IRE Transactions on AUDIO

Volume AU-10 JANUARY-FEBRUARY, 1962 Number 1

Published Bi-Monthly

TABLE OF CONTENTS

PUBLISHED BY THE

Professional Group on Audio

IRE PROFESSIONAL GROUP ON AUDIO

The Professional Group on Audio is an organization, within the framework of the IRE, of members with principal professional interest in Audio Technology. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of an annual fee of $2.00.

IRE TRANSACTIONS® ON AUDIO

Published by The Institute of Radio Engineers, Inc., for the Professional Group on Audio, at 1 East 79 Street, New York 21, N.Y. Responsibility for the contents rests upon the authors, and not upon the IRE, the Group, or its members. Individual copies of this issue may be purchased at the following prices: IRE members (one copy) $2.25, libraries and colleges $3.25, all others $4.50. Annual subscription price: non-members $17.00; colleges and public libraries $12.75.

The Editor's Corner

HIGH-FREQUENCY BIAS

CONCERNING magnetic recording, the question I have been asked most frequently in the past twenty years is "How does high-frequency bias work?" "Can it be explained by a simple analogy?" There is good reason for this curiosity, because high-frequency bias is one of those things that you would not expect to work, but it does; which is a happy departure from the usual situation where things always turn out less well than anticipated.

If you should try to record directly on a magnetic tape by feeding only an audio signal to a recording head, the result would be extremely distorted because the permanent-magnet material used for the tape is non-linear. But if you add a strong steady high-frequency current to the audio during recording, the magnetization retained by the tape is a faithful replica of the audio signal and gives practically undistorted playback. The high-frequency bias disappears somewhere in the process, and not even a trace is detected in the usual output.

To explain this, I used to draw B-H curves and minor hysteresis loops, but soon realized that most people don't want you to "get technical." So I changed my story and told them that high-frequency bias shakes up the magnetic domains, and loosens them until they can be molded by the audio signal. Most nonspecialists were happy with this explanation. I even detected that some were more confident of their mastery of the subject than I myself was.

But the specialists wanted something with more meat on it. Although several different theories had been published, nobody seemed to like anyone else's theory; so everybody proceeded to cook up one of his own. Some used major and minor hysteresis loops. Others chose the B_r-H curve. There was a mathematical explanation using Fourier series, and a graphical explanation using the "theorem of the mean." Still another tack was that a pulse-width modulation was produced in the tape. It was pointed out that high frequency bias resembled the anhysteretic magnetization process; however the anhysteretic process itself was not understood.

Recent work has thrown additional light on anhysteretic magnetization. The picture is complicated with Preisach diagrams, particle interactions, and single-domains. But when these are all considered they account for the fact that residual magnetization is proportional to the low frequency component of the applied field. Or a rough way to describe this latest most advanced theory is that the high frequency bias shakes up the magnetic domains and loosens them until they can be molded by the audio signal. (Where have we heard this before?)

You might think the detective story ended here, with all clues solved and missing persons found. But an interesting fact is that high frequency bias will work even *without the anhysteretic process*. That is to say, if anhysteretic magnetization gave a curved instead of a straight line, or had an infinite slope with no gradation between positive and negative saturation we would still be able to obtain linear distortionless magnetic recordings.

Evidently more than one linearizing process is at work here. It may be more accurate to combine some of the explanations instead of relying on a single one, and excluding the others. After all, the product of several linearizing operations is still linear and we all agree that we can't have too much linearity in magnetic recording.

MARVIN CAMRAS, *Editor*

PGA News

ACOUSTICAL SOCIETY AND PGA HOLD JOINT SESSIONS

Audio news from the recent Acoustical Society meeting:

The 62nd Meeting of the Acoustical Society was held in the Netherland Hilton Hotel in Cincinnati, Ohio, on November 8–11, with sessions at Wright-Patterson Air Force Base, near Dayton, on November 9. Of special interest to the PGA was the all-day session on November 9, which was arranged with the cooperation of the IRE Professional Group on Audio. The papers submitted were as follows:

Joint Session G. Communication

Cyril Harris, *Chairman*

G1. Nonacoustical Means of Communication. J. R. Pierce, *Bell Telephone Labs., Inc., Murray Hill, N. J.*—The human voice is a favored means of communication because we are all trained to use it face-to-face, and we can speak over the telephone without any further training. Nonetheless, writing, typing, diagrams, and pictures play an important role in human communication, and today we have data communication between machines. All of these nonvocal forms of communication will come increasingly into use. This paper will discuss some of their possible uses and some of the obstacles in their path.

G2. Voice Communication in Aircraft and Aerospace Systems. P. S. Veneklasen, *Western Electro-Acoustic Lab., Los Angeles, Calif.*—As the boundaries of man's missions above, around, and beyond the earth enlarge into unlimited space, new problems are presented in voice communications. The solutions to these problems are essential to the continued effectiveness of man in military aircraft and spacecraft. Man will be an essential and controlling factor in the space environment, and his usefulness and effectiveness there will be, as in all his historic past, utterly dependent upon communications: language, voice, and hearing, received and transmitted. Missions near the earth will be of shorter duration, higher speed, and higher noise level. Space missions will probably be less noisy, but their duration will make the 40-hour limit appear as a moment. Extreme discomfort associated with communication equipment has critically hampered the operational effectiveness of flyers on long range missions. We must discover new transducers which will permit reduction in the encumbrance of combined communication, protection, and respiration equipment worn by flyers during high altitude, long range missions. The noise field largely determines the possible techniques for speech projection and reception. The existing and prospective noise fields for aircraft communications are reviewed. Both projection and reception systems consist of a transducer, a coupling means from transducer to voice or hearing, and a noise exclusion device. A morphological chart of available devices and techniques is given. Criteria are described for evaluation of the effectiveness of a transducer system, involving not only advanced psychophysical parameters of speech intelligibility, but also equivalent physical factors peculiar to flight communication in high ambient noise. Evaluation of encumbrance or wearability in terms of long-term pain tolerance greatly needs quantitative procedure. New transduction techniques are described and compared with conventional methods to illustrate the state of the art and the problems and prospects for future advancement.

Session K. Speech Compression Systems

Michel Copel, *Chairman*

K1. Low Bit Rate Digital Speech Communication. L. G. Stead, E. T. Jones, and R. C. Weston, *Signals Research and Development Estab., Christchurch, England.*—Preliminary results and experience with a two-way analysis/synthesis formant tracking speech system are reported. Parametric information from the analyzer is coded into binary digital form at rates in the range 600 to 1200 bits per sec and transmitted over audio channels by high-speed voice-frequency telegraph equipment. (Paper presented by Walter Lawrence.)

K2. Bandwidth Compression of Speech by Spectrum Sampling. K. D. Kryter, *Bolt Beranek and Newman Inc., Cambridge, Mass.*—An experimental communications system has been built that filters speech into a number of narrow bands. The width of the bands, at the 3-db downpoints on the filter skirts, can be set at 25, 50, 100, 300, or 500 cps; the positions of the center frequencies of the filters are variable. A number of tests were conducted of the intelligibility of speech transmitted over a wide variety of combinations of filters having different bandwidths and different center frequencies. A few listening tests also were made of the effects on intelligibility of sequentially sampling in time the output of the various narrow band filters. The results of these tests show that intelligibility is a joint function of the number, width, and position on the frequency scale of the narrow pass band filters. The most efficient combination of filters we have found to date, in terms of total effective bandwidth required for 80 per cent PB word intelligibility, consists of five filters each 100 cps wide measured at the 3-db downpoints on the filter skirts. This represents a bandwidth saving over a low-pass system (simulated telephone) by a factor of about 3 or 4 to 1; we previously reported (*J. Acoust. Soc. Am.*, vol. 32, pp. 547–556; 1960) a bandwidth reduction of 2 to 1 with three filters having bandwidths of 500 cps. (This work was performed under contract with the U. S. Army Signal Research and Development Agency.)

K3. Design vs Performance Factors for Some Speech Compression Systems. C. P. Smith, *Electronics Research Directorate, AF Cambridge Research Labs., Bedford, Mass.*—In selecting the design parameters for a speech compression system, perhaps one of the most difficult problems involved is that of deciding what compromise to make between the projected data rate of the system and the performance to be achieved by the system (in terms of such factors as intelligibility achieved with words, syllables, and connected speech; voice quality factors; ability to recognize a familiar talker over the system; interactions of rate of talking with other performance factors, and so on). Some aspects of this problem are discussed and illustrated with data.

K4. A Military Semi-Vocoder for Analog Telephone Transmission. A. S. Howell, G. O. K. Schneider, and T. M. Stump, *General Dynamics/Electronics, Rochester, N. Y.*—A semi-vocoder for telephone speech is described. It was developed for an Air Force-envisioned system of multiple analog vocoders fed from any telephone line to operate over a single voice frequency channel. Such a semi-vocoding plan is shown to offer advantages in the maintenance of voice quality, an important factor in user acceptance. It comprises 500 cps of unvocoded baseband and 250 cps, or 13 spectrum channels, of vocoded information. The direct baseband provides spectrum and excitation information; it facilitates the derivation of pitch when the fundamental frequencies of talkers are missing. Also, the processing of the direct baseband to provide automatic voice excitation obviates the need for an arbitrary voiced-unvoiced decision and excitation by buzz-hiss switching. Errors associated with the latter method are accentuated by the limited frequency range of telephone inputs. The vocoded information is processed by the direct time sampling of the unrectified spectrum channel energy. This method of identifying the speech spectral pattern is entirely feasible when certain requirements are recognized. The frequency occupancy of the complete semi-vocoded speech channel is directed toward multiplexing within a 900-cps total bandwidth inclusive of guard space.

K5. Analog Multiplexing of a Telephone Semi-Vocoder. A. S. Howell, G. O. K. Schneider and T. M. Stump, *General Dynamics/Electronics, Rochester, N. Y.*—Narrow-band analog multiplexing means accommodating three telephone semi-vocoders, each comprising 500 cps of unvocoded baseband and 250 cps of vocoded spectrum information, within a nominal 3-kc bandwidth are discussed. Developed under Air Force contract, the plan involves analog level companding of the vocoded and time-sampled spectrum energy. Transmission of the latter information is accomplished by negative modulation in conjunction with vestigial sideband techniques. The unvocoded baseband content is transmitted by single sideband. A compatible synchronizing system constitutes an essential part of this plan. Design problems accentuated by narrow-band confinement of the multiplexed information are pointed out. It is shown that severe response requirements are imposed upon the vestigial filters. Associated sampling techniques are necessary in order to minimize inter-channel spectrum cross-talk. A TDM control system providing the necessary freedom from mutual interference between the transmitted spectrum intelligence and the receiver synchronizing information is described.

Session H. Speech Communication Equipment

B. B. Bauer, *Chairman*

H1. General Design and Performance Characteristics of a Military Voice Communication System. State of the Art Appraisal. W. B. Snow, *Bissett-Berman Corp., Santa Monica, Calif.*—The general requirements, plan, and equipment for on-board voice communication in military aircraft are described. The characteristics and functions of essential components, including speech projectors, receptors, and intervening electronic equipment are illustrated by conventional current and improved equipment. This equipment and its design problems are demonstrated in relation to requirements dictated by personnel protection, respiration, and altitude provisions. The electrical, mechanical, and acoustical characteristics of prospective transducers, conventional and unusual, including microphones, earphones, loudspeakers, and accelerometers, are reviewed as part of a morphological study of speech projection and reception techniques. Evaluation is based upon standard electrical and acoustical ratings and newly devised ratings appropriate to the problem, as well as engineering judgment regarding adaptability to flight systems, including wearability. The comparisons show great promise for the electrostatic transducer as an earphone or loudspeaker.

H2. Noise Environment and Control—Present and Prospects; Evaluation of Speech Projection and Reception Systems. P. S. Veneklasen, *Western Electro-Acoustic Labs., Los Angeles, Calif.*—The noise spectra in present military aircraft are reviewed, describing the characteristic sources. Objectives for noise control based upon communications, aural damage risk, and comfort are proposed. The probable maximum increases in noise exposure are described, and the prospects for control in future aircraft and space vehicles are discussed. Requirements for personal protection are deduced. Latest methods of evaluation of speech projection and reception systems are based both on physical measurements and intelligibility testing. Highly efficient articulation testing is based upon most difficult consonant sounds and upon simulated flight command sentences. Results permit rating of systems in terms of noise tolerance as a function of speaking effort.

H3. Speech Projection and Reception Systems—Developments, Characteristics, Prospects. J. P. Christoff, *Western Electro-Acoustic Lab., Los Angeles, Calif.*—The search for transducers which would permit reduction in the encumbrance of communication equipment for use in high-altitude long-range missions has led to the discovery of new techniques for the projection and reception of voice communication. Absolute and quantitative measurements have been made of the relative performance of speech projection methods—conventional and radical—to define speech spectral sensitivity, noise spectral sensitivity, speech articulation capability, and noise tolerance as a function of speaking effort. Similarly, speech reception techniques have been studied quantitatively to find promising new approaches to eliminate head contact or discomfort while maintaining adequate noise tolerance. The performance of two new approaches, the forehead contact transmitter and the horn-coupled receiver, are described in detail.

CHAPTER NEWS

Chicago

On November 14 the PGA had a joint meeting with the Chicago Chapter of the Professional Group on Military Electronics at the Armour Research Foundation. The paper presented was by Henry Karplus, Research Physicist of Armour, on the subject "Seeing with Sound Waves," abstract of which is as follows:

Novel techniques will be described for taking X-ray-like pictures of internal flaws in opaque materials using ultrasonic waves. Sound waves have greater versatility than X rays because they can be reflected and refracted so that it is possible to "see" around corners; images may be formed by the use of suitable lenses and mirrors. The heart of the system is an image converter which works like a TV iconoscope with the light-sensitive film replaced with a piezoelectric tranducer. The final picture is seen on a synchronously driven kinescope screen.

The basic principles of acoustic propagation in solids and liquids will be reviewed and other techniques of making sound waves visible will be described. These involve direct interaction of sound and light.

The Chicago Section of IRE had a lecture-demonstration of FM stereophonic broadcasting for its opening meeting for the 1961–1962 season on September 12, 1961. The principal speaker was Carl G. Eilers, Research Engineer at the Zenith Corporation in Chicago. Mr. Eilers demonstrated the first broadcast of stereophonic FM from Station WEFM, Zenith's FM Radio Station. The stereophonic broadcasting began at midnight, June 1, 1961, the earliest time authorized by the FCC.

Philadelphia

A meeting of the Philadelphia Chapter was held on December 8, at Studio One—WCAU. Mrs. E. L. R. Corliss, the National Bureau of Standards, talked on the subject "Electrostatic Loudspeakers—Transient Distortion in Transducers." During the talk a demonstration was given of the newly developed electrostatic loudspeaker which has high excursion and frequency range between 200 and 10,000 cps.

ANNOUNCEMENTS

Call for Papers

A call for papers is being made for the 1962 Chicago Spring Conference on Broadcast and Television Receivers. Special sessions are being arranged on audio, particularly as it applies to the home entertainment radio and television industry. Deadline is February 15, 1962. At their earliest convenience, potential authors are asked to send in triplicate a 50 to 100 word summary including title of the paper, name of author(s), together with company affiliation and position. Papers should be limited to approximately 2500 words and the presentation to twenty minutes. Papers should be submitted to:

Mr. Al Cotsworth
Papers Committee
Zenith Radio Corporation
6001 West Dickens Avenue
Chicago 39, Ill.

International Congress on Acoustics

The Fourth International Congress on Acoustics, sponsored by the International Commission on Acoustics, will meet August 21–28, 1962, under the auspices of the Acoustical Society of Scandinavia. The meeting will be held at the Royal Technical University, Copenhagen, Denmark. Contributed papers will be given in English, French, and German. Papers which have already been published cannot be accepted and preference will be given to new material of scientific interest. Summaries of not more than 50 words should be submitted by April 1, 1962, to the Secretariat of the Congress:

Fourth International Congress on Acoustics
Oestervoldgade 10
Copenhagen K
Denmark.

Current Research in Sound*

VINCENT SALMON†, SENIOR MEMBER, IRE

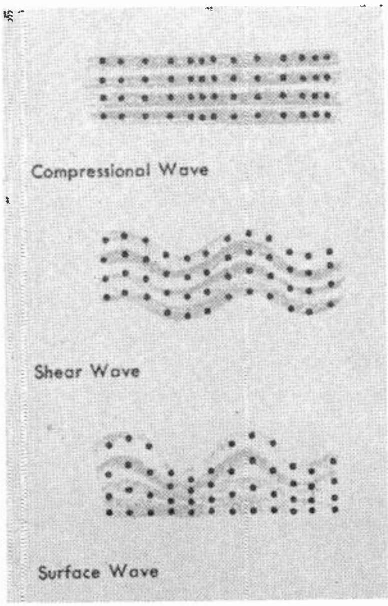
Summary—Sounds in industry may go unheard by the ear, but they are all-important to "listening" machines, compounds, and processes. Already with sound we can test, mix, measure, clean, drill, and control. And research indicates broad new applications in the tomorrow.

THE SKILLED MUSICIAN utilizes his talents and instruments to evoke, alter, and control the emotions of his audience. To a similar extent the well-informed production engineer can call on the resources of acoustics to measure, alter, and control the behavior of many recalcitrant industrial processes and materials. These applications of sonic energy we may class under industrial acoustics, so named out of analogy to industrial electronics.

Because industrial acoustics is as yet not a well-defined activity, estimates of its size vary. The annual dollar volume of pertinent equipment manufactured in the United States is probably in the $60–75 million range. Most of this is in sonar, underwater telephone, and fathometer devices. Equipment for other purposes accounts for scarcely $20 million, most of which is for flaw detection devices. But developments are coming so rapidly that these figures may be out of date when printed. Much of the expansion is in nonmilitary applications, especially those involving energy for industrial processing.

The worker in industrial acoustics has many sonic parameters to bring to bear on the problems brought to him. He sometimes can select the medium in which he works and its wave transmitting properties of sound velocity and absorption. More often than not, this medium is selected for him, because it is actually the workpiece to be tested or processed; and the appropriate wavelength and sound intensity range must be selected to achieve the desired effect. In a solid, functional requirements of the problem will often dictate whether the medium should be squeezed by compressional waves, rubbed by shear waves, or rippled by surface waves. In nondestructive testing and acoustic control, the information desired may lie in measurements on absorbed, reflected, bent, or scattered sound waves; often transit time, resonant frequency, and mechanical impedance are measured. The parameters are usually selected to maximize the information in an unequivocal and repeatable fashion.

In sonic processing, the desired effects usually depend on cavitation—that is, the generation of shock waves by rapidly collapsing bubbles in a liquid. The industrial acoustics engineer has all these major sonic parameters at his command and is also alert to the results from pure research that he may add to his acoustic armamentarium.



Types of sound waves in solids.

Testing with Sound

Sonic testing may be destructive or nondestructive. The necessity for the destructive testing arose in part from damage to early jet aircraft caused by intense noise from improperly located engines. This noise can also cause mal- or nonfunction of electronic gear. The test problem is to recreate jet noise of the desired intensity and frequency range without a jet engine. At present, sirens and modulated airstream loudspeakers (a mechanical version of human voice production) are employed, but neither is completely satisfactory. What is needed is a device with the efficiency of a siren and the ability of a modulated airstream loudspeaker to deliver several frequencies simultaneously.

The present research attack entails giving each device some of the characteristics of the other. An intriguing possibility lies in using the fact that a vacuum tube amplifier is in many ways analogous to a modulated airstream loudspeaker. This analogy suggests the use of the

* Received by the PGA, January 30, 1961. Reprinted from the *Stanford Res. Inst. J.*, vol. 4, 1st Quart., 1960. This is the second in a series of articles reviewing current research in acoustics and audio. The first article, entitled "Current research in communication acoustics," appeared in IRE TRANSACTIONS ON AUDIO, vol. AU-9, pp. 37–40; March–April, 1961.

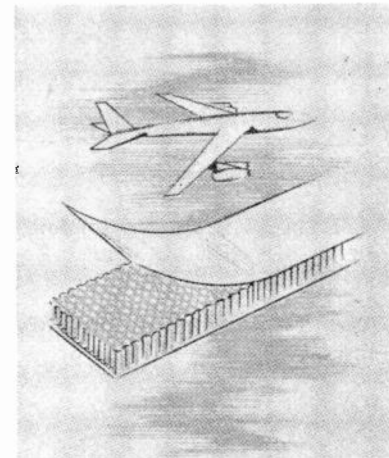
† Stanford Research Institute, Menlo Park, Calif.

positive portion of an acoustic wave to control the blowing of air through one modulator and the negative portion to control air sucked into another modulator. Thus, the total acoustic wave is synthesized by assembling the positive and negative portions. The realization of this blow-suck modulator depends on improvements in our knowledge of still intractable flow equations. Perhaps the next decade will see this roadblock removed or at least reduced to a minor obstacle.

Nondestructive testing is concerned with uncovering such properties as the strength of adhesive or welded bonds, viscosity, elastic constants, or grain size. The major application is the location of flaws in solids. Here, the principal sonic parameter is wavelength, for when the wavelength is much larger than the flaw, the wave is unaware of the presence of the flaw. For the flaws of industrial interest (in metal parts), this usually means that ultrasonic frequencies are necessary. For example, to delineate defects larger than $\frac{1}{8}$-inch extent in steel, the frequency of the probe wave should be over one megacycle per second (one million cps). Ordinarily, compressional pulses of the sound are transmitted, and the instrument "listens" for reflections from defects. The time of transit to and from the defect locates it. Many such devices are on the market, but they are hard pressed when the flaw is close to the source of the ultrasonic pulse. In this case, the echo may return while the transmitted pulse is still continuing and thus may be hidden in the transmitted pulse. At present pulse-echo devices using some species of surface wave appear fairly promising in the inspection of near-surface flaws. The precise type of wave used depends on whether the flaw is laminar (parallel to the surface) or a crack (perpendicular to the surface). Operating difficulties remain, and we shall have to rely on further research for an answer.

Glued Aircraft

The appearance of high-strength organic adhesives has made possible aircraft (such as the B-58) that are almost entirely "glued" together. Nondestructive, nondamaging inspection of such bonds in both manufacturing and maintenance involves a technique using mechanical impedance. (This quantity is the ratio of vibratory force to vibratory velocity, and plays the same part in vibration and sound that electrical impedance plays in its field.) If we vibrate an adhesively-bonded skin, its mechanical impedance depends on the quality of the bond, and several means are available for displaying this dependence. In addition to exposing a flaw, in many instances the instrument readings show good correlation with the actual bond strength, determined destructively on a statistically significant number of calibration samples. The technique has applications to welding inspection and possibly to determining the temper or anneal in glass and metal.



Much structure in supersonic aircraft is made from a honeycombed metal "sandwich," *i.e.*, the aircraft is "glued" together. The bonds of the structure can be tested nondestructively with sonic instruments.

Other Nondestructive Tests

The elasticity of a material depends on both the constitution of its structural elements and how these elements are assembled. For example, in a polymer, both the elastic properties of each molecule and the degree of crosslinking between molecules control the overall elastic properties. Usually these properties must be described in terms of energy storage and energy loss, under the general heading of viscoelastic behavior. These properties can be measured with sound. Such tests are commonplace in the elastomer and textile industries, and often serve in production control as well as in product development. The extension of similar sonic testing techniques to problems in ceramics, glass and metals technology is but a matter of time. Even our present limited probing in space flight environments has exposed some severe problems. Radiation found in space can markedly alter the internal structure of materials. These effects can be simulated in the laboratory where sonic testing may be used to measure the degree of alteration.

An important example of nondestructive testing is the location of sizable foreign objects in a large body of liquid, as a lurking submarine. We know this technology better as sonar, but in principle it utilizes much the same equipment as that used in the laboratory for pulse-echo ultrasonic defectoscopy. However, the defect may now be tens and even hundreds of feet long, and the distances may be in miles. Instead of a stable and excellently-conducting medium, we have to deal with a restless, changing ocean that shatters sound beams, and bends and reflects them into an almost unrecognizable form. Research to overcome or circumvent these natural limitations is of prime importance in national defense. Perhaps the sense of urgency felt by many workers in sonar technology will communicate itself not only to acoustical scientists but also to researchers in the allied contributing fields of fluid dynamics, oceanography, information theory, computers, and what might be termed bathymeteorology, or "underwater weather study." The sonar scientist needs all the interdisciplinary help he can get.

Almost everyone has run into situations for which nondestructive tests would have prevented dangerous or unpleasant results. The ripe egg in the frying pan, the fermenting canned fruit, the windshield that dissolves at the first pebble strike, the knothole that may appear when plywood is cut, the welded Liberty ship that cracks in two—all of these involve defects that may possibly be located by sonic testing. Laboratories engaged in such research can be sure of one thing—there is no dearth of problems.

Measuring with Sound

Sound has been successfully applied to the measurement of liquid flow in pipelines, using the fact that the motion of the medium changes the apparent sound wave velocity. Still, there does not exist a general-purpose portable, sonic flowmeter that can quickly be clamped on a pipeline. The chief difficulties lie in turbulent flow in the pipe and transmission of signals in the walls of the pipe, rather than in the liquid. The researcher who first develops an easily-applied, portable, sonic flowmeter will meet an enthusiastic reception from the chemical and petroleum industries.

A portable sonic flowmeter that can be clamped on a pipeline is a research goal.

The Doppler effect is used to measure the velocity of a moving surface. This is the change of apparent pitch of a rapidly moving sound source that passes us, such as the drop in pitch we hear in the whistle of an approaching train as it comes alongside us. If a beam of sound is directed at a moving surface, then the reflected sound returns altered in pitch. Thus, the velocity of the surface can be detected without touching it. Traffic police radar does this with radio waves that reach farther than possible with ordinary sound.

With a modified sonic instrument, the spacing between fixed and movable objects may be measured. Such devices should have all sorts of application where dial or feeler gages are now used; for example, the eccentricity of rotating parts can easily be measured. Or a sonic altimeter could help helicopters land or maintain a fixed altitude above, say, a heaving sea. A similar device, horizontally directed, in automobiles would serve to indicate the distance of the car ahead in heavy traffic or dense fog.

Chemical Processing Methods

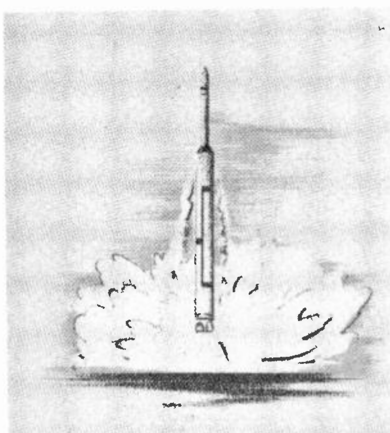
Much of sonic processing involves working in a liquid medium with sound intense enough to produce cavitation. As explained, this action results in the repeated production of intense short-range shock waves, which in turn are responsible for many of the effects observed. Cavitating sound can cause chemical reactions to begin or to proceed faster. Thus, controlled chemical attack on hard metals, known as chemical milling, can be speeded. Indeed, with proper direction of the sound beam, the chemical attack may possibly be made more selective and controllable.

In electrolysis, sonic energy can promote more uniform dissolution of the anode, and can alter the character of the deposit. Many of the effects are specific to the particular electrolyte, anode, and cathode employed and results have often not been what was desired. Ordinarily, sonic energy is regarded as another additive to the bath, when, in reality, it affects the electrode processes so profoundly that the optimum bath and current density may differ with and without sonic activation. Moreover, all too frequently sonic devices have been evaluated by immersing a sound source of unknown characteristics and output, with no quantitative evaluation. Only careful quantitative research can supply data on which engineering and economic decisions can be based.

Intense cavitation can rip apart long molecules. Thus, enzymes can be deactivated, and liquids having a large range of high molecular weights can be reduced to smaller molecular weights. This has primary application in finishes such as lacquers and varnishes, where the film-forming properties are improved by molecular weight control. However, no large-scale exploitation of this application of sonic processing has been tried yet, probably because of the high cost of the sonic energy.

An interesting application occurs in the extraction of fat from bone meal in the manufacture of glue and gelatin. Ordinarily, hot solvent extraction is used, but the extraction is incomplete, and the end product may be degraded by heat. Cavitating sound has been found to effect more complete separations at temperatures low enough so that the fat is unharmed. But the cost of sonic energy generated by conventional means is so great that it is infeasible.

One important chemical effect lies in the apparent ability of intense sound to control combustion. The glamorous example of the moment is the increase in burning rate occurring in solid-fuel rocket motors where intense self-generated oscillations can build up, sometimes until the motor is destroyed. With controlled use of this effect, we may be better able to direct rockets or to obtain more energy from oil burner flames. Research is barely getting under way, and 1961 may see the first concrete results. A major difficulty is the choice of the means of sound generation; sirens, whistles, and modulated airstream loudspeakers have been suggested.



Intense sound can control combustion, thus, perhaps, leading to more efficient rockets.

Physical Processing Methods

Physical effects of cavitating sonic energy are best typified by the many sonic cleaners on the market. In these cleaners, tiny, intense shock waves caused by cavitation literally blast the soil off the surfaces to be cleaned. Sonic cleaners provide expensive energy, and hence are most feasible where the piece part cost is high. Ball bearings, instrument bearings, gyros, hypodermic needles, and shaver heads are typical examples. Effective application of sonic energy to home dishwashers and the like does not now appear promising.

Watch cleaning is a good example in which sonic cleaners will probably supersede others. The watch may be cleaned, without disassembly, much more thoroughly than by conventional means. The decreased handling also means less chance of damaging the delicate works.

The sonic cleaner industry is gradually divesting itself of the emphasis on ultrasonic energy, for better effects are often obtained at audible frequencies. A major unsolved problem is how to specify cleaning efficiency or even intensity of cavitation. At present we do not even know what questions to ask, much less the answers. But the problem has been recognized.

Cavitation has been used in a demonstration of the emulsification of mercury and water, a laboratory curio only. Cavitation-induced shock waves shatter the mercury into very fine droplets that remain suspended for a relatively long time. Ordinarily, the energy costs too much for such mundane tasks as homogenizing milk or making mayonnaise. Probably sonically-produced emulsions will always be a laboratory accomplishment, and industrial applications will be only for the really difficult tasks.

A physical effect of cavitation that is of industrial importance is the increased transfer of heat through a liquid-solid interface. The sound produces an intense stirring action that is concentrated at the interface, and the heat transfer rate may be increased by 200 to 300 per cent. There is potential application to heat transfer in nuclear reactors, where the local heating problem on fuel elements is still incompletely solved. However, large amounts of sound are needed, and as yet we do not have inexpensive sources.

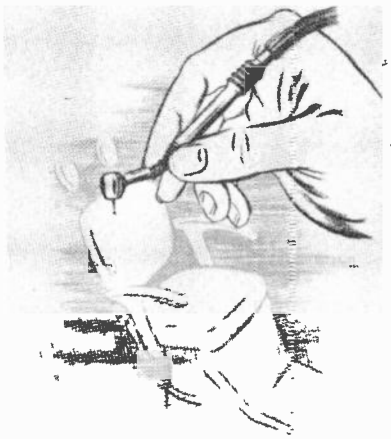
Uses with Metals

In metals technology, the use of intense sonic energy during the solidification of a metal from its melt will prevent the coarse, tree-like structures called dendrites. The final metal grains are small, and more of the resulting casting is sound. A Czech firm is rumored to have a sonic casting process in operation; it should be particularly useful in casting turbine buckets and other high-cost, high-performance parts. Isolation of the sound source from the heat of the melt is perhaps the largest problem. However, there is really no research uncertainty, and only straightforward engineering is needed to make the process practical for small, expensive parts.

Action of sound waves is greatest at the corners of a solid object when it is immersed in a liquid. This fact is being used in sonic deburring of cast or machined parts. The parts are immersed in a stirred abrasive slurry supplied with cavitating sound. Edges are speedily removed, with minimum change of critical dimensions. There should be all sorts of applications of this idea, which at present is restricted to small expensive parts because of cost.

Sonic impact grinding involves the vibratory motion of a tool to drive abrasive diamond chips against a hard, brittle material. Thus, die cavities can be sunk in very hard metals, or glass can be cut in intricate designs. Use of sonic impact grinding is now firmly established, and efforts have been made to apply it to dental drills. The misnomer, "drilling with sound," has led to some disappointment, for the drilling is done by the abrasive, which must be continuously fed in and removed from the mouth. Also, there is an unresolved controversy as to whether or not the repeated impacts will damage tooth pulp and nerve tissue. The most that can be said is that good, objective, quantitative evidence for one view or the other is still needed.

In forming metal in a die, considerable wear of the die walls occurs. It has been proposed to vibrate the walls of the die so that the metal will flow more easily, as does flour in a vibrating chute. The idea has not received a fully definitive test, but it seems likely that some useful effects should be obtained.



Sound has recently been tried as an aid in abrasive drilling in dentistry. Results are inconclusive without further objective study.

Textiles, Food, and Bugs

In textile production, one of the most fabulous fibers is the product of the oriental plant, ramie. Its stout fiber produces such long wearing cloth that even the Man in the White Suit would raise his eyebrows. Socks lasting 20 years have been claimed, though the proof is rather tenuous. Production of the ramie fibers is hampered by their affinity for the other portions of the plant. Ordinarily, retting (exposing the plant to the action of the elements) is used to separate fibers from the other plant material. However, the treatment weakens the fiber. It would seem that cavitation attack, in combination with chemical treatment, might hasten the process without undue damage to the fiber. At present there is no economic necessity for this material, so the idea has had no full and careful investigation.

Cavitating sound increases the flow of liquids through semi-permeable membranes. Thus, intense sonic energy might well decrease dramatically the time required for tanning leather. Again, the optimum tanning substance would not necessarily be the same as those used in the conventional process.

In food technology and pharmaceuticals, sonic processing has already been used on a small scale. Below the cavitation threshold, sonic agitation has increased the rate of bacteria growth by a factor of 50. However, above the threshold, the cell walls are ruptured, killing the microorganisms (or human cells) in the treatment zone.

It has already been mentioned that cavitating sound may inactivate enzymes. This could be of importance in problems involving the discoloration of foods by enzymic action.

Zoological acoustics indicates that many insects are quite sensitive to sounds in particular frequency ranges, some ultrasonic. Thus, if we know the sensitive range, it should be possible to create an exodus of small, crawling insects from foods. Then you may eat your brussels sprouts without fear of ingesting unwelcome protein. Or weevils might be chased from stored grain. Of course, the primary knowledge must come from insect studies; promiscuous use of sound on foodstuffs just to see what will happen would be a waste of money.

Source Needed

In all the foregoing processing applications of sonic energy, a large obstacle to their full exploitation is the absence of a suitable, economical sonic source. Such a source should yield high power with good efficiency, should have low installed and operating cost, should be able to be serviced by a plant plumber and electrician, and should be suitable for flow processing. No present sonic generator satisfies all these requirements.

The need for a cheap high-power device has prompted an examination of the difficulty of producing cavitation as a function of frequency. In water, below a frequency of 10,000 cps, a relatively constant amount of energy suffices, whereas above this frequency the energy requirement increases sharply. Working below 10,000 cps loses the magic appeal of ultrasonics, but gains the use of rotating machinery to produce the energy. Compact, 30 kw 10,000 cps motor-generators are readily available. Such a power source should be more attractive than a vacuum tube supply.

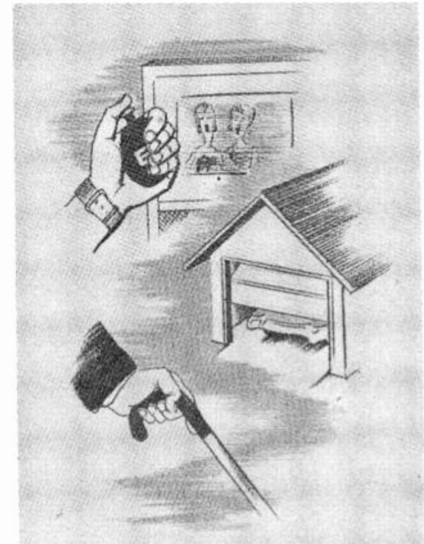
Another possible improvement in means of producing intense sonic energy is the replacement of magnetostrictive and ferro-electric (ceramic) devices by purely mechanical ones. A possible high-power source now being developed employs the same mechanism as a water valve that screams; it uses a motor, a pump, and a special valve. Called a hydrodynamic oscillator, it may well make possible a host of applications that await such a simple source.

For operation at frequencies below 1000 cps, unbalanced rotating mass force generators are available. Some are capable of acoustic outputs in excess of 100 hp when properly coupled to the liquid. Finally, in a fluid modulator, the steady stream of liquid from a pump can be changed into a pulsatory flow; this energy can readily be piped to the point of use. All these ideas are being investigated now, and should one turn out to be the answer, sonic processing should become a standard tool for industry.

Sonic Control

The use of sound waves for control involves an instrument's sensing a change and acting on it. Here we shall briefly mention a few. A popular TV remote control uses struck bars to produce ultrasonic control signals. A garage door opener employs an ultrasonic whistle operated from the car's intake manifold. Guide canes for the blind use reflections of short-wavelength sound to locate obstacles such as curbs. And an inaudible burglar alarm uses criss-crossing sound waves that detect any motion in their field. This device also detects flames and mice; the latter sensitivity had to be thwarted by a "mouse recognition" circuit. Finally, a system of sonar beacons has been suggested to replace lighthouses and foghorns. All of these control uses of sound are relatively straightforward, and need more ingenuity than research.

A phenomenon noted in all branches of industrial acoustics has been the reduction in time lag between obtaining fundamental results and applying them to the solution of industrial problems. Because fundamental research furnishes our scientific capital, we must take care that we keep it working—with a sufficient amount of its profits returned for generating more capital.



TV remote control devices, automatic garage doors, canes for the blind, are some of the many new ways to use sonic energy.

Syllable Analyzer, Coder and Synthesizer for the Transmission of Speech*

H. F. OLSON†, FELLOW, IRE, AND H. BELAR†

Summary—The analysis of speech by machine, the transmission of the coded message and the synthesis of speech from the code offer many advantages. For instance, if it is desired to transmit only "what" is spoken and not "how" it was spoken, very great savings in channel capacity to transmit information can be made. The digital form of the message permits more secure coding and decoding for secrecy. Work done at the RCA Laboratories on phonetic typewriters, machines that type in response to words spoken into a microphone, has shown that the transmission of speech analyzer output and synthesis of speech therefrom is feasible when performed by units of speech of the order of a syllable. Smaller phonetic entities, phonemes or speech sounds, cannot always be recognized by machine or even by ear without the knowledge of adjoining sounds. This information is always available when the analysis proceeds by syllables. Feasibility studies have also shown that intelligible speech can be synthesized from prerecorded syllables.

Introduction

Of THE different means of communication between individuals each endowed with a brain, eyes, hand, ears and a mouth, speech is the most important method of transmitting information. A graphic presentation[1] of a number of different systems is shown in Fig. 1. In the over-all plan of Fig. 1 there are many systems for the transmission of information that have not yet been developed or commercialized. Block diagrams of the undeveloped systems are shown in Fig. 2. This paper will present in more detail the category of systems which involves the analysis of speech, the transmission of a code and the synthesis of speech from the code as depicted by item five of Fig. 2. Such a system may be placed in the classification of digitized transmission of speech. An example, often cited, is the transmission of the printed version of the spoken language. The saving in channel capacity has been pointed out by Shannon.[2] The reason for the saving in channel capacity is due to the elimination of redundancy in speech and the simple fact that less information is transmitted. In the printed speech all personal inflections are removed and the information concerning how a word was spoken or by whom is omitted. Only the word itself is defined. There are many communication problems where the transmission of the equivalent of printed speech is sufficient, but where it is nevertheless desired that the sender speak into the transmitter and that the individual at the receiver hear the message again in the form of speech. To bring such a service and other new services to realization requires automatic machines to analyze speech and means to synthesize speech from a digital code. These considerations prompted the work done at the RCA Laboratories on phonetic typewriters[3,4] and electronic sound synthesizers.[5,6]

The Syllabic Approach to Speech Recognition

The classification of the English language into words, syllables and sounds, the latter represented by phonetic symbols, has no exact acoustical counterparts. Certain characteristics of the envelopes of speech can be associated with subdivisions of speech that are of the order of syllables, but the nearest to a phonetic entity or phoneme would be a spectrum in one time step. This, however, is true only insomuch that both represent the smallest temporal unit of analysis from the respective viewpoints. Actually, a phoneme or sound may require a time series of acoustic spectra to describe it. Moreover, the spectra for a given sound are found to depend upon what sounds precede or follow. But there are sounds that can be recognized out of context. This fact and the knowledge that English can be represented by less than 50 sounds has led many to work in that direction. The syllabic approach to speech recognition, however, has made better progress. By using speech segments of the order of a syllable, the analyzer presents the sounds of speech in their proper setting and portrays more realistically the transient nature of speech. The phonetic typewriters[3] developed at the RCA Laboratories were based on the syllabic approach from the outset. A block diagram of the first model built in 1954 to test the principle is shown in Fig. 3. In this model, the spectral analysis of speech is quantized as to frequency and amplitude and is transferred to the spectral memory in discrete steps. The spectral memory holds a time sequence of spectra and is not read out un-

* Received by the PGA, July 21, 1961. This paper was presented with a demonstration of speech synthesized from syllables at the National Aerospace Electronics Conference, Dayton, Ohio, May 9, 1961.

† RCA Laboratories, Princeton, N. J.

[1] H. F. Olson, "Acoustical Engineering," D. Van Nostrand Co., Inc., Princeton, N. J.; 1957.

[2] C. E. Shannon, "Developments in communications theory," *Electronics*, vol. 23, pp. 80–83; April, 1950.

[3] H. F. Olson and H. Belar, "Phonetic typewriter," *J. Acoust. Soc. Am.*, vol. 28, pp. 1072–1081; November, 1956.

[4] H. F. Olson and H. Belar, "Time compensation for speed of talking in speech recognition machines," IRE Trans. on Audio, vol. AU-8, pp. 87–90; May-June, 1960.

[5] H. F. Olson, H. Belar and J. Timmens, "Electronic music synthesis," *J. Acoust. Soc. Am.*, vol. 32, pp. 311–319; March, 1960.

[6] H. F. Olson and H. Belar, "Electronic music synthesizer," *J. Acoust. Soc. Am.*, vol. 27, pp. 595–612; May, 1955.

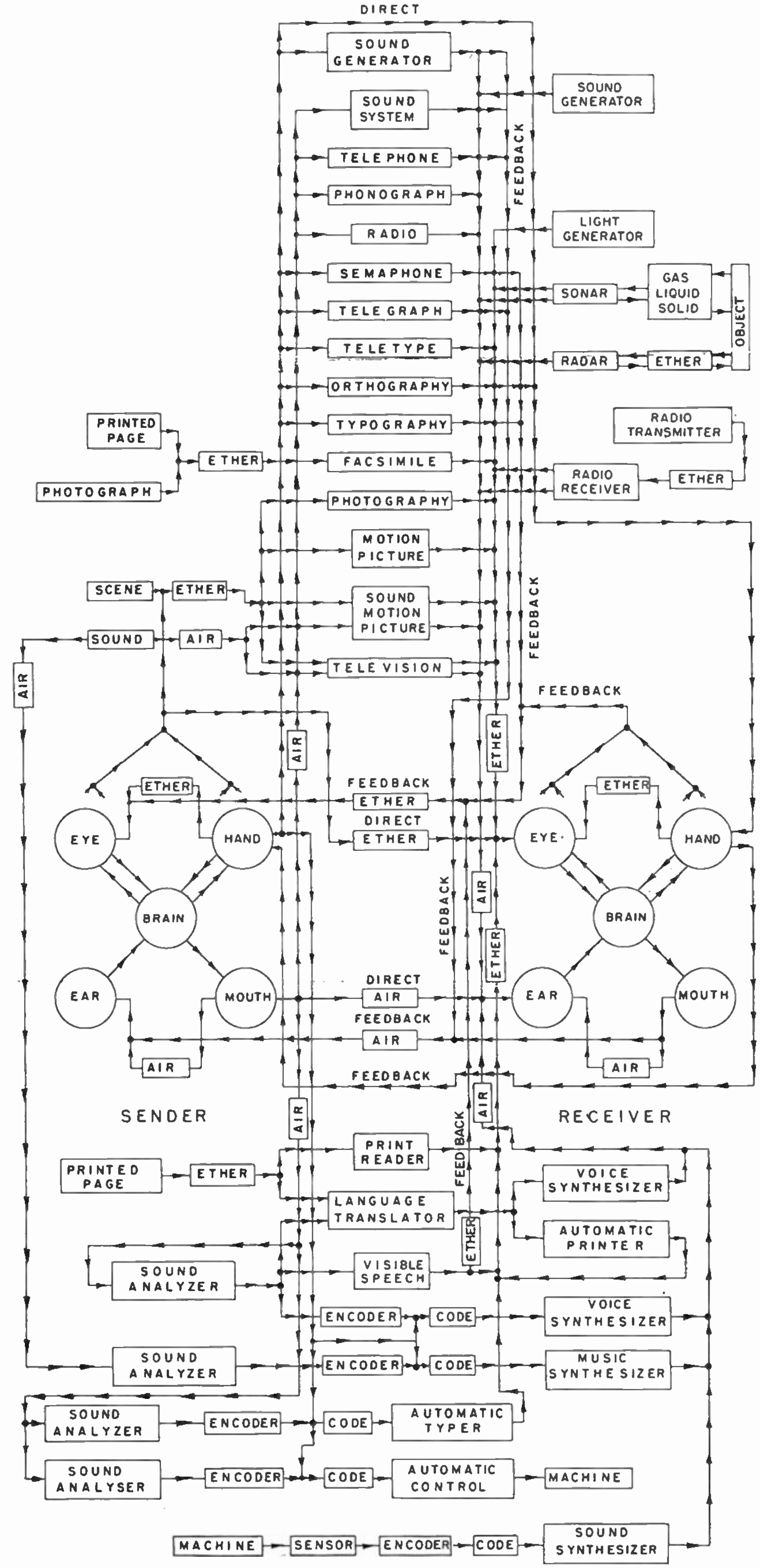


FIG. 1—Systems for the communication of information between two individuals.

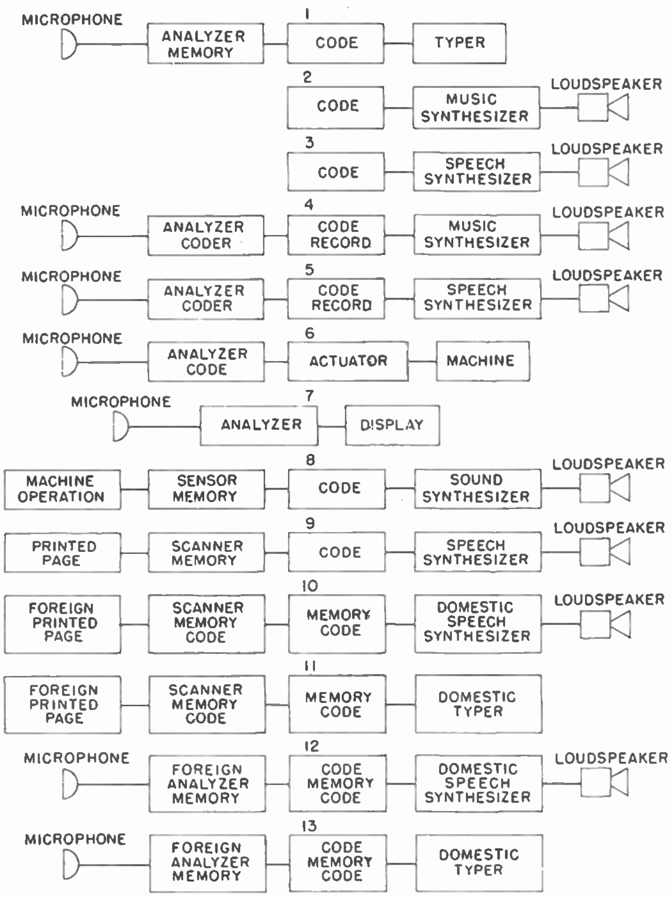


Fig. 2—Systems for the communication of information that have not been developed or commercialized.

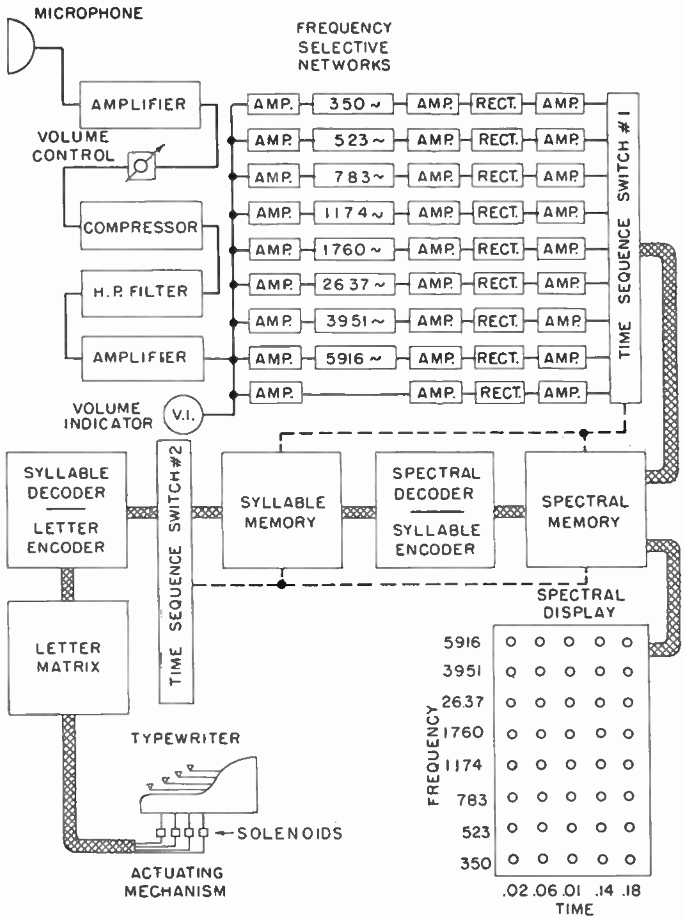


Fig. 3—Block diagram of phonetic typewriter model I.

til all the time steps making up a syllable have been completed, or in other words, until the setup of the spectral memory is completed. Model I phonetic typewriter has eight frequency channels, five time steps, and the amplitude was presented by a single binary digit. If the time steps are depicted as columns and the frequency channels as rows in a 5 by 8 matrix, then the spectral memory information can be visualized as a pattern. By the same reasoning process, the syllable memory can then be called a pattern recognizer. But there is actually no physical pattern involved. The forty fields in the display of the spectral memory represent, also, a binary number of that many digits. The different displays resulting from different voicings of the same syllable are therefore a set of numbers. The syllable memory consists of switches and wires that make contact to a respective syllable bus when one of the numbers of the set associated with that syllable is fed to the spectral memory input.

There are other advantages to the syllabic approach. Some have to do with methods of normalizing the speech signal before it is processed by the analyzer. This is done to reduce variation due to the manner of speaking. For instance, specially designed audio compressors and limiters are used to reduce the effect of changes in the loudness of speaking. Good results were also obtained with a system developed to compensate for the time rate of speaking.[4] In the system which has been incorporated in the later model phonetic typewriters, the spectral information is not transferred to the spectral memory in predetermined fixed time steps, but instead in steps that are determined by significant changes in spectrum. Thus, a slow or more rapid voicing of the same syllable results in more nearly the same input to the spectral memory. The improvement is particularly striking when the duration of the component sounds are differently apportioned, like "ssssee" as against "seeee." With a time compensated spectral memory, those two different voicings are recognized as more closely the same.

Other improvements were made in the detector circuits that perform the quantizing functions. In earlier models the amplitude was detected against a fixed or variable threshold. In phonetic typewriter Model III, comparison networks are employed that compare the amplitude of the signal in each channel with that in the two adjacent ones. An indication is obtained only if the amplitude in the respective channel is larger than the mean of the two adjacent ones. In this manner, results are obtained by a means somewhat approaching the

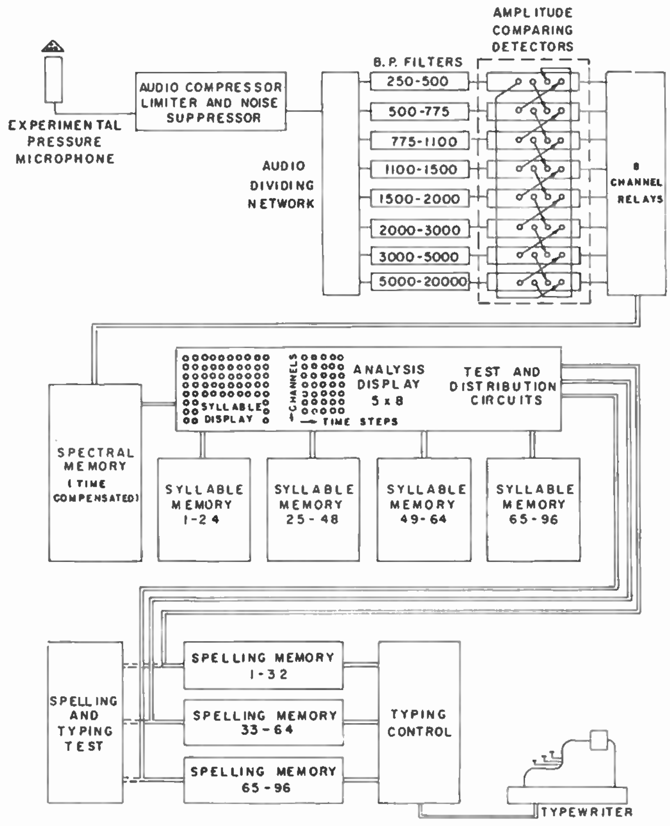


Fig. 4—Block diagram of phonetic typewriter model II.

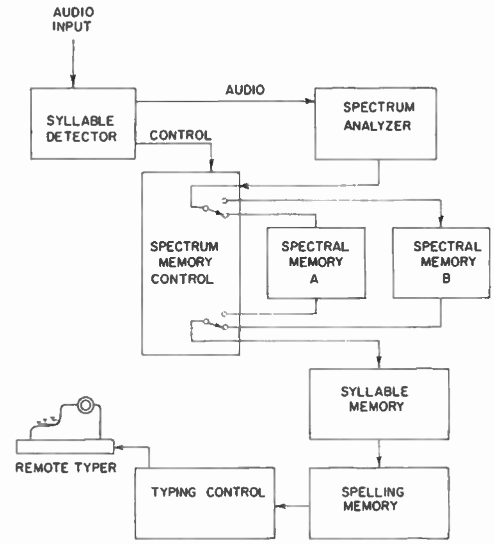


Fig. 5—Continuous speech analyzing system schematic.

process of obtaining the second derivative by finite differences. The bands of maximum energy are between points of inflection. This method is independent of the over-all levels. Minor peaks in the spectral distribution are also indicated more clearly.

A block diagram of the phonetic typewriter Model III[7] is shown in Fig. 4. The equipment consists of a microphone, audio amplifier and compressor-limiter, eight sets of high-pass low-pass filters, amplitude comparing detectors, a time-compensated means for transferring spectrum information into a spectral memory and associated display, a syllable memory that connects the proper syllable bus when one of a set of numbers associated with that syllable is fed into the syllable memory input, a spelling memory that supplies a predetermined spelling information for the syllable number called for, and a typing control unit which reads out the spelling information and operates the electric typewriter. The machine, as described, receives voice input spoken one syllable at a time. To function with continuously spoken speech requires the additional spectral memory as is shown in Fig. 5. The operation will then be as follows: continuously spoken speech is divided into segments of the order of syllables and the spectral analyses of alternate syllables are fed to alternate spectral memories. Thus, while one spectral memory is set up, the other is read out and vice versa.

Syllable Communications System

The use of a machine like the phonetic typewriter described in the preceding paragraph is a logical step toward realizing the "bandwidth" savings that should accrue from the transducing of the information contained in the spoken to that in the printed language. But it is not necessary to convert syllable information into spelling information before transmission; on the contrary, added advantages are realized by transmitting just the syllable number. This number is identified in the output of the syllable memory. The reasons for the added advantages are explained as follows. According to the findings of Dewey[8] 98 per cent of the syllables making up the English language can be represented by 2000 different syllables. Dewey based his findings on a study of 100,000 words found in newspapers, magazine articles, texts, famous speeches, the Bible, etc. It is also well known, as shown by Fletcher,[9] that a syllable articulation index of 98 per cent results in a sentence intelligibility of over 99 per cent. Thus, it is very likely that 2000 syllables would suffice for intelligible communication.

Now, 2000 different syllables can be transmitted by eleven binary digits per syllable, whereas to transmit the corresponding printed information, assuming five letters of 26 possible different characters each, requires more than twice that many binary digits. Thus, a syllable communications system offers very great savings in the capacity required to transmit the information.

Another advantage of the system is security. For the

[7] H. F. Olson and H. Belar, "Phonetic typewriter III," *J. Acoust. Soc. Am.*, vol. 33, pp. 1610–1615, November, 1961.

[8] G. Dewey, "Relative Frequency of English Speech Sounds," Harvard University Press, Cambridge, Mass.; 1928.

[9] H. Fletcher, "Speech and Hearing in Communication," D. Van Nostrand Co., Inc., Princeton, N. J.; year.

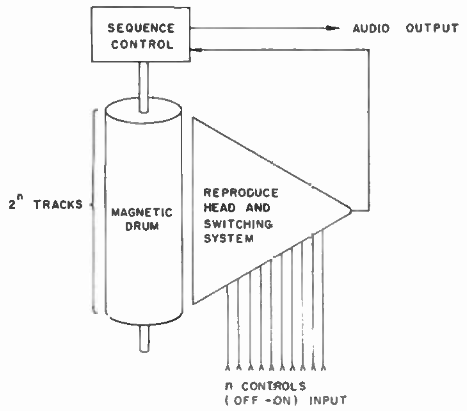


Fig. 6—Binary code syllable speech synthesizer.

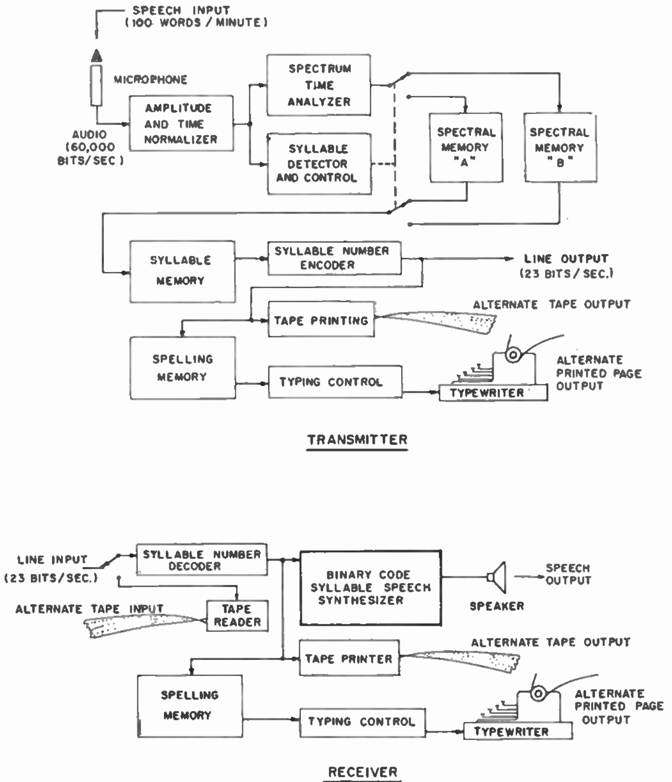


Fig. 7—Syllable communications system.

same amount of sophistication of encoding and decoding, the probability for breaking a code based on 2000 syllables as against 26 letters is in the ratio of

$$\frac{26!}{2000!}$$

which is as one is to several thousand powers of ten. The advantage of security becomes even more apparent when compared to scrambled speech transmission.

Synthesis of Speech for a Syllable Communications System

In the foregoing the advantages of a syllable communications system have been established. In addition, a syllable transmitting system has been described. The receiving system will now be described. Again, one of the logical alternatives would be to employ the spelling memory, etc., of the phonetic typewriter to print out the message. The person at the receiving end could then read the message spoken into the transmitter. This would, no doubt, serve many purposes but there is still a need for voice-to-voice communication, so that the individual at the receiver can hear the message spoken by the individual at the transmitter. For example, if speech can be segmented into particles of the order of syllables, analyzed in such steps and an intelligible message derived therefrom, then it should also be possible to synthesize speech from corresponding particles of speech and understand it. This was indeed found to be the case. Tests were made by recording clearly enunciated syllables in alphabetic order. The recordings were then rearranged to form messages which proved highly intelligible. Using prerecorded spoken syllables has the advantage for this application, namely, that the speaker and the recordings can be selected for highest intelligibility, which depends upon and includes many complex factors. Moreover, the use of prerecorded syllables reduces the complexity of the receiving equipment. It is foreseen that the synthesis of speech from prerecorded syllables may employ a drum with as many sound tracks as there are syllables to be transmitted and that the reception of a syllable number would open the corresponding gate in order to reproduce that syllable. Thus, syllables would be reproduced in the order in which they are spoken, analyzed and transmitted. A schematic diagram of such a reproducer is shown in Fig. 6.

Those responsible for working out systems will find the concept of transmitting information by syllable number adaptable to other ends. Tape perforators and tape readers may be included to store information of syllable numbers for reading later or transmitting at some other speed. Spelling memories, etc., could be used with typewriters to record messages sent or received and so on. Such an arrangement of equipment is shown in Fig. 7, which has been taken from an RCA Laboratory proposal[10] for research on a speech synthesizer for this application. Considerable progress has been made in the field of analysis and recognition of speech by machine. There is still more work to be done.

The Flow of Information in a Syllable Communications System

Beginning with the first letter of the alphabet, which is also the indefinite article, a, and therefore a word, let it be assumed that it is spoken into a microphone. The signal output of the microphone will then be a function of amplitude versus time as shown by the trace depicted

[10] RCA Laboratory Descriptive Specification PP60–1089, "Speech Synthesizer for a Syllable Communications System," RCA, Princeton. N. J.

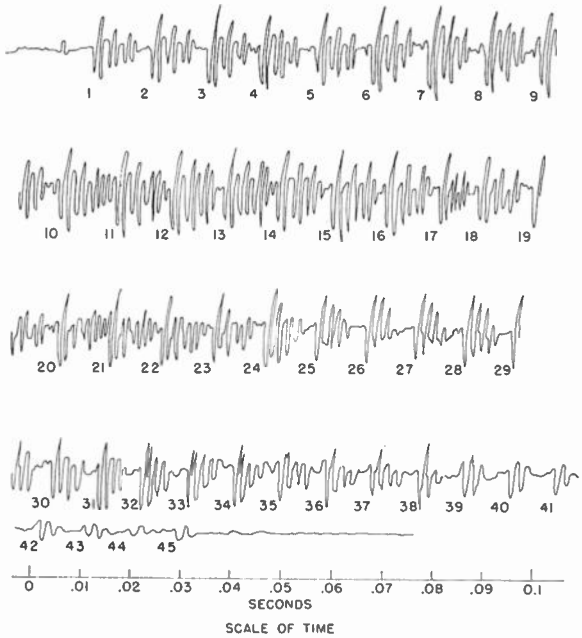


Fig. 8—Trace of speech wave of "a" spoken by H.B.

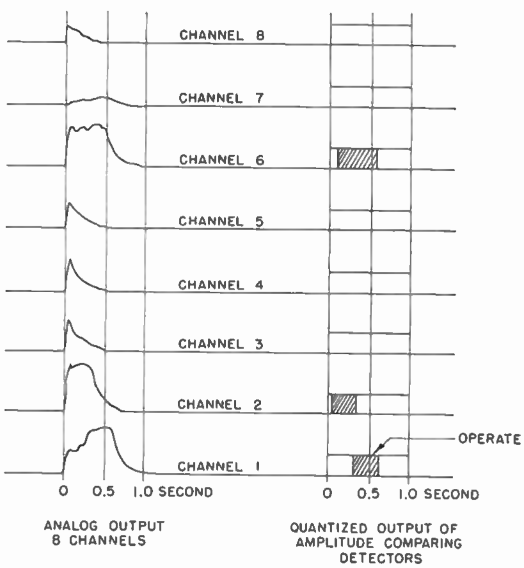


Fig. 9—Signal amplitudes and relay operations spectral analyzer, model III, voicing "a" by H.B.

in Fig. 8. When this signal is fed to the spectral analyzer, where it is transformed into a quantized frequency-amplitude-time function, the rectified output in the various frequency bands will be found to vary as shown in Fig. 9. In order to explain the action of the amplitude-comparing detector circuits, the signals existing in eight channels before and after amplitude comparison are shown for one instance of time in Fig. 10. The advantage of the amplitude comparing system can be seen on Fig. 10 by the sharper definition of the peaks. The quantized output of the spectral analyzer and the information fed into the spectral memory, in steps determined by changes in the quantized spectrum, are both shown in Fig. 11. Figs. 9 to 11, inclusive, portray

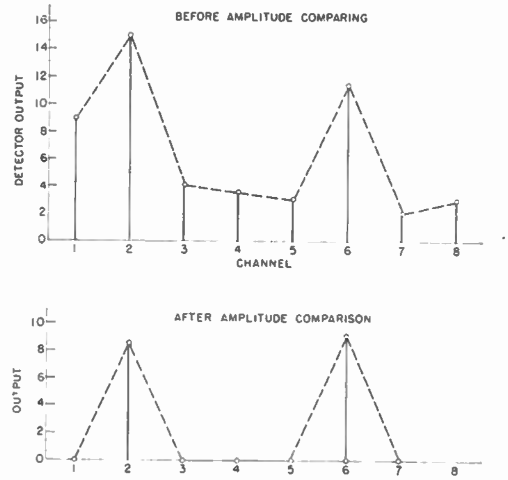


Fig. 10—Spectral analyzer model III. Signal levels in detector circuits, 0.2 sec after start of voicing of "A" by H.B.

the acoustically derived information from a unique but typical voicing of "a." If the same speaker spoke it again, he may, or may not, again produce the same results. Speech is subject to variations, not all of which are under the control of the speaker. The effects of these variations are reduced by the normalizing processes but some still remain. The number of variations in the analyzer output depends on the number of steps in which the information is quantized. In general, the greater the number of steps, the greater is the number of variations. They also depend on the speaker, on the language and on the word or syllable. For example, 100 voicings of "a" by H.B. result in 14 different displays, as depicted in Fig. 12 in the order of their frequency of occurrence. The most frequent display occurring in 70 out of 100 voicings is also the same as the typical voicing described in Figs. 9 to 11. A study of the 14 different displays on Fig. 12 shows that there are some features in common to all 14. For instance, the field second row from the bottom in column one is always operated. The same is true for row six from the bottom where either column one or two, or both, are always operated. On the other hand, there are four other fields that may be either operated or not operated. They are denoted by X's on Fig. 13. All other fields must not be operated. A wiring schematic that serves to recognize the set of numbers resulting in the above mentioned displays for the 100 voicings of "a" is shown on Fig. 13. As stated before, the output of the syllable memory consists of a connection to a bus associated with the syllable that was recognized. The conversion of information contained in a singular connection to one of many buses, to information encoded into a signal for a single channel can be accomplished by known means and will not be elaborated here. At the receiving end, after suitable decoding, the information would again be an audio signal with no inherent restrictions as to fidelity other than practical considerations. The seeming paradox, that the transmission of information at the rate of say 25 binary

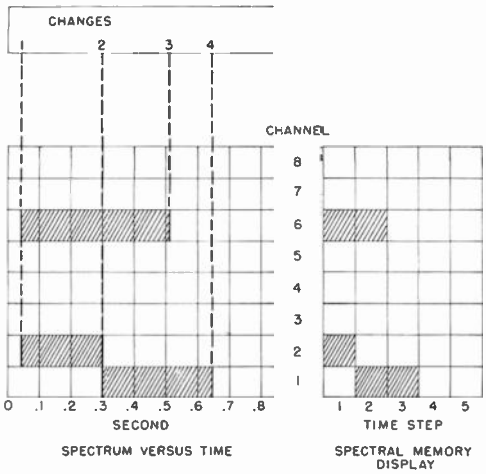


Fig. 11—Derivation of a time compensated memory display. Voicing of "a" by H.B.

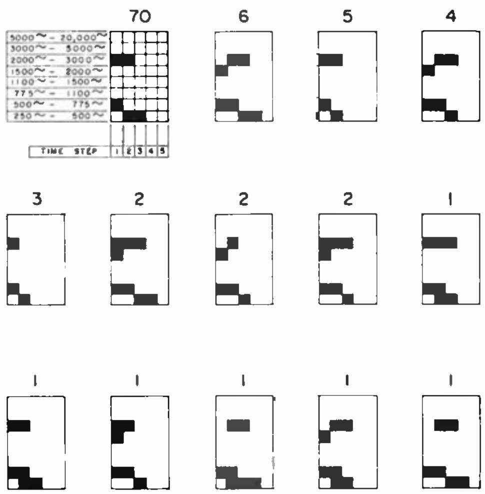


Fig. 12—Fourteen different displays obtained in 100 voicings of "a" as in RCA, H.B. voice.

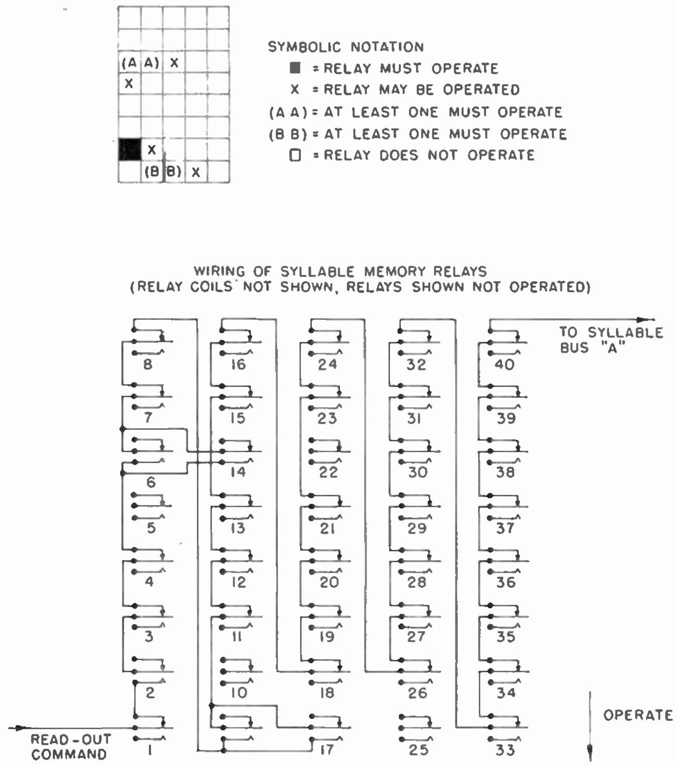


Fig. 13—Syllable memory wiring to recognize "a" spoken by H.B. Phonetic typewriter model III.

1. come	13. on
2. course	14. Pete
3. dee	15. po
4. get	16. part
5. I	17. pro
6. in	18. rea
7. Kate	19. see
8. Kay	20. seed
9. knee	21. shun
10. low	22. tar
11. mu	23. Ted
12. of	24. your

Fig. 14—Prerecorded spoken syllables in alphabetic order.

digits per second may result in an audio output that requires a channel with a capacity of over 100,000 binary digits per second, is of course resolved when it is considered that the transmitting system is required to trigger only 2000 different messages. The explanation given above for the circuits that perform the recognition has been greatly simplified. Many other factors are included in the course of work to design these circuits to perform at a predicted accuracy, etc.

Demonstration of Synthesized Speech from Prerecorded Syllables

It is not practical to demonstrate the automatic analysis of speech by machine such as by the phonetic typewriter Model III other than at its location for the present. However, it is possible to show that intelligible speech can be synthesized from prerecorded syllables. Syllables were recorded on a tape, one syllable at a time, in alphabetic order as listed in Fig. 14. The syllables were then rearranged to make sentences, which indicated in a demonstration that the sentence intelligibility was the same as for conventional reproduced speech.

Phasor Analysis of Some Stereophonic Phenomena*

BENJAMIN B. BAUER†, FELLOW, IRE

Summary—An improved understanding of some stereophonic phenomena may be obtained by use of acoustical pressure phasors to portray sound pressure at the ears of the observer. With the help of phasors, it is possible to expand and modify certain conclusions of previous observers and to validate some previously unpublished observations: a stereophonic "law of sines" is derived. The existence and location of the "out-of-bounds" stereophonic image is analyzed and verified. The "allowed maximum out-of-phase ratio" is derived, together with the observation that this maximum is exceeded by certain microphone arrays. The motion and elevation of the center image in stereophonic reproduction is observed and explained.

Introduction

THE directional perception and localization of sound is a complicated phenomenon, which depends upon the psychological state of the observer (*e.g.*, "Cocktail Party Effect"), and upon visual as well as aural cues. With respect to the aural effects, it is well known that interaural delays, as well as intensity differences for steady-state and transient sounds, play an important role in establishing directional perception.[1] Within limits these functions may be portrayed in terms of magnitude and phase angle of sinusoidal waves and, thus, may be readily represented by phasor ("vector") notation. The use of phasors has been found to be a convenient tool for the analysis of stereophonic phenomena.[2] The method of phasor analysis described in this paper has served to predict and explain certain stereophonic effects and to modify some of the conclusions of previous investigators.

Single Sound Source

Consider a source of sound S in Fig. 1 emitting a sinusoidal wave at some frequency f below, say, 1000 cps and moving in a semicircle around the observer from M to N. As a first approximation, assuming an interaural distance D, the effective interaural delay for sounds arriving from an angle θ is $(D/c)\sin\theta$ seconds; and, therefore, the phase angle between the sound pressure phasors at the left and right ears is given by

$$\phi_\theta = \omega(D/c)\sin\theta \text{ radians} \quad (1)$$

where ω is $2\pi f$, and c is the speed of sound.

The portrayal of the sound-pressure phasors corresponding to different positions of the source is shown at the bottom of Fig. 1. Starting with the extreme left position at M, ϕ_{90} is the maximum phase angle that is attained at the particular frequency, with the left-ear phasor L leading the right-ear phasor R by the angle $\omega D/c$. As the source moves clockwise, the phase angle diminishes with the $\sin\theta$, and in the center position C, L and R are equal and in phase. Evidently the mechanism of hearing has learned to accept the resulting time delay and relative pressure relationships as signifying "the sound is arriving from the particular direction θ." As the source continues to move toward N, the phasor diagram repeats its previous geometry; however, this time the phasor R leads the phasor L.

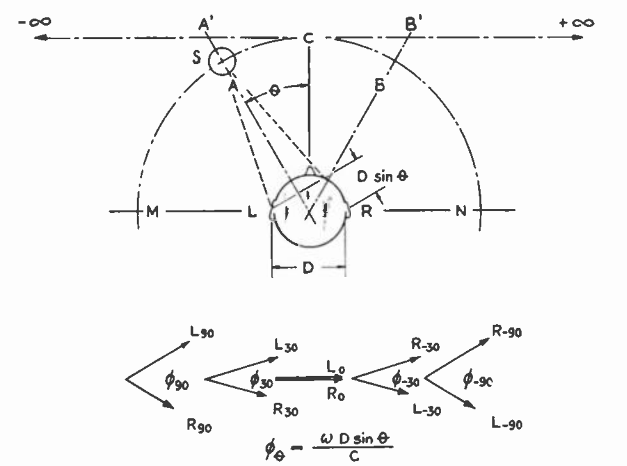


Fig. 1—Pressure phasors for single sound source.

Except for the symmetrical case $(\theta = 0)$, there is a relative increase in magnitude of sound pressure at the ear nearest to the source, and diminution at the ear farthest from the source.[3] These differences stem from the diffraction of sound around the head of the observer and from the inverse pressure vs distance law. Below about 200 cps the diffraction is negligible, and the localization appears to stem mainly from the temporal phenomena. Between about 200 and roughly 1200 cps, diffraction effects become progressively more important, and they may be assumed to add to the localization ability. These are the frequency regions in which phasor analysis of stereophonic phenomena appears to achieve its best success. The effect of pressure differences cannot be neglected, as otherwise in some instances, incorrect conclusions may result.

It has been assumed previously that the source of sound moves around a semicircle. If the sound were

* Received by the PGA, July 20, 1961. Presented at the Acoustical Society of America, Philadelphia, Pa., May 12, 1961. This article appeared in the *Journal of the Acoustical Society of America;* November, 1961.

† CBS Laboratories, Stamford, Conn.

[1] For an excellent summary see B. H. Deatherage and I. J. Hirsh, *J. Acoust. Soc. Am.*, vol. 31, pp. 486–492; 1959.

[2] T. T. Sandel, D. C. Teas, W. E. Feddersen, and L. A. Jeffress, *J. Acoust. Soc. Am.*, vol. 27, p. 842; 1955.

[3] F. M. Weiner, *J. Acoust. Soc. Am.*, vol. 19, pp. 143–146; 1947.

to move in a straight line, from $-\infty$ to $+\infty$, similar phasors could be drawn for the corresponding angular positions, but the absolute values of the pressures would vary inversely with the distance from the source. This variation of relative intensity, together with the time delay and perhaps Doppler effect, would appear to be one way in which a blindfolded observer could tell the difference between a source which moves in a straight line and one moving around him in a circle.

Two cautions are needed here: 1) The phase angle is viewed merely as a convenient mathematical device for describing time delay, and not as a cause of directional perception *per se*, and 2) The absolute magnitude of the phase angle is proportional to the frequency. Therefore, to be valid at all the frequencies in the range of interest, whatever conclusions are drawn should stem from analysis which holds for a wide range of the frequency-dependent variable $(\omega D/c)$.

Two Sound Sources

Having established the use of pressure phasors for a single source, it follows that the same impression of direction and relative position of the source will be manifest no matter how these phasors are obtained. It is known that for an observer placed centrally between two stationary loudspeakers, it is possible to produce the illusion of a moving source by varying the intensity of in-phase signals supplied to the loudspeakers.[4] In phasor form this is demonstrated in Fig. 2. The observer is located symmetrically between the two loudspeakers r and l. Let the loudspeaker l be energized: The phasor diagram at A is identical to that previously obtained for a moving source; the phase angle between R_l and L_l is ϕ_A, and the direction of arrival perceived by the observer will stem from l. With two equal and in-phase signals applied to the loudspeakers, the phasor diagram is as shown in B. The two sets of phasors R_r and L_r and L_l and R_l, from the right and the left loudspeakers, are congruent, and when the corresponding phasors are added, the resultant total pressures R and L are equal and in phase; this explains the manifestation of a central virtual source at B. Assuming that the left signal is smaller than the right signal, the phasor diagram at C applies; the resultant phasors R and L are displaced at an angle ϕ_I which defines a virtual image C at an angle θ_I intermediate between B and r. An analysis of the position of this image was made by Clark, Dutton and Vanderlyn,[5] assuming equality of absolute sound pressures at the two ears. With the aid of their

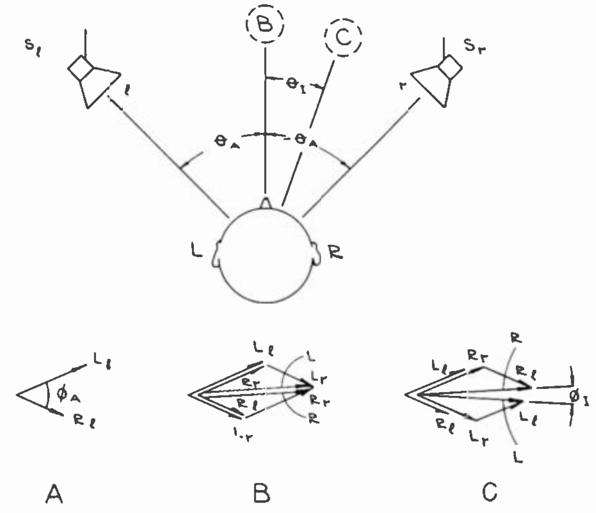


Fig. 2—Pressure phasors for two sound sources with inphase sound signals. A: $S_r=0$. B: $S_l=S_r$. C: $S_l<S_r$.

equation, (2), the following important relationship can be derived:

$$\sin\theta_I/\sin\theta_A=(S_l-S_r)/(S_l+S_r), \text{ approximately}, \quad (2)$$

where

θ_I is the azimuth angle of the virtual image, and
θ_A is the azimuth angle of the real sources.
S_l and S_r are the strengths of the signals applied to the left and right loudspeakers, respectively.

This we call the "stereophonic law of sines," and it shows that through appropriate distribution of in-phase signals to the loudspeakers, the position of the virtual image for the centrally placed observer may be adjusted anywhere relative to the loudspeakers.

Direct utilization of this principle occurs in the widely used present-day method of stereophonic recording with *proximate microphones*. Each group of performers is covered by one or more microphones, and the outputs of these microphones are recorded and distributed in proportions related to their appropriate placement between the two channels. Upon reproduction, the performer positions are portrayed as virtual sources between the loudspeakers [Fig. 3(a)].

A similar result was obtained by Blumlein[4] and subsequently by Clark, *et al.*,[5] through the use of two velocity microphones crossed at 90 degrees to each other and oriented toward the sound stage in such manner that the outputs transmitted to the left and the right channels were cos $(45°-\theta)$ and cos $(45°+\theta)$, respectively. This arrangement, which is shown in Fig. 3(b), has not gained as wide popularity in the concert hall or recording studio as has the proximate-microphone method, because of at least two factors: 1) To cover a sound stage or large group, the crossed microphones must be placed at a considerable distance from the group thus making it difficult to control the, signal/reverberation ratio; and, 2) A substantial amount of "out-of-phase" signal exists due to reverberant sounds arriving from angles considerably greater than 45 degrees, causing the recorded

[4] For example see A. D. Blumlein, Brit. Patent No. 394325.

[5] H. A. M. Clark, G. F. Dutton, P. B. Vanderlyn, "The 'Stereosonic' recording and reproducing system," IRE Trans. on Audio, vol. AU-5, pp. 96–111; July–August, 1957. Clark, *et al.*, derive the following equation: $\phi_2=[(L-R)/(L+R)/(\omega D/c)\sin\theta_A$, where L and R are average instantaneous pressures at both ears from the left and right loudspeakers and ϕ_2 is the resulting phase angle. By the further step of defining $\phi_2=(\omega D/c)\sin\theta_I$, where θ_I is the location of the virtual image and additionally remembering that $L=KS_l$ and $R=KS_r$, our equation (2) is obtained.

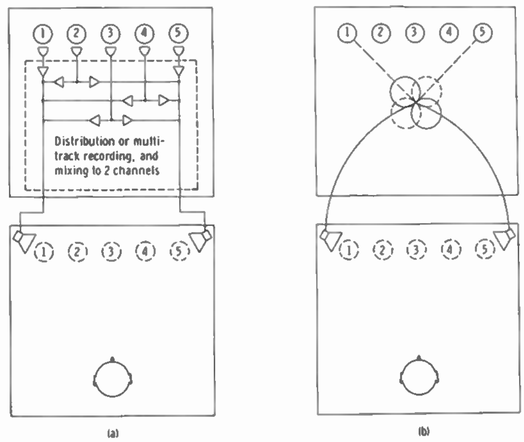


Fig. 3—Microphone arrays for stereophonic recording. (a) Proximate microphones with intensity distribution. (b) Remote crossed-microphones for intensity distribution.

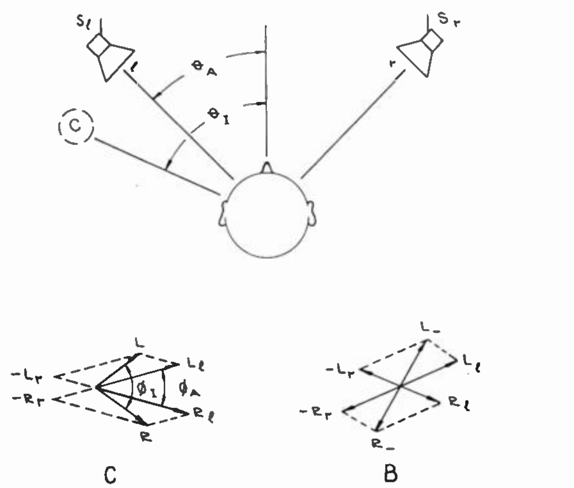


Fig. 4—Pressure phasors for two sound sources, with out-of-phase signals. B: Equal out-of-phase signals. C: Unequal out-of-phase signals.

reverberation to be reproduced unnaturally with a consequent unpleasant auditory sensation, for reasons later to be explained. Several other distant-microphone pickup systems pose similar problems to the recording engineer.

Out-of-Phase Signals

At this time we can examine the interesting question of what happens when the two similar signals are applied to the loudspeakers in an out-of-phase manner. Taking first the equal-signal case, shown in Fig. 4 at A, it is evident that the reversal of one signal, say the right, will reverse the direction of the phasors R_r and L_r; they become $-R_r$ and $-L_r$; the resulting total phasors at the ears will be R_- and L_-, which are equal and opposite identically for all values of the modulus $\omega D/c$. This is a most unnatural situation which has no counterpart in the normal hearing experience. First, there is a reduction of response at low frequency. Next, there is a loss of localization and, with some observers, a feeling of "pressure in the ears." One can only conclude that integrity of phase relations must be carefully maintained in stereophonic sound reproduction.

In the case of unequal out-of-phase signals, another interesting condition occurs. Referring to C, when the right signal is smaller than the left as well as reversed in phase, the phasors R_r and L_r become $-R_r$ and $-L_r$ and the resultant angle ϕ_I is greater than ϕ_A, predicting the existence of an image "out of bounds" of the loudspeakers. Virtual images out of bounds of the loudspeakers were noticed by Offenhauser and Israel.[6] Subsequently, this phenomenon was reported by Clark, *et al.*, as "not readily observable in practice,"[5] and the extent of it does not appear to have been previously studied. Examination of phasor diagrams in Figs. 1 and 4 suggested that it should be possible to achieve any desired phase relationship between the resultant sound pressures. Study of the phenomenon experimentally was facilitated by the setup in Fig. 5.

[6] W. H. Offenhauser, Jr., and J. J. Israel, *J. Soc. Motion Picture Engrs.*, vol. 43, pp. 139–155; February, 1939.

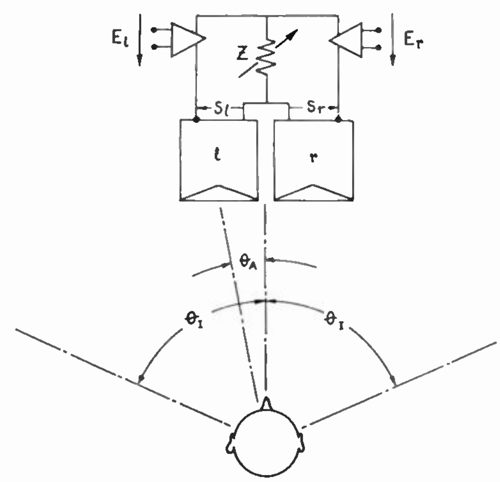


Fig. 5—Arrangement of apparatus for study of out-of-bounds stereophonic images.

Two carefully matched loudspeakers l and r are placed close to each other and are connected to two amplifiers so that with the rheostat Z set to 0, equal input program signals E_l and E_r will produce equal pressures at the observer location. Next, with $E_r = 0$ the rheostat Z is adjusted so that a part of the current through l will also circulate through r in a sense opposite to that considered "positive." The virtual image will then shift from θ_A to a θ_I depending upon the adjustment of Z. If E_l proceeds from instrumental music or voice, the performer can be steered to appear to arrive from any angle up to 90 degrees. If stereophonic signals are now applied to the amplifiers, the apparent field of sound arrival will become $\pm\theta_I$. This effect is critically determined by the accurate placement of the listener along the line of symmetry, and it is most strikingly observed in an anechoic chamber.

It is interesting to compare these results with those of Sandel, *et al.*,[2] who investigated the more limited case of equal out-of-phase signals stemming from dissymmetrically placed loudspeakers. Under these conditions Sandel and his collaborators found that the position of the virtual image was frequency dependent. In our tests, with the symmetrical loudspeaker placement and unequal signals, the position of the external virtual image appeared to be independent of frequency, and its position may be predicted by (2) which does not contain a frequency term.

At this juncture we can answer the question, "How much out-of-phase signal is allowable?" This can be answered by reference to (2). Let S_r become reversed, and hence equal to $-S_r$, substituting in (2)

$$\sin \theta_I / \sin \theta_A = (S_l + S_r)/(S_l - S_r). \quad (3)$$

Since $\sin \theta_I$ cannot exceed 1, θ_r cannot exceed 90 degrees whereupon (2) may be solved as follows:

$$S_r/S_l(1 - \sin \theta_A)/(1 + \sin \theta_A). \quad (4)$$

For instance, with the loudspeakers placed at 30 degrees azimuth angle, the allowed maximum negative ratio of signals is $(1-0.5)/(1+0.5) = 0.333$. Similarly with the loudspeakers at 45 degrees $-S_r/S_l \leqq 0.17$. With the crossed microphone method $-S_r/S_l$ may greatly exceed these allowable maxima, becoming unity for a 90 degree direction of the source.

Elevation of the Central Image

Anyone listening carefully to stereophonic reproduction of sound becomes aware of the fact that the central images appear to move somewhat with the turning of the head. As this is in contradiction to the assertion made by Clark, *et al.*,[5] this phenomenon was studied by means of phasor analysis.

If a sound source is located horizontally in front of the observer, and the latter turns his head about the vertical axis through a small angle θ_P (Fig. 6A), by (1), the phase angle $\phi_P = (\omega D/c)\theta_P$, and the rate of change of this phase angle with the motion of the head is

$$d\phi_P/d\theta_P = \omega(D/c). \quad (5)$$

The relative motion of the source is opposite to the head motion, and we have been accustomed to accepting this condition as signifying that the source is stationary, and located directly in front.

Consider next the stereophonic array in Fig. 6B with the loudspeakers at θ_A and $-\theta_A$ located horizontally in the plane of the head of the observer and receiving equal signals. Adding up the individual phasors due to each loudspeaker at each ear, it will be seen that for a small angle θ_P, $\phi_r = (\omega D/c)(\sin \theta_A + \theta_P \cos \theta_A)$, and $\phi_P \doteq (\omega D/c)(\sin \theta_A - \theta_P \cos \theta_A)$, therefore, $\phi_P \doteq (\phi_r/2) - (\phi_l/2) = (\omega D/C)\theta_P \cos \theta_A$, and the rate of change of the phase angle with the observer head motion is,

$$d\phi_P/d\theta_P = -\omega(D/c) \cos \theta_A; \quad (6)$$

Therefore, the opposing virtual image motion is not quite fast enough to compensate for the head motion and it can be expected to move with the head, at a rate $(1-\cos \theta_A)$. The cosine factor also occurs in the instance of single central source elevated at an angle θ_A.[7] Therefore, alternatively, the virtual image may be imagined to be elevated above the line connecting the loudspeakers. The actual apparent elevation visualized by most observers is about ½ of the predicted amount. However, when the observer is placed very close to the center line connecting the loudspeakers, $\cos \theta = 0$, and $d\phi_P/d\theta_P$ vanishes, which is also the situation experienced with a source elevated at 90 degrees. At this point the central virtual image tends to appear directly overhead.

[7] H. Wallach, *J. Exptl. Psychol.*, vol. 27, pp. 339–368; 1940.

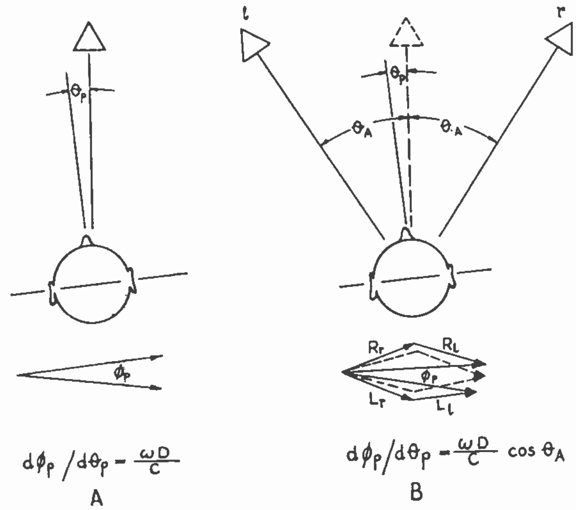


Fig. 6—Effect of head motions. A: Single sound source. B: Two sound sources.

Conclusion

Phasor representation of stereophonic phenomena has been found to be a useful tool for explanation and prediction of several aspects of stereophonic sound reproduction. By means of phasor analysis several conclusions of previous observers were expanded and modified, and certain heretofore unpublished observations were validated:

A "stereophonic law of sines" for virtual image location was stated; the pragmatic method of stereophonic recording with proximate microphones was validated; the existence and location of out-of-bounds stereophonic sound was predicted and verified; the "allowed maximum out-of-phase ratio" for stereophonic signals was derived; the existence of excessive out-of-phase ratio with the crossed-microphone method of stereophonic pickup was encountered; the motion and elevation of center image in stereophonic reproduction was observed and explained.

Acknowledgment

The author is grateful to Dr. P. C. Goldmark for his interest and encouragement of this study; to G. W. Sioles for many helpful discussions and A. Gust and W. Koskuba for arranging the illustrative experiments; and to Dr. I. Hirsh of the Central Institute for the Deaf and a CBS Laboratories Consultant for suggestions in connection with this paper and help with the location of some of the references.

The Hamograph

A New Amplitude Rhythm Control Device for the Production of Electronic Music*

MYRON SCHAEFFER†

Summary—**The Hamograph was designed and built to fulfill the need for a practical and relatively inexpensive means for controlling the rhythm and dynamic envelope of raw electronic sounds, as a tool for the electronic music composer. The machine consists of control loops of variable speed, which carry the rhythmic and envelope message in the form of conductive contours, read by a series of contacts of graduated resistance, which in turn modulate the signal voltage. Any type of sound may be subjected to such control which results in a wide variety of musical textures and melodies. Corrections can be made easily, and the number of splices required for a total electronic composition is materially reduced.**

THE history of man is the story of countless invasions. Some have affected his economic life, others his way of thinking and still others, his behavior and sense of values. In the arts new styles and forms have been the result of invasions within the same medium, that is, whatever changes have occurred, painting remained painting, the theatre was always live whatever the nature of the play, and music required musicians to perform it whether it was a simple plain chant melody or a complicated orchestration of Schoenberg. During the past hundred years the arts have experienced a new kind of invasion, the invasion by completely foreign elements which have changed creative techniques and crafts from one day to another. The first area invaded was painting in Paris in 1839 by Daguerre with his wet plate technique of photography. By the end of the century the legitimate theatre was invaded by the movies, and now music, the oldest and perhaps the most robust of all the arts, is undergoing a similar experience. The invader is electronics, in itself hardly expressive of aesthetic values. First electronics were used to make possible the successful competition of a solo voice with a full orchestra, then to allow a Bing Crosby to sing a duet with himself, and, within the past thirty years, electronics have invaded the sacred domain of the composer.

The composer once created exclusively in the medium of sound produced by traditional musical instruments or voices and was forced to rely on the intervention of one or more human performers in order to have his compositions heard. With the advent of the player piano and organ, the initial performer was required, but in subsequent uses of the perforated paper roll control, the performer was eliminated and the performance became automatic. In this case the sound source, the piano or organ, was still traditional. Yet, as early as 1929, at the Paris Exposition[1] there was a public demonstration of "synthetic music" made by electronic devices. A roll operated device consisting of four monophonic electronic oscillators worked very successfully. Other similar French patents[2] involving natural as well as synthetic sounds followed the basic patent covering device.

With the development of the magnetic-tape recorder, the modest beginnings of electronic music demonstrated in Paris made possible a completely new musical personality, the electronic music composer, whose technique of composition involves many electronic as well as new mechanical devices. His craft consists of using a wide range of sounds of traditional and new timbres obtained through a microphone or from electronic sound generators. He filters, molds and manipulates these sounds stored on magnetic tape and, finally, cuts, splices, and combines them in such a way that an organized musical composition results. The traditional musical score, which must first be notated and then realized through one or more performing musicians, is by-passed and the composer both creates and performs simultaneously. But the production of electronic music forces its composer to tackle certain new problems which require unusual and often difficult technical solutions. Not among the least of these are the synchronization of melodic units into a contrapuntal texture; the organization of sounds in extremely complex rhythmic continuities; the shaping of individual sounds in relation to their attack, continuum and decay patterns; and perhaps most difficult, the exact repetition of rhythmic patterns with varying tinbres and at the same time, increasing or decreasing the tempo without altering the pitch of the constituent sounds.

All of these problems can and have been solved with a greater or lesser degree of success, but their solutions almost always involve many hours of splicing, listening, splicing and finally trying all over again with even greater care. The processes employed require repeated transfers from one tape to another with the resulting increase of noise level, which, in some cases, may be

* Received by the PGA, November 13, 1961.
† Electronic Music Studio, University of Toronto, Ontario, Can.

[1] E. E. Coupleux and J. A. Givelet, French Patent; U. S. Patent No. 1,957,392, 1930.
[2] L. Lavalleen, French Patent No. 806,076; 1936.

filtered out, but often the result is a distortion of the composer's original intentions.

The solving of these problems and the difficulties involved in their solution prompted the study and construction of an apparatus which is now an integral part of the Electronic Studio of the University of Toronto. The Toronto Studio is fortunate to have acquired from the National Research Council of Canada the multitrack tape recorder with a six channel touch-sensitive keyboard.[3] This provides twelve channels for continuous independent sound sources, which may be varied in pitch and volume. Through the multi-track recorder, melodic units can be created at will with any previously selected timbre and performed by the operator. By using a metronome, various rhythmic patterns can be synchronized into a contrapuntal texture, but only with tremendous care in measuring and splicing the required lengths of recorded tape. The growth and decay envelope of individual tones can be varied at will through the touch-sensitive keyboard, but here again the human element enters into the picture, and it is practically impossible to repeat exactly the same intensity pattern and to perform with absolute rhythmic accuracy the units which the composer finally wishes to combine and synchronize.

The basic idea in designing the Hamograph was to provide, in the simplest possible form, a means of:

1) Producing and repeating rhythmic patterns of any complexity without the need for splicing or metronomic control.
2) Combining and repeating numerous and distinct rhythmic patterns of any complexity without the need for re-recording and synchronizing spliced montages.
3) Repeating exactly any chosen growth and decay envelope pattern for the same or different individual tone timbre.
4) Repeating a calculated proportionate distribution of sound objects, melodic units or individual sounds with varying timbres, or programming a portion of or a total composition.
5) Increasing or decreasing the tempo of pre-determined rhythmic organizations without the alteration of pitch in the process.
6) Modifying unit elements of the control pattern at will without the need for reconstructing the total control system.

The Olson music synthesizer,[4] at present in the New York studios of the Columbia-Princeton Electronic Music Project, offers an ample solution to the majority of these problems through an elaborate and highly ingenious sound generator and control system based on a perforated paper roll. The coded information required to produce a tone of specific timbre, pitch, duration and attack and decay pattern, is punched out on a continuous paper scroll and in turn is transmitted to the sound generator. One tone at a time is coded and when produced in series makes possible the sounding of single melodic lines. By combining various recorded continuities, any harmonic or contrapuntal texture can be synthesized.

The disadvantages of this system for most studios is the initial cost of the equipment and the tremendous amount of time required for coding of the composer's musical intentions.

The principle of the Hamograph is to provide a simple mechanical control of the voltage in the electronic stage of the signal, providing thus an absolute control of the envelope as well as a finite rhythmic organization in the aural stage. Conductive silver ink was first used with good results, since it is easy to apply and offers excellent conductivity. Greater accuracy and simplicity was finally obtained by using aluminum foil which allows for rapid correction or replacement of individual units. As shown in Fig. 1, these metal foil contours are made to adhere to 35-mm continuous loop carriers and are read by a series of silver brushes (Fig. 2) so arranged that the signal circuit is modified in intensity by means of a resistor chain, which, in turn, controls the volume of the final signal in the aural monitor or recording device. In the present model six control loops provide six independent channels with the resulting maximum texture of six simultaneous voices.

The electric current emanating from the resistor chain is channeled through a secondary bias circuit which by means of variable capacitors, makes possible a planned delay response in the reading of the control contours. A circuit diagram of this arrangement is shown in Fig. 3.

A variable speed motor directly coupled to the crankshaft of the Hamograph control-loop drive through a worm reduction gear makes possible a constant or variable speed of the control loops. Individual envelope patterns may be removed, added to, or otherwise modified by simply removing the original contour and substituting a new design.

Rhythmic patterns are determined by the linear placement of the control designs at measured intervals along the control loop. This takes the place of cutting and of splicing corresponding lengths of already recorded tape. By increasing or decreasing the speed of the motor, the tempo of the rhythmic patterns may be correspondingly altered.

The control designs provide a dynamic range from zero to minus twenty db at a standard of the bias-plate voltage, but may be altered to provide a greater or lesser over-all range. Each control channel is provided with an individual volume control to compensate for sensible dynamic differences of varying timbres and/or pitch ranges.

A signal, continuous tone, melodic fragment or polyphonic complex is fed into the input side of the Hamograph, modulated rhythmically and/or dynamically by

[3] Designed and build by Dr. Hugh LeCaine, Natl. Research Council, Ottawa, Ontario, Canada.

[4] H. F. Olson and H. Belar, "Electronic music synthesizer," *J. Acoust. Soc. Am.*, vol. 27, pp. 597–612; May, 1955.

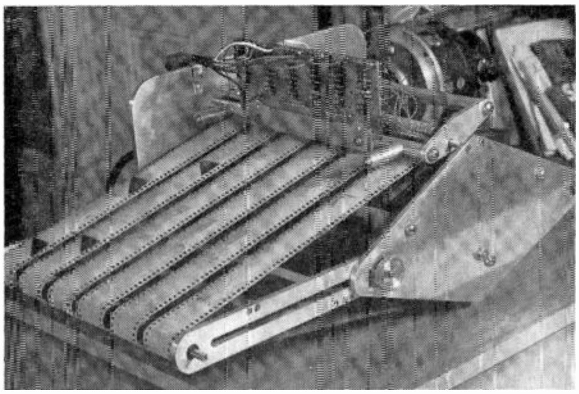


Fig. 1—The Hamograph showing placement of metallic contours on endless loops for dynamic control.

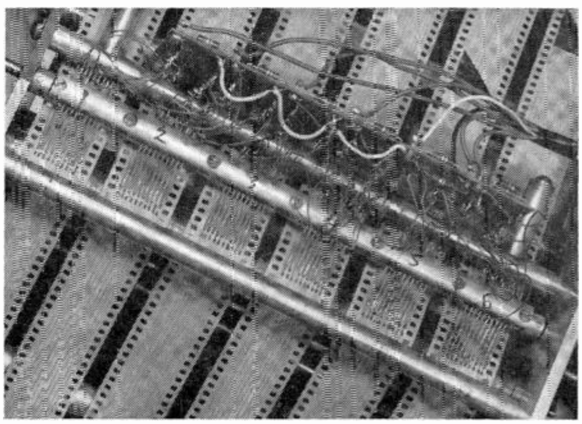


Fig. 2—Closeup of the silver brush system for reading the control contours.

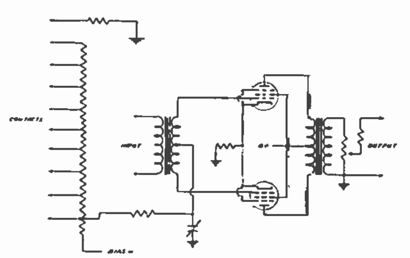


Fig. 3—Amplifier circuit diagram used in the Hamograph.

the control design as read by the silver brushes and resistor chain, and finally released through the capacitor control for monitoring or recording. This resulting continuity may be repeated as often as required for aural study, modified and perfected and finally recorded in the desired form without the need for numerous splices or trial recordings which are difficult to edit and more difficult to reproduce a second time.

A particularly interesting aspect of the Hamograph is the arrangement by which the control designs are placed which makes it possible to vary in a matter of seconds the growth and decay ratios, as well as rhythmic position of the individual sounds.

A total duration of each attack is determined by the length of the control design. Any constant unit may be used, but 1 cm has been found to be minimum for tempi not exceeding sixteenth notes at MM60 with maximum response delay. For faster tempi, larger attack time units may be used. The total length of the control loops may vary from approximately 80 cm to 120 cm for continuous operation, and from 80 cm to any required length for manual replacement and repetition. Thus, a rhythm of a dotted eighth, sixteenth, eighth, eighth, triplet of eighths and a quarter in a quadruple meter required control forms of 3 plus 1 plus 2 plus 2 plus 1.33 plus 1.33 plus 1.33 plus 4 cm in length. These ratios remain constant for any tempo and provide a means of altering the tempo without altering the pitch of the tones fed into the circuit.

If all the forms are placed on one control channel, rhythmic pattern of repeated tones may be produced, but by placing the control forms on various channels melodic contours can be produced by feeding different pitches into the respective channels of the Hamograph.

The dynamic envelope of each attack is controlled by the vertical dimension of the control design which varies from 2.5 cm maximum to 0.5 minimum. The graduation of the height of the control design in proportion to the length produces the attack envelope. This may vary from a momentary maximum with a gradual decay to mimimum, characteristic of a bell or piano tone, to a gradual growth from minimum with a sharply cut off maximum, characteristic of the sounds often heard produced by the brass instrument of the orchestra.

By varying the growth and decay ratios in relation to the length of the sound, the attack patterns of known musical instruments may be simulated and also interesting effects which are new and not associated with any known musical instruments can be achieved.

These variations mentioned can be made by simply substituting previously prepared envelope designs which can be interchanged as quickly as a stamp can be attached to a letter. The adhesive used is non-drying, thus as many changes may be made as required without delay for applying adhesive or allowing it to dry.

Once an interesting Hamograph score has been established, the composer is in a position to instrumentate with sounds of different timbres which are fed through the control channels and recorded for future use, modified for immediate use or simply studied and accepted.

A larger model of the Hamograph with twenty-four control channels is to be built during the coming year. It is expected that with this more complete design, preplanned melodic and harmonic continuities of greater complexity will be possible as well as more subtle variations in natural or synthetic tone spectra.

No claim is made that the Hamograph accomplishes its mission more completely than several other known methods. It does, however, represent a great saving in time, and greater simplicity and flexibility in the coding stage of mechanically controlled musical continuities as compared with other devices at present known to the author, who has prepared a number of musical compositions on the Hamograph for demonstration purposes.

Correspondence

Ferrites for Audio and Video Contact Recording Heads*

In the past, ferrites have been used only as a substitute for laminated metal heads in noncontact-type applications such as magnetic-drum-storage systems for computers. This limitation on use was imposed first because the ferrite composition was, until recently, of such a nature that a smooth surface, free of irregularities, could not be obtained. Secondly, its hard composition, together with surface roughness, would deteriorate the tape. Third, because of the mechanical property of ferrite, hard, and brittle, a crumbling effect or erosion of the gap edge was produced as the head was lapfinished in manufacturing and as the tape moved across the head in use. This erosion produced a deterioration of gap definition.

The fact that ferrites offer so many advantages over laminated heads inspired considerable development effort on the part of many experts in the field. Recently, after many years of research and experimentation, a suitable material and bonding technique has been obtained which eliminates prior difficulties and enables the ferrite to be used in contact recording head applications. Long recognized advantages of ferrite over metal heads include: a) lower power losses, b) reduced eddy current flow which gives both a more even flux distribution across the gap region and no deterioration of the field in the neighborhood of the gap. This improves the performance and efficiency of the head, especially in recording heads which are used at high flux levels, c) low intrinsic head noise producing a better SNR ratio on reproduction because of the low head losses, d) ease of construction, e) more constant permeability as a function of frequency, and f) higher permeability in the high-frequency range. Along with these advantages, this newly developed material and construction technique for the manufacture of ferrite contact heads has produced additional merits:

1) Extremely high durability, giveng almost infinite life for slow speed audio recording.
2) Elimination of any cold working effect (the loss of the magnetic property of the material in the vicinity of the gap edge as a result of lapping the faces of the pole pieces) and the associated loss of gap definition.
3) Extremely small gaps (<1 micron).
4) Better SNR ratio on reproduction over metallic heads because the magnetic properties of the high-density skin at the contact side of the head resulting from the polishing process are not different from the bulk material.
5) An extremely consistent high quality product.
6) Considerable reduction in construction and assembling cost of the complete head.

* Received by the PGA, September 11, 1961.

TABLE I

Physical Properties of Transducer Head Materials

Property \ Material	Fxc IV A2 (1)	Mumetal (2)	Permalloy 4-79 (2) (3)	Vacodur 16 (2) (4)	Dimension
Composition	NiO 17.5 ZnO 32.5 Fe_2O_3 50.0	Ni 77 Cu 5 Cr 2	Ni 79 Mo 4	Al 16	for ferrite: mol %; for alloys: % by weight, balance Fe
Initial permeability	0.85	30	25	5 a) 7 b)	$\times 10^3$ at 0.4 A/m = 5 m Oe
Maximum permeability	2.5	70	60	30 a) 60 (b)	$\times 10^3$
Coercive force	32	2.4	3.2	2 (a) 4.8 (b)	A/m = 0.0125 Oe
Saturation induction	0.39	0.8	0.9	0.9	$Wb/m^2 = 10^4$ G
Specific resistance	$>10^8$	0.50	0.55	1.45	$\mu\Omega m = \Omega mm^2/m$
Curie temperature	125	400	460	350	°C
Relative hardness	600	118	132	220–280 (a) 350 (b)	Vickers HV 5

(1) tentative data
(2) data supplied by Vacuumschmelze AG, Hanau a.M., Germany
(3) known in U.S.A. under other trade names such as: Hy Mu 80, Mo-Permalloy, Permalloy C
(4) this material is similar to the Alfenol material described in Electronics, June, 1955.
(a) cold-rolled
(b) hot-rolled

Material Properties

Electrical

The new ferrite material used for this application is Ferroxcube 4A2. Its magnetic properties are listed in Table I, along with corresponding data on some more popular type metal materials, applied in or recommended for head construction. The most significant difference between Ferroxcube 4A2 and the alloy materials is its value of the specific resistance. This extremely high value of resistivity accounts for the reduction in eddy currents in ferrite and the consequent improvement of operation as mentioned above. The price we pay for the improvement in the lower permeability and lower saturation density of the ferrite. It is in general desirable, particularly for reproducing heads, to have a high permeability material. It should be pointed out, however, that the alloy values in Table I are valid only for well-annealed laminated toroidal samples completely free from stress and strains. In practice, if heads are constructed of this metallic material, much lower values of permeability are realized. In addition, the cold working effect of metal further deteriorates the magnetic properties of the lamination in the most important region—at the gap.

Another significant consideration to be given such technical information as published in Table I is that usually data of this type for alloys refer to measurements made at 60 cps. The permeability decreases with frequency much more rapidly for alloys than it does for ferrites. Ferroxcube 4A2 has a constant self-inductance together with a high constant Q over a wide range of frequencies. Fig. 1 shows how the real and the imaginary part of the permeability of Ferroxcube 4A2 varies with frequency.

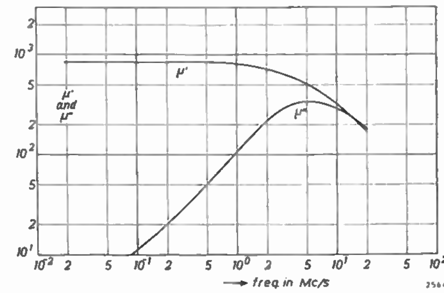


Fig. 1—Real part (μ') and imaginary part (μ'') of initial permeability, together with loss factor (tan δ)/μ', plotted as a function of frequency for Ferroxcube 4A2. Note the differences in vertical scale.

We can obtain a better insight on the usefulness of ferrites in the HF range by examining the loss factor tan δ/μ. For a laminated head with lamination thickness of 0.004 in (the smallest commercially available thickness which can reasonably be handled for head construction) with an effective permeability equal to that of Ferroxcube 4A2 and a specific resistance corresponding to Ni-Fe alloys, the loss factor tan δ/μ is equal to 1 at a frequency somewhat below 100 kc. As the effective permeability of the laminated metal head is increased, the limiting frequency at which tan δ/μ will have a value of 1 decreases. On the other hand, for Ferroxcube 4A2, tan δ/μ is smaller than 10^{-3} at frequencies up to 4 Mc/s.

With regard to the low saturation induction of ferrite, which is important for recording heads because saturation effects at the gap edges could give rise to field deterioration particularly for smaller gap lengths, experiments with Ferroxcube 4A2 recording heads gave no evidence of saturation effects. It appears, therefore, the low value of saturation induction is not a serious handicap.

Mechanical

Unlike other ferrite applications, mechanical properties of ferrite used for recording head applications are of utmost importance. Well-defined sharp gap edges producing a high resolving power can only be produced if a material has a uniform fine grain structure. Ferroxcube 4A2 meets this requirement with a grain size varying from 10 to 30 microns. In addition to the fine grain structure, ferrites used for recording heads must have a homogenous composition. Nonhomogenous materials will give rise to the presence of voids, inclusions and pores. These will not only effect the gap edges in heads, but also make it impossible to obtain perfectly polished surfaces. Holes in the contact surface of the head act as a site from which erosion begins. This erosion, being a cumulative effect, is detrimental to the head performance as soon as it attacks the gap edges. Ferroxcube 4A2 was therefore developed with as small a value of porosity as practical, with pores as small as possible and uniformly dispersed. Ferroxcube 4A2 has a porosity of 3.5 per cent of the roentgen density, which is 5.33. This value of 3.5 per cent is for the bulk material and is decreased enormously at the polished surface to a value of about 0.2 per cent. The polishing technique developed for this material yields an ultra-thin, extremely hard and wear-resistant skin. The photomicrographs shown in Fig. 2 illustrate surface conditions of the ferrite before and after polishing. The average size of the pores for Ferroxcube 4A2 is less than 2 microns.

Although the hardness of ferrite presents many difficulties in the fabrication, such as in cutting, grinding, lapping and polishing, it is obvious that this is a desirable property in terms of good resistance to wear. Table I shows the relative hardness of Ferroxcube 4A2 to be definitely superior to that of the alloy materials.

Glass Bonding

Although the material development eliminated many of the previous problems associated with the use of ferrite in contact recording, subsequent tests indicated that complete erosion effect could not be prevented unless some mechanical support could be added to the gap edges in the construction of the completed head. The main reason for the continued difficulty was thought to lie in the fact that in the construction, the two halves of the head are necessarily pressed together rather tightly in order to keep the spacer between them in position. Consequently, strong forces are exerted at the gap faces and edges and chipping occurs initiating the erosion process. The required mechanical support of the edges of the gap was obtained by employing a special bonding-glass spacer technique consisting of sandwiching a shim of glass, slightly thicker than the gap length ultimately desired, between the polished mating surfaces of the two halves of the head. A complete core and shim are then subjected to a process which effectively bonds the two core pieces together. Tests on heads constructed of cores bonded together in this fashion showed no signs of erosion. Fig. 3 illustrates, by photomicrographs, gap lengths of about 7.5 and 1.8 microns suitable for audio recording heads and a gap length of about 0.7 micron, suitable for a video recording head.

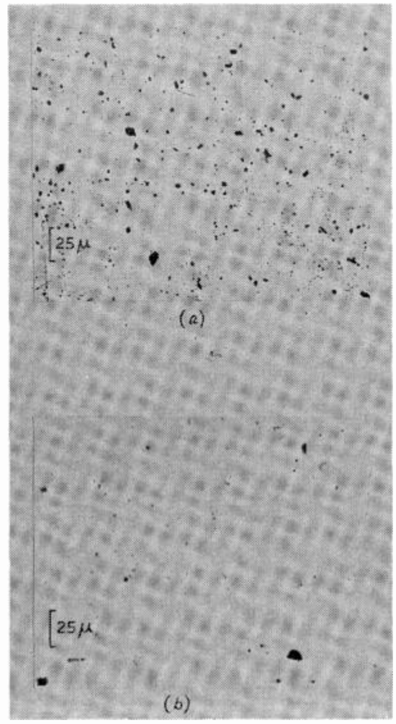


Fig. 2—Photomicrographs showing surface of Ferroxcube 4A2. (a) After lapping and before polishing. (b) After polishing.

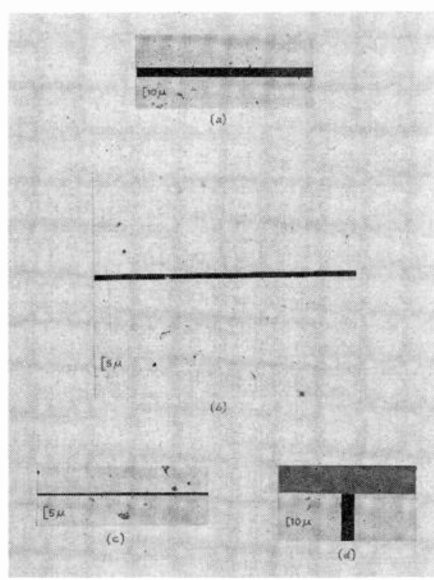


Fig. 3—Photomicrographs of high-resolution recording heads made with Ferroxcube 4A2 material. (a) Audio-reproducing head with 7.5 micron gap. (b) Audio-reproducing head with 1.8 micron gap. (c) Video head with approximately 0.7 micron gap. (d) Detail of a head which is enameled at the sides for protection and containing a 12 micron gap.

Compared optically, the glass-gapped heads are highly superior to the gaps obtained in metallic heads. In metallic heads there is always some irregularity as the result of the laminated structure and the smearing out of both the spacer and the head material during lapping. Irregularities in the gap contribute to lower SNR ratio, especially when separate heads with, in general, incongruent gap irregularities are used for both recording and reproducing.

Actual Performance

The actual behavior of a head, particularly for reproducing, is not governed by the mechanical quality of the gap as optically observed, but rather by the effective gap length which determines the resolving power, that is, the shortest wavelength that can be reproduced. In this respect, the ferrite heads with bonding glass faces are superior to metallic heads. As previously mentioned, the effective permeability of laminated nickel-iron alloys appreciably diminishes as the result of the assembly operation. This may not be too serious a problem, since the permeability still remains at a high value. However, what is much more serious, owing to the lapping of the gap faces and of the contact surface, is that these surfaces become work hardened. A skin results which has a depth of about 1.5 microns and in which the permeability gradually drops to a low value and the magnetic properties almost vanish. This effect, which cannot be eliminated by subsequent annealing of the completed head, sets a limit to the resolving power that can be obtained with metallic heads. Reproduced wavelengths smaller than about 3 microns are impossible to achieve. Furthermore, the nonmagnetic skin at the contact surface produces an inherent spacing between the head and tape which is detrimental to performance and reduces the output of reproduce heads.

For metallic heads with nonlaminated pole pieces, we may expect, on account of the aforementioned factors, that the effective gap is only somewhat larger than the optically observed one and with about the same irregularities. For laminated heads, however, the situation is worse since for each lamination the thickness of the working skin will increase from the center of the lamination to the nonmagnetic insulating interlamination layer. These heads, therefore, even if their optically observed gap is satisfactory, will have an effective gap of a herringbone structure of which the periodicity is determined by the pitch of the laminations. For high resolution heads with gap lengths on the order of a few microns, these irregularities will be at least of the same order of magnitude.

For heads made of Ferroxcube 4A2, no such working skin could be observed. The optically-estimated gap length corresponded very accurately to the effective gap length which was determined experimentally by recording a range of frequencies at various tape speeds and observing the first zero in the response. The shortest gap length that can be determined in this way is about 2 microns. This is about the shortest wavelength to be recorded on tapes for which the mean length of the iron oxide particles is of the order of 1 micron.

Deductions from these measurements indicate that apparently the working skin of the gap faces, if any initially existed, is completely nullified during the bonding process. Furthermore, no evidence has been found that the magnetic properties of the extremely dense skin at the contact side of the head resulting from the polishing process, are different from those of the bulk material. From this, it is evident that a better SNR ratio may be expected for ferrite reproducing heads than for metallic heads.

Wear Characteristics

Transducer heads, which are subject to wear, must be replaced from time to time. However, what is perhaps more serious, especially in continuous duty professional equipment, is that during the wearing away of the head its electrical characteristics are continually changing.

For recording heads, a decrease of gap height results in a decrease of the HF biasing field which is superimposed on the signal to be recorded in order to secure linearity. In addition, the head sensitivity, *i.e.*, the strength of the LF signal necessary for full magnetization over the entire depth of the magnetic layer at a given frequency and a given bias, increases and the pre-emphasis to be applied to the signal for correction of the over-all frequency characteristic also increases. If, at the same time, the wear increases the gap width or deteriorates the sharpness of the gap edges, then the resolving power or gap definition of the head also decreases since the latter is largely determined by the gradient of the stray flux in front of the gap. The shorter and better defined the gap is, the larger the stray flux gradient will be.

For reproducing heads, the sensitivity, *i.e.*, the level of the output voltage resulting from a signal of some specific wavelength recorded on a reference tape which possesses a specific surface flux density, becomes larger as the gap height decreases. Since the gap length must always be small compared to the shortest wavelength to be reproduced, the crumbling or the rounding off of the gap edges has a detrimental effect on the resolving power of the head.

These factors imply that a frequent readjustment of the electrical equipment is necessary in order to maintain a certain standard of quality for the reproduced signals. It is hardly necessary to say that an improvement in the durability of transducer heads is of great importance. The problem of head wear is extremely complicated since a large number of factors are involved. These include such things as surface smoothness and hardness of the head and tape material, contact pressure, wrap angle, and tape speed. These factors may differ considerably from one application to another. It is not possible, therefore, to extrapolate data on head wear obtained at high tape speeds to lower tape speeds. Extremely time-consuming experiments are required to obtain exact data on the life of heads for each particular field of application. To make general statements about the durability of transducer heads is almost impossible.

Many laboratory tests have been made to obtain data on the wear characteristics of ferrite heads containing bonding-glass spacers in various applications during a period of several years. Results of these tests indicate that these ferrite heads are highly superior to metallic heads. The results of two of the tests performed follow.

In one experiment, a Mu-metal head and a ferrite head, both having a gap length of 7 microns and similar characteristics, were tested simultaneously in a standard professional recorder for 500 hours at a tape speed of 15 in per second. After this period of time, the metal head was completely worn down, whereas the ferrite head showed no electrically measurable evidence of wear. In another experiment concerning video recording, heads with extremely narrow track width (150 microns) were tested at a relative tape speed of about 100 ft per second and at still higher re-wind speeds. After 50 hours (corresponding to the passing of approximately 18 million feet of tape) the output of the head started to fall off. This was not as a result of gap erosion but because of the track width becoming narrower due to a crumbling off of the sides of the head. These heads were initially imbedded in a casting resin and gradually became unsupported because the resin wore away. A much improved lifetime of the head can be expected by adequately supporting the sides of the head. This can be done by dipping the top of the head, which contains the glass gap of the correct length and which has already been polished down to the required curvature, gap height and track width, in molten enamel. The contact surface is then reprocessed by grinding, lapping and polishing to its original state. The photomicrograph shown in Fig. 3 illustrates such an imbedded head (note: the one shown is not a video head).

As a tentative and conservative conclusion, it is estimated from these and other experiments that the life of ferrite heads with bonding glass spacers for professional applications at medium tape speeds ($7\frac{1}{2}$ to 15 in/sec) will be at least several thousand hours. This is equivalent to the passing of several millions of feet of tape. For video applications at extremely high relative tape speeds, it appears that the handling of a total length of tape of at least 650 million feet is possible. This compares favorably with the figures quoted for other types of video heads made with Alfenol 16 or High-Mu 80 pole tips. For audio entertainment equipment with its low tape speed of $1\frac{7}{8}$ in/sec and less and with its less stringent electrical requirements, it is expected that the life of the head will be practically infinite.

From the foregoing, it should not be concluded that ferrite heads will have a longer life under all operating conditions. There are certain exposures that have a detrimental effect on ferrites, for which metallic alloys are perhaps much less sensitive. In discussing the life expectancy of heads in various fields and applications, it is assumed the heads will be subjected to normal operating conditions in which care will be taken to avoid circumstances that are particularly harmful to ferrite heads. An example in which ferrite heads might give less satisfactory results is in the operation in conjunction with sound tracks on movie films. These are much different than the ordinary magnetic tape application. Here a much higher contact pressure is necessary in order to obtain the necessary wrap angle. It should be noted that once the polished contact surface is damaged by scratching or by chipping, nothing can be done to prevent it from wearing down rather quickly.

Acknowledge

The author is indebted, and wishes to express his thanks to, N. V. Philips Gloeilampenfabrieken, Eindhoven, The Netherlands, and Dr. S. Duinker, for permission to use material and illustrations originally published in the *Philips Research Reports*, vol. 15; August, 1960.

C. J. Kunz, Jr.
Chief Engineer
Ferroxcube Corporation of America
Saugerties, N. Y

Contributors

Vincent Salmon (SM'46), for a photograph and biography, please see page 59 of the March/April, 1961, issue of these TRANSACTIONS.

Harry F. Olson (A'37-VA'39-SM'48-F'49) was born in Mt. Pleasant, Iowa on December 28, 1902. He received the B.S. degree in 1924, the M.S. degree in 1925, the Ph.D. degree in 1928 and the E.E. degree in 1932, all from the University of Iowa, Iowa City. He received an Honorary D.Sc. degree from Iowa Wesleyan College, Mount Pleasant, in 1959.

He has been affiliated with the Research Department of Radio Corporation of America, the Engineering Department of RCA Photophone, the Research Division of RCA Manufacturing Company and RCA Laboratories. Currently he is Director of the Acoustical and Electromechanical Research Laboratory of the RCA Laboratories.

He is Past President of the Audio Engineering Society, Past President of the Acoustical Society of America, and Past Chairman of the Administrative Committee of the IRE Professional Group on Audio.

He has received the folowing honors: the Modern Pioneer Award of the National Association of Manufacturers, the John H. Potts Medal of the Audio Engineering Society, the Samuel L. Warner Medal of the Society of Motion Picture and Television Engineers, the John Scott Medal of the City of Philadelphia and the Achievement Award of the IRE Professional Group on Audio.

Dr. Olson is a member of Tau Beta Pi, Sigma Xi, and the National Academy of Sciences. He is a fellow of the Society of Motion Picture and Television Engineers, the American Physical Society, and the Acoustical Society of America, and an honorary member of the Audio Engineering Society.

Benjamin B. Bauer (S'37–A'39–SM'44–F'53), for a photograph and biography please see page 181 of the September/October, 1961, issue of these TRANSACTIONS.

Myron Schaeffer was born in Barberton, Ohio, on November 11, 1908. He received the Mus.B. degree from Oberlin Conservatory of Music, Oberlin, Ohio, in 1930 and the A.B. degree from Oberlin College in 1931. In 1937 he received the Ph.D. degree in Musicology from Western Reserve University, Cleveland, Ohio.

He spent three years doing musicological research in Europe on grants from the Western Reserve University and the Belgian-American Educational Foundation, respectively. From 1940 to 1941 he taught musicology at the Graduate School, Columbia University, New York, N. Y. From 1941 to 1945 he was Dean of the College of Arts and Sciences at the University of Panama, Panama City, Republic of Panama, and from 1945 to 1958 he was engaged in lecturing and research in the field of folk-music throughout Latin-America. In 1958 he came to the faculty of music, University of Toronto, Toronto, Ontario, Can., to teach musicology and to develop the Electronic Music Studio of the University.

Dr. Schaeffer has applied for membership in the Acoustical Society of America.