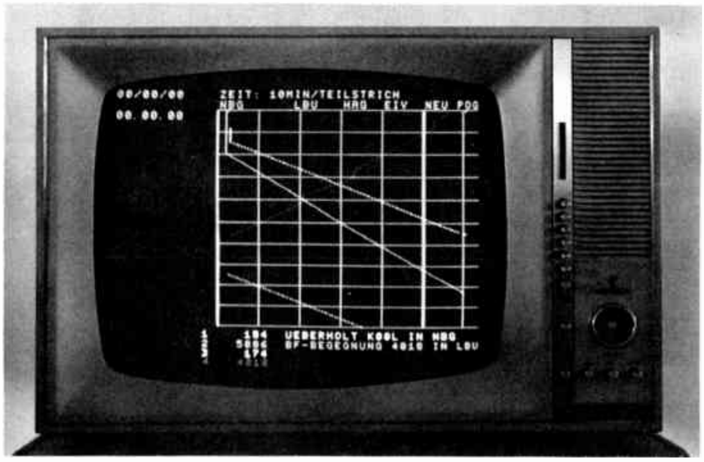
1967: DB field tests of an integrated transport-control computer, with input and output devices, at Kassel, Seelze, and Uelzen. Also test conducted on a process-control computer at the Siemens Braunschweig works for the solution of on-line and off-line tasks.

1968: Microwave system for reading of 12-digit vehicle numbers placed in experimental service on the DB's lines in the Munich area.

1970: First "superexpress" locomotive (DB series 103) was run under full automatic control (CATC) by a Siemens process-control computer, series 304.

equipped with track conductors, and provides all trains within this zone with necessary data. But since a number of trains may occupy this zone simultaneously, a message must be discretely addressed to ensure that information transmitted is received only by the relevant train. The elements of this discrete form could be

[4] CRT display of a Siemens computerized line supervision situation involving three trains.

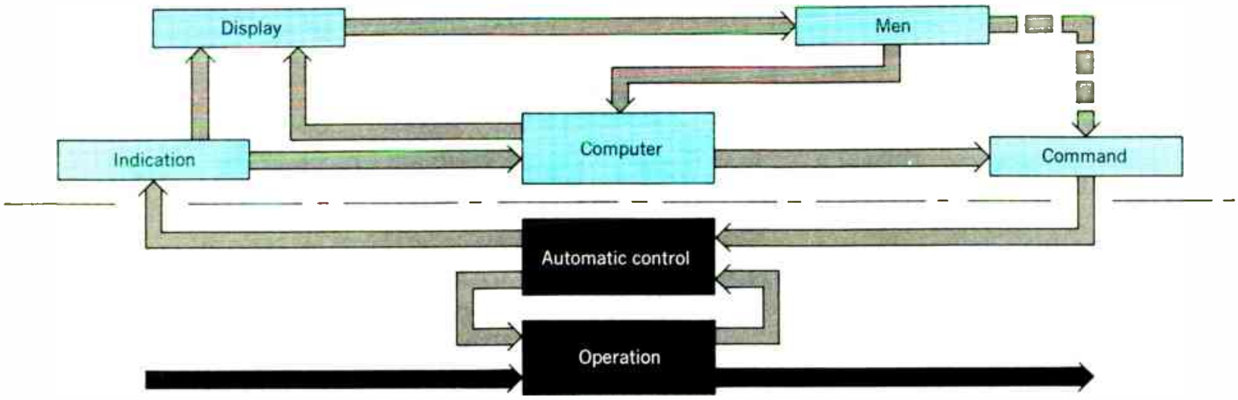


1. The train's number.
2. The number of the locomotive.
3. The train's position (the number of the last-reported 100-meter-long section of track conductor).
4. The track number (near a terminal or interlocking zone).

The control point, however, must be continuously informed of the condition at variable signals (drawbridge crossings, grade crossings, etc.). If this information cannot be obtained directly from the relays at the signal box, then these data must be collected from backup devices via channel 3 in Fig. 2.

An information range of about twice the emergency braking distance is considered both safe and practicable; in the case of passenger express trains, this would be a distance of 5 km.

To achieve true CATC, under the DB's criteria, the supervisory and safety system must be completely "cybernetic" and function automatically, even if a train's operator is retained in the cab. Furthermore, the supervisory system must enable fixed-block wayside signaling to be replaced by "moving blocks" to provide the highest degree of centralized supervision required by CATC. (Up to the present time, mainline trains on the DB have been operated under CATC on a test basis; the system has been—and is being—installed, however, on various U-Bahn and S-Bahn systems that either exist or are being built.)



[5] **Organizational interface of man–machine system philosophy that is in effect on DB mainline operations.**

(For further details on CATC in the Munich area, see pp. 68–69 of this author's article, "Progress in Rail Transportation," in the January 1974 issue of *IEEE Spectrum*.)

EDP for integrated transport control

The concept of integrated transport control on the DB is—to a large extent—predicated upon existing automatic installations such as all-relay interlockings, and automatic signal towers ("cabins" as they are generally referred to in Europe) used as data centers, in which electronic data processors are rapidly becoming the focal features. In the DB's future, EDPs loom large in the exchange of information for district operational control.

Because there are three basic tasks of a comprehensive rail transportation system, three interconnected computer subsystems were introduced to form a complete cybernetic system capable of

1. *Data acquisition* in automating the task of mercantile service (essentially the conveyance of freight).
2. *Operational control of a freight marshaling yard* by means of prepared data and software.
3. *Control under EDP, of mainline train runs* as an operational task that is directly connected to the remote-control (LBZ) equipment.

Because of space limitation, we shall only discuss the last-mentioned function—train control of mainlines. In order to include EDP on mainline rights-of-way, "remote ordering" had to be developed, in which the operating personnel in the signal tower cabs (although still responsible for the safe handling of train movements), need not announce the passage of trains because indication elements—tripped by a passing train—transmit the pertinent information by means of a data-transmission system to the computer. The amount of such information for control of a train's run is not inordinately high; thus, a limited number of indication elements per wayside station is sufficient.

The computer also gives orders, by means of teletypewriters to the operating personnel for automatic record keeping for train-schedule evaluation. Simple operational situations that are clearly defined are handled by EDP alone. But more complex conflict situations must be solved by human judgment, based upon their machine presentation. In other words, the scheduling, analysis, and normal running functions are left to machine control, but the *decisions* in emergency situations are still the human operators' responsibility. Figure 4 is an example of a situation, shown on a CRT display, in which train no. 5886 is being overtaken in "block NBG" by train no. 184. Therefore train no. 184 is delayed so that a platform occupancy conflict with train no. 4018 in "block LBU" is avoided. However, the delay to no. 184 is not so long as to delay the following train no. 174. The data, presented by the visual display, was evaluated by the operator for his just-mentioned decision.

The organization of the man–machine interfacing is shown in the Fig. 5 diagram.

The proof is in the riding

As a matter of historical record, West German rail systems (the DB, S-Bahnen, and U-Bahnen) rate among the safest in the world. And, it is a "point of honor" to German railroaders that "all trains must run on time"—barring some hellish winter snow avalanche or spring-flood high waters!

German ingenuity remains in the forefront in advancing state-of-the-art technology in CATC, communications, and passenger safety. With an ongoing mainline electrification program, the future looks bright for high-speed rail travel in the Federal Republic.

Of things to come

Our next adventure in European railroading will be in Switzerland—that small, Alpine country that is very big on all-electric railways, and features some of the Continent's most modern and sophisticated traction equipment and electronic "command and control" systems. So, we have a date—with Zurich as the point of departure—in a forthcoming issue of *IEEE Spectrum*. All aboard!

EEs' salaries: on the inflationary rack

With employees' real disposable income plummeting, some corporations are reevaluating traditional wage policies

It is true, as a relatively stable salary levelled employee, the engineer in an inflationary spiral will not fare as well as the hourly or union worker. So, relatively, engineers lose in net income during this period.

Harold S. Goldberg
President
Data Precision Corp.
Wakefield. Mass.

This potentially explosive statement—to be published in the forthcoming *IEEE Manpower Report 1974* entitled, "Your Job in E/E Engineering"—might have attracted little attention as recently as a year ago. But as U.S. engineers and their fellow citizens know, the Government's cost-of-living index has been rising at an unprecedented rate. And outside of Washington, few economists see quick relief in sight.

In itself, an inflationary spiral like the current one may seem of little concern to the worker. If his wage package is similarly inflationary, he may feel he is losing nothing. But for at least 15 months, the cost of living in the U.S. has outpaced the average increase in the U.S. worker's wage.

The beginning of what has become an ominous trend can be traced to the second quarter of last year. From April through June, 1973, the real disposable income of the average U.S. worker, as measured by the Survey Research Center of the University of Michigan, Ann Arbor, declined by 2.8 percent from the preceding quarter. The third and fourth quarters of 1973 saw continued declines of 0.1 and 0.6 percent respectively. But the crusher came during the first quarter of this year: from January through March 1974, the average U.S. worker's real disposable income declined at a seasonally adjusted annual rate of 8.7 percent! Against this backdrop, new corporate wage policies have begun to emerge and *Spectrum*, in an attempt to keep its readers informed on a subject vital to their livelihood, has surveyed companies both within and outside the electrical/electronics industry.

Wage hikes in the general economy

Beginning with the May 1, 1974, dropping of wage and price controls, many corporations, particularly those that have prospered recently, began to review the plight of their employees—both exempt (salaried) and nonexempt (hourly). The fruits of these reviews have begun to crop up in industries throughout the U.S. economy. Banks and oil companies have been among the first to break with traditional pay practices and hike employee salaries to counter the continued decline in real disposable income. The methods employed have varied. The Mobil Oil Corporation, for example, having recently reported record earnings, offered all exempt U.S. employees an extraordinary bonus amounting to one month's salary. The Chase Manhattan and First National City Banks instituted, effective May 1, an across-the-board increase of 7 percent for all salaried workers. And within two weeks, Allied Chemical hiked its salaried employees' wages 4 percent across-the-board.

But what of engineers?

A mixed bag for EEs

As Harold Goldberg notes in the prefatory quotation, the engineer, in a time of inflationary spiral, may be more vulnerable than the hourly worker. Union workers, for example, in many cases have managed to write "escalator clauses" into their contracts. Such provisions provide the union employee with a measure of security in inflationary periods by ensuring wage increases that parallel unforeseen rises in the cost of living.

However, few engineers are unionized. As exempt employees, they must depend mostly on the responsiveness of management for assistance in the face of inflation. A highly competitive talent market can help by forcing companies to raise salaries to retain their skilled professionals, but where supply is greater than demand and when living costs fluctuate unpredictably, corporate goodwill toward the exempt employee becomes an unavoidably vital factor.

Based on *Spectrum*'s interviews, such goodwill exists in many, if not all, sectors of the electrical/electronics industry. No corporate representative with whom *Spectrum* spoke denied the seriousness of inflationary conditions. But while many companies are continuing to rely on competitive pressures to inflate employee salaries appropriately, some have, like the oil companies and banks, decided to take unprecedented measures.

In the electronics field, perhaps the quickest on its feet was Hewlett-Packard. Speaking of the May 1 cutoff of controls, a company spokesman reported that "eight hours and one minute after the lid blew off, Hewlett-Packard instituted a 5-percent across-the-board raise for all employees irrespective of merit considerations."

By May 26, Tektronix had followed suit, instituting a similar 5-percent hike across-the-board for all U.S. employees. Explained company president Earl Wantland, "We have always had a strong orientation toward

Ellis Rubinstein Associate Editor