IEEE spectrum

features

The Bold New Breed in Digital Thermometry

Digitec offers a new "breed" of digital thermometers to serve the scientific community. Many years as leaders in digital thermometry assures you of the benefits of our "know-how." Accessory probes and data acquisition components are also offered to fulfill your complete requirements.

SOLID STATE THERMISTOR THERMOMETER
MODELS 251A/551A-1 & 2
°F and °C Models •
Thermilinear' Probes Available •
Systems Compatible •
from $715

SOLID STATE PLATINUM RESISTANCE THERMOMETER
MODELS 251A/551-3 & 4
- Wide Ranges, from −150°F to +750°F, and −100°C to +400°C
- High Accuracy, ±.4°C, ±.7°F
- Systems Compatible

$810

DIGITAL THERMISTOR THERMOMETER
SERIES 1500
Accuracies to ±.05°C •
Low Cost Probes Available •
4 Selectable Input Channels •
°F and °C Models •
$395

Measure the others by

DIGITEC®

by UNITED SYSTEMS CORPORATION

918 Woodley Road • Dayton, Ohio 45403 • (513) 254-6251

Circle No. 2 on Reader Service Card

ISA Silver Jubilee International SHOW
Visit Booth 217

Request new catalog
Representatives throughout the world

the cover

Computer synthesis of speech has been the subject of considerable recent study, as described in the article beginning on page 22. This month's cover depicts multiple images of a typical vocal-tract model used for synthesizing basic speech sounds, or "phonemes"

departments

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 can be made effective with any particular issue if received by IEEE Headquarters by the 1st of the preceding month. Annual subscription: IEEE members, first subscription $3.00 included in dues. Single copies $1.50. 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.

BPA

OTHER IEEE PUBLICATIONS: IEEE also publishes the PROCEEDINGS OF THE IEEE, the IEEE STUDENT JOURNAL, and more than 30 Transactions for IEEE 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 page 17 of the January 1970 issue. When in doubt, send the manuscript to E. K. Gannett, Editorial 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*.

Forum

Readers are invited to comment in this department on material previously published in IEEE SPECTRUM; on the policies and operations of the IEEE; and on technical, economic, or social matters of interest to the electrical and electronics engineering profession.

Air-traffic control

I was extremely impressed with Gordon Friedlander's most competent article in the June issue of *IEEE Spectrum*. He must have done a great deal of research in order to get so many of the facts on this complex system straight and explain them so well. There is only one part wherein he departs slightly from fact.

In Fig. 13, the block diagram of the fundamental ARTS III system indicates that the primary radar is processed along with the beacon data. In the present contract, only beacon processing is achieved, and no radar processing is performed. Consequently, the computer is aware only of beacon targets. In the very near future, radar processing will be added to the ARTS III system, which is, of course, a modularly expandable system.

I am looking forward to the future articles in the series, and I hope that they will include the activities of the Air Traffic Control Advisory Committee (Alexander Committee).

James C. Nelson, Manager
Advanced Transportation Systems
Univac FSD
St. Paul, Minn.

Inventors' rights

I was encouraged to read in the July 1970 Forum that Congressman John E. Moss of California has introduced H.R. 15512. From the information available, it seems that this bill corresponds to the spirit of an article, "The Legal Rights of Employed Inventors" by Theodore H. Lassagne, in the September 1965 issue of the *American Bar Association Journal*.

The main concluding points of Mr. Lassagne's article are as follows:

"A trend is being evidenced in every technologically advanced country in the world today: a trend toward requiring that employed inventors be compensated in proportion to the economic productiveness of their inventions, instead of by payments, whether salary or bonuses, which are not based upon any consideration of such productiveness." Incentives for employed inventors are very different today from what the framers of the Constitution had in mind. Ninety-five percent of professional engineers and scientists are employees of industrial enterprises or government. There is a question whether employment contracts promote or impede the constitutionally declared public policy of promoting the progress of science and the useful arts.

Inventors have security of regular paychecks. In fact, they are better off if they avoid proposing anything untried. A resounding failure might cost a man his job but mediocrity never will.

Big government contends with big business for title to patent rights on government-funded projects on the naive premise that dollars rather than brains create inventions.

Because of the present policy, the U.S. ranks seventh in the number of new inventions. Switzerland and West Germany have the foremost policy and lead in the number of new inventions. West Germany has a grading point system in which circumstances are evaluated toward establishing special compensation.

It would seem that when an employer owns anything that an employee might dream up on his own time and which is inspired from his previous life, something that is outside his employer's business interests but which that employer might choose to include in plans for the future—then this smacks of mental and economic servitude. Also, other situations would seem to fit as well in the absence of some proportional compensation for significant contributions.

Motivation has been left to the individual conscience and, as per Mr. Lassagne's article, mediocrity is promoted. One can only surmise that a present drought for new innovations must exist since it would seem prudent for an employee to save any significant new thing to prove himself in future unanticipated employment.

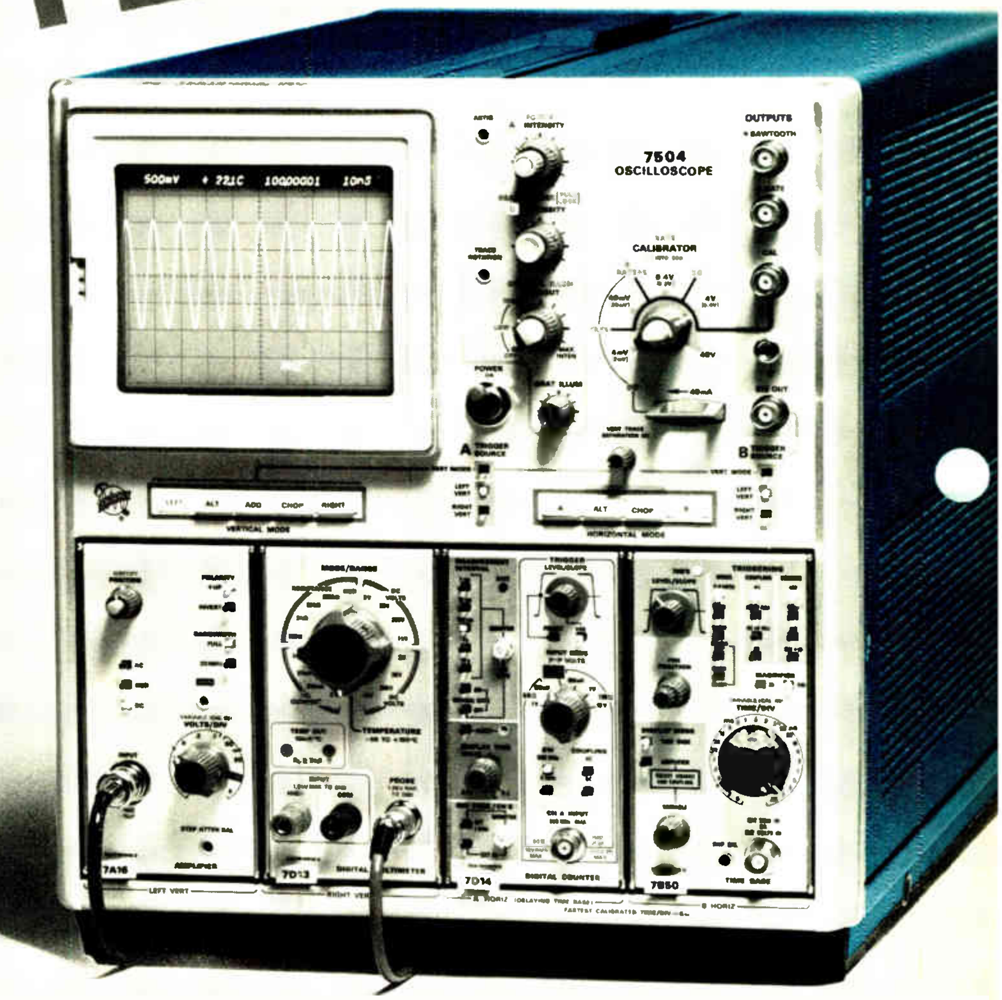
Thomas B. Albright, Santa Ana, Calif.

Tape-to-film conversion

The Focal Points article in July *Spectrum*, which announced the development by 3M of a new videotape-to-film conversion process, was interesting, but misleading. I do not question the development of a new process by 3M, but I do take exception to the

INTRODUCING THE . . .

INTEGRATED TEST SYSTEM



Measures:

- ***Frequency***
- ***Temperature***
- ***Time***
- ***Voltage***
- ***Resistance***
- ***Current***

With the introduction of two new digital plug-ins, a 500-MHz Direct Counter (no prescaling) and a versatile Digital Multimeter (center plug-ins above), it is even more apparent that the Tektronix 7000 Series was intended to be more than just another oscilloscope . . . ITS much more!

There are 5 mainframes and 17 plug-ins in the 7000 Series . . . with more to come.

Consult your August Catalog Supplement for complete information. If you do not have a Supplement, check the inquiry card, we will send you one promptly.

Circle No. 4 on Reader Service Card

TEKTRONIX®

committed to technical excellence

tone of the article, to the implication that a new, not heretofore available, technique has been developed.

LOGOS engineers in our Teleproduction Center developed and assembled a tape-to-film transfer device in 1963. Our technique, Trinescope, uses three high-resolution subscreen kinescope tubes, with adjustable color fidelity available on each tube. Our best results are obtained with P11, P16, and P24 phosphors to produce a white-level color temperature of 3200 degrees. The 600-footlambert brightness available from each kinescope tube permits the use of very low ASA rated film.

Our system will handle any NTSC signal source, either live or from 1-inch helical or 2-inch quadrature scanned tape. The NTSC signals are conventionally decoded into the basic colors, presented to the kinescopes, triplexed through dichroic mirrors into a composite color image, and focused into either a 16-mm or 35-mm camera. Synchronization is maintained to convert 30 video frames to 24 film frames.

We are proud of our technique and of the product we produce. Many satisfied clients will testify to the degree of our expertise. For example, our Education Systems Division recently completed a U.S. Navy contract in which training films were converted to video tape, edited and revised using video techniques, and converted back to film without loss of quality. You do a disservice to those of us in the communications profession when you announce a "new" or "better" technique without qualifying your statements. I am sure the 3M system has merit and will perform as stated but there are other systems, ours for example, that will perform equally well.

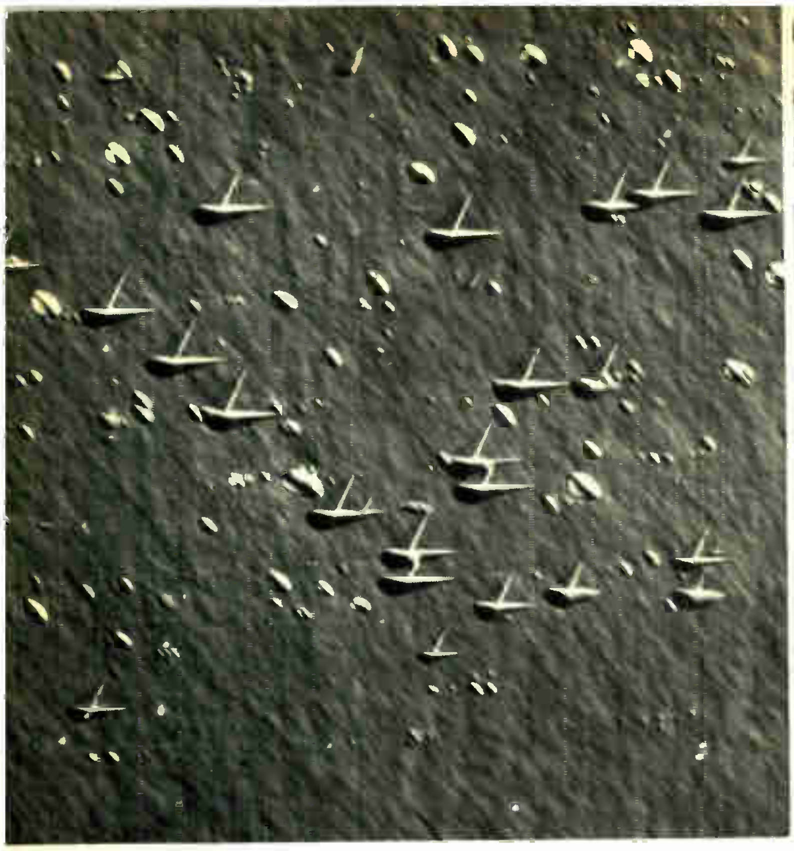
W. D. Clendenon
LOGOS, Arlington, Va.

A remedy to the population explosion?

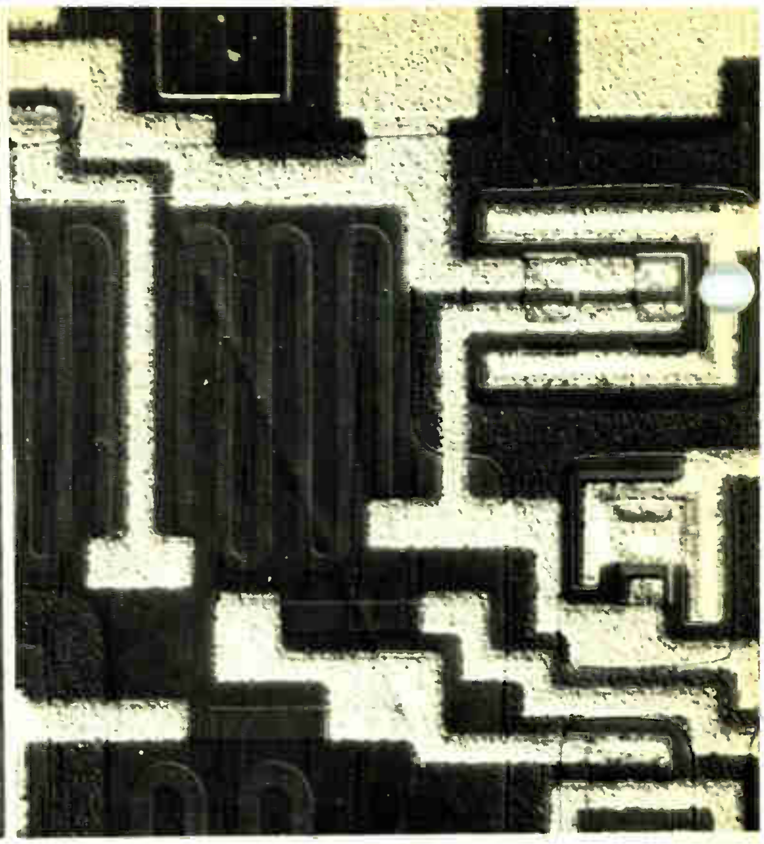
In the process of editing my article, "The Exponential Crisis," Spectrum (August) miraculously succeeded in reducing the population growth rate from 0.1 dB/yr to 0.1 percent/decade (a very neat trick!) and in awarding me a B.S. degree I do not hold.

F. N. Eddy, Weston, Mass.

The growth rate error resulted when we misread our log tables. In the original manuscript, the growth rate was specified as 1 dB/decade. This should have translated as 12 percent/decade—still low in relation to other available figures. Before Mr. Eddy's biography was amended, it read: "F. N. Eddy (M) attended Harvard College (applied physics, 1954) and did graduate work at Boston University."—*Ed.*



Only with Epi-Nomarski Interference Contrast do faults such as these in unetched silicon crystals stand out. (200x)



Diagonal fault seen here was detected only with Nomarski Interference Contrast.

From slice to circuit, Zeiss optics keep small troubles from becoming big problems.

At every stage in the production of integrated circuits, from slice to chip, the application of optical inspection equipment can help you find and cure those small faults that can result in big, expensive problems. Zeiss has a number of instruments that are proven money-savers in the microelectronics industry. Among them are:

1. Epi-Nomarski Interference Contrast Equipment for all Zeiss Standard and Research Microscopes — This microscope technique has been termed by one of the industry's leading microscopists as "the greatest invention in optics for the semiconductor industry." With Epi-Nomarski, particles, holes, stacking faults, deviations from flatness, and other surface irregularities can be made to stand out. Even an untrained observer can see irregularities by adjusting the optics so that all differences in surface and materials are displayed in brilliant colors. You can use this technique for evaluating every step of the manufacturing process—particularly the etching, metal deposition, and photoresist operations — and in quality control of the finished product.

The new Zeiss KK-08 microscope is now a best buy. Developed specifically for the electronics industry, it is well-worth looking at, both for the production line and the laboratory.

2. Interference Flatness Tester—A unique instrument that by strictly optical, non-contact means checks the flatness of silicon slices and the glass surfaces of photographic masks down to 1/10 of an interference fringe (300Å).

3. Interference Microscope — For measuring thickness of evaporated plating films and determining junction depth. A non-destructive measuring tool with accuracy to 300Å.

4. Light-Section Microscope — For thick-film measurement (1-400 microns). Completely non-destructive, this instrument enables continuous measurement of the uniformity of wet films prior to firing. Measurements of great precision can be made in 10-12 seconds.

5. Micro-Hardness Tester — Used to test hardness, and uniformity of hardness, of silicon wafers and of gold or aluminum coatings. Great accuracy with loads of less than 1 gram.

Besides these inspection instruments, Zeiss is a world leader in such production equipment as Mask-Alignment Microscopes, Stereo Microscopes, and Reduction Objectives.

For information on any Zeiss instrument for electronics, write Carl Zeiss, Inc., 444 Fifth Avenue, New York, N. Y. 10018.

Nationwide Service

ZEISS

THE GREAT NAME IN OPTICS

CARL ZEISS WEST GERMANY

ATLANTA, BOSTON, CHICAGO, COLUMBUS, DENVER, HOUSTON, LOS ANGELES, PHILADELPHIA, ST. LOUIS, SAN FRANCISCO, SEATTLE, WASHINGTON, D. C.

Circle No. 6 on Reader Service Card

Spectral lines

What more can IEEE do for young engineers? Electrical engineers today are feeling their responsibilities to society and mankind more strongly than ever before. Expanded IEEE activity to help encourage further development of these feelings has been suggested. In response to some of these evolving concerns, the IEEE Board of Directors has appointed an *ad hoc* Committee on Professional Concerns of Young Engineers. This committee, which includes ten young engineers, has been asked to seek ways in which the Institute more vigorously can serve the technical, societal, and career objectives of electrical engineering students and young professionals. The purpose of these paragraphs is to solicit correspondence on these matters from interested members and friends of the Institute.

In the realm of technical needs, the committee is examining means by which some of the results published in IEEE journals can be made more meaningful and accessible to the majority of members. Also of interest are ways by which more Student Branches could be encouraged to operate at the high level of the most energetic Groups, and in which technical activities of local IEEE Sections could be made more useful and attractive to young engineers. A more active role for IEEE in sponsoring and/or evaluating programs of continuing education has been suggested. Finally, some members feel that more vigorous action by the IEEE could magnify the mutual technical advantages to members of the Institute's unique transnational character.

Discussion in the *ad hoc* committee has elicited suggestions for IEEE activity to inform students and young engineers of developing opportunities in various electrical engineering specialities, and of the potential social impact of work in fields utilizing these specialities. Possibly the Institute could take part by stimulating efforts to identify promising areas for new research and product development related to human needs. Summer task forces staffed with engineering students might be involved in such investigations.

Career needs of present and future engineers range from basic matters of equal opportunity for women and others underrepresented in present ranks of the profession, to economic matters and effective utilization of qualified engineering personnel.

In examining new areas for possible IEEE action, the *ad hoc* committee is anxious to provide well-formulated plans for specific programs or other activity, within the grasp of the IEEE, accompanying every specific recommendation it may make. In other words, we are not going to be satisfied simply with declarations of principles or codes of ethics. In this spirit, we invite communication from all interested persons. You may write to:

Committee on Professional Concerns of Young Engineers
Ralph W. Wyndrum, Jr., Secretary
Room 1C-425
Bell Telephone Laboratories, Inc.
Holmdel, N.J. 07733

Or, if you prefer, contact one or more of the other committee members:

James C. Castle
General Electric Company
Bethesda, Md. 20014

James B. Farison
University of Toledo
Toledo, Ohio 43606

Gerald J. Herskowitz
Stevens Institute of Technology
Hoboken, N.J. 07030

David A. Hodges
University of California
Berkeley, Calif. 94720

Robert W. Johnston
Northern Electric Company Ltd.
London, Ont., Canada

Robert E. Larson
Systems Control Inc.
Palo Alto, Calif. 94306

Fred A. Lindholm
University of Florida
Gainesville, Fla. 32601

James R. Rowland
Georgia Institute of Technology
225 North Avenue
Atlanta, Ga. 30332

Harvey Rubin
Bell Telephone Laboratories, Inc.
Holmdel, N.J. 07733

Steve E. Schwarz
University of California
Berkeley, Calif. 94720

F. Karl Willenbrock
State University of New York at Buffalo
Buffalo, N.Y. 14214

David A. Hodges

Synthetic voices for computers

The concept of giving voices to computers has been the subject of considerable study in recent years. Formant synthesis and text synthesis represent two approaches to the problem

J. L. Flanagan, C. H. Coker, L. R. Rabiner, R. W. Schafer, N. Umeda
Bell Telephone Laboratories, Inc.

The two methods described for giving voices to computers recognize the importance of economical storage of speech information and extensive vocabularies, and consequently are based on principles of speech synthesis. The first, formant synthesis, generates connected speech from low-bit-rate representations of spoken words. The second, text synthesis, produces connected speech solely from printed English text. For both methods the machine must contain stored knowledge of fundamental rules of language and acoustic constraints of human speech. Formant synthesis from an input information rate of about 1000 bits per second is demonstrated, as is text synthesis from a rate of about 75 bits per second. To give the reader an opportunity to evaluate some of the results described, a sample recording is available; see Appendix A for details.

Voice output from machines

If computers could speak their answers—as well as print them and display them graphically—digital machines could be applied effectively to an expanded range of problems. Typical uses would include automatic information services, computer-based instruction, reading machines for the blind, and spoken status reports from aircraft and space-vehicle systems. Vast amounts of information would be only as far away as the closest push-button telephone. For example, a physician sitting in his office might need information on some obscure disease. It would be convenient if he could dial a computer, key in a reference number, and hear a page or two "read" to him out of a medical encyclopedia. A prospective air traveler might dial a computer, enter destination and desired departure time, and have the computer make combinational searches through timetables and report verbally the convenient connecting flights.

For such applications, the computer must have a large and flexible vocabulary. It therefore must store sizable quantities of speech information, and it must have the information in a form amenable to producing a great variety of messages.* Speech generated by the machine must be as intelligible as natural speech. It need not, however, sound like any particular human and might even be permitted to have a "machine accent." Toward these objectives we have explored two methods of obtaining voice response from a computer, both of which at present appear feasible and attractive.

The first method is called *formant synthesis*. It depends upon an initial, automatic analysis of human speech to produce a synthetic vocabulary. Word libraries are analyzed and stored in terms of formant frequencies. Formants are the natural resonances of the vocal tract, and they take on different frequency values as the vocal tract changes its shape during talking. Typically for nonnasal, voiced sounds three such resonances occur in the frequency range 0 to 3 kHz. The word-length formant data are accessed upon program demand, and are concatenated to form complete formant functions for an utterance. The formant functions have to be interpolated naturally across word boundaries, and voice pitch and word duration have to be calculated according to linguistic rules. Economy in storage derives from the fact that the formant and excitation parameters change relatively slowly and can be specified by fewer binary numbers (bits) per second than can, for example, the speech waveform.

The second method is *synthesis from printed text*—that is, speech synthesis literally from the printed page. In this method, no element of human speech is involved.

* Voice response is already being used in a number of limited-vocabulary applications. Present methods mainly employ prerecorded messages, which are stored and accessed on demand. The limitations of storage and vocabulary size are factors that synthetic speech aims to overcome.

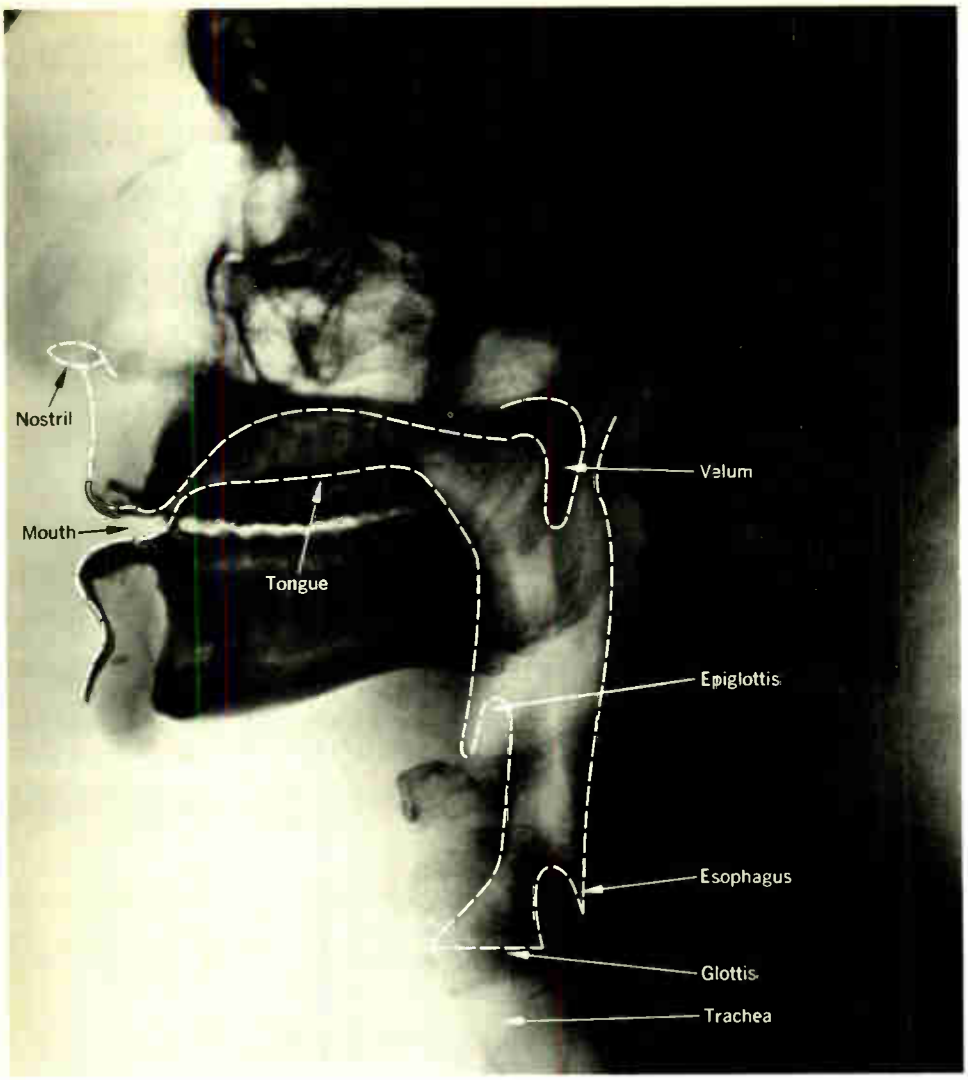


FIGURE 1. Sagittal-plane X ray of a man's vocal tract.

The method depends solely upon the machine knowing, as best it can, the rules and constraints of human speech production and language. An automatic syntax analysis is first made of the text to be spoken. Sound pitch and duration are computed from stored rules about English language. A sequence of vocal tract shapes is then calculated to correspond to the message. Economy of storage results from being able to represent the alphabetic text characters by very few bits per second.

The complexities and ambitions of the two methods can be put into focus at the outset. This is conveniently done by comparing their typical data rates to the data rate for a digitized speech waveform.

I. Information rates for different forms of digital speech data

Stored Data	Bit Rate, b/s	Duration of Speech per 10^6 Bits of Storage
Digitized waveform (PCM)	50 000	20 seconds
Formant data	1 000	17 minutes
Printed text	75	4 hours

Parametric (reduced-data) storage of speech information. A comparison of information rates corresponding to digitized waveform, formant data, and printed text encodings of speech information is given in Table I. The duration of speech that can be stored in 10^6 bits is also shown for the three cases.

It can be seen that the waveform, without further coding, requires around 50 000 b/s (bits per second); that is, the signal is typically sampled at the Nyquist rate and quantized to about 7 bits on a logarithmic scale. Storage capacity of 10^6 bits can therefore accommodate only about 20 seconds of speech in this form. Further, this signal cannot be satisfactorily chopped up and used to fabricate messages different from that originally spoken.

Formant data, on the other hand, require an information rate around 1000 b/s—a reduction of 50:1 over the waveform. (The subsequent discussion will reveal the nature of the compression.) In this case, a store of 10^6 bits can accommodate about 17 minutes of continuous speech. Equally important as the saving in storage is the

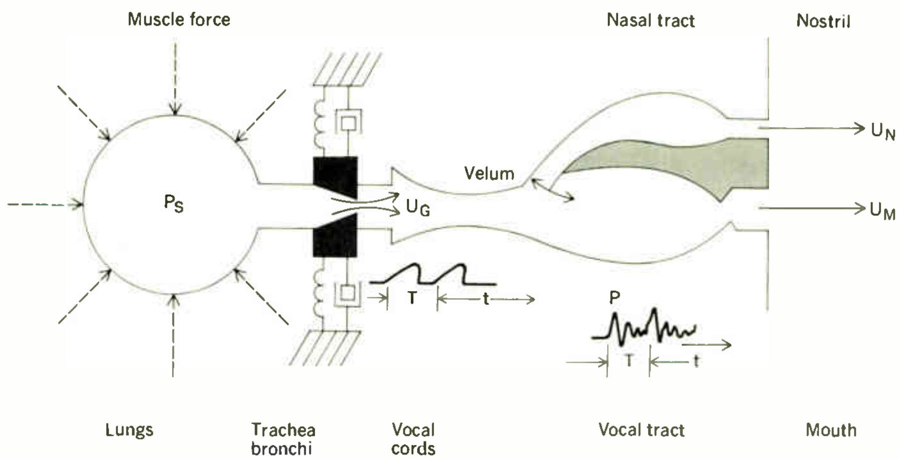
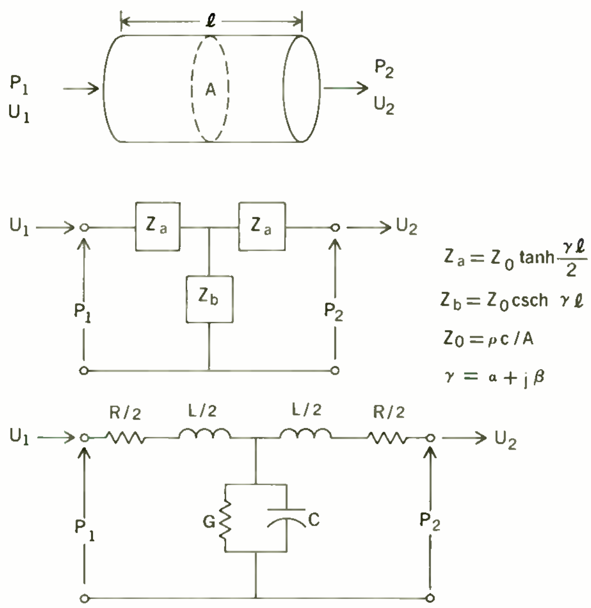


FIGURE 2. Schematic diagram of the human vocal system.

FIGURE 3. Equivalent network for plane-wave propagation in a right circular pipe.



fact that formant-coded words and phrases can be connected together in a relatively natural way to form a variety of messages. This possibility makes formant synthesis attractive for a range of voice services, such as inventory reporting, automatic information operators, and machine instruction.

The ultimate in data reduction and flexibility is printed text, the third entry of Table I. Printed text converted to conversational speech corresponds to an information rate around 75 b/s. About 5 bits are required to specify an alphabetic character and punctuation; words on the average have about 5 characters; and conversational-rate speech corresponds to about 3 words per second. A storage of 10^6 bits therefore provides about four hours of speech. If the available storage were, say, 10 million 36-bit words (not uncommon with larger machines), then about two months of continuous voice output could be provided. In this form, then, the storage of encyclopedic material is quite feasible. The synthesis problem, however, is relatively complex, and the computations that are used to obtain the synthetic output are more consuming of machine time.

In the following sections we explain these two methods and indicate their stages of development. The acoustic and linguistic fundamentals that underlie both synthesis methods are the same. A convenient point of departure is a summary of these relations.

The speech signal

Acoustics of speech production. The major parts of a man's vocal apparatus are shown in the sagittal-plane X ray of Fig. 1. The vocal tract proper is a nonuniform acoustic tube about 17 cm in length. It is terminated at one end by the vocal cords (or by the opening between them, the glottis) and at the other by the lips. The cross-sectional area of the tract is determined by placement of the lips, jaw, tongue, and velum, and can vary from zero (complete closure) to about 20 cm^2.

An ancillary cavity, the nasal tract, can be coupled to the vocal tract by the trapdoor action of the velum. The nasal tract begins at the velum and terminates at the nostrils. In man the cavity is about 12 cm long and has a volume of about 60 cm^3. In nonnasal sound the velum seals off the nasal cavity and no sound is radiated from the nostrils.

Sound can be generated in the vocal system in three ways. Voiced sounds are produced by elevating the air pressure in the lungs, forcing a flow through the vocal-cord orifice (the glottis) and causing the cords to vibrate. The interrupted flow produces quasiperiodic, broad-spectrum pulses, which excite the vocal tract. Fricative sounds are generated by forming a constriction at some point in the tract, usually toward the mouth end, and forcing air through the constriction at a sufficiently high Reynolds number to produce turbulence. A noise source of sound pressure is thereby created. Plosive sounds result from making a complete closure, again usually toward the front, building up pressure behind the closure and abruptly releasing it. All these sources are relatively broad in spectrum. The vocal system acts as a time-varying filter to impose its spectral characteristics on the sources.

The vocal system can be schematized as shown in Fig. 2. The lungs are represented by the air reservoir at the left. The force of the rib-cage muscles raises the air in the lungs to subglottal pressure P_S. This pressure expels a flow of air with volume velocity U_G through the glottal orifice and produces a local Bernoulli pressure. The vocal cords are represented as a mechanical oscillator composed of a mass, spring, and viscous damping. The cord oscillator is actuated by a function of the subglottal and Bernoulli pressures. The sketched waveform shows the form of the U_G flow during voiced sounds. The vocal tract and nasal tract are shown as tubes whose cross-sectional areas change with distance. The acoustic volume velocities at the mouth and nostrils are U_M and U_N, respectively. The sound pressure P, in front of the mouth, is approximately a linear superposition of the time derivatives $\dot{U}_M$ and $\dot{U}_N$.

A factor of primary interest is the transmission characteristic of the vocal system. The tract length is comparable to a wavelength at all speech frequencies of interest. Its cross dimensions, however, are relatively small. If the tract is considered hardwalled, lossless, and with no side branches, a numerical solution of a one-dimensional steady-state wave equation with nonconstant coefficients (Webster's horn equation) yields the undamped eigenfrequencies (or resonances) of the system.[1,2] On the other hand, a bilateral transmission-line equivalent, useful for digital simulation of actual pressure and velocity relations, including those of the vocal cords, nasal tract, and the subglottal system, can also be obtained.[1]

Consider the variable-area pipe to be composed of elemental right-circular pieces, one of which is shown in Fig. 3. Sound pressure and volume velocity at the two ends are represented by P_1, U_1 and P_2, U_2, respectively. For plane-wave propagation, the pipe element of length l has an equivalent T-section in which the impedance elements Z_a and Z_b are hyperbolic functions of the complex acoustic propagation constant γ. (The acoustic network has exactly the same form as that for a uniform electrical line.) The characteristic impedance Z_0 is the product of air density ρ and sound velocity c divided by the cross-sectional area A. For a given quantal length, then, the impedance elements are determined solely by the cross-sectional area. The first terms in the series expansions of the hyperbolic functions for Z_a and Z_b give the acoustic elements R, L, G, C of Fig. 3. The loss R arises from viscous loss at the walls of the pipe; the inertance L is due to the mass of air in the elemental cylinder; the loss G results from the heat conduction at the walls; and the capacity C arises from the compressibility of air in the volume Al.*

The schematic system of Fig. 2 can therefore be decomposed into elemental right-circular pieces and represented for computation and simulation by the bilateral network of Fig. 4. Network elements correspond to those parts shown in Fig. 2. Consider voltage analogous to pressure and current analogous to volume velocity. The lung volume is represented by a capacity and loss whose sizes depend upon the state of lung inflation. The lungs are connected to the vocal cords by the trachea and bronchi tubes, represented in Fig. 4 as a single T-section. The impedance of the vocal cords Z_G is both time-varying and dependent upon the glottal volume velocity U_G.[3] The vocal tract is approximated as a cascade of T-sections in which the element impedances are determined by the cross-sectional areas $A_1 \cdots A_N$. The line is terminated in a radiation load Z_M at the mouth, which is taken as the radiation impedance of a circular piston in a plane baffle. U_M is the mouth current and, for simulation of dc quantities, the battery P_A represents atmospheric pressure.

The nasal tract is coupled by the variable velar impedance Z_V. The nasal tract is fixed in shape, and the nostril current U_N flows through the radiation impedance Z_N.

This formulation of the vocal system actually can simulate respiration and phonation. The glottis is opened (Z_G is reduced), the rib-cage muscles enlarge the lung capacitor (volume), and the atmospheric pressure forces a charge of air through the tract and onto the capacitor. The glottis is then clenched and increased in impedance; the rib-cage muscles contract, raising the voltage (pressure) across the lung capacity, and force out a flow of air.

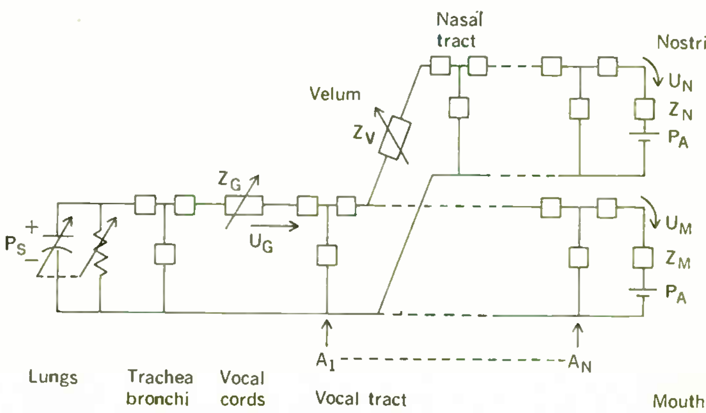
Under proper conditions, the vocal-cord oscillator is set into stable vibration, and the network is excited by periodic pulses of volume velocity. The lung pressure, cord parameters, velar coupling, and vocal-tract area all vary with time during an utterance. A difference-equation specification of the network, with these variable coefficients, permits calculation of the Nyquist samples of the output sound pressure.*

For purposes of computer synthesis, sound may be computed from a bilateral network such as Fig. 4, or the eigenfrequencies (formants) of the system may be obtained and used in a "terminal analog" synthesis.[1] The latter involves control of the transfer function of a variable network so that its unilateral transmission simulates that of the vocal tract. The nature of the unilateral transmission is simply illustrated for a vocal tract in the shape of a straight pipe in Fig. 5. The hard-walled pipe is assumed to be excited by a high-impedance volume velocity source U_G, and the mouth radiation impedance is assumed negligible for simplicity. The Fourier transform of the ratio of mouth and glottal currents, U_M/U_G, is the transmission function of interest.

* Mechanical yielding of the vocal-tract walls modifies the shunt parameters of the equivalent circuit (see Ref. 1).

* One of our computer programs is a vocal-tract synthesizer represented just this way.

FIGURE 4. Network simulation of the vocal system.



As shown, $|U_M/U_G|$ has peaks, or formants, at frequencies where the pipe is an odd quarter-wavelength—that is, at frequencies $f_n = (2n - 1)c/4l$, for $n = 1, 2, \cdots$. For a tract length of 17 cm and a sound velocity of 340 m/s, these frequencies are 500, 1500, 2500, $\cdots$ Hz. The resonances are simple and appear as single complex-conjugate poles in the transmission function. The half-power bandwidths of the resonances are conditioned by the losses in the system and are roughly constant for each formant. The phase shift of the transmission passes through π radians as each formant is traversed in frequency. The phase response is a perturbation about a line of constant slope, namely the transit delay through the tube, l/c. This transmission characteristic is the "filter" that operates on the vocal sound source. Hence the radiated sound bears these resonances.

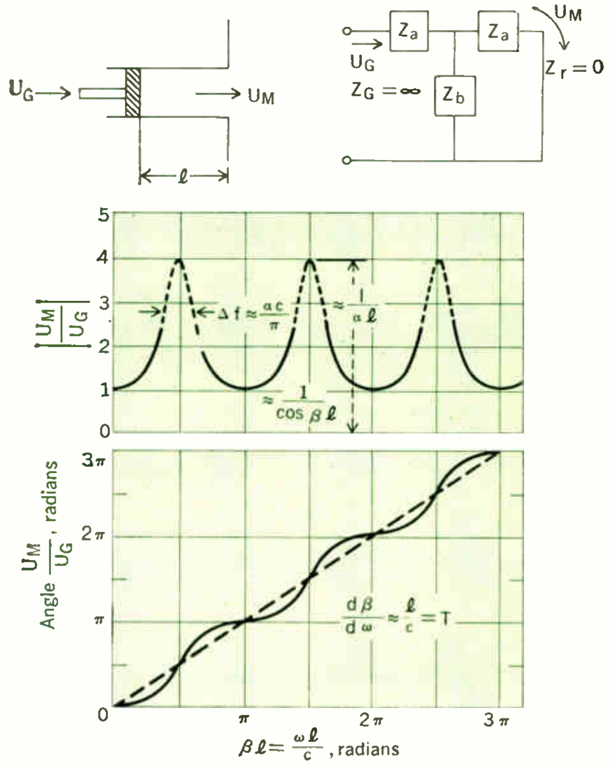
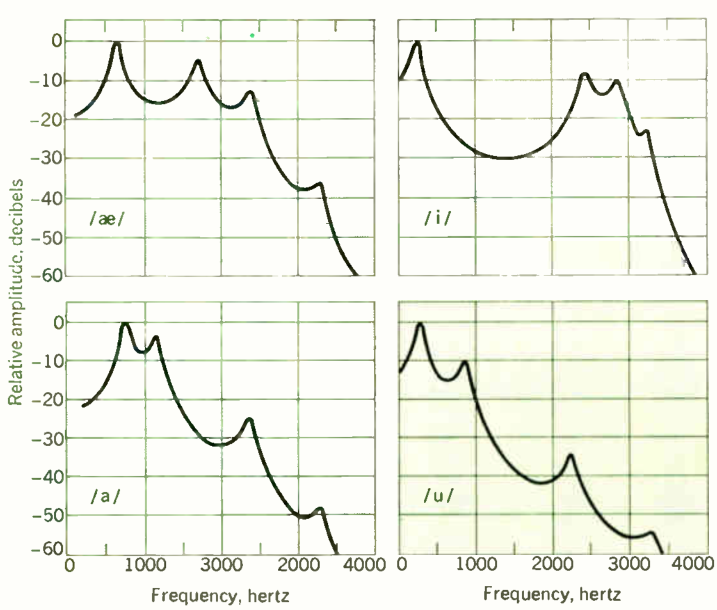


FIGURE 5. Acoustic transmission properties of a straight hardwalled pipe.

FIGURE 6. Frequency spectrums of vocal-tract transmission for four vowel sounds.



For nonnasal voiced sounds, three formants typically fall in the frequency range 0–3 kHz. Because the resonances are simple and have relatively constant bandwidths, the formant frequencies effectively specify the spectrum everywhere. For voiceless sounds, because the sound source is located forward in the tract, one resonance (pole) and one antiresonance (zero) typically describe the transmission in the frequency range 0–3 kHz.

Every shape of the vocal tract has a unique set of formant frequencies and the distinctive sounds of a language have perceptually distinctive formant positions. Idealized vocal-tract transfer functions for several vowels are shown in Fig. 6. Note, for example, the vowel /i/ (as in eat) has typically a low first formant frequency and a high second formant.* By contrast, the vowel /a/ (as in father) has a high first formant proximate to its low second formant. The overall spectral shapes of the two sounds are notably different.

In continuous speech the formant resonances move around as the vocal tract changes shape. Figure 7 shows a sound spectrogram (time–frequency–intensity plot) of a sentence in which the first three formant frequencies are traced. (Dashed lines in Fig. 7 are idealized and may not accurately plot formant transitions—especially in the context of stop and nasal consonants.) These parameters vary slowly because of the physical limitations on how quickly the vocal-tract shape can be changed. Hence they occupy only a small bandwidth. If these resonances can be determined accurately, and preferably automatically, they can be used with data about fundamental voice frequency and intensity to synthesize signals similar to natural speech.

Discrete and dynamic aspects of speech. Speech has long been viewed as a discrete process. The pronunciation key of any dictionary expresses the sounds of the language as a finite set of discrete symbols, each having a relatively invariant sound and vocal-tract shape. One might think, therefore, that a simple means could be devised to represent speech sounds by a truly small inventory of these basic speech sounds, the *phonemes*, and that messages might be composed by concatenation of these small units.

A look at a speech spectrogram, such as Fig. 7, however reveals that speech, at the acoustic level, is not particularly discrete. Spectrograms and X-ray motion pictures show that the articulations of adjacent phonemes interact and that transient movements of the vocal tract for the production of any phoneme last much longer than the average duration of the phoneme (total time divided by number of phonemes); that is, the articulatory gestures overlap and are superposed.

The transient motions of the vocal tract are perceptually important. Experiments show that much information about the identity of a consonant is carried not by the

* Symbols enclosed in slashes are those for the International Phonetic Alphabet.

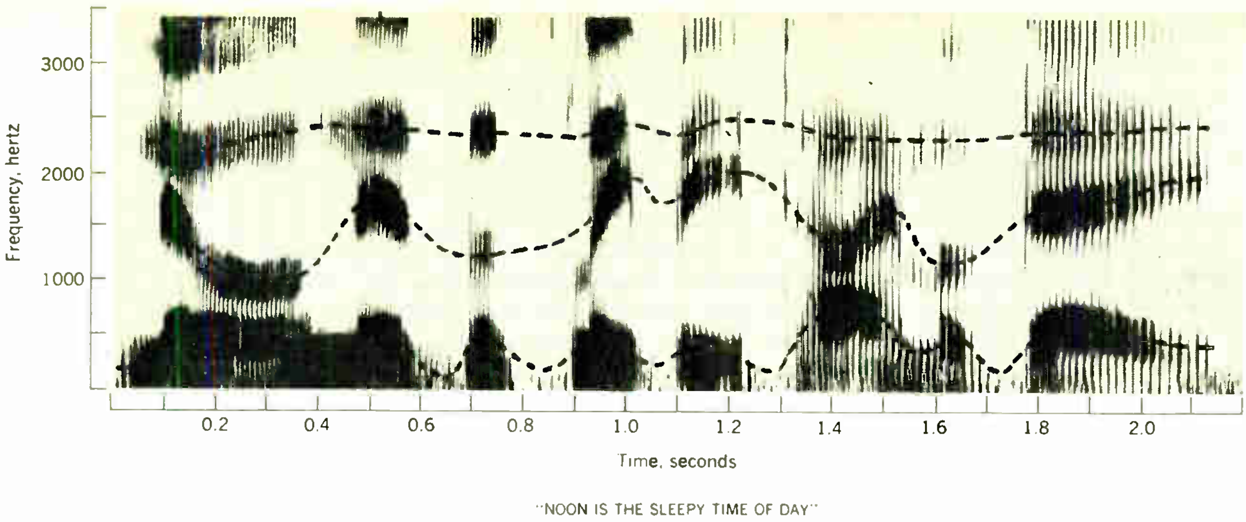


FIGURE 7. Dynamic variation of formant frequencies in connected speech.

spectral shape at the "steady-state" time of the consonants, but by its dynamic interactions with adjacent phonemes.

Speech synthesis is therefore strongly concerned with dynamics. A synthesizer must reproduce not only the characteristics of sounds when they most nearly represent the ideal of each phoneme, but also the dynamics of vocal-tract motion as its progresses from one phoneme to another.

This fact highlights a difference between speech synthesis from word or phrase storage and synthesis from more elemental speech units. If the library of speech elements is to be a small number of short units, such as phonemes, then the concatenation procedures must approach the complexity of the vocal tract itself.

Conversely, if the library of speech elements is a much larger number of longer segments of speech, such as words or phrases, then concatenations can be made at points in the message where information in transients is minimal.

The form in which speech is represented for either long-or short-element storage is somewhat flexible. For short elements, a representation as coordinates of the articulatory system seems advantageous. For word- or phrase-length storage, a formant characterization is especially appropriate.*

Speech synthesis from formant data

Automatic formant analysis of speech.[12] Because speech parameters vary slowly with time, the concept of the short-time spectrum is a basic tool in speech analysis. The Fourier transform of a short segment of the speech waveform reflects features of the excitation and formant frequencies for that segment. Figure 8 illustrates the way in which short-time spectral analysis can be employed in the estimation of speech parameters.

Figure 8(A) depicts the analysis of voiced speech. The waveform at the left is a segment of voiced speech of approximately 40-ms duration, which has been multiplied by a data window (which reduces the undesirable effects of analyzing a finite amount of data). Over such a short time interval, the speech waveform looks like a segment of a periodic waveform. The detailed time variation of the waveform during a single period is determined primarily by the vocal-tract response, whereas the fundamental period (pitch period) reflects the rate of vibration of the vocal cords.

The logarithm of the magnitude of the Fourier transform of this segment of speech is shown as the rapidly varying spectrum at the right. This function can be thought of as consisting of an additive combination of two components: a rapidly varying periodic component associated primarily with the vocal-cord excitation, and a slowly varying component primarily attributable to the vocal-tract transmission function. Therefore, the excitation and vocal-tract components are mixed and must be separated to facilitate estimation of the parameter values. The standard approach to the problem of separating a slowly varying signal and a rapidly varying signal is to employ linear filtering. One technique for achieving this filtering is through the intermediate computation of the *cepstrum*,[13,14] the inverse Fourier transform of the log-magnitude spectrum.

The cepstrum is plotted in the middle diagram of Fig. 8(A). The rapidly varying component of the log magnitude corresponds to the cepstral peak at about 8 ms (the value of the pitch period). The slowly varying component corresponds to the low-time portion of the cepstrum. Therefore, the slowly varying component can be extracted by first smoothly truncating the cepstrum values to zero above about 4 ms, and then computing the inverse transform. This yields the slowly varying curve that is superimposed on the short-time spectrum, shown at the right in Fig. 8(A). The figure of 4 ms was chosen to be representative of the lower limit of the pitch period for male speakers.

The formant frequencies correspond closely with the resonance peaks in the smoothed spectrum. Therefore, a good estimate of the formant frequencies is obtained by simply determining which peaks in the smoothed spec-

* Many people have investigated approaches to storage of speech for computer voice response. These methods have varied widely in their efficiency and flexibility; see Refs. 4–11.

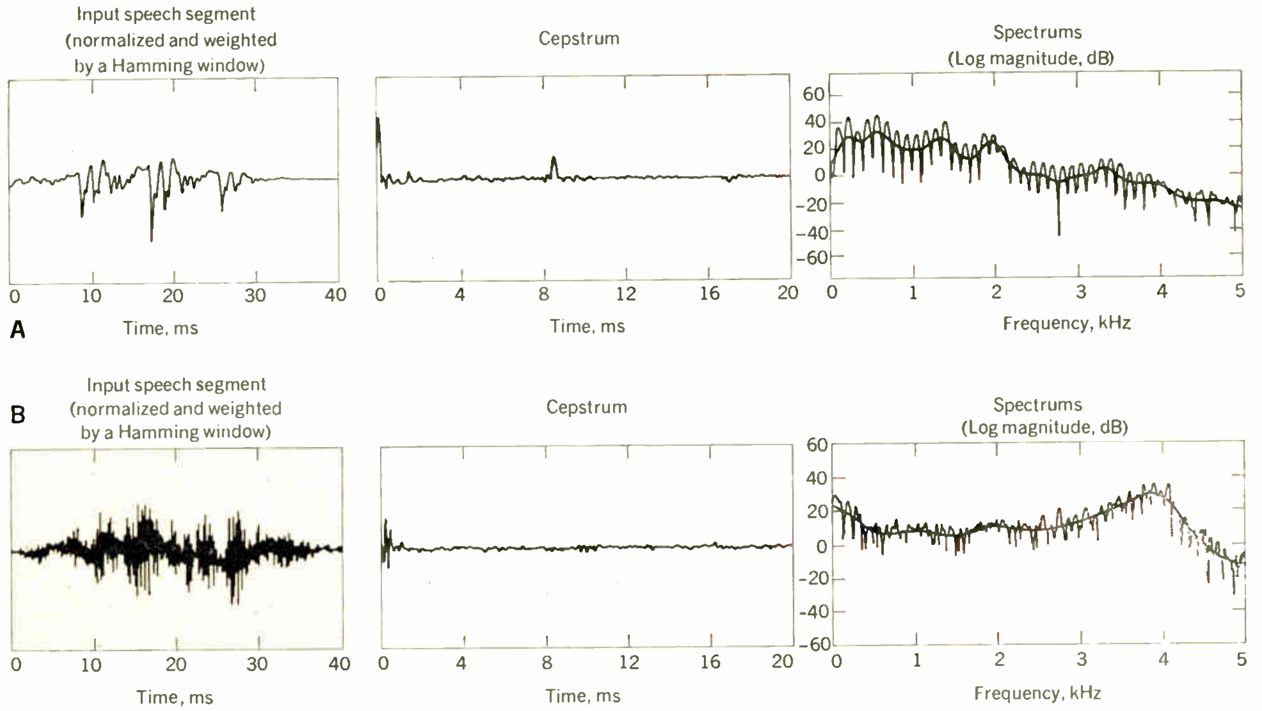


FIGURE 8. Short-time spectrum and cepstrum analysis of (A) voiced and (B) unvoiced speech.

trum are vocal-tract resonances. Acoustic constraints on formant frequencies and amplitudes are incorporated into an algorithm that locates the formant peaks in the smoothed spectrum.[12]

The analysis of unvoiced speech segments is depicted in Fig. 8(B). In this case, the input speech resembles a segment of a random-noise signal. As before, the log-magnitude spectrum of the speech segment can be thought of as consisting of a rapidly varying component associated with excitation, plus a slowly varying component due to the spectral shaping of the vocal tract. In this case, however, the rapidly varying component is not periodic but random. Again the low-time part of the cepstrum corresponds to the slowly varying component of the transform, but the high-time peak present in the cepstrum of voiced speech is absent for unvoiced speech. Therefore, the cepstrum can also be used in deciding whether an input speech segment is voiced or unvoiced. If voiced, the pitch period can be estimated from the location of the cepstral peak.[15] Truncation of the cepstrum and subsequent Fourier transformation produce the smoothed spectrum curve that is superimposed on the short-time transform at the right of Fig. 8(B). An adequate specification of the spectrum of an unvoiced sound can be achieved by estimating the frequency locations of a single wide-bandwidth resonance and a single antiresonance—that is, a single pole and zero.

Continuous speech is analyzed by performing these operations on short segments of speech, which are selected at equally spaced time intervals. Figure 9 illustrates this process for a section of voiced speech. The short-time spectrum and smoothed spectrum corresponding to each cepstrum are plotted adjacent to the cepstrum. Time increases from top to bottom, and each set of curves corresponds to a segment of speech offset 20 ms from the preceding segment. The formant peaks are connected by straight lines. One notices that formants occasionally come close together in frequency and pose a special problem in automatic estimation.

In the third and fourth spectrums* from the top of Fig. 9, the second and third formants are so close together that there are no longer two distinct peaks in the Fourier spectrum. A similar situation occurs in the last four spectrums, where the first and second formants are not resolved. A procedure for detecting such situations and for enhancing the resolution of the formants is shown in Fig. 10.[12]

The upper curve is the smooth spectrum as evaluated along the $j\omega$-axis of the s-plane. (The lowest three eigenfrequencies are depicted in their approximate locations.) Because formants two and three (F_2 and F_3) are quite close together, only one broad peak is observed in the spectrum. However, when the spectrum is evaluated on a contour that passes closer to the poles, two distinct peaks are in evidence, as shown in the lower curve. A computation algorithm known as the Chirp z-transform algorithm facilitates this additional spectral analysis.[16]

Formant synthesis. Once the excitation and transmission parameters are obtained, they are used to synthesize a waveform that approximates the original speech signal. Numerous systems, both analog and digital, have been devised for formant synthesis.[17-22] A digital system is illustrated in Fig. 11. The upper branch produces voiced speech. Its excitation source produces a tra'n of impulses with spacing equal to τ (the fundamental pitch period). The signal A_V, also estimated from the natural speech, controls the intensity of the pulse excitation

* The authors take no credit for the plural terminology—such as spectrums, cepstrums, and phenomenons—used in this article.

applied to a cascade of variable digital resonators. The resonator system is specified (under steady conditions) by the system function

$$H_V(z) = \prod_{k=1}^{4} \frac{1 - 2e^{-\alpha_k T}\cos(2\pi F_k T) + e^{-2\alpha_k T}}{1 - 2e^{-\alpha_k T}\cos(2\pi F_k T)\, z^{-1} + e^{-2\alpha_k T} z^{-2}}$$

where T is the sampling period, the F_k's are the formant frequencies (only three of which are time-varying), and the α_k's are the formant bandwidths (all of which are fixed). The output of this system excites a fixed system whose transfer function is

$$S(z) = \frac{(1 - e^{-aT})(1 + e^{-bT})}{(1 - e^{-aT}z^{-1})(1 + e^{-bT}z^{-1})}$$

This network is a cascade of two simple poles, and is designed to approximate the spectral shaping due to radiation and source properties.

The lower branch of Fig. 11 produces unvoiced speech. A random-noise generator, whose intensity is controlled by the signal A_N, excites a digital filter whose steady-state transfer function is given by the relation

$$H_U(z) = \frac{AB}{CD}$$

where

$$A = 1 - 2e^{-\beta T}\cos(2\pi F_P T) + e^{-2\beta T}$$
$$B = 1 - 2e^{-\beta T}\cos(2\pi F_Z T)\, z^{-1} + e^{-2\beta T}z^{-2}$$
$$C = 1 - 2e^{-\beta T}\cos(2\pi F_Z T) + e^{-2\beta T}$$
$$D = 1 - 2e^{-\beta T}\cos(2\pi F_P T)\, z^{-1} + e^{-2\beta T}z^{-2}$$

In these expressions, F_P and F_Z are, respectively, the time-varying pole and zero center frequencies for the unvoiced sound and β is the fixed bandwidth of both the pole and zero. The output of this system is passed to the fixed spectral compensation filter to provide the unvoiced speech output. Control parameters are supplied to the synthesizer at least at their Nyquist rates, and output samples are generated at a 10-kHz rate.

Figure 12 illustrates an example of automatic analysis and synthesis. Figure 12(A) shows the pitch period and formant signals (each band-limited to 16 Hz) as automatically estimated from a natural utterance. Figure 12(C) shows the spectrogram of speech synthesized from the estimated control parameters. For comparison, Fig. 12(B) shows the spectrogram of the original signal. Figure 13 shows spectrograms of original and synthetic versions of another utterance. Comparison of the spectrograms for the original and synthetic signals indicates that spectral properties are reasonably well preserved. [The best way to evaluate the results is actually to listen to them. For this purpose, a recording is available in conjunction with this article. Appendix A gives the contents of the record and information on how it may be obtained. Section 1 of the recording illustrates automatic analysis and synthesis.]

Efficiency of formant synthesis. The storage efficiency of the formant representation of speech depends on the precision with which the basic parameters must be specified. Synthetic speech of high quality can be obtained if the pitch period is specified to the nearest 0.1 ms, the gain specified to one place in 100, and the formant frequencies to the nearest 1 Hz. Since the parameters are estimated and supplied to the synthesizer 100 times per

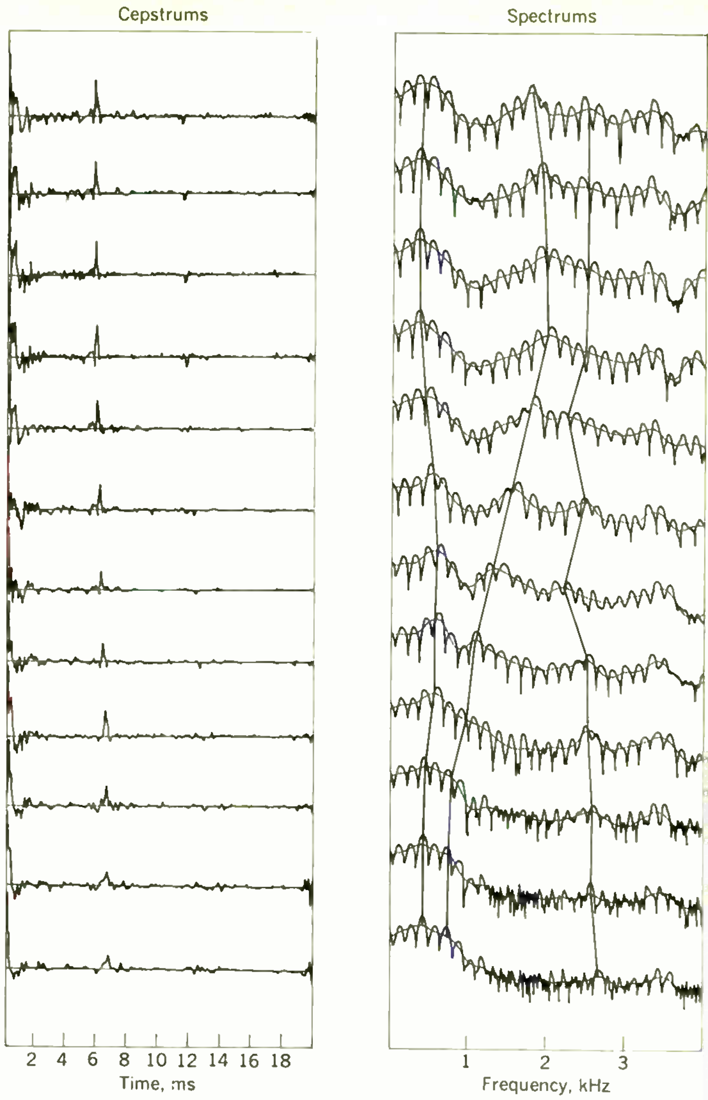
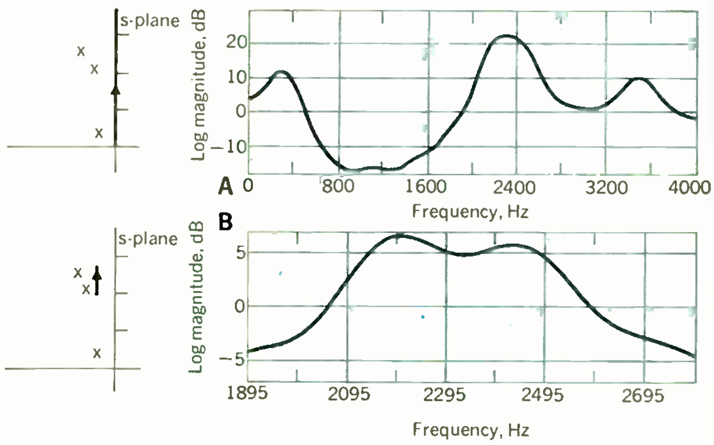


FIGURE 9. Cepstrum analysis of continuous speech. The left row shows cepstrums of consecutive segments of speech separated by 20 ms. The right row shows the corresponding short-time and cepstrally smoothed spectrums.

FIGURE 10. Illustration of the enhancement of formant resonances. A—Cepstrally smoothed spectrum in which F_2 and F_3 are not resolved. B—Narrow-band analysis along a contour passing closer to the poles.



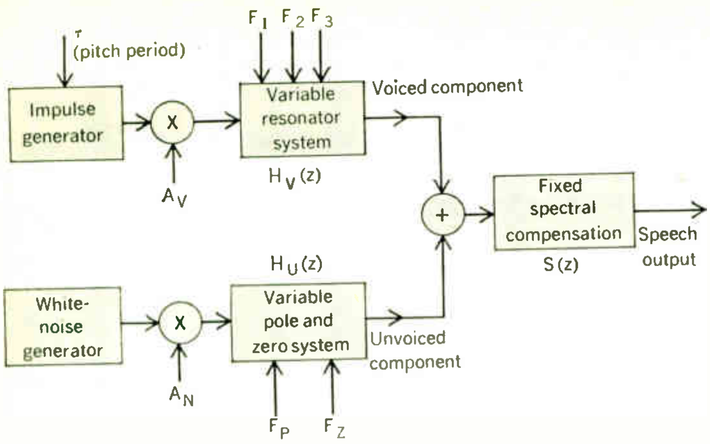


FIGURE 11. Digital formant synthesizer.

second, the required information rate can be shown to be 4600 b/s. (See Appendix A for a derivation of this figure.) Although this represents a considerable saving over the PCM representation, it is important to determine how much the information rate can be lowered before there is a significant perceptual difference between speech synthesized at 4600 b/s and speech synthesized at a lower rate.

In investigating this question, two considerations are important. The first is the bandwidth required to preserve the essential variations of the parameters. The second is the degree to which the synthesis parameters may be quantized. Sampling the parameters 100 times per second implies that they occupy a bandwidth of 50 Hz. In fact, their bandwidth occupancy is much smaller.

Figure 14(A) shows the formant frequencies as automatically estimated from natural speech at a rate of 100 Hz—that is, allowing a 50-Hz bandwidth for each parameter. Figure 14(B) shows the same data smoothed by a 16-Hz low-pass filter. Similarly, Fig. 14(C) shows the results for a 4-Hz low-pass cutoff. An auditory comparison of conditions represented by Fig. 14(A) and (B) shows that 16-Hz smoothing has negligible effect on the parameter variations. However, Fig. 14(C) shows that the 4-Hz filter has smeared out the waveforms significantly. Preliminary perceptual experiments indicate that the parameters can be band-limited to approximately 12 to 16 Hz with no noticeable degradation.[23] These results imply that the required sampling rate may be no higher than about 35 times per second. [Section 2 of the record illustrates the effect of low-pass-filtering the speech parameters.]

The second consideration involves the quantization of the parameter samples. Experiments indicate that pitch must be quantized to approximately 6 bits, whereas the remaining parameters each require 4 bits or less. Figure 15 shows a comparison between formant functions determined at 100 Hz and quantized to 11 bits and those sampled at 35 Hz and quantized to 3, 4, and 2 bits, respectively. The total bit rates for the syntheses are 4600 and 600 b/s, respectively. The two results are nearly indistinguishable. [Sections 3 and 4 of the record illustrate quantizing effects and give an example of speech synthesized at 600 b/s.]

It should be pointed out that these results are not meant to be taken as completely general. The exact bit rate depends on many factors, such as speaking rate and the nature of a particular speech utterance, which can vary widely. The results do give an idea of how low the information rate can become, and demonstrate that intelligible and natural-sounding synthetic speech can be generated from data rates of the order of 1000 b/s.

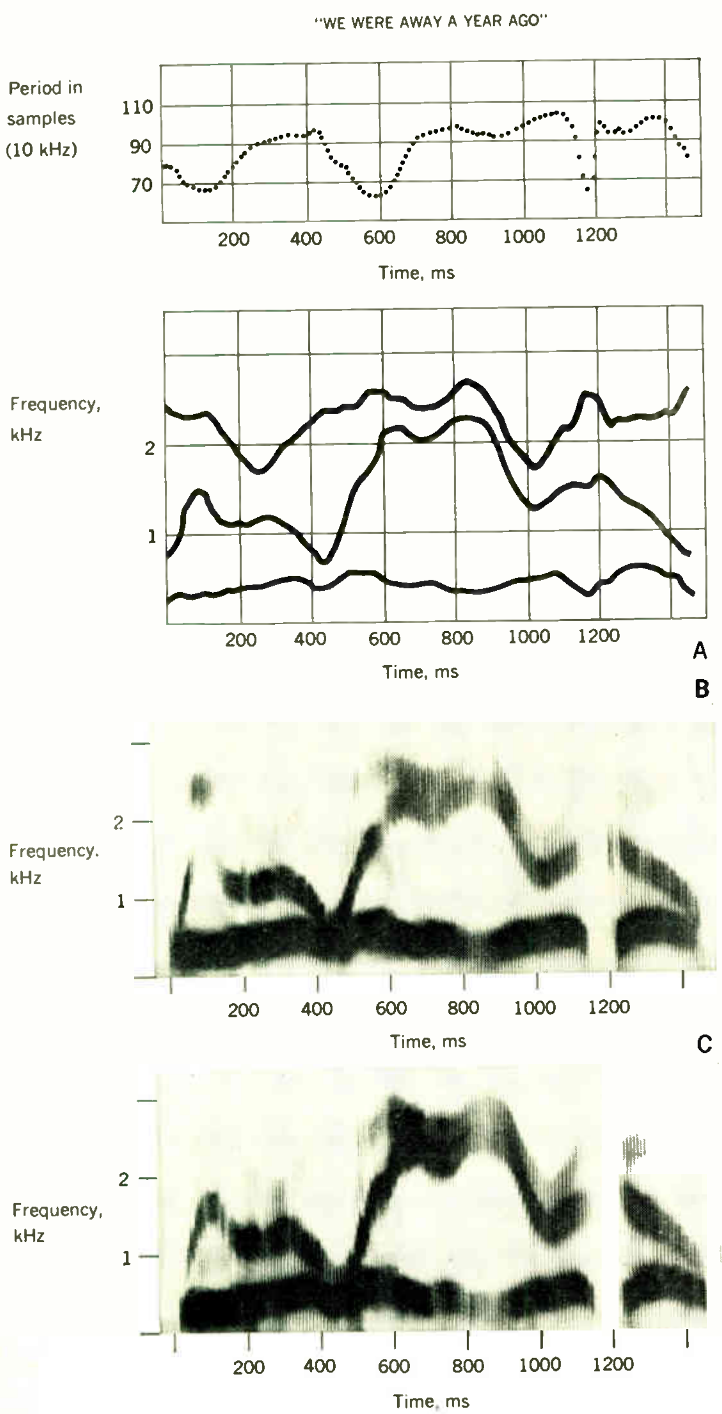


FIGURE 12. Automatic analysis and synthesis. A—Pitch period and formant frequencies estimated from natural speech. B—Spectrogram of the original speech. C—Spectrogram of synthetic speech.

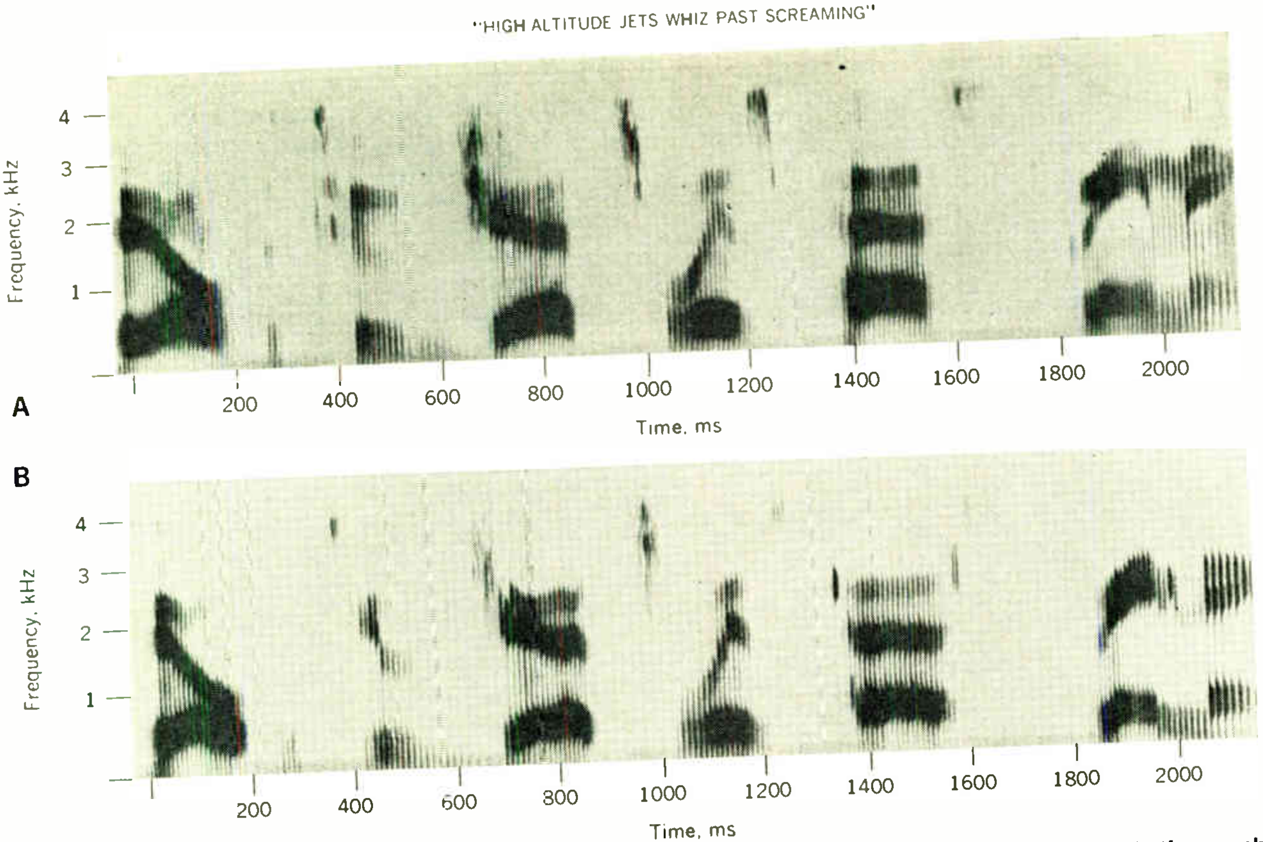
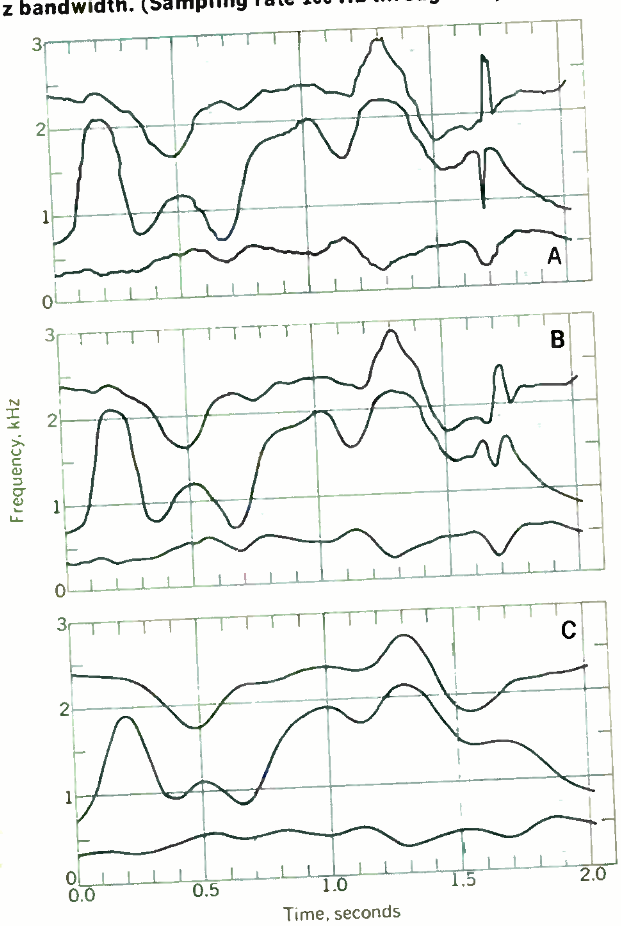


FIGURE 13. Automatic analysis and synthesis. A—Spectrogram of original speech. B—Spectrogram of synthetic speech.

FIGURE 14. Smoothing of formant signals. A—50-Hz bandwidth (no smoothing). B—16-Hz bandwidth. C—4-Hz bandwidth. (Sampling rate 100 Hz throughout.)



Speech synthesis by concatenation of formant data. The previous discussion emphasized the storage economy that can be realized by coding speech in terms of formant and excitation data. Another advantage of formant-coded speech is equally important—namely, its flexibility for fabricating messages from preanalyzed, naturally spoken, isolated words.

In the formant representation of an utterance, formant frequencies, voice pitch, amplitude, and timing can all be manipulated independently. Thus in synthesizing an utterance one can substitute an artificial pitch contour* for the natural contour. A steady-state sound can be lengthened or shortened, and even the entire utterance can be speeded up or slowed down with little or no loss in intelligibility. Formants can be locally distorted, and the entire formant contour can be uniformly raised or lowered to alter voice quality. [Section 5 of the record demonstrates synthetic speech where one or more of the basic parameters have been manipulated.]

A concatenation program that uses the flexibility of formant-coded speech to synthesize message-length utterances is shown in Fig. 16. Words spoken in isolation are automatically analyzed to estimate the parameters required by the synthesizer of Fig. 11. Quantized Nyquist rate samples of these control signals are stored in a word catalog. To synthesize a message composed of words from the catalog, the printed word string is supplied to the concatenation program. The program uses separate strategies for deriving the "segmental features" of the

* "Contour" is used to mean the time course of the relevant parameter.

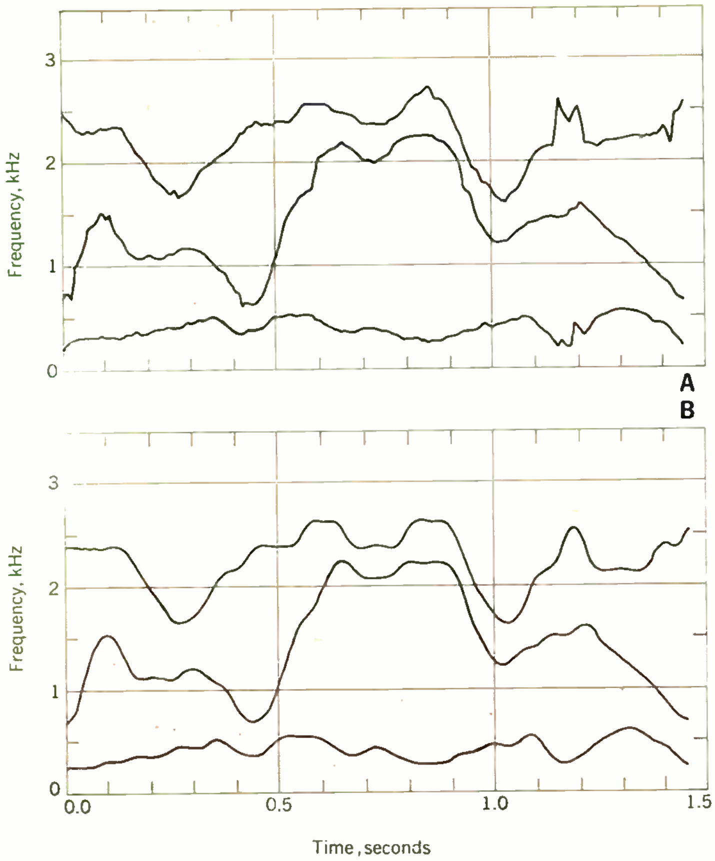
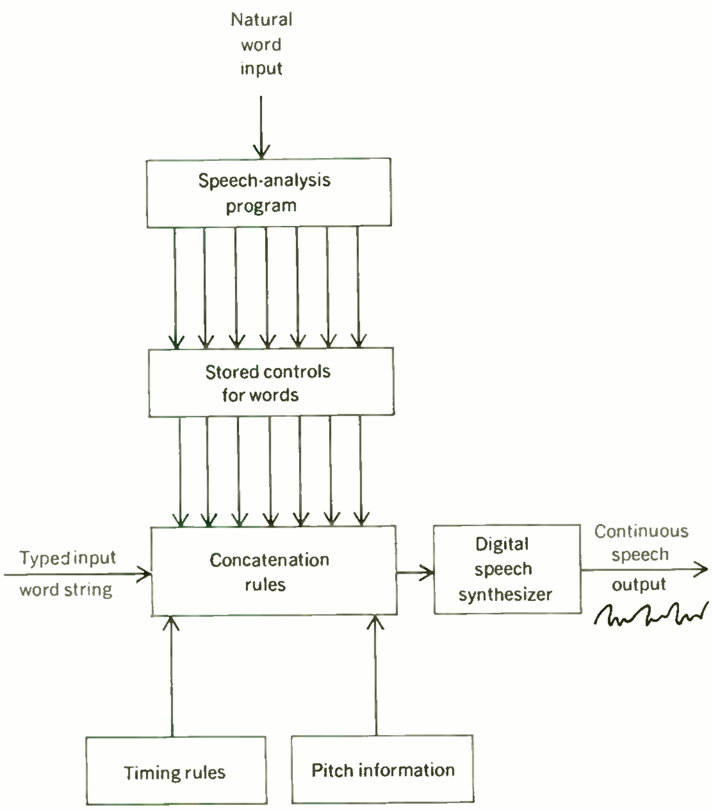


FIGURE 15. Quantization and smoothing of formant signals. A—Unquantized formants (4600 b/s). B—Quantized and smoothed formants (600 b/s).

FIGURE 16. Block-diagram representation of concatenation program.



message (formant frequencies and unvoiced pole and zero frequencies) and the "prosodic features" (timing, amplitude, and pitch). Figure 16 shows separate inputs for timing rules and pitch information. The program strategy for deriving segmental features is included in the box labeled "concatenation rules." It is the flexibility in manipulating formant-coded speech that permits breaking the synthesis problem into two parts. The output of the concatenation program is a set of smoothly varying formant-synthesis parameters, which are supplied to a digital synthesizer of the type shown in Fig. 11.*

To synthesize a continuous message, timing, pitch, and formant information must be generated. Timing information is derived in several ways. The techniques employed include

1. External specification of the duration of each word in the input string to be synthesized. In this case, word duration is chosen according to some external criterion (e.g., it can be measured from a naturally spoken version of the message to be synthesized) and in no way is meant to be a typical duration, independent of context.

2. Calculation of word duration by rules based on models of English language. Rules of this type are described in the next section and are used for synthesis from printed text.

3. Specification of word duration from tables of stored data. For limited-context messages, such as sequences of digits, such specification of word duration often is acceptable.

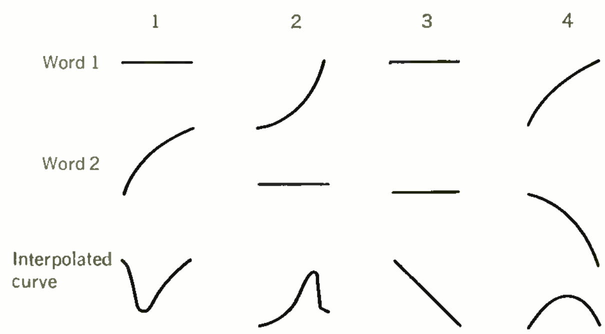
The next step in the synthesis procedure is to determine the pitch contour (e.g., the pitch period as a function of time) for the message to be synthesized. Pitch information can be obtained in several ways. The use of the pitch information measured from the originally spoken words and the use of monotone pitch are possibilities, but both give unacceptably unnatural results. Three remaining alternatives are

1. Supplying a pitch contour extracted from a naturally spoken version of the message. Data of this type would normally be used when word durations have been obtained in a similar manner and supplied externally. Pitch and timing information obtained in this manner are the most natural.

2. Calculating a pitch contour by rule, as discussed in the next section.

* Both software and hardware synthesizers are available with the DDP-516 computers employed.

FIGURE 17. Interpolation of formant contours.



3. Using an archetypal pitch contour. For limited-context messages, this pitch contour is modified (i.e., locally shortened or lengthened) to match the word durations as determined from the timing rules. There are obviously many ways in which the prosodic information for the message can be obtained and the choice of the foregoing alternatives depends strongly on the desired quality of the synthetic speech and the specific application in mind.

Once the timing pattern for the message is established, the segmental data in the word catalog must be altered to match the timing. This means that the formant data for a word in the catalog must be either lengthened or shortened. The formant contours in successive voiced words must also be merged to form smooth, continuous transitions.

The choice of place in a word to alter duration is based on the dynamics of the formant contours. For every 10-ms voiced interval of each word in the catalog, a measure of the rate of change of the formant contours is computed. This measure is called the "spectral derivative." Regions of the word in which the spectral derivative is small are regions wherein the word can be shortened or lengthened with the least effect on the word intelligibility. Thus to shorten a word by a given amount, an appropriate number of 10-ms intervals are deleted in the region of the smallest spectral derivative. To lengthen a word, the region of the lowest spectral derivative is lengthened by adding an appropriate number of 10-ms intervals. Unvoiced regions of words are never modified.

Whenever the end of one word is voiced and the beginning of the next word is also voiced, a smooth transition must be made from the formants of the first word to those of the second word. This transition is made over the last 100 ms of the first word and the first 100 ms of the second. The transition rate depends on the relative rates of spectrum change of the two words, over the merging region. Figure 17 gives examples of interpolations for typical shapes of formant curves. (In the figure it is assumed that, for both words 1 and 2, all three formants have identical shapes; only one formant is shown in the illustration.)

In case 1, word 1 has a very small spectrum change over its last 100 ms of voicing, but word 2 has a very large spectrum change. The interpolated curve shown at the bottom of the first column, though beginning at the formants of word 1, rapidly makes a transition and follows the formants of word 2. Case 2 shows the reverse of case 1; word 2 has little spectrum change, whereas word 1 has a large spectrum change. The interpolated curve now follows the formants of word 1 for most of the merging region, making the transition to the formants of word 2 at the end of the region. Cases 3 and 4 show examples where spectrum changes in both words are relatively the same. When they are small, as in case 3, the interpolated curve becomes essentially linear. When they are large (case 4), the interpolated curve tends to follow the formants of the first word for half the merging region, and the formants of the second word for the other half of the merging region.*

The final step in producing the message is to synthesize the speech using the chosen prosodic features and segmental features generated by the preceding rules. A hardware digital speech synthesizer performs this task in real time at a sampling rate of 10 kHz. Its control signals are updated pitch-synchronously from the interpolated stored data.

Figure 18 is a spectrographic illustration of the synthesis technique. At the top is a spectrogram of a natural utterance of "We were away a year ago." At the bottom is a spectrogram of the same sentence produced by abutting individually spoken, formant-synthesized words with no alteration of pitch or timing of the isolated words. The lack of continuity of formants is obvious. The resulting synthetic output is choppy and completely unacceptable. At the center of Fig. 18 is a spectrogram of the individual words concatenated according to the rules discussed previously. For this example the pitch and timing were obtained from the natural version of the utterance. It can be seen that the rule-interpolated formant functions of the middle spectrogram closely resemble the natural transition of the top spectrogram.* [Examples are on section 6 of the record.]

Figure 19 compares original and synthesized versions of the limited-context message "The number is 836-1246." Pitch and timing for the synthetic signal were calculated completely by rule. Timing for the complete message was derived from a stored table and is specific to sequences of 7-digit telephone numbers. An archetypal pitch contour was calculated and fitted to the timing of the word elements.† [Several examples of this type are given on section 7 of the record.]

Present success with rule-concatenated, word-length formant data makes it appear attractive for computer voice response. Simple rules for timing and pitch appear to suffice for certain limited-context applications. Broader application will depend upon the success of general prosodic analysis, as described in some detail in the next section.

Synthesis from printed text

Synthesis of speech from printed text offers virtually unlimited message capacity. The expense is the large storage capacity required for a pronouncing dictionary of word data.

A typical desk-size abridged dictionary lists 130 000 words. If the storage form is to be PCM or formant-encoded words, the dictionary must be expanded to include variations of most entries as pronounced with different common endings: plurals for nouns; *-s*, *-ed*, *-ing*, and *-er* for verbs; *-er*, *-est*, and *-ly* for adjectives; and as well, verbs with a number of prefixes, such as *re-*, *de-*, and *un-*, and nouns with invented word endings, such as *-ize*, *-ish*, and *-y*, and, as well, *-ized*, *-izes*, *-izing*, *-izer*, etc.

The dictionary would contain many infrequent words of long duration. Such words contain, on the average, 3½ syllables per word, one of which is generally stressed, and may be up to 1 second per word in length. If we assume that the dictionary contains from 500 000 to

* This interpolation function is a preliminary one, chosen for initial experiments. The ideal interpolation is likely to be context-dependent. This question is the subject of continuing research.

* Note particularly the way the sound "*a*" at about 1400 ms in the bottom spectrogram is smoothly interpolated into the formant tracks at about 1000 ms in the middle spectrogram.

† The ideal pitch contour, like the formant interpolation function, is likely to be context-dependent. The archetypal contour used here represents a preliminary function for experiment.

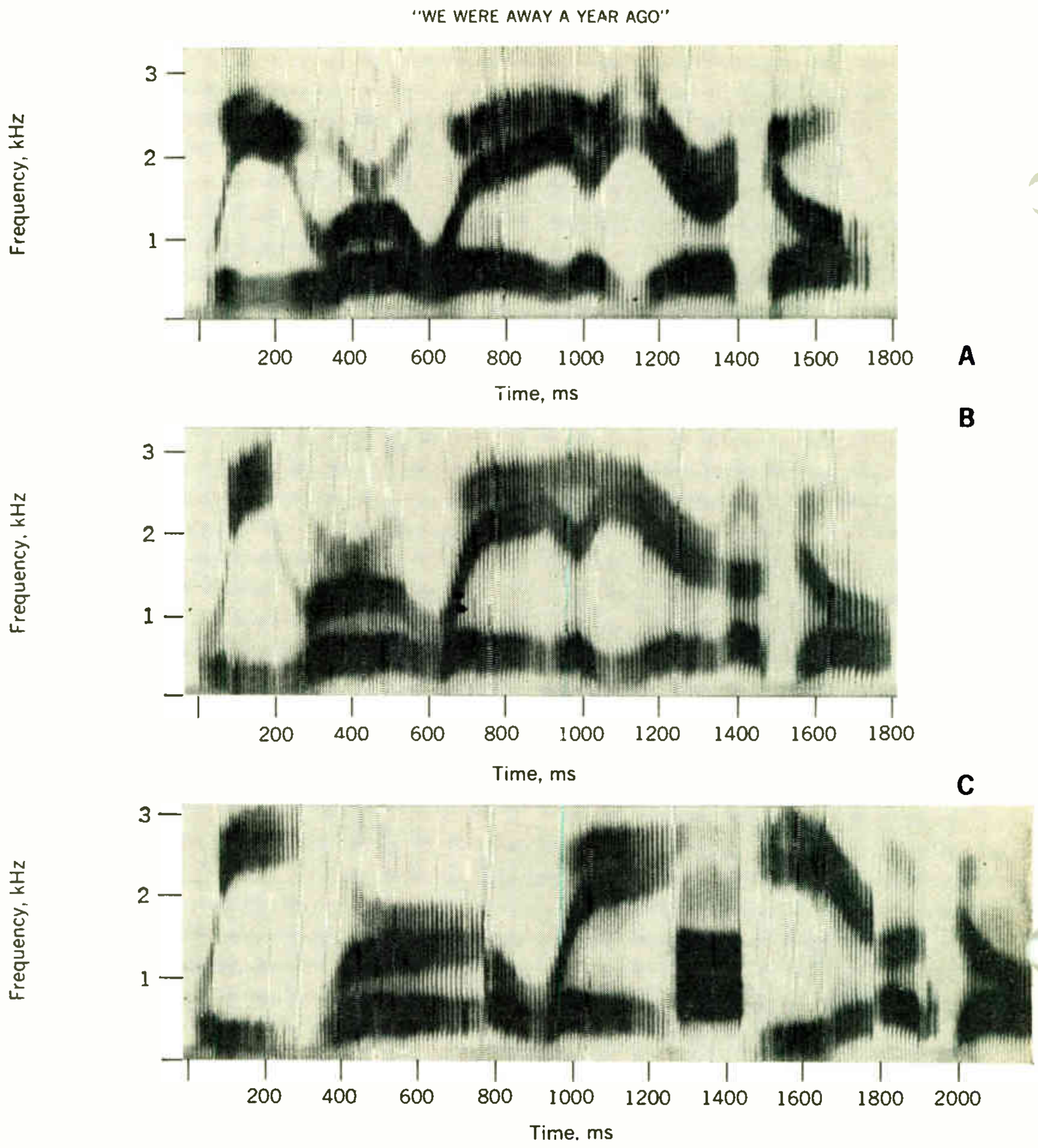
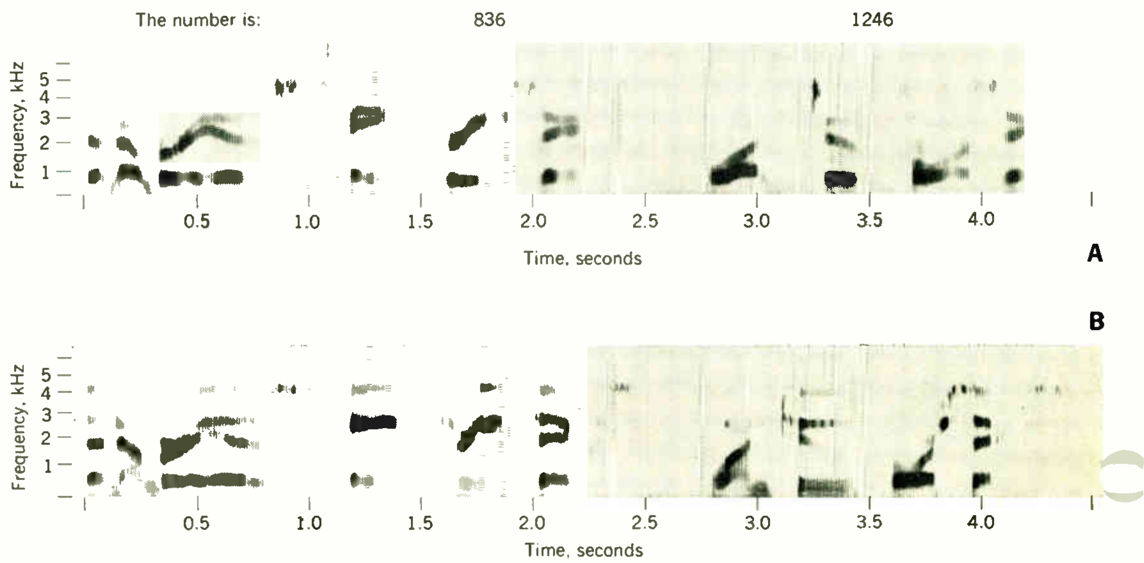


FIGURE 18. Sound spectrograms of (A) natural speech, (B) concatenated speech, and (C) isolated words.

FIGURE 19. Sound spectrograms of (A) concatenated digits and (B) naturally spoken digit string.



700 000 words including derived forms, we can project a PCM pronouncing dictionary to require around 100 hours of speech or, equivalently, 25 to 37 billion bits! With formant storage, the estimate drops to perhaps 500 million bits, which is still rather large but necessary for talking encyclopedias. A more compact form is obviously desirable; the most compact form is probably phonemic transcription.

Phonemic transcriptions of single-word entries may be encoded with about 72 bits per second (6 bits per phoneme; 12 phonemes and lexical stress marks per second). Thus the phonemic dictionary offers a 13:1 direct saving over formant storage and a 700:1 saving over PCM. In addition, it permits simple means for generation, rather than storage, of derived word forms, usually by simple concatenation of phonemes.* Using such schemes to derive words, rather than store them, the saving might approach 100:1 over formant data, or about 5000:1 over PCM.

We shall describe a working experimental system for synthesis from text, using a phoneme dictionary. The system differs from other work in two significant ways: (1) It contains fully automatic rules for conversion from English text (without special marks) to speech with reasonably natural timing and intonation. (2) It is a synthesis of unrestricted speech from a dynamic characterization of the human articulatory system.

A block diagram of the text to speech conversion program is shown in Fig. 20.† Blocks A through D convert from printed text to discrete symbols representing phonemes with detailed data for pitch and duration. Blocks E through I convert the discrete phonemic symbols to sequences of articulatory motion and thence to sound.

Articulatory synthesis. Figure 21 is a diagram of the vocal-tract model used for phoneme synthesis. Seven parameters are used to describe the cross-sectional area of the vocal tract as a function of distance along it. The parameters are: two coordinates each for the configuration of the lips (W, L), the position of the tongue tip (R, B), and the position of the tongue body (X, Y); and one coordinate for the position of the velum and uvula (N). The width of the pharynx and position of the teeth (jaw angle) are dependent variables inferred from the position of the tongue body.

The design of such a model seeks to satisfy two somewhat conflicting goals: (1) The model should be very general; that is, it should approximate well all of the vocal-tract shapes that occur in normal speech. (2) The model should be strongly constrained. It should exhibit such natural constraints of the vocal mechanism as continuity of the tongue surface, curvature of the vocal tract, and discontinuities of the tract at the teeth, velum, and esophagus. Most important, the model should incorporate the temporal constraints and dynamic behavior of the speech mechanism.

In the present model, there exists a good balance between constraints to exclude unnatural vocal-tract shapes, and flexibility to match the shapes that do occur in speech. Figure 22 shows comparisons between actual human vocal-tract shapes and those of the model,

* This subject was studied extensively by Lee.[24]

† Work on this system began and developed on an older computer, a DDP-24. Continuing research is implemented on a DDP-516, described in Appendix B.

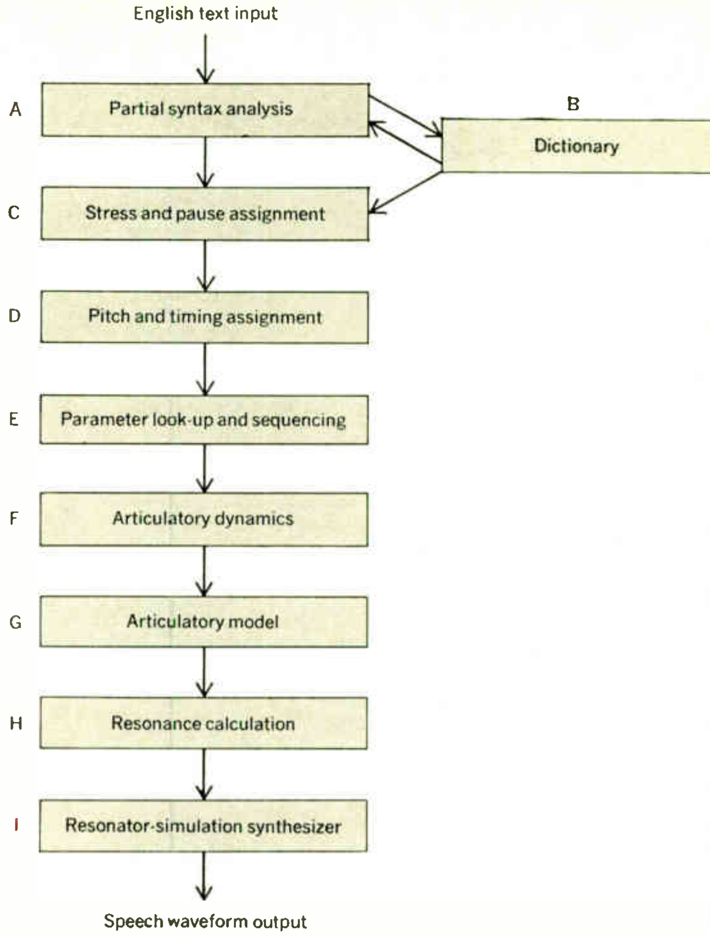
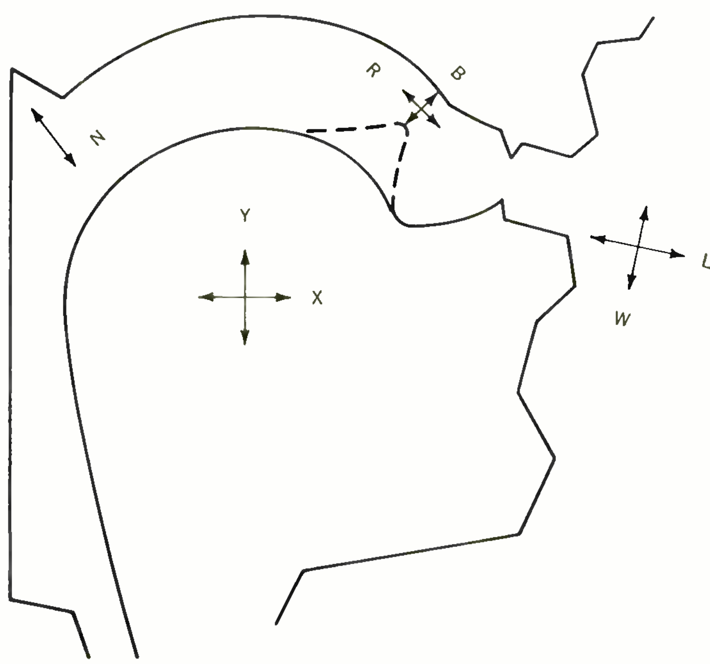


FIGURE 20. Block diagram of a complete system for synthesis of speech from English text. Blocks A through D convert from conventional text to detailed phonetic text. Blocks E through I achieve phoneme synthesis by modeling human articulation.

FIGURE 21. Schematic diagram of the computational articulatory model used in phoneme synthesis.



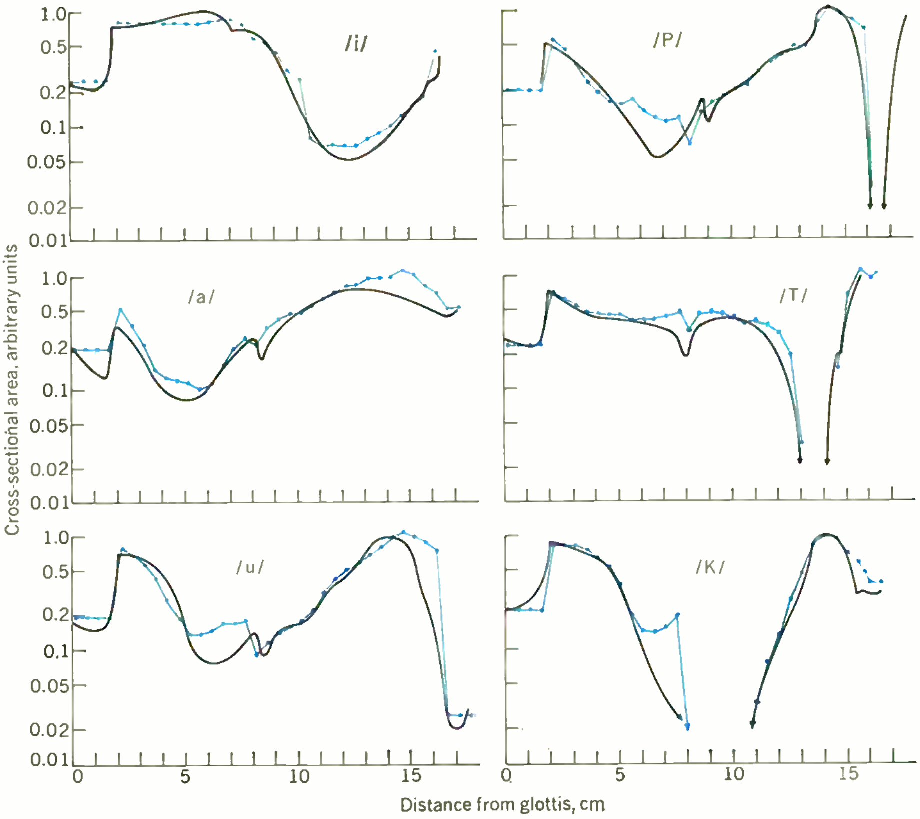
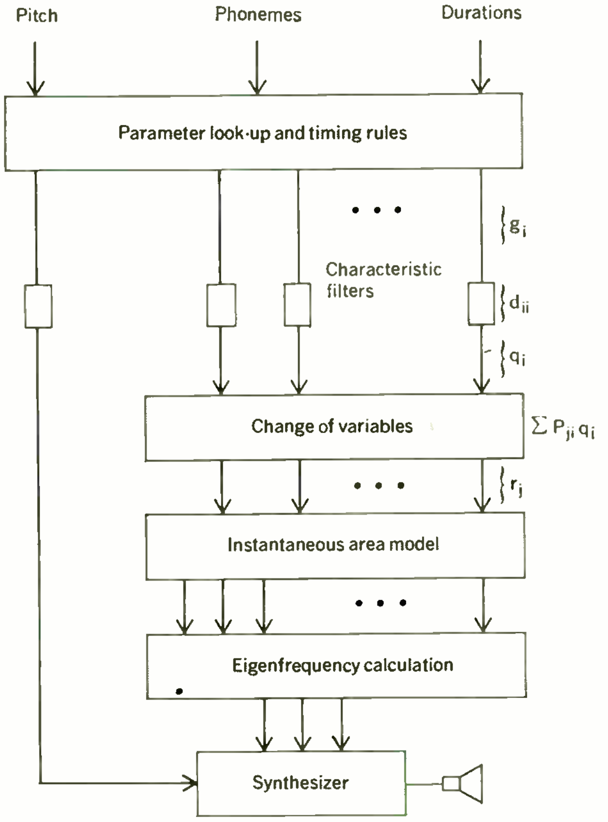


FIGURE 22. Examples of the ability of the articulatory model to fit human articulatory data as taken from X rays.

FIGURE 23. Block diagram of the synthesis system showing the method of supplying dynamics for the articulatory model. The variables g_i identify closely with "command variables" for independent rounding or closing the lips, motion of the tongue tip and tongue body. "Characteristic filters" simulate dynamics, and a change of variables accounts for relatively minor control interactions. (See Fig. 25.)



for a representative group of speech sounds. The solid lines are plots of cross-sectional area against distance along the vocal tract. The data are derived from multiple X-ray photographs of a human. The dotted points are corresponding functions computed from the vocal-tract model.

An important property of the model is its ability to represent accurately the shape of the vocal tract where area is small. It provides specific coordinates for control of place and degree of constriction in most of the consonants of English. This leads to a direct way of establishing the dynamic behavior of the model or, at least, of optimizing the dynamics of each part of the model for the specific phonemes in which that part is most significant. For example, control of lip protrusion (L) is optimized for the rounded vowels and the sound /*w*/; control of lip opening (W) is chosen for good reproduction of the consonants /*p*/, /*b*/, and /*m*/.

Figure 23 shows how the model is used to synthesize connected speech from sequences of discrete symbols. The computation of the time-dependence of the model coordinates is logically divisible into two separate parts: the sequencing of "position commands" to the model, and the simulation of responses to these various commands.

Basic speech sound elements of the synthesizer vocabulary, most of them corresponding to conventional phonemes, are stored for the model as sets of "idealized position coordinates" or target positions. These are presented as sequences of step functions (g_i) to a set of "characteristic filters" (d_{ii}), which simulate the dynamics of the system. These filters have second-order overdamped responses, with rise times ranging from 50 to 200 ms.

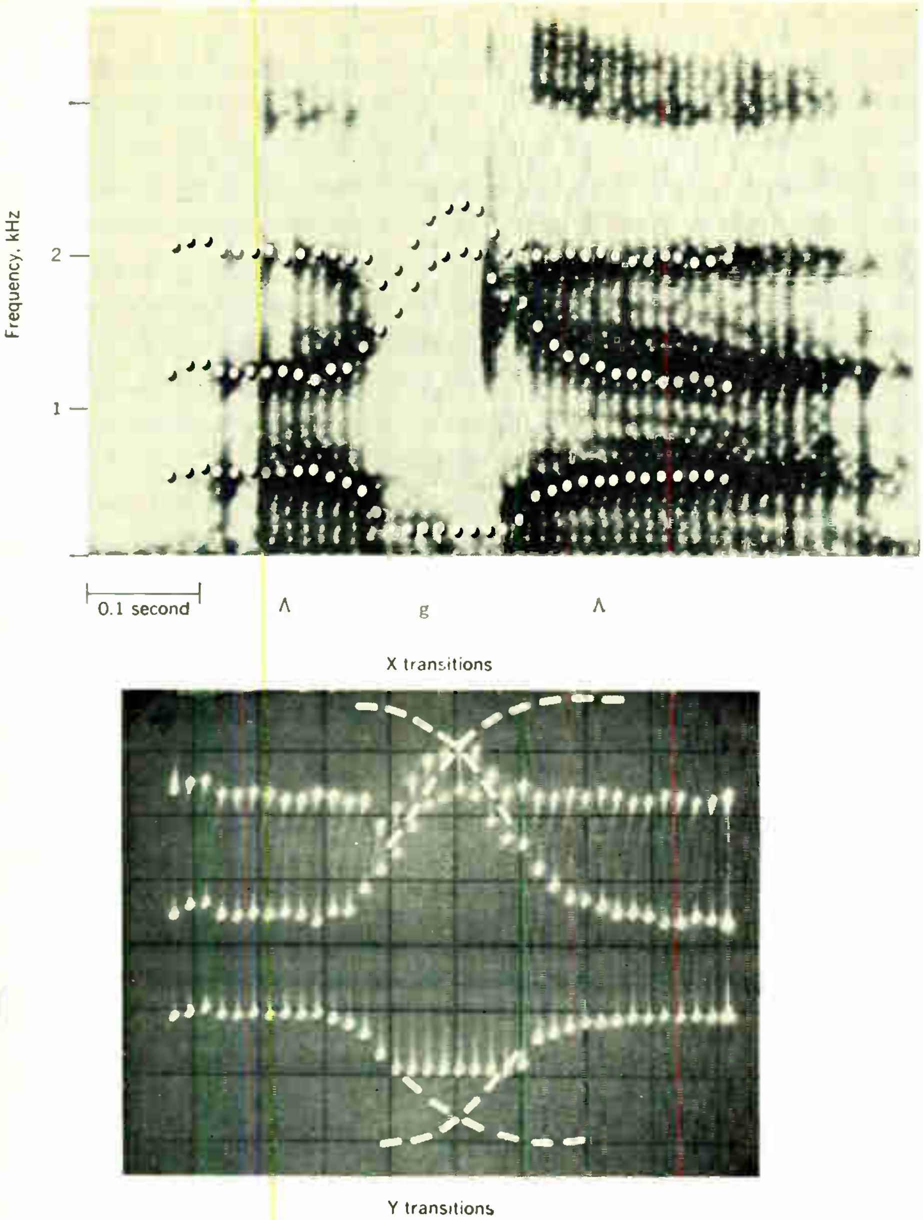


FIGURE 24. Comparisons of formants calculated from the articulatory model with a spectrogram of natural speech for the same utterance. Dots superimposed on the spectrogram were traced by the computer scope display. Sketched on the lower photograph are transitions in control variables to the articulatory model.

The responses for commands for lip rounding, velar motion, and front-to-back motion of the tongue tip are slowest; those for simple lip closure and raising of the tongue tip are fastest. The outputs q_i of the "characteristic filters" are converted to the position coordinates of the model (r_j) by a transformation of variables. The transformation accounts for components of tongue tip and lip motion due to tongue-body motion, and for a component of lip closure due to lip rounding.

The position coordinates r_j specify the tract area function at 10-ms intervals of time. These coordinates are applied to the computational model to derive sets of data representing area as a function of distance along the tract. These area functions are incorporated into a one-dimensional lossless solution of Webster's horn equation. The equation is iterated to solve for the first three eigenfrequencies of the tract for each 10-ms interval of time. The data at this point refer to the variables of formant synthesis. These computed formant frequencies are outputted to a hardware synthesizer, of the type shown in Fig. 11, to generate the sound.

Figure 24 is a comparison of a natural-speech utterance of a nonsense syllable */ígagi/* (*eégahgee*) and the formant frequencies for a similar synthetic sequence computed from the vocal-tract model. The figure also illustrates the basic approach to establishing dynamic response of the model. The response of the lowest resonance is found to depend primarily on the control variable dominating the size of the constriction for the consonant */g/*; the transients in the second and third formants are found to depend mainly on the parameter affecting the place of the constriction. Similar relationships are found for parameters controlling the lips and tongue tip in other consonants.

Besides the dynamics of articulatory motion, another important feature is incorporated in control of the model. The fixed-target, step-function method of control matches real speech only if some asynchrony is introduced into the timing of different articulatory parameters. Consider, for example, the transition from a vowel to a lip or tongue-tip consonant. The specific gesture of the lips or tongue tip to form the consonant occurs primarily before any significant motion of the tongue body. In a transition between two consonants, the specific gesture for each consonant overlaps that for the other. The result is that a constriction for one or the other of the consonants is consistently maintained. At the "midpoint" of the transition, the significant gestures of both consonants are almost fully articulated.

This phenomenon is incorporated in the model by a priority strategy. Each phoneme has a table of factors designating the relative importance of each articulatory parameter to the formation of the phoneme. The timing of step-function changes in the control variables g_i is staggered according to comparisons of priorities between adjacent phonemes. Transitions to a critical value therefore occur early, and those away from a critical value occur late.

Figure 25 illustrates results of this process. The plotted data are the filtered variables q_i, before the change of variables to actual model coordinates r_j. Transitions in the nasality, tongue, and lip variables are caused to occur primarily during the neighboring phonemes. Nasality, for example, is unimportant for the phoneme /*s*/ in "once more." The action of the priority algorithm allows the velum target positions of the preceding phoneme /*n*/ and following phoneme /*m*/ to dominate during /*s*/. A similar example occurs for the tongue-tip front/back control in the sequence "more from." In the tongue-tip front/back control for "the beginning" and in the lip extension control for "once," the control transitions can be seen to extend across not one but two adjacent phonemes.

Although the action of the model must be considered only approximate to that of a real vocal tract, Fig. 25 illustrates a rather remarkable property of speech. Discrete symbols (phonemes) are transmitted through a multivariable articulatory system whose parts are too sluggish individually to resolve discrete values at the natural rate of phoneme production. The sequential constraints of the language and the control strategies of articulation, however, allow sufficient time for each individual articulator to reach its goal *when it becomes necessary*.

Section 8 of the recording is an example of the capabilities of the articulatory synthesizer when controlled with hand-supplied data. For this example, discrete phonemes with additional symbols for timing and pitch control were fed into the computer by typewriter. Timing and pitch were selected by repeated listening and retyping as necessary to make the synthesized sound more natural. Experiments of this type have been the primary source of data upon which rules for the prosodic features of English have been developed. Additional resources include the dynamic matching of X-ray motion pictures and the visual matching of sound spectrograms.[25]

Conversion of text to detailed phonetic transcription. The articulatory model just described (blocks E through I of Fig. 20) requires discrete phonetic symbols and pitch and duration data to effect synthesis. Printed alphabetic text must therefore be transformed into this phonetic form (blocks A through D of Fig. 20).

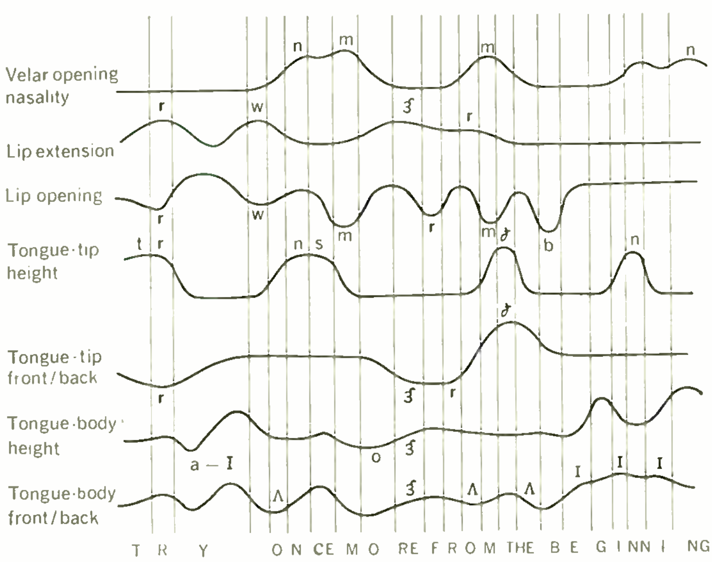
Automatic conversion of printed English text into discrete phonetic symbols must provide sufficient information about the prosodic features of speech to effect a natural-sounding synthesis. A human speaker never gives the same importance to every word. Some words are made prominent by giving them higher pitch, increased intensity, and lengthened duration. Some words are so reduced (shortened) that phonemes become very short and weak, and may even be dropped completely. Pauses may be inserted in a sentence at places not marked by punctuation.

Recently, with the development of digital computers, several attempts have been made to program systems for the automatic conversion of printed text to synthetic speech.* These efforts depend upon an analysis of the surface structure of sentences to derive information about grammatical boundaries and about the patterns of stress within sentences.

The first part of the system described here (blocks A through C of Fig. 20) generates the minimum information for pause and stress assignment. The remaining part of the system (block D) generates pitch and timing assignments for each phoneme and produces discrete symbols that are used directly as the input to the articulatory model described in the previous section. The rules for pitch and timing assignment, however, are also applicable to the formant synthesis technique mentioned previously, as well as to similar synthesis systems.

Syntax analyzer and dictionary. The dictionary (block B, Fig. 20) provides a phonemic transcription of each word with lexical stress marks, and the possible usages

FIGURE 25. Outputs of "characteristic filters" for an example utterance. The variables B and R for the articulatory model are B′ and R′ above, plus components of tongue-body position, X and Y. Actual lip opening W is the sum of W′ above, a component of lip extension L and components of X and Y representing jaw angle.



* Holmes, Mattingly, and Shearme developed semiautomatic rules in which the phonemic transcriptions and stress and pause assignments were made by a human operator.[9,10] A fully automatic system for accomplishing the conversion was first developed by Teranishi and Umeda.[26] The syntax analyzer reported here is an outgrowth of the latter work. Treatments of stress and pause assignment were made by Vanderslice.[27] Also, the work of Lee[24] and Allen[25] is contemporary with the work reported here.

of the word, such as noun, adjective, or verb. It also provides a content–function distinction.* Content words convey substantial meaning in the sentence. They relate to things, actions, or attributes. Function words, usually monosyllabic, serve mainly to establish grammatical relationships. Function words include articles, prepositions, conjunctions, and personal pronouns. There are also "intermediate" words, which are neither function nor content. Polysyllabic, less-frequent prepositions and conjunctions, and frequently used verbs (such as "get," "take," "give," etc.) fall into this category, as do some pronouns, adverbs, and adjectives.

The role of the syntax analyzer (shown as block A in Fig. 20) is multiple. For each sentence it must (1) assign the probability of a break (or more exactly, the potential of a pause) at every grammatical boundary (such as a phrase or clause), (2) select alternative pronunciation of certain words according to their usage, and (3) change the function–content distinction for specific conditions to assign proper stress to important words in the sentence (e.g., an auxiliary verb at the end of a clause has to be changed to a content word). Using the information on word class contained in the dictionary, the analyzer groups words into phrases. For each word, it assigns a phrase category such as introductory modifier, subject, verb, object (or complement), or tail modifier, according to the possible order of occurrence of the phrase categories. The potential of a pause or of a weaker break is assigned between all words. Words in the same phrase have zero probability of break. Break probability is higher between subject and predicate than between verb and object. Break probability is relatively high between an introductory prepositional phrase and the subject. Any reverse order of occurrence among phrase categories indicates a clause boundary at the reverse point. Clause boundaries have higher probability of a break than any phrase boundary within the clause. Punctuation marks require the highest probabilities of a break.

Stress-pause assignment. Using the information on probabilities of a break obtained in the syntax analyzer, the program element of this stage (block C, Fig. 20) decides what kind of break and pitch contour is to be assigned at the end of the grammatical unit. The threshold for putting an actual pause or pseudo-pause† in the synthetic speech shifts depending on the length of the sentence, and on the speech rate. The unit separated by a pause or break with one focus stress will be called a "pause group." Full stops indicate the longest possible pause and have a falling pitch contour at the end. Commas indicate a pause and a rising pitch contour, implying continuation. For deliberate speech and for sentences of reasonable size, clause boundaries without punctuation are terminated without pause and with rising pitch. Phrase boundaries with higher probability of a break are accompanied by the elongation of the last phoneme in the phrase.

Stress levels are assigned to words mainly according to the content-function distinction. The stress levels are used in computing pitch and duration information. The stress assignments are:

1. No stress: function words
2. Weak stress: intermediate words
3. Primary stress: content words
4. Focus stress: focus of the sentence☆

A typical result of the entire conversion of printed text to discrete symbols for the articulatory synthesizer input is shown in Fig. 26. In the left column the input English text is shown word by word. (In this case, the text is the first line of Aesop's fable of the North Wind and the Sun.) Column 2, labeled *cat*, represents the phrase category of each word. Column 3, labeled *wc*, gives word class. Column 4, labeled *pp*, represents the probability of break. Column 5, labeled *cf*, makes a content–function distinction in the role of each word. Column 6, labeled *ic*, indicates the shape of the pitch contour for each word. Column 7, the rightmost column, shows the form of the input information to the synthesizer. These symbols are the result of pitch and timing assignments applied to data determined from the syntax analysis, dictionary look-up, and stress-pause assignment. These output symbols are described next.

Pitch-timing assignment. From the previously derived information, timing-control marks and pitch marks are assigned to each phoneme. In the rightmost column in Fig. 26, numerals and minus signs are timing controls, and the special marks *, $, and & are pitch controls that increase voice pitch from a nominal value. *q* is a pitch mark that lowers the pitch in a prescribed decrement. All alphabetic characters, except *q*, specify English phonemes. The symbols ⟨ and ⟩ indicate the second part of diphthongs—front and back, respectively.

The program has a timing table for all phonemes. Each phoneme has a fixed minimum duration and an additive variable duration. Numbers for timing control represent the duration in terms of the sums of the fixed portion and some multiple of the variable portion. Duration values are specific to individual phonemes. A given number can therefore play the same role in prosodic rules independent of absolute duration.

Two major rules are used to determine the timing and pitch controls: (1) a word boundary rule, which determines consonant durations, and (2) a stress and termination rule, which determines pitch marks and vowel durations.

Word boundary. Consonant durations at word boundaries are adjusted according to the relations of Fig. 27. Function words are merged together (like unstressed syllables inside a polysyllabic word) and take minimum duration of consonants at their boundaries (0-boundary). Average durations are assigned to consonants at the boundaries of words with intermediate stress (1-boundary). Content words always receive long consonants at the boundary so that they might be prominently separated in the stream of the utterance. Content-word boundaries (2-boundary) produce consonants about 24 ms longer than average (seen as the numeral 6 in Fig. 26). Intermediate boundaries (seen as 1-boundary in Fig. 27 and as the numeral 4 in the initial and/or final consonant in Fig. 26) produce consonants of average duration, and occur between intermediate and compound content words. Function word boundaries (0-boundary in Fig.

* This distinction is given intuitively by Pike.[29]

† A kind of break with rising pitch contour for termination without an actual pause following.

☆ In actual speech, any word could be the focus. In our system, however, the last content word in the pause group is assigned focus stress, unless the focus word is especially marked in the input text.

English Text	cat	wc	pp	cf	ic	
the	s	tce	0			4dh 4a
north	s	a	0	++	-	6n $4aw 2er 6th
wind	s	n	0	++	-	6w *qq5i 4n 4d
and	s	cla	0			4aa -n -d
the	s	tce	0			-dh 4a
sun	s	n	0	++	-	6s *qq5uh 6n
were	v	vbp	5			4w 4er
arguing	v	vg	0	++	-	4: $q6ah -r -g -y 4uu 4i 6ng
one	o	aq	2	+		4w &5uh 4n
day	o	n	0	++	/	6d *q9ay qq9<
,	p	comm	11	++	/	$,
when	i	whn	0	+		2h 2w &5eh 4n
a	s	tca	6			4a
traveler	s	n	0	++	*	4t 4tr *q7aa -v 4o -l 4er
came	v	vp0	5	+		4k &4ay 4< 4m
along	t	p2	6	+	/	4a 4l 8aw 4ng
,	p	comm	11	++	/	$,
wrapped	v	vp0	0	++	-	6r $q8aa 4p 4t
in	t	pl	6			4i -n
a	t	tca	0			4a
warm	t	a	0	++	-	6w $5ah 2er 6m
coat	t	n	0	++		6k *q2oh qq20h 6t
	p	peri	12	++		

FIGURE 26. Printout of a program that converts from English text to discrete phonetic symbols. The leftmost column is the input. The columns cat (phrase category) and wc (word class or "part of speech") are internal decisions of the syntax analyzer; the columns pp (pause probability) and cf (content/functional distinction) are its output. The column ic (intonation contour) describes the stress pattern assignments: rising, sustained, or falling pitch. Data in the rightmost column, phonemic symbols, pitch, and timing, control the articulatory synthesizer to produce speech.

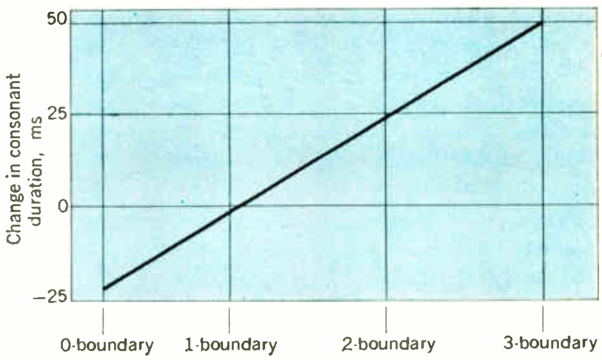


FIGURE 27. Consonant duration in terms of degree of word separation. 0-boundary is a weak separation typical of the juncture of two monosyllabic functional words. The 2-boundary is typical of the juncture of two important content words.

27) produce the minimum consonant duration at the boundary, about 18 ms shorter than average (shown as minus sign in Fig. 26). A more prominent boundary (3-boundary, Fig. 27) is applied to a phrase boundary where greater separation is needed.

Inside polysyllabic words the initial consonant (or consonant cluster) of a stressed syllable is assigned 1-boundary duration, and any other consonant gets 0-boundary duration.

Stress and termination. In accordance with the stress levels assigned to words, pitch marks are put on the stressed vowels as follows:

No stress (every unstressed vowel)		no mark
Weak stress		&
Primary stress:	reduced (verb)	$*q*$
	normal (adj. and adv.)	$
	high (noun)	**qq*
Focus stress		**q*

The higher the pitch, the longer the vowel is made.

Two kinds of termination are assumed. Termination is used here to denote the pitch contour used to terminate a pause group.

1. Falling pitch: sentence end (unless yes–no question)
2. Rising pitch: middle of the sentence, followed by a comma, or at a point at which the break probability is high

Combinations of stress and termination assignment, used with the phonetic context, provide a range of timing control for each vowel. The continuation pitch rise forces a vowel to be quite elongated. Focus stress also produces a long vowel. When these two factors fall on the same vowel, it is forced to have a steep falling pitch followed by a rising pitch; consequently the vowel becomes very long. Voiced consonants are also a factor in elongating a preceding vowel. When all three factors occur on the same vowel at the end of the pause group, the vowel is lengthened greatly, typically three times as long as its normal duration in ordinary context.

Figure 28 shows duration modification of vowels. Each line represents a different phonetic condition of stress, with or without termination. Different lines represent different conditions of stress and termination as a function of the phonetic contexts. Figure 28 is specific for the vowels /a/ and /c/. Other vowels have different slopes and ranges of variation for their duration increments. The pitch and timing assignments complete the information derived by the text-conversion program. The resulting data of the rightmost column in Fig. 26 are then supplied to the articulatory model and converted to connected speech. A spectrogram of the output synthetic speech is shown in Fig. 29. For comparison, a spectrogram of a corresponding natural utterance is also shown. [The reader is directed finally to sections 9 and 10 of the recording. These passages demonstrate speech synthesized automatically from printed-text input.]

Summary

The methods of speech synthesis from stored formant data and from printed text have advanced to a promising point. Their potential for automatic information services and for computer voice response appears good. At this point in time the quality of formant-synthesized speech is generally better than that of text synthesis. In the former the so-called segmental information is derived from naturally spoken speech. The suprasegmental (prosodic) information is calculated either by rule or from stored measurements on real speech. The saving in required storage is approximately 50:1 compared with digitized speech waveform. In text synthesis, both segmental and suprasegmental information are calculated, and no shred of naturally derived data is relied on. The economy in storage is concomitantly greater, being of the order of 700:1 compared with the digitized speech waveform.

Both techniques appear to have places in the range

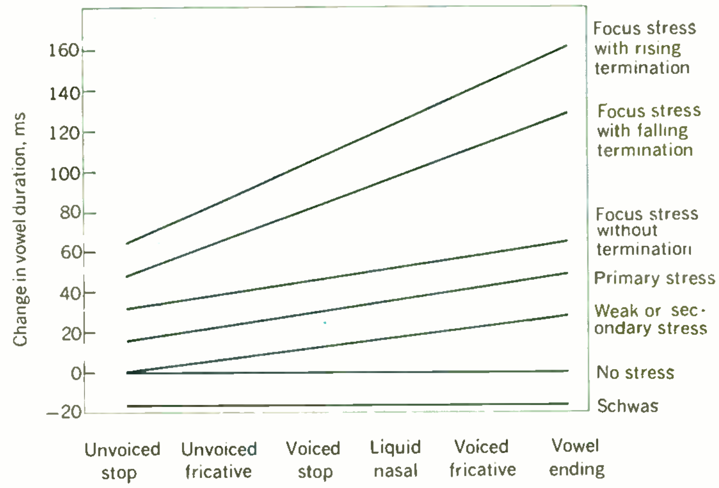
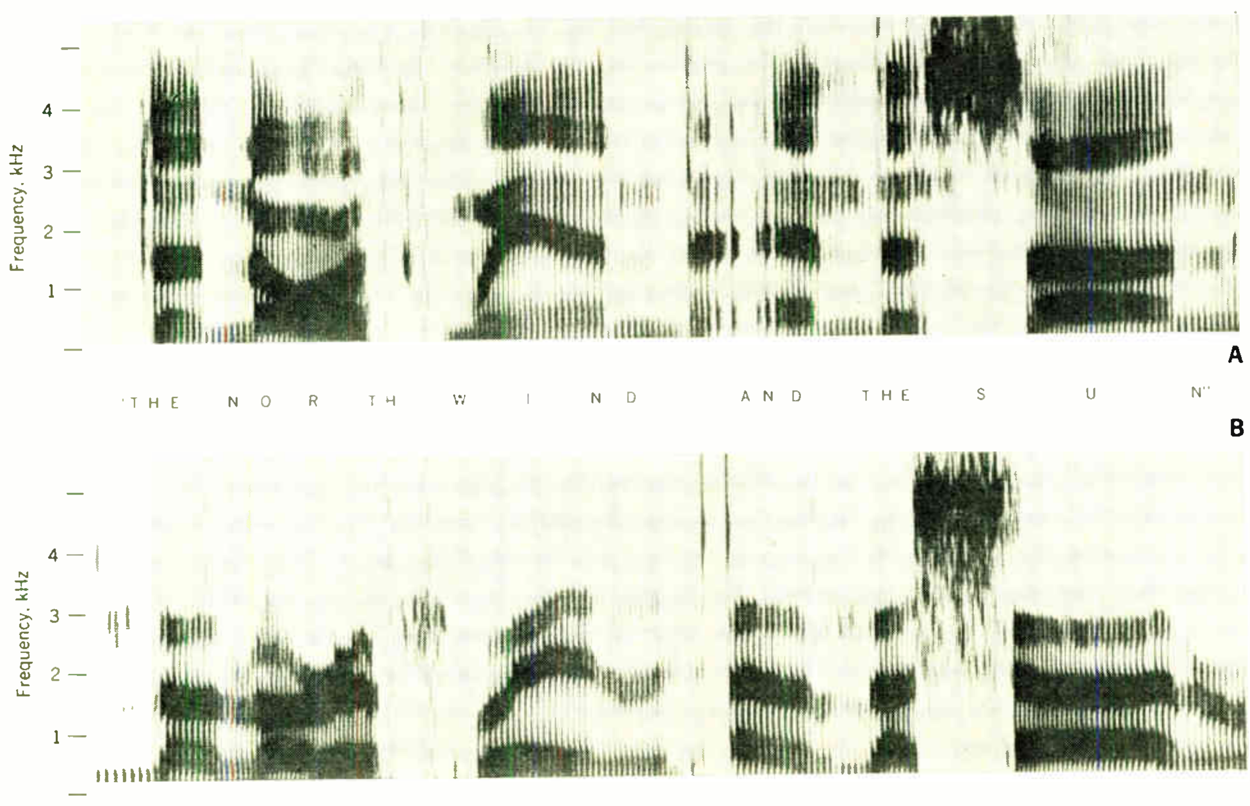


FIGURE 28. Vowel duration as a function of stress and phonetic context. The term "focus stress" indicates the major stressed word in a pause group.

FIGURE 29. Spectrogram of (A) natural speech and (B) speech synthesized from phoneme rules, for the same sentence.



of computer answer-back applications envisioned. Formant synthesis is a natural for cases requiring moderately large vocabularies of high information content (low redundancy) and good intelligibility. Typical applications might be automatic intercept and information operators, inventory reporting, weather forecasts, and stock market quotations. In contrast, automatic information services such as the talking encyclopedia would clearly depend upon text synthesis, as would uses such as reading machines for the blind and certain computer-based teaching techniques.

The synthesis methods described in this article are being researched using two laboratory computers especially tailored for this type of acoustic problem. Questions of feasibility in computer implementation therefore constitute an integral part of the work. Computation times, access times, and methods of computer control of digital hardware for synthesis are all relevant questions. For the reader interested in the computer techniques being used, an outline description of the computer facilities we have constructed is given in Appendix B.

Appendix A

Description of record

The recording mentioned in this article is an 8-mil-thick, 7-inch, 33⅓-r/min microgroove disk containing the demonstration items listed below.

To purchase the record, please complete the address label below and forward it with 50 cents in coin to IEEE, RC Unit, 345 East 47 Street, New York, N.Y. 10017.

Side 1—Formant analysis and synthesis

1. Automatic analysis and synthesis—3 utterances: synthetic—original—synthetic
2. Smoothing of pitch and formant controls—1 utterance, 4 smoothing filters; 16-Hz, 12-Hz, 8-Hz, 4-Hz low-pass cutoff
3. Quantization and smoothing of pitch and formant controls—1 utterance, 16-Hz cutoff low-pass filter used on all versions
   a. Pitch quantizing: 7, 4, 3, 2, 1 bits
   b. Formant 1 quantizing: 4, 3, 2, 1 bits
   c. Formant 2 quantizing: 4, 3, 2, 1 bits
   d. Formant 3 quantizing: 3, 2, 1 bits

4. Bit rate comparison—1 utterance

	4600 b/s	600 b/s
Pitch	7 bits	6 bits
F_1	10 bits	3 bits
F_2	11 bits	4 bits
F_3	11 bits	3 bits
A_V	7 bits	2 bits
Sampling rate	100 per second	32 per second

   4600 b/s—original—4600 b/s—600 b/s
5. Manipulation of pitch, timing, and formant controls—1 utterance
   a. Pitch manipulations
      (1) Synthetic unmanipulated
      (2) Monotone pitch—100 Hz
      (3) Pitch doubled
      (4) Pitch squared
   b. Timing manipulations
      (1) Synthetic unmanipulated
      (2) Vowels in "we" and "year" lengthened by 100 ms
      (3) Synthetic unmanipulated
      (4) Total duration = 75% natural duration
      (5) Total duration = 50% natural duration
      (6) Synthetic unmanipulated
      (7) Total duration = 150% natural duration
      (8) Total duration = 200% natural duration
   c. Formant manipulations
      (1) Synthetic unmanipulated
      (2) Formants raised by 10%, pitch raised
      (3) Formants raised by 20%, pitch raised
      (4) Formants raised by 30%, pitch raised by 50%
      (5) Synthetic unmanipulated
      (6) Formants lowered by 10%
      (7) Formants lowered by 20%
6. Concatenation of words
   a. Isolated words in sequence
   b. Words concatenated by rule; natural pitch and timing from speaker 1
   c. Words concatenated by rule; natural pitch and timing from speaker 2
   d. Words concatenated by rule; natural pitch and timing from speaker 3
7. Concatenated digit strings—four comparisons of strings of isolated digits followed by concatenated digits

Side 2—Synthesis from printed text

8. Articulatory synthesis from manual phonetic input
9. Automatic synthesis from printed text, "Parable of the North Wind and the Sun"
10. Synthesis from printed text

Appendix B

Laboratory computer facility for interactive studies of speech analysis and synthesis

The speech research described in this article is being carried out on an interactive laboratory computer especially configured for problems in acoustic signal processing. Because some of the capabilities are unique, engineering details are outlined for the reader interested in computer implementation.

The facility employed in these investigations includes two Honeywell DDP-516s. The machines communicate with each other and with a central GE-635 computer via data-phone connections. The two machines and their software systems are identical; programs are completely interchangeable. One machine is normally dedicated to problems in speech analysis and synthesis and digital filtering, and it typically serves nine research staff members. The second machine is presently dedicated to perceptual experiments on synthetic speech, auditory acuity, and acoustic signal processing. It serves about eight staff members.

The Honeywell DDP-516 is an integrated circuit machine with a 0.96-μs cycle time and 16-bit word length. As shown in Fig. 30, our configurations include 16 k of core memory, hardware multiply and divide, direct multiplex control (DMC) with 16 data channels (0.25 MHz each), and direct memory access (DMA) channel (1.0 MHz). An ASR-33 teletypewriter is standard (and was the only peripheral delivered with the machine). Fortran IV compiler, DAP-16 machine-language assembler, math libraries, and various utility software are supplied by the manufacturer.

For our range of problems we have interfaced the peripherals shown in Fig. 30:

1. Two fixed-head disks for each machine. Each disk provides 394 k words of storage with a 33-ms maximum access time and 180-kHz word transfer rate.

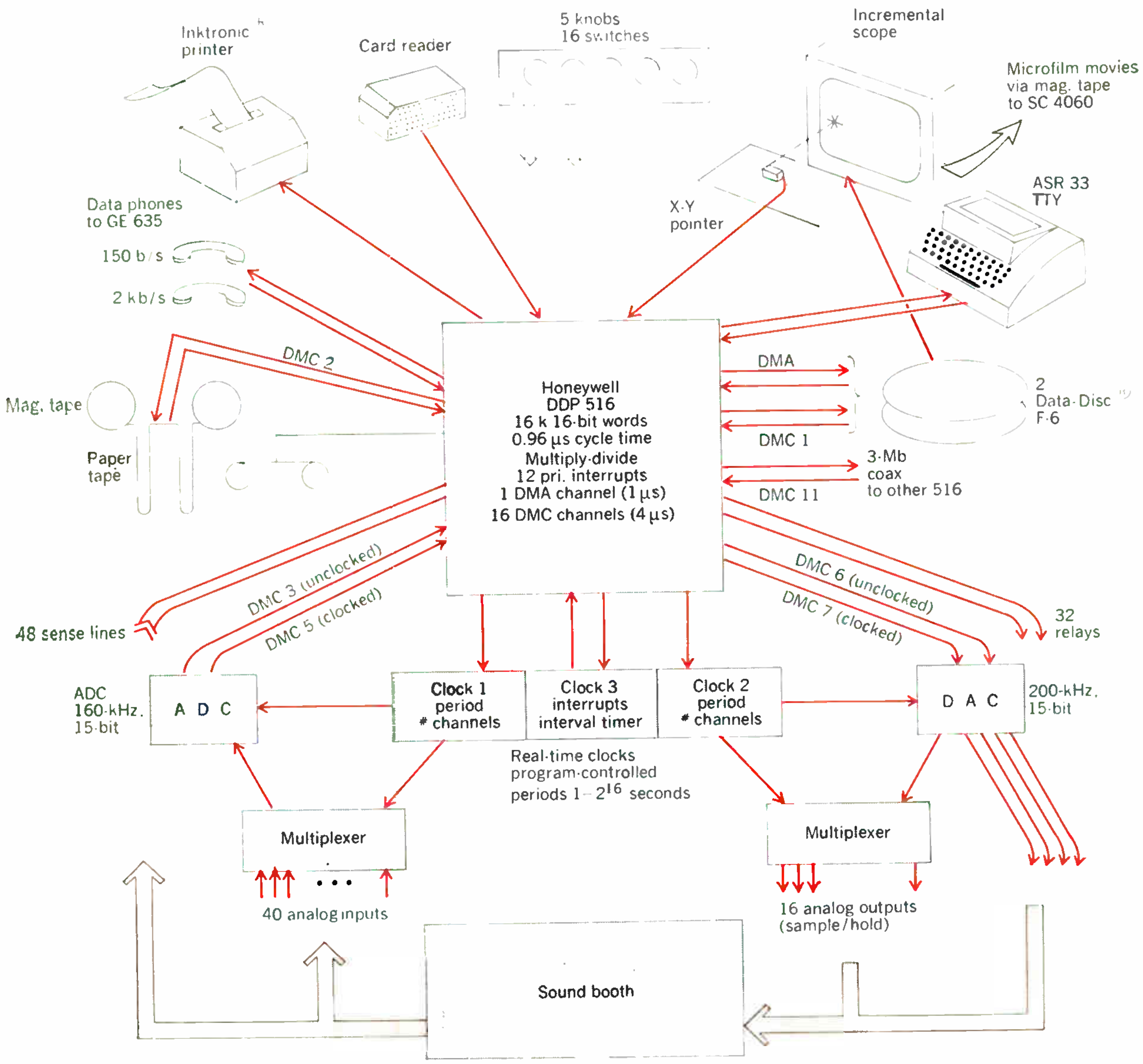
2. D/A converters. Four independent channels with program-controlled clocks for rates from dc to 180 kHz; 16 analog sample/hold channels with practical sampling rates from 50 Hz to 20 kHz.

3. A/D converters. 180-kHz 15-bit A/D converter; 40-channel 100-kHz multiplexer. The converter can operate through two data channels at the same time, sampling one group of channels synchronously at rates set by a program-controllable clock, and a different group asynchronously under direct program control.

4. Display scope. An incremental scope driven from a track on the disk, not from memory. The scope can reproduce 24 000 dots at an 825-kHz rate, thus providing a full page of text or about 10 meters of almost continuous line. An x-y pointer gives graphic input and manipulation.

5. Card reader. A 300-per-minute card reader is in-

FIGURE 30. **Laboratory computer facility for speech analysis and synthesis.**



corporated into the console. A keypunch is located beside the console. Card equipment and character set are compatible with the central GE-635 computer, and with nearby card editing and off-line listing facilities.

6. Inktronic® 120-character-per-second printer. Software allows the user to scan a listing with the display scope. With the press of a switch selected portions of the listing (such as the data-location map) are printed at a rate of 2 to 5 pages per minute.

7. Paper tape reader (150 characters per second). Used as a backup and for compatibility with manufacturer checkout software.

8. Magnetic tape drive (64 k characters per second). This nine-track drive provides selection of read and write densities. Tapes are compatible with nine-track drives on the central GE-635.

9. Three 16-bit input registers and two 16-bit output registers are used to control external electronic equipment.

10. A double-wall sound booth is installed adjacent to each machine for high-quality recording and for listening tests.

The operating system is based on the fast disk store, and the system programs reside permanently in 48-k write-protected words of the lower disk (⅛ of the total disk capacity). Users may write on all remaining disk and normally retain no permanent disk storage. Walk-away storage is through magnetic tape. Compiled programs may be conveniently saved on tape. Compiler, assembler, linking relocating loader, and all libraries are immediately available from permanent disks. Loading a Fortran source deck, compiling, listing on scope and/or line printer, loading compilation, and library all proceed expeditiously, with the disk serving as the storage medium at every step.

The scope system allows high-resolution line displays, and printing via Fortran-formatted I/O statements. The display is unique in that it provides a continually replenished display from the disk. The display consumes no core space and is operative even when the computer contains a different program or is halted. A feature of the display program causes, at the programmer's option, all display data to be echoed onto magnetic tape. A companion program in the local central computer allows regeneration of identical display frames via Xerox quick hard copy and SC4060 microfilm or 16-mm movies.

A display utility allows paging or scrolling through assembly and compilation listings (automatically written on disk), octal and decimal display of disk and core, quick hard copy of any printed display via the Inktronic printer or magnetic tape for microfilm, waveform display of disk or core, and simultaneous audio output.

A simple overlay procedure obviates serious limitations of the 16-k memory size. Through a subroutine call, a program can save itself onto disk, bring in another program, and transfer to it in about 0.1 second. Overlays may be nested to any depth and returns are made successively as with simple subroutine calls. A breakpoint and debugging package links the general display program as an overlay. A user can quickly alternate between running his program; looking at or listening to data on disk or inside his program; repositioning or printing selected portions of his listing; restoring his program to core, while keeping a page of listing on the scope; altering locations in his program or data; restoring registers; and running to another breakpoint or looping through the same one.

In evolving these laboratory systems, we followed a philosophy which to us seems important. Engineering of peripherals was done on a schedule that permitted maximum use of the machines at any stage of development. The first peripherals added were D/A converters. These required trivial effort and immediately allowed us to carry out three planned research projects. At the same time work on disks and tape proceeded. As a result it was possible to complete three separate research problems during the first year of operation of the first machine. By adding peripherals only when we were certain they were what we needed and would work, we insured that the facility was always a low-risk investment. Expenditures at each stage have been accompanied by productive output. This strategy also minimizes vulnerability to manufacturer delays or to new software difficulties. Considering the rate of depreciation of computer equipment, the overhead for space and support, this kind of planning, we feel, is valuable in obtaining the maximum return per computer dollar invested.

Our two DDP-516 systems are being applied to a range of problems in acoustic signal processing. Typical projects, besides the speech analysis and synthesis discussed here, include speech quality studies,[23,30,31] auditory detection,[32] vocal-cord modeling,[33] adaptive delta modulation, deconvolution of acoustic reverberation, and interactive design of digital filters.

REFERENCES

1. Flanagan, J. L., *Speech Analysis, Synthesis, and Perception.* New York: Academic Press, 1965.

2. Fant, C. G. M., *Acoustic Theory of Speech Production.* The Hague: Mouton and Co., 1960.

3. Flanagan, J. L., and Landgraf, L. L., "Self-oscillating source for vocal-tract synthesizers," *IEEE Trans. Audio and Electroacoustics,* vol. AU-16, pp. 57–64, Mar. 1968.

4. Chapman, W. D., "Prospectives in voice response from computers," *Proc. Internatl. Conf. Commun.,* 1970.

5. Dixon, N. R., and Maxey, H. D., "Terminal analog synthesis of continuous speech using the diphone method of segment assembly," *IEEE Trans. Audio and Electroacoustics,* vol. AU-16, pp. 40–50, Mar. 1968.

6. Rabiner, L. R., "Speech synthesis by rule: an acoustic domain approach," *Bell System Tech. J.,* vol. 47, pp. 17–37, Jan. 1968.

7. Liberman, A. M., Ingemann, F., Lisker, L., DeLattre, P., and Cooper, F. S., "Minimal rules for synthesizing speech," *J. Acoust. Soc. Am.,* vol. 31, pp. 1490–1499, 1959.

8. Kelly, J. L., Jr., and Gerstman, L. J., "An artificial talker driven from a phonetic input," *J. Acoust. Soc. Am.,* vol. 33, p. 835 (A), 1961.

9. Holmes, J. N., Mattingly, I. G., and Shearme, J. N., "Speech synthesis by rule," *Language and Speech,* vol. 7, pt. 3, pp. 127–143, July–Sept. 1964.

10. Mattingly, I. G., "Synthesis by rule of prosodic features," *Language and Speech,* vol. 9, pt. 1, pp. 1–13, Jan.–Mar. 1966.

11. Cooper, F. S., Gaitenby, J. H., Mattingly, I. G., and Umeda, N., "Reading aids for the blind: a special case of machine-to-man communication," *IEEE Trans. Audio and Electroacoustics,* vol. AU-17, pp. 266–270, Dec. 1969.

12. Schafer, R. W., and Rabiner, L. R., "System for automatic formant analysis of voiced speech," *J. Acoust. Soc. Am.,* vol. 47, pt. 2, pp. 634–648, Feb. 1970.

13. Bogert, B. P., Healy, M. J. R., and Tukey, J. W., "The quefrency alanysis of time-series for echoes," *Proc. Symp. Time Series Analysis,* M. Rosenblatt, ed., chap. 15, pp. 209–243, 1963.

14. Oppenheim, A. V., Schafer, R. W., and Stockham, T. G., "Nonlinear filtering of multiplied and convolved signals," *Proc. IEEE,* vol. 56, pp. 1264–1291, Aug. 1968.

15. Noll, A. M., "Cepstrum pitch determination," *J. Acoust. Soc. Am.,* vol. 41, pp. 293–309, Feb. 1967.

16. Rabiner, L. R., Schafer, R. W., and Rader, C. M., "The Chirp

z-transform algorithm and its application," *Bell System Tech. J.*, vol. 48, pp. 1249–1292, May–June 1969.

17. Tomlinson, R. S., "SPASS—an improved terminal-analog speech synthesizer," *J. Acoust. Soc. Am.*, vol. 38, p. 940(A), 1965.

18. Fant, G., Martony, J., Rengman, U., and Risberg, A., "OVE II synthesis strategy," Paper F5, Speech Commun. Seminar, Stockholm, 1962.

19. Flanagan, J. L., "Note on the design of terminal-analog speech synthesizers," *J. Acoust. Soc. Am.*, vol. 29, pp. 306–310, 1957.

20. Coker, C. H., and Cummiskey, P., "On-line computer control of a formant synthesizer," *J. Acoust. Soc. Am.*, vol. 38, p. 940(A), 1965.

21. Holmes, J. N., "Notes on synthesis work," Speech Transmission Lab. Quart. Progr. Rept., Stockholm, Apr. 1961.

22. Rabiner, L. R., "Digital-formant synthesizer for speech synthesis studies," *J. Acoust. Soc. Am.*, vol. 43, pp. 822–828, Apr. 1968.

23. Rosenberg, A. E., Schafer, R. W., and Rabiner, L. R., "An investigation of the effects of smoothing and quantization of the parameters of formant coded speech," to be presented at 80th Meeting of the Acoustical Society of America.

24. Lee, F. F., "Reading machine: from text to speech," *IEEE Trans. Audio and Electroacoustics*, vol. AU-17, pp. 275–282, Dec. 1969.

25. Houde, R. A., "A study of tongue body motion during selected speech sounds," Doctoral dissertation, Univ. of Michigan, 1967.

26. Teranishi, R., and Umeda, N., "Use of pronouncing dictionary in speech synthesis experiments," *Reports of the 6th ICA*, vol. II, pp. 155–158, 1968.

27. Vanderslice, R., "Synthetic elocution—consideration in automatic orthographic to phonetic conversion of English with special reference to prosodic features," Working Papers in Phonetics 8, U.C.L.A., Feb. 1968.

28. Allen, J., "Machine-to-man communication by speech. Part 2: Synthesis of prosodic features of speech by rule," *1968 Spring Joint Comput. Conf., AFIPS Proc.*, vol. 32. Washington, D.C.: Thompson, 1968, pp. 339–344.

29. Pike, K. L., *The Intonation of American English.* Ann Arbor: Univ. of Michigan Press, 1945.

30. Rosenberg, A. E., "A computer-controlled system for the subjective evaluation of speech samples," *IEEE Trans. Audio and Electroacoustics*, vol. AU-17, pp. 216–221, Sept. 1969.

31. Rosenberg, A. E., "Effect of glottal pulse shape on the quality of natural vowels," *J. Acoust. Soc. Am.*, in press.

32. Hall, J. L., "Maximum-likelihood sequential procedure for estimation of psychometric functions," *J. Acoust. Soc. Am.*, vol. 44, p. 370(A), 1968.

33. Flanagan, J. L., "Use of an interactive laboratory computer to study an acoustic-oscillator model of the vocal cords," *IEEE Trans. Audio and Electroacoustics*, vol. AU-17, pp. 2–6, Mar. 1969.

James L. Flanagan (F) received the B.S. degree from Mississippi State University in 1948 and the S.M. and Sc.D. degrees from the Massachusetts Institute of Technology in 1950 and 1955 respectively, all in electrical engineering. He was a member of the electrical engineering faculty at Mississippi State from 1950 to 1952, returning to M.I.T. for doctoral study as a Rockefeller Foundation Fellow. After joining Bell Laboratories in 1957, he became head of its Speech and Auditory Research Department in 1961 and of the Acoustics Research Department in 1967. His interests have centered on voice communication and digital techniques for signal analysis and transmission. He holds patents in speech coding, digital processing, and underwater acoustics.

Dr. Flanagan has written a number of technical papers and a book, "Speech Analysis, Synthesis and Perception." He is chairman of the IEEE Audio and Electroacoustics Group and is a Fellow of the Acoustical Society of America. He is a member of the Committee on Hearing and Bioacoustics of the National Academy of Sciences, the Sensory Aids Subcommittee of the National Academy of Engineering, Tau Beta Pi, and Sigma Xi.

Cecil H. Coker (SM) is a supervisor in the Acoustics Research Department at Bell Laboratories. He received the B.S. and M.S. degrees in electrical engineering from Mississippi State University in 1954 and 1956 respectively, and in 1960 he received the Ph.D. degree in electrical engineering, with a minor in physics, from the University of Wisconsin. He then taught for a year in the university's Electrical Engineering Department. Dr. Coker joined Bell Laboratories in 1961 and worked for several years on formant analysis and synthesis of speech. Subsequent work included supervision of the development of two laboratory computer facilities. His work on modeling of the articulatory process was begun in 1966. He holds several patents and has written a number of papers on speech analysis and synthesis.

Lawrence Rabiner (M) received the S.B. and S.M. degrees simultaneously in 1964 and the Ph.D. degree in electrical engineering in 1967, all from the Massachusetts Institute of Technology. From 1962 to 1964 he participated in the cooperative plan in electrical engineering at Bell Laboratories, in Murray Hill and Whippany, N.J. He worked on digital circuitry, military communications problems, and problems in binaural hearing. At present he is engaged in research on speech communications and digital signal-processing techniques at Bell Laboratories. He is a member of Eta Kappa Nu, Sigma Xi, Tau Beta Pi, and the Acoustical Society of America.

Ronald W. Schafer (M) received the B.Sc.E.E. and the M.Sc.E.E. degrees in 1961 and 1962 from the University of Nebraska, Lincoln. In 1968 he earned the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology. While attending M.I.T. he was associated with its Electronic Systems Laboratory in 1965 and, between 1966 and 1968, he was a member of its Research Laboratory of Electronics. He joined the Acoustics Research Department of Bell Laboratories in 1968 and is now engaged there in research on speech analysis and synthesis, and digital signal-processing techniques.

Noriko Umeda received the B.S. and the M.S. degrees in linguistics from the University of Tokyo, Japan, in 1957 and 1959. She completed the doctorate course requirements in the same field and at the same university in 1962. That year, she joined the research staff of the Electrotechnical Laboratory, Ministry of International Trade and Industry, Japan, and there worked on speech analysis and synthesis—which included the problem of converting printed text to speech. She joined Bell Laboratories after taking a leave of absence from Electrotechnical in 1969. Her primary interest has been to study speech phenomenons (articulatory, prosodic, and syntactic) using the technique of synthesis by rule.

Integrated-circuit digital logic families

I—Requirements and features of a logic family; RTL, DTL, and HTL devices

In this first part of a three-part article comparing the major IC logic families, the emphasis is on three types: resistor–transistor, diode–transistor, and high-threshold logic

Lane S. Garrett Motorola Semiconductor Products Inc.

The purpose of this article is to categorize the needs and desired features that a logic or system designer should consider when selecting a family of digital integrated circuits. The various major digital IC families are evaluated and compared with these needs in view. The intended results are that the reader will gain facility in IC technology, terminology, and usage in various areas of application. Part II of this article will cover transistor–transistor logic (TTL) devices. Part III is devoted to emitter-coupled logic (ECL) and metal oxide semiconductor (MOS) devices.

Each user of digital logic has different needs, which vary in importance with each design and application. The eight major requirements and desired features of a logic family are as follows:

1. Logic flexibility
2. Speed
3. Availability of complex functions
4. High noise immunity
5. Wide operating-temperature range
6. Low power dissipation
7. Lack of generated noise
8. Low cost

In the following sections, these needs will be examined individually.

Logic flexibility

Logic flexibility is a measure of the capability and versatility or the amount of work or variety of uses that can be obtained from a logic family; in other words, it is a measure of the utility of a logic family in meeting various system needs.

Logic flexibility can be compared between the families on the basis of several factors, as follows:

1. Wired logic capabilities
2. Complement outputs
3. Line-driving capability
4. Indicator driving
5. I/O interfacing
6. Driving other logic forms
7. Multiple gates

Wired logic refers to the capability of tying the outputs of gates or functions together to perform additional logic without extra hardware or components. Examples of two types of output circuitry common in integrated circuits are illustrated in Figs. 1 and 2. Note how both the positive logic AND and OR functions can be obtained, without the use of additional gates, simply by tying the outputs of the gates together.

Frequently, both a variable and its complement are required in a logic system. If the family of logic being used has gates with complement outputs, the need for inverters can be avoided. Gates with complementary outputs perform the OR/NOR and AND/NAND functions, as shown symbolically in Fig. 3.

Also included in the category of logic flexibility is the need for driving nonstandard loads, such as long signal lines, lamps, and indicator tubes.

In the design of a logic system, the gate count is minimized if AND, NAND, OR, NOR, and "exclusive OR" gates (Fig. 4) are all available in the family. This is not always the case; but the more types of gate included, the easier the design implementation.

Logic flexibility can be summed up as the capability of the family to meet the needs of the logic and system designer and to provide a wide selection of building blocks.

Speed

An increasingly important requirement in many areas is for getting more things done in the minimum amount of time. The simplest way to increase machine capability

logic families

is to speed up the logic and information access rates. Moreover, to accomplish a given task in a given time it may be possible to go to a high-speed serial implementation instead of a parallel design with a larger number of functions. This approach obviously can save money through the use of fewer gates.

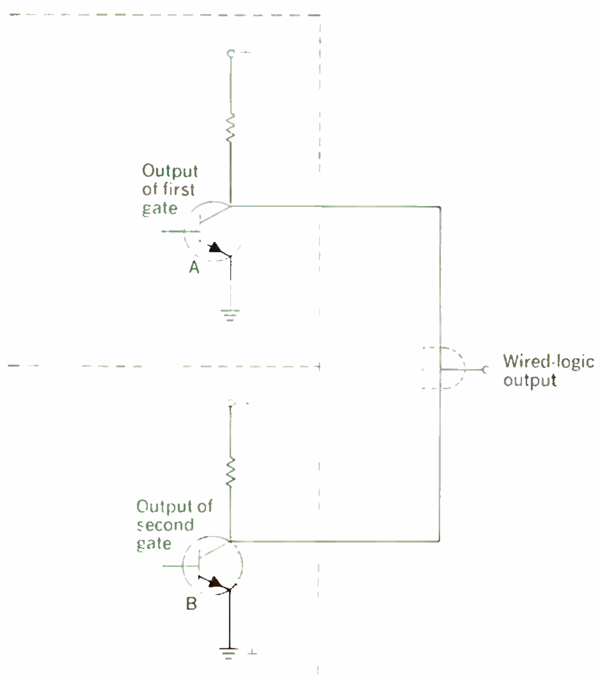
The most important measures of speed are gate propagation delay and pair delay. The latter may be defined as the propagation delay through two successive inverting gates. Pair delay averages out propagation delays for positive- and negative-going output transitions, which may be unequal. Another important factor is the maximum shift frequency of flip-flops, which may be significant in determining the maximum cycle time of a system.

Complex functions

Complex functions are becoming more and more common as the sophistication of designs increases. A complex function may be described as a grouping of basic gates for performing a relatively high level of integration, usually involving more than 12 but less than 100 gates.

As complexity increases, the number of input/output pins also increases—but usually at a decreasing rate.

FIGURE 1. Example of "implied AND" or "○ AND" wired logic (frequently but improperly called "wired OR.") The output is high (most positive level) if transistors A and B are off. Conversely, we may say that the output is low if A or B is on (negative logic).



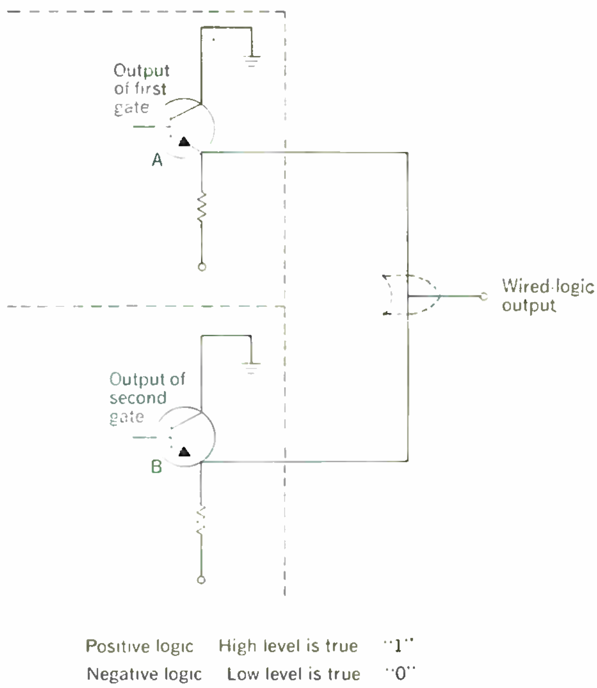


FIGURE 2. Example of "implied OR" wired logic. The output is high if transistor A or B has a high output. Conversely, the output is low if A and B outputs are both low (negative logic).

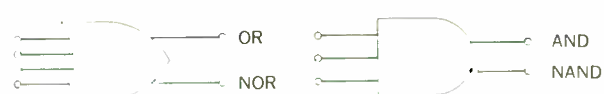
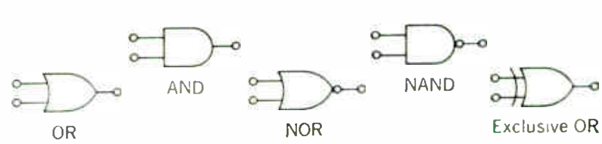


FIGURE 3. Complementary outputs. Standard symbols for OR/NOR and AND/NAND gates.

FIGURE 4. Standard symbols for various gate types.



Gate-to-pin ratios that normally increase with complexity give the benefit of decreasing assembly costs per gate while also increasing the reliability per gate.

At times, multiple groupings of a common logic function are available in the same package. A better understanding of the term "complex function" is obtained by noting several examples:

A parity tree is a grouping of gates designed to perform the "exclusive OR" function; the exclusive OR's are then grouped to form a logic tree. A typical example is the eight-bit parity tree, which checks eight different binary bits at a time to see whether the sum is odd or even. (Refer to Fig. 5, where an eight-bit parity tree is shown in logic-notation form.) A quad latch (Fig. 6) is a four-bit storage element, frequently used as buffer storage or temporary memory. Each of the four elements employs several gates. A "quad D" is a complex function containing four D-type flip-flops, which logically delay data by a clock period and also are useful in forming shift registers. A typical quad-D block diagram is illustrated in Fig. 7.

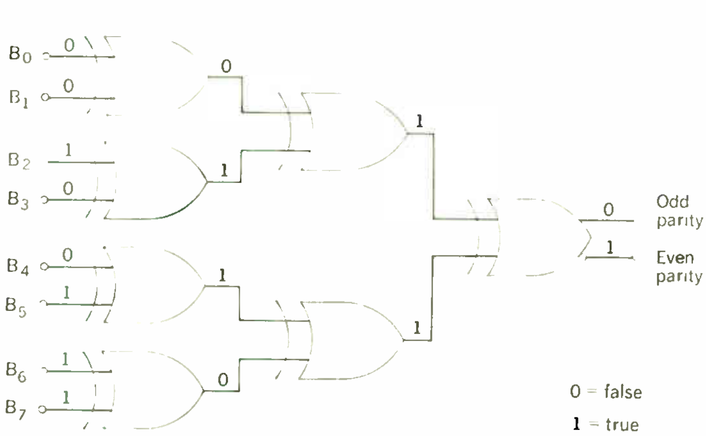
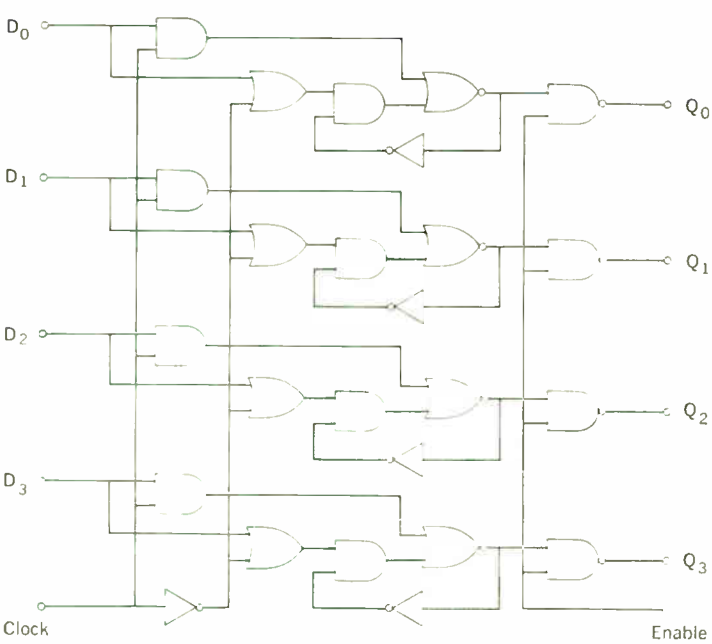


FIGURE 5. Eight-bit parity tree. The output of each exclusive OR is true whenever the inputs are different. The odd output is true whenever the sum of B_0 through B_7 is odd. The even output is true when the sum is even.

FIGURE 6. Block diagram of quad latch.



Counters also may be classified as complex functions. They are groupings of flip-flops and gates, which perform such functions as "divide by 10," "divide by 16," or "divide by *N*." Counters may also be able to count up or down according to logic-control inputs. Three examples are shown in Fig. 8.

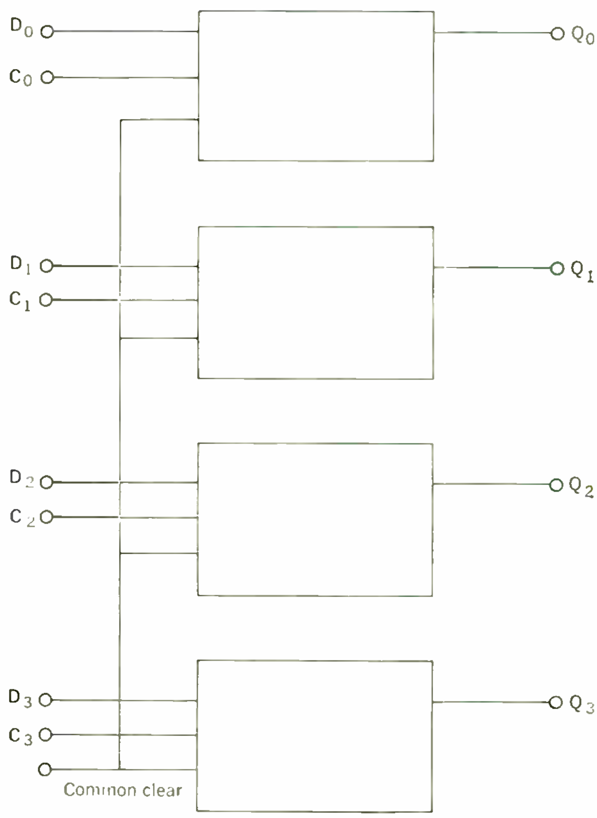
Decoders, such as the one shown in Fig. 9, may hold one of ten output lines at a low level based upon the states of four input control lines. The channel or data selectors comprise another group of complex functions. A typical device is an eight-channel data selector that picks one of eight input lines and passes its information on to the output. The input selected is determined by the states of the three input-control lines (see Fig. 10).

Complex functions, which are rapidly increasing in types and numbers in the major logic families, represent a second generation of complexity in the integrated-circuit art and often use the latest developments in technology, such as double-layer metalization. They are frequently the key to minimizing system size, cost, and parts count. Many IC families are chosen on the basis of the complex functions available.

High noise immunity

In order to prevent the occurrence of false logic signals in a system, high immunity to noise is desired. Erroneous signals may be caused by switching transients, excessive coupling between signal leads, and external sources such as relays, circuit breakers, and powerline transients. In general, the higher the noise immunity of the circuits, the

FIGURE 7. Block diagram of quad-D function.



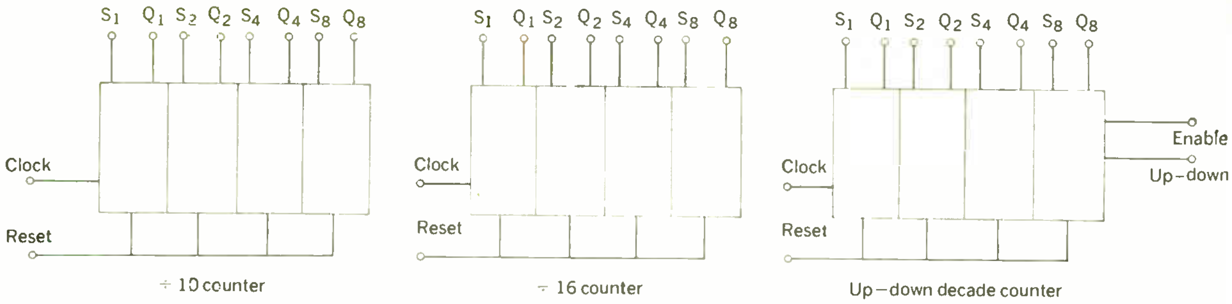
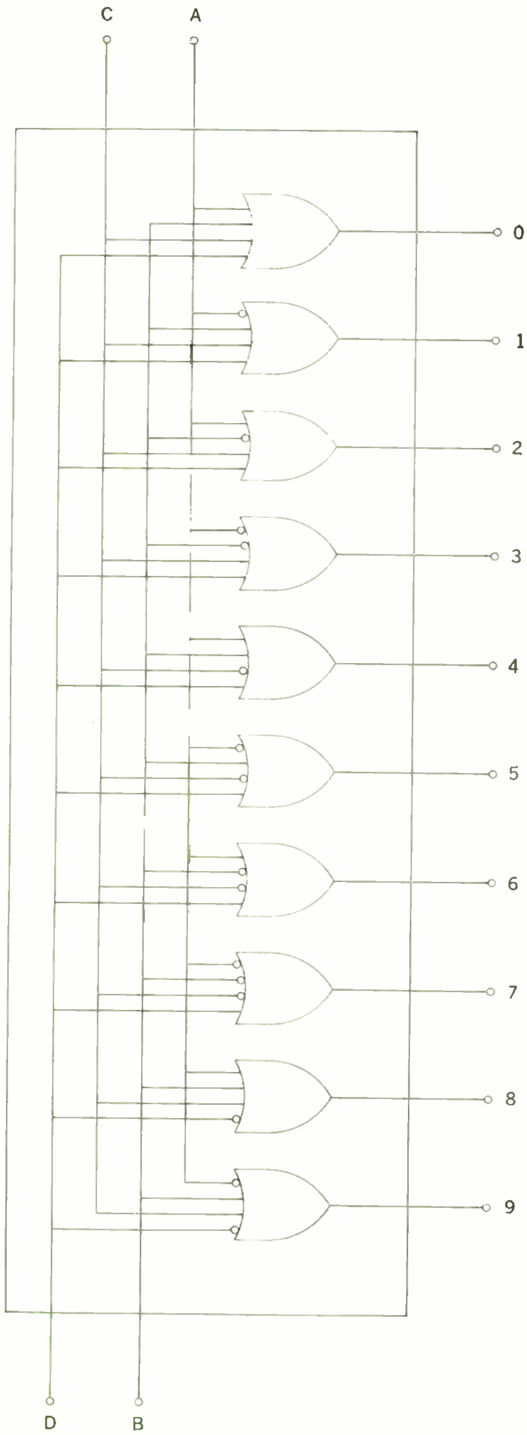


FIGURE 8. Three basic counters.

FIGURE 9. Four-to-ten-line decoder.



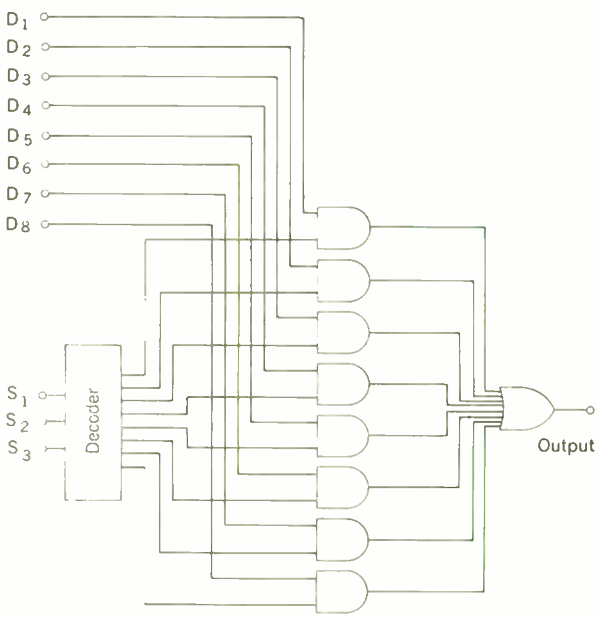
fewer the precautions required to prevent false logic signals. This becomes an important advantage in those areas, such as in small industrial logic control systems, that are subject to high noise levels. Voltage-noise immunity, or noise margin, is normally specified in terms of millivolts or volts—more specifically, the amount of voltage that can be added algebraically to a worst-case output level before a worst-case gate tied to that output will begin to switch. Noise margins for two gates are illustrated in Fig. 11, in which

$V_{OH(\min)}$ = minimum specified output voltage when the gate is in the "high" state

$V_{OL(\max)}$ = maximum specified output voltage when the gate is in the "low" state

$V_{IH(\min)}$ = minimum specified high-level input voltage at

FIGURE 10. Eight-channel data selector.



Input select	Data line selected
$\bar{S}_1$ $\bar{S}_2$ $\bar{S}_3$	D_1
S_1 S_2 S_3	D_2
S_1 S_2 $\bar{S}_3$	D_3
S_1 S_2 S_3	D_4
S_1 $\bar{S}_2$ S_3	D_5
S_1 $\bar{S}_2$ S_3	D_6
S_1 S_2 $\bar{S}_3$	D_7
S_1 S_2 S_3	D_8

which the gate output voltage will still meet specified limits

$V_{IL(max)}$ = maximum specified low-level input voltage at which the gate output voltage will still meet specified limits

The high-level and low-level noise margins are, respectively, $[V_{OH(min)} - V_{IH(min)}]$ and $[V_{IL(max)} - V_{OL(max)}]$.

Wide operating-temperature range

A wide operating-temperature range is always desired and is often a design requirement. For commercial and industrial needs, temperatures usually range from 0°C or −30°C to 55, 70, or 75°C. The military has an almost universal requirement for operability from −55°C to +125°C. In most cases, a logic line specified for −55°C to +125°C will exhibit better characteristics at room ambient conditions than a line specified for commercial requirements; that is, fanout, noise immunity, and tolerance to power supply variations are usually better, since the circuits must still be within specifications even when the inherent degradation due to temperature extremes occurs. The advantages of wide temperature specification are often offset by increased IC cost.

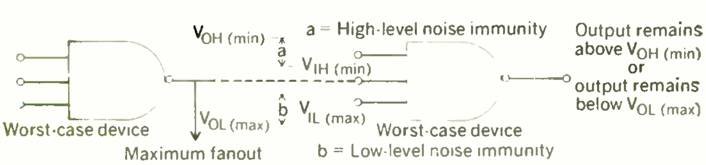
Low power dissipation

Logic with low power dissipation is desired in large systems because it lowers cooling costs and power supply and distribution costs, thereby reducing mechanical design problems. In an airborne or satellite application, power dissipation may be the most critical parameter because of power-source limitations. Although power dissipation may not be a large financial factor in the economic design of a system, it must certainly be considered along with other factors, such as logic speed. As chip complexity and packing density continue to increase, power dissipation will decrease on a per-gate basis. This is dictated by heat-dissipation restrictions arising from system design and maximum allowable semiconductor junction temperatures.

Minimum noise generation

The lack of generated noise is an important requirement. All power-supply leads in a system must be bypassed. The amount of capacitive bypassing and methods of power supply and ground distribution depend heavily upon the form of logic utilized. Supply distribution is less expensive if the logic family generates minimal noise. Also, maximum line lengths in the back plane and wiring of the computer are a function of crosstalk generated by the logic when it switches states. The ideal would be a logic family that draws constant current in either a logic 0 or 1 state and does not change supply drain current when switching states. Also, signal line currents would approach 0 with very slow rise and fall times.

FIGURE 11. Noise margins for two gates.



Low cost

The last consideration, and often the most important one, is the cost of a given logic family. The first approximate cost comparison can be obtained by pricing a common function such as a dual four-input or quad two-input gate. This is only approximate at best. Among other cost considerations are package count for a given system speed, layout and shielding costs, and interface costs. An effective cost comparison involves a paper design of a system with two or more different families. If the system employs a number of circuits, it is preferable to obtain factory quotes rather than rely on book prices.

Basic design problems

The question arises, why not design a family that best meets these needs and then mass produce it and drive costs down? This cannot be done, since unfortunately there is no universal logic family that does a good job of meeting all of the previously stated needs. Silicon technology, though better understood and studied than any other solid-state technology, still has limitations. These limitations, along with the inherent restrictions on circuit design, allow only some of the desired parameters to be optimized. First, the silicon wafers used in processing integrated circuits have faults in the crystalline structure. Although modern technology has greatly reduced their number, these faults are still a large contributing factor to the number of bad die or circuits on a given wafer of silicon. Photoresist, metalization, and photographic masks themselves are all sources of frequent errors and faults that contribute to the number of bad spots on a wafer.

The larger the size of a given die, the higher is the probability that it will incorporate a fault resulting in a bad circuit. Integrated-circuit bipolar die larger than about 80 mils (2 mm) on a side exhibit rapidly increasing costs, due to reduced yields. The fact that a lot fewer large die can fit on a wafer also increases expense. Thus, the complexity that can be obtained on a single chip is limited by simple economics. Silicon IC technology has limits in absolute values of resistivity obtainable, maximum voltage breakdowns, parasitic capacitance per given area, current density, and even operating temperatures.

In addition, there are several major restrictions on circuit design. For example, to design a given function requires a certain number of components, which in turn demand a given chip area. The best circuit design eliminates only redundant components, and further reducing components usually results in performance degradation, but this may be required due to die-size limitations. Good circuit design can alleviate some processing problems, such as increasing allowable transistor β variations or permitting larger changes in resistor values, but unfortunately this usually comes at the expense of more components. Moreover, bipolar design has a disadvantage in that current can effectively pass through the transistors in only one direction. In other words a transistor does not act as a conventional switch, which passes current in either direction. Metal oxide semiconductor devices on the other hand, have the advantage that once the device is turned on, current will pass in either direction. The problem here is that MOS devices have high "on" impedanees, usually a few thousand ohms, and impedance can be reduced only if the device is made large, thereby causing an area problem.

Another limitation on circuit design is the interrelationship of parameters. As transistor β is increased, collector breakdown voltage decreases, and the storage time of the device also increases. High-speed parameters require low circuit impedances, small devices, and narrow metalization, which in turn tends to be incompatible with the high currents that are required by the low-impedance circuitry. It is obvious that good circuit design involves extensive juggling of parameters. Different design parameters have been emphasized in each of the digital logic families. It will be seen that each logic family is geared to its own particular market, meeting a specific set of needs.

Most of the semiconductor digital efforts are centered in seven major categories, each with its own inherent advantages and disadvantages. Let's look at each of the seven major categories briefly before going into detail on each individual category. The first three of these categories will be discussed in this month's installment.

1. Resistor–transistor logic (RTL) results in the smallest die size (minimum space on a silicon wafer) for standard bipolar functions. It is easy to process and has low to medium power dissipation. RTL is known primarily for its economy.

2. Modified diode–transistor logic (DTL) is low in cost, has logic familar to most designers, is available from many sources, and can be used for most general-purpose designs.

3. High-threshold logic (HTL) is designed for noise immunity and finds application in industrial environments and locations likely to have high electrical noise levels. It is noted for its ability to interface easily with discrete devices and electromechanical components.

4. Transistor–transistor logic (TTL) has characteristics that are similar to DTL, and is noted for many complex functions and the highest available speed of any saturated logic. Moreover, many sources are available with an excellent rate of complex-function introductions.

5. Emitter-coupled logic (ECL) is known for performance and logic flexibility. It has the highest speed of any of the logic forms.

6. Metal oxide semiconductor (MOS) devices are noted for small die and large complex repetitive circuits, such as memories and shift registers. Power dissipation is low to moderate, with low cost per gate.

7. Complementary metal oxide semiconductor (CMOS) devices employ both p- and n-channel MOS components and are known for very low power dissipation with moderate delay times.

The characteristics of a family of digital ICs are set primarily by the design of the basic gate. A simple four-input gate will be used as the common dominator for comparisons between families. Input, transfer, and output characteristic curves are essential in the understanding of device parameters and will be illustrated. Gates will be discussed using MIL-STD-806B conventions. A high level (*H*) represents the most positive or least negative logic level, whereas *L* represents the most negative logic level. A high- and low-level table (as opposed to a truth table) will be shown for the various gates. Gates are named for their positive logic function; that is, a high level is taken as true. The positive logic equation generated by the gate is shown along with the equivalent or "dual" negative logic equation; the latter is generated by calling the low entries in the level table true. The saturated logic families will be studied in order of their complexity, followed by ECL, MOS, and CMOS logic.

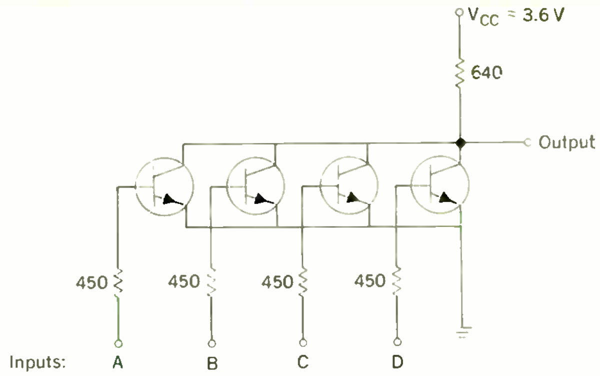


FIGURE 12. Four-input RTL gate.

FIGURE 13. Level table for four-input RTL gate.

D	C	B	A	Out.
L	L	L	L	H
L	L	L	H	L
L	L	H	L	L
L	L	H	H	L
L	H	L	L	L
L	H	L	H	L
L	H	H	L	L
L	H	H	H	L
H	L	L	L	L
H	L	L	H	L
H	L	H	L	L
H	L	H	H	L
H	H	L	L	L
H	H	L	H	L
H	H	H	L	L
H	H	H	H	L

RTL

Resistor–transistor logic is well named, since it contains resistors and transistors exclusively. The schematic diagram of a four-input RTL gate (Fig. 12) and its level table (Fig. 13) illustrate basic gate operation. The gate may be symbolized as a four-input NOR (positive logic) or a four-input NAND (negative logic), as shown in Fig. 14. In Fig. 12, it can be seen that if input A goes to a high level, current will flow through the 450-ohm input resistor into the base of the first transistor, turning it on. The input voltage is then the sum of the base–emitter diode drop (about 0.75 volt) and the drop across the input resistor. Note that if any one of the inputs, or even if all inputs, go to a high level, current will flow from V_{CC} through the 640-ohm resistor, resulting in a low-level output. Gain of the transistors is high, giving an output (V_{sat}) that is usually about 0.1 volt. The circuit performs the NOR function, since if one or more inputs are high the output is not high.

A unique situation occurs when *all* inputs are at a low level. Each of the transistors is turned off, and therefore the 640-ohm pull-up resistor is allowed to bring the output, and associated capacitance, to a relatively high voltage level. Thus, using negative logic notation, the circuit performs the NAND function; that is, when and only when *all* inputs are low the output is not low.

Figure 15 shows how the outputs of two gates would function when tied together. Assuming that each of the transistors has the capability of sinking the current through parallel pull-up resistors, the output node will be low if one or more of the transistors are turned on. The output will be high only when both transistors are off. This generates the positive logic AND function simply by tying outputs together. This connection is represented by the dashed AND gate drawn around the tie point to show that the function is "implied." This connection is sometimes improperly called the "wired OR," but should be called the "implied AND." Note that the resultant function is the AND of the two functions generated by each gate. For example, if

$$f_1 = \overline{A + B + C} \qquad \text{and} \qquad f_2 = \overline{D + E}$$

we would have

$$\text{Output} = (f_1)(f_2) = \overline{(A + B + C)}\,\overline{(D + E)}$$

or

$$\text{Output} = (ABC)(DE) = ABCDE$$
$$= \overline{A + B + C + D + E}$$

by the use of DeMorgan's rule. It is seen that we have effectively expanded the inputs of the RTL gate. This limits the usefulness of the "implied AND" capability of RTL except with some of the more advanced functions.

To provide a better understanding of the circuit, the input, transfer, and output characteristics are shown in Figs. 16–19. From the input characteristics (Fig. 16) it can be seen that the RTL gate starts to turn on and draw current when the input voltage reaches one diode voltage drop. RTL transistors are designed with fairly high β, and since the input resistor is fairly low in value, a slight increase in input voltage above a diode drop is sufficient to turn the pertinent transistor on completely. This is observed in Fig. 17, which shows the transfer characteristics for a fanout of 1; that is, the RTL output is supplying current to only one other RTL input. The maximum, typical, and minimum curves illustrate the normal distribution of devices caused by processing variations. Note that about 150 mV of change in the input voltage is sufficient to cause the output to switch completely. Figure 18 shows the transfer characteristics for a maximum RTL fanout of 5. Note how the high level is much more negative due to the loading of the five inputs. Figure 19 illustrates the high-level output characteristics in the region of RTL input voltages. The slope of the curve is simply the measure of the pull-up resistor value (640 ohms).

Figure 20 illustrates the equivalent circuit and curve for the gate output voltage. (Note, however, that the Fig. 20 curve is based on a V_{CC} value of 3.6 volts, whereas Figs. 16–19 are based on a V_{CC} of 3.0 volts.) At a fanout of 5 the curve becomes flat, giving a relatively constant voltage drop across the 640-ohm resistor. There is now about 3.9 mA available to be divided among the various gate inputs. The input resistors help the various gate inputs to share the input current evenly; otherwise a low threshold gate would "hog" more than its share of input current. Due to temperature variation, differences in absolute resistor values, and variations in β and input

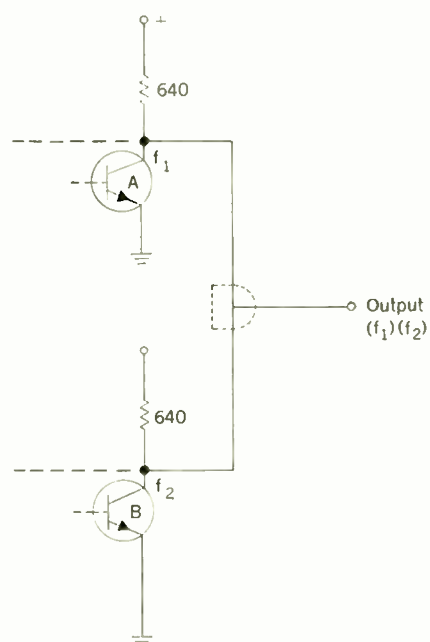


FIGURE 15. Gate operation with two outputs tied together.

FIGURE 14. Symbolic representations of Fig. 12 gate.

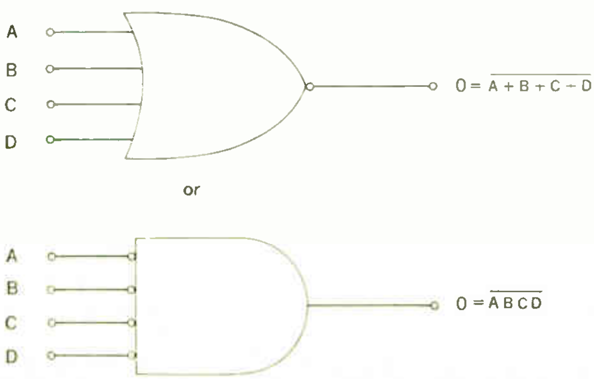
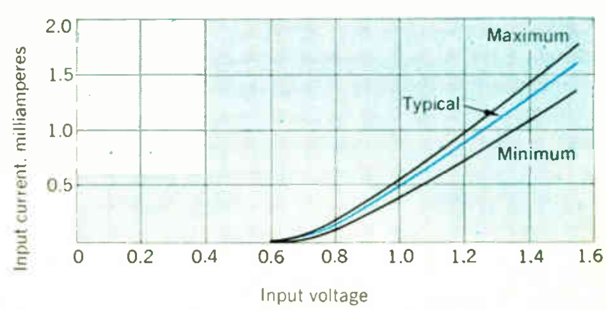


FIGURE 16. RTL input characteristics. V_{CC} = 3.0 volts, T_A = 25°.



thresholds, RTL is limited to a worst-case fanout of 5. Note that as fanout increases, the base overdrive decreases, which reduces energy noise immunity.

High voltage-noise immunity for 3-volt operation and maximum fanout can be obtained from Fig. 18. It takes a maximum input of 0.80 volt or less to cause a low-β transistor to switch the output below a maximum threshold. The worst output voltage is 0.90 volt at a fanout of 5. The result is a voltage-noise margin of greater than 100 mV at 25°C. Energy-noise immunity is better than anticipated due to the low output impedance. Low-level noise immunity is better because of the greater difference between the saturation voltage and the lowest threshold. As temperature increases, the gate threshold decreases at about 1.5 mV/°C, resulting in more constant thresholds than found in DTL and TTL designs. Transfer characteristics at a fanout of 1 and 5 versus temperature are shown in Figs. 21(A) and (B). Note that the thresholds and the high-level output decrease with increasing temperature. This effect is attributable to diffused resistor values that increase with temperature and the "clamping" effect of the gate inputs.

The basic RTL gate is easily fabricated into flip-flops and complex functions. In fact, RTL design results in the simplest and smallest-area bipolar method of obtaining complex functions. If many RTL gates are fabricated on the same chip, all transistors will have the same thresholds. Thus the "current hogging" problem mentioned previously is eliminated. For this reason, direct-coupled logic (RTL without input resistors) shows excellent promise for LSI arrays. If comparable device and processing technology are employed, the speed–power product of RTL is actually better than that of TTL—a factor that again shows promise for LSI functions. In addition, the resistor values of the basic gate (Fig. 12) are easily increased to reduce power dissipation. An RTL version called milliwatt RTL has been used extensively in low-power applications. Characteristics are similar to and compatible with standard RTL at one fifth the power dissipation.

Like any other family, RTL has its advantages, disadvantages, and areas of preferable application. These are listed as follows for easy comparison with the other families:

RTL disadvantages

1. Low voltage-noise immunity

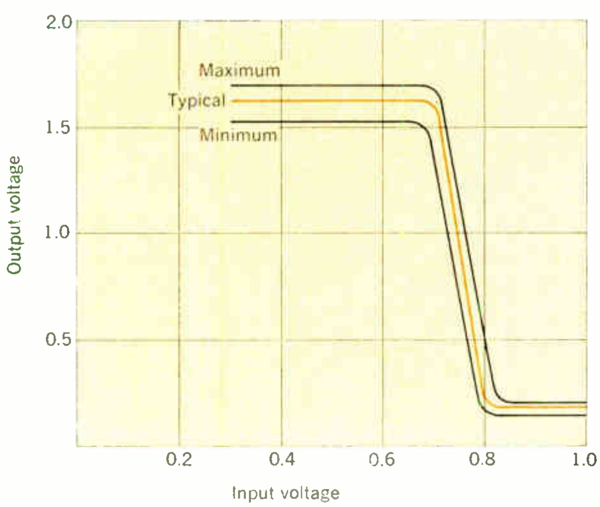
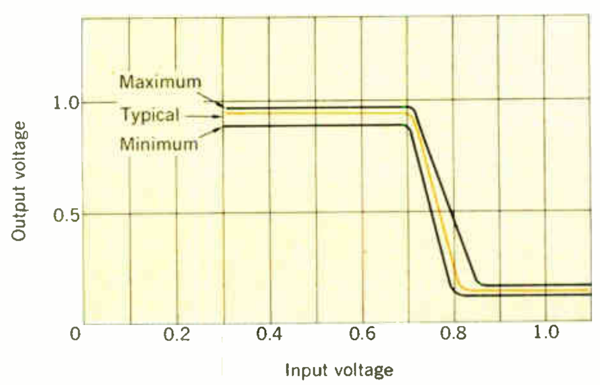


FIGURE 17. RTL transfer characteristics. V_{CC} = 3.0 volts, T_A = 25°C, fanout = 1.

FIGURE 18. RTL transfer characteristics. V_{CC} = 3.0 volts, T_A = 25°C, fanout = 5.



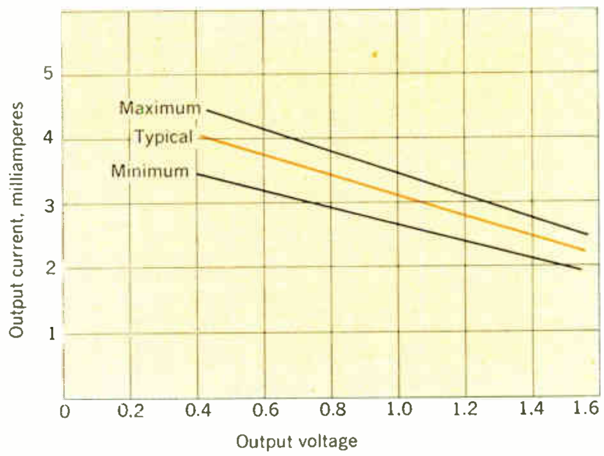
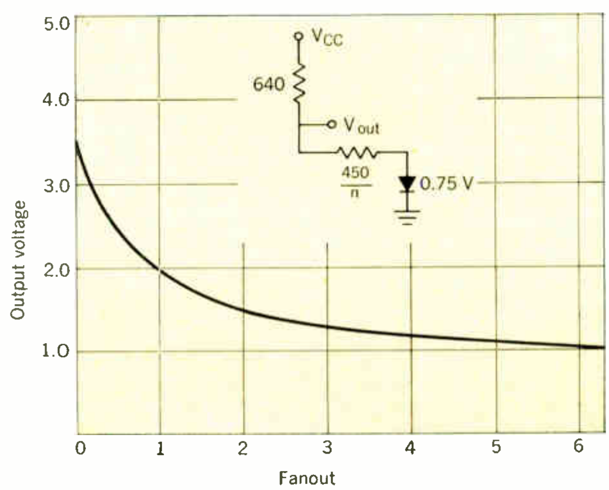


FIGURE 19. RTL output characteristics. V_{CC} = 3.0 volts, T_A = 25°C.

FIGURE 20. Output voltage vs. fanout and equivalent circuit for a typical RTL device. (n = fanout.)



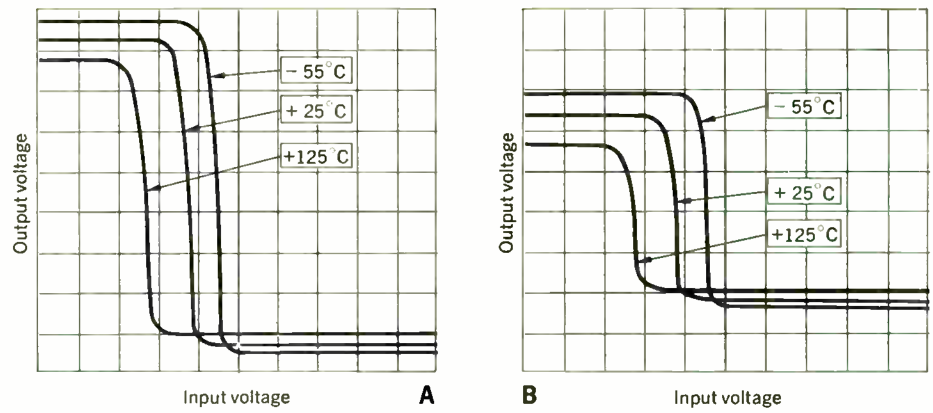


FIGURE 21. Modified RTL for three different temperatures. A—Fanout = 1. B—Fanout = 5. Scales: 200 mV/div.

2. Relatively low fanout

Advantages

1. Very low cost
2. Ease of design and manufacture
3. Ease of use in system designs
4. Very good speed–power product
5. Wide family of devices available. AND, OR, NAND, NOR, exclusive OR, exclusive NOR, flip-flops, and complex functions are now being manufactured
6. Smallest die size for a given amount of bipolar circuitry
7. Excellent promise for future bipolar LSI
8. Ease of interface with discrete components
9. Relatively low noise generation
10. Capability of tying most devices together to perform the "implied AND" function

Areas of application

1. Low-cost systems
2. Counters
3. Instrumentation
4. Small- to medium-size computers
5. Medical electronics
6. Data sets and computer peripherals
7. Printers
8. Calculators
9. Systems that are intermixed with discrete transistors
10. Military and aerospace systems (pertains primarily to R-13, a high-reliability specially specified form of RTL)

DTL

The schematic diagram shown in Fig. 22 represents the first form of integrated diode–transistor logic and is a carryover from discrete design, where transistors were expensive compared with diodes. Circuit operation is rather simple. If one or more of the inputs are brought to ground or a low level, current through the 2-kΩ input resistor will be shunted to ground, thereby eliminating base drive from the output transistor. Since this method keeps the transistor turned off, the output is maintained at a high level. A unique input condition occurs when all the inputs are high. Current will then flow through the 2-kΩ input resistor and the two standoff diodes into the base of the output transistor, turning it on. The saturation voltage of the output is low—normally 0.2 volt.

The *X* input simply stands for "expander," where additional diodes may be added to increase fan-in. The 20-kΩ pull-down resistor provides a discharge path for stored charge in the output transistor, thus speeding up the turn-off time of the transistor. The pull-down

FIGURE 22. Basic DTL gate design. Power dissipation = 10 mW, fanout = 4, propagation delay = 10 ns, 10-MHz flip-flops.

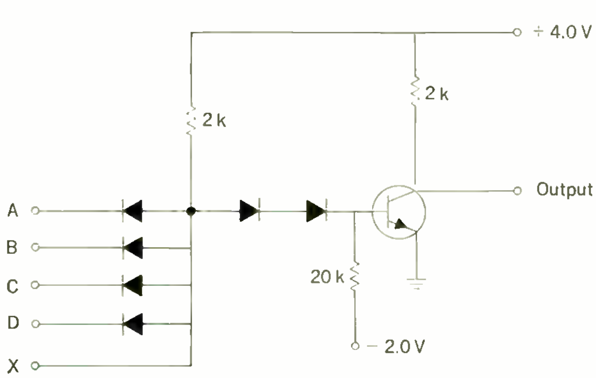
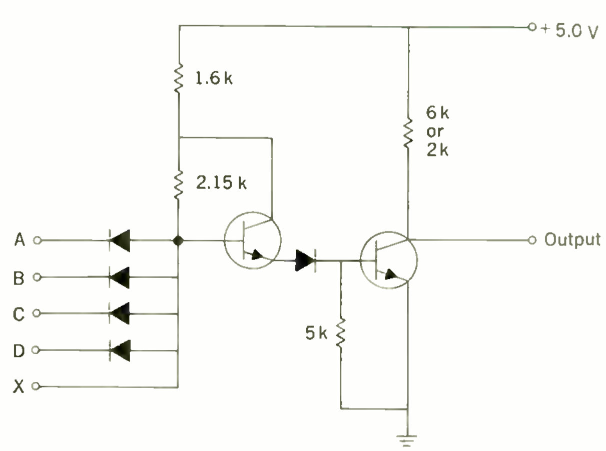


FIGURE 23. Modified DTL gate design. For 6-kΩ resistor, power dissipation = 8 mW, fanout = 8, propagation delay = 30 ns, frequency = 12 MHz. For 2-kΩ resistor, power dissipation = 12 mW, fanout = 7, propagation delay = 25 ns, frequency = 20 MHz.



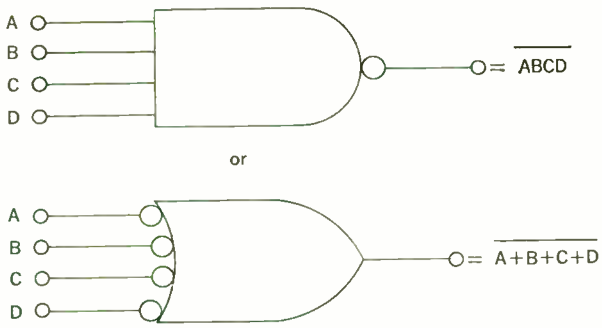


FIGURE 24. Modified DTL positive and negative logic symbols.

FIGURE 25. Modified DTL level table.

D	C	B	A	Out.
L	L	L	L	H
L	L	L	H	H
L	L	H	L	H
L	L	H	H	H
L	H	L	L	H
L	H	L	H	H
L	H	H	L	H
L	H	H	H	H
H	L	L	L	H
H	L	L	H	H
H	L	H	L	H
H	L	H	H	H
H	H	L	L	H
H	H	L	H	H
H	H	H	L	H
H	H	H	H	L

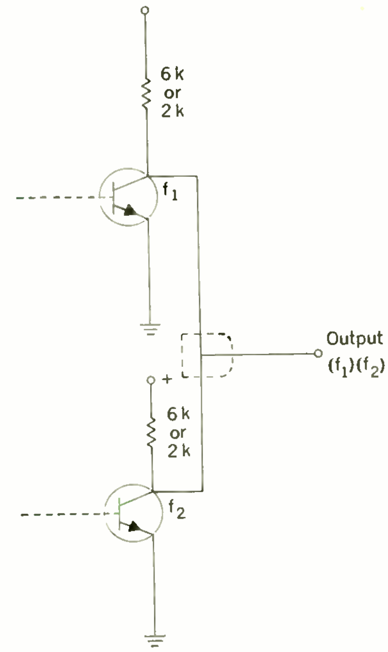


FIGURE 26. DTL gate operation with two outputs tied together.

resistor also helps to keep the transistor turned off for short positive-going input transients, thus aiding noise immunity.

The basic DTL gate design has been modified to lend itself more easily to IC processing and also to enhance performance characteristics. This design, shown in Fig. 23, is called modified DTL. The design has greatly increased functionality and is largely responsible for the wide acceptance of DTL. The input resistor has been split into two parts in the modified design. Typical input current drawn through an input diode is slightly more than 1 mA for the modified circuit and about 1.5 mA for the older circuit. On the other hand, the available base drive for the output transistor is increased in the modified circuit.

The input transistor acts as an emitter follower with collector current limited only by the 1.6-kΩ resistor. Base drive in the modified circuit is approximately 1.7 mA while only about 0.9 mA in the older circuit. Since the modified circuit provides more base drive for the output transistor (which, in turn, is required to sink a smaller current from the gates it may be driving), the minimum allowable β of the output transistors is lowered. For this reason the additional transistor is often referred to as a β saver. Thus the circuit modification provides a much better ratio of base-drive current to input current. This improvement allows the fanout to be doubled and also improves testing yields.

A 5-kΩ pull-down resistor, instead of the 20-kΩ value, takes up less area on the IC chip while still providing similar pull-down current. An important advantage to the user is that only one power supply is now required. The output pull-up resistor can be either 2 kΩ or 6 kΩ for a convenient speed–power tradeoff. Figures 24 and 25 give the positive and negative logic symbols and a level table, to describe both gate designs from a logic point of view.

Like RTL outputs, DTL outputs may be tied together to perform additional logic through the "implied AND" connection, which is illustrated in Fig. 26, with an example in Fig. 27. The implied AND permits both NAND and AND logic with DTL, which results in more flexibility than was present with RTL. Tying RTL NOR outputs together effectively expands the inputs.

Although DTL input impedance is very high for a high-level input (a reverse-biased diode), it is essentially the value of the input resistor for a low level. Input impedance changes at the threshold level, which is two diode voltage drops above ground. Output impedance is 6 kΩ or 2 kΩ for a high output level and R_{sat} for a low level. Since the input characteristics and current-sinking capabilities of DTL and TTL are very similar, their discussion will be postponed to the section devoted to TTL, which appears in Part II of this article.

Transfer characteristics for a typical DTL gate are shown in Fig. 28. Note that as the input voltage is increased from a low level, the high-level output will start to change at an input voltage of about 1.3 volts. At an input of 1.7 volts the output has completely switched to a low level. Noise immunity, from a voltage point of view, is easily specified from the transfer characteristics. $V_{IL(\max)}$, the maximum input low-level voltage, and $V_{IH(\min)}$, the minimum input high-level voltage, are the test voltages at which the worst-case output voltage levels are measured. With these test inputs, $V_{OH(\min)}$ and $V_{OL(\max)}$, the minimum and maximum low-level output voltages, respectively, are obtained. These points determine the specified voltage noise margin of the DTL circuit (930 series type).

Since the threshold voltage is equivalent to two diode voltage drops, changes will occur with temperature. Threshold decreases with increasing temperature by about 3 mV/°C. The resulting changes in worst-case low-level noise immunity NI_L and high-level noise immunity NI_H are as follows:

25°C (830 series)

$V_{IL(\max)}$ =	1.10 V	$V_{OH(\min)}$ =	2.60 V
$V_{OL(\max)}$ =	−0.45 V	$V_{IH(\min)}$ =	−1.90 V
NI_L =	0.65 V	NI_H	0.70 V

75°C (830 series)

$V_{IL(\max)}$ =	0.95 V	$V_{OH(\min)}$ =	2.50 V
$V_{OL(\max)}$ =	−0.50 V	$V_{IH(\min)}$ =	−1.80 V
NI_L =	0.45 V	NI_H =	0.70 V

125°C (930 series)

$V_{IL(\max)}$ =	0.80 V	$V_{OH(\min)}$ =	2.50 V
$V_{OL(\max)}$ =	−0.45 V	$V_{IH(\min)}$ =	−2.00 V
NI_L =	0.35 V	NI_H =	0.50 V

Note how the noise immunity decreases with increasing temperature.

DTL has many advantages and is used widely in the industry. The various disadvantages, advantages, and areas of application are as follows:

DTL disadvantages

1. Low noise immunity, especially in the high state because of the high output impedance. (2-kΩ pull-up is an improvement over 6-kΩ pull-up)
2. Rapid change in thresholds with temperature
3. Rapid slowdown with capacitive loading
4. Rather poor rate of complex-function introductions
5. Lower speed capabilities than those of some other families

Advantages

1. Output circuitry allows the implied AND. This can result in significant savings
2. Low power dissipation (especially with 6-kΩ pull-ups)
3. Compatibility with TTL (easily mixed)

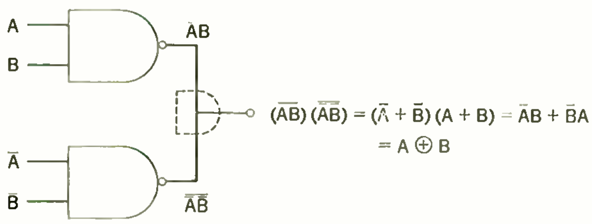


FIGURE 27. "Implied AND" example.

FIGURE 28. Modified DTL transfer characteristic (25°C).

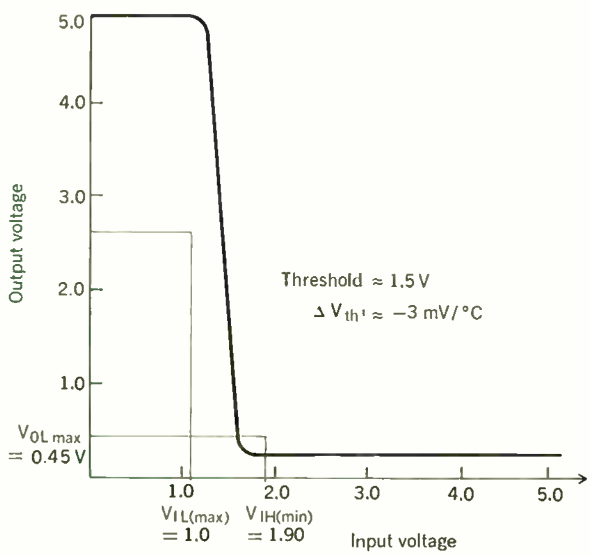
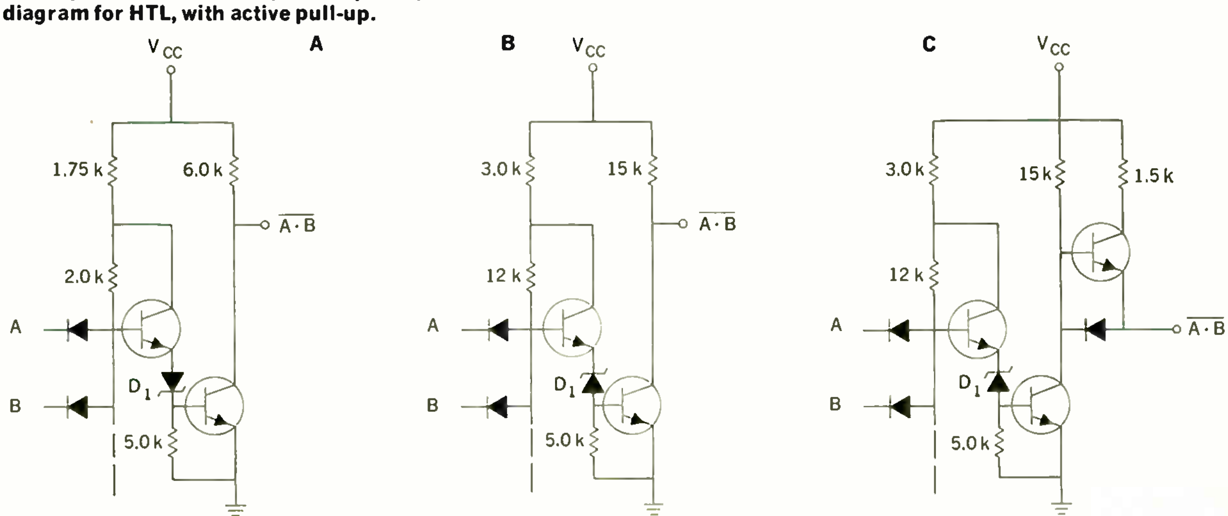


FIGURE 29. A—Schematic diagram for DTL. B—Schematic diagram for HTL, with passive pull-up. C—Schematic diagram for HTL, with active pull-up.



4. Low cost
5. Ease of manufacture
6. Wide range of positive logic devices available, including AND, NAND, OR, NOR, and exclusive OR
7. Ease of use in system designs
8. Ease of interface with discrete circuits
9. Low noise generation
10. Large number of sources in the industry
11. Good fanout

Areas of application

1. Small computers
2. Calculators
3. Instrumentation
4. Counters
5. Medical electronics
6. Computer electronics
7. Military systems
8. Control systems
9. Aerospace systems
10. Ground-support systems

HTL

There are many applications in noisy environments where logic with appreciably greater noise immunity than DTL is a requirement. High-threshold logic has been designed for these areas. Individual designs vary to some extent between manufacturers, but this form of logic is characterized by higher supply voltages, high noise immunity, and high thresholds obtained by adding a Zener breakdown voltage to the normal circuit diode drops. The circuit form is the same as DTL except for the Zener, which replaces a diode, and increased resistor values, which prevent excessive power dissipation. Figure 29(A) repeats the basic DTL gate for comparison to the HTL gate with a passive pull-up; see Fig. 29(B). Figure 29(C) shows an HTL gate with an active pull-up, which results in lower high-level output impedance.

The HTL gate operation is the same as for DTL except for voltage levels. The input diode drop and emitter-base voltage of the β saver cancel out, giving a threshold dependent upon the Zener and the base-emitter drop of the output transistor. The Zener, actually a base-emitter junction operating in the breakdown avalanche mode, has a voltage drop that is dependent on processing. For a 300-ohm-per-square base diffusion the "Zener" breakdown will be approximately 6.9 volts, which when added to the base-emitter turn-on potential gives a 7.5-volt gate threshold. Some designs eliminate the β saver and have a lower threshold, but the characteristics are similar. The designs shown in Fig. 29(B) and (C) permit a worse-case voltage-noise immunity of 5 volts—as can be seen from Fig. 30, which also shows the DTL transfer characteristics for the sake of comparison. Thresholds are constant with temperature, since the Zener diode has a positive temperature characteristic that matches the negative characteristic of the base-emitter junction. Actual HTL transfer characteristics are illustrated in Fig. 31. Note the "square" characteristics and relatively narrow threshold region.

The active pull-up design has both advantages and disadvantages. Since the high-level output impedance is much lower than with the passive pull-up, outputs should not be tied together. Therefore, the "implied AND" function is restricted to passive pull-ups. However, the active pull-up does give appreciable better energy-noise immunity because of its lower impedance. Moreover, the output diode eliminates the requirement for a "phase-splitter" transistor, which would drive the two output transistors out of phase; that is, only one transistor would be on at a time. This will be explained in the TTL section, which appears in next month's installment. The output diode has the disadvantage of giving a low-level output of a diode voltage drop plus V_{sat} of the transistor. Current sourcing and sinking characteristics are given in Figs. 32 and 33 respectively. The major HTL parameters for the β-saver type of design are as follows:

Temperature range = −30 to 75°C (special, −55 to 125°C)
Upper clock frequency ≈ 3 MHz
V_{CC} = 15 ± 1 volt (special, 18 volts)
Rise time (passive) ≈ 120 ns, 10% to 90%
Rise time (active) ≈ 100 ns, 10% to 90%
Fall time ≈ 30 ns, 10% to 90%

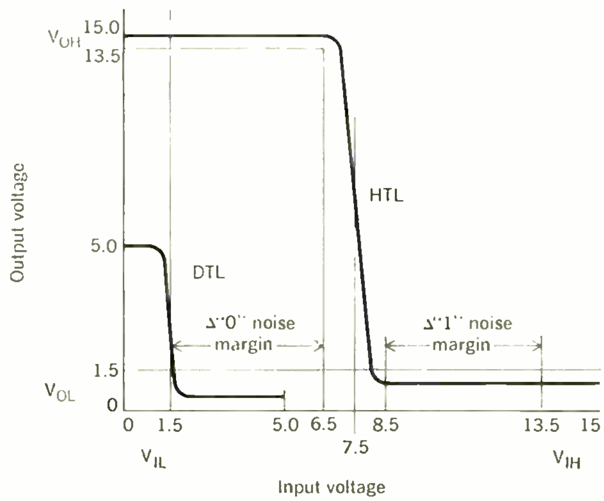
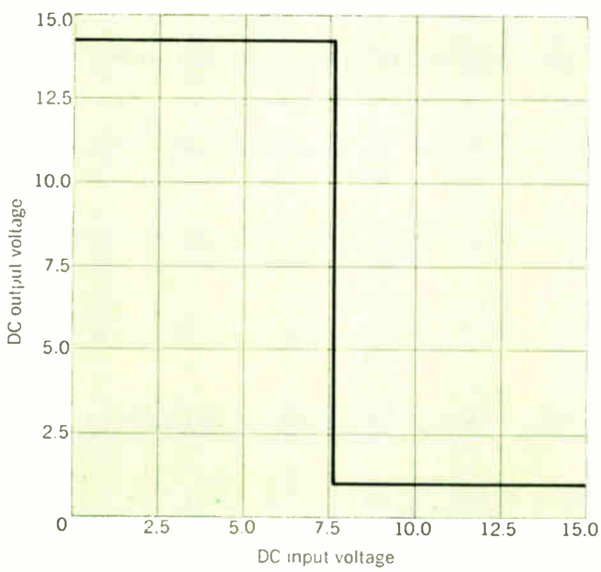


FIGURE 30. Worst-case noise margins. Temperature = −30°C to 75°C, V_{CC} = 15 volts.

FIGURE 31. HTL transfer characteristics. Active pull-up, T_A = 25°C, fanout = 1, V_{CC} = 15 volts.



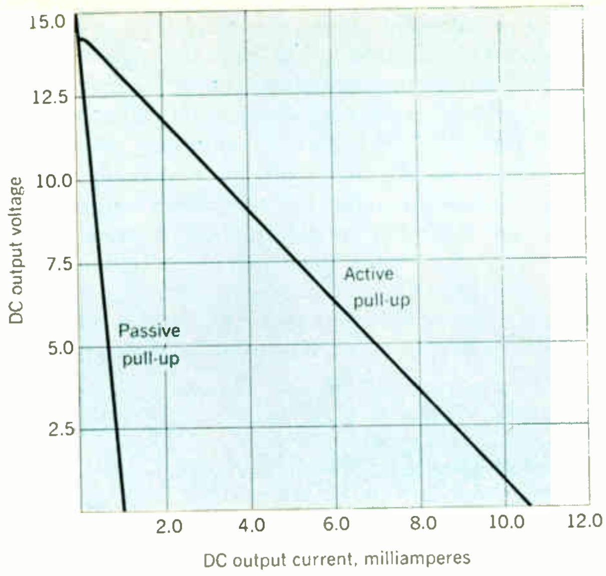
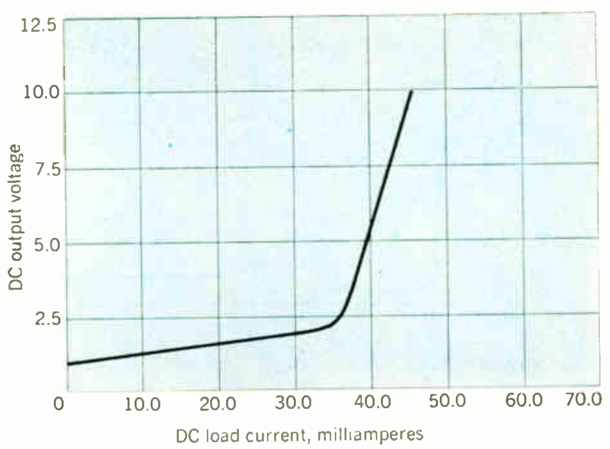


FIGURE 32. HTL output voltage vs. output current. $T_A = 25°C$, $V_{CC} = 15$ volts.

FIGURE 33. HTL output voltage vs. load current. $T_A = 25°C$, $V_{CC} = 15$ volts.



Power dissipation ≈ 55 mW	
Fanout = 10	Z_{in} (high) > 8 MΩ
Propagation delay ≈ 90 ns	Z_{in} (low) ≈ 15 kΩ
Z_o (high, active) ≈ 1.4 kΩ	Z_o (low) ≈ 30 ohms
Z_o (high, passive) ≈ 15 kΩ	Threshold ≈ 7.5 volts
V_{OH} (active) ≈ 14.3 volts	V_{OH} (passive) ≈ 15 volts
V_{OL} (active) ≈ 1.0 volt	V_{OL} (passive) ≈ 0.3 volt

HTL circuits are finding use in areas that previously belonged to discretes. They can be used with electromechanical components and operate from the higher-voltage supplies usually found in industrial environments.

HTL disadvantages

1. Uneconomical for large processing systems
2. Costs more to make than low-level logic because of the higher values of resistors, which take up more room on the chip
3. Needs more functions for the industrial market
4. Has relatively high power dissipation

Advantages

1. High voltage-noise immunity (6 volts typical, 5 volts worst-case)
2. 13- to 14-volt logic swing
3. Slow propagation delay (about 90 ns), which results in insensitivity to short noise pulses
4. The 15-volt power supply required is commonly available in industrial areas
5. −30°C to 75°C temperature range
6. Interfaces easily with discrete components
7. High power-switching capability
8. Highest energy-noise immunity
9. "Implied AND" capabilities with many functions
10. Can be used to drive long lines with 6 volts of common-mode noise immunity
11. Interfaces easily with linear functions such as operational amplifiers and multipliers
12. Constant threshold versus temperature
13. Interfaces easily with electromechanical components
14. New functions are being added to increase flexibility

Areas of application

1. Process control systems
2. Interface with linear circuits
3. Transducer interface
4. Numerical control systems
5. Industrial environments
6. Motor control systems
7. Interface with discrete components
8. Line driving and receiving
9. Replace relays
10. Solenoid valve control
11. SCR circuits
12. Telephone interface
13. Automated machinery
14. Premade printed-circuit boards for building-block systems

Lane S. Garrett (M) joined Motorola Semiconductor Products in 1965 as a senior applications engineer. In 1967 he became section manager of computer applications, in which capacity he worked on the application of integrated-circuit technology to the computer field. In 1968 he moved into production as project engineer for hybrid digital complex functions. From this position he became project manager of different IC families, including both standard and custom lines. In 1969 he received his present position of training manager of digital IC product marketing. Prior to joining Motorola he was with the U.S. Naval Air Development Center in Johnsville, Pa., where he was project leader on the design and fabrication of a special-purpose high-speed computer for testing analog-to-digital conversion devices.

Mr. Garrett received the B.S.E.E. degree from Drexel University in 1962. He has done graduate work at the University of Pennsylvania and Pennsylvania State University. He received the M.S.E. degree from Arizona State University in June 1970. He has previously published two IEEE papers and written numerous application notes and training courses for Motorola. He is a registered professional engineer in the state of Arizona and a member of the American Management Association.

Acoustic communication is better than none

Sound waves that are transmitted through the ocean are prone to bending and scattering between source and receiver. But, in contrast to radio waves, they can be used to convey intelligible information through water over considerable distances

Victor C. Anderson

Marine Physical Laboratory of the Scripps Institution of Oceanography

The transmission of electromagnetic energy through seawater is so heavily attenuated that it is necessary to turn to acoustic energy to transmit information over any appreciable range of ocean. In the sea, the absorption of acoustic energy is several orders of magnitude less than that of electromagnetic waves. Still, this absorption constitutes a limitation on the bandwidth of the communication link. In addition, the upper and lower boundaries of the ocean, and the multipaths caused by them, distort the communication path and further reduce the usable bandwidth.

Throughout the ages, man has gone down to the sea to confront the elements—lured by an almost irresistible urge that in many ways defies understanding. Lately, this urge has carried him down into the depths of the sea, again with a fascination that is difficult to fathom.

Perhaps one fascination is the "security blanket" that the sea affords him. In the depths, he is divorced from light; his vision is restricted severely—even with artificial light; and with his unaided ear, he can hear only poorly.

Whatever the fascination, once below the surface of the sea, man again finds a need to communicate with his fellowmen and to extend his senses to greater depths and greater distances so as to unravel the sea's mysteries. The blanket of water that was to have been his consolation becomes a barrier: This layer of water transmits neither light nor other electromagnetic energy any great distance. Man, therefore, is forced to return to his most primitive mode of communication: acoustics.

Better than radio waves

Unlike electromagnetic propagation in the atmosphere, the transmission of acoustic energy through the ocean is subject to absorption that is strongly frequency dependent. This frequency dependence is the dominant factor determining the usable bandwidth of an acoustic channel for any particular range in the ocean. Acoustic propagation loss is plotted in Fig. 1 as a function of frequency for three different distances. The dotted line in the figure represents the 5-dB octave slope of the classical Knudsen curves for deep-sea ambient-background noise. A bandwidth, about one decade in frequency around the point of tangency of the background-noise curve and the propagation-loss curve, defines an optimum operating-frequency band. It is apparent that the width of the band will depend on the operating range of the system.

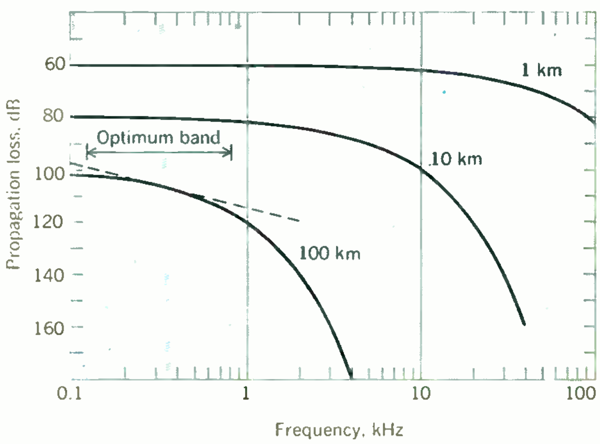


FIGURE 1. Acoustic propagation loss versus frequency for several ranges assuming inverse-square-law spreading and frequency-dependent absorption only.

Although this frequency-dependent attenuation imposes severe restrictions on the bandwidth of an acoustic communication channel in the ocean, the attenuation suffered by an electromagnetic wave would be many orders of magnitude greater. In contrast to the land domain where radio transmissions extend man's communication range from earshot to around the world and out into space, in the ocean depths the range of acoustic transmission far transcends that for electromagnetic communications.

Figure 2 illustrates the disparity between electromagnetic and acoustic communication in the sea. Considerable liberty has been taken in the use of simplifying assumptions for the purpose of comparison. However, the overall comparison is valid, although the range and bandwidths that have been plotted for the systems considered are far too general for quantitative use. From the diagram one can conclude that, at least for distances in excess of 300 meters, acoustic transmission channels in the ocean offer bandwidths four to five orders of magnitude broader than those associated with electromagnetic propagation. The acoustic link, on the other hand, is several orders of magnitude short of the bandwidth of a comparable length of coaxial cable.

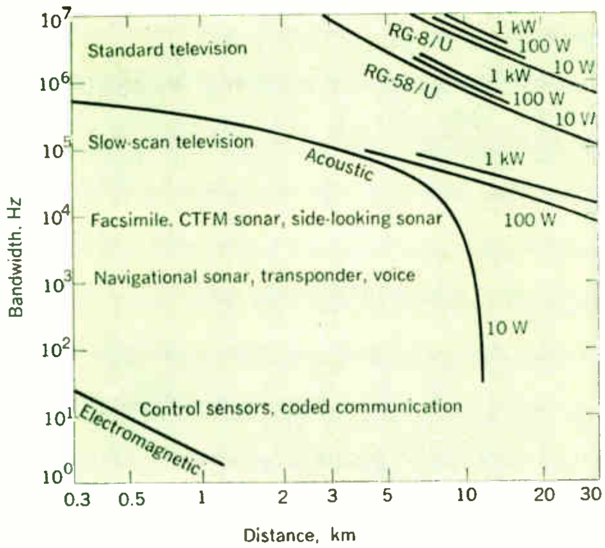
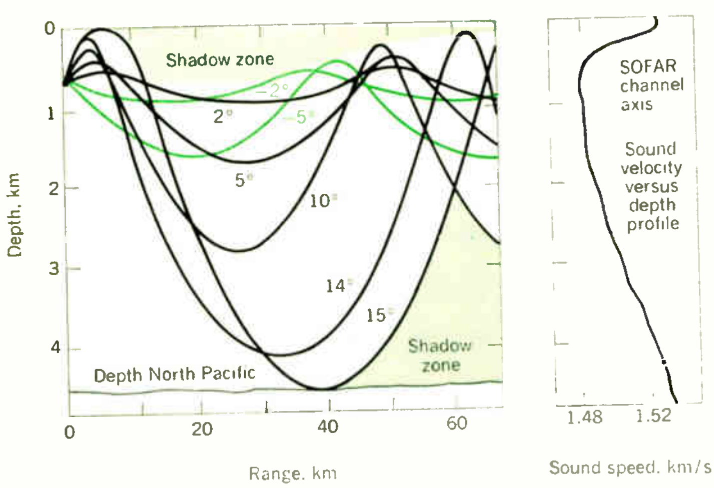


FIGURE 2. Bandwidth limitations of information channels in the ocean. Bandwidth requirements for information systems are also shown. The "RG" designated curves in the illustration refer to types of coaxial cable.

FIGURE 3. Refraction effects in the ocean. SOFAR is the acronym for SOund Fixing And Ranging and is a deep sound channel.



In discussing the "capricious" ocean medium, communication through it will be considered, in its broadest sense, as the transmission of information between two points. The variety of information transmission requirements is illustrated by the examples given in Fig. 2. The vertical positions at which the various forms of communication have been placed in the illustration correspond to the approximate bandwidth requirement for the communication link in each case. It is seen that bandwidth requirements range from the order of 10 Hz for control sensors and coded communication links to 10^5 Hz for slow-scan television systems—the latter taxing the maximum capability of acoustic links over any appreciable distance. The types of information listed in the illustration can be considered as representative of the practical requirements for communication in the ocean. (Other types of underwater acoustic systems—in particular, military and commercial sonars and echo sounders—are not listed because they are environmental probes rather than communicators that transmit information between two points.)

What happens

Of course, the simplified characterization of the acoustic transmission link represented in Fig. 2 falls far short of describing the limitations of acoustics for communication in the ocean. To look more closely at the problems besetting communication links, start by considering voice communication and transponder operations. These two types are related by both their bandwidth requirement—of the order of 1 kHz—and operating-range requirement —which is of the order of several hundred meters to a few kilometers.

For deep water where one can safely ignore the effect of either the bottom or the surface of the ocean—but not both, the primary limitation on operating range is the refraction of the sound waves in water. This effect is illustrated in the ray-path diagram of Fig. 3 in which a 700-meter-deep sound source is located at the depth of the minimum velocity point in the typical deep-water source-velocity profile. This depth coincides with the area of the so-called SOFAR deep sound channel.

Near the surface, a normal downward refraction, caused by the presence of negative thermal gradients in the upper layer of the water, occurs in many localities much of the time. Consequently, a near-surface shadow zone is created beyond the extreme ray that just grazes the surface.

For ray paths leaving a transducer at angles lower than that for the extreme ray, the range to a receiver located at the source depth is reduced. If sound is propagated over ray paths of higher angles, it is reflected at the air-water interface, and its carrying range again falls short of that for the maximum ray. At greater depths where the water is nearly isothermal, the pressure coefficient of sound velocity dominates the velocity profile and gives

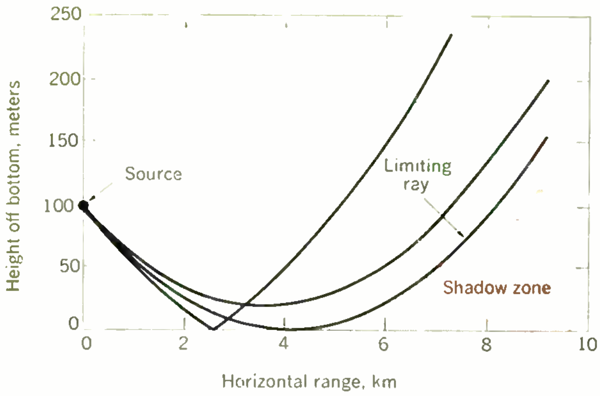


FIGURE 4. Near-bottom refraction effects.

rise to a weaker positive gradient, and causes an upward refraction.

Imposing shadow

The presence of a shadow zone severely thwarts any attempt to transmit beyond ranges greater than that for the limiting ray because penetration into the shadow zone only can occur by volume scattering (similar to that of tropospheric-scatter propagation in the upper atmosphere), by bottom reflection, or by diffraction.

In any case, a four-order-of-magnitude power increase is required to obtain an adequate signal-to-noise ratio at the receiver in the shadow zone. Even if an adequate signal-to-noise ratio is achieved, the propagation path is severely distorted by multipath effects, which include volume scattering that produces an extremely diffuse multipath propagation, and multifaceted reflection from the sea floor that gives rise to multiple discrete to overlapping arrivals that severely degrade intelligibility of the communications link.

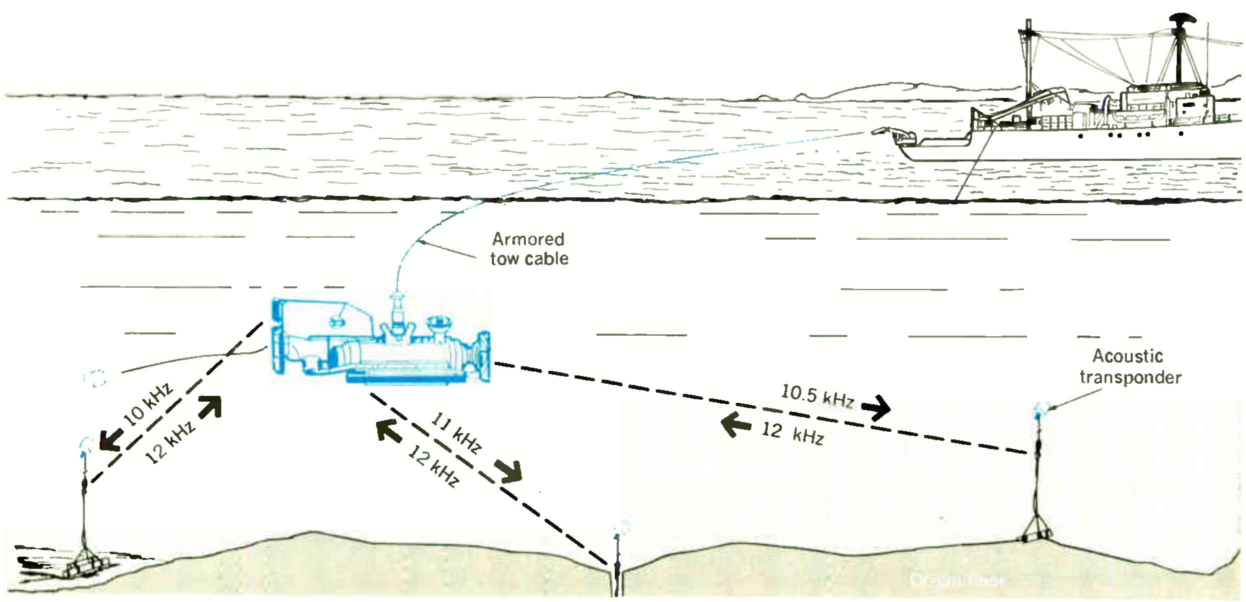
The radius of curvature of the rays shown in Fig. 3 is accentuated by a 12-to-1 compression of the horizontal scale. The radius in the ocean is actually very large and the limiting range of the shadow zone, as can be seen, is of the order of several thousand meters—depending on the depths of the source and receiver. Obviously, as the transducer's operating depth increases, a greater slant range can be supported by the refractive ocean medium.

It is easy to be misled into a complacent acceptance of this "typical" ray diagram. In reality, the near-surface thermal gradients are highly variable showing such a strong dependence on the weather, time of day, time of year, and geographic location, that the effective operating range of any communications system may vary more than an order of magnitude, depending on these factors.

In contrast to this near-surface variability, near the deep sea floor, the velocity profile is dominated by the pressure effect, which is highly predictable and gives rise to the characteristic upward refraction shown in Fig. 4. Although this refraction is not as severe as that in the upper layers of the ocean because of the smaller value of the pressure-dependent velocity gradient, it nevertheless does influence the effective range of transponders in use near the sea floor. As in the surface region, a shadow zone is created establishing a maximum range that essentially is independent of the transmitted acoustic power. The maximum operating range is achieved for that ray which just grazes the bottom. Any energy that is scattered, reflected, or diffracted into the shadow zone is too severely attenuated or distorted for practical use.

To illustrate by example, ray-path calculations show that a submersible operating 30 meters above the sea floor can interrogate a transponder to a range of approximately 9 km, provided the transponder–transducer also is elevated a few hundred meters above the sea floor. From the geometry of Fig. 4, it is seen that the range to a bottom-located transducer from a submersible floating at a given height from ocean bottom is essentially only half the range achievable for a transponder–transducer suspended at least as far above the bottom as the submersible.

FIGURE 5. Marine Physical Laboratory's deep-tow navigational system.



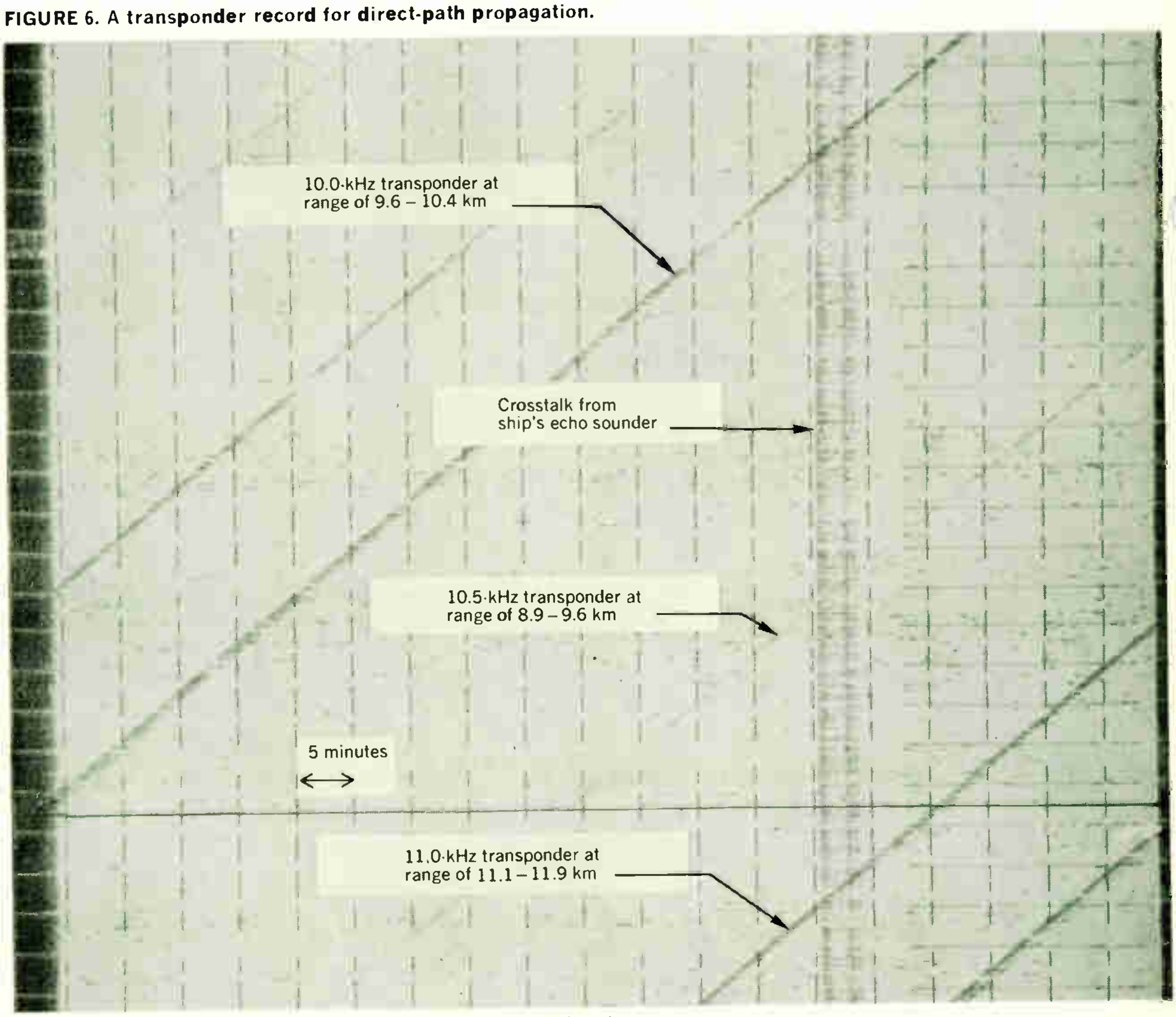
The existence of the shadow zone may be illustrated by citing examples of signals received from a near-bottom-transponder navigation system—that of the deep-tow vehicle of the Scripps Institution's Marine Physical Laboratory. The geometry used in this system is illustrated in Fig. 5. A "fish," towed from a surface ship at the end of a long cable, carries a transducer that projects a pulse of sound omnidirectionally in the ocean. This sound is received by transponders moored off the bottom that, in turn, retransmit a pulse at an offset frequency for reception by the "fish" and for transmission up the coaxial tow cable to the ship. The travel-time difference between the transmitted interrogating pulse and the received response is a measure of the range of the vehicle to a particular transponder. In actual practice, the succession of pulses received from the transponder is displayed on a graphic recorder of the type used for ordinary echo-sounding recording.

A good response

Figure 6 represents a record made in the region for which direct-path propagation occurred well clear of the shadow zone discussed previously. (The start of information being transferred over this transponder communication link is, of course, the *epoch* of the pulse.) The return pulses from two transmitters are shown on this particular record and it can be seen that the succession of arrival times presents a smooth, straight-line plot indicating an undistorted transmission of the information. Figure 7, by contrast, was obtained for a region at the boundary of the shadow zone. The distortion of the communication path by the multipath interference (caused by bottom reflection) is manifested by both a wide variation in amplitude and a wide variation in the estimate of the arrival time of the pulse epoch. Not only does one have to face the restriction of sound-absorption limitations (Fig. 2); but one also must deal with the effect of path-length distortions caused by the multipath propagation associated with marginal bottom-bounce propagation (Fig. 7), which produces erratic travel-time estimates.

A more detailed look at the path distortion caused by bottom-bounce propagation is given in Fig. 8, which shows oscillographic traces of three pulses recorded in 2000-fathom (4500-meter-deep) water at a range of about 30 km—a range just short of that associated with purely

FIGURE 6. A transponder record for direct-path propagation.



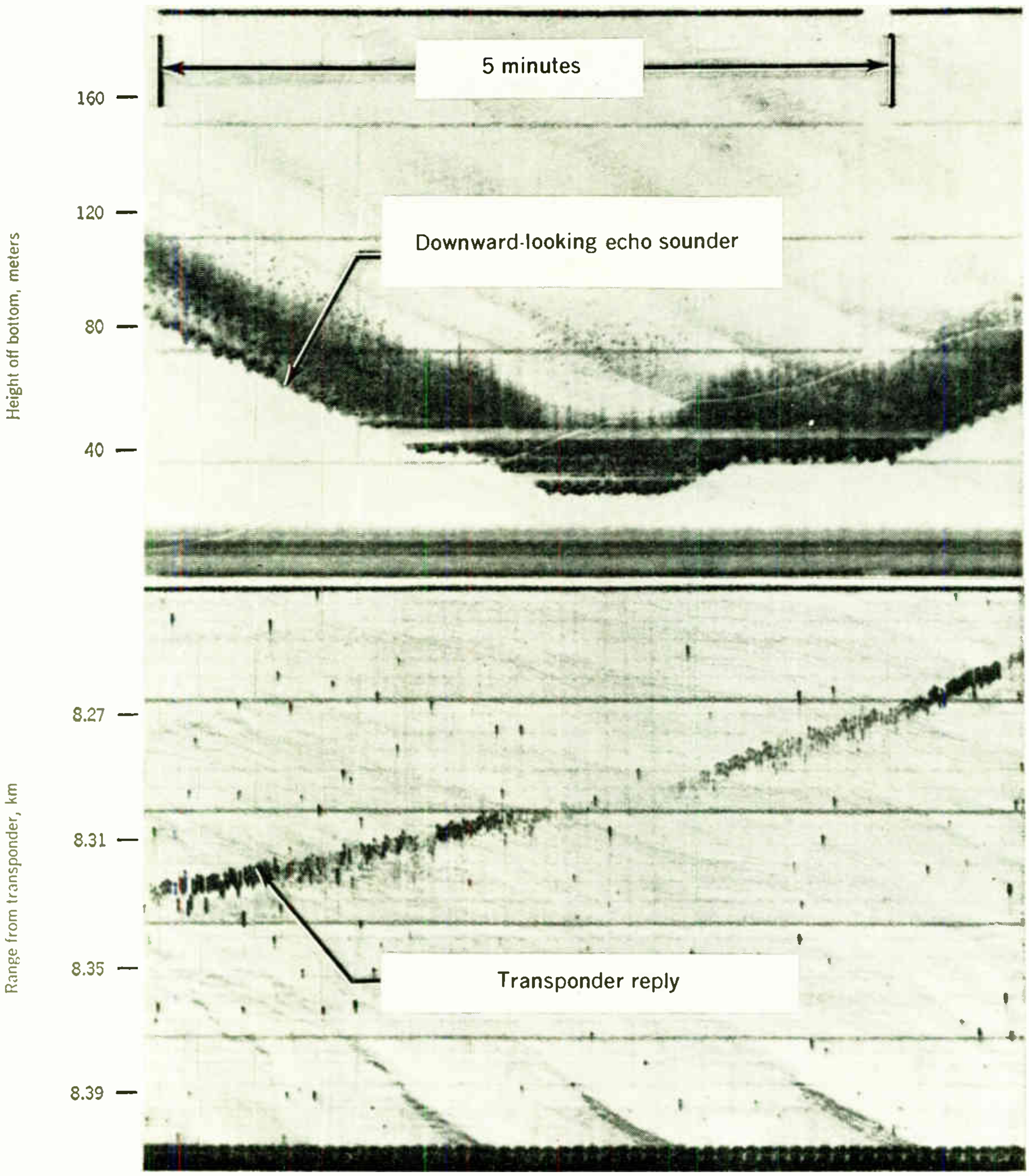
refracted paths. The sound pulse in this record has been reflected from the bottom. The three traces represent sample pulses taken at intervals of 10 minutes. The sweep rate of the oscillograph was 20 ms/cm; the transmitted pulse length was 3 ms.

The bottom trace shows a nearly undistorted pulse. This transmission corresponds either to the reflection from an extremely smooth portion of the bottom or else it may represent a pure, refracted transmission taken at a time when the range "opened up" to the refraction-limited range. The other two traces are seen to exhibit very strong amplitude modulation and pulse stretching, with pulses stretched out over 60 ms—representing an increase in pulse length by a factor of 20. The amplitude modulation of the pulse is in the vicinity of several hundred hertz. In the presence of this amplitude modulation and phase distortion, the transmission of transponder-pulse information (which requires resolution to a few milliseconds) or the transmission of voice communication (which requires reasonable fidelity over a bandwidth of the order of a kilohertz) will degrade seriously.

Since multipath distortion is one of the most serious problems in the transmission of information over reasonably long distances in the ocean, the structure of the signal must be designed to accommodate the pulse stretching and amplitude fluctuations without loss of the desired information.

The pulse-stretching phenomenon can range from the relatively minor effect shown for bottom reflections

FIGURE 7. A transponder record near the shadow zone.



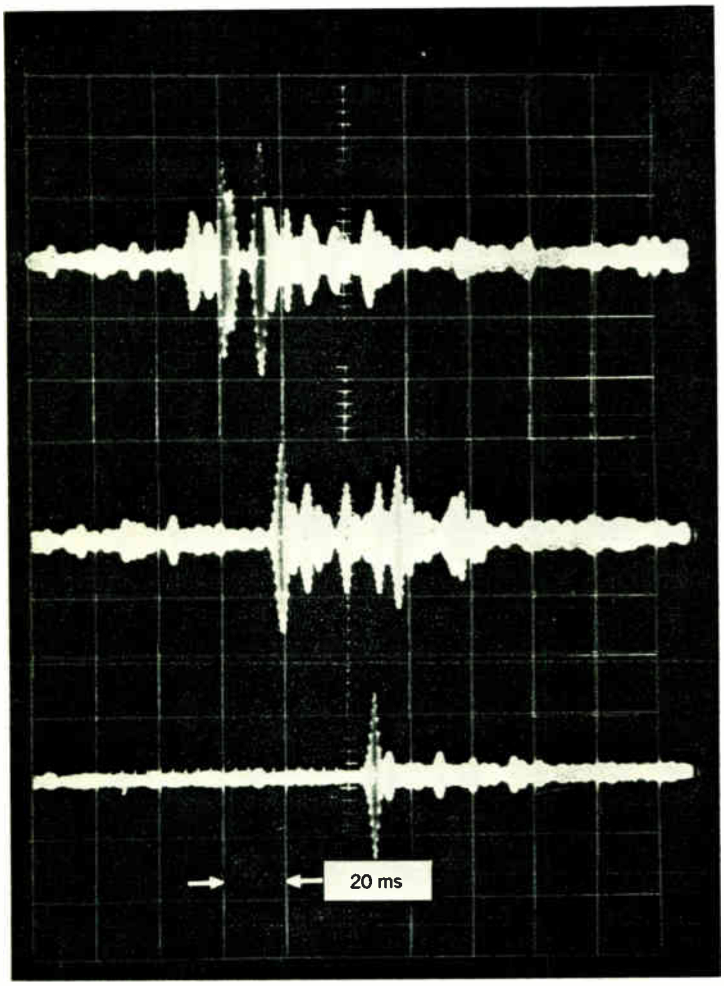
(Fig. 8), to a stretching that is more than two orders of magnitude more severe. For intermediate-range propagation in a duct such as a surface channel, it can be an order of magnitude greater than that shown in Fig. 8. The worst effects may occur for very-long-range transmission (hundreds of kilometers) within the deep sound channel where a very large number of pure refracted paths exist and pulse stretching can be as much as several seconds.

Despite multipath distortion, it is possible to transmit information over very long distances, providing an adequate reduction in data rate is imposed on the system. Data rates for tens of kilometers over a surface-channel propagation path typically can be of the order of 10 b/s. Data rates for deep-sound-channel propagation, over hundreds of kilometers using low frequencies, will fall as low as 0.1 to 0.01 b/s.

Another point

In addition to multipath disturbance, there is one other phenomenon that greatly affects the coding method used in acoustic data transmission. Within the ocean, there are inhomogeneities that give rise to a patchiness in the sound-velocity structure. The patch size of these sound-velocity inhomogeneities ranges from a few centimeters in the near-surface regions to tens of centimeters in deeper water. These patches give rise to a scintillation during transmission, in effect, superimposing a randomly fluctuating component on the normal-absorption and inverse-square spreading loss.

FIGURE 8. Received pulses from a 3-ms transmitted pulse in 4500-meter-deep water at 30 km.



The standard deviation of these amplitude disturbances is range dependent, and, of course, depends on the body of water through which the sound is transmitted. But, typically, standard deviations of the order of several decibels are observed. Because of this wide amplitude fluctuation, both frequency-modulation and frequency-shift-keying techniques have proved quite effective for information transmission in the ocean environment. These techniques also help in some severe multipath propagation situations as long as the frequency-shift rates or the transmitted frequency modulation is slow compared with the time dispersion of the multipath propagation.

A synopsis

By bowing to the whims of the ocean as a transmission medium, man has succeeded in transmitting a wide variety of information through it acoustically. The transmitted information ranges from the direct-path FM acoustic television link demonstrated by the Ball Brothers Company in 1968—where information of tens of kilohertz was achieved for several kilometers of water path—to some of the more secure command control links—where data rates as slow as 1 b/s or less were reliably achieved in a severe multipath environment.

This, then, is the nature of the acoustic communication channel that man is forced to use—and has used—as he penetrates the depths of the sea. The communication requirements cover a broad spectrum and it is certain that refinement of the application of acoustic-communication techniques to the difficult medium of the ocean will be a continuing demand upon ocean scientists in the future.

This article is a revised version of a text presented at the Northeast Electronics Research and Engineering Meeting, November 7, 1969.

Victor C. Anderson (SM) received the A.B. degree from the University of Redlands, Redlands, Calif., in 1943. Then, until 1946 when he enrolled as a graduate student at the University of California, Los Angeles, he worked in various capacities on atomic-energy developments. After a year of graduate work, he accepted a position at the newly formed Marine Physical Laboratory of the University of California. He received the M.A. from the Los Angeles school in 1950 and subsequently received the Ph.D. in 1953. In 1954, he was granted a postdoctoral fellowship at the Acoustics Research Laboratory, Harvard, and for the next year worked on digital time-compression techniques for application to acoustic-signal processing. Since returning to the Marine Physical Laboratory, he has continued to research the general field of signal processing, instrumenting new techniques such as dielectric recording and digital beam forming (receiving patents for both). Dr. Anderson was responsible for the design and construction of the Benthic Laboratory used for Sealab II. He is now assistant director at the Marine Laboratory and is a Fellow of ASA.

Delta modulation

Better communications are possible when the signals produced by the human voice are changed into a series of binary digits. One conversion method offers special advantages

H. R. Schindler IBM Zurich Research Laboratory

The idea of coding the human voice into digital pulses was conceived more than three decades ago; exploitation had to await the ushering in of the transistor era. Since then, coding systems gradually have improved. One of the latest and most efficient schemes is delta modulation. Compared with earlier analog-to-digital-pulse conversion systems, an increase in the voice-handling capacity of telephone equipment by a factor of two or more has been made possible. At the same time, equipment requirements have become less stringent.

Anyone who has talked on the telephone no doubt, at one time or another, has been able to eavesdrop on the conversation of others. This crosstalk is one of the unfortunate consequences of grouping many analog-signal conversations over telephone facilities. It and other deficiencies can be lessened by translating the voice signal into a form less susceptible to the vagaries of analog transmission.

One method is to sample the analog signal at regular discrete intervals and code the signal amplitude into a digital format. This procedure commonly is called pulse-code modulation (PCM). A variation of this ploy is to compare successive signal samples and transmit only their differences. This method (logically) is called differential pulse-code modulation. A very special form of this latter technique is known as delta modulation, and it—after some brief introduction—forms the subject of this article.

Reasons for digitizing

More and more during the past decade, analog signals have been digitized for transmission and processing. This trend is observed in almost all fields of electronics, but specifically in communications, process control, and data processing. There are many reasons for this.

- The digital format allows transmission of information over long distances without deterioration, since digital signals, unlike analog signals, can be regenerated with only small probability of error.
- Time-division multiplexing of digital information frequently leads to economical use of cables. Compared with frequency-division multiplexing, no complex filters are required in the digital case since all the multiplexing functions can be accomplished with digital circuitry.
- Switching of digital information can be easily realized with digital building blocks. This leads to fully electronic exchanges. Many problems of present-day exchanges, such as analog crosstalk and mechanical contacts, can thus be avoided.
- Information in digital form can be processed with powerful digital computers.
- Information in digital form can be scrambled easily. This is important wherever privacy or secrecy must be guaranteed—specifically in military communications.
- Transmission of information over futuristic carriers, such as laser beams or long-distance waveguides, is best accomplished in digital form.

Historical review

A. H. Reeves first proposed the conversion of analog signals into digital format in 1938. In doing so, he clearly defined the concept of PCM and recognized its advantages. During World War II, a successful PCM system was built but, because it was constructed of low-reliability vacuum tubes, it had severe drawbacks. With the invention of the transistor in 1948, the active electronic component that was ideally suited for digital systems became available—its main advantages being smallness, low power consumption, and, especially, suitability as a switching element.

In 1955 the development of the first commercial PCM system was begun (Bell Telephone T1 carrier system), and in 1962 production was started. Today several hundred thousand PCM links are in operation in the United States. There are also many installations in Japan and England, and a number of other countries are introducing PCM links at the present time.

PCM is still the most well-known method for the

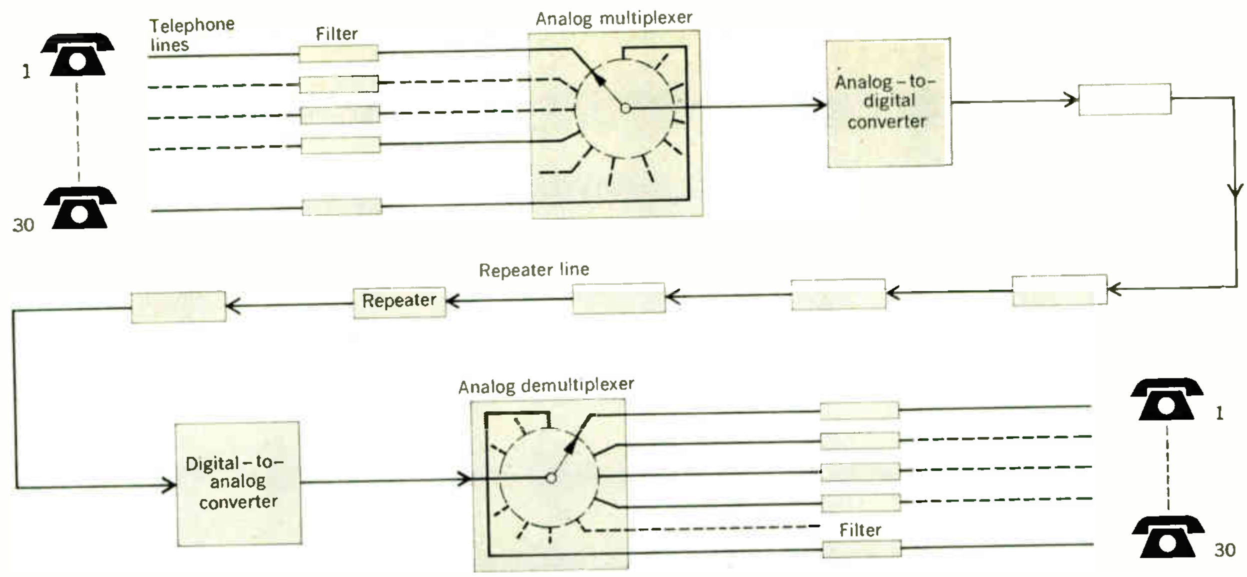
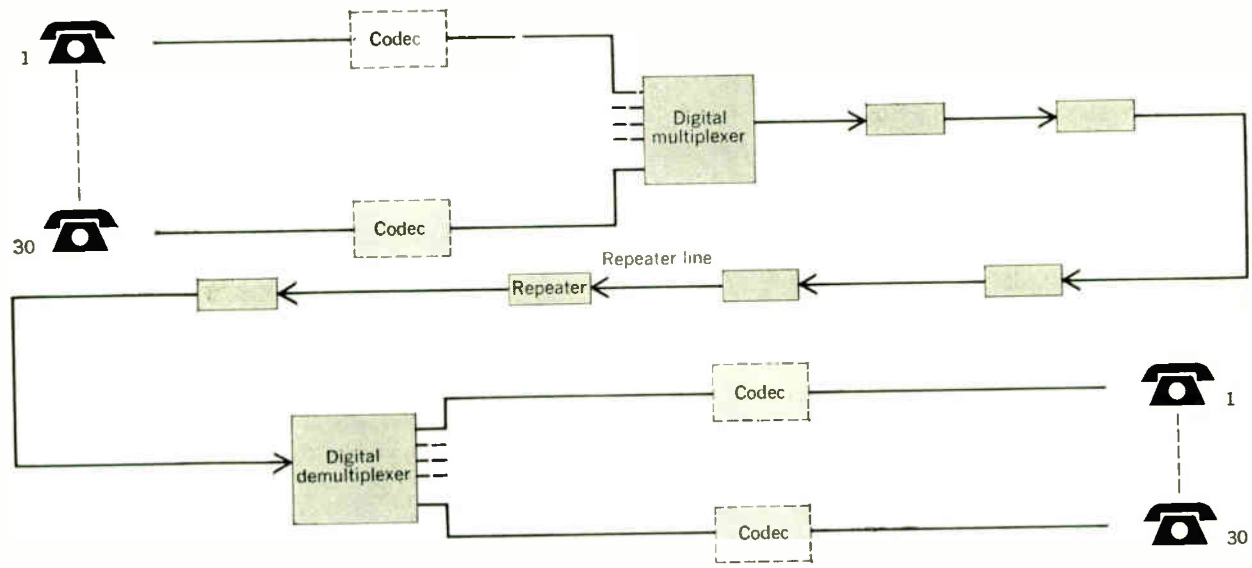


FIGURE 1. Typical PCM system.

FIGURE 2. Block diagram of a delta-modulation system.



analog-to-digital conversion of voice signals. A typical example of a PCM system, as specified by the CCITT (International Telegraph and Telephone Consultative Committee), is shown in Fig. 1.*

The audio signals from subscriber sets are sent through telephone lines to the PCM concentrator. There they first pass through a low-pass filter, which eliminates all frequency components exceeding one half of the sampling rate, or all components above 4 kHz, in order to conform to the sampling theorem. (Without this filter a high amount of foldover distortion due to the folding of the frequency spectrums would occur.) The output from these line filters is sampled sequentially by an analog multiplexer (stepping switch), which returns to the same filter output every 125 μs. The multiplexer output is connected to the PCM coder. The complexity of this coder is rather high; however, it is shared by 30 telephone channels.

The output from the coder is sent to the repeater line. The remote end of the line ties into the decoder, which feeds an analog demultiplexer. There, each line has to be filtered a second time to eliminate the undesired folded-frequency spectrums.

* The system is based on a symmetrical, quasi-logrithmic compression–expansion characteristic, consisting of 13 segments. Each segment (with the exception of the center one) is divided into eight quantizing steps. The total number of quantizing steps is 128, described by seven binary bits. The sampling rate is 8 kHz; the bit rate is 56 kb/s. Thirty telephone conversations and two control channels are simultaneously transmitted over a cable via time-division multiplexing.

The delta system

The development of delta modulation is more recent. The basic principle was described for the first time in a French patent issued in 1946.[1] More detailed descriptions by de Jager (Philips) and Libois (CNET)[2-4] of the many aspects of delta modulation appeared in 1952.

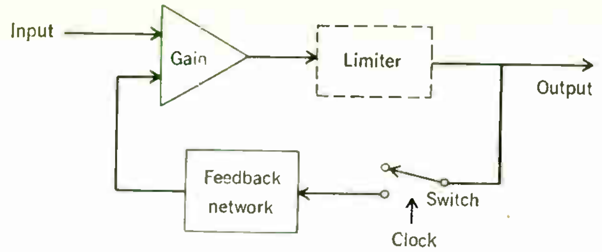


FIGURE 3. The electronic feedback system forms the backbone of the delta modulator.

I. Feedback systems

Type of Amplifier	Switch Closed	Feedback Signal	Circuit Used In
linear	permanently	continuous	normal feedback
linear	periodically	pulsed	sampling servo
clipping	permanently	clipped	on–off servo
clipping	periodically	quantized	delta modulation

A system corresponding to the PCM scheme that appears in Fig. 1, but based on delta modulation, is shown in Fig. 2. Here, each telephone channel is coded individually. The line multiplexing is done digitally.

With respect to PCM the following important differences exist:

- There is no need for expensive filters with steep-slope characteristics.
- The problem of analog multiplexing, requiring cross-talk levels below 70 dB, is eliminated.
- Many identical building blocks are contained but there are few varieties of blocks. Considerable component redundancy improves the economical aspect.

However, such systems remained experimental because

- The number of coders is much higher. A system of this kind, therefore, only can be economical if the cost per coder is low.

Also, with regard to earlier coders,

- Many delta-modulation systems suffered from idle noise caused by an asymmetry of the two current sources normally used in delta coders. If these current sources are not perfectly matched, an audible idling pattern appears in pauses between words of a telephone conversation that, when demodulated, produces an audible sound that is very disturbing.
- The dynamic range of delta coders operating at bit rates comparable to those used in PCM systems was inadequate. Voice signals of low amplitude were quantized very coarsely. Although the dynamic range readily is improved by increasing bit rate, this procedure allows fewer voice conversations to be multiplexed on a cable.

Therefore, despite the attractive simplicity of these delta coders,[5-7] their drawbacks had prevented much use. Delta modulation remained simply an interesting field for theoretical studies. This situation began to change when more refinements were suggested. In 1963, Winkler (RCA)[8] proposed a first method for improving the dynamic range. In 1967 and 1968 more propositions followed.[9-11] These refined delta-modulation systems were given different designations—the most common today being "companded delta modulation." [The word companded is composed of (signal dynamic range) *com*pressing at the transmitter and ex*pand*ing at the receiver.]

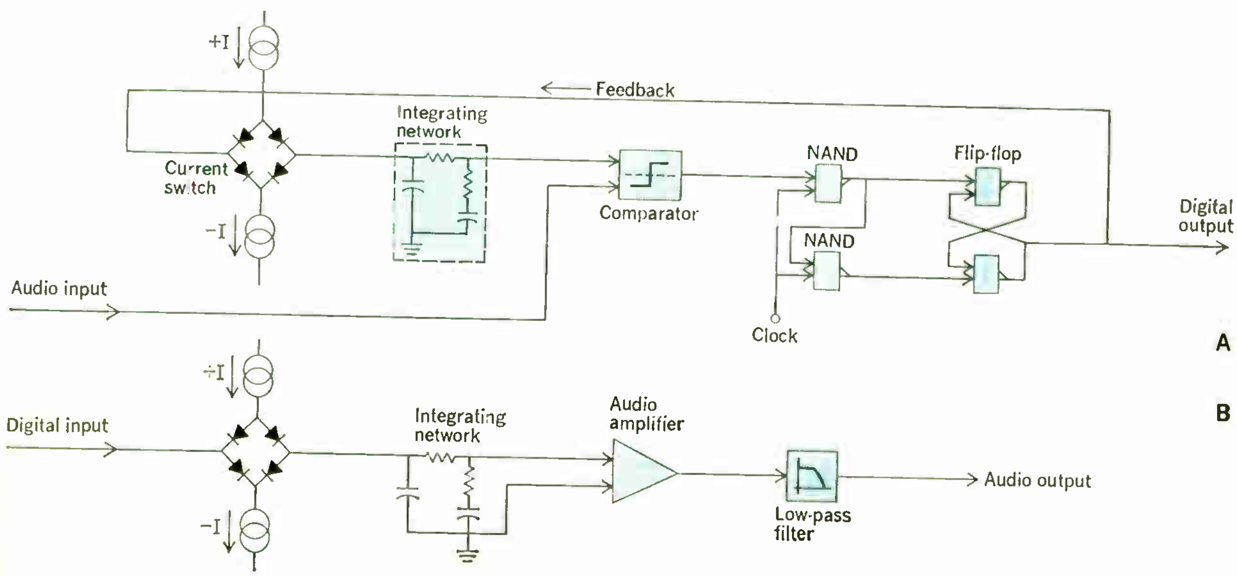
A total telephone system based on delta modulation was first described by a group at Nippon Electric.[12] Later, a similar system was built in France (Société Télécommunications Radioélectriques et Téléphoniques)[13,14] but, by companding, it operated at a lower bit rate per telephone channel.

Today the development of delta modulation is in full progress. Many communications research laboratories are engaged in exploring in depth the theory of delta modulation.[15-28]

The concept of delta modulation

Basically, the delta coder is an electronic feedback system as shown in Fig. 3. Table I lists the different kinds

FIGURE 4. A—Circuitry for the basic delta-modulation encoder. B—Circuitry for the decoder.



of feedback and their application. In the case of the delta coders, the amplifier is limiting—thereby quantizing the output—and the feedback path is closed periodically.

The most basic delta coder [Fig. 4(A)] consists of a comparator, two NAND gates, a flip-flop, two switchable current sources, and an integrating network. The decoder [Fig. 4(B)] contains similar building blocks—two switchable current sources, an integrating network, an audio gain element, and a low-pass filter. The encoder ciphers the voice signal into a single-digit binary code; the decoder reproduces the voice signal.

In operation an analog signal—derived from the digital output signal of the encoder—is compared with an input audio signal. This comparison is made in a comparator, which operates as a limiting amplifier. (If the reconstructed signal is more positive than the input signal, the comparator's output is equivalent to a digital binary "one" and vice versa.) At intervals controlled by the clock, the comparator output signal is switched to a storage flip-flop through two NAND gates and it is held there until updated by a subsequent signal. The inverted flip-flop output controls the two current sources $\pm I$ through feedback. If it is positive, corresponding to a binary "one," the positive current source is connected into the integrating network and vice versa. The integrating circuit, through ramp-type increments, then reconstructs the audio input signal—and the cycle is repeated. However, there are small differences between the original and reproduced signals and these differences are audible at the decoder output in the form of quantization noise.

The decoder, in a similar manner, controls one of two current sources, which feeds an integrating network identical to that in the encoder. The signal from this integrating network is amplified and filtered. The bandwidth of this filter corresponds to the bandwidth of the input signal and the filter suppresses the frequency components of the quantization noise that lie outside the audio band. In this way, the ratio of signal-to-quantization noise (SNR) can be improved.

In the most simple case, the integrating network consists of a capacitor and the reconstructed waveform, as shown in Fig. 5. This single-integration approximation is rather coarse and, therefore, the quantization-noise level is quite high. It can be shown that for a sine-wave input the optimum SNR is[3]

$$\frac{S}{N} = \frac{0.2\, f_s^{3/2}}{f f_0^{1/2}} \tag{1}$$

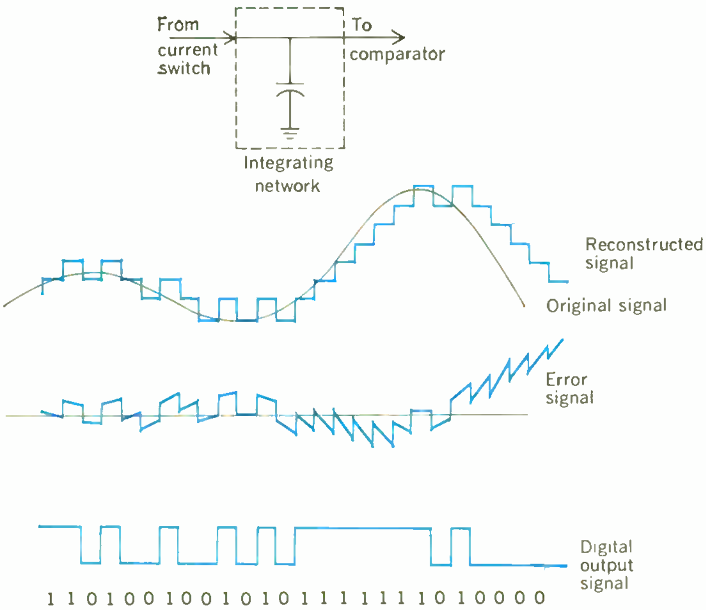


FIGURE 5. Waveforms for the delta coder with single integration.

FIGURE 6. Waveforms for the delta coder with double integration.

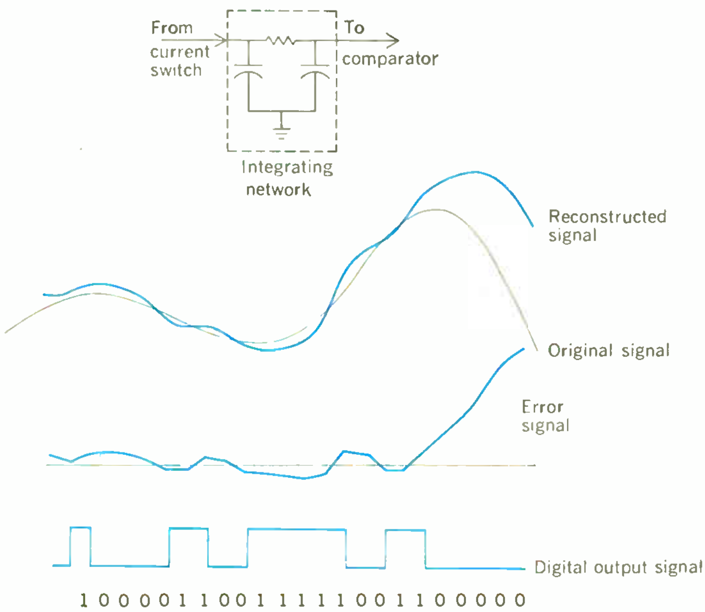
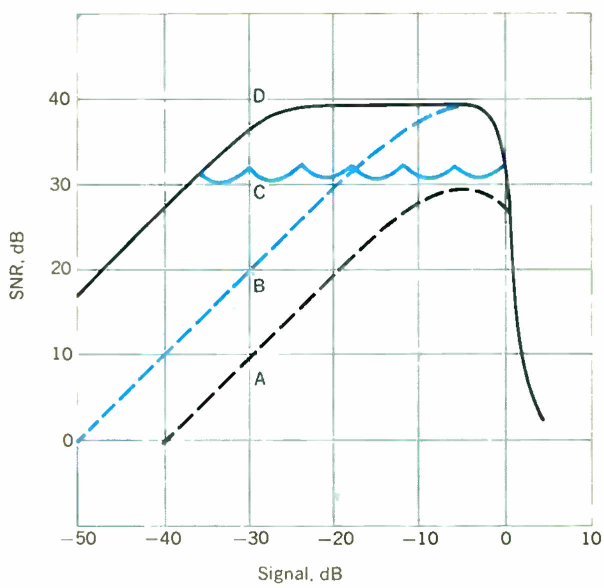


FIGURE 7. Comparison performance of coding methods. A—Delta modulation, 56 kbaud, single integration, f (see text) = 800 Hz. B—Delta modulation, 56 kbaud, double integration, f = 800 Hz. C—PCM, 13 segment CCITT, 56 kbaud. D—Companded delta modulation, 56 kbaud, f = 800 Hz.



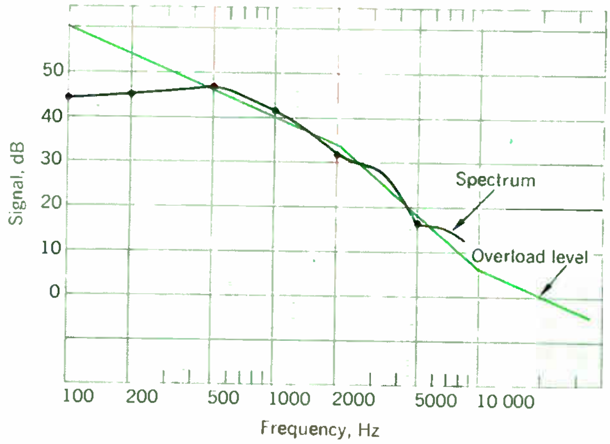


FIGURE 8. Spectrum of the human voice compared with delta-coder overload level.

where f_s = sampling frequency, f = audio frequency, and f_0 = audio bandwidth.

In order to improve the coding properties of the delta coders, a scheme was proposed to encode differences in slope instead of differences in amplitude. This concept led to coders with pure double integration. Figure 6 shows the original and the reconstructed waveforms for this system. However, systems of this kind lack stability and can oscillate under certain conditions.

By combining single and double integration, improved performance is obtained with respect to single integration and the stability problem of pure double integration is avoided. Systems of this type frequently also are called in the literature delta modulators with double integration.

For these delta coders the maximum ratio of signal-to-quantization noise is (for a sine wave)

$$\frac{S}{N} = \frac{0.026 f_s^{5/2}}{f f_0^{3/2}} \tag{2}$$

The SNR as a function of the signal amplitude S is shown in Fig. 7 for a sampling rate of 56 kHz, an audio bandwidth of 3.4 kHz, and a sine-wave input of 800 Hz.* The SNR gain due to the addition of double integration is obvious.

Overload occurs when the input-signal amplitude exceeds the maximum possible amplitude of the reconstructed signal. In the case of single integration, the overload level is inversely proportional to the frequency. The overload level for a properly designed coder with mixed integration as a function of frequency is shown in Fig. 8. In the same diagram the spectrum of the human voice (male) has been plotted. It can be seen that the two curves match very closely, indicating that such a delta coder is well adapted to human-voice input.

The ideal coder

Before discussing the more refined delta-coding schemes in detail, it might be useful to summarize the properties of an ideal coding and decoding system (CODEC).

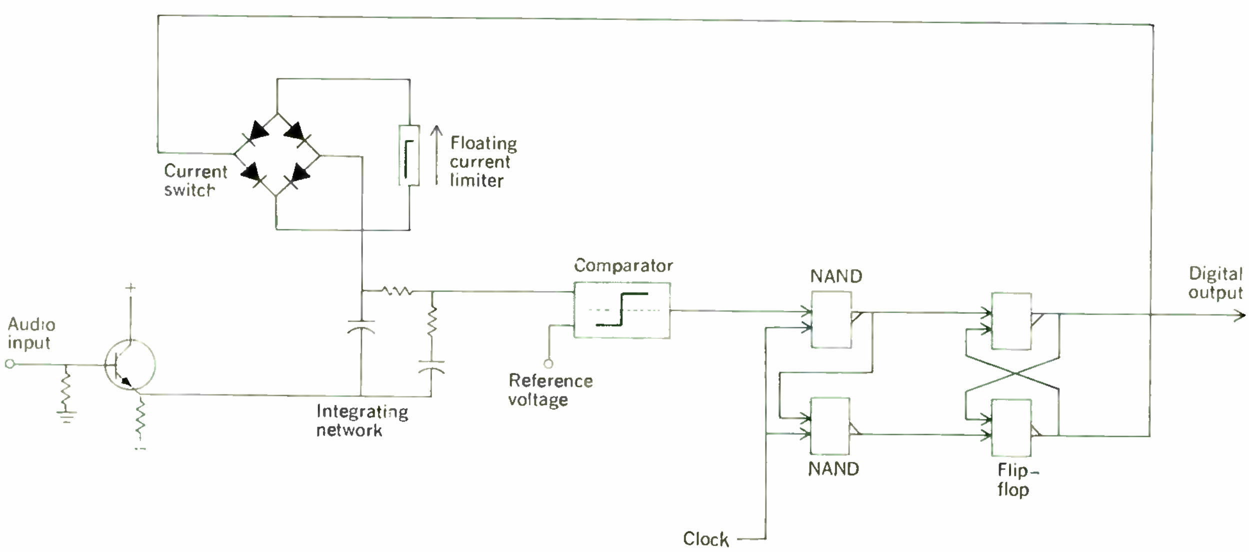
- Simplicity (i.e., a minimum of hardware). This requirement is closely connected with the suitability of the coder for monolithic circuit implementation.
- Very low idle-noise level.
- Large dynamic range. Logarithmic companding is desirable.
- High signal-to-quantization-noise ratio (e.g., 32 dB or better).
- High audio bandwidth (3 dB or better at 3.4 kHz).
- Natural sound: no sharp limiting of the upper voice spectrum.
- Low bit rate.
- A short compression time constant in case of dynamic compression.
- Good reproducibility of the compression characteris-

* The normalized standard deviation frequency $\bar{f}$ is indicative of the long-time spectrum $A(f)$ of the human voice (although the coder may perform differently for real [short-time] speech sound).

$$\bar{f}^2 = \frac{\int_0^{f_0} A^2(f) f^2 \, df}{\int_0^{f_0} A^2(f) \, df}$$

In the frequency range 0 to 3.5 kHz, $\bar{f}$ is quite accurately 800 Hz. Above 3.5 kHz the voice spectrum falls off rapidly.

FIGURE 9. Delta-coder scheme for eliminating idle noise.



tic. The number of precision components defining this characteristic should be low.

Eliminating idle noise

Idle noise[29] can be eliminated by replacing the two current sources in the delta coder by a single current limiter in a floating arrangement (Fig. 9). The currents are switched through a diode bridge. In this way, the positive and the negative currents are matched in magnitude. The current limiter can be a field-effect current-limiter diode or simple circuit using bipolar transistors that provides the same function. Any remaining unbalance can be easily eliminated using a simple servo loop.

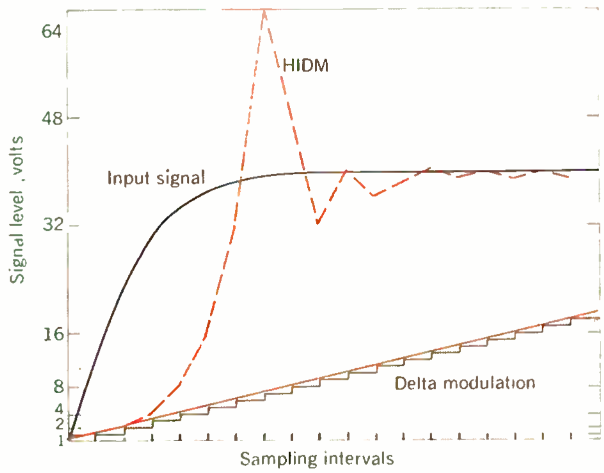
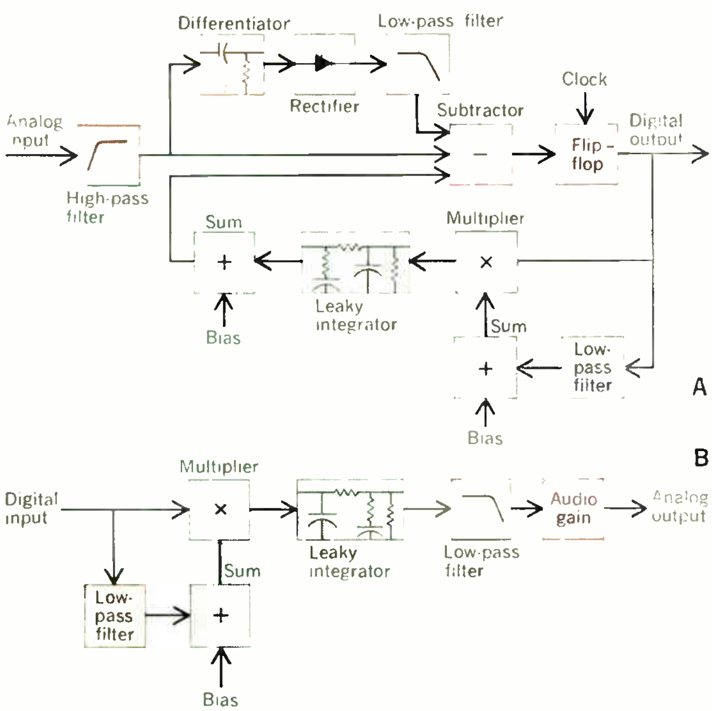


FIGURE 10. Step response for a high-information delta-modulation setup.

FIGURE 11. Continuous delta-modulation system. A—coder. B—decoder.



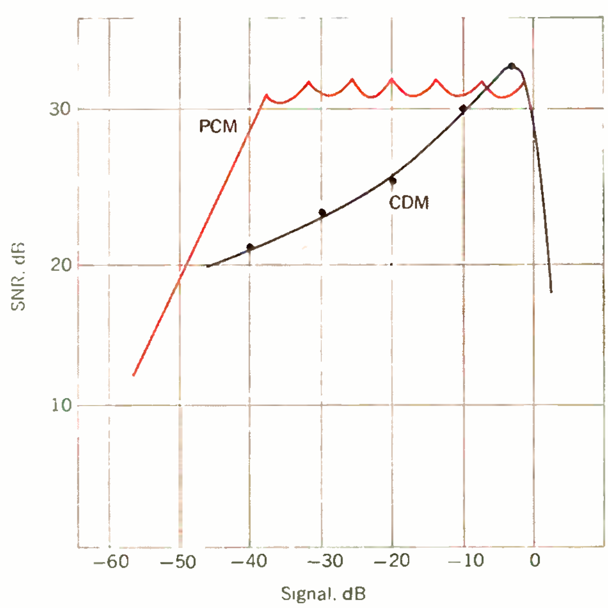
Improving the dynamic range

The dynamic range of a delta-coding system is appreciably improved by companding. In delta coders, the most suitable companding method consists of adapting the quantization step in coder and decoder to the input signal. The relationship between input signal and step size can be either static or dynamic. Whereas a static, time-invariant relationship between signal amplitude and quantization step is used in most PCM systems, in companded delta-modulation systems the quantization step size is always derived dynamically from certain overload criterions of the coder. In pauses between words, the coder operates well below overload and, therefore, the step size drifts in a controlled manner toward its minimum value.

The first proposition for a companded delta-modulation system came from M. R. Winkler (RCA) in 1963. His method consists of doubling the size of the quantization step whenever two identical, consecutive binary values appear at the coder output. The step size is divided in half after each transition from one binary number to another (see Fig. 10)—a method called "high-information delta modulation" (HIDM). By choosing this name, Winkler tried to indicate that the binary sequence produced by this technique contains more information than straightforward delta modulation. The method is capable of giving a finer description of the input signal, or reproducing the signal with smaller quantization and overload errors. The dynamic range also is improved appreciably.

More proposals for companding schemes followed in the years 1966 to 1968. Greefkes and de Jager (Philips) developed a system called "continuous delta modulation," wherein the quantization step is varied continuously. The block diagram is shown in Fig. 11. The basic concept uses the frequency band below 200 Hz to transmit the information describing quantization step size. For proper system operation, the audio spectrum has to be filtered to remove all components below 300 Hz in order

FIGURE 12. SNR versus signal level for continuous delta modulation.



to keep the two spectrums separated.

After the incoming audio signal is sent through a high-pass filter and split in two, one part is put through a branch consisting of a differentiator, a full-wave rectifier, and a smoothing filter. In this branch, the slope of the signal is determined and a measure of proximity to the overload level is obtained. This slowly varying signal is combined in a difference network with the high-pass-filtered audio signal. The feedback path from the coder also leads into the same network. The coder is therefore fed by a combined input of three signals, one of which is a time-varying dc component.

The integrating network of one feedback path is designed to have leakage, which makes possible an average ratio of "ones" to "zeros" different from unity at the coder output. With the addition of a bias voltage, the coder forces this ratio of "ones" to "zeros" to be 1 to 2 for low input levels, and to increase to 1 to 1 for large input signals. In the second feedback path, a voltage, derived from this ratio, controls a multiplier circuit and thus the quantization step size.

These operations are repeated in the decoder. The results prove that good SNRs and high dynamic range can be achieved.

The continuous delta-modulation scheme was used in France (Société Télécommunication Radioélectriques et Téléphoniques) for the design of an all-digital telephone link and exchange that doubles the number of telephone conversations over a multiplex cable without increasing bit rate. At a multiplex rate of 2.048 Mb/s, 30 telephone conversations were transmitted using regular PCM coding, whereas 60 conversations were transmitted using delta modulation. The quality of the digital link is shown in Fig. 12. In this version the coding method was called controlled-slope delta modulation (modulation en delta à pente asservie).

Doubling the number of telephone channels in a multiplex link provides a great advantage. However, this particular system has some drawbacks, the first being that the compression response time is rather long. This can lead to transient overload distortion, which does not show up during measurements of the quantization (granular) noise with an 800-Hz sine-wave input. The second drawback is the need to filter the audio and the compression signals in order to avoid overlap of the two frequency spectrums. Several potcore inductors per coder are used for this purpose; however, it should be possible to find a circuit configuration compatible with integrated-circuit technology.

At the same time that the French were developing their system, Brolin and Brown of the Bell Telephone Laboratories reported a companded delta-modulation system (Fig. 13). The delta coder is composed of two individual coders. The first contains in its feedback path a multiplier controlled by the feedback signal of the second loop. The second coder is fed a signal derived by differentiating, rectifying, and smoothing the audio signal. This signal is a means for determining the quantization step size.

The second delta coder operates at a rate of 6 kb/s, whereas the first coder operates at a rate of 96.5 kb/s. In this manner, the size of the quantization step is controlled and the coder operates over a wide range of input-signal levels without overloading. The two bit streams are multiplexed into the transmission line.

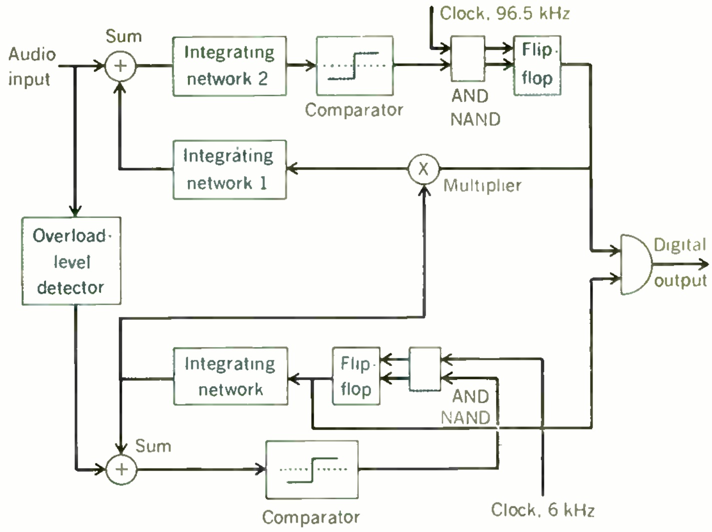


FIGURE 13. Bell Telephone Labs' companded delta coder.

FIGURE 14. Nippon Electric's companded delta coder.

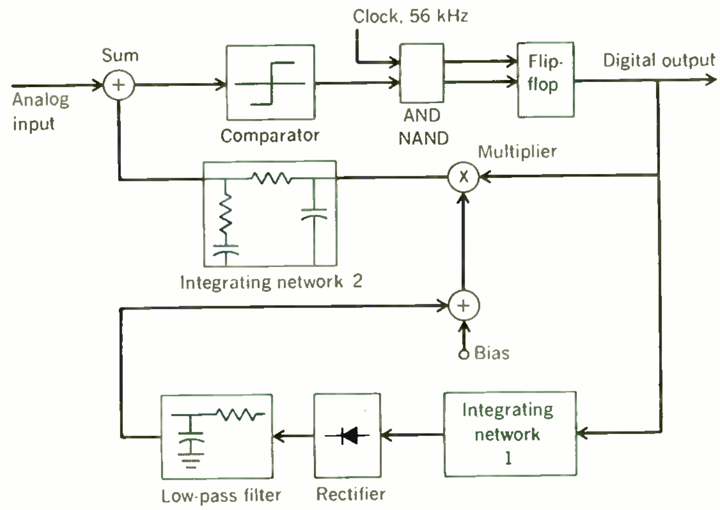
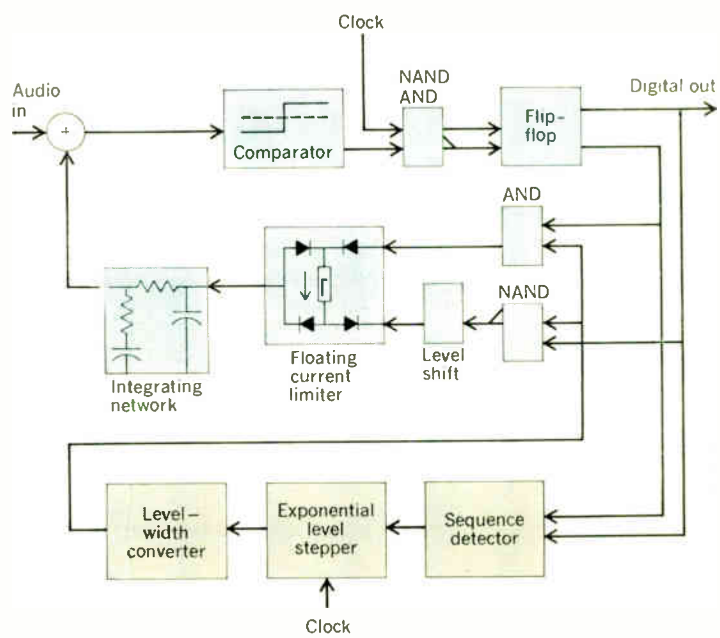


FIGURE 15. IBM's companded delta coder.



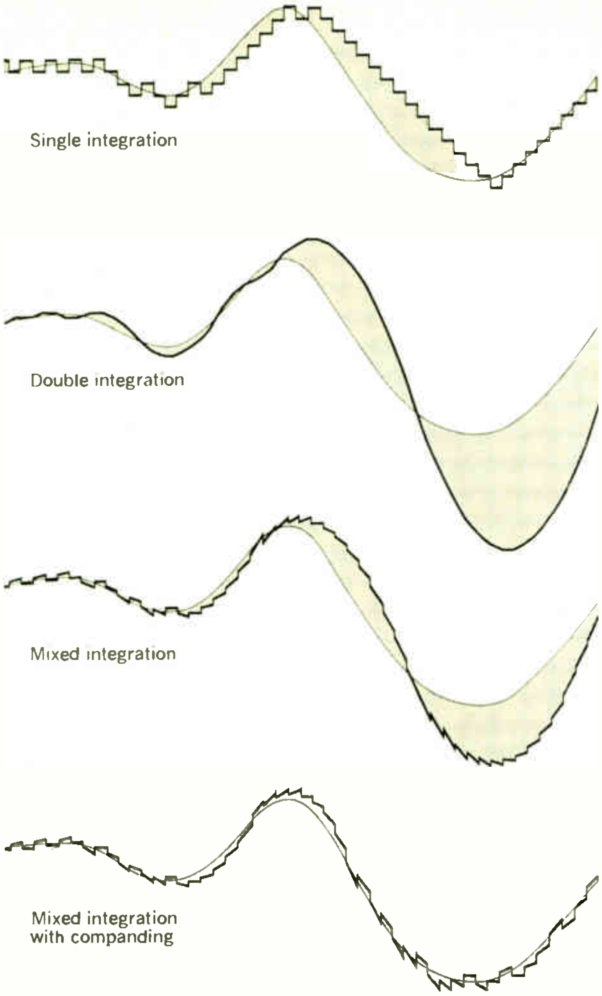


FIGURE 16. Waveforms obtained for various integrating systems.

At the receiver, the two bit streams are separated and fed to two decoders similar to the two coders. The output signal to the second decoder controls the quantization step size of the first decoder, which reproduces the original audio signal.

Good results were achieved with this coding system. A dynamic range of 35 dB at an SNR of 40 dB was measured. Drawbacks are the high bit rate and the need to interleave two bit streams.

Other schemes

Another companding scheme, by Tomozawa and Kaneko (Nippon Electric), is shown in Fig. 14. The digital signal at the coder output is detected, rectified, and smoothed. The voltage produced is added to a bias voltage and both are fed, as one input, to a multiplier located in the feedback path of the coder. The coder rate is 56 kb/s, the companding ratio (the ratio of the largest to the smallest quantization step) is 20 dB; the effective compression characteristic is a square-root function. At an audio frequency of 800 Hz, an SNR of 35 dB can be measured over a dynamic range of 25 dB.

A companded delta-modulation system (Fig. 15) was developed at IBM by the writer in 1969. Its compression characteristic is logarithmic—a condition that, like logarithmic PCM, produces a constant ratio of signal-to-quantization noise over a wide dynamic range. Furthermore, logarithmic encoding eases the problem of matching the compression characteristic in the coder to the expansion characteristic in the decoder.

In addition to the standard coder circuitry, the system contains four blocks for the companding mechanism:

The *sequence detector* analyzes the digital coder output. If in the coding process the number of consecutive "ones" or "zeros" exceeds four, the sequence detector responds with a positive output.

As a consequence, the *exponential level stepper* increases its voltage level (retained on a hold capacitor) by a given factor K_1 (such as 2 dB) once each sampling interval until a transition occurs in the coder output. For a negative output of the sequence detector, the voltage is reduced by a second factor K_2 (which is less than K_1—for example, 0.2 dB) during each sampling interval.*

In the *level-width converter* the output from the level stepper is translated into a sequence of pulses of variable width. These pulses control the duty cycle of the current flowing into the integrating network of the coder. By this method, the quantization step depends on the digital-output sequence. The dependence is chosen so that the coder operates just a few decibels below the overload level; therefore, an optimum ratio of signal-to-quantization noise is obtained over a wide dynamic range.

Finally, two *logic gates* control the current flow through a diode quad to the integrating network.

The same functions are performed in the decoder, which reproduces the original signal. The response time of the companding scheme is sufficiently short to ensure low-transient overload distortion.

The only parameters determining the compression characteristic are K_1 and K_2. The factor K_1 depends on the ratio of two components, whereas K_2 is determined by an RC time constant. K_1 and K_2 are chosen independently for optimum performance. Figure 16 shows a typical original and reconstructed waveform produced by the coder–decoder combination. For comparison, the waveforms produced by single, double, and mixed-integration coders are also shown.

The performance of the logarithmic coder is shown in Figs. 17 and 18. The coder rate is 56 kb/s and the audio bandwidth (at the 3-dB point) is 3.4 kHz. For an 800-Hz input, the SNR exceeds 38 dB over a range of 25 dB. For equal bit rates, a reduction in quantization noise of approximately 6 dB has been realized compared with the CCITT PCM coder.

In comparing the frequency response of companded delta modulation to PCM, a much smoother roll-off at

* The exponential-level stepping of the quantization step size using unequal factors K_1 and K_2 might appear to be a rather complicated process. However, it is implemented with a minimum of hardware, which adds very little to the complexity of the coder. The designed circuit is based on the principle of charge pumping. There are two inputs to the exponential-level stepper and these are fed by a negative clock pulse and an inverted-sequence-detector signal. If both inputs are negative—indicating that the step size is to be increased—charge is pumped through a pump capacitor into a hold capacitor. The amount of pumped charge is a fixed fraction of the charge stored in the hold capacitor. A decrease in step size is realized by a resistance leakage path across the holding capacitor. Both the charging and the discharging processes follow an exponential law, thus producing the equivalent of a logarithmic compression of the audio signal. (For the delta coder, the ratio of input signal to quantization step size is the important element.)

frequencies above 4 kHz is evident. This results in a more natural sound, with all fricatives still present and well-distinguished.†

Theoretical aspects

It becomes clear that companded delta modulation has better coding properties than PCM when used at rates below 70 kb/s for human-voice signals (see Fig. 19).‡ A similar conclusion is valid for differential PCM compared with straight PCM. The differential PCM coding method is based on a procedure whereby differences are taken between consecutive samples, which are then quantized into 2^n levels. The samples are taken at the Nyquist rate—that is, at twice the frequency of the signal bandwidth—and n binary digits are used per sample. However, differential PCM still requires rather sharp limiting of the audio band since the sampling rate is much below the bit rate.

One reason for the good coding properties of companded delta modulation is that voice signals are very different from band-limited random ones; they have certain known properties that can be described by the spectrum or the autocorrelation function of the voice signal. These properties can be used to design a coding method that is adapted to this type of signal. More specifically, this means that coders can be designed on the principle of prediction.

In the companded delta coder, two kinds of prediction are used: one is in the form of the integrating network, which combines single and double integration matching the long-time speech spectrum (first- and second-order linear prediction); and the other is based on a algorithm for the derivation of the quantization step size. In the IBM approach the physical implementation of the algorithm involves the sequence detector and the exponential-level stepper—the prediction device being the hold capacitor. Both types of prediction extrapolate the audio signal. Since only the *deviations* from the predicted (extrapolated) value are encoded, lower quantization noise can be achieved.

Looking at predictive coding from the aspect of information theory, it is clear that for best reproduction of the analog waveform, the digital output signal must contain as much information as possible: Its properties must be those of a random sequence. Since the spectrum of a random binary sequence is flat over the range of interest, the integrating network must have a response identical to the spectrum of the human voice. However, the proper choice of the integrating network alone does not ensure the best possible SNR. A second important parameter is the relationship between input signal level and quantization step size. If the input level is too low (near idling), the sequences of consecutive "ones" or "zeros" at the coder output are very short. For high input-signal levels (near overload), many long sequences of consecutive "ones" or "zeros" occur. Only at the proper input level does the distribution of the length of these sequences resemble that of a random binary signal.

The adaptive process of sensing the length of these

† Since submission of the manuscript, several papers describing delta coders using digital-output sequence for step-size control have been published.[30-33] With one exception, none make reference to exponential-level stepping and thus are not companded logarithmically.

‡ Recent investigations using real speech signals indicate a crossover rate closer to 60 kb/s.

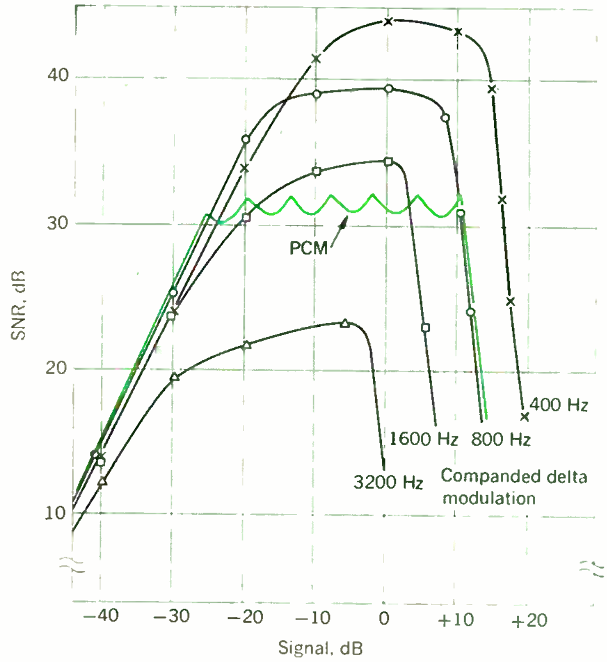


FIGURE 17. Actual companded delta modulation SNR performance of an IBM system compared with a theoretical PCM system.

FIGURE 18. Small-signal frequency response for companded delta modulation and PCM.

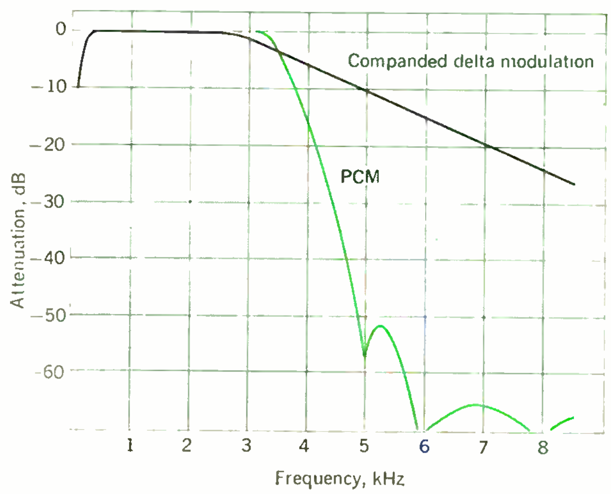
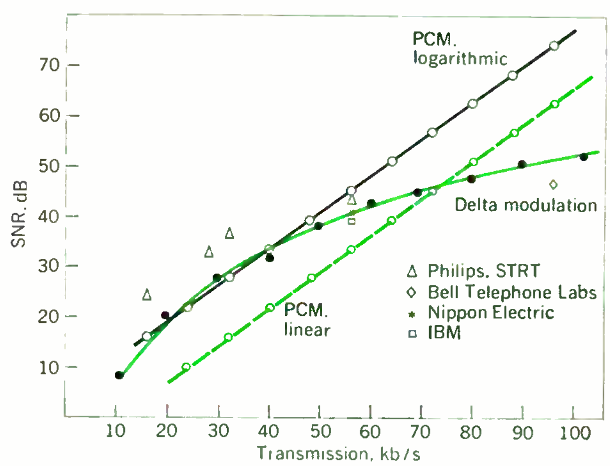


FIGURE 19. Comparison of SNRs for various pulse schemes.



sequences and of controlling the quantization step size therefore is justified.

Applications

As a consequence of the advantages described, it appears reasonably certain that delta modulation will find a much wider range of application than it has in the past—and mainly in the following areas.

- High-quality digital speech transmission and storage (e.g., audio-response units).
- Telephone-quality digital speech transmission at bit rates lower than those used in PCM systems. This could be of great importance in satellite communication of the type PCM-FDMA (pulse-code modulation, frequency division multiple access).[34]
- Very-low-bit-rate telephone communication for military applications.
- Encoding of single voice channels (e.g., a service line in a data link).

Finally it should be noted that delta-coding schemes using similar companding laws (besides coders without companding) have been applied to video with good results[35, 36] and delta coding is expected to play an important part in Bell Telephone's Picturephone service.

The author is indebted to many colleagues in IBM's Zurich Research Laboratory and in IBM Centre d'Etudes et Recherches, La Gaude, France. Particularly, the author thanks E. Hopner, A. Croisier, C. Jacquart, A. Frei, D. Maiwald, and P. Vettiger for many exciting discussions.

REFERENCES

1. Deloraine, E. M., van Miero, S., and Derjavitch, B., "Methode et système de transmission par impulsions," French Patent No. 932 140, Aug. 1946.

2. Libois, L. J., "Un nouveau procédé de modulation codée la modulation en delta," *L'Onde Electronique*, vol. 32, pp. 26–31, Jan. 1952.

3. de Jager, F., "Delta modulation, a method of PCM transmission using the 1-unit code," *Philips Res. Rept.*, no. 7, pp. 442–466, 1952.

4. Cutler, C. C., "Differential quantization of communication signals," U.S. Patent 2 605 361, July 29, 1952.

5. Sekimoto, T., and Kaneko, H., "A method of delta modulation," *Record Nat'l Conv. IECEJ*, pp. 317–319, 1958.

6. Nakamaru, Y., and Kaneko, H., "Delta modulation encoder," *Nippon Electric Co. Research and Development*, no. 1, pp. 46–54, Oct. 1960.

7. Inose, H., and Yasuda, Y., "A unity bit coding method by negative feedback," *Proc. IEEE*, vol. 51, pp. 1524–1535, Nov. 1963.

8. Winkler, M. R., "High information delta modulation," *1963 IEEE Internat'l Conv. Record*, pt. 8, pp. 260–265.

9. Tomozawa, A., and Kaneko, H., "Companded delta modulation for telephone transmission," *IEEE Trans. Communication Technology*, vol. COM-16, pp. 149–157, Feb. 1968.

10. Brolin, S. J. and Brown, J. H., "Companded delta modulation for telephony," *IEEE Trans. Communication Technology*, vol. COM-16, pp. 157–162, Feb. 1968.

11. Greefkes, J. A., and de Jager, F., "Continuous delta modulation," *Philips Res. Rept.*, no. 23, pp. 233–246, 1968.

12. Enomoto, O., Tomozawa, A., Katayama, H., Kaneko, H., and Sekimoto, T., "An experimental high speed electronic switching system using delta modulation," *1965 IEEE Internat'l Conv. Record*, pt. 1, pp. 41–52.

13. Daguet, J., Bréant, P., Gaquère, C., Lamoine, P., and Bellanger, M., "Multiplex téléphonique à 60 voies, à modulation en delta à pente asservie," *Câble et Transmission*, vol. 23A, pp. 178–197, 1969.

14. Bergeron, J. R., "Controlled slope delta modulation," *Telecommunications*, vol. 3, pp. 31–33, Dec. 1969.

15. Bennett, W. R., "Spectra of quantized signals," *Bell Sys. Tech. J.*, vol. 27, pp. 446–472, July 1948.

16. Van de Weg, H., "Quantizing noise of a single integration delta modulation system with an n-digit code," *Philips Res. Rept.*, no. 8, pp 367–385, Oct. 1953.

17. Zetterberg, L. H., "A comparison between delta and pulse code modulation," *Ericsson Technics*, vol. 11, pp. 95–154, 1955.

18. Debart, H., "Etude informationelle de la modulation en delta," *Câble et Transmission*, vol. 15A, pp. 198–204, 1961.

19. Halijak, C. A., and Tripp, J. S., "A deterministic study of delta modulation," *1963 IEEE Internat'l Conv. Record*, pt. 8, pp. 247–259.

20. O'Neal, J. B., "Delta modulation quantizing noise analytical and computer simulation results for Gaussian and television input signals," *Bell Sys. Tech. J.*, vol. 45, pp. 117–141, Jan. 1966.

21. Protonotarios, E. N., "Slope overload noise in differential pulse code modulation systems," *Bell Sys. Tech. J.*, vol. 46, pp. 2119–61, Nov. 1967.

22. Abate, J. E., "Linear and adaptive delta modulation," *Proc. IEEE*, vol. 55, pp. 298–308, Mar. 1967.

23. Johnson, F. B., "Calculating delta modulator performance," *IEEE Trans. Audio and Electroacoustics*, vol. AU-16, pp. 121–129, Mar. 1968.

24. Hosokawa, S., and Yamashita, K., "Companded delta-modulation coders of the 1/2-power and 2/3-power types," *Electron. Commun. Japan*, vol. 51-A, pp. 18–26, 1968.

25. Donaldson, R. W., and Douville, R. J., "Analysis, subjective evaluation, optimization, and comparison of the performance capabilities of PCM, DPCM, ΔM, AM, and PM voice communication systems," *IEEE Trans. Communication Technology*, vol. COM-17, pp. 421–431, Aug. 1969.

26. Goodman, D. J., "Delta modulation granular quantizing noise," *Bell Sys. Tech. J.*, vol. 48, pp. 1197–1218, May–June 1969.

27. Aaron, M. R., Fleischmann, J. S., McDonald, R. A., and Protonotarios, E. N., "Response of delta modulation to Gaussian signals," *Bell Sys. Tech. J.*, vol. 48, pp. 1167–96, May–June 1969.

28. Goodman, D. J., "The application of delta modulation to analog-to-PCM encoding," *Bell Sys. Tech. J.*, vol. 48, pp. 321–343, Feb. 1969.

29. Wang, P. P., "Idle channel noise of delta modulation," *IEEE Trans. Communication Technology*, vol. COM-16, pp. 737–742, Oct. 1968.

30. Greefkes, J. A., "A digitally companded delta modem for speech transmission," *1970 IEEE Internat'l Conf. Commun. Record*, pp. 7-33– 7-48.

31. Deschenes, P. A., and Villeret, M., "A continuous delta modulator and a numerical converter for an integrated telecommunications network," *1970 IEEE Internat'l Conf. Commun. Record*, pp. 7-27– 7-32.

32. Petford, B., and Clarke, C. M., "A companded delta-sigma speech digitizer, *IEE Conf. Publ.*, no. 64, pp. 59–63, May 1970.

33. Jayant, N. S., "Adaptive delta modulation with one-bit memory," *Bell. Sys. Tech. J.*, vol. 49, pp. 321–342, Mar. 1970.

34. Puente, J. G., "A PCM–FDMA demand assigned satellite multiple access experiment," *Telecommun. J.*, vol. 36, pp. 158–170, 1969.

35. Winkler, M. R., "Pictorial transmission with HIDM," *1965 IEEE Internat'l Conv. Record*, pt. 1, pp. 260–265.

36. Bosworth, R. H., and Candy, J. C., "A companded one-bit coder for television transmission," *Bell Sys. Tech. J.*, vol. 48, pp. 1459–80. July 1969.

H. R. Schindler (M) is a research engineer at the IBM Zurich Research Laboratory, Switzerland, where he is concerned with aspects of digital-communication systems such as PCM, differential PCM, and delta modulation. Prior to joining IBM, Mr. Schindler held various engineering positions at the Electronics Laboratory of the General Electric Co., Syracuse, N.Y. There he did research and development work on ultrahigh-speed circuitry, instrumentation for biological research, and ranging systems for laser rangefinder applications. From 1956 to 1960, he was a research assistant at the Institute of Technical Physics of the Swiss Federal Institute of Technology. He holds the M.S.E.E. degree from that institute. Mr. Schindler is the author of several papers on UHF A/D conversion, microelectronics, delta modulation, and ranging systems. He holds two patents.

Planning the coordination of ground transport

The need for transportation depends to a great extent on the life style of a society, and this need changes as the society changes in response to the pressures of technology

J. E. Gibson **Oakland University**

Proper coordination in the design of a transportation system requires consideration of general land use in the area served, and indeed of the entire sociological and ecological fabric of society. Moreover, the designer must help weave this fabric while it is being used. A possible starting place for a transportation study, and some would argue the only starting place, is an overall regional plan. More and more it is being clearly recognized that a transportation link not only satisfies established needs in a region but also develops and anticipates others.

Transportation planners who view themselves as merely responding to needs are tied to the past. Extrapolation from the past mortgages the future, whereas forward planning helps create it. This phenomenon perhaps may be most clearly recognized as characteristic of the American west during the latter half of the 19th century. The period dating roughly from 1850 to 1900 could be called the age of the railroads. The future of a town or region depended on whether the railroad would pass through it. Government land grants to the railroads were justified by arguing that this would "open up the territory" and aid in the development of the west.

"Between 1850 and 1872 the federal government gave an empire of land to promoters who promised to build railroads across the sparsely settled territory of the west. While Congress dispensed over 100 000 000 acres of the public domain, state and local governments provided an additional $280 000 000 in cash or credit (about 30 percent of the total capitalization of railroads) in the decades preceding the Civil War. Spurred by such inducements the railroad network expanded swiftly. The legacy left to the Gilded Age included a system of track over 35 000 miles long—almost half the world's total."[1] Railroad construction brought about the development of towns, cities, and whole regions of the west.

(It is to be noted, however, that the railroads did not develop unopposed. A report to the New York State legislature in the 1830s stated: "We are led to the conclusion that in regard to cost of construction and maintenance and also in reference to the expense of conveyance at moderate velocities, canals are clearly the most advantageous means of communication. On the other hand, where high velocities are required, as for the conveyance of passengers, and under some circumstances of competition, for light goods of great value, in proportion to their weight, the preference would be given to a railroad."[1])

Transportation plans that create the future will have a heavy impact on the city. Planners who think in terms of 20- to 30-year periods have good cause to ignore the shape of the cities as they are at present in order to concentrate on how they should be. By this logic, transportation engineers now despairing about the interchange problem in future urban mass-transit schemes for grid or radial cities are concerned with the wrong problem. Transportation and land use should be so planned as to take advantage of high-speed ground transport (HSGT) rather than sacrificing almost all of its benefits in order to save present city ground plans. The conventional round city is appropriate if vehicular speeds are independent of direction, but with HSGT perhaps other criterions for the shape of a city are operative. Today, a city 80 km (50 miles) long and four blocks wide could be square—in the sense of time. A spinal cord containing a 400-km/h transit system would permit a person to move between points on it in 20 minutes, the same length of time it takes to walk two blocks to the green fields and trees beside the city. In order to aid in planning future cities it seems imperative that we determine what the modern city is for.

The function of a city

Of course, one should not mindlessly do violence to a classical city in order to adapt its form to its future uses but one also cannot expect classical cities to remain functional under the pressure of future urban technology. We are not talking here of regional trading centers that seem to grow to about 50 000 persons, but rather of urban centers with populations of 500 000 or more. Classical cities are extant examples of centers of empire developed before the 18th century. The "center of empire" is an obsolete function for modern cities, although a number of classical cities must be maintained as art treasures. Following the industrial revolution, a city could perform as a "center of industrialization." Great factories attracted workers, merchants, and auxiliary activities to a growing center of industrialization, and this in turn made the center a more attractive place in which to locate more factories. Europe has cities of both of these types and some that are a mixture, but North American cities are almost all of the latter class. As a living and functional entity it would seem that the center of industrialization is also obsolescent. Modern factories require considerable land area, and because of the high cost of city land and the difficulty of arranging for transportation of raw material and finished goods it is not even economically feasible to locate them in repossessed slums. (Although a few small factories have recently been located in slum areas as evidence of social concern by large corporations, none have been free of economic difficulty. Even the few that have made a profit have not shown an adequate return on investment.) Encouraged by such factors and the easy availability of private transportation, future city development may tend toward low-density suburban sprawl laced with expressways, of which Los Angeles is the best (or worst) example. Alternatively, a new centralizing force may develop certain cities into "centers of communication."

The center of communication seems to be the final functional form in the long development of the city. If one ignores the function of communication it is impossible to explain the continued increase in density in New York City. Why build the Pan Am building where construction costs will be maximized? Why add the International Trade Center to an apparently already crowded lower Manhattan? Those elements of commerce involved in manufacturing rather than in communication, such as the garment industry, are being driven from Manhattan by the increasing cost of space. These are random examples, and others could be cited; however, more convincing evidence of the evolution of the city into a center of communication is available.

I. Floor-space usage in Detroit central business district*

Group	Percent Used in 1954	Percent Used in 1968	Trend
Commercial	32	50	up
Industrial	8	2	down
Residential	3	3	unchanged
Retail	20	14	down
Hotels	10	8	down
Warehousing	10	7	down
Parking	13	13	unchanged

*Area bounded by Detroit River and Fisher, Lodge, and Chrysler Freeways.

(Calculated from data supplied by Paul Bork, CBD Study, Detroit City Planning Commission.)

Table I shows the most recent data on floor-space utilization in Detroit's central business district. Note the rather dramatic increase in space devoted to the communications facilities.

Classification of transportation needs with regard to function

The need for transportation depends greatly upon life style, and this need changes even in a well-developed technological society. Seifert[2] indicates that "...on the average each man, woman, and child in the United States now travels twice as many miles each year and requires twice as much cargo transportation to satisfy his needs as he did in 1940." Thus, one would be wise to inquire into the basic function served by transportation and to consider the effect of technology in providing substitutes as well as new modes, if he wishes to engage in forward planning.

How might the various functions served by transportation be classified? One way would be by the object served. This is not as obvious as it would appear. It is the rare industrial executive who really understands his mission, and transportation planners are no exception. For example, the business of IBM is not to make computers or typewriters or other objects. The more sophisticated answer "IBM makes money" is also unsatisfactory. Ultimately, IBM serves certain data and information processing functions for business customers. The electric utility doesn't merely make and distribute electricity; it distributes energy in convenient form for wholesale and retail. The automobile companies don't merely make cars; they design and manufacture private, personal transportation components. We aim by this redefinition to shift the emphasis from a specific object or device to determining how a company may fulfill a continuing function for society. Excessive concern for a current product may cause a corporation to overlook a revolution under its nose. The automobile manufacturers replaced the horse-drawn-carriage manufacturers, and they in turn can be replaced if they fail to understand their real function. With this in mind, let us consider the object of three basic kinds of intercity travel.

Freight. Obviously the object is to move goods from the shipper to the consignee. This truism seems to tell us little but it could serve to focus our attention on the overall trip rather than merely the line-haul segment of it. Such elements as warehouse and inventory costs to both shipper and consignee, pilferage, and packing costs must be included in the total cost of transportation. Certain large shippers use railroad cars as rolling warehouses. Automobiles, for example, are often started for the west coat of the U.S. without a specific consignee. During the trip the autos are sold and the railroad is informed of their specific destination. Certainly collection and distribution are part of the transportation costs as is all the time lost by the consignee.

Business travel. The second major kind of transportation is business passenger travel. We must examine the object of such business trips if we are to understand how future changes in technology and society are likely to affect this segment of transportation. This function will be discussed in detail later, but the conclusion is that much business travel is for the purpose of communication and will be eliminated by wide-band communication systems now under development.

Personal travel. The passenger traveling for personal reasons makes up the last major kind of transportation use. We can identify two subclasses of personal travel—tourism and all other purposes. The object of tourism is the journey itself. Transportation is fundamental to this object and cannot be replaced. However, for mass transportation systems to profit from increased tourism in the future, certain fundamental freedoms, usually associated with the private auto, must be supplied. First, a family fare plan must drastically reduce the unit cost for additional members of the family after one full fare has been paid. This can be accomplished in much the way as airlines in the United States now operate. Off-peak loading can be required without significantly detracting from the acceptability of the vacation plan.

Second, almost unlimited baggage for use on the vacation must be permitted. Third, convenient arrangements for movement by rental cars, bicycles, interurbans, and so forth, must be available at the resort area if people are to be lured from their automobiles. The present European rail pass for U.S. tourists contains the germ of this idea. It may be that aggressive promotion will raise per capita tourist travel on public facilities to offset the relative decline I foresee in per capita business travel.

Much nontourist personal travel is for the purpose of communication, and except for extremely important social situations such as funerals and marriages, can be eliminated. Suppose, for example, there were available during nonbusiness hours a "videophone visit plan." The cost for an hour's "visit" might be quite modest, perhaps $20 or so. Equipment would be delivered to the homes of the two parties and self-operated, perhaps through the television set. A wide-angle lens on a hand-held camera and a relatively weak transmitter beamed at a stationary satellite are all that would be required. Would not such a service provide an adequate substitute for many family visits?

Passenger transportation and communication

In the foregoing discussion we said that most business travel and nontourist personal travel are designed to serve a communication function. How can one demonstrate this? Start by asking the businessman's object in traveling. Almost without exception he is going to an office or conference room to communicate with a few other people. From internal evidence it will be impossible to establish where the office or conference room is geographically. There is nothing intrinsic to the conduct of business in the geographical location of that room. The mission is to perform an interpersonal communication function. This is the source of the common remark, "I could have accomplished as much by telephone as I did at that conference or meeting." As it stands, however, such a remark is incorrect.

More than objective verbal information and documents are exchanged at a typical business meeting. Nonverbal cues greatly increase the effective bandwidth and thus the communication rate. To take a minor example, the element of homage is unspoken but very real. One projects homage through one's actions—by the way one holds one's body as he enters his superior's office, by waiting to be asked to be seated, by not smoking until invited to do so or by asking permission, by not taking the lead in the conversation until invited to do so. One does not learn these rituals easily, and once learned there is a reluctance to relinquish them for a situation in which a mistake could be made.*

The reader can extend this scenario to salesmen's calls and other examples. Of course, all this will be nerve wracking in the transition period. It will only be accomplished by business leaders who force the transition after they realize how much of their time and company travel funds can be saved by using a wide-band communication system such as a videophone.

Videophone conference calls as a substitute for meetings will require care if they are to be successful. It will *not* be satisfactory to project by videophone the image of a single speaker to all other viewers. Each viewer will desire a visual mosaic of all other participants since much communication is nonverbal. Even here there is a difficulty, however, since spontaneous semiprivate conversations and nonverbal signals before, during, and after the conference by any two participants or among several of the participants will be difficult or impossible. The impact of wide-band communications on business and professional conventions is difficult to predict for the same reason. Spontaneity is lost as well as extracurricular participation. Probably structured, businesslike conferences in typical northeastern industrial cities will be replaced by videophone conference calls sooner than those involving boondoggling junkets to Acapulco or Honolulu.

Examples of noninterpersonal communication business journeys include examination of goods at a buyers' mart or exposition and inspection of a factory or factory site. Such activities are extremely wide-bandwidth functions. One either doesn't know quite what he is looking for or wants to avoid allowing someone to serve as an editor or censor. He wishes to open himself to a whole range of sensory perceptions. In this sense such a journey is analogous to a vacation trip. A movie or videophone tour is a reminder but not a substitute for direct participation.

We conclude therefore that an important change in business travel will occur with the introduction of economical videophone service.[4]

Coordination of transportation and future needs

One can argue that although coordination of transportation is essential, such coordination has been notable by its absence. I would argue that not only is this true but it is economically unavoidable under present constraints. As I have pointed out elsewhere,[5] rationalization of intercity transportation by mode so as to optimize transit time can be accomplished in principle, but economic pressure will always tend to blur the nominal modal boundaries. Short-haul carriers will press to be allowed to involve themselves in the more-profitable-per-unit long hauls and the long-haul carriers will seek to penetrate the larger short-haul market.

Intracity transportation coordination is even more complex. We suggest that by the year 2000 many major cities must be rebuilt if they are to become and remain functional. Yet who is to do this massive job? Neither local government nor private business can be expected to com-

* Does this seem a bit overdrawn? See the recent book on *Business Adventures*,[3] which describes the negotiations between the New York Stock Exchange and the U.S. and European creditors of Ira Haupt and Company, a stock brokerage firm that went bankrupt due to a vegetable-oil swindle in 1963. Note how many journeys were involved in the negotiations even though time was extremely short. Ask yourself why a telephone call wouldn't have done as well.

mit itself to what is more properly a national goal.

How much more worthwhile than diverting the best technical talent in the U.S. for a decade in the glittering but essentially trivial occupation of placing a man on Mars would it be to extend ourselves to construct *one* truly habitable city within the United States. I do not wish to denigrate the magnificent technical achievement of Apollo 11 and its predecessors, nor do I argue that Kennedy's establishment of the Moon Goal ten years ago did not then seem the right thing to do—but now our goal must be to revitalize the city.

Let us consider some strategies for dealing with the coordination of future transportation needs. Planning on all of these proposals can be initiated immediately but there are three distinct, though overlapping, time periods for realization. One of the tests for a valid thesis is to require of it a list of those things forbidden or contradicted by it. An all-pervasive and permissive thesis is no thesis at all. Thus, I will list counterproductive activities as well as activities to be encouraged.

The elements of social existence in the next 30 years that one must project in order to arrive at an estimate of future transportation requirements are really quite simple and few in number. The beginnings of these trends surround us. They are independent of detailed political developments on a national and international level. Here are examples of the elements upon which projections should be based.

- Population in developed nations will continue to increase, although at a decreasing rate. Birth control will be more effective and widely used but the U.S. will have a population of 300 million before the 21st century.
- The population will be distributed in three primary megalopoli in the U.S. rather than scattered uniformly through the nation.[6] One of the simplest arguments for this position is the following. People go where employment opportunities develop. Employment opportunities will continue to become more and more specialized with more and more complex equipment involved. Industry will be forced in this direction because of the demand for higher wages and a better standard of living. Such industry experiences a high pressure for agglomeration and very little pressure for disbursion. (Agglomeration here is meant in a national sense and not as an argument for an increased density in a specific core city.) People will live near their jobs as they always have because, contrary to idle speculation by those who should know better, the work week is increasing, not decreasing, and there is no evidence yet of a revolution in human nature that would encourage industry to allow a worker to work at home or a place of his own choosing, free of supervision.
- It is impossible to change in a radical way within a few years the basic structure of the cities. A modified city form will develop gradually, if at all. Thus, modification of the city form must be made a long-range goal although system analysis should start immediately. The so-called "rush to the suburbs" has been evident for more than half a century. It used to be called "moving out to the end of the trolley car line."
- Private industry will continue to seek short-term profit-making opportunities and be guided by the marketplace. This seems obvious, but if it were, we would not expect the auto industry to take the lead in safety standards or in eliminating the internal combustion engine. Nor would we expect the management of U.S. railroads voluntarily to embrace passenger travel.

Near-range national goals. In the 0- to 5-year range, national transportation policy should be directed toward proper utilization of existing facilities. For example, by proper balancing of coercion and rewards it should be possible to involve U.S. railroads in more "experiments" such as the New York–Washington Penn Central Metroliner.* In reality, of course, this is no experiment at all, since it lags by almost a decade modern European and Japanese rail practice. It is essential that present U.S. rail management be given an opportunity to move into the modern era voluntarily. There is little doubt that they will fail to meet the challenge. For instance, *The Wall Street Journal* for September 3, 1969, reports that the U.S. Department of Transportation is considering a plan for government operation of passenger railroads in certain urban regions such as the Boston–Washington corridor.

Vast sums of money should not be spent on a network of new airports because present aircraft-use growth trends will not be sustained if HSGT is available. Since almost 80 percent of all commercial airline flights in the U.S. are between cities less than 800 km apart, and this is precisely the range into which HSGT ideally fits, one must analyze the situation carefully before extrapolating existing aircraft-use trends.

We moved into the air because of the mess we had made with ground travel, and now we are making a mess of the airways. I was in the air on the night of the New York City blackout and cold sweat still breaks out at the thought of several hundred aircraft converging on Kennedy Airport on individual visual control. The far more efficient ground mode (HSGT) will save time, be lower in cost, eliminate airport traffic congestion, and is generically safer due to its absolute constraint to fixed guideways.

Medium-range national goals. Planning should begin immediately on programs to be activated in the 3- to 15-year time period. Consolidation of existing railroad lines into a smaller number of more efficient lines is already well under way in the United States. It seems inevitable that the work of the Interstate Commerce Commission will be rationalized in this time period and recourse to the Commission will not be available to one mode of transport as a means of stifling competition from other modes. (See the delightful "Yak Fat Rate Case" in Mayer's *The Lawyers*.[7]) It seems just as likely, however, that present rail management will continue to concentrate on freight traffic.

It appears that a number of railroad rights of way will become redundant due to the consolidation mentioned above. It would be useful to rebuild a number of such lines for automatically controlled, medium-speed, intercity passenger service. Only the most densely populated areas should be considered for this interim measure and present technology involving the steel wheel on the steel rail should be employed.

Experience on the New Tokaido Line and the New

* One must fight very hard to overcome one's feeling of depression at the lack of enthusiasm and cooperation on the part of railroad management in this exercise. H. C. Kohout, Penn Central vice president for passenger service, pronounces himself "delightfully surprised" with the 72 percent increase in traffic and the average load factor of 81 percent. As well he might! The July 1969 issue of *Fortune* states: "The results are even more impressive in the face of the limited schedule and stone-age seat-reservation facilities."

San Yo Line extension should be of great value.[8] Fujii states that a minimum radius of 4000 meters (1200 feet) and a maximum grade of 1.5 percent will permit operating speeds of 250 km/h to be realized. With a minimum radius of 2500 meters on the existing portions of the New Tokaido Line the maximum speed permitted is 210 km/h. This will *triple* present U.S. rail operating speeds, which have been progressing steadily downward in recent years due to improper maintenance of way. The sums of money required for grade relocation will be large but not astronomical. Most unacceptably small curvature is located in existing built-up areas but, due in no small part to the railroad itself, these are mostly areas of urban decay. Often it will be possible to combine urban renewal with the grade relocation. Present expressway construction runs in excess of $10 million per mile with some downtown expressway replacement costs at more than $25 million per mile. A medium-speed rail line requires less land than a four- or six-lane expressway and is quicker and simpler to build. Since a double-track rail line under automatic control can carry more than five times the traffic of a typical expressway, a cost of $10 million per mile would not be unreasonable. Such a line for the Boston–Washington corridor would cost less than $5 billion, or about the cost of one year of the U.S. space program. Even though this medium-range program would be conducted entirely with presently available hardware (and relatively unskilled labor), it would be far beyond the capability of present railroad management. Fortunately, however, a very large number of highly trained and capable engineers, familiar with supervising much larger and more complex systems, are soon to become redundant in our manned-spaceflight program and the very large U.S. airframe industry will be available for the design and construction of the control computers, vehicles, and guideway.

Long-range national goals. For the 10- to 25-year time period, national transportation policy should be directed toward immediate initiation of R&D on a national HSGT system. Indeed, such work has already been initiated but not on the scale and in the direction most suitable. The magnificent technical team assembled by NASA should be unleashed on this task.

Present studies for DOT are beginning to take the "system approach" to the overall problem of medium-stage-length transportation (80 to 800 km) in that most densely populated portion of the U.S. known as the Northeast corridor. Surprising to some is one critical area that developed out of DOT task force discussions. Research and development of new means of tunneling will be needed as this almost surely will be a serious bottleneck. Independent of the final mode of transport chosen, part of the journey probably will be underground. Mining engineers and others should be called upon to develop new methods of boring.[9] Government sponsorship of research in tunneling should stimulate sectors of our industrial base other than transportation. Other research and development needs established by these studies, such as fail-safe computer control of the vehicles and new methods of vehicle suspension and baggage handlings, are, perhaps, more to be expected.

Just as we will not call the personal vehicle described in the following an automobile, even though it operates more or less as a conventional auto when off the guideway, we will not call the HSGT vehicle a "train." A "train" as presently conceived is a vehicle under semiautomatic control operating on a fixed guideway* with a localized power unit and consisting of a number of connected but more or less separate carriages containing passengers and freight. The train stops periodically to take on and discharge its cargo and to disconnect or connect carriages to itself. Why, we might ask, is there such a unity then as a "train?" Its engine and cars and passengers and crew may change but the "train" still exists. The concept of a train was convenient because early engineers did not know how to design an efficient power source small enough to propel one car. Thermal efficiency is generally higher in the huge engines that pull long trains. Then, as the complexity of controlling several vehicles simultaneously began to mount, it was a comfort to the railroad dispatcher to concern himself only with one train of 100 cars rather than 100 trains of one car each. (At least one early U.S. railroad ran only east–west trains before noon and west–east trains after.) More recently, it must be admitted, labor practices on modern railroads have argued against short trains.

In the new technology, however, individual units carrying perhaps 50 to 100 passengers under continuous computer control, moving nonstop from origin to destination, will be quite practical. A ground vehicle under computer control, with a peak speed of 480 km/h and averaging 400 km/h, could compete successfully with present-day jet planes for stage lengths of less than 1200 km, city center to city center.

The concept of a "local" making intermediate stops will not exist. This will require, of course, a carefully designed interface with personal vehicle transportation and the HSGT terminal since only under exceptional circumstances will the HSGT terminals be closer than 160 km apart. The scheduling could be more like present-day airlines than present-day trains, with the added advantage that very small vehicles carrying as few as ten to 20 passengers for off-peak times or little-used stations could be economical.

It is in the ten-year time scale that for the first time one must take account of the reshaping and relocation of our cities in the design and installation of the national transportation system. Location of terminals can have a major long-term effect on the viability of these cities. It should be the aim of new city planning to provide by the year 2000 an HSGT link between all major cities less than 800 km apart. The development of substitutes for passenger travel such as economical videophone service must be carefully considered in this time scale. See Table II.

Coordination of metropolitan transportation with future needs

Near-range goals. Urban transportation improvements designed for implementation in the 0- to 5-year time period should be restricted to economical utilization of available technology. One should decide which projects to support by developing an understanding through system studies of the function to be served rather than by making random tests of available equipment. Suppose it is determined that articulation of the central business district is the immediate functional goal. It should be

* We use nonconnotative words such as "personal vehicle" and "guideway" as generic terms rather than more specific words because we are dealing with generalized concepts here. A new form of personal vehicle would use one type of guideway whereas HSGT might use quite another.

possible to move about between buildings within this area as easily as one now moves between floors of a given building. Then one surveys available technology to find the best match between the functional needs and performance characteristics.

Medium-range goals (3–15 years). Perhaps free minibus service along curbside lanes should be instituted in the central business district of several major cities. If so, commitments for long test periods (two years) should be made because in the past tests often were of such short duration that no impact could be made on the transportation habits of the typical user. The same is true for Dial-a-bus[10] tests in low- and moderate-traffic-density areas. Perhaps freeway flier service should be integrated into the Dial-a-bus service, which is a computerized dynamic-dispatched user-demand service using minibuses with drivers. Since Meyer, Kain, and Wohl[11] conclude that the private auto will retain its present high-preference position against existing modes of mass transport in existing cities (they say nothing about new modes in new city forms) and previous tests of existing mass-transit modes have not proved especially successful, one might ask why we should bother with further experiments with newer mass-transit modes. I would argue that very seldom has a transportation experiment been properly designed and controlled. Thus, the results are suspect. Past (and present) experiments seem involved with technology rather than being goal-oriented. The location and the duration of the tests seem to be chosen at random and restricted by available funds. Such tests may be worse than useless since their failure may be used to justify unwarranted conclusions.

Tests of new mass-transit proposals will provide some easing of traffic congestion and reintroduce a number of citizens to public transportation. They will also serve to show the irate urban dweller that something is being done about his plight.

II. Summary of national transport planning goals

Range	General	Specific	Contra-indicated
Short term (0 to 5 years)	Initiate immediately long-range studies (HSGT, etc.)	Metroliner tests, etc.	Massive network of new airports
	Properly use existing facilities	Begin tests of devices proposed for medium-term installations	
Medium (3 to 15 years)	Consolidation of existing railroad lines	Initiate new medium-speed rail passenger service	Continued restraint of competition among modes in freight service
Long (10 to 25 years)	HSGT	Connect major cities less than 800 km apart	Present manual, noncomputerized control of transport vehicles

Although modest extension and improvement of existing rail rapid-transit lines should not be ruled out, vast sums should *not* be spent in the near future on rail rapid transit. The failure of the new subways in Toronto and Chicago to attract new riders to public transport illustrates a stern lesson in this regard. Some 85 percent or more of the 1960 riders on the new Yonge Street subway in Toronto and the new Congress Street subway in Chicago formerly rode surface mass transit. Only about 12 percent were attracted from autos.[12]

The extension of the Cleveland rail rapid transit to connect to the metropolitan airport has proved moderately successful, but there are grave doubts about the financial success of BART in San Francisco. Calculated on the rather narrow grounds of investment amortized at the fare box, I base this (possibly presumptuous) prediction on the expectation that one could provide private taxi service for about the same total cost per passenger as BART. There is evidence, however, that fixed-guideway rapid transit encourages urban development with private capital along the route. This is to be wished, of course, and it may be that such a factor, along with reduced congestion, plays an important role in the agreed-upon effectiveness criterion upon which the success of BART will be judged. Both Toronto and Chicago plan to continue to invest in subway modifications and additions in spite of the ridership data reported.[12]

Of all the transportation strategies to be developed by system studies, perhaps the most difficult to establish objectively will be the program of urban transportation for the 3- to 15-year time period. It is quite apparent that the short-term modified status quo will not suffice nor are we ready now to opt unequivocally for the new city form that seems to be implied by long-range opportunities. The politician argues, "when in doubt, run in place and shout." In other words, do nothing.

On the other hand, a group of "new enthusiasts" for urban public transit in present cities has been spawned in the U.S. by the fertilizing effect of a recent round of HUD study contracts. A number of these reports ignore previous economic concerns about urban mass transit in existing cities such as those expressed by Meyer *et al.*, and, on the basis of little or no new evidence, recommend massive new systems that seem to contain all of the old difficulties. A recent Stanford Research Institute study appears typical of these.[13] In this study a determinedly optimistic note is maintained. On page 43, vol. I, for example, one reads that "any number of studies can be consulted that give pessimistic predictions for the future of new technology as applied to transit. In spite of the fact that this argument is widespread in the transportation literature, it is in error." Beyond this flat statement, no evidence is advanced as to why the pessimism is unwarranted; furthermore, this study itself, in my opinion, contains the following flaws.

- It badly underestimates capital and operating costs and entirely omits new mass-transit system R&D costs. My rough estimates yield costs two to five times as great as those given by SRI.
- Although accepting the present city form, the study ignores the causative effect of new transportation elements on changing the structure of the city. The whole analysis fails to consider the continuing change in function of the central business district and the spread of the city.

III. Summary of urban transport planning goals

Range	General	Specific	Contra-indicated
Short term (0 to 5 years)	Economic utilization of present technology	Scientifically design experiments for future field tests	Massive investment in rail rapid transit
Medium (3 to 15 years)	Test installation of new transport modes	Minibus, Dial-a-bus freeway flier, fixed guideway installations	Massive transport investment contributing to maintenance of present city form
Long (10 to 25 years)	Redesign of the city	Integrated, computer-controlled mass and personal transport modes	That the life of leisure will supplant the world of work

• Transfers are an integral part of the proposed SRI system. On page 292, vol. I, we read: "Passengers traveling on NET-1 systems first walk to PAS (Public Auto Services) stands; drive PAS cars to NET-1 (extended area services, similar to Westinghouse or BART system) stations; board the appropriate NET car, transfer to other NET cars at loop interfaces, if necessary, and travel to their destination stations; ride PAS cars to stand nearest their destinations; and walk the rest of the way." Then on page 147, vol. I, we continue: "Passengers arriving from external points would board MAC-1 (Major Activity Center systems based on moving sidewalk variation) at the outer edge of the CBD—on a typical trip a passenger would pass through one or two intermediate stops and transfer once [*sic*]. At the station nearest his destination, he would leave the MAC-1 conveyer, descend to the street level, and walk one or two blocks to his destination."

It seems unlikely, to say the least, that the public will accept a journey to work requiring five or six transfers plus perhaps a dozen or more intermediate stops and more than a four-block walk.

• Multistop or local service is the usual mode throughout the proposed system. The effect on discomfort and increased mean travel time are ignored.

• Several elements of the proposed system require central computer control in real time and on-line. Such devices do not exist. Size, speed, and cost are not the only problems. Safety, fail-safe operation, and graceful degradation of service are design concepts not realized by the present generation of computer architects. This is not as serious an objection as some others since such devices could be developed within ten years.

Perhaps the best approach to medium-range urban transportation strategy would be to implement elements of those networks that a system analysis indicates will meet goals in the 10- to 25-year time period, somewhat sooner than might otherwise be wise. Such field tests should be labeled frankly as experimental.

Early implementation might mean installation in or near existing metropolitan centers. These test installations should be designed not only to further certain agreed-upon goals for the new city by acting as a vector of change for the existing city, but also to be absorbed into the existing transport network of the city should the envisioned change in city structure be unsuccessful. This is important because a meaningful test could cost several billions of dollars and because we must face the fact that most attempts at urban planning to the present have been unsuccessful.

Long-range goals: ten years and beyond. It seems apparent that the mode of transport in a city and the city's form are initimately related. It is counterproductive to superimpose a particular transport mode on a geographic form that thwarts the very attributes of what caused the mode to be selected. The conventional round city with rectangular or radial grid implies transit velocity independent of direction. In addition, the rectangular grid seems to imply a democratic egalitarian sameness whereas a radial grid implies a centrality or dominance at the hub. It would be interesting to examine the religious and political persuasions of those who laid out the street plans of various cities.

To exploit properly the attributes of high-speed fixed-guideway mass transport seems to imply high-density nodes like separated beads on a string. A high-density population node argues against individually controlled personal vehicles within the node. The modern American will not easily give up his love affair with his harem of autos.

The foregoing is not a series of disconnected thoughts but a sample of precepts on transportation that are susceptible of systematic verification. These and others like them should be tested and, when verified, factored into the design of the new city. This is not the place to discuss in detail the design of the new city, but I think one can show that transport plays a more critical role in the articulation of the city than is generally understood by architects. For example, one can demonstrate in fairly concrete terms, by means of a system analysis of transportation, the financial diseconomy of allowing a city to exceed a certain population range. Thus, one important long-range goal is to redesign the city to suit its inhabitants and to exploit the features of modern means of communication and transportation; see Table III.

Planning, laissez-faire, or the new empiricism?

Results of long-range studies will be to require the redesign of the city. By now such a suggestion has lost its shock value. Those who despair of the present city have advocated redesign, and arrogant planners fresh from their most recent failure in rearranging the life style of the masses have advocated it. Usually such plans are sterile and lifeless from the moment of their conception on the drawing boards of the high priests of planning. They often are based on concepts that have been proved failures *in vivo* around the world or are futuristic fantasies designed to display one or two ideosyncratic notions by which the chief planner hopes to be remembered by posterity. An example of the former is Brasilia and of the latter Le Corbusier's Marseilles stilt apartment dwellings.

Let me propose two lemmas, by which I hope to lead to a theorem.

Lemma I. There is no man living who has a mind adequate to develop a structured city plan that will serve the city's citizens into the indefinite future. The proof is twofold. First, there has never been an exception. Planning has influenced cities, of course. Rome, Moscow, Washington are examples of this. But in no case has the plan been uninfluenced by the reality. Second, psychologists tell us that one can place little credence in expressed preferences based on hypothetical choices rather than experience. Thus, no planner, nor even future residents themselves, can predict which proposed city plans will "work" and which will not. To freeze the form of a new city on the basis of unrelieved failure in the past and the unknowable future desires of its citizens surely earns our planners the rank of high priests of mysticism.

Lemma II. If no planning is done, continued degeneration is in store. The proof lies in recent history, and in another simple observation. Kenneth Boulding, the articulate economist, recently discovered what he calls the real name of the devil. It is not Lucifer, but "Suboptimization." Boulding defines suboptimization as that plan or program which, by concentrating exclusively on a narrow goal and ignoring subsidiary or ancillary effects, contributes overall evil rather than good. Such narrow artifacts as DDT, the Asswan Dam, and deep-well dispersion of atomic wastes, not to mention urban superhighways and "low-cost" mass housing, are in this category. Boulding is right, but he has discovered only the tip of the iceberg. Our whole economic structure is geared to short-run gain and our whole technological philosophy is based on subdividing a problem until its elements seem to be manageable, i.e., suboptimizable. Thus laissez-faire is not the answer.

Theorem. Planned empiricism will save us. We must extract those elements of cities that work, whether they appeal to our personal professional persuasion or not, and use them. Somehow, Doxiadis,[14] with his inward-turned shells of human scale designed for walking, and Mumford,[15] with his worship of the decaying centers of empire and their monumentalism, must come to grips with Jane Jacobs,[16] her antigrass fetish, her concept of the multiple use of urban space, and her ballet of the sidewalks.

In practice this means that we start by admitting that we do not know all of the answers for the new city. Without imposing a rigid structure a priori, we must find out what works by systematic and planned experimentation. The dialogue must be continuing. It must include those practical philosophers, the politicians; as well as planners, technologists, and citizens. The system engineer will be needed to construct the models and make projections showing the end results of various options. If the traditional planners and "civic leaders" accept this rather unusual role of cooperation rather than attempt domination, the system engineer must be very careful that he does not assume the mantle of the former high priests of planning. It is easy for the system analyst to confuse his proper activity of analysis and exploration of options with the improper usurpation of unilateral decision-making authority. Perhaps the most obvious example of this confusion of roles is exhibited by the urban freeway planner. The common result is the highway revolt (for example, the Baltimore cross-harbor proposal, the San Francisco Embarcadero Freeway, the New Orleans French Quarter Freeway proposal, the Everglades Airport proposal, the proposed Hudson River Expressway, the I-698 route location in Detroit, and controversies in Washington, D.C., Indianapolis, Cleveland, and Philadelphia). Successful coordination of communication and transport planning for the closing decades of the 20th century will involve the system engineer in the political process. Neither he nor it is likely to emerge unchanged.

This article is based on an invited lecture delivered at the 17th International Congress of Transport and Communication, Genoa, Italy, October 8, 1969.

REFERENCES

1. Fogel, R. W., *Railroads and American Economic Growth*. Baltimore, Md.: The Johns Hopkins Press, 1964.

2. Seifert, W. W., "Status of transportation," *Proc. IEEE*, vol. 56, pp. 385–395, Apr. 1968.

3. Brook, J., *Business Adventures*. New York: Weybright and Talley, 1969, chap. 6.

4. Hult, J. L., "Cheap communications," *The Futurist*, pp. 70–72, June 1969.

5. Gibson, J. E., "National goals in transportation," *Proc. IEEE*, vol. 56, pp. 380–384, Apr. 1968.

6. Doxiadis, C. A., "Emergence and growth of an urban region, vol. 1," Analysis, Detroit Edison Co., 1966.

7. Mayer, M., *The Lawyers*. New York: Dell, 1968, p. 351.

8. Fugii, M., "New Tokaido Line," *Proc. IEEE*, vol. 56, pp. 625–645, Apr. 1968.

9. McGarry, F. J., "Underground tunnels for transportation systems," *Proc. IEEE*, vol. 56, pp. 535–544, Apr. 1968.

10. Guenther, K. W., and Givens, W. E., "Dial-a-bus," SAE Paper 700184, Jan. 1970.

11. Meyer, J. R., Kain, J. F., and Wohl, M., *The Urban Transportation Problem*. Cambridge, Mass.: Harvard University Press, 1966.

12. Burch, P. H., Jr., *Highway Revenue and Expenditure Policy in the U.S.* New Brunswick, N.J.: Rutgers University Press, 1962.

13. "Future urban transport systems," final report, vols. I and II, Stanford Research Inst. for U.S. Dept. of Housing and Urban Development, Mar. 1968.

14. Doxiadis, C. A., *Ekistics*. New York: Oxford, 1968.

15. Mumford, Lewis, *The City in History*. New York: Harcourt, 1961.

16. Jacobs, Jane, *The Death and Life of Great American Cities*. New York: Random House, 1961.

John E. Gibson (SM) received the B.S. degree from the University of Rhode Island in 1950 and the M.S.E.E. and Ph.D. degrees from Yale University in 1952 and 1956 respectively. From 1952 to 1956 he was an instructor in electrical engineering at Yale and in 1956 was named assistant professor. In 1957 he joined Purdue University as associate professor of electrical engineering, where he organized and taught a number of graduate courses in automatic control. His research activities have been in dual-mode control systems, nonlinear system stability theory, describing-function studies, adaptive control, and optimum control systems. He was appointed professor of electrical engineering at Purdue in 1960 and in 1961 was appointed director of the newly formed Control and Information Systems Laboratory. In July 1965 he was appointed to his present position of dean of engineering at Oakland University, Rochester, Mich.

Dr. Gibson has published approximately 40 papers and 25 technical reports in the field of automatic control. He is a coauthor of the book *Control System Components* (1958) and the author of the books *Nonlinear Automatic Control* (1963) and *Introduction to Engineering Design* (1968). His memberships include Tau Beta Pi, Sigma Xi, ASME, and the ASEE.